OUTLIERS IN THE SPECTRUM FOR PRODUCTS OF INDEPENDENT RANDOM MATRICES

NATALIE COSTON, SEAN O'ROURKE, AND PHILIP MATCHETT WOOD

ABSTRACT. For fixed $m \geq 1$, we consider the product of m independent $n \times n$ random matrices with iid entries as $n \to \infty$. Under suitable assumptions on the entries of each matrix, it is known that the limiting empirical distribution of the eigenvalues is described by the m-th power of the circular law. Moreover, this same limiting distribution continues to hold if each iid random matrix is additively perturbed by a bounded rank deterministic error. However, the bounded rank perturbations may create one or more outlier eigenvalues. We describe the asymptotic location of the outlier eigenvalues, which extends a result of Tao [60] for the case of a single iid matrix. Our methods also allow us to consider several other types of perturbations, including multiplicative perturbations.

CONTENTS

1

1. Introduction

This paper is concerned with the asymptotic behavior of outliers in the spectrum of bounded-rank perturbations of large random matrices. We begin by fixing the following notation and introducing an ensemble of random matrices with independent entries.

The *eigenvalues* of an $n \times n$ matrix M_n are the roots in $\mathbb{C}$ of the characteristic polynomial $\det(M_n - zI)$, where I is the identity matrix. We let $\lambda_1(M_n), \ldots, \lambda_n(M_n)$ denote the eigenvalues of M_n counted with (algebraic) multiplicity. The *empirical spectral measure* μ_{M_n} of M_n is given by

$$\mu_{M_n} := \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(M_n)}.$$

If M_n is a random $n \times n$ matrix, then μ_{M_n} is also random. In this case, we say μ_{M_n} converges weakly in probability (resp. weakly almost surely) to another Borel probability measure μ on the complex plane $\mathbb{C}$ if, for every bounded and continuous function $f : \mathbb{C} \to \mathbb{C}$,

$$\int_{\mathbb{C}} f \, d\mu_{M_n} \longrightarrow \int_{\mathbb{C}} f \, d\mu$$

in probability (resp. almost surely) as $n \to \infty$.

Throughout the paper, we use asymptotic notation (such as O, o) under the assumption that $n \to \infty$; see Section 4.1 for a complete description of our asymptotic notation.

1.1. iid random matrices.
In this paper, we focus on random matrices whose entries are independent and identically distributed.

Definition 1.1 (iid random matrix)**.** Let ξ be a complex-valued random variable. We say X_n is an $n \times n$ *iid random matrix* with atom variable ξ if X_n is an $n \times n$ matrix whose entries are independent and identically distributed (iid) copies of ξ.
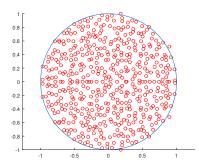
The circular law describes the limiting distribution of the eigenvalues of an iid random matrix. For any matrix M, we denote the Hilbert-Schmidt norm $\|M\|_2$ by the formula

$$\|M\|_2 := \sqrt{\operatorname{tr}(MM^*)} = \sqrt{\operatorname{tr}(M^*M)}. \tag{1}$$

Theorem 1.2 (Circular law; Corollary 1.12 from [61])**.** *Let ξ be a complex-valued random variable with mean zero and unit variance. For each $n \geq 1$, let X_n be an $n \times n$ iid random matrix with atom variable ξ, and let A_n be a deterministic $n \times n$ matrix. If $\operatorname{rank}(A_n) = o(n)$ and $\sup_{n \geq 1} \frac{1}{n}\|A_n\|_2^2 < \infty$, then the empirical measure $\mu_{\frac{1}{\sqrt{n}}X_n + A_n}$ of $\frac{1}{\sqrt{n}}X_n + A_n$ converges weakly almost surely to the uniform probability measure on the unit disk centered at the origin in the complex plane as $n \to \infty$.*

This result appears as [61, Corollary 1.12], but is the culmination of work by many authors. We refer the interested reader to the excellent survey [28] for further details.

From Theorem 1.2, we see that the low-rank perturbation A_n does not affect the limiting spectral measure (i.e., the limiting measure is the same as the case when $A_n = 0$). However, the perturbation A_n may create one or more outliers. An example of this phenomenon is illustrated in Figure 1.
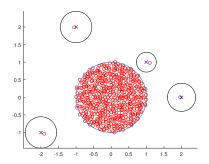
FIGURE 1. On the left, we have plotted the eigenvalues of a 500×500 random matrix with iid standard Gaussian entries scaled by $\frac{1}{\sqrt{500}}$. Additionally, the unit circle is plotted for reference. The image on the right contains the eigenvalues of $\frac{1}{\sqrt{500}}X + A$, where X is a 500×500 random matrix with iid symmetric ± 1 Bernoulli entries and $A = \operatorname{diag}(1 + i, 2i - 1, 2, -i - 2, 0, \ldots, 0)$. For reference, we have also plotted each nonzero eigenvalue of A with a cross.

Recall that the spectral radius of a square matrix M is the largest eigenvalue of M in absolute value. Among other things, Theorem 1.2 implies that with probability tending to one, the spectral radius of $\frac{1}{\sqrt{n}}X_n$ is at least $1 - o(1)$. When the atom variable ξ has finite fourth moment, it is possible to improve the lower bound on the spectral radius and give a matching upper bound.

Theorem 1.3 (No outliers for iid matrices; Theorem 5.18 from [15]). *Let ξ be a complex-valued random variable with mean zero, unit variance, and finite fourth moment. For each $n \geq 1$, let X_n be an iid random matrix with atom variable ξ. Then the spectral radius of $\frac{1}{\sqrt{n}}X_n$ converges to 1 almost surely as $n \to \infty$.*

Remark 1.4. It is conjectured in [27] that the spectral radius of $\frac{1}{\sqrt{n}}X_n$ converges to 1 in probability as $n \to \infty$ only assuming that ξ has mean zero and unit variance.

Theorem 1.3 asserts that almost surely all eigenvalues of $\frac{1}{\sqrt{n}}X_n$ are contained in the disk of radius $1 + o(1)$ centered at the origin. However, as we saw in Figure 1, this need not be the case for the eigenvalues of the additive perturbation $\frac{1}{\sqrt{n}}X_n + A_n$. In this case, it is possible for eigenvalues of the perturbation to be larger than $1 + o(1)$, and Tao precisely describes the location of these outlying eigenvalues in [60].

Theorem 1.5 (Outliers for small low-rank perturbations of iid matrices; Theorem 1.7 from [60]). *Let ξ be a complex random variable with mean zero, unit variance, and finite fourth moment. For each $n \geq 1$, let X_n be a $n \times n$ random matrix whose entries are iid copies of ξ, and let A_n be a deterministic matrix with rank $O(1)$ and operator norm $O(1)$. Let $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no eigenvalues of A_n in the band $\{z \in \mathbb{C} : 1 + \varepsilon < |z| < 1 + 3\varepsilon\}$, and there are j eigenvalues $\lambda_1(A_n), \ldots, \lambda_j(A_n)$ for some $j = O(1)$ in the region $\{z \in \mathbb{C} : |z| \geq 1 + 3\varepsilon\}$. Then, almost surely, for sufficiently large n, there are precisely j eigenvalues $\lambda_1(\frac{1}{\sqrt{n}}X_n + A_n), \ldots, \lambda_j(\frac{1}{\sqrt{n}}X_n + A_n)$ of $\frac{1}{\sqrt{n}}X_n + A_n$ in the*

region $\{z \in \mathbb{C} : |z| \geq 1 + 2\varepsilon\}$, *and after labeling these eigenvalues properly,*

$$\lambda_i \left(\frac{1}{\sqrt{n}} X_n + A_n \right) = \lambda_i(A_n) + o(1)$$

as $n \to \infty$ *for each* $1 \leq i \leq j$.

Analogous results describing the location and fluctuation of the outlier eigenvalues have been obtained for many ensembles of random matrices; we refer the reader to [17, 16, 18, 19, 20, 21, 22, 23, 26, 29, 34, 35, 36, 37, 44, 45, 49, 52, 54, 55, 56, 57, 58, 60] and references therein. In particular, the results in [56] extend Theorem 1.5 by also describing the joint fluctuations of the outlier eigenvalues about their asymptotic locations.

1.2. **Products of independent iid matrices.** In this paper, we focus on the product of several independent iid matrices. In this case, the analogue of the circular law (Theorem 1.2) is the following result from [50], due to Renfrew, Soshnikov, Vu, and the second author of the current paper.

Theorem 1.6 (Theorem 2.4 from [50])**.** *Fix an integer* $m \geq 1$, *and let* $\tau > 0$. *Let* $\xi_1, \ldots, \xi_m$ *be real-valued random variables with mean zero, and assume, for each* $1 \leq k \leq m$, *that* ξ_k *has nonzero variance* σ_k^2 *and satisfies* $\mathbb{E}|\xi_k|^{2+\tau} < \infty$. *For each* $n \geq 1$ *and* $1 \leq k \leq m$, *let* $X_{n,k}$ *be an* $n \times n$ *iid random matrix with atom variable* ξ_k, *and let* $A_{n,k}$ *be a deterministic* $n \times n$ *matrix. Assume* $X_{n,1}, \ldots, X_{n,m}$ *are independent. If*

$$\max_{1 \leq k \leq m} \operatorname{rank}(A_{n,k}) = O(n^{1-\varepsilon}) \quad and \quad \sup_{n \geq 1} \max_{1 \leq k \leq m} \frac{1}{n} \|A_{n,k}\|_2^2 < \infty$$

for some fixed $\varepsilon > 0$, *then the empirical spectral measure* μ_{P_n} *of the product*[1]

$$P_n := \prod_{k=1}^{m} \left(\frac{1}{\sqrt{n}} X_{n,k} + A_{n,k} \right)$$

converges weakly almost surely to a (non-random) probability measure μ *as* $n \to \infty$. *Here, the probability measure* μ *is absolutely continuous with respect to Lebesgue measure on* $\mathbb{C}$ *with density*

$$f(z) := \begin{cases} \frac{1}{m\pi} \sigma^{-2/m} |z|^{\frac{2}{m}-2}, & if \ |z| \leq \sigma, \\ 0, & if \ |z| > \sigma, \end{cases} \tag{2}$$

where $\sigma := \sigma_1 \cdots \sigma_m$.

Remark 1.7. When $\sigma = 1$, the density in (2) is easily related to the circular law (Theorem 1.2). Indeed, in this case, f is the density of ψ^m, where ψ is a complex-valued random variable uniformly distributed on the unit disk centered at the origin in the complex plane.

---

[1]Here and in the sequel, we use Pi (product) notation for products of matrices. To avoid any ambiguity, if $M_1, \ldots, M_m$ are $n \times n$ matrices, we define the order of the product

$$\prod_{k=1}^{m} M_k := M_1 \cdots M_m.$$

In many cases, such as in Theorem 1.6, the order of matrices in the product is irrelevant by simply relabeling indices.

Theorem 1.6 is a special case of [50, Theorem 2.4]. Indeed, [50, Theorem 2.4] applies to so-called elliptic random matrices, which generalize iid matrices. Theorem 1.6 and the results in [50] are stated only for real random variables, but the proofs can be extended to the complex setting. Similar results have also been obtained in [25, 40, 51]. The Gaussian case was originally considered by Burda, Janik, and Waclaw [30]; see also [32]. We refer the reader to [3, 4, 5, 6, 7, 8, 31, 38, 39, 46, 59] and references therein for many other results concerning products of random matrices with Gaussian entries.

2. Main results

From Theorem 1.6, we see that the low-rank deterministic perturbations $A_{n,k}$ do not affect the limiting empirical spectral measure. However, as was the case in Theorem 1.2, the perturbations may create one or more outlier eigenvalues. The goal of this paper is to study the asymptotic behavior of these outlier eigenvalues. In view of Theorem 1.5, we will assume the atom variables $\xi_1, \ldots, \xi_m$ have finite fourth moment.

Assumption 2.1. The complex-valued random variables $\xi_1, \ldots, \xi_m$ are said to satisfy Assumption 2.1 if, for each $1 \leq k \leq m$,

- the real and imaginary parts of ξ_k are independent,
- ξ_k has mean zero and finite fourth moment, and
- ξ_k has nonzero variance σ_k^2.

We begin with the analogue of Theorem 1.3 for the product of m independent iid matrices.

Theorem 2.2 (No outliers for products of iid matrices). *Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables which satisfy Assumption 2.1. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Define the products*

$$P_n := n^{-m/2} X_{n,1} \cdots X_{n,m}$$

and $\sigma := \sigma_1 \cdots \sigma_m$. Then, almost surely, the spectral radius of P_n is bounded above by $\sigma + o(1)$ as $n \to \infty$. In particular, for any fixed $\varepsilon > 0$, almost surely, for n sufficiently large, all eigenvalues of P_n are contained in the disk $\{z \in \mathbb{C} : |z| < \sigma + \varepsilon\}$.

Remark 2.3. A version of Theorem 2.2 was proven by Nemish in [47] under the additional assumption that the atom variables $\xi_1, \ldots, \xi_m$ satisfy a sub-exponential decay condition. In particular, this condition implies that all moments of $\xi_1, \ldots, \xi_m$ are finite. Theorem 2.2 only requires the fourth moments of the atom variables to be finite.

Remark 2.4. In view of Remark 1.4, it is natural to also conjecture that the spectral radius of P_n is bounded above by $\sigma + o(1)$ in probability as $n \to \infty$ only assuming the atom variables $\xi_1, \ldots, \xi_m$ have mean zero and unit variance. Here, we need the result to hold almost surely, and hence require the atom variables have finite fourth moment.

In view of Theorem 1.6, it is natural to consider perturbations of the form

$$P_n := \prod_{k=1}^{m} \left(\frac{1}{\sqrt{n}} X_{n,k} + A_{n,k} \right).$$

However, there are many other types of perturbations one might consider, such as multiplicative perturbations

$$P_n := \frac{1}{\sqrt{n}}X_{n,1}(I + A_{n,1})\frac{1}{\sqrt{n}}X_{n,2}(I + A_{n,2})\cdots\frac{1}{\sqrt{n}}X_{n,m}(I + A_{n,m}) \qquad (3)$$

or perturbations of the form

$$P_n := n^{-m/2}\prod_{k=1}^{m}X_{n,k} + A_n.$$

In any of these cases, the product P_n can be written as

$$P_n = n^{-m/2}X_{n,1}\cdots X_{n,m} + M_n + A_n,$$

where A_n is deterministic and M_n represents the "mixed" terms, each containing at least one random factor and one deterministic factor. Our main results below show that only the deterministic term A_n determines the location of the outliers. The "mixed" terms M_n do not effect the asymptotic location of the outliers.

This phenomenon is most easily observed in the case of multiplicative perturbations (3), for which there is no deterministic term (i.e., $A_n = 0$ and the perturbation consists entirely of "mixed" terms). In this case, the heuristic above suggests that there should be no outliers, and this is the content of the following theorem.

Theorem 2.5 (No outliers for products of iid matrices with multiplicative perturbations). *Let $m \geq 1$ be a fixed integer, and assume $\xi_1,\ldots,\xi_m$ are complex-valued random variables which satisfy Assumption 2.1. For each $n \geq 1$, let $X_{n,1},\ldots,X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1,\ldots,\xi_m$, respectively. In addition for any fixed integer $s \geq 1$, let $A_{n,1}, A_{n,2},\ldots, A_{n,s}$ be $n \times n$ deterministic matrices, each of which has rank $O(1)$ and operator norm $O(1)$. Define the product P_n to be the product of the terms*

$$\frac{1}{\sqrt{n}}X_{n,1},\ldots,\frac{1}{\sqrt{n}}X_{n,m},(I + A_{n,1}),\ldots,(I + A_{n,s})$$

in some fixed order. Then for any $\delta > 0$, almost surely, for sufficiently large n, P_n has no eigenvalues in the region $\{z \in \mathbb{C} : |z| > \sigma + \delta\}$ where $\sigma := \sigma_1\cdots\sigma_m$.

We now consider the case when there is a deterministic term and no "mixed" terms.

Theorem 2.6 (Outliers for small low-rank perturbations of product matrices). *Let $m \geq 1$ be a fixed integer, and assume $\xi_1,\ldots,\xi_m$ are complex-valued random variables which satisfy Assumption 2.1. For each $n \geq 1$, let $X_{n,1},\ldots,X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1,\ldots,\xi_m$, respectively. In addition, let A_n be an $n \times n$ deterministic matrix with rank $O(1)$ and operator norm $O(1)$. Define*

$$P_n := n^{-m/2}\prod_{k=1}^{m}X_{n,k} + A_n \qquad (4)$$

and $\sigma := \sigma_1\cdots\sigma_m$. Let $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no eigenvalues of A_n in the band $\{z \in \mathbb{C} : \sigma + \varepsilon < |z| < \sigma + 3\varepsilon\}$, and there are j eigenvalues $\lambda_1(A_n),\ldots,\lambda_j(A_n)$ for some $j = O(1)$ in the region $\{z \in \mathbb{C} : |z| \geq \sigma + 3\varepsilon\}$. Then, almost surely, for sufficiently large n, there are precisely j

eigenvalues $\lambda_1(P_n), \dots, \lambda_j(P_n)$ of P_n in the region $\{z \in \mathbb{C} : |z| \geq \sigma + 2\varepsilon\}$, and after labeling these eigenvalues properly,

$$\lambda_i(P_n) = \lambda_i(A_n) + o(1)$$

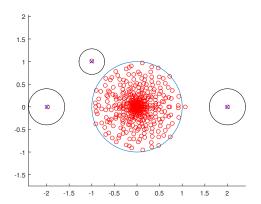*as $n \to \infty$ for each $1 \leq i \leq j$.*

FIGURE 2. In this figure, we have plotted the eigenvalues of $(500)^{-2} X_1 X_2 X_3 X_4 + A$, where $X_1, \dots, X_4$ are independent 500×500 iid random matrices with symmetric ± 1 Bernoulli entries and $A = \operatorname{diag}(-1 + i, -2, 2, 0, \dots, 0)$. The majority of the eigenvalues cluster inside the unit disc with the exception of three outliers. These outliers are close to the eigenvalues of A, each of which is marked with a cross.

Figure 2 presents a numerical simulation of Theorem 2.6. In the case that all the entries of A_n take the same value, the product P_n in (4) can be viewed as a product matrix whose entries have the same nonzero mean. Technically, Theorem 2.6 cannot be applied in this case, since such a matrix A_n does not have operator norm $O(1)$. However, using a similar proof, we establish the following result.

Theorem 2.7 (Outliers for a product matrix with nonzero mean)**.** *Let $m \geq 1$ be an integer, and let $\mu \in \mathbb{C}$ be nonzero. Assume $\xi_1, \dots, \xi_m$ are complex-valued random variables which satisfy Assumption 2.1. For each $n \geq 1$, let $X_{n,1}, \dots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \dots, \xi_m$, respectively. Let $\phi_n := \frac{1}{\sqrt{n}}(1, \dots, 1)^*$ and fix $\gamma > 0$. Define*

$$P_n := n^{-m/2} \prod_{k=1}^{m} X_{n,k} + \mu n^\gamma \phi_n \phi_n^*$$

and $\sigma := \sigma_1 \cdots \sigma_m$, and fix $\varepsilon > 0$. Then, almost surely, for n sufficiently large, all eigenvalues of P_n lie in the disk $\{z \in \mathbb{C} : |z| \leq \sigma + \varepsilon\}$ with a single exception taking the value $\mu n^\gamma + o(1)$.

Lastly, we consider the case of Theorem 1.6, where there are both "mixed" terms and a deterministic term.

Theorem 2.8. *Let $m \geq 1$ be an integer, and assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables which satisfy Assumption 2.1. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. In addition, for each $1 \leq k \leq m$, let $A_{n,k}$ be a deterministic $n \times n$ matrix with rank $O(1)$ and operator norm $O(1)$. Define the products*

$$P_n := \prod_{k=1}^{m} \left(\frac{1}{\sqrt{n}} X_{n,k} + A_{n,k} \right), \quad A_n := \prod_{k=1}^{m} A_{n,k}, \tag{5}$$

and $\sigma := \sigma_1 \cdots \sigma_m$. Let $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no eigenvalues of A_n in the band $\{z \in \mathbb{C} : \sigma + \varepsilon < |z| < \sigma + 3\varepsilon\}$, and there are j eigenvalues $\lambda_1(A_n), \ldots, \lambda_j(A_n)$ for some $j = O(1)$ in the region $\{z \in \mathbb{C} : |z| \geq \sigma + 3\varepsilon\}$. Then, almost surely, for sufficiently large n, there are precisely j eigenvalues $\lambda_1(P_n), \ldots, \lambda_j(P_n)$ of the product P_n in the region $\{z \in \mathbb{C} : |z| \geq \sigma + 2\varepsilon\}$, and after labeling these eigenvalues properly,

$$\lambda_i\left(P_n\right) = \lambda_i(A_n) + o(1)$$

as $n \to \infty$ for each $1 \leq i \leq j$.

Theorem 2.8 can be viewed as a generalization of Theorem 1.5. In fact, when $m = 1$, Theorem 2.8 is just a restatement of Theorem 1.5. However, the most interesting cases occur when $m \geq 2$. Indeed, in these cases, Theorem 2.8 implies that the outliers of P_n are asymptotically close to the outliers of the product A_n. Specifically, if even one of the deterministic matrices $A_{n,k}$ is zero, asymptotically, there cannot be any outliers for the product P_n. Figure 3 presents a numerical simulation of Theorem 2.8.
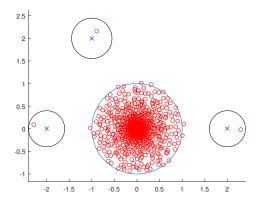
FIGURE 3. In the above figure, we display the eigenvalues of products of random matrices of the form $\prod_{k=1}^{5} \left(\frac{1}{\sqrt{1000}} X_k + A_k \right)$, where $X_1, \ldots, X_5$ are 1000×1000 independent iid matrices with symmetric ± 1 Bernoulli entries, and the product of the deterministic matrices $A_1, \ldots, A_5$ is $\mathrm{diag}(-2, -1 + 2i, 2, 0, \ldots, 0)$. Each nonzero eigenvalue of the product $A_1 \cdots A_5$ is marked with a cross.

2.1. **Outline.** The main results of this paper focus on iid matrices. However, one may ask if these results may be generalized to other matrix ensembles. These questions and others are discussed in Section 3. In Sections 4 and 5, we present some preliminary results and tools. In particular, Section 4 presents our notation and some standard linearization results that will be used throughout the paper. Section 5 contains a key eigenvalue criterion lemma and a brief overview of the proofs of the main results. In Section 6, we state an isotropic limit law and use it to prove our main results. The majority of the paper (Sections 7–11) is devoted to the proof of this isotropic limit law. Section 12 contains the proofs of the related results from Section 3. A few auxiliary results and proofs are presented in the appendices.

3. Related results, applications, and open questions

There are a number of related results which are similar to the main results of this paper, but do not directly follow from the theorems in Section 2. We discuss these results as well as some applications of our main results in this section.

3.1. **Related Results.** While our main results have focused on independent iid matrices, it is also possible to consider the case when the random matrices $X_{n,1}, \ldots, X_{n,m}$ are no longer independent. In particular, we consider the extreme case where $X_{n,1} = \cdots = X_{n,m}$ almost surely. In this case, we obtain the following results, which are analogs of the results from Section 2.

Theorem 3.1 (No outliers for multiplicative perturbations)**.** *Assume that ξ is a complex-valued random variable which satisfies Assumption 2.1 with $\sigma^2 := \mathrm{Var}(\xi)$. For each $n \geq 1$, let X_n be an $n \times n$ iid random matrix with atom variable ξ. In addition for any finite integer $s \geq 1$, let $A_{n,1}, A_{n,2}, \ldots, A_{n,s}$ be $n \times n$ deterministic matrices, each of which has rank $O(1)$ and operator norm $O(1)$. Define the product P_n to be the product of m copies of $\frac{1}{\sqrt{n}} X_n$ with the terms*

$$(I + A_{n,1}), (I + A_{n,2}), \ldots, (I + A_{n,s})$$

in some fixed order. Then for any $\delta > 0$, almost surely, for sufficiently large n, P_n has no eigenvalues in the region $\{z \in \mathbb{C} : |z| > \sigma^m + \delta\}$.

For a single additive perturbation, we have the following analog of Theorem 2.6.

Theorem 3.2 (Outliers for a single additive perturbation)**.** *Assume ξ is a complex-valued random variable which satisfies Assumption 2.1 with $\sigma^2 := \mathrm{Var}(\xi)$. For each $n \geq 1$, let X_n be an $n \times n$ iid random matrix with atom variable ξ. In addition, let A_n be an $n \times n$ deterministic matrix with rank $O(1)$ and operator norm $O(1)$. Define*

$$P_n := n^{-m/2} X_n^m + A_n. \tag{6}$$

Let $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no eigenvalues of A_n in the band $\{z \in \mathbb{C} : \sigma^m + \varepsilon < |z| < \sigma^m + 3\varepsilon\}$, and there are j eigenvalues $\lambda_1(A_n), \ldots, \lambda_j(A_n)$ for some $j = O(1)$ in the region $\{z \in \mathbb{C} : |z| \geq \sigma^m + 3\varepsilon\}$. Then, almost surely, for sufficiently large n, there are precisely j eigenvalues $\lambda_1(P_n), \ldots, \lambda_j(P_n)$ of P_n in the region $\{z \in \mathbb{C} : |z| \geq \sigma^m + 2\varepsilon\}$, and after labeling these eigenvalues properly,

$$\lambda_i(P_n) = \lambda_i(A_n) + o(1)$$

as $n \to \infty$ for each $1 \leq i \leq j$.

Note that Theorem 2.7 can also be generalized in an analogous way to Theorem 3.2 above.

Lastly, we have the following analog of Theorem 2.8.

Theorem 3.3. *Assume ξ is a complex-valued random variable which satisfies Assumption 2.1 with $\sigma^2 := \mathrm{Var}(\xi)$. For each $n \geq 1$, let X_n be an $n \times n$ iid random matrix with atom variable ξ. In addition, let $m \geq 1$ be an integer and for each $1 \leq k \leq m$, let $A_{n,k}$ be a deterministic $n \times n$ matrix with rank $O(1)$ and operator norm $O(1)$. Define the products*

$$P_n := \prod_{k=1}^{m} \left(\frac{1}{\sqrt{n}} X_n + A_{n,k} \right), \quad A_n := \prod_{k=1}^{m} A_{n,k}. \tag{7}$$

Let $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no eigenvalues of A_n in the band $\{z \in \mathbb{C} : \sigma^m + \varepsilon < |z| < \sigma^m + 3\varepsilon\}$, and there are j eigenvalues $\lambda_1(A_n), \ldots, \lambda_j(A_n)$ for some $j = O(1)$ in the region $\{z \in \mathbb{C} : |z| \geq \sigma^m + 3\varepsilon\}$. Then, almost surely, for sufficiently large n, there are precisely j eigenvalues $\lambda_1(P_n), \ldots, \lambda_j(P_n)$ of the product P_n in the region $\{z \in \mathbb{C} : |z| \geq \sigma^m + 2\varepsilon\}$, and after labeling these eigenvalues properly,

$$\lambda_i(P_n) = \lambda_i(A_n) + o(1)$$

as $n \to \infty$ for each $1 \leq i \leq j$.

The proofs of these results are presented in Section 12 and use similar techniques to the proofs of the main results from Section 2.

3.2. **Applications.**

Random matrices are useful tools in the study of many physically motivated systems, and we note here two potential applications for products of perturbed random matrices. First, iid Gaussian matrices can be used to model neural networks as in, for example, [1, 2, 11]. In the case of a linear version of the feed-forward networks in [2], the model becomes a perturbation of a product of iid random matrices, which, if the interactions in the model were fixed, could potentially be analyzed using approaches in the current paper. Second, one can conceive of a dynamical system (see, for example, [43]) evolving according to a matrix equation, which, when iterated, would lead a matrix product of the form discussed in Theorem 2.8.

3.3. **Open Questions.**

While our main results have focused on iid matrices, it is natural to ask if the same results can be extended to other matrix models. For example, Theorem 1.6 and the results in [50] also hold for products of so-called elliptic random matrices. However, the techniques used in this paper (in particular, the combinatorial techniques in Section 11) rely heavily on the independence of the entries of each matrix. It is an interesting question whether an alternative proof can be found for the case when the entries of each matrix are allowed to be dependent.

In this paper, we have focused on the asymptotic location of the outlier eigenvalues. One can also ask about the fluctuations of the outliers. For instance, in [56], the joint fluctuations of the outlier eigenvalues from Theorem 1.5 were studied. We plan to pursue this question elsewhere.

4. Preliminary tools and notation

This section is devoted to introducing some additional concepts and notation required for the proofs of our main results. In Section 5, we present a brief overview of our proofs and explain how these concepts will be used.

4.1. **Notation.** We use asymptotic notation (such as O, o, Ω) under the assumption that $n \to \infty$. In particular, $X = O(Y)$, $Y = \Omega(X)$, $X \ll Y$, and $Y \gg X$ denote the estimate $|X| \le CY$, for some constant $C > 0$ independent of n and for all $n \ge C$. If we need the constant C to depend on another constant, e.g. $C = C_k$, we indicate this with subscripts, e.g. $X = O_k(Y)$, $Y = \Omega_k(X)$, $X \ll_k Y$, and $Y \gg_k X$. We write $X = o(Y)$ if $|X| \le c(n)Y$ for some sequence $c(n)$ that goes to zero as $n \to \infty$. Specifically, $o(1)$ denotes a term which tends to zero as $n \to \infty$. If we need the sequence $c(n)$ to depend on another constant, e.g. $c(n) = c_k(n)$, we indicate this with subscripts, e.g. $X = o_k(Y)$.

Throughout the paper, we view m as a fixed integer. Thus, when using asymptotic notation, we will allow the implicit constants (and implicit rates of convergence) to depend on m without including m as a subscript (i.e. we will not write O_m or o_m).

An event E, which depends on n, is said to hold with *overwhelming probability* if $\mathbb{P}(E) \ge 1 - O_C(n^{-C})$ for every constant $C > 0$. We let $\mathbf{1}_E$ denote the indicator function of the event E, and we let E^c denote the complement of the event E. We write a.s. for almost surely.

For a matrix M, we let $\|M\|$ denote the spectral norm of M, and we let $\|M\|_2$ denote the Hilbert-Schmidt norm of M (defined in (1)). We denote the eigenvalues of an $n \times n$ matrix M by $\lambda_1(M), \ldots, \lambda_n(M)$, and we let $\rho(M) := \max\{|\lambda_1(M)|, \ldots, |\lambda_n(M)|\}$ denote its spectral radius. We let I_n denote the $n \times n$ identity matrix and 0_n denote the $n \times n$ zero matrix. Often we will just write I (or 0) for the identity matrix (alternatively, zero matrix) when the size can be deduced from context.

The singular values of an $n \times n$ matrix M_n are the non-negative square roots of the eigenvalues of the matrix $M_n^* M_n$ and we will denote their ordered values $s_1(M_n) \ge s_2(M_n) \ge \cdots \ge s_n(M_n)$.

We let C and K denote constants that are non-random and may take on different values from one appearance to the next. The notation K_p means that the constant K depends on another parameter p. We allow these constants to depend on the fixed integer m without explicitly denoting or mentioning this dependence. For a positive integer N, we let $[N]$ denote the discrete interval $\{1, \ldots, N\}$. For a finite set S, we let $|S|$ denote its cardinality. We let $\sqrt{-1}$ denote the imaginary unit and reserve i as an index.

4.2. **Linearization.** Let $M_1, \ldots, M_m$ be $n \times n$ matrices, and suppose we wish to study the product $M_1 \cdots M_m$. A useful trick is to linearize this product and instead consider the $mn \times mn$ block matrix

$$\mathcal{M} := \begin{bmatrix} 0 & M_1 & & & 0 \\ 0 & 0 & M_2 & & 0 \\ & & \ddots & \ddots & \\ 0 & & & 0 & M_{m-1} \\ M_m & & & & 0 \end{bmatrix}. \tag{8}$$

The following proposition relates the eigenvalues of $\mathcal{M}$ to the eigenvalues of the product $M_1 \cdots M_m$. We note that similar linearization tricks have been used previously; see, for example, [9, 30, 41, 50, 51] and references therein.

Proposition 4.1. *Let $M_1, \ldots, M_m$ be $n \times n$ matrices. Let $P := M_1 \cdots M_m$, and assume $\mathcal{M}$ is the $mn \times mn$ block matrix defined in (8). Then*

$$\det(\mathcal{M}^m - zI) = [\det(P - zI)]^m$$

for every $z \in \mathbb{C}$. In other words, the eigenvalues of $\mathcal{M}^m$ are the eigenvalues of P, each with multiplicity m.

Proof. A simple computation reveals that $\mathcal{M}^m$ is a block diagonal matrix of the form

$$\mathcal{M}^m = \begin{bmatrix} Z_1 & & 0 \\ & \ddots & \\ 0 & & Z_m \end{bmatrix},$$

where $Z_1 := P$ and

$$Z_k := M_k \cdots M_m M_1 \cdots M_{k-1}$$

for $1 < k \leq m$. Since each product $Z_2, \ldots, Z_m$ has the same characteristic polynomial[2] as P, it follows that

$$\det(\mathcal{M}^m - zI) = \prod_{k=1}^m \det(Z_k - zI) = [\det(P - zI)]^m$$

for all $z \in \mathbb{C}$. $\qquad\square$

We will exploit Proposition 4.1 many times in the coming proofs. For instance, in order to study the product $X_{n,1} \cdots X_{n,m}$, we will consider the $mn \times mn$ block matrix

$$\mathcal{Y}_n := \begin{bmatrix} 0 & X_{n,1} & & & 0 \\ 0 & 0 & X_{n,2} & & 0 \\ & & \ddots & \ddots & \\ 0 & & & 0 & X_{n,m-1} \\ X_{n,m} & & & & 0 \end{bmatrix} \tag{9}$$

and its resolvent

$$\mathcal{G}_n(z) := \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right)^{-1}, \tag{10}$$

defined for $z \in \mathbb{C}$ provided z is not an eigenvalue of $\frac{1}{\sqrt{n}} \mathcal{Y}_n$. We study the location of the eigenvalues of $\frac{1}{\sqrt{n}} \mathcal{Y}_n$ in Theorem 6.1 below.

Similarly, when we deal with the deterministic $n \times n$ matrices $A_{n,1}, \ldots, A_{n,m}$, it will be useful to consider the analogous $mn \times mn$ block matrix

$$\mathcal{A}_n := \begin{bmatrix} 0 & A_{n,1} & & & 0 \\ 0 & 0 & A_{n,2} & & 0 \\ & & \ddots & \ddots & \\ 0 & & & 0 & A_{n,m-1} \\ A_{n,m} & & & & 0 \end{bmatrix}. \tag{11}$$

---

[2]This fact can easily be deduced from Sylvester's determinant theorem; see (12).

4.3. **Matrix notation.** Here and in the sequel, we will deal with matrices of various sizes. The most common dimensions are $n \times n$ and $N \times N$, where we take $N := mn$. Unless otherwise noted, we denote $n \times n$ matrices by capital letters (such as M, X, A) and larger $N \times N$ matrices using calligraphic symbols (such as $\mathcal{M}, \mathcal{Y}, \mathcal{A}$).

If M is an $n \times n$ matrix and $1 \le i, j \le n$, we let M_{ij} and $M_{(i,j)}$ denote the (i,j)-entry of M. Similarly, if $\mathcal{M}$ is an $N \times N$ matrix, we let $\mathcal{M}_{ij}$ and $\mathcal{M}_{(i,j)}$ denote the (i,j)-entry of $\mathcal{M}$ for $1 \le i, j \le N$. However, in many instances, it is best to view $N \times N$ matrices as block matrices with $n \times n$ entries. To this end, we introduce the following notation. Let $\mathcal{M}$ be an $N \times N$ matrix. For $1 \le a, b \le m$, we let $\mathcal{M}^{[a,b]}$ denote the $n \times n$ matrix which is the (a,b)-block of $\mathcal{M}$. For convenience, we extend this notation to include the cases where $a = m + 1$ or $b = m + 1$ by taking the value $m + 1$ to mean 1 (i.e., modulo m). For instance, $\mathcal{M}^{[m+1,m]} = \mathcal{M}^{[1,m]}$. For $1 \le i, j \le n$, the notation $\mathcal{M}^{[a,b]}_{ij}$ or $\mathcal{M}^{[a,b]}_{(i,j)}$ denotes the (i,j)-entry of $\mathcal{M}^{[a,b]}$.

Sometimes we will deal with $n \times n$ matrices notated with a subscript such as M_n. In this case, for $1 \le i, j \le n$, we write $(M_n)_{ij}$ or $M_{n,(i,j)}$ to denote the (i,j)-entry of M_n. Similarly, if $\mathcal{M}_n$ is an $N \times N$ matrix, we write $\mathcal{M}^{[a,b]}_{n,(i,j)}$ to denote the (i,j)-entry of the block $\mathcal{M}^{[a,b]}_n$.

In the special case where we deal with a vector, the notation is the same, but only one index or block will be specified. In particular, if v is a vector in $\mathbb{C}^N$, then v_i denotes the i-th entry of v. If we consider v to be a block vector with m blocks of size n, then $v^{[1]}, \ldots, v^{[m]}$ denote these blocks, i.e., each $v^{[a]}$ is an n-vector, and

$$v = \begin{pmatrix} v^{[1]} \\ \vdots \\ v^{[m]} \end{pmatrix}.$$

In addition, $v^{[a]}_i$ denotes the i-th entry in block a.

4.4. **Singular Value Inequalities.** For an $n \times n$ matrix M, recall that $s_1(M) \ge \cdots \ge s_n(M)$ denote its ordered singular values. We will need the following elementary bound concerning the largest and smallest singular values.

Proposition 4.2. *Let M be an $n \times n$ matrix and assume $E \subseteq \mathbb{C}$ such that*

$$\inf_{z \in E} s_n(M - zI) \ge c$$

for some constant $c > 0$. Then

$$\sup_{z \in E} \|G(z)\| \le \frac{1}{c}$$

where $G(z) = (M - zI)^{-1}$.

Proof. First, observe that for any $z \in E$, we have that z is not an eigenvalue of M and so $M - zI$ is invertible and $G(z)$ exists. Recall that if s is a singular value of $M - zI$ and $M - zI$ is invertible, then $1/s$ is a singular value of $(M - zI)^{-1}$. Thus,

we conclude that

$$\sup_{z \in E} s_1(G(z)) = \sup_{z \in E} \frac{1}{s_n(M - zI)}$$

$$= \frac{1}{\inf_{z \in E} s_n(M - zI)}$$

$$\leq \frac{1}{c},$$

as desired.   □

5. Eigenvalue criterion lemma and an overview of the proof

Let us now briefly overview the proofs of our main results. One of the key ingredients is the eigenvalue criterion lemma presented below (Lemma 5.1), which is based on Sylvester's determinant theorem:

$$\det(I + AB) = \det(I + BA) \tag{12}$$

whenever A is an $n \times k$ matrix and B is a $k \times n$ matrix. In particular, the left-hand side of (12) is an $n \times n$ determinant and the right-hand side is a $k \times k$ determinant.

For concreteness, let us focus on the proof of Theorem 2.8. That is, we wish to study the eigenvalues of

$$P_n := \prod_{k=1}^{m} \left(\frac{1}{\sqrt{n}} X_{n,k} + A_{n,k} \right)$$

outside the disk $\{z \in \mathbb{C} : |z| \leq \sigma + 2\varepsilon\}$. We first linearize the problem by invoking Proposition 4.1 with the matrix $\frac{1}{\sqrt{n}} \mathcal{Y}_n + \mathcal{A}_n$, where $\mathcal{Y}_n$ and $\mathcal{A}_n$ are defined in (9) and (11). Indeed, by Proposition 4.1, it suffices to study the eigenvalues of $\frac{1}{\sqrt{n}} \mathcal{Y}_n + \mathcal{A}_n$ outside of the disk $\{z \in \mathbb{C} : |z| \leq \sigma^{1/m} + \delta\}$ for some $\delta > 0$ (depending on σ, ε, and m). Let us suppose that $\mathcal{A}_n$ is rank one. In other words, assume $\mathcal{A}_n = vu^*$ for some $u, v \in \mathbb{C}^{mn}$. In order to study the outlier eigenvalues, we will need to solve the equation

$$\det \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n + \mathcal{A}_n - zI \right) = 0 \tag{13}$$

for $z \in \mathbb{C}$ with $|z| > \sigma^{1/m} + \delta$. Assuming z is not eigenvalue of $\mathcal{Y}_n$, we can rewrite (13) as

$$\det \left(I + \mathcal{G}_n(z) \mathcal{A}_n \right) = 0,$$

where the resolvent $\mathcal{G}_n(z)$ is defined in (10). From (12) and the fact that $\mathcal{A}_n = vu^*$, we find that this reduces to solving

$$1 + u^* \mathcal{G}_n(z) v = 0.$$

Thus, the problem of locating the outlier eigenvalues reduces to studying the resolvent $\mathcal{G}_n(z)$. In particular, we develop an isotropic limit law in Section 6 to compute the limit of $u^* \mathcal{G}_n(z) v$. This limit law is inspired by the isotropic semicircle law developed by Knowles and Yin in [44, 45] for Wigner random matrices as well as the isotropic law verified in [49] for elliptic matrices.

The general case, when $\mathcal{A}_n$ is not necessarily rank one, is similar. In this case, we will exploit the following criterion to characterize the outlier eigenvalues.

Lemma 5.1 (Eigenvalue criterion). *Let Y and A be $n \times n$ matrices, and assume $A = BC$, where B is an $n \times k$ matrix and C is a $k \times n$ matrix. Let z be a complex number which is not an eigenvalue of Y. Then z is an eigenvalue of $Y + A$ if and only if*

$$\det\left(I_k + C \left(Y - zI_n \right)^{-1} B \right) = 0.$$

Remark 5.2. The proof of Lemma 5.1 actually reveals that

$$\det\left(I_k + C \left(Y - zI_n \right)^{-1} B \right) = \frac{\det\left(Y + A - zI \right)}{\det\left(Y - zI \right)} \tag{14}$$

provided the denominator does not vanish. Versions of this identity have appeared in previous publications including [10, 19, 20, 21, 34, 35, 44, 45, 49, 52, 55, 57, 60].

Proof of Lemma 5.1. Assume z is not an eigenvalue of Y. Then $\det(Y - zI) \neq 0$ and

$$\det(Y + A - zI) = \det(Y - zI) \det(I + (Y - zI)^{-1}A)$$
$$= \det(Y - zI) \det(I + (Y - zI)^{-1}BC).$$

Thus, by (12), z is an eigenvalue of $Y + A$ if and only if

$$\det(I + C(Y - zI)^{-1}B) = 0,$$

as desired. $\qquad\square$

Another identity we will make use of is the Resolvent Identity, which states that

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1} \tag{15}$$

whenever A and B are invertible.

6. Isotropic limit law and the proofs of the main theorems

This section is devoted to the proofs of Theorems 2.2, 2.5, 2.6, 2.7, and 2.8. The key ingredient is the following result concerning the properties of the resolvent $\mathcal{G}_n(z)$.

Theorem 6.1 (Isotropic limit law). *Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables with mean zero, unit variance, finite fourth moments, and independent real and imaginary parts. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Recall that $\mathcal{Y}_n$ is defined in (9) and its resolvent $\mathcal{G}_n(z)$ is defined in (10). Then, for any fixed $\delta > 0$, the following statements hold.*

(i) Almost surely, for n sufficiently large, the eigenvalues of $\frac{1}{\sqrt{n}}\mathcal{Y}_n$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. In particular, this implies that almost surely, for n sufficiently large, the matrix $\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI$ is invertible for every $z \in \mathbb{C}$ with $|z| > 1 + \delta$.

(ii) There exists a constant $c > 0$ (depending only on δ) such that almost surely, for n sufficiently large,

$$\sup_{z \in \mathbb{C} : |z| > 1 + \delta} \|\mathcal{G}_n(z)\| \leq c.$$

(iii) *For each $n \geq 1$, let $u_n, v_n \in \mathbb{C}^{mn}$ be deterministic unit vectors. Then*

$$\sup_{z \in \mathbb{C}: |z| > 1+\delta} \left| u_n^* \mathcal{G}_n(z) v_n + \frac{1}{z} u_n^* v_n \right| \longrightarrow 0$$

almost surely as $n \to \infty$.

We conclude this section with the proofs of Theorems 2.2, 2.5, 2.6, 2.7, and 2.8 assuming Theorem 6.1. Sections 7–11 are devoted to the proof of Theorem 6.1.

6.1. **Proof of Theorem 2.2.** Consider

$$P_n := n^{-m/2} X_{n,1} \cdots X_{n,m}$$

and note that by rescaling by $\frac{1}{\sigma}$, it is sufficient to assume that $\sigma_i = 1$ for all $1 \leq i \leq m$. By Proposition 4.1, the eigenvalues of P_n are precisely the eigenvalues of $n^{-m/2} \mathcal{Y}_n^m$, each with multiplicity m. Additionally, the eigenvalues of $n^{-m/2} \mathcal{Y}_n^m$ are exactly the m-th powers of the eigenvalues of $n^{-1/2} \mathcal{Y}_n$. Thus, it is sufficient to study the spectral radius of $n^{-1/2} \mathcal{Y}_n$. By part (i) of Theorem 6.1, we conclude that almost surely,

$$\limsup_{n \to \infty} \rho\left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right) \leq 1$$

where $\rho(M)$ denotes the spectral radius of the matrix M. This completes the proof of Theorem 2.2.

6.2. **Proof of Theorem 2.5.** By rescaling by $\frac{1}{\sigma}$, it is sufficient to assume that $\sigma_i = 1$ for $1 \leq i \leq m$. Observe that if two deterministic terms

$$I + A_{n,i} \text{ and } I + A_{n,j}$$

appeared consecutively in the product P_n, then they could be rewritten

$$(I + A_{n,i}) \cdot (I + A_{n,j}) = I + A'_{n,j}$$

where all non-identity terms are lumped into the new deterministic matrix $A'_{n,j}$ which still satisfies the assumptions on rank and norm. Additionally, if two random matrices $X_{n,i}$ and $X_{n,j}$ appeared in the product P_n consecutively, then we could write

$$\left(\frac{1}{\sqrt{n}} X_{n,i} \right) \left(\frac{1}{\sqrt{n}} X_{n,j} \right) = \left(\frac{1}{\sqrt{n}} X_{n,i} \right) (I_n + 0_n) \left(\frac{1}{\sqrt{n}} X_{n,j} \right)$$

where 0_n denotes the $n \times n$ zero matrix. Therefore, it is sufficient to consider products in which terms alternate between a random term $\frac{1}{\sqrt{n}} X_{n,i}$ and a deterministic term $I + A_{n,j}$. Next, observe that the eigenvalues of the product P_n remain the same when the matrices in the product are cyclically permuted. Thus, without loss of generality and up to reordering the indices, we may assume that the product P_n appears as

$$P_n = (I + A_{n,1}) \left(\frac{1}{\sqrt{n}} X_{n,1} \right) (I + A_{n,2}) \left(\frac{1}{\sqrt{n}} X_{n,2} \right) \cdots (I + A_{n,m}) \left(\frac{1}{\sqrt{n}} X_{n,m} \right). \tag{16}$$

Next, define the $2mn \times 2mn$ matrix

$$\mathcal{L}_n := \begin{bmatrix} 0_{mn} & I_{mn} + \mathrm{diag}(A_{n,1}, \ldots, A_{n,m}) \\ \frac{1}{\sqrt{n}} \mathcal{Y}_n & 0_{mn} \end{bmatrix}, \tag{17}$$

where 0_{mn} denotes the $mn \times mn$ zero matrix, $\mathcal{Y}_n$ is as defined in (9), and $\mathrm{diag}(A_{n,1}, \ldots, A_{n,m})$ is defined so that

$$I_{mn} + \mathrm{diag}(A_{n,1}, \ldots, A_{n,m}) := \begin{bmatrix} I + A_{n,1} & 0 & \ldots & 0 \\ 0 & I + A_{n,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & I + A_{n,m} \end{bmatrix}.$$

Here, we have to slightly adjust our notation to deal with the fact that $\mathcal{L}_n$ is a $2mn \times 2mn$ matrix instead of $mn \times mn$. For the remainder of the proof, we view $\mathcal{L}_n$ as a 2×2 matrix with entries which are $mn \times mn$ matrices, and we denote the $mn \times mn$ blocks as $\mathcal{L}_n^{[a,b]}$ for $1 \leq a, b \leq 2$. We will use analogous notation for other $2mn \times 2mn$ matrices and $2mn$-vectors.

For $1 \leq k \leq 2m$, let W_k denote the product P_n, but with terms cyclically permuted so that the product starts on the kth term. Note that there are $2m$ such products and each results in an $n \times n$ matrix. For instance,

$$W_2 = \left(\frac{1}{\sqrt{n}} X_{n,1}\right) (I + A_{n,2}) \left(\frac{1}{\sqrt{n}} X_{n,2}\right) (I + A_{n,3}) \cdots \left(\frac{1}{\sqrt{n}} X_{n,m}\right) (I + A_{n,1})$$

and

$$W_3 = (I + A_{n,2}) \left(\frac{1}{\sqrt{n}} X_{n,2}\right) (I + A_{n,3}) \cdots \left(\frac{1}{\sqrt{n}} X_{n,m}\right) (I + A_{n,1}) \left(\frac{1}{\sqrt{n}} X_{n,1}\right).$$

A simple computation reveals that

$$\mathcal{L}_n^{2m} = \begin{bmatrix} \mathcal{W} & 0_{mn} \\ 0_{mn} & \tilde{\mathcal{W}} \end{bmatrix}$$

where

$$\mathcal{W} = \begin{bmatrix} W_1 & 0_n & \ldots & 0_n \\ 0_n & W_3 & \ldots & 0_n \\ \vdots & \vdots & \ddots & \vdots \\ 0_n & 0_n & \ldots & W_{2m-1} \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{W}} = \begin{bmatrix} W_2 & 0_n & \ldots & 0_n \\ 0_n & W_4 & \ldots & 0_n \\ \vdots & \vdots & \ddots & \vdots \\ 0_n & 0_n & \ldots & W_{2m} \end{bmatrix}.$$

Thus, the eigenvalues of $\mathcal{L}_n^{2m}$ are precisely the eigenvalues of the product P_n, each with multiplicity $2m$.

Define

$$\mathcal{X}_n := \begin{bmatrix} 0_{mn} & I_{mn} \\ \frac{1}{\sqrt{n}} \mathcal{Y}_n & 0_{mn} \end{bmatrix} \quad \text{and} \quad \mathcal{A}_n^{\square} := \begin{bmatrix} 0_{mn} & \mathrm{diag}(A_{n,1}, \ldots, A_{n,m}) \\ 0_{mn} & 0_{mn} \end{bmatrix}.$$

Then we can rewrite $\mathcal{L}_n = \mathcal{X}_n + \mathcal{A}_n^{\square}$.

For $1 \leq k \leq m$, let $Z_k := X_{n,k} X_{n,k+1} \cdots X_{n,m} X_{n,1} \cdots X_{n,k-1}$. Then

$$\mathcal{X}_n^{2m} = n^{-m/2} \begin{bmatrix} \mathcal{Z}_n & 0_{mn} \\ 0_{mn} & \mathcal{Z}_n \end{bmatrix} \tag{18}$$

where

$$\mathcal{Z}_n = \begin{bmatrix} Z_1 & \ldots & 0_n \\ \vdots & \ddots & \vdots \\ 0_n & \ldots & Z_m \end{bmatrix}.$$

Thus, the eigenvalues of $\mathcal{X}_n^{2m}$ are precisely the eigenvalues of $n^{-m/2} X_{n,1} \cdots X_{n,m}$, each with multiplicity $2m$.

Fix $\delta > 0$. By part (i) of Theorem 6.1 and Proposition 4.1, we find that almost surely, for n sufficiently large, the eigenvalues of $n^{-m/2}X_{n,1}\cdots X_{n,m}$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. By (18), we conclude that almost surely, for n sufficiently large, the eigenvalues of $\mathcal{X}_n$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. Therefore, almost surely, for n sufficiently large $\mathcal{X}_n - zI$ is invertible for all $|z| > 1 + \delta$. For all such values of z, we define

$$\mathcal{R}_n(z) := (\mathcal{X}_n - zI_{2mn})^{-1}.$$

Since $\mathcal{A}_n^{\square}$ has rank $O(1)$ and operator norm $O(1)$, we can decompose (by the singular value decomposition) $\mathcal{A}_n^{\square} = \mathcal{B}_n\mathcal{C}_n$, where $\mathcal{B}_n$ is a $mn \times k$ matrix, $\mathcal{C}_n$ is a $k \times mn$ matrix, $k = O(1)$, and both $\mathcal{B}_n$ and $\mathcal{C}_n$ have rank $O(1)$ and operator norm $O(1)$.

Thus, for $|z| > 1 + \delta$, almost surely, for n sufficiently large,

$$\det(\mathcal{L}_n - zI_{2mn}) = \det(\mathcal{X}_n + \mathcal{A}_n^{\square} - zI_{2mn}) = 0$$

if and only if

$$\det(I_k + \mathcal{C}_n\mathcal{R}_n(z)\mathcal{B}_n) = 0 \tag{19}$$

by Lemma 5.1.

Using Schur's Compliment to calculate the $mn \times mn$ blocks of $\mathcal{R}_n(z)$, we can see that $\mathcal{R}_n(z) = \begin{pmatrix} z\mathcal{G}_n(z^2) & \mathcal{G}_n(z^2) \\ I + z^2\mathcal{G}_n(z^2) & z\mathcal{G}_n(z^2) \end{pmatrix}$, where $\mathcal{G}_n(z) := \left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right)^{-1}$ (which is defined for $|z| > 1 + \delta$ by Theorem 6.1); hence

$$\mathcal{R}_n(z)^{[a,b]} = \begin{cases} z\mathcal{G}_n(z^2) & \text{if } a = b \\ \mathcal{G}_n(z^2) & \text{if } a = 1, \ b = 2 \\ I + z^2\mathcal{G}_n(z^2) & \text{if } a = 2, \ b = 1. \end{cases}$$

Note that for $u = u_{2mn}$ and $v = v_{2mn}$ in $\mathbb{C}^{2mn}$, we have

$$u^*\mathcal{R}_n(z)v = \sum_{1 \leq a,b \leq 2} (u^*)^{[a]}\,\mathcal{R}_n(z)^{[a,b]}v^{[b]}$$

where $v^{[1]}, v^{[2]}$ denote the $mn \times 1$ sub-blocks of the vector v and $(u^*)^{[1]}, (u^*)^{[2]}$ denote the $1 \times mn$ sub-blocks of the vector u^*. Additionally, if u and v have uniformly bounded norm for all n, then by Theorem 6.1, almost surely

$$\sup_{|z|>1+\delta} \left| (u^*)^{[1]} z\mathcal{G}_n(z^2)v^{[1]} - z\left(-\frac{1}{z^2}\right)(u^*)^{[1]}v^{[1]} \right| = o(1),$$

$$\sup_{|z|>1+\delta} \left| (u^*)^{[2]} z\mathcal{G}_n(z^2)v^{[2]} - z\left(-\frac{1}{z^2}\right)(u^*)^{[2]}v^{[2]} \right| = o(1),$$

$$\sup_{|z|>1+\delta} \left| (u^*)^{[1]}\mathcal{G}_n(z^2)v^{[2]} - \left(-\frac{1}{z^2}(u^*)^{[1]}v^{[2]}\right) \right| = o(1), \text{ and}$$

$$\sup_{|z|>1+\delta} \left| (u^*)^{[2]}(I + z^2\mathcal{G}_n(z^2))v^{[1]} - \left((u^*)^{[2]}v^{[1]} - z^2\frac{1}{z^2}(u^*)^{[2]}v^{[1]}\right) \right| = o(1).$$

We note that the off-diagonal blocks of $\mathcal{R}_n(z)$ have a much different behavior than $\mathcal{G}_n(z)$. In order to keep track of this behavior, define

$$\mathcal{H}_n = \begin{bmatrix} 0_{mn} & I_{mn} \\ 0_{mn} & 0_{mn} \end{bmatrix}.$$

Then, for deterministic u and v in $\mathbb{C}^{2mn}$ with uniformly bounded norm for all n, almost surely

$$\sup_{|z|>1+\delta}\left|u^*\mathcal{R}_n(z)v - \left(-\frac{1}{z}u^*v - \frac{1}{z^2}u^*\mathcal{H}_nv\right)\right| = o(1).$$

Applying this to (19), we obtain that

$$\sup_{|z|>1+\delta}\left\|\mathcal{C}_n\mathcal{R}_n(z)\mathcal{B}_n - \left(-\frac{1}{z}\mathcal{C}_n\mathcal{B}_n - \frac{1}{z^2}\mathcal{C}_n\mathcal{H}_n\mathcal{B}_n\right)\right\| = o(1)$$

almost surely. Hence, Lemma D.2 reveals that almost surely

$$\sup_{|z|>1+\delta}\left|\det\left(I - \mathcal{C}_n\mathcal{R}_n\mathcal{B}_n\right) - \det\left(I - \frac{1}{z}\mathcal{C}_n\mathcal{B}_n - \frac{1}{z^2}\mathcal{C}_n\mathcal{H}_n\mathcal{B}_n\right)\right| = o(1).$$

By another application of Sylvester's determinant formula (12) and by noticing that $\mathcal{H}_n\mathcal{A}_n^{\square}$ is the zero matrix, this can be rewritten as

$$\sup_{|z|>1+\delta}\left|\det\left(I + \mathcal{R}_n(z)\mathcal{A}_n^{\square}\right) - \det\left(I - \frac{1}{z}\mathcal{A}_n^{\square}\right)\right| = o(1).$$

Finally, since the eigenvalues of $\mathcal{A}_n^{\square}$ are all zero, $\det\left(I - \frac{1}{z}\mathcal{A}_n^{\square}\right) = 1$ so we can write the above statement as

$$\sup_{|z|>1+\delta}\left|\det\left(\mathcal{R}_n\right)\det\left(\mathcal{L}_n - zI\right) - 1\right| = o(1)$$

almost surely. Almost surely, for n sufficiently large, and for $|z| > 1 + \delta$, we know that $\det(R_n(z))$ is finite, and thus $\det\left(\mathcal{L}_n - zI\right)$ is nonzero for all $|z| > 1 + \delta$, implying that $\mathcal{L}_n$ has no eigenvalues outside the disk $\{z \in \mathbb{C} : |z| \le 1 + \delta\}$. By the previous observations, this implies the same conclusion for P_n (since δ is arbitrary), and the proof is complete.

6.3. **Proof of Theorem 2.8.** Recall that

$$P_n := \prod_{k=1}^m \left(\frac{1}{\sqrt{n}}X_{n,k} + A_{n,k}\right), \qquad A_n := \prod_{k=1}^m A_{n,k}.$$

By rescaling P_n by $\frac{1}{\sigma}$, we may assume that $\sigma_i = 1$ for $1 \le i \le n$. Let $\varepsilon > 0$, and assume that for sufficiently large n, no eigenvalues of A_n fall in the band $\{z \in \mathbb{C} : 1 + \varepsilon < |z| < 1 + 3\varepsilon\}$. Assume that for some $j = O(1)$, there are j eigenvalues $\lambda_1(A_n), \ldots, \lambda_j(A_n)$ that lie in the region $\{z \in \mathbb{C} : |z| \ge 1 + 3\varepsilon\}$.

Let $\mathcal{Y}_n$ and $\mathcal{A}_n$ be defined as in (9) and (11). Using Proposition 4.1, it will suffice to study the eigenvalues of $\frac{1}{\sqrt{n}}\mathcal{Y}_n + \mathcal{A}_n$. In particular, let $\varepsilon' > 0$ such that $(1 + 2\varepsilon)^{1/m} = 1 + \varepsilon'$. Then we want to find solutions to

$$\det\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n + \mathcal{A}_n - zI\right) = 0 \tag{20}$$

for $|z| \ge 1 + \varepsilon'$. By part (i) of Theorem 6.1, almost surely, for n sufficiently large $\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI$ is invertible for all $|z| \ge 1 + \varepsilon'$. By supposition, we can decompose (using the singular value decomposition) $\mathcal{A}_n = \mathcal{B}_n\mathcal{C}_n$, where $\mathcal{B}_n$ is $mn \times k$, $\mathcal{C}_n$ is $k \times mn$, $k = O(1)$, and both $\mathcal{B}_n$ and $\mathcal{C}_n$ have rank $O(1)$ and operator norm $O(1)$. Thus, by Lemma 5.1, we need to investigate the values of $z \in \mathbb{C}$ with $|z| \ge 1 + \varepsilon'$ such that

$$\det(I_k + \mathcal{C}_n\mathcal{G}_n(z)\mathcal{B}_n) = 0,$$

where $\mathcal{G}_n(z)$ is defined in (10).

Since $k = O(1)$, Theorem 6.1 implies that

$$\sup_{|z| \geq 1+\varepsilon'} \left\| \mathcal{C}_n \mathcal{G}_n(z) \mathcal{B}_n - \left(-\frac{1}{z} \right) \mathcal{C}_n \mathcal{B}_n \right\| \longrightarrow 0 \tag{21}$$

almost surely as $n \to \infty$. By Lemma D.2, this implies that

$$\sup_{|z| \geq 1+\varepsilon'} \left| \det\left(I_k + \mathcal{C}_n \mathcal{G}_n(z) \mathcal{B}_n \right) - \det\left(I_k - \frac{1}{z} \mathcal{C}_n \mathcal{B}_n \right) \right| \longrightarrow 0$$

almost surely. By an application of Sylvester's determinant theorem (12), this is equivalent to

$$\sup_{|z| \geq 1+\varepsilon'} \left| \det\left(I_k + \mathcal{C}_n \mathcal{G}_n \mathcal{B}_n \right) - \det\left(I_n - \frac{1}{z} \mathcal{A}_n \right) \right| \longrightarrow 0 \tag{22}$$

almost surely as $n \to \infty$. Define

$$g(z) := \det\left(I_n - \frac{1}{z} \mathcal{A}_n \right) = \prod_{i=1}^{k} \left(1 - \frac{\lambda_i(\mathcal{A}_n)}{z} \right).$$

Since the eigenvalues of $\mathcal{A}_n^m$ are precisely the eigenvalues of A_n, each with multiplicity m, it follows that g has precisely $l := jm$ roots $\lambda_1(\mathcal{A}_n), \ldots, \lambda_l(\mathcal{A}_n)$ outside the disk $\{ z \in \mathbb{C} : |z| < 1 + \varepsilon' \}$. Thus, by (22) and Rouché's theorem, almost surely, for n sufficiently large,

$$f(z) := \det\left(I_k + \mathcal{C}_n \mathcal{G}_n \mathcal{B}_n \right)$$

has exactly l roots outside the disk $\{ z \in \mathbb{C} : |z| < 1 + \varepsilon' \}$ and these roots take the values $\lambda_i(\mathcal{A}_n) + o(1)$ for $1 \leq i \leq l$.

Returning to (20), we conclude that almost surely for n sufficiently large, $\frac{1}{\sqrt{n}} \mathcal{Y}_n + \mathcal{A}_n$ has exactly l roots outside the disk $\{ z \in \mathbb{C} : |z| < 1 + \varepsilon' \}$, and after possibly reordering the eigenvalues, these roots take the values

$$\lambda_i\left(\frac{1}{\sqrt{n}} \mathcal{Y}_n + \mathcal{A}_n \right) = \lambda_i(\mathcal{A}_n) + o(1)$$

for $1 \leq i \leq l$.

We now relate these eigenvalues back to the eigenvalues of P_n. Recall that $\left(\frac{1}{\sqrt{n}} \mathcal{Y}_n + \mathcal{A}_n \right)^m$ has the same eigenvalues as P_n, each with multiplicity m; and $\mathcal{A}_n^m$ has the same eigenvalues of A_n, each with multiplicity m. Taking this additional multiplicity into account and using the fact that

$$(\lambda_i(\mathcal{A}_n) + o(1))^m = \lambda_i(\mathcal{A}_n^m) + o(1)$$

since $\mathcal{A}_n$ has spectral norm $O(1)$, we conclude that almost surely, for n sufficiently large, P_n has exactly j eigenvalues in the region $\{ z \in \mathbb{C} : |z| \geq 1 + 2\varepsilon \}$, and after reordering the indices correctly

$$\lambda_i(P_n) = \lambda_i(A_n) + o(1)$$

for $1 \leq i \leq j$. This completes the proof of Theorem 2.8.

6.4. **Proof of Theorems 2.6 and 2.7.** In the proofs of Theorems 2.6 and 2.7 we will make use of an alternative isotropic law which is a corollary of Theorem 6.1. We state and prove the result now.

Corollary 6.2 (Alternative Isotropic Law). *Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables with mean zero, unit variance, finite fourth moments, and independent real and imaginary parts. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Then, for any fixed $\delta > 0$, the following statements hold.*

(i) *Almost surely, for n sufficiently large, the eigenvalues of $n^{-m/2}X_{n,1} \cdots X_{n,m}$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. In particular, this implies that almost surely, for n sufficiently large, the matrix $n^{-m/2}X_{n,1} \cdots X_{n,m} - zI$ is invertible for every $z \in \mathbb{C}$ with $|z| > 1 + \delta$.*

(ii) *There exists a constant $c > 0$ (depending only on δ) such that almost surely, for n sufficiently large,*

$$\sup_{z \in \mathbb{C}: |z| > 1+\delta} \left\| \left(n^{-m/2}X_{n,1} \cdots X_{n,m} - zI \right)^{-1} \right\| \leq c.$$

(iii) *For each $n \geq 1$, let $u_n, v_n \in \mathbb{C}^n$ be deterministic unit vectors. Then*

$$\sup_{z \in \mathbb{C}: |z| > 1+\delta} \left| u_n^* \left(n^{-m/2}X_{n,1} \cdots X_{n,m} - zI \right)^{-1} v_n + \frac{1}{z} u_n^* v_n \right| \longrightarrow 0$$

almost surely as $n \to \infty$.

Proof. Part (i) follows from Theorem 2.2. Let $\mathcal{Y}_n$ be defined by (9), and let $\mathcal{G}_n(z)$ be defined by (10). Then the last two parts of Corollary 6.2 follow from the last two parts of Theorem 6.1 due to the fact that

$$\mathcal{G}_n^{[1,1]}(z) = z^{m-1}(n^{-m/2}X_{n,1} \cdots X_{n,m} - z^m I)^{-1}.$$

$\square$

With this result in hand, we proceed to the remainder of the proofs.

Proof of Theorem 2.6. By rescaling by $\frac{1}{\sigma}$ it suffices to assume that $\sigma_i = 1$ for $1 \leq i \leq m$. Let $\varepsilon > 0$. Assume that for sufficiently large n there are no eigenvalues of A_n in the band $\{z \in \mathbb{C} : 1 + \varepsilon < |z| < 1 + 3\varepsilon\}$ and there are j eigenvalues, $\lambda_1(A_n), \ldots, \lambda_j(A_n)$, in the region $\{z \in \mathbb{C} : |z| \geq 1 + 3\varepsilon\}$.

By Corollary 6.2, almost surely, for n sufficiently large, $n^{-m/2}X_{n,1} \cdots X_{n,m} - zI$ is invertible for all $|z| > 1 + \varepsilon$. We decompose (using the singular value decomposition) $A_n = B_n C_n$, where B_n is $n \times k$, C_n is $k \times n$, $k = O(1)$, and both B_n and C_n have rank $O(1)$ and spectral norm $O(1)$. By Lemma 5.1, the eigenvalues of P_n outside $\{z \in \mathbb{C} : |z| < 1 + 2\varepsilon\}$ are precisely the values of $z \in \mathbb{C}$ with $|z| > 1 + 2\varepsilon$ such that

$$\det \left(I_k + C_n \left(n^{-m/2}X_{n,1} \cdots X_{n,m} - zI_n \right)^{-1} B_n \right) = 0.$$

By Corollary 6.2,

$$\sup_{|z| > 1+\varepsilon} \left\| C_n \left(n^{-m/2}X_{n,1} \cdots X_{n,m} - zI_n \right)^{-1} B_n + \frac{1}{z} C_n B_n \right\| \longrightarrow 0$$

almost surely as $n \to \infty$. By applying (12) and Lemma D.2, this gives

$$\sup_{|z|>1+\varepsilon} \left| \det\left(I_k + C_n \left(n^{-m/2} X_{n,1} \cdots X_{n,m} - zI_n \right)^{-1} B_n \right) - \det\left(I_n - \frac{1}{z} A_n \right) \right| \longrightarrow 0$$

almost surely as $n \to \infty$. By recognizing the roots of $\det\left(I_n - \frac{1}{z} A_n \right)$ as the eigenvalues of A_n, Rouché's Theorem implies that almost surely for n sufficiently large P_n has exactly j eigenvalues in the region $\{z \in \mathbb{C} : |z| > 1 + 2\varepsilon\}$, and after labeling the eigenvalues properly,

$$\lambda_i(P_n) = \lambda_i(A_n) + o(1)$$

for $1 \le i \le j$. $\qquad\square$

The proof of Theorem 2.7 will require the following corollary of [60, Lemma 2.3].

Lemma 6.3. *Let φ_n and $X_{n,1}, \ldots, X_{n,m}$ be as in Theorem 2.7. Then almost surely,*

$$n^{-m/2} |\varphi_n^* X_{n,1} X_{n,2} \cdots X_{n,m} \varphi_n| = o(1).$$

Proof. Let $u := n^{-(m-1)/2} \varphi_n^* X_{n,1} X_{n,2} \cdots X_{n,m-1}$. In view of [60, Theorem 1.4], it follows that $\|u\| = O(1)$ almost surely. We now condition on $X_{n,1}, \ldots, X_{n,m-1}$ so that $\|u\| = O(1)$. As $X_{n,m}$ is independent of u, we apply [60, Lemma 2.3] to conclude that

$$u \left(\frac{1}{\sqrt{n}} X_{n,m} \right) \varphi_n = o(1)$$

almost surely, concluding the proof. $\qquad\square$

With this result, we may proceed to the proof of Theorem 2.7.

Proof of Theorem 2.7. By rescaling by $\frac{1}{\sigma}$ it suffices to assume that $\sigma_i = 1$ for $1 \le i \le m$. Fix $\gamma > 0$ and let $\varepsilon > 0$. By Corollary 6.2 and Lemma 5.1, almost surely, for n sufficiently large, the only eigenvalues of

$$P_n := n^{-m/2} X_{n,1} \cdots X_{n,m} + \mu n^\gamma \phi_n \phi_n^*$$

in the region $\{z \in \mathbb{C} : |z| > 1 + \varepsilon\}$ are the values of $z \in \mathbb{C}$ with $|z| > 1 + \varepsilon$ such that

$$1 + \mu n^\gamma \phi_n^* \left(n^{-m/2} X_{n,1} \cdots X_{n,m} - zI_n \right)^{-1} \phi_n = 0. \qquad (23)$$

Define the functions

$$f(z) := 1 + \mu n^\gamma \phi_n^* \left(n^{-m/2} X_{n,1} \cdots X_{n,m} - zI_n \right)^{-1} \phi_n$$

$$g(z) := 1 - \frac{\mu n^\gamma}{z}.$$

Observe that $g(z)$ has one zero located at μn^γ which will be outside the disk $\{z \in \mathbb{C} : |z| \le 1 + \varepsilon\}$ for large enough n. By Corollary 6.2, it follows that almost surely

$$\sup_{|z|>1+\varepsilon} |f(z) - g(z)| = o(n^\gamma).$$

Thus, almost surely

$$f(z) = g(z) + o(n^\gamma)$$

for all $z \in \mathbb{C}$ with $|z| > 1 + \varepsilon$.

Observe that if z is a root of f with $|z| > 1 + \varepsilon$, then $|g(z)| = \left| 1 - \frac{\mu n^\gamma}{z} \right| = o(n^\gamma)$. We conclude that if z is a root of f outside the disk $\{z \in \mathbb{C} : |z| > 1 + \varepsilon\}$, then the root must tend to infinity with n almost surely. We will return to this fact shortly.

For the next step of the proof, we will need to bound the spectral norm of $n^{-m/2}X_{n,1}\cdots X_{n,m}$. To do so, we apply [60, Theorem 1.4] and obtain that, almost surely, for n sufficiently large,

$$n^{-m/2}\|X_{n,1}\cdots X_{n,m}\| \leq (2.1)^m. \tag{24}$$

Thus, by a Neumann series expansion, for all $|z| > (2.5)^m$, we have

$$f(z) = 1 - \frac{\mu n^\gamma}{z} + \frac{\mu n^\gamma}{z^2}\phi_n^*\left(n^{-m/2}X_{n,1}\cdots X_{n,m}\right)\phi_n + O\left(\frac{n^\gamma}{|z|^3}\right)$$

$$= g(z) + \frac{\mu n^\gamma}{z^2}\phi_n^*\left(n^{-m/2}X_{n,1}\cdots X_{n,m}\right)\phi_n + O\left(\frac{n^\gamma}{|z|^3}\right).$$

By Lemma 6.3, $\varphi_n^*\left(n^{-m/2}X_{n,1}\cdots X_{n,m}\right)\varphi_n = o(1)$ almost surely, so one can see

$$f(z) = g(z) + o\left(\frac{n^\gamma}{|z|^2}\right) + O\left(\frac{n^\gamma}{|z|^3}\right)$$

uniformly for all $|z| > (2.5)^m$. In particular, when $|z| \to \infty$, we obtain

$$|z||f(z) - g(z)| = o\left(\frac{n^\gamma}{|z|}\right) \tag{25}$$

almost surely. Since any root of $zf(z)$ outside $\{z \in \mathbb{C} : |z| \leq 1 + \varepsilon\}$ must tend to infinity with n, it follows from Rouché's theorem that almost surely, for n sufficiently large, $zf(z)$ has precisely one root outside the disk $\{z \in \mathbb{C} : |z| \leq 1 + \varepsilon\}$ and that root takes the value $\mu n^\gamma + o(n^\gamma)$.

It remains to reduce the error from $o(n^\gamma)$ to $o(1)$. Fix $\delta > 0$, and let Γ be the circle around μn^γ with radius δ. Then from (25) we see that almost surely

$$\sup_{z \in \Gamma} |z||f(z) - g(z)| = o(1).$$

Hence, almost surely, for n sufficiently large,

$$|z||f(z) - g(z)| < \delta = |z||g(z)|$$

for all $z \in \Gamma$. Therefore, by another application of Rouché's theorem, we conclude that almost surely, for n sufficiently large, $zf(z)$ contains precisely one root outside of $\{z \in \mathbb{C} : |z| \leq 1 + \varepsilon\}$ and that root is located in the interior of Γ. Since δ was arbitrary, this completes the proof. $\square$

7. Truncation and useful tools

We now turn to the proof of Theorem 6.1. We will require the following standard truncation results for iid random matrices.

Lemma 7.1. *Let ξ be a complex-valued random variable with mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. Let $\mathrm{Re}(\xi)$ and $\mathrm{Im}(\xi)$ denote the real and imaginary parts of ξ respectively, and let $\sqrt{-1}$ denote the imaginary unit. For $L > 0$, define*

$$\tilde{\xi} := \mathrm{Re}(\xi)\mathbf{1}_{\{|\mathrm{Re}(\xi)|\leq L/\sqrt{2}\}} - \mathbb{E}\left[\mathrm{Re}(\xi)\mathbf{1}_{\{|\mathrm{Re}(\xi)|\leq L/\sqrt{2}\}}\right]$$

$$+ \sqrt{-1}\left(\mathrm{Im}(\xi)\mathbf{1}_{\{|\mathrm{Im}(\xi)|\leq L/\sqrt{2}\}} - \mathbb{E}\left[\mathrm{Im}(\xi)\mathbf{1}_{\{|\mathrm{Im}(\xi)|\leq L/\sqrt{2}\}}\right]\right)$$

and

$$\hat{\xi} := \frac{\tilde{\xi}}{\sqrt{\mathrm{Var}(\tilde{\xi})}}.$$

Then there exists a constant $L_0 > 0$ (depending only on $\mathbb{E}|\xi|^4$) such that the following statements hold for all $L > L_0$.

    (i) $\mathrm{Var}(\tilde{\xi}) \geq \frac{1}{2}$
    (ii) $|1 - \mathrm{Var}(\tilde{\xi})| \leq \frac{4}{L^2}\mathbb{E}|\xi|^4$
    (iii) *Almost surely, $|\hat{\xi}| \leq 4L$*
    (iv) *$\hat{\xi}$ has mean zero, unit variance, $\mathbb{E}|\hat{\xi}|^4 \leq C\mathbb{E}|\xi|^4$ for some absolute constant $C > 0$, and the real and imaginary parts of $\hat{\xi}$ are independent.*

The proof of this theorem is a standard truncation argument. The full details of the proof can be found in Appendix A.

Let X be an $n \times n$ random matrix filled with iid copies of a random variable ξ which has mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. We define matrices $\tilde{X}$ and $\hat{X}$ to be the $n \times n$ matrices with entries defined by

$$\tilde{X}_{(i,j)} := \mathrm{Re}(X_{(i,j)})\mathbf{1}_{\{|\mathrm{Re}(X_{(i,j)})|\leq L/\sqrt{2}\}} - \mathbb{E}\left[\mathrm{Re}(X_{(i,j)})\mathbf{1}_{\{|\mathrm{Re}(X_{(i,j)})|\leq L/\sqrt{2}\}}\right]$$
$$+ \sqrt{-1}\left(\mathrm{Im}(X_{(i,j)})\mathbf{1}_{\{|\mathrm{Im}(X_{(i,j)})|\leq L/\sqrt{2}\}} - \mathbb{E}\left[\mathrm{Im}(X_{(i,j)})\mathbf{1}_{\{|\mathrm{Im}(X_{(i,j)})|\leq L/\sqrt{2}\}}\right]\right)$$

$$\tag{26}$$

and

$$\hat{X}_{(i,j)} := \frac{\tilde{X}_{(i,j)}}{\sqrt{\mathrm{Var}(\tilde{X}_{(i,j)})}} \tag{27}$$

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

Lemma 7.2. *Let X be an iid random matrix with atom variable ξ which has mean zero, unit variance, $m_4 := \mathbb{E}|\xi|^4 < \infty$, and independent real and imaginary parts. Let $\hat{X}$ be as defined in (27). Then, there exist constants $C, L_0 > 0$ (depending only on m_4) such that for all $L > L_0$*

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}}\left\|X - \hat{X}\right\| \leq \frac{C}{L}$$

almost surely.

Proof. Let $\tilde{X}$ be defined as in (26), and let L_0 be the value from Lemma 7.1. Begin by noting that

$$\left\|X - \hat{X}\right\| \leq \left\|X - \tilde{X}\right\| + \left\|\tilde{X} - \hat{X}\right\|$$

and thus it suffices to show that

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}}\left\|X - \tilde{X}\right\| \leq \frac{C}{L}$$

and

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}}\left\|\tilde{X} - \hat{X}\right\| \leq \frac{C}{L}$$

almost surely. First, by Lemma 7.1,

$$\frac{1}{\sqrt{n}}\left\|\tilde{X}-\hat{X}\right\| = \frac{1}{\sqrt{n}}\left\|\tilde{X}\right\|\left|1-\frac{1}{\sqrt{\mathrm{Var}(\tilde{\xi})}}\right|$$

$$\leq \frac{1}{\sqrt{n}}\left\|\tilde{X}\right\|\sqrt{2}\left|\sqrt{\mathrm{Var}(\tilde{\xi})}-1\right|$$

$$\leq \frac{1}{\sqrt{n}}\left\|\tilde{X}\right\|\sqrt{2}\left|\mathrm{Var}(\tilde{\xi})-1\right|$$

$$\leq \frac{1}{\sqrt{n}}\left\|\tilde{X}\right\|\sqrt{2}\left(\frac{4}{L^2}\mathbb{E}|\xi|^4\right). \tag{28}$$

By [60, Theorem 1.4], we find that almost surely

$$\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\left\|\tilde{X}\right\|\leq 2,$$

and thus by (28)

$$\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\left\|\tilde{X}-\hat{X}\right\|\leq\frac{C}{L}$$

almost surely for all $L\geq\max\{1,L_0\}$.

Next consider $\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\left\|X-\tilde{X}\right\|$. Note that $X-\tilde{X}$ is an iid matrix with atom variable

$$\mathrm{Re}(X_{(i,j)})\mathbf{1}_{\{|\mathrm{Re}(X_{(i,j)})|>L/\sqrt{2}\}} - \mathbb{E}\left[\mathrm{Re}(X_{(i,j)})\mathbf{1}_{\{|\mathrm{Re}(X_{(i,j)})|>L/\sqrt{2}\}}\right]$$
$$+ \sqrt{-1}\left(\mathrm{Im}(X_{(i,j)})\mathbf{1}_{\{|\mathrm{Im}(X_{(i,j)})|>L/\sqrt{2}\}} - \mathbb{E}\left[\mathrm{Im}(X_{(i,j)})\mathbf{1}_{\{|\mathrm{Im}(X_{(i,j)})|>L/\sqrt{2}\}}\right]\right).$$

Thus, each entry has mean zero, variance

$$\mathrm{Var}((X-\tilde{X})_{i,j})\leq\frac{8}{L^2}\mathbb{E}|\xi|^4,$$

and finite fourth moment. Thus, again by [60, Theorem 1.4],

$$\limsup_{n\to\infty}\frac{1}{\sqrt{n}}\left\|X-\tilde{X}\right\|\leq\frac{C}{L}$$

almost surely, and the proof is complete. $\qquad\square$

We now consider the iid random matrices $X_{n,1},\ldots,X_{n,m}$ from Theorem 6.1. For each $1\leq k\leq m$, define the truncation $\hat{X}_{n,k}$ as was done above for $\hat{X}$ in (27). Define $\hat{\mathcal{Y}}_n$ by

$$\hat{\mathcal{Y}}_n := \begin{bmatrix} 0 & \hat{X}_{n,1} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{X}_{n,m-1} \\ \hat{X}_{n,m} & 0 & \ldots & 0 \end{bmatrix}. \tag{29}$$

Using Theorem 7.1, we have the following corollary.

Corollary 7.3. *Let $X_{n,1},\ldots,X_{n,m}$ be independent iid random matrices with atom variables $\xi_1,\ldots,\xi_m$, each of which has mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. Let $\hat{X}_{n,1},\ldots,\hat{X}_{n,m}$ be the truncations of $X_{n,1},\ldots,X_{n,m}$ as was done in (27). In addition, let $\mathcal{Y}_n$ be as defined in*

(9) and $\hat{\mathcal{Y}}_n$ be as defined in (29). Then there exist constants $C, L_0 > 0$ (depending only on the atom variables $\xi_1, \ldots, \xi_m$) such that

$$\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left\| \mathcal{Y}_n - \hat{\mathcal{Y}}_n \right\| \leq \frac{C}{L} \tag{30}$$

almost surely for all $L > L_0$.

Proof. Due to the block structure of $\mathcal{Y}_n$ and $\hat{\mathcal{Y}}_n$, it follows that

$$\left\| \mathcal{Y}_n - \hat{\mathcal{Y}}_n \right\| \leq \max_k \left\| X_{n,k} - \hat{X}_{n,k} \right\|.$$

Therefore, the claim follows from Lemma 7.2. $\square$

8. Least singular value bounds

In this section, we study the least singular value of $\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI$. We begin by recalling Weyl's inequality for the singular values (see, for example, [24, Problem III.6.5]), which states that for $n \times n$ matrices A and B,

$$\max_{1 \leq i \leq n} |s_i(A) - s_i(B)| \leq \|A - B\|. \tag{31}$$

We require the following theorem, which is based on [47, Theorem 2].

Theorem 8.1. *Fix $L > 0$, and let $\xi_1, \ldots, \xi_m$ be complex-valued random variables, each having mean zero, unit variance, independent real and imaginary parts, and which satisfy*

$$\sup_{1 \leq k \leq m} |\xi_k| \leq L$$

almost surely. Let $X_{n,1}, \ldots, X_{n,m}$ be independent iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Define $\mathcal{Y}_n$ as in (9), and fix $\delta > 0$. Then there exists a constant $c > 0$ (depending only on δ) such that

$$\inf_{|z| \geq 1 + \delta} s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \geq c$$

with overwhelming probability.

A similar statement was proven in [47, Theorem 2], where the same conclusion was shown to hold almost surely rather than with overwhelming probability. However, many of the intermediate steps in [47] are proven to hold with overwhelming probability. We use these intermediate steps to prove Theorem 8.1 in Appendix B.

Lemma 8.2. *Let $X_{n,1}, \ldots, X_{n,m}$ satisfy the assumptions of Theorem 6.1, and let $\mathcal{Y}_n$ be as defined in (9). Fix $\delta > 0$. Then there exists a constant $c > 0$ such that almost surely, for n sufficiently large,*

$$\inf_{|z| \geq 1 + \delta} s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \geq c.$$

Proof. Let $L > 0$ be a large constant to be chosen later. Let $\hat{X}_{n,1}, \ldots, \hat{X}_{n,m}$ be defined as in (27), and let $\hat{\mathcal{Y}}_n$ be defined as in (29). By Theorem 8.1 and the Borel–Cantelli lemma, there exists a constant $c' > 0$ such that almost surely, for n sufficiently large,

$$\inf_{|z| \geq 1 + \delta} s_{mn} \left(\frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n - zI \right) \geq c'.$$

By Corollary 7.3 and (31) we may choose L sufficiently large to ensure that almost surely, for n sufficiently large,

$$\left| s_{mn}\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right) - s_{mn}\left(\frac{1}{\sqrt{n}}\hat{\mathcal{Y}}_n - zI\right)\right| \le \frac{c'}{2},$$

uniformly in z. We conclude that almost surely, for n sufficiently large,

$$\inf_{|z|\ge 1+\delta} s_{mn}\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right) \ge \frac{c'}{2},$$

and the proof is complete. □

With this result we can prove the following lemma.

Lemma 8.3. *Let $X_{n,1},\ldots,X_{n,m}$ satisfy the assumptions of Theorem 6.1, and fix $\delta > 0$. Then there exists a constant $c > 0$ such that almost surely, for n sufficiently large,*

$$\inf_{|z|\ge 1+\delta} s_{mn}\left(n^{-m/2}X_{n,1}\cdots X_{n,m} - zI\right) \ge c.$$

Proof. Let $\mathcal{Y}_n$ be defined as in (9). Then Lemma 8.2 implies that almost surely, for n sufficiently large, $\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI$ is invertible for all $|z| \ge 1 + \delta$. By computing the block inverse of this matrix, we find

$$\left(\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right)^{-1}\right)^{[1,1]} = z^{m-1}\left(n^{-m/2}X_{n,1}\cdots X_{n,m} - z^m I\right)^{-1}.$$

Thus, for $|z| \ge 1 + \delta$,

$$\left\|\left(n^{-m/2}X_{n,1}\cdots X_{n,m} - z^m I\right)^{-1}\right\| \le |z|^{m-1}\left\|\left(n^{-m/2}X_{n,1}\cdots X_{n,m} - z^m I\right)^{-1}\right\|$$

$$\le \left\|\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right)^{-1}\right\|,$$

where the last step used the fact that the operator norm of a matrix bounds above the operator norm of any sub-matrix.

Recall that if M is an invertible $N \times N$ matrix, then $s_N(M) = \|M^{-1}\|^{-1}$. Applying this fact to the matrices above, we conclude that

$$s_{mn}\left(n^{-m/2}X_{n,1}\cdots X_{n,m} - z^m I\right) \ge s_{mn}\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right),$$

and the claim follows from Lemma 8.2. □

Remark 8.4. Observe that z is an eigenvalue of $\frac{1}{\sqrt{n}}\mathcal{Y}_n$ if and only if $\det\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right) = 0$. Also, recall that $\left|\det\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right)\right| = \prod_i s_i\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI\right)$; this product is zero if and only if the smallest singular value is zero. Since Lemma 8.2 and Lemma 8.3 bound the least singular values of $\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI$ and $n^{-m/2}X_{n,1}\cdots X_{n,m} - zI$ away from zero for $|z| \ge 1 + \delta$, we can conclude that such values of z are almost surely, for n sufficiently large, not eigenvalues of these matrices.

9. Reductions to the proof of Theorem 6.1

In this section, we will prove that it is sufficient to reduce the proof of Theorem 6.1 to the case in which the entries of each matrix are truncated and where we restrict z to the band $5 \leq |z| \leq 6$.

Theorem 9.1. *Let $X_{n,1}, \ldots, X_{n,m}$ be as in Theorem 6.1. Then there exists a constant L_0 such that the following holds for all $L > L_0$. Let $\hat{X}_{n,1}, \ldots, \hat{X}_{n,m}$ be defined as in (27), let $\hat{\mathcal{Y}}_n$ be given by (29), and let $\hat{\mathcal{G}}_n(z) := \left(\frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n - zI \right)^{-1}$.*

(i) For any fixed $\delta > 0$, almost surely, for n sufficiently large, the eigenvalues of $\frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. In particular, $\frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n - zI$ is almost surely invertible for every $z \in \mathbb{C}$ with $|z| > 1 + \delta$.

(ii) For any fixed $\delta > 0$, there exists a constant $c > 0$ (depending only on δ) such that almost surely, for n sufficiently large

$$\sup_{z \in \mathbb{C}: |z| > 1 + \delta} \|\hat{\mathcal{G}}_n(z)\| \leq c.$$

(iii) For each $n \geq 1$, let $u_n, v_n \in \mathbb{C}^{mn}$ be deterministic unit vectors. Then

$$\sup_{z \in \mathbb{C}: 5 \leq |z| \leq 6} \left| u_n^* \hat{\mathcal{G}}_n(z) v_n + \frac{1}{z} u_n^* v_n \right| \longrightarrow 0$$

almost surely as $n \to \infty$.

We now prove Theorem 6.1 assuming Theorem 9.1.

Proof of Theorem 6.1. Part (i) of Theorem 6.1 follows from Lemma 8.2 (see Remark 8.4). In addition, part (ii) of Theorem 6.1 follows from an application of Lemma 8.2 and Proposition 4.2.

We now turn to the proof of part (iii). Fix $0 < \delta < 1$. Let $\mathcal{Y}_n$ be given by (9), and let $\mathcal{G}_n$ be given by (10). Let $\varepsilon, \varepsilon' > 0$, and observe that there exists a positive constant M_1 such that for $|z| \geq M_1$,

$$\left\| \left(-\frac{1}{z} \right) u^* v \right\| \leq \left| \frac{1}{z} \right| \|u^*\| \|v\| \leq \frac{\varepsilon}{2}.$$

Also, by [60, Theorem 1.4] and Lemma D.3, there exists a constant $M_2 > 0$ such that almost surely, for n sufficiently large,

$$\sup_{|z| \geq M_2} \|u^* \mathcal{G}_n(z) v\| \leq \sup_{|z| \geq M_2} \|\mathcal{G}_n(z)\| \leq \frac{\varepsilon}{2}.$$

Let $M := \max\{M_1, M_2, 6\}$. Then, almost surely, for n sufficiently large,

$$\sup_{|z| \geq M} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \leq \varepsilon. \tag{32}$$

We now work on the region where $1 + \delta < |z| \leq M$. By the resolvent identity (15), we note that

$$\left\| \mathcal{G}_n(z) - \hat{\mathcal{G}}_n(z) \right\| \leq \|\mathcal{G}_n(z)\| \left\| \hat{\mathcal{G}}_n(z) \right\| \frac{1}{\sqrt{n}} \left\| \hat{\mathcal{Y}}_n - \mathcal{Y}_n \right\|.$$

Thus, by part (ii) of Theorem 6.1 (proven above), Theorem 9.1, and Corollary 7.3, there exist constants $C, c > 0$ such that

$$\limsup_{n \to \infty} \sup_{|z| > 1 + \delta} \left\| \mathcal{G}_n(z) - \hat{\mathcal{G}}_n(z) \right\| \leq \limsup_{n \to \infty} c^2 \frac{1}{\sqrt{n}} \left\| \hat{\mathcal{Y}}_n - \mathcal{Y}_n \right\| \leq c^2 \frac{C}{L} \leq \frac{\varepsilon'}{2} \quad (33)$$

almost surely for L sufficiently large.

From (33) and Theorem 9.1, almost surely, for n sufficiently large,

$$\sup_{5 \leq |z| \leq 6} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \leq \sup_{5 \leq |z| \leq 6} \left| u^* \mathcal{G}_n(z) v - u^* \hat{\mathcal{G}}_n(z) v \right|$$

$$+ \sup_{5 \leq |z| \leq 6} \left| u^* \hat{\mathcal{G}}_n(z) v + \frac{1}{z} u^* v \right|$$

$$\leq \varepsilon'.$$

Since $\varepsilon' > 0$ was arbitrary, this implies that

$$\limsup_{n \to \infty} \sup_{5 \leq |z| \leq 6} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| = 0$$

almost surely. Since the region $\{z \in \mathbb{C} \ : \ 1 + \delta \leq |z| \leq M\}$ is compact and contains the region $\{z \in \mathbb{C} \ : \ 5 \leq |z| \leq 6\}$, Vitali's Convergence Theorem[3] (see, for instance, [15, Lemma 2.14]) implies that we can extend this convergence to the larger region, and we obtain

$$\limsup_{n \to \infty} \sup_{1 + \delta \leq |z| \leq M} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| = 0$$

almost surely. In particular, this implies that, almost surely, for n sufficiently large,

$$\sup_{1 + \delta \leq |z| \leq M} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \leq \varepsilon.$$

Combined with (32), this completes the proof of Theorem 6.1 (since $\varepsilon > 0$ was arbitrary). $\qquad\qquad\square$

It remains to prove Theorem 9.1. We prove parts (i) and (ii) of Theorem 9.1 now. The proof of part (iii) is lengthy and will be addressed in the forthcoming sections.

Proof of Theorem 9.1 (i) and (ii). Let $\delta > 0$, and observe that by Theorem 8.1 and the Borel–Cantelli lemma, there exists a constant $c > 0$ (depending only on δ) such that almost surely, for n sufficiently large,

$$\inf_{|z| > 1 + \delta} s_{mn} \left(\frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n - zI \right) \geq c. \quad (34)$$

This implies (see Remark 8.4) that almost surely, for n sufficiently large, the eigenvalues $\frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$, proving (i). From (34) and Proposition 4.2, we conclude that almost surely, for n sufficiently large,

$$\sup_{|z| > 1 + \delta} \left\| \hat{\mathcal{G}}_n(z) \right\| \leq \frac{1}{c},$$

---

[3] The hypothesis of Vitali's Convergence Theorem are satisfied almost surely, for n sufficiently large, by parts (i) and (ii) of Theorem 6.1. In addition, one can check that $(u_n)^* \mathcal{G}_n(z) v_n$ is holomorphic in the region $\{z \in \mathbb{C} : |z| > 1 + \delta\}$ almost surely, for n sufficiently large, using the resolvent identity (15).

proving (ii). □

10. CONCENTRATION OF BILINEAR FORMS INVOLVING THE RESOLVENT $\mathcal{G}_n$

Sections 10 and 11 are devoted to the proof of part (iii) of Theorem 9.1. Let $\hat{X}_{n,1}, \ldots, \hat{X}_{n,m}$ be the truncated matrices from Theorem 9.1, and let $u_n, v_n \in \mathbb{C}^{mn}$ be deterministic unit vectors. For ease of notation, in Sections 10 and 11, we drop the decorations and write $X_{n,1}, \ldots, X_{n,m}$ for $\hat{X}_{n,1}, \ldots, \hat{X}_{n,m}$. Similarly, we write $\mathcal{Y}_n$ for $\hat{\mathcal{Y}}_n$ and $\mathcal{G}_n$ for $\hat{\mathcal{G}}_n$. Recall that all constants and asymptotic notation may depend on m without explicitly showing the dependence.

Define the following event:

$$\Omega_n := \left\{ \frac{1}{\sqrt{n}} \|\mathcal{Y}_n\| \leq 4.5 \right\}. \tag{35}$$

Lemma 10.1 (Spectral Norm Bound for $\mathcal{Y}_n$). *Under the assumptions above, the event Ω_n holds with overwhelming probability.*

Proof. Based on the block structure of $\mathcal{Y}_n$, it follows that

$$\|\mathcal{Y}_n\| \leq \max_i \|X_{n,i}\|.$$

Therefore, the claim follows from [15, Theorem 5.9] (alternatively, [60, Theorem 1.4]). In fact, the constant 4.5 can be replaced with any constant strictly larger than 2; 4.5 will suffice for what follows. □

By Lemma 10.1, Ω_n holds with overwhelming probability, i.e., for every $A > 0$,

$$\mathbb{P}\left(\Omega_n\right) = 1 - O_A(n^{-A}). \tag{36}$$

We will return to this fact several times throughout the proof. The remainder of this section is devoted to proving the following lemma.

Lemma 10.2 (Concentration). *Let $u_n, v_n \in \mathbb{C}^{mn}$ be deterministic unit vectors. Then, under the assumptions above, for any $\varepsilon > 0$, almost surely*

$$\sup_{5 \leq |z| \leq 6} |u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n} - \mathbb{E}\left[u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n}\right]| < \varepsilon \tag{37}$$

for n sufficiently large.

The proof of Lemma 10.2 follows the arguments of [13, 49]. Before we begin the proof, we present some notation. Define $\mathcal{Y}_n^{(k)}$ to be the matrix $\mathcal{Y}_n$ with all entries in the kth row and the kth column filled with zeros. Note that $\mathcal{Y}_n^{(k)}$ is still an $mn \times mn$ matrix. Define

$$\mathcal{G}_n^{(k)} := \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n^{(k)} - zI \right)^{-1}, \tag{38}$$

let r_k denote the kth row of $\mathcal{Y}_n$, and let c_k denote kth column of $\mathcal{Y}_n$. Also define the σ-algebra

$$\mathcal{F}_k := \sigma(r_1, \ldots, r_k, c_1, \ldots, c_k)$$

generated by the first k rows and the first k columns of $\mathcal{Y}_n$. Note that $\mathcal{F}_0$ is the trivial σ-algebra and $\mathcal{F}_{mn}$ is the σ-algebra generated by all rows and columns. Next define

$$\mathbb{E}_k[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\,\mathcal{F}_k] \tag{39}$$

to be the conditional expectation given the first k rows and columns, and

$$\Omega_n^{(k)} := \left\{ \frac{1}{\sqrt{n}} \|\mathcal{Y}_n^{(k)}\| \le 4.5 \right\}. \tag{40}$$

Observe that $\Omega_n \subseteq \Omega_n^{(k)}$ and therefore, by Lemma 10.1, $\Omega_n^{(k)}$ holds with overwhelming probability as well.

Remark 10.3. By Lemma D.3, we have that

$$\sup_{5 \le |z| \le 6} \|\mathcal{G}_n(z)\| \le 2, \qquad \sup_{5 \le |z| \le 6} \left\|\mathcal{G}_n^{(k)}(z)\right\| \le 2$$

on the event Ω_n, and $\sup_{5 \le |z| \le 6} \left\|\mathcal{G}_n^{(k)}(z)\right\| \le 2$ on $\Omega_n^{(k)}$.

We will now collect some preliminary calculations and lemmata that will be required for the proof of Lemma 10.2.

Lemma 10.4 (Rosenthal's Inequality; [33]). *Let $\{x_k\}$ be a complex martingale difference sequence with respect to the filtration $\mathcal{F}_k$. Then for $p \ge 2$,*

$$\mathbb{E} \left| \sum x_k \right|^p \le K_p \left(\mathbb{E} \left(\sum \mathbb{E} \left[|x_k|^2 \mid \mathcal{F}_{k-1} \right] \right)^{p/2} + \mathbb{E} \sum |x_k|^p \right)$$

for a constant $K_p > 0$ which depends only on p.

Lemma 10.5. *Let A be an $n \times n$ Hermitian positive semidefinite matrix, and let $S \subset [n]$. Then $\sum_{i \in S} A_{ii} \le \operatorname{tr} A$.*

Proof. The claim follows from the fact that, by definition of A being Hermitian positive semidefinite, the diagonal entries of A are real and non-negative. $\square$

Lemma 10.6. *Let A be an $N \times N$ Hermitian positive semidefinite matrix with rank at most one. Suppose that ξ is a complex-valued random variable with mean zero, unit variance, and which satisfies $|\xi| \le L$ almost surely for some constant $L > 0$. Let $S \subseteq [N]$, and let $w = (w_i)_{i=1}^N$ be a vector with the following properties:*

   (i) $\{w_i : i \in S\}$ is a collection of iid copies of ξ,
   (ii) $w_i = 0$ for $i \notin S$.
Then for any $p \ge 1$,

$$\mathbb{E} |w^* A w|^p \ll_{L,p} \|A\|^p. \tag{41}$$

Proof. Let w_S denote the $|S|$-vector which contains entries w_i for $i \in S$, and let $A_{S \times S}$ denote the $|S| \times |S|$ matrix which has entries $A_{(i,j)}$ for $i, j \in S$. Then we observe

$$w^* A w = \sum_{i,j} \bar{w}_i A_{(i,j)} w_j = w_S^* A_{S \times S} w_S.$$

By Lemma D.1, we get

$$\mathbb{E} |w^* A w|^p \ll_p (\operatorname{tr} A_{S \times S})^p + \mathbb{E}|\xi|^{2p} \operatorname{tr} A_{S \times S}^p \le (\operatorname{tr} A_{S \times S})^p + L^{2p} \operatorname{tr} A_{S \times S}^p.$$

Since the rank of $A_{S \times S}$ is at most one, we find

$$\operatorname{tr} A_{S \times S} \le \|A\|$$

and

$$\operatorname{tr} A_{S \times S}^p \le \|A\|^p,$$

where we used the fact that the operator norm of a matrix bounds the operator norm of any sub-matrix. We conclude that

$$\mathbb{E}\left|w^* A w\right|^p \ll_p \|A\|^p + L^{2p} \|A\|^p \ll_{L,p} \|A\|^p,$$

as desired. $\qquad\square$

Lemma 10.7. *Let A be a deterministic complex $N \times N$ matrix. Suppose that ξ is a complex-valued random variable with mean zero, unit variance, and which satisfies $|\xi| \leq L$ almost surely for some constant $L > 0$. Let $S, R \subseteq [N]$, and let $w = (w_i)_{i=1}^N$ and $t = (t_i)_{i=1}^N$ be independent vectors with the following properties:*

(i) $\{w_i : i \in S\}$ and $\{t_j : j \in R\}$ are collections of iid copies of ξ,

(ii) $w_i = 0$ for $i \notin S$, and $t_j = 0$ for $j \notin R$.

Then for any $p \geq 1$,

$$\mathbb{E}\left|w^* A t\right|^p \ll_{L,p} (\mathrm{tr}(A^* A))^{p/2}. \tag{42}$$

Proof. Let w_S denote the $|S|$-vector which contains entries w_i for $i \in S$, and let t_R denote the $|R|$-vector which contains entries t_j for $j \in R$. For an $N \times N$ matrix B, we let $B_{S \times S}$ denote the $|S| \times |S|$ matrix with entries $B_{(i,j)}$ for $i, j \in S$. Similarly, we let $B_{R \times R}$ denote the $|R| \times |R|$ matrix with entries $B_{(i,j)}$ for $i, j \in R$.

We first note that, by the Cauchy–Schwarz inequality, it suffices to assume p is even. In this case, since w is independent of t, Lemma D.1 implies that

$$\begin{aligned}
\mathbb{E}|w^* A t|^p &= \mathbb{E}|w^* A t t^* A^* w|^{p/2} \\
&= \mathbb{E}\left|w_S^* (A t t^* A^*)_{S \times S} w_S\right|^{p/2} \\
&\ll_p \mathbb{E}\left[(\mathrm{tr}(A t t^* A^*)_{S \times S})^{p/2} + L^p \, \mathrm{tr}(A t t^* A^*)_{S \times S}^{p/2}\right].
\end{aligned}$$

Recall that for any matrix B, $\mathrm{tr}(B^* B)^{p/2} \leq (\mathrm{tr}(B^* B))^{p/2}$. By this fact and by Lemma 10.5, we observe that

$$\mathbb{E}\left[(\mathrm{tr}(A t t^* A^*)_{S \times S})^{p/2} + L^p \, \mathrm{tr}(A t t^* A^*)_{S \times S}^{p/2}\right] \ll_{L,p} \mathbb{E}\left[(\mathrm{tr}(A t t^* A^*))^{p/2}\right].$$

By a cyclic permutation of the trace, we have

$$\mathbb{E}\left[(\mathrm{tr}(A t t^* A^*))^{p/2}\right] = \mathbb{E}\left[(t^* A^* A t)^{p/2}\right] \leq \mathbb{E}\left|t^* A^* A t\right|^{p/2}.$$

By Lemma D.1, Lemma 10.5, and a similar argument as above, we obtain

$$\begin{aligned}
\mathbb{E}\left|t^* A^* A t\right|^{p/2} &= \mathbb{E}\left|t_R^* (A^* A)_{R \times R} t_R\right|^{p/2} \\
&\ll_p (\mathrm{tr}(A^* A)_{R \times R})^{p/2} + L^p \, \mathrm{tr}(A^* A)_{R \times R}^{p/2} \\
&\ll_{L,p} (\mathrm{tr}(A^* A))^{p/2},
\end{aligned}$$

completing the proof. $\qquad\square$

Lemma 10.8. *Let r_k be the kth row of $\mathcal{Y}_n$, c_k be the kth column of $\mathcal{Y}_n$, $\mathcal{G}_n^{(k)}(z)$ be the resolvent of $\mathcal{Y}_n^{(k)}$, and $u_n \in \mathbb{C}^{mn}$ be a deterministic unit vector. Then, under the assumptions above,*

$$\sup_{5 \leq |z| \leq 6} \mathbb{E}_{k-1} \left|\frac{1}{n} r_k \mathcal{G}_n^{(k)}(z) u_n u_n^* \mathcal{G}_n^{(k)*}(z) \mathbf{1}_{\Omega_n^{(k)}} r_k^*\right|^p \ll_{L,p} n^{-p} \tag{43}$$

and

$$\sup_{5 \leq |z| \leq 6} \mathbb{E}_{k-1} \left|\frac{1}{n} c_k^* \mathcal{G}_n^{(k)*}(z) u_n u_n^* \mathcal{G}_n^{(k)*}(z) \mathbf{1}_{\Omega_n^{(k)}} c_k\right|^p \ll_{L,p} n^{-p}. \tag{44}$$

almost surely.

Proof. We will only prove the bound in (43) as the proof of (44) is identical. For each fixed z in the band $5 \leq |z| \leq 6$, we will show that

$$\mathbb{E}_{k-1} \left| \frac{1}{n} r_k \mathcal{G}_n^{(k)}(z) u_n u_n^* \mathcal{G}_n^{(k)*}(z) \mathbf{1}_{\Omega_n^{(k)}} r_k^* \right|^p \ll_{L,p} n^{-p}$$

surely, where the implicit constant does not depend on z. The claim then follows by taking the supremum over all z in the band $5 \leq |z| \leq 6$.

To this end, fix z with $5 \leq |z| \leq 6$. Throughout the proof, we drop the dependence on z in the resolvent as it is clear from context. Note that

$$\mathbb{E}_{k-1} \left| \frac{1}{n} r_k \mathcal{G}_n^{(k)} u_n u_n^* \mathcal{G}_n^{(k)*} \mathbf{1}_{\Omega_n^{(k)}} r_k^* \right|^p = \frac{1}{n^p} \mathbb{E}_{k-1} \left| r_k \left(\mathcal{G}_n^{(k)} u_n u_n^* \mathcal{G}_n^{(k)*} \mathbf{1}_{\Omega_n^{(k)}} \right) r_k^* \right|^p,$$

and $\mathcal{G}_n^{(k)} u_n u_n^* \mathcal{G}_n^{(k)*} \mathbf{1}_{\Omega_n^{(k)}}$ is independent of r_k. In addition, observe that $\mathcal{G}_n^{(k)} u_n u_n^* \mathcal{G}_n^{(k)*} \mathbf{1}_{\Omega_n^{(k)}}$ is Hermitian positive semidefinite matrix with rank at most one. Applying Lemma 10.6 and Remark 10.3, we obtain

$$\mathbb{E}_{k-1} \left| \frac{1}{n} r_k \mathcal{G}_n^{(k)} u_n u_n^* \mathcal{G}_n^{(k)*} \mathbf{1}_{\Omega_n^{(k)}} r_k^* \right|^p \ll_{L,p} \frac{1}{n^p} \mathbb{E}_{k-1} \left\| \mathcal{G}_n^{(k)} u_n u_n^* \mathcal{G}_n^{(k)*} \mathbf{1}_{\Omega_n^{(k)}} \right\|^p \ll_{L,p} n^{-p}$$

surely, and the proof is complete. $\square$

Lemma 10.9. *Let $\zeta_1, \ldots, \zeta_{mn}$ be complex-valued random variables (not necessarily independent) which depend on $\mathcal{Y}_n$. Assume*

$$\sup_k |\zeta_k| \mathbf{1}_{\Omega_n^{(k)}} = O(1)$$

almost surely. Then, under the assumptions above, for any $p \geq 1$,

$$\mathbb{E} \left| \sum_{k=1}^{mn} \left(\zeta_k \mathbf{1}_{\Omega_n^{(k)}} - \zeta_k \mathbf{1}_{\Omega_n \cap \Omega_n^{(k)}} \right) \right|^p = O_p(n^{-p}).$$

Proof. We will exploit the fact that $\Omega_n \subseteq \Omega_n^{(k)}$ for any $1 \leq k \leq mn$. Indeed, we have

$$\mathbb{E} \left| \sum_{k=1}^{mn} \left(\zeta_k \mathbf{1}_{\Omega_n^{(k)}} - \zeta_k \mathbf{1}_{\Omega_n \cap \Omega_n^{(k)}} \right) \right|^p \leq \mathbb{E} \left| \sum_{k=1}^{mn} \zeta_k \mathbf{1}_{\Omega_n^{(k)} \cap \Omega_n^c} \right|^p$$

$$\ll_p \mathbb{E} \left(\sum_{k=1}^{mn} \mathbf{1}_{\Omega_n^{(k)} \cap \Omega_n^c} \right)^p$$

$$\ll_p n^p \mathbb{P}(\Omega_n^c),$$

and the claim follows from (36). $\square$

We will made use of the Sherman–Morrison rank one perturbation formula (see [42, Section 0.7.4]). Suppose A is an invertible square matrix, and let u, v be vectors. If $1 + v^* A^{-1} u \neq 0$, then

$$(A + uv^*)^{-1} = A^{-1} - \frac{A^{-1} uv^* A^{-1}}{1 + v^* A^{-1} u} \tag{45}$$

and

$$(A + uv^*)^{-1} u = \frac{A^{-1} u}{1 + v^* A^{-1} u}. \tag{46}$$

Now, we proceed to prove the main result of this section, Lemma 10.2.

Proof of Lemma 10.2. Define

$$\mathcal{Y}_n^{(k1)} := \mathcal{Y}_n^{(k)} + e_k r_k, \quad \mathcal{Y}_n^{(k2)} := \mathcal{Y}_n^{(k)} + c_k e_k^* \tag{47}$$

where $e_1, \ldots, e_{mn}$ are the standard basis elements of $\mathbb{C}^{mn}$. Also define

$$\mathcal{G}_n^{(kj)} := \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n^{(kj)} - zI \right)^{-1}, \quad j = 1, 2, \tag{48}$$

and set

$$\alpha_n^{(k)} := \frac{1}{1 + z^{-1} n^{-1} r_k \mathcal{G}_n^{(k)} c_k \mathbf{1}_{\Omega_n}},$$

$$\zeta_n^{(k)} := n^{-1} r_k \mathcal{G}_n^{(k)} c_k,$$

$$\eta_n^{(k)} := n^{-1} r_k \mathcal{G}_n^{(k)} v_n u_n^* \mathcal{G}_n^{(k)} c_k.$$

Using these definitions, we make the following observations.

(i) Since the only nonzero element in the k-th row and k-th column of $\mathcal{Y}_n^{(k)} - zI$ is on the diagonal,

$$e_k^* \mathcal{G}_n^{(k)} e_k = -z^{-1}, \quad e_k^* \mathcal{G}_n^{(k)} v_n = -z^{-1} v_{n,k}, \quad u_n^* \mathcal{G}_n^{(k)} e_k = -z^{-1} \bar{u}_{n,k}$$

where $u_n = (u_{n,k})_{k=1}^{mn}$ and $v_n = (v_{n,k})_{k=1}^{mn}$.

(ii) Since the k-th elements of c_k and r_k are zero,

$$e_k^* \mathcal{G}_n^{(k)} c_k = 0, \quad r_k \mathcal{G}_n^{(k)} e_k = 0.$$

(iii) By (45), (i), and (ii),

$$e_k^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k = e_k^* \mathcal{G}_n^{(k)} n^{-1/2} c_k - \frac{e_k^* \mathcal{G}_n^{(k)} e_k n^{-1/2} r_k \mathcal{G}_n^{(k)} n^{-1/2} c_k}{1 + n^{-1/2} r_k \mathcal{G}_n^{(k)} e_k}$$

$$= z^{-1} n^{-1} r_k \mathcal{G}_n^{(k)} c_k,$$

so that

$$\frac{1}{1 + e_k^* \mathcal{G}_n^{(k1)} \mathbf{1}_{\Omega_n} n^{-1/2} c_k} = \alpha_n^{(k)}.$$

(iv) By the same argument as (iii),

$$n^{-1/2} r_k \mathcal{G}_n^{(k2)} e_k = z^{-1} n^{-1} r_k \mathcal{G}_n^{(k)} c_k,$$

so that

$$\frac{1}{1 + n^{-1/2} r_k \mathcal{G}_n^{(k2)} e_k \mathbf{1}_{\Omega_n}} = \alpha_n^{(k)}.$$

(v) By Schur's compliment,

$$(\mathcal{G}_n)_{(k,k)} = -\frac{1}{z + n^{-1} r_k \mathcal{G}_n^{(k)} c_k}$$

provided the necessary inverses exist (which is the case on the event Ω_n). Thus, on $\Omega_n = \Omega_n \cap \Omega_n^{(k)}$ and uniformly for $5 \le |z| \le 6$, by Remark 10.3,

$$\left| \alpha_n^{(k)} \right| = \left| \frac{z}{z + n^{-1} r_k \mathcal{G}_n^{(k)} c_k} \right|$$

$$= |z| \left| (\mathcal{G}_n)_{(k,k)} \right|$$

$$\le 12.$$

On Ω_n^c, $\alpha_n^{(k)} = 1$, so we have that, almost surely,

$$\left|\alpha_n^{(k)}\right| \le 12.$$

(vi) By (46) and (iii),

$$
\begin{aligned}
u_n^* \mathcal{G}_n n^{-1/2} c_k \mathbf{1}_{\Omega_n} &= \frac{u_n^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k \mathbf{1}_{\Omega_n}}{1 + e_k^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k} \\
&= \frac{u_n^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k \mathbf{1}_{\Omega_n}}{1 + e_k^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k \mathbf{1}_{\Omega_n}} \\
&= u_n^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k \mathbf{1}_{\Omega_n} \alpha_n^{(k)}.
\end{aligned}
$$

(vii) By (46) and (iv), a similar argument as above gives

$$u_n^* \mathcal{G}_n e_k \mathbf{1}_{\Omega_n} = u_n^* \mathcal{G}_n^{(k2)} e_k \mathbf{1}_{\Omega_n} \alpha_n^{(k)}.$$

(viii) By (45) and (ii),

$$
\begin{aligned}
\mathcal{G}_n^{(k1)} &= \mathcal{G}_n^{(k)} - \frac{\mathcal{G}_n^{(k)} e_k n^{-1/2} r_k \mathcal{G}_n^{(k)}}{1 + n^{-1/2} r_k \mathcal{G}_n^{(k)} e_k} \\
&= \mathcal{G}_n^{(k)} - \mathcal{G}_n^{(k)} e_k n^{-1/2} r_k \mathcal{G}_n^{(k)}.
\end{aligned}
$$

(ix) By (45) and (ii), and by the same calculation as in (viii),

$$\mathcal{G}_n^{(k2)} = \mathcal{G}_n^{(k)} - \mathcal{G}_n^{(k)} n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)}.$$

(x) By definition of $\alpha_n^{(k)}$,

$$z^{-1}(\mathbb{E}_k - \mathbb{E}_{k-1})[n^{-1} r_k \mathcal{G}_n^{(k)} c_k \alpha_n^{(k)}] = -(\mathbb{E}_k - \mathbb{E}_{k-1})[\alpha_n^{(k)}].$$

(xi) By definition of $\alpha_n^{(k)}$ and $\zeta_n^{(k)}$,

$$
\begin{aligned}
\alpha_n^{(k)} - 1 &= \frac{-z^{-1} n^{-1} r_k \mathcal{G}_n^{(k)} \mathbf{1}_{\Omega_n} c_k}{1 + z^{-1} n^{-1} r_k \mathcal{G}_n^{(k)} \mathbf{1}_{\Omega_n} c_k} \\
&= -z^{-1} \zeta_n^{(k)} \alpha_n^{(k)} \mathbf{1}_{\Omega_n}.
\end{aligned}
$$

(xii) The entries of c_k and r_k have mean zero, unit variance, and are bounded by $4L$ almost surely. In addition, (r_k^T, c_k) and $\mathcal{G}_n^{(k)} \mathbf{1}_{\Omega_n^{(k)}}$ are independent. By Remark 10.3, $\left\| \mathcal{G}_n^{(k)*} u_n v_n^* \mathcal{G}_n^{(k)*} \mathcal{G}_n^{(k)} v_n u_n^* \mathcal{G}_n^{(k)} \right\| \le 16$ on $\Omega_n^{(k)}$. Thus, by Lemma 10.7, for any $p \ge 2$,

$$
\begin{aligned}
\sup_{1 \le k \le n} \mathbb{E}_{k-1}[|\eta_n^{(k)}|^p \mathbf{1}_{\Omega_n}] &\le \sup_{1 \le k \le n} \mathbb{E}_{k-1}[|\eta_n^{(k)}|^p \mathbf{1}_{\Omega_n^{(k)}}] \\
&\ll_{L,p} \sup_{1 \le k \le n} n^{-p} \mathbb{E}_{k-1}\left[\left(\mathrm{tr}(\mathcal{G}_n^{(k)*} u_n v_n^* \mathcal{G}_n^{(k)*} \mathcal{G}_n^{(k)} v_n u_n^* \mathcal{G}_n^{(k)}) \right)^{p/2} \mathbf{1}_{\Omega_n^{(k)}} \right] \\
&\ll_{L,p} n^{-p}
\end{aligned}
$$

since

$$\mathcal{G}_n^{(k)*} u_n v_n^* \mathcal{G}_n^{(k)*} \mathcal{G}_n^{(k)} v_n u_n^* \mathcal{G}_n^{(k)}$$

is at most rank one. Similarly, Remark 10.3 give the almost sure bound $\left\| \mathcal{G}_n^{(k)*} \mathcal{G}_n^{(k)} \right\| \leq 4$ on $\Omega_n^{(k)}$, which gives

$$
\sup_{1 \leq k \leq n} \mathbb{E}_{k-1}[|\zeta_n^{(k)}|^p \mathbf{1}_{\Omega_n}] \leq \sup_{1 \leq k \leq n} \mathbb{E}_{k-1}[|\zeta_n^{(k)}|^p \mathbf{1}_{\Omega_n^{(k)}}]
$$

$$
\ll_{L,p} \sup_{1 \leq k \leq n} n^{-p} \mathbb{E}_{k-1}\left[\left(\operatorname{tr}(\mathcal{G}_n^{(k)*} \mathcal{G}_n^{(k)}) \right)^{p/2} \mathbf{1}_{\Omega_n^{(k)}} \right]
$$

$$
\ll_{L,p} n^{-p/2}.
$$

With the above observations in hand, we now complete the proof. By the resolvent identity (15) and Remark 10.3, it follows that the function

$$
z \mapsto |u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n} - \mathbb{E}\left[u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n} \right]|
$$

is Lipschitz continuous in the region $\{z \in \mathbb{C} : 5 \leq |z| \leq 6\}$. Thus, by a standard net argument, it suffices to prove that almost surely, for n sufficiently large,

$$
|u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n} - \mathbb{E}\left[u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n} \right]| < \varepsilon
$$

for each fixed $z \in \mathbb{C}$ with $5 \leq |z| \leq 6$. To this end, fix such a value of z. Throughout the proof, we drop the dependence on z in the resolvent as it is clear from context. Note that by Markov's inequality and the Borel–Cantelli lemma, it is sufficient to prove that

$$
\mathbb{E} |u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} - \mathbb{E}\left[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} \right]|^p = O_{L,p}(n^{-p/2}) \tag{49}
$$

for all $p > 2$. We now rewrite the above expression as a martingale difference sequence:

$$
u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} - \mathbb{E}[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n}] = \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n}].
$$

Since $u_n^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n^{(k)}}$ is independent of r_k and c_k, one can see that

$$
(\mathbb{E}_k - \mathbb{E}_{k-1})[u_n^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n^{(k)}}] = 0,
$$

and so

$$
u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} - \mathbb{E}[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n}] = \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} - u_n^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n^{(k)}} \right].
$$

In view of Lemma 10.9 and since $\Omega_n \cap \Omega_n^{(k)} = \Omega_n$, it suffices to prove that

$$
\mathbb{E} \left| \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} - u_n^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right] \right|^p = O_{L,p}(n^{-p/2})
$$

for all $p > 2$. Define

$$
W_k := (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} - u_n^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right] \tag{50}
$$

and observe that $\{W_k\}_{k=1}^{mn}$ is a martingale difference sequence with respect to $\{\mathcal{F}_k\}$.

From the resolvent identity (15), we observe that

$$\sum_{k=1}^{mn} W_k = \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n \frac{1}{\sqrt{n}} (\mathcal{Y}_n^{(k)} - \mathcal{Y}_n) \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right]$$

$$= -\sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n \left(\frac{1}{\sqrt{n}} e_k r_k + \frac{1}{\sqrt{n}} c_k e_k^* \right) \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right]$$

$$:= -\sum_{k=1}^{mn} (W_{k1} + W_{k2})$$

where we define

$$W_{k1} := (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n n^{-1/2} e_k r_k \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right],$$

and

$$W_{k2} := (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right].$$

By (i), (vii), and (ix), we can further decompose

$$\sum_{k=1}^{mn} W_{k1} = \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n n^{-1/2} e_k r_k \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right]$$

$$= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n^{(k2)} n^{-1/2} e_k r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]$$

$$= -\sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} (\bar{u}_{n,k} - u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k) n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]$$

$$= -\sum_{k=1}^{mn} (W_{k11} + W_{k12})$$

where

$$W_{k11} := (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} \bar{u}_{n,k} n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]$$

and

$$W_{k12} := - (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} u_n^* \mathcal{G}_n^{(k)} n^{-1} c_k r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right].$$

Similarly, by (i),(vi), (viii), and (x),

$$\sum_{k=1}^{mn} W_{k2} = \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right]$$

$$= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \frac{\mathcal{G}_n^{(k1)} n^{-1/2} c_k}{1 + r_k^* \mathcal{G}_n^{(k1)} c_k} e_k^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right]$$

$$= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \frac{\mathcal{G}_n^{(k1)} n^{-1/2} c_k}{1 + r_k^* \mathcal{G}_n^{(k1)} c_k \mathbf{1}_{\Omega_n}} e_k^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right]$$

$$= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]$$

$$= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]$$

$$= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} v_{n,k} u_n^* (\mathcal{G}_n^{(k)} - \mathcal{G}_n^{(k)} e_k n^{-1/2} r_k \mathcal{G}_n^{(k)}) n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]$$

$$= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} v_{n,k} (u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} - u_n^* \mathcal{G}_n^{(k)} e_k n^{-1} r_k \mathcal{G}_n^{(k)} c_k \alpha_n^{(k)}) \mathbf{1}_{\Omega_n} \right]$$

$$= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} v_{n,k} (u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} + \bar{u}_{n,k} z^{-1} n^{-1} r_k \mathcal{G}_n^{(k)} c_k \alpha_n^{(k)}) \mathbf{1}_{\Omega_n} \right]$$

$$= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n} - z^{-1} \bar{u}_{n,k} v_{n,k} \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]$$

$$= - \sum_{k=1}^{mn} (W_{k21} + W_{k22})$$

where

$$W_{k21} := (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]$$

and

$$W_{k22} := - (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} \bar{u}_{n,k} v_{n,k} \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right].$$

Thus, in order to complete the proof, it suffices to show that for all $p > 2$,

$$\mathbb{E} \left| \sum_{k=1}^{mn} W_{k11} \right|^p + \mathbb{E} \left| \sum_{k=1}^{mn} W_{k12} \right|^p + \mathbb{E} \left| \sum_{k=1}^{mn} W_{k21} \right|^p + \mathbb{E} \left| \sum_{k=1}^{mn} W_{k22} \right|^p = O_{L,p} \left(n^{-p/2} \right).$$

$$(51)$$

We bound each term individually. To begin, observe that by Lemma 10.4, Lemma 10.8, and (v), for any $p > 2$,

$$
\mathbb{E} \left| \sum_{k=1}^{mn} W_{k11} \right|^p = \mathbb{E} \left| \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[z^{-1} \bar{u}_{n,k} n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right] \right|^p
$$

$$
\ll_p \mathbb{E} \left[\sum_{k=1}^{mn} \mathbb{E}_{k-1} \left| (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} \bar{u}_{n,k} n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \right|^2 \right]^{p/2}
$$

$$
+ \mathbb{E} \left[\sum_{k=1}^{mn} \left| (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} \bar{u}_{n,k} n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \right|^p \right]
$$

$$
\ll_p \mathbb{E} \left[\sum_{k=1}^{mn} |\bar{u}_{n,k}|^2 \mathbb{E}_{k-1} \left| n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right|^2 \right]^{p/2}
$$

$$
+ \sum_{k=1}^{mn} |\bar{u}_{n,k}|^p \mathbb{E} \left| n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right|^p
$$

$$
\ll_p \mathbb{E} \left[\sum_{k=1}^{mn} |\bar{u}_{n,k}|^2 \mathbb{E}_{k-1} \left| n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n^{(k)}} \right|^2 \right]^{p/2}
$$

$$
+ \sum_{k=1}^{mn} |\bar{u}_{n,k}|^p \mathbb{E} \left| n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n^{(k)}} \right|^p
$$

$$
\ll_{L,p} \left(\mathbb{E} \left[\sum_{k=1}^{mn} |\bar{u}_{n,k}|^2 \cdot n^{-1} \right]^{p/2} + \sum_{k=1}^{mn} |\bar{u}_{n,k}|^p \cdot n^{-p/2} \right)
$$

$$
\ll_{L,p} n^{-p/2},
$$

where we also used Jensen's inequality and the fact that $|z| \geq 5$. Similarly, by Lemma 10.4, Lemma 10.8, and (v), for any $p > 2$,

$$
\begin{aligned}
\mathbb{E}\left|\sum_{k=1}^{mn} W_{k21}\right|^p &= \mathbb{E}\left|\sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1})\left[z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n}\right]\right|^p \\
&\ll_p \mathbb{E}\left[\sum_{k=1}^{mn} \mathbb{E}_{k-1}\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n}]\right|^2\right]^{p/2} \\
&\qquad + \mathbb{E}\left[\sum_{k=1}^{mn}\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n}]\right|^p\right] \\
&\ll_p \mathbb{E}\left[\sum_{k=1}^{mn} |v_{n,k}|^2 \mathbb{E}_{k-1}\left|u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \mathbf{1}_{\Omega_n}\right|^2\right]^{p/2} \\
&\qquad + \sum_{k=1}^{mn} |v_{n,k}|^p \mathbb{E}\left|u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \mathbf{1}_{\Omega_n}\right|^p \\
&\ll_p \mathbb{E}\left[\sum_{k=1}^{mn} |v_{n,k}|^2 \mathbb{E}_{k-1}\left|u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \mathbf{1}_{\Omega_n^{(k)}}\right|^2\right]^{p/2} \\
&\qquad + \sum_{k=1}^{mn} |v_{n,k}|^p \mathbb{E}\left|u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \mathbf{1}_{\Omega_n^{(k)}}\right|^p \\
&\ll_{L,p} \left(\mathbb{E}\left[\sum_{k=1}^{mn} |v_{n,k}|^2 \cdot n^{-1}\right]^{p/2} + \sum_{k=1}^{mn} |v_{n,k}|^p \cdot n^{-p/2}\right) \\
&\ll_{L,p} n^{-p/2}.
\end{aligned}
$$

Next, by Lemma 10.4, (v), and (xii), for any $p > 2$,

$$
\mathbb{E}\left|\sum_{k=1}^{mn} W_{k12}\right|^p = \mathbb{E}\left|\sum_{k=1}^{mn}(E_k - \mathbb{E}_{k-1})[z^{-1}u_n^*\mathcal{G}_n^{(k)}n^{-1}c_k r_k \mathcal{G}_n^{(k)}v_n\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]\right|^p
$$

$$
= \mathbb{E}\left|\sum_{k=1}^{mn}(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}\alpha_n^{(k)}\eta_n^{(k)}\mathbf{1}_{\Omega_n}]\right|^p
$$

$$
\ll_p \mathbb{E}\left[\sum_{k=1}^{mn}\mathbb{E}_{k-1}\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}\alpha_n^{(k)}\eta_n^{(k)}\mathbf{1}_{\Omega_n}]\right|^2\right]^{p/2}
$$

$$
+ \mathbb{E}\left[\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}\alpha_n^{(k)}\eta_n^{(k)}\mathbf{1}_{\Omega_n}]\right|^p\right]
$$

$$
\ll_p \left(\mathbb{E}\left[\sum_{k=1}^{mn}\mathbb{E}_{k-1}\left|\eta_n^{(k)}\mathbf{1}_{\Omega_n}\right|^2\right]^{p/2} + \sum_{k=1}^{mn}\mathbb{E}\left|\eta_n^{(k)}\mathbf{1}_{\Omega_n}\right|^p\right)
$$

$$
\ll_{L,p} \left(\mathbb{E}\left[\sum_{k=1}^{mn}n^{-2}\right]^{p/2} + \sum_{k=1}^{mn}n^{-p}\right)
$$

$$
\ll_{L,p} n^{-p/2} + n^{-p+1}
$$

$$
\ll_{L,p} n^{-p/2},
$$

where we also used Jensen's inequality and the fact that $|z| \geq 5$; the last inequality follows from the fact that $p > 2$.

Finally, moving on to W_{k22}, by (xi) we can decompose this further as

$$
\sum_{k=1}^{mn} W_{k22} = \sum_{k=1}^{mn}(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}\bar{u}_{n,k}v_{n,k}\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]
$$

$$
= \sum_{i=1}^{mn}(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}\bar{u}_{n,k}v_{n,k}(1 - z^{-1}\zeta_n^{(k)}\alpha_n^{(k)}\mathbf{1}_{\Omega_n})]
$$

$$
= \sum_{k=1}^{mn}(W_{k221} + W_{k222})
$$

where

$$
W_{k221} := (\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}\bar{u}_{n,k}v_{n,k}]
$$

and

$$
W_{k222} := -(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-2}\bar{u}_{n,k}v_{n,k}\zeta_n^{(k)}\alpha_n^{(k)}\mathbf{1}_{\Omega_n}].
$$

Since $z^{-1}\bar{u}_{n,k}v_{n,k}$ is deterministic, it follows that

$$
\left|\sum_{k=1}^{mn} W_{k221}\right|^p = 0.
$$

Thus, it suffices to show that $\mathbb{E}\left|\sum_{k=1}^{mn} W_{k222}\right|^p = O_{L,p}(n^{-p/2})$ for $p > 2$. By Lemma 10.4, (v), and (xii), we have that for any $p > 2$,

$$
\mathbb{E}\left|\sum_{k=1}^{mn} W_{k222}\right|^p = \mathbb{E}\left|\sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-2}\bar{u}_{n,k}v_{n,k}\zeta_n^{(k)}\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]\right|^p
$$

$$
\ll_p \mathbb{E}\left[\sum_{k=1}^{mn} \mathbb{E}_{k-1}\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-2}\bar{u}_{n,k}v_{n,k}\zeta_n^{(k)}\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]\right|^2\right]^{p/2}
$$

$$
+ \mathbb{E}\left[\sum_{k=1}^{mn}\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-2}\bar{u}_{n,k}v_{n,k}\zeta_n^{(k)}\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]\right|^p\right]
$$

$$
\ll_p \mathbb{E}\left[\sum_{k=1}^{mn}|\bar{u}_{n,k}|^2|v_{n,k}|^2\mathbb{E}_{k-1}\left|\zeta_n^{(k)}\mathbf{1}_{\Omega_n}\right|^2\right]^{p/2} + \sum_{k=1}^{mn}|\bar{u}_{n,k}|^p|v_{n,k}|^p\mathbb{E}\left|\zeta_n^{(k)}\mathbf{1}_{\Omega_n}\right|^p
$$

$$
\ll_{L,p} \left[\sum_{k=1}^{mn}|\bar{u}_{n,k}|^2|v_{n,k}|^2 n^{-1}\right]^{p/2} + \sum_{k=1}^{mn}|\bar{u}_{n,k}|^p|v_{n,k}|^p n^{-p/2}
$$

$$
\ll_{L,p} n^{-p/2},
$$

where we again used Jensen's inequality, the bound $|z| \geq 5$, and the fact that u_n and v_n are unit vectors. This completes the proof of (51), and hence the proof of Lemma 10.2 is complete. $\qquad\square$

11. Proof of Theorem 9.1

In this section we complete the proof of Theorem 9.1. We continue to work under the assumptions and notation introduced in Section 10.

It remains to prove part (iii) of Theorem 9.1. In view of Lemma 10.2 and (36), it suffices to show that

$$
\sup_{5 \leq |z| \leq 6}\left|\mathbb{E}[u_n^*\mathcal{G}_n(z)v_n\mathbf{1}_{\Omega_n}] + \frac{1}{z}u_n^*v_n\right| = o(1). \tag{52}
$$

11.1. **Neumann series.** We will rewrite the resolvent, $\mathcal{G}_n$, as a Neumann series. Indeed, for $|z| \geq 5$,

$$
\frac{1}{\sqrt{n}}\left\|\frac{\mathcal{Y}_n}{z}\right\| \leq \frac{9}{10} < 1 \tag{53}
$$

on the event Ω_n, so we may write

$$
\mathcal{G}_n(z)\mathbf{1}_{\Omega_n} = -\frac{1}{z}\left(I\mathbf{1}_{\Omega_n} + \sum_{k=1}^{\infty}\left(\frac{1}{\sqrt{n}}\frac{\mathcal{Y}_n\mathbf{1}_{\Omega_n}}{z}\right)^k\right) = -\frac{1}{z}I\mathbf{1}_{\Omega_n} - \sum_{k=1}^{\infty}\frac{\left(\frac{\mathcal{Y}_n}{\sqrt{n}}\mathbf{1}_{\Omega_n}\right)^k}{z^{k+1}}
$$

almost surely. Therefore, using (36), we obtain

$$\mathbb{E}\left[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n}\right] = \mathbb{E}\left[-\frac{1}{z}u_n^* v_n \mathbf{1}_{\Omega_n} - u_n^* \sum_{k=1}^{\infty} \frac{\left(\frac{\mathcal{Y}_n}{\sqrt{n}}\right)^k}{z^{k+1}} v_n \mathbf{1}_{\Omega_n}\right]$$

$$= -\frac{1}{z}u_n^* v_n \mathbb{P}(\Omega_n) - \sum_{k=1}^{\infty} \frac{1}{z^{k+1}} \mathbb{E}\left[u_n^* \left(\frac{\mathcal{Y}_n}{\sqrt{n}}\right)^k v_n \mathbf{1}_{\Omega_n}\right]$$

$$= -\frac{1}{z}u_n^* v_n + o(1) - \sum_{k=1}^{\infty} \frac{1}{z^{k+1}} \mathbb{E}\left[u_n^* \left(\frac{\mathcal{Y}_n}{\sqrt{n}}\right)^k v_n \mathbf{1}_{\Omega_n}\right]. \qquad (54)$$

Showing the sum in (54) converges to zero uniformly in the region $\{z \in \mathbb{C} : 5 \leq |z| \leq 6\}$ will complete the proof of (52).

11.2. **Removing the indicator function.** In this subsection, we prove the following.

Lemma 11.1. *Under the assumptions above, for any integer $k \geq 1$,*

$$\left|\mathbb{E}\left[u_n^* \left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v_n\right] - \mathbb{E}\left[u_n^* \left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v_n \mathbf{1}_{\Omega_n}\right]\right| = o_{k,L}(1).$$

Proof. Since the entries of $\mathcal{Y}_n$ are truncated, it follows that

$$\|\mathcal{Y}_n\| \leq \|\mathcal{Y}_n\|_2 \ll_L n$$

almost surely. Therefore, as $\mathbb{P}(\Omega_n^c) = O_A(n^{-A})$ for any $A > 0$, we obtain

$$\left|\mathbb{E}\left[u_n^* \left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v_n\right] - \mathbb{E}\left[u_n^* \left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v_n \mathbf{1}_{\Omega_n}\right]\right| \leq \mathbb{E}\left[\|\mathcal{Y}_n\|^k \mathbf{1}_{\Omega_n^c}\right]$$

$$\ll_L n^k \mathbb{P}(\Omega_n^c)$$

$$\ll_{L,A} n^{k-A}.$$

Choosing A sufficiently large (in terms of k), completes the proof. $\qquad \square$

11.3. **Combinatorial arguments.** In this section, we will show that

$$\sup_{5 \leq |z| \leq 6} \left|\sum_{k=1}^{\infty} \frac{\mathbb{E}\left[u_n^* \left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v_n \mathbf{1}_{\Omega_n}\right]}{z^{k+1}}\right| = o_L(1). \qquad (55)$$

In view of (53), the tail of the series can easily be controlled. Thus, it suffices to show that, for each integer $k \geq 1$,

$$\mathbb{E}\left[u_n^* \left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v_n \mathbf{1}_{\Omega_n}\right] = o_{L,k}(1).$$

By Lemma 11.1, it suffices to prove the statement above without the indicator function. In particular, the following lemma completes the proof of (52) (and hence completes the proof of Theorem 9.1).

Lemma 11.2 (Moment Calculations). *Under the assumptions above, for any integer $k \geq 1$,*

$$\mathbb{E}\left[u_n^*\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v_n\right] = o_{L,k}(1).$$

To prove Lemma 11.2, we will expand the above expression in terms of the entries of the random matrices $X_{n,1},\ldots,X_{n,m}$. For brevity, in this section we will drop the subscript n from our notation and just write $X_1,\ldots,X_m$ for $X_{n,1},\ldots,X_{n,m}$. Similarly, we write the vectors u_n and v_n as u and v, respectively.

To begin, we exploit the block structure of $\mathcal{Y}_n$ and write

$$\mathbb{E}\left[u^*\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v\right] = n^{-k/2}\mathbb{E}\left[\sum_{1 \leq a,b \leq m}(u^*)^{[a]}\left(\mathcal{Y}_n^k\right)^{[a,b]}v^{[b]}\right]$$

$$= n^{-k/2}\sum_{1 \leq a,b \leq m}(u^*)^{[a]}\mathbb{E}\left[\left(\mathcal{Y}_n^k\right)^{[a,b]}\right]v^{[b]}. \tag{56}$$

Due to the block structure of $\mathcal{Y}_n$, for each $1 \leq a \leq m$, there exists some $1 \leq b \leq m$ which depends on a and k such that the $[a,b]$ block of $\mathcal{Y}_n^k$ is nonzero and all other blocks, $[a,c]$ for $c \neq b$, are zero. Additionally, the nonzero block entry $(\mathcal{Y}_n^k)^{[a,b]}$ is $X_a X_{a+1} \cdots X_{a+k}$ where the subscripts are reduced modulo m and we use modular class representatives $\{1,\ldots,m\}$ (as opposed to the usual $\{0,..,m-1\}$). To show that the expectation of this sum is $o_{L,k}(1)$, we need to systematically count all terms which have a nonzero expectation. To do so, we use graphs to characterize the terms which appear in the sum. In particular we develop path graphs, each of which corresponds uniquely to a term in the expansion of (56). For each term, the corresponding path graph will record the matrix entries which appear, the order in which they appear, and the matrix from which they come. We begin with the following definitions.

Definition 11.3. We consider graphs where each vertex is specified by the ordered pair (t, i_t) for integers $1 \leq t \leq k+1$ and $1 \leq i_t \leq n$. We call t the *time* coordinate of the vertex and i_t the *height* coordinate of the vertex. Let $(i_1, i_2, \ldots, i_{k+1})$ be a $k+1$ tuple of integers from the set $\{1, 2, \ldots, n\}$ and let $1 \leq a \leq m$. We define an *m-colored k-path graph* $G^a(i_1, i_2, \ldots, i_{k+1})$ to be the edge-colored directed graph with vertex set $V = \{(1, i_1), (2, i_2), \ldots, (k+1, i_{k+1})\}$ and directed edges from (t, i_t) to $(t+1, i_{t+1})$ for $1 \leq t \leq k$, where the edge from (t, i_t) to $(t+1, i_{t+1})$ is color $a + t - 1$, with the convention that colors are reduced modulo m, and the modulo class representatives are $\{1, \ldots, m\}$. The graph is said to *visit* a vertex (t, i_t) if there exists an edge which begins or terminates at that vertex, so that $(t, i_t) \in V$. We say an edge of $G^a(i_1, \ldots, i_{k+1})$ is of *type I* if it terminates on a vertex (t, i_t) such that $i_t \neq i_s$ for all $s < t$. We say an edge is of *type II* if it terminates on a vertex (t, i_t) such that there exists some $s < t$ with $i_t = i_s$.

We make some observations about this definition.

Remark 11.4. We view k and n as specified parameters. Once these parameters are specified, the notation $G^a(i_1, \ldots, i_{k+1})$ completely determines the graph. The vertex set V is a subset of the vertices of the $(k+1) \times n$ integer lattice and each graph has exactly k directed edges. In each graph, there is an edge which begins on vertex $(1, i_1)$ for some $1 \leq i_1 \leq n$ and there is an edge which terminates on vertex

$(k + 1, i_{k+1})$ for some $1 \leq i_{k+1} \leq n$. Each edge begins at (t, i_t) and terminates on $(t + 1, i_{t+1})$ for some integers $1 \leq t \leq k$ and $1 \leq i_t, i_{t+1} \leq n$. Additionally, each edge is one of m possible colors. Note that if we think about the edges as ordered by the time coordinate, then the order of the colors is a cyclic permutation of the coloring $1, 2, \ldots, m$, beginning with a. This cycle is repeated as many times as necessary in order to cover all edges.

Notice that we call $G^a(i_1, i_2, \ldots, i_{k+1})$ a path graph because it can be thought of as a path through the integer lattice from vertex $(1, i_1)$ to $(k + 1, i_{k+1})$ for some $1 \leq i_1, i_{k+1} \leq n$. Indeed, by the requirements in the definition, this graph must be one continuous path and no vertex can be visited more than once. We may call an m-colored k-path graph a *path graph* when m and k are clear from context.

Finally, we can think of an edge as type I if it terminates at a height not previously visited. It is of type II if it terminates at a height that has been previously visited.

Definition 11.5. For a given m-colored k-path graph $G^a(i_1, i_2, \ldots, i_{k+1})$, we say two edges e_1 and e_2 of $G^a(i_1, i_2, \ldots, i_{k+1})$ are *time-translate parallel* if e_1 begins at vertex (t, i_t) and terminates at vertex $(t + 1, i_{t+1})$ and edge e_2 begins at vertex $(t', i_{t'})$ and terminates at vertex $(t' + 1, i_{t'+1})$ where $i_t = i_{t'}$ and $i_{t+1} = i_{t'+1}$ for some $1 \leq t, t' \leq k$ with $t \neq t'$.

Remark 11.6. Intuitively, two edges are time-translate parallel if they span the same two hight coordinates at different times. Throughout this section, we shorten the term "time-translate parallel" and refer to edges with this property as "parallel" for brevity. We warn the reader that by this, we mean that the edges must span the same heights at two different times. For instance, the edge from $(1, 2)$ to $(2, 4)$ is not parallel to the edge from $(3, 1)$ to $(4, 3)$ since they don't span the same heights, although these edges might appear parallel in the geometric interpretation of the word. Also note that two parallel edges need not have the same color. See Figure 6 for examples of parallel and non parallel edges.

Definition 11.7. For a fixed k, we say two m-colored k-path graphs $G^a(i_1, \ldots, i_{k+1})$ and $G^a(i'_1, \ldots, i'_{k+1})$ are *equivalent*, denoted $G^a(i_1, \ldots, i_{k+1}) \sim G^a(i'_1, \ldots, i'_{k+1})$, if there exists some permutation σ of $\{1, \ldots, n\}$ such that $(i_1, \ldots, i_{k+1}) = (\sigma(i'_1), \ldots, \sigma(i'_{k+1}))$. Note here that for two path graphs to be equivalent, the color of the first edge, and hence the color of all edges sequentially, must be the same in both.

One can check that the above definition of equivalent m-colored k-path graphs is an equivalence relation. Thus, the set of all m-colored k-path graphs can be split into equivalence classes.

Definition 11.8. For each equivalence class of graphs, the *canonical m-colored k-path graph* is the unique graph from an equivalence class which satisfies the following condition: If $G^a(i_1, \ldots, i_{k+1})$ visits vertex (t, i_t), then for every $0 < i < i_t$ there exists $s < t$ such that $G^a(i_1, \ldots, i_{k+1})$ visits (s, i).

Remark 11.9. Observe that, intuitively, the canonical representation for $G^a(i_1, \ldots, i_{k+1})$ is the graph which does not "skip over" any height coordinates. Namely, the canonical graph necessarily begins at vertex $(1, 1)$ and at each time step the height of the next vertex can be a most one larger than the maximum height of all previous vertices.
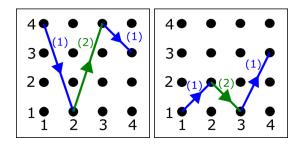
FIGURE 4. The graphs featured here are two possible 2-colored 3-path graphs, $G^1(4, 1, 4, 3)$ and $G^1(1, 2, 1, 3)$, which correspond to the terms $X_{1,(4,1)}X_{2,(1,4)}X_{1,(4,3)}$ and $X_{1,(1,2)}X_{2,(2,1)}X_{1,(1,3)}$, respectively. These two graphs are equivalent. The first graph is not a canonical graph while the second graph is a canonical graph. For ease of notation and clarity, we have drawn all vertices on the integer lattice as dots, with the time axis appearing horizontally and the height axis vertically. While each dot represents a vertex in the integer latter, not all of the dots represent vertices in the path graphs. Only dots from which an edge begins or terminates are vertices of the path graph. We have also colored the edges with blue and green to represent the two colors. The colors of each edge are also represented by the number in parenthesis, e.g., (1) or (2). In either graph, the first and third edges are type I edges, while the second edge in each graph is type II. See Example 11.12 for further discussion.

Now, we return to the task at hand: the proof of Lemma 11.2. We fix a positive integer k, and we expand as in (56).

For a fixed value $1 \le a \le m$, consider the nonzero block $\left(\mathcal{Y}_n^k\right)^{[a,b]}$. We further expand to see

$$(u^*)^{[a]}\mathbb{E}\left[\left(\mathcal{Y}_n^k\right)^{[a,b]}\right]v^{[b]} = \sum_{i_1,\dots,i_{k+1}} \overline{u}_{i_1}^{[a]}\mathbb{E}[X_{a,(i_1,i_2)}X_{a+1,(i_2,i_3)}\cdots X_{a+k,(i_k,i_{k+1})}]v_{i_{k+1}}^{[b]},$$

(57)

where the subscripts $a, \dots, a+k$ are reduced mod m with representatives $\{1, \dots, m\}$. Observe that by the structure of $\mathcal{Y}_n$, these subscripts must appear cyclically in the order $a, a+1, \dots, m, 1, \dots, a-1$, with the order repeating as many times as necessary before ending at b. In particular, the subscripts are uniquely determined by the starting subscript a and the value of k.

We now consider the expectation on the right-hand side of (57). Since all entries of each matrix are independent, if an index appears only once in a product, that product will have expectation zero. Therefore, only terms in which every index appears more than once will have a nonzero expectation, and only such terms will contribute to the expected value of the sum. Note that for an entry to appear more than once, we not only need the index of the entries to match but also which of the m matrices the entries came from. The following definition will assist in encoding each entry on the right-hand side of (57) as a unique m-colored k-path graph.

Definition 11.10. We say the term $X_{a,(i_1,i_2)}X_{a+1,(i_2,i_3)}\cdots X_{a+k,(i_k,i_{k+1})}$ from the expansion of (57) *corresponds* to the m-colored k-path graph $G^a(i_1,\ldots,i_{k+1})$. We use the notation

$$x_G := X_{a,(i_1,i_2)}X_{a+1,(i_2,i_3)}\cdots X_{a+k,(i_k,i_{k+1})}$$

whenever $X_{a,(i_1,i_2)}X_{a+1,(i_2,i_3)}\cdots X_{a+k,(i_k,i_{k+1})}$ corresponds to the path graph G. In this case, we also write u_G^* and v_G for $\bar{u}_{i_1}^{[a]}$ and $v_{i_{k+1}}^{[a+k]}$, respectively.

Each term in the expansion of (57) corresponds uniquely to an m-colored k-path graph. In terms of the corresponding m-colored k-path graph, if an edge spans two vertices (t,i_t) and $(t+1,i_{t+1})$, and has color a then the corresponding matrix product must contain the entry $X_{a,(i_t,i_{t+1})}$. Thus, the color corresponds to the matrix from which the entry came and the height coordinates correspond to the matrix indices. Repeating indices is analogous to parallel edges, and entries coming from the same matrix corresponds to edges sharing a color. For example, if $X_{4,(3,5)}$ appears at some point in a term, the corresponding m-colored k-path graph will have an edge from $(t,3)$ to $(t+1,5)$ for some t, and the edge will be colored with color 4. Thus, a graph corresponds to a term with nonzero expectation if for every edge e_1, there exists at least one other edge in the graph which is parallel to e_1 and which has the same color as e_1. We must systematically count the terms which have nonzero expectation.

Since two equivalent graphs correspond to two terms which differ only by a permutation of indices, and since entries in a given matrix are independent and identically distributed, the expectation of the corresponding terms will be equal. This leads to the following lemma.

Lemma 11.11. *If two path graphs $G^a(i_1,\ldots,i_{k+1})$ and $G^a(i_1',\ldots,i_{k+1}')$ are equivalent, then*

$$\mathbb{E}[x_{G^a(i_1,\ldots,i_{k+1})}] = \mathbb{E}[x_{G^a(i_1',\ldots,i_{k+1}')}].$$

This lemma allows us to characterize graphs with non-zero expectation based on their canonical representation. Before we begin counting the graphs which correspond to terms with nonzero expectation, we present some examples.

Example 11.12. Consider two 2-colored 3-path graphs: $G^1(4,1,4,3)$ and $G^1(1,2,1,3)$. $G^1(4,1,4,3)$ is the leftmost graph in Figure 4 and $G^1(1,2,1,3)$ is the rightmost graph in Figure 4. They correspond to the terms $X_{1,(4,1)}X_{2,(1,4)}X_{1,(4,3)}$ and $X_{1,(1,2)}X_{2,(2,1)}X_{1,(1,3)}$ respectively from the expansion of $u^*\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^3 v$ where $m=2$. Note that these graphs are equivalent by the permutation which maps $4\mapsto 1\mapsto 2\mapsto 4$. $G^1(4,1,4,3)$ is not a canonical graph, while $G^1(1,2,1,3)$ is a canonical graph. Observe that since $X_{1,(4,1)}$ appeared first in the product $X_{1,(4,1)}X_{2,(1,4)}X_{1,(4,3)}$, the first edge in the corresponding path graph is an edge of color 1 spanning from height coordinate 4 to height coordinate 1.

Example 11.13. Consider a product of the form

$$X_{1,(i_1,i_2)}X_{1,(i_2,i_1)}X_{1,(i_1,i_2)}X_{1,(i_2,i_1)}X_{1,(i_1,i_2)}X_{1,(i_2,i_1)}, X_{1,(i_1,i_2)}$$

where i_1,i_2 are distinct. This corresponds to a 1-colored 7-path graph whose canonical representative can be seen in Figure 5. Since all entries come from matrix X_1, we know $m=1$. Since there are 7 terms in the product, $k=7$. This term

has expected value $\mathbb{E}\left[(X_{1,(i_1,i_2)})^4\right]\mathbb{E}\left[(X_{1,(i_2,i_1)})^3\right]$. Since X_1 is an iid matrix, the particular choice of $i_1 \neq i_2$ is irrelevant to the expected value.
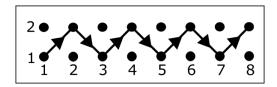

FIGURE 5. This 1-colored 7-path graph corresponds to the product in Example 11.13. Since all edges are the same color, they are not labeled with distinct colors. Observe that every edge in this graph is parallel to at least one other edge.

Example 11.14. Consider a product of the form

$$X_{1,(i_1,i_2)}X_{1,(i_2,i_3)}X_{1,(i_3,i_1)}X_{1,(i_1,i_3)}X_{1,(i_3,i_4)}X_{1,(i_4,i_3)}X_{1,(i_3,i_4)}$$

where i_1, i_2, i_3 and i_4 are distinct. This corresponds to a 1-colored 7-path graph whose canonical representative is featured in Figure 6. Since the only index pair which is appears more than once is (i_3, i_4), the corresponding term will have zero expectation.
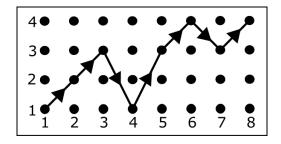
FIGURE 6. This 1-colored 7-path graph is the canonical representative to the graph which corresponds to the product in Example 11.14. Note that the edge from $(5,3)$ to $(6,4)$ is parallel to the edge from $(7,3)$ to $(8,4)$, but no other edges in the graph are parallel. This implies that the corresponding term will have zero expectation.

Example 11.15. Let i_1, i_2, i_3, i_4 be distinct, and consider the product

$$X_{1,(i_1,i_2)}X_{2,(i_2,i_1)}X_{3,(i_1,i_2)}X_{4,(i_2,i_1)}X_{1,(i_1,i_2)}X_{2,(i_2,i_1)}X_{3,(i_1,i_2)}X_{4,(i_2,i_1)}.$$

Since there are 8 entries in this product, $k = 8$, and as there are entries from 4 matrices, $m = 4$. The corresponding canonical 4-colored 8-path graph representative is shown in Figure 7. This term has expectation

$$\mathbb{E}\left[(X_{1,(i_1,i_2)})^2(X_{2,(i_2,i_1)})^2(X_{3,(i_1,i_2)})^2(X_{4,(i_2,i_1)})^2\right].$$
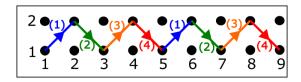
FIGURE 7. This 4-colored 8-path graph corresponds to the product in Example 11.15. Since there are eight edges, this is one possible product from $u^* \left(n^{-1/2} \mathcal{Y}_n \right)^8 v$.

Example 11.16. Consider the product

$$X_{3,(i_1,i_2)} X_{1,(i_2,i_3)} X_{2,(i_3,i_4)} X_{3,(i_4,i_1)} X_{1,(i_1,i_2)} X_{2,(i_2,i_4)} X_{3,(i_4,i_2)} X_{1,(i_2,i_1)} X_{2,(i_1,i_5)}.$$

Since there are 9 entries in this product, $k = 9$ and we can see that there are 3 different matrices so that $m = 3$. For any distinct indices i_1, i_2, i_3, i_4, i_5, the corresponding canonical 3-colored 9-path graph representative is shown in Figure 8. In this product, every entry appears only once. Thus the expectation of this product factors, and the term will have expectation zero.
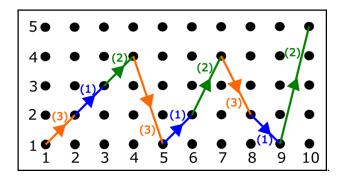

FIGURE 8. This 3-colored 9-path graph is the canonical representative of the graph corresponding to the term in Example 11.16. There are nine edges that appear in this graph and 3 colors on those edges, indicating that $k = 9$ and $m = 3$. Since no edges are parallel to another edge of the same color, this path graph corresponds to a term which has zero expectation.

We now complete the proof of Lemma 11.2, which will occupy the remainder of the section. Let

$$\Delta_{n,k}^a := \{ G^a(i_1, \ldots, i_{k+1}) \ : \ 1 \le i_1, \ldots, i_{k+1} \le n \},$$

and let $\tilde{\Delta}_{n,k}^a$ be the set of all canonical graphs in $\Delta_{n,k}^a$. We now divide the proof into cases based on the value of k.

Case where k is a multiple of m: If k is a multiple of m, then by the block structure of $\mathcal{Y}_n$, it follows that $\mathcal{Y}_n^k$ is a block diagonal matrix. Since the

diagonal blocks are the only nonzero blocks in this case, (56) simplifies to

$$\mathbb{E}\left[u^*\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n\right)^k v\right] = n^{-k/2}\sum_{1\le a\le m}(u^*)^{[a]}\mathbb{E}\left[(\mathcal{Y}_n^k)^{[a,a]}\right]v^{[a]}$$

$$= n^{-k/2}\sum_{1\le a\le m}\sum_{G\in\Delta_{n,k}^a}u_G^*\mathbb{E}[x_G]v_G \qquad (58)$$

Recall that if $G\in\Delta_{n,k}^a$, then G is an m-colored k-path graph which starts with color, a, i.e., $G = G^a(i_1,\ldots,i_{k+1})$ for some $i_1,\ldots,i_{k+1}\in[n]$. By Lemma 11.11, we can reduce the task of counting all terms with nonzero expectation to counting canonical graphs and the cardinality of each equivalence class.

Observe that if $k = m$, then any term in (58) will be of the form

$$\overline{u}_{i_1}\mathbb{E}[X_{a,(i_1,i_2)}X_{a+1,(i_2,i_3)}\cdots X_{a-1,(i_k,i_{k+1})}]v_{i_{k+1}}$$

where each matrix contributes only one entry to the above expression. In this case, all terms are independent and the expectation in (58) is zero.

Now consider the case where $k = cm$ for some integer $c\ge 2$. Define

$$h(G) =: \max\{i_t \ : \ (t,i_t)\in V_{\tilde{G}}\}, \qquad (59)$$

where $\tilde{G}$ is the canonical representative for the graph G and $V_{\tilde{G}}$ is the vertex set for the graph $\tilde{G}$. We call $h(G)$ the *maximal height* (or sometimes just *height*) of a graph G. Intuitively, $h(G)$ is the number of distinct height coordinates G visits. In terms of the canonical graph, $\tilde{G}$, this is the largest height coordinate visited by an edge in $\tilde{G}$. For each a, define

$$(\Delta_{n,k}^a)_1 := \{G\in\Delta_{n,k}^a \ : \ h(G) > k/2\} \qquad (60)$$

$$(\Delta_{n,k}^a)_2 := \{G\in\Delta_{n,k}^a \ : \ h(G) = k/2\} \qquad (61)$$

$$(\Delta_{n,k}^a)_3 := \{G\in\Delta_{n,k}^a \ : \ h(G) < k/2\}. \qquad (62)$$

This partitions $\Delta_{n,k}^a$ into disjoint subsets. Without loss of generality, we assume that $a = 1$ since the argument will be the same for any permutation of the coloring. In this case, all path graphs start with color 1 and since $k = cm$, the colors $1, 2, \ldots, m$ will each repeat c times. We analyze each set of graphs separately.

Graphs in $(\Delta_{n,k}^1)_1$:

First, consider the set $(\Delta_{n,k}^1)_1$ and recall that each $G\in(\Delta_{n,k}^1)_1$ must have exactly k edges. Since the expectation of all equivalent graphs is the same, it is sufficient to assume that G is canonical. If $h(G) > k/2$, there must be more than $k/2$ type I edges. If each edge were parallel to at least one other edge in G, then there would be more than k edges, a contradiction. Hence there will be at least one edge that is not parallel to any other edge. This implies $\mathbb{E}[x_G] = 0$ whenever $G\in(\Delta_{n,k}^1)_1$ and thus

$$\sum_{G\in(\Delta_{n,k}^a)_1}u_G^*\mathbb{E}[x_G]v_G = 0. \qquad (63)$$

Graphs in $(\Delta_{n,k}^1)_2$:

Note that if k is odd, then this set will be empty; so assume k is even. Now consider a graph $G \in (\Delta^1_{n,k})_2$. By Lemma 11.11, we can assume that G is canonical. If G has any edges which are not parallel to any other edges, then $\mathbb{E}[x_G] = 0$ and it does not contribute to the expectation. Thus we can consider only graphs in which every edge is parallel to at least one other edge. Since any $G \in (\Delta^1_{n,k})_2$ must visit exactly $k/2$ distinct height coordinates and since there must be precisely k edges in G, a counting argument reveals that every edge in G must be parallel to exactly one other edge in G. This gives way to the following lemma.

Lemma 11.17. *Let $k \geq 2$ be any even integer (not necessarily a multiple of m). Then there is only one canonical k-path graph in $\Delta^1_{n,k}$ for which $h(G) = \frac{k}{2}$ and in which each edge is parallel to exactly one other edge.*

The proof of this lemma, which relies on a counting argument and induction, is detailed in Appendix C. In fact, the proof reveals that this one canonical m-colored k-path graph starting with color 1 is

$$G^1(1, 2, \ldots, k/2, 1, 2, \ldots, k/2, 1).$$

If two edges are parallel but are not the same color then the expectation of terms with corresponding canonical graph will be zero.

If c is odd and m is even, then the edge from $(k/2, k/2)$ to $(k/2+1, 1)$ will have color $\frac{m}{2}$ and thus edge from $(k/2+1, 1)$ to $(k/2+2, 2)$ will have color $\frac{m}{2}+1$. This edge is necessarily parallel to the edge from $(1, 1)$ to $(2, 2)$, and it is the only edge parallel to the edge from $(1, 1)$ to $(2, 2)$. But note that the edge from $(1, 1)$ to $(2, 2)$ had color 1 and $\frac{m}{2}+1$ is not congruent to 1 mod m. Therefore in the case where c is odd and m is even, the canonical m-colored k-path graph corresponds to a term in the product which has expectation zero.

Finally, if c is even, then the edge from $(k/2, k/2)$ to $(k/2+1, 1)$ must have color m. Hence the edge from $(k/2+1, 1)$ to $(k/2+2, 2)$ will have color 1, which is the same color as the edge from $(1, 1)$ to $(2, 2)$. This means that when $k = cm$ and c is even, every edge in $G^1(1, 2, \ldots, k/2, 1, 2, \ldots, k/2, 1)$ will be parallel to exactly one other edge of the same color. In particular, note that for this graph

$$
\begin{aligned}
|u_G^* \mathbb{E}[x_G] v_G| &\leq \left| \bar{u}_1^{[1]} \right| \mathbb{E} \left| X_{1,(1,2)} \cdots X_{m,(k/2,1)} \right| \left| v_1^{[m]} \right| \\
&\leq \left| \bar{u}_1^{[1]} \right| \mathbb{E} \left| X_{1,(1,2)} \right|^2 \cdots \mathbb{E} \left| X_{m,(k/2,1)} \right|^2 \left| v_1^{[m]} \right| \\
&\leq \left| \bar{u}_1^{[1]} \right| \left| v_1^{[m]} \right|.
\end{aligned}
\tag{64}
$$

For ease of notation, let $\tilde{G} := G^1(1, 2, \ldots, k/2, 1, 2, \ldots, k/2, 1)$, and consider $G^1(i_1, \ldots, i_{k+1}) \in (\Delta^1_{n,k})_2$ such that $G^1(i_1, \ldots, i_{k+1}) \sim \tilde{G}$. Observe that there are n options for the first coordinate i_1 of $G^1(i_1, \ldots, i_{k+1})$. If we fix the first coordinate, then there are at most $(n-1)(n-2) \cdots (n-k/2-1) \leq n^{k/2-1}$ graphs with first coordinate i_1 which are equivalent to $\tilde{G}$. If we repeated the computation of the expectation of any of these equivalent graphs, we would get a term similar to (64) but with different starting and ending coordinates, yielding an upper bound of $|\bar{u}_{i_1}||v_{i_1}|$. Therefore, by the above

argument and the Cauchy–Schwarz inequality, we obtain

$$\left| \sum_{G \in (\Delta^1_{n,k})_2} u_G^* \mathbb{E}[x_G] v_G \right| \le \sum_{1 \le i_1 \le n} n^{k/2-1} |\overline{u}_{i_1}| |v_{i_1}|$$
$$\le n^{k/2-1} \|u\| \|v\|$$
$$\le n^{k/2-1}. \tag{65}$$

Graphs in $(\Delta^1_{n,k})_3$:

Consider an m-colored k-path graphs $G \in (\Delta^1_{n,k})_3$, and assume that G is canonical. If G contains any edges which were not parallel to another edge, then the graph will correspond to a term with expectation zero. So consider a canonical graph $G \in (\Delta^1_{n,k})_3$ such that all edges are parallel to at least one other edge. If $h(G) = 1$, then $G = G^1(1, 1, \ldots, 1)$ and so

$$|\mathbb{E}[x_G]| \le \mathbb{E}\left| X_{1,(1,1)} \cdots X_{1,(1,1)} \right| = \mathbb{E}\left| X_{1,(1,1)} \right|^k \le (4L)^k.$$

Note that this is the highest possible moment in a term. Let $M := (4L)^k$. For any canonical m-colored k-path graph $G \in (\tilde{\Delta}^1_{n,k})_3$,

$$\mathbb{E}|x_G| \le M.$$

Also note that this bound holds for graphs of all starting colors, not just starting color 1. In addition, for any G with maximal height $h(G)$, there are $n(n-1) \cdots (n-h(G)-1) < n^{h(G)}$ graphs in the equivalence class of G. By over counting, we can bound the number of distinct equivalence classes by k^k since there are k edges and at each time coordinate, the edge which starts at that time coordinate can terminate at most one height coordinate larger than it started, so any edge has a most k options for an ending coordinate.

Based on the above observations, we have

$$\left| \sum_{G \in (\Delta^1_{n,k})_3} u_G^* \mathbb{E}[x_G] v_G \right| \le \sum_{G \in (\tilde{\Delta}^1_{n,k})_3} \left| n^{h(G)} u_G^* \mathbb{E}[x_G] v_G \right|$$
$$\le n^{\frac{k}{2}-\frac{1}{2}} \sum_{G \in (\tilde{\Delta}^1_{n,k})_3} |u_G^*| \mathbb{E}|x_G| |v_G|$$
$$\le n^{\frac{k}{2}-\frac{1}{2}} \sum_{G \in (\tilde{\Delta}^1_{n,k})_3} M$$
$$\ll_{L,k} n^{\frac{k}{2}-\frac{1}{2}}. \tag{66}$$

Combining the bounds:

By (63), (65), and (66) we conclude that

$$\left| \sum_{G \in \Delta^1_{n,k}} u_G^* \mathbb{E}[x_G] v_G \right| \le \sum_{i=1}^3 \left| \sum_{G \in (\Delta^1_{n,k})_i} u_G^* \mathbb{E}[x_G] v_G \right|$$
$$\ll_{L,k} n^{k/2-1/2}.$$

While the bounds above were calculated for $a = 1$, the same argument applies for any a by simply permuting the colors. Therefore, in the case where k is a multiple of m, from (58) we have

$$\left| \mathbb{E}\left[u^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v \right] \right| \leq n^{-k/2} \sum_{1 \leq a \leq m} \left| \sum_{G \in \Delta_{n,k}^a} u_G^* \mathbb{E}[x_G] v_G \right|$$

$$\ll_{L,k} n^{-k/2} \sum_{1 \leq a \leq m} n^{k/2 - 1/2}$$

$$\ll_{L,k} \frac{1}{\sqrt{n}}.$$

Case where k is not a multiple of m: Now assume that k is not a multiple of m. If $k < m$, then each matrix has at most one entry in the product on the right-hand side of (57) and all terms will be independent. Hence the expectation will be zero. Therefore, consider the case when $k > m$. Then there must exist some positive integer c such that

$$cm < k < (c+1)m.$$

We can write $k = cm + r$ for some $0 < r < c$ and in this case, a computation reveals that the only nonzero blocks in $\mathcal{Y}_n^k$ are blocks of the form $[a, a+r]$ where a and $a+r$ are reduced modulo m, and the modulo class representatives are $\{1, 2, \ldots, m\}$. In this case we can write

$$\mathbb{E}\left[u^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v \right] = n^{-k/2} \sum_{1 \leq a \leq m} (u^*)^{[a]} \mathbb{E}\left[\left(\mathcal{Y}_n^k \right)^{[a, a+r]} \right] v^{[a+r]}$$

$$= n^{-k/2} \sum_{1 \leq a \leq m} \sum_{G \in \Delta_{n,k}^a} u_G^* \mathbb{E}[x_G] v_G \qquad (67)$$

Again, define $h(G)$ as in (59) and define $(\Delta_{n,k}^a)_1$, $(\Delta_{n,k}^a)_2$, and $(\Delta_{n,k}^a)_3$ as in (60), (61), and (62), respectively. Without loss of generality, assume that $a = 1$.

If a graph G has height greater than $k/2$, by the same argument in the previous case we can see that there must be an edge which is not parallel to any other edge. Therefore when k is not a multiple of m we still have

$$\sum_{G \in (\Delta_{n,k}^1)_1} u_G^* \mathbb{E}[x_G] v_G = 0. \qquad (68)$$

If $G \in (\Delta_{n,k}^1)_3$ has height less than $k/2$, then we may still bound $\mathbb{E}\,|x_G| \leq M$. Therefore, we may use the same argument as in the previous case to conclude that

$$\left| \sum_{G \in (\Delta_{n,k}^1)_3} u_G^* \mathbb{E}[x_G] v_G \right| \ll_{L,k} n^{k/2 - 1}. \qquad (69)$$

Thus, we need only to consider graphs in $(\Delta_{n,k}^a)_2$.

Graphs in $(\Delta_{n,k}^1)_2$:

If k is odd, then $(\Delta^1_{n,k})_2$ is empty, so assume that k is even. Consider a graph $G \in (\Delta^1_{n,k})_2$ and by Lemma 11.11, we may assume that G is canonical. If G has any edges which are not parallel to any other edge, then $\mathbb{E}[x_G] = 0$, so assume each edge is parallel to at least one other edge. A counting argument reveals that in fact each edge must be parallel to exactly one other edge and by Lemma 11.17, we can conclude that in fact $G = G^1(1, 2, \ldots, k/2, 1, 2, \ldots, k/2, 1)$.

In order for this graph to correspond to a term with nonzero expectation, the colors on the pairs of parallel edges must match. In order for this to happen, we would need the edge from $(k/2+1, 1)$ to $(k/2+2, 2)$ to have color 1. This would force the edge from $(k/2, k/2)$ to $(k/2 + 1, 1)$ to have color m. Note that if we think about drawing edges sequentially with the time coordinate, then this implies that the $k/2$th edge drawn from $(k/2, k/2)$ to $(k/2 + 1, 1)$ is of color m, forcing $k/2$ to be a multiple of m. However, this would imply that k is also a multiple of m, a contradiction. Hence in this case, if $G \in (\tilde{\Delta}^a_{n,k})_2$, then $\mathbb{E}[x_G] = 0$. By Lemma 11.11, this gives

$$\sum_{G \in (\Delta^a_{n,k})_2} u^*_G \mathbb{E}[x_G] v_G = 0. \tag{70}$$

Combining the bounds:

By (68), (69), and (70) we can see that

$$\left| \sum_{G \in \Delta^1_{n,k}} u^*_G \mathbb{E}[x_G] v_G \right| \leq \sum_{i=1}^{3} \left| \sum_{G \in (\Delta^1_{n,k})_i} u^*_G \mathbb{E}[x_G] v_G \right|$$

$$\ll_{L,k} n^{k/2-1}.$$

While the bounds above were calculated for $a = 1$, the same arguments apply for any a by simply permuting the colors. Thus, from (67) we have

$$\left| \mathbb{E}\left[u^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v \right] \right| \leq n^{-k/2} \sum_{1 \leq a \leq m} \left| \sum_{G \in \Delta^a_{n,k}} u^*_G \mathbb{E}[x_G] v_G \right|$$

$$\ll_{L,k} n^{-k/2} \sum_{1 \leq a \leq m} n^{k/2-1}$$

$$\ll_{L,k} \frac{1}{n}$$

in the case where k is not a multiple of m.

Combining the cases above completes the proof of Lemma 11.2.

Remark 11.18. Note that if $m = 1$, then k is trivially a multiple of m. Hence, the case where $\mathcal{Y}_n$ is an $n \times n$ matrix follows as a special case of the above argument.

12. Proofs of results from Section 3

Before we prove the results from Section 3, we must prove an isotropic limit law for products of repeated matrices.

Theorem 12.1 (Isotropic limit law for repeated products)**.** *Assume ξ is a complex-valued random variable with mean zero, unit variance, finite fourth moment, and*

independent real and imaginary parts. For each $n \geq 1$, let X_n be an $n \times n$ iid random matrix with atom variable ξ. Define $\mathcal{Y}_n$ as in (9) and define $\mathcal{G}_n(z)$ as in (10) but with $X_{n,1} = X_{n,2} = \cdots = X_{n,m} = X_n$. Then, for any fixed $\delta > 0$, the following statements hold.

(i) *Almost surely, for n sufficiently large, the eigenvalues of $\frac{1}{\sqrt{n}}\mathcal{Y}_n$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. In particular, this implies that almost surely, for n sufficiently large, the matrix $\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI$ is invertible for every $z \in \mathbb{C}$ with $|z| > 1 + \delta$.*

(ii) *There exists a constant $c > 0$ (depending only on δ and m) such that almost surely, for n sufficiently large,*

$$\sup_{z \in \mathbb{C}: |z| > 1 + \delta} \|\mathcal{G}_n(z)\| \leq c.$$

(iii) *For each $n \geq 1$, let $u_n, v_n \in \mathbb{C}^{mn}$ be deterministic unit vectors. Then*

$$\sup_{z \in \mathbb{C}: |z| > 1 + \delta} \left| u_n^* \mathcal{G}_n(z) v_n + \frac{1}{z} u_n^* v_n \right| \longrightarrow 0$$

almost surely as $n \to \infty$.

Proof. Fix $\delta > 0$. From [60, Theorem 1.4], the spectral radius of $\frac{1}{\sqrt{n}}X_n$ converges to 1 almost surely as $n \to \infty$. Thus, $n^{-m/2}(X_n)^m$ has spectral radius converging to 1 almost surely as well. It follows that the spectral radius of $n^{-m/2}(\mathcal{Y}_n)^m$ converges to 1 almost surely, which in turn implies that the spectral radius of $\frac{1}{\sqrt{n}}\mathcal{Y}_n$ converges to 1 almost surely as $n \to \infty$, proving claim (i).

To prove part (ii), we consider two events, both of which hold almost surely. By [60, Theorem 1.4], there exists a constant $K > 0$ such that almost surely, for n sufficiently large, $n^{-1/2}\|X_n\| \leq K$, and hence, on the same event, $n^{-1/2}\|\mathcal{Y}_n\| \leq K$. By Lemma D.3 this implies that almost surely, for n sufficiently large,

$$\sup_{z \in \mathbb{C}: |z| \geq K+1} \left\| \left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI \right)^{-1} \right\| \leq 1. \tag{71}$$

To deal with $1 + \delta \leq |z| \leq K + 1$, we observe that

$$\left(\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI \right)^{-1} \right)^{[a,b]} = z^{(m-1)-\alpha} n^{-\alpha/2} X_n^\alpha \left(n^{-m/2} X_n^m - z^m I \right)^{-1} \tag{72}$$

by a block inverse computation, where $\alpha = (b - a) \pmod{m}$. Thus, we have

$$\left\| \left(\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI \right)^{-1} \right)^{[a,b]} \right\|$$

$$\leq |z|^{(m-1)-\alpha} \left\| n^{-\alpha/2} X_n^\alpha \right\| \left\| \left(n^{-m/2} X_n^m - z^m I \right)^{-1} \right\|$$

$$\leq |z|^{(m-1)-\alpha} \left\| n^{-\alpha/2} X_n^\alpha \right\| \prod_{k=1}^{m} \left\| \left(n^{-1/2} X_n - z e^{2\pi k \sqrt{-1}/m} I \right)^{-1} \right\|.$$

We now bound

$$\sup_{z \in \mathbb{C}: 1+\delta < |z| < K+1} |z|^{(m-1)-\alpha} \left\| n^{-\alpha/2} X_n^\alpha \right\| \prod_{k=1}^{m} \left\| \left(n^{-1/2} X_n - z e^{2\pi k \sqrt{-1}/m} I \right)^{-1} \right\|.$$

Note that almost surely, for n sufficiently large $\left\| n^{-\alpha/2} X_n^\alpha \right\| \leq K^\alpha \leq K^{m-1}$. Hence, we obtain

$$\sup_{z \in \mathbb{C}: 1+\delta < |z| < K+1} |z|^{(m-1)-\alpha} \left\| n^{-\alpha/2} X_n^\alpha \right\| \prod_{k=1}^m \left\| \left(n^{-1/2} X_n - z e^{2\pi k \sqrt{-1}/m} I \right)^{-1} \right\|$$

$$\leq (K+1)^{m-1} K^{m-1} \sup_{z \in \mathbb{C}: 1+\delta < |z| < K+1} \prod_{k=1}^m \left\| \left(n^{-1/2} X_n - z e^{2\pi k \sqrt{-1}/m} I \right)^{-1} \right\|$$

almost surely, for n sufficiently large. The bound for

$$\sup_{z \in \mathbb{C}: 1+\delta < |z| < K+1} \prod_{k=1}^m \left\| \left(n^{-1/2} X_n - z e^{2\pi k \sqrt{-1}/m} I \right)^{-1} \right\|$$

follows from Lemma 8.2 (taking $m = 1$). Returning to (72), we conclude that almost surely, for n sufficiently large

$$\sup_{z \in \mathbb{C}: 1+\delta \leq |z| \leq K+1} \left\| \left(\left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right)^{-1} \right)^{[a,b]} \right\| \leq c$$

for some constant $c > 0$ (depending only on δ and m). Since $1 \leq a, b \leq m$ are arbitrary, the proof of property (ii) is complete.

For (iii), [60, Theorem 1.4] yields that almost surely, for n sufficiently large,

$$\sup_{|z| \geq 5} \frac{1}{\sqrt{n}} \left\| \frac{\mathcal{Y}_n}{z} \right\| \leq \frac{9}{10} < 1. \tag{73}$$

Thus, we expand the resolvent as a Neumann series to obtain

$$\mathcal{G}_n(z) = -\frac{1}{z} \left(I + \sum_{k=1}^\infty \left(\frac{1}{\sqrt{n}} \frac{\mathcal{Y}_n}{z} \right)^k \right) = -\frac{1}{z} I - \sum_{k=1}^\infty \frac{\left(\frac{\mathcal{Y}_n}{\sqrt{n}} \right)^k}{z^{k+1}}.$$

Thus, we have almost surely, for n sufficiently large,

$$u^* \mathcal{G}_n(z) v = -\frac{1}{z} u^* v - \sum_{k=1}^\infty \frac{u^* \left(\frac{\mathcal{Y}_n}{\sqrt{n}} \right)^k v}{z^{k+1}}.$$

We will show that the series on the right-hand side converges to zero almost surely uniformly in the region $\{z \in \mathbb{C} : 5 \leq |z| \leq 6\}$. Indeed, from (73), the tail of the series is easily controlled. Thus, it suffices to show that, for each fixed integer $k \geq 1$,

$$\left| u_n^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v_n \right| = o_k(1).$$

But this follows from the block structure of $\mathcal{Y}_n$ and [60, Lemma 2.3].

We now extend this convergence to the region $\{z \in \mathbb{C} : |z| \geq 1 + \delta\}$. Let $\varepsilon > 0$. Let $M \geq 6$ be a constant to be chosen later. By Vitali's convergence theorem (see, for instance [15, Lemma 2.14]), it follows that

$$\sup_{1+\delta \leq |z| \leq M} \left| u_n^* \mathcal{G}_n(z) v_n + \frac{1}{z} u_n^* v_n \right| \longrightarrow 0$$

almost surely. In particular, almost surely, for n sufficiently large,

$$\sup_{1+\delta \leq |z| \leq M} \left| u_n^* \mathcal{G}_n(z) v_n + \frac{1}{z} u_n^* v_n \right| \leq \varepsilon. \tag{74}$$

Choose $M_1 > 0$ such that, for all $|z| \geq M_1$,

$$\left\| \left(-\frac{1}{z} \right) u^* v \right\| \leq \left| \frac{1}{z} \right| \|u^*\| \, \|v\| \leq \frac{\varepsilon}{2}.$$

Also there exists a constant $M_2 > 0$ such that

$$\sup_{|z| \geq M_2} \|u^* \mathcal{G}_n(z) v\| \leq \frac{\varepsilon}{2}$$

almost surely, for n sufficiently large, by Lemma D.3 and [60, Theorem 1.4]. Take $M := \max\{M_1, M_2, 6\}$. Then almost surely, for n sufficiently large,

$$\sup_{|z| \geq M} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \leq \varepsilon. \tag{75}$$

Combining (74) and (75), we obtain almost surely, for n sufficiently large,

$$\sup_{|z| \geq 1+\delta} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \leq \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the proof is complete. $\qquad\square$

We also need the following lemma in order to prove Theorem 3.2.

Lemma 12.2. *Let ξ be a complex-valued random variable with mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. For each $n \geq 1$, let X_n be an $n \times n$ iid random matrix with atom variable ξ. Let m be a positive integer. Then, for any fixed $\delta > 0$, the following statements hold.*

(i) *Almost surely, for n sufficiently large, the eigenvalues of $n^{-m/2} X_n^m$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. In particular, this implies that almost surely, for n sufficiently large, the matrix $n^{-m/2} X_n^m - zI$ is invertible for every $z \in \mathbb{C}$ with $|z| > 1 + \delta$.*

(ii) *There exists a constant $c > 0$ such that almost surely, for n sufficiently large,*

$$\sup_{z \in \mathbb{C}:|z|>1+\delta} \left\| \left(n^{-m/2} X_n^m - zI \right)^{-1} \right\| \leq c.$$

(iii) *For each $n \geq 1$, let $u_n, v_n \in \mathbb{C}^n$ be deterministic unit vectors. Then*

$$\sup_{z \in \mathbb{C}:|z|>1+\delta} \left| u_n^* \left(n^{-m/2} X_n^m - zI \right)^{-1} v_n + \frac{1}{z} u_n^* v_n \right| \longrightarrow 0$$

almost surely as $n \to \infty$.

The proof of this lemma is similar to the proof of Corollary 6.2; we omit the details.

With these results, we may proceed to the proofs of the results in Section 3. The proofs of Theorems 3.1, 3.2, and 3.3 follow the proofs of Theorems 2.5, 2.6, and 2.8, respectively, verbatim, except for the following changes:

- Take $X_{n,1} = \cdots = X_{n,m} = X_n$,
- Replace all occurrences of Theorem 6.1 by Theorem 12.1
- Replace all occurrences of Corollary 6.2 by Lemma 12.2.

- The scaling factor of σ needs to be replaced by σ^m.

APPENDIX A. PROOF OF LEMMA 7.1

In this section, we present the proof of Lemma 7.1.

Proof of Lemma 7.1. Take $L_0 := \sqrt{8\mathbb{E}|\xi|^4}$. We begin by proving *(ii)*. Observe that

$$
\begin{aligned}
1 &= \mathbb{E}|\xi|^2 \\
&= \mathbb{E}|\operatorname{Re}(\xi)|^2 + \mathbb{E}|\operatorname{Im}(\xi)|^2 \\
&= \mathbb{E}\left[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)|\leq L/\sqrt{2}\}}\right] + \mathbb{E}\left[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)|>L/\sqrt{2}\}}\right] \\
&\quad + \mathbb{E}\left[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)|\leq L/\sqrt{2}\}}\right] + \mathbb{E}\left[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)|>L/\sqrt{2}\}}\right] \\
&= \operatorname{Var}(\operatorname{Re}(\tilde{\xi})) + \operatorname{Var}(\operatorname{Im}(\tilde{\xi})) \\
&\quad + \left|\mathbb{E}\left[\operatorname{Re}(\xi)\mathbf{1}_{\{|\operatorname{Re}(\xi)|\leq L/\sqrt{2}\}}\right]\right|^2 + \mathbb{E}\left[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)|>L/\sqrt{2}\}}\right] \\
&\quad + \left|\mathbb{E}\left[\operatorname{Im}(\xi)\mathbf{1}_{\{|\operatorname{Im}(\xi)|\leq L/\sqrt{2}\}}\right]\right|^2 + \mathbb{E}\left[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)|>L/\sqrt{2}\}}\right],
\end{aligned}
$$

which implies

$$
\begin{aligned}
1 - \operatorname{Var}(\tilde{\xi}) &= \left|\mathbb{E}\left[\operatorname{Re}(\xi)\mathbf{1}_{\{|\operatorname{Re}(\xi)|\leq L/\sqrt{2}\}}\right]\right|^2 + \mathbb{E}\left[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)|>L/\sqrt{2}\}}\right] \\
&\quad + \left|\mathbb{E}\left[\operatorname{Im}(\xi)\mathbf{1}_{\{|\operatorname{Im}(\xi)|\leq L/\sqrt{2}\}}\right]\right|^2 + \mathbb{E}\left[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)|>L/\sqrt{2}\}}\right].
\end{aligned}
$$

Thus, using the fact that $\operatorname{Re}(\xi)$ and $\operatorname{Im}(\xi)$ both have mean zero (so, for example, $\mathbb{E}\left[\operatorname{Re}(\xi)\mathbf{1}_{\{|\operatorname{Re}(\xi)|\leq L/\sqrt{2}\}}\right] = -\mathbb{E}\left[\operatorname{Re}(\xi)\mathbf{1}_{\{|\operatorname{Re}(\xi)|>L/\sqrt{2}\}}\right]$) and then applying Jensen's inequality, we obtain

$$
\begin{aligned}
|1 - \operatorname{Var}(\tilde{\xi})| &\leq 2\mathbb{E}[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)|>L/\sqrt{2}\}}] + 2\mathbb{E}[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)|>L/\sqrt{2}\}}] \\
&\leq 2\mathbb{E}[|\xi|^2 \mathbf{1}_{\{|\xi|>L/\sqrt{2}\}}] \\
&\leq \frac{4}{L^2}\mathbb{E}|\xi|^4.
\end{aligned}
$$

This concludes the proof of *(ii)*.

Property *(i)* follows easily from *(ii)* by the choice of L_0. Next we move onto the proof of *(iii)*. One can see that since $\operatorname{Var}(\tilde{\xi}) \geq \frac{1}{2}$,

$$
\begin{aligned}
\left|\hat{\xi}\right| &\leq \frac{|\operatorname{Re}(\xi)|\,\mathbf{1}_{\{|\operatorname{Re}(\xi)|\leq L/\sqrt{2}\}} + \mathbb{E}\left[|\operatorname{Re}(\xi)|\,\mathbf{1}_{\{|\operatorname{Re}(\xi)|\leq L/\sqrt{2}\}}\right]}{\sqrt{\operatorname{Var}(\tilde{\xi})}} \\
&\quad + \frac{|\operatorname{Im}(\xi)|\,\mathbf{1}_{\{|\operatorname{Im}(\xi)|\leq L/\sqrt{2}\}} + \mathbb{E}\left[|\operatorname{Im}(\xi)|\,\mathbf{1}_{\{|\operatorname{Im}(\xi)|\leq L/\sqrt{2}\}}\right]}{\sqrt{\operatorname{Var}(\tilde{\xi})}} \\
&\leq 4L
\end{aligned}
$$

almost surely.

For *(iv)*, we observe that $\hat{\xi}$ has mean zero and unit variance by construction. Additionally, since the real and imaginary parts of $\hat{\xi}$ depend only on the real and

imaginary parts of ξ respectively, they are independent by construction. For the fourth moment, we use that $\mathrm{Var}(\tilde{\xi}) \geq \frac{1}{2}$ and Jensen's inequality inequality to obtain

$$\mathbb{E}|\hat{\xi}|^4 \ll \frac{1}{\mathrm{Var}(\tilde{\xi})^2} \left(\mathbb{E}\left[|\mathrm{Re}(\xi)|^4 \, \mathbf{1}_{\{| \mathrm{Re}(\xi)| \leq L/\sqrt{2}\}} \right] + \mathbb{E}\left[|\mathrm{Im}(\xi)|^4 \, \mathbf{1}_{\{| \mathrm{Im}(\xi)| \leq L/\sqrt{2}\}} \right] \right)$$
$$\ll \mathbb{E}|\xi|^4,$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Appendix B. Proof of Theorem 8.1

This section is devoted to the proof of Theorem 8.1. We begin with Lemma B.1 below, which is based on [47, Theorem 4]. Throughout this section, we use $\sqrt{-1}$ for the imaginary unit and reserve i as an index.

Lemma B.1. *Let μ be a probability measure on $[0, \infty)$, and for each $n \geq 1$, let*

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_{n,i}}$$

for some triangular array $\{\lambda_{n,i}\}_{i \leq n}$ of nonnegative real numbers. Let m_n by the Stieltjes transform of μ_n and m be the Stieltjes transform of μ, i.e.,

$$m_n(z) := \int \frac{d\mu_n(x)}{x - z}, \quad m(z) := \int \frac{d\mu(x)}{x - z}$$

for all $z \in \mathbb{C}$ with $\mathrm{Im}(z) > 0$. Assume

- *(i) $\mu_n \to \mu$ as $n \to \infty$,*
- *(ii) there exists a constant $c > 0$ such that $\mu([0, c]) = 0$,*
- *(iii) $\sup_{E \in [0,c]} \left| m_n(E + \sqrt{-1}n^{-1/2}) - m(E + \sqrt{-1}n^{-1/2}) \right| = o(n^{-1/2})$.*

Then there exists a constant $n_0 \geq 1$ such that $\mu_n([0, c/2]) = 0$ for all $n > n_0$.

Proof. Observe that

$$\mathrm{Im}\, m_n(E + \sqrt{-1}n^{-1/2}) - \mathrm{Im}\, m(E + \sqrt{-1}n^{-1/2}) = \int \frac{n^{-1/2} d(\mu_n - \mu)(x)}{(E - x)^2 + n^{-1}}.$$

From assumption (iii), we conclude that

$$\sup_{E \in [0,c]} \left| \int \frac{d(\mu_n - \mu)(x)}{(E - x)^2 + n^{-1}} \right| = o(1).$$

We decompose this integral into two parts

$$\int \frac{d(\mu_n - \mu)(x)}{(E - x)^2 + n^{-1}} = \int_0^c \frac{d\mu_n(x)}{(E - x)^2 + n^{-1}} + \int_c^\infty \frac{d(\mu_n - \mu)(x)}{(E - x)^2 + n^{-1}},$$

where we used the assumption that $\mu([0, c]) = 0$.

Observe that

$$\int_c^\infty \frac{d(\mu_n - \mu)(x)}{(E - x)^2 + n^{-1}} \longrightarrow 0$$

uniformly for any $E \in [0, c/2]$ by the assumption that $\mu_n \to \mu$. Therefore, it must be the case that

$$\sup_{E \in [0,c/2]} \int_0^c \frac{d\mu_n(x)}{(E - x)^2 + n^{-1}} \longrightarrow 0.$$

Take $n_0 \geq 1$ such that

$$\sup_{E \in [0, c/2]} \int_0^c \frac{d\mu_n(x)}{(E - x)^2 + n^{-1}} \leq 1/2 \tag{76}$$

for all $n \geq n_0$.

In order to reach a contradiction, assume there exists $n > n_0$ and $i \in [n]$ such that $\lambda_{n,i} \in [0, c/2]$. Then

$$\sup_{E \in [0, c/2]} \int_0^c \frac{d\mu_n(x)}{(E - x)^2 + n^{-1}} = \sup_{E \in [0, c/2]} \frac{1}{n} \sum_{j=1}^n \frac{1}{(E - \lambda_{n,j})^2 + n^{-1}}$$

$$\geq \sup_{E \in [0, c/2]} \frac{1}{n} \frac{1}{(E - \lambda_{n,i})^2 + n^{-1}}$$

$$\geq 1,$$

a contradiction of (76). We conclude that $\mu_n([0, c/2]) = 0$ for all $n > n_0$. $\square$

With Lemma B.1 in hand, we are now prepared to prove Theorem 8.1. The proof below is based on a slight modification to the arguments from [47, 48]. As such, in some places we will omit technical computations and only provide appropriate references and necessary changes to results from [47, 48].

Fix $\delta > 0$. It suffices to prove that

$$\inf_{1 + \delta \leq |z| \leq 6} s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \geq c \tag{77}$$

and

$$\inf_{|z| > 6} s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \geq c' \tag{78}$$

with overwhelming probability for some constants $c, c' > 0$ depending only on δ.

The second bound (78) follows by Lemma D.3. Indeed, a bound on the spectral norm of $\mathcal{Y}_n$ (which follows from standard bounds on the spectral norms of $X_{n,k}$; see, for example, [60, Theorem 1.4]) gives

$$\|\mathcal{Y}_n\| \leq 3\sqrt{n}$$

with overwhelming probability. The bound in (78) then follows by applying Lemma D.3.

We now turn to the bound in (77). To prove this bound, we will use Lemma B.1. Let $\mu_{n,z}$ be the empirical spectral measure constructed from the eigenvalues of

$$\left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right)^*.$$

From [48, Theorem 2.6], for all $|z| \geq 1 + \delta$, there exists a probability measure μ_z supported on $[0, \infty)$ such that $\mu_{n,z} \to \mu_z$ with overwhelming probability. Moreover, from [12, Lemma 4.2] there exists a constant $c > 0$ (depending only on δ) such that $\mu_z([0, c]) = 0$ for all $|z| \geq 1 + \delta$. Lastly, condition (iii) in Lemma B.1 follows for all $1 + \delta \leq |z| \leq 6$ with overwhelming probability from [47, Theorem 5]. Applying Lemma B.1, we conclude that

$$s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \geq c/2$$

with overwhelming probability uniformly for all $1 + \delta \leq |z| \leq 6$. The bound for the infimum can now be obtained by a simple net argument and Weyl's inequality (31). The proof of Theorem 8.1 is complete.

APPENDIX C. PROOF OF LEMMA 11.17

This section is devoted to the proof of Lemma 11.17.

Proof of Lemma 11.17. The proof proceeds inductively. We begin with a graph only containing the vertex $(1,1)$ and then add vertices and edges sequentially with time. First, edge 1 is added; it will span from $(1,1)$ to $(2,i_2)$. Next, edge 2 is added and will span from $(2,i_2)$ to $(3,i_3)$, and so on. We use induction to prove that at each time step t, there is only one possible choice for i_{t+1}, resulting in a unique canonical graph with maximal height $k/2$ and in which each edge is parallel to exactly one other edge.

The edge starting at vertex $(1,1)$ can either be of type II (terminating on $(2,1)$) or of type I (terminating on $(2,2)$). By way of contradiction, assume the edge is type II. Since G still has $k/2 - 1$ more height coordinates left to reach, it would require at least $k/2 - 1$ type I edges to reach hight coordinate $k/2$. Since each edge must be parallel to exactly one other edge, at some point there must be a type II edge, returning to a height coordinate previously visited. This edge will also need to be parallel to another edge. Counting all pairs of parallel edges shows that G must have at least $k + 1$ more edges, a contradiction. Hence the edge starting at vertex $(1,1)$ must be of type I.

Assume that all edges up to time coordinate t, where $1 \leq t < k/2 - 1$, are type I edges. Then G must have an edge starting at vertex $(t+1, t+1)$. This edge can either of type I or type II. In order to reach a contradiction, assume that the edge is type II. Then G must have at least $k/2 - t - 1$ more type I edges in order to reach the height $k/2$, and G has exactly $k - t - 1$ more edges to be added. Visiting each unvisited height coordinate would require at least $k/2 - t - 1$ more type I edges, and at some point after visiting new height coordinates, G must return to a smaller hight coordinate, resulting in a type II edge. None of these edges could be parallel to any previous edges. Thus, overall G would need to have at least $k - t + 1$ more edges, a contradiction to the fact that G must have exactly $k - t - 1$ more edges.

We conclude that each edge of G must be type I until the hight coordinate $k/2$ is reached. Namely, we have vertices $(1,1), \ldots, (k/2, k/2)$.

At this point G must have an edge starting at vertex $(k/2, k/2)$. Note that G has $k/2 - 1$ edges up to this point, none of which are parallel to any other edge. G must have edges parallel to the edges previously introduced and G has exactly $k/2 + 1$ edges remaining to do so. Since there are no remaining unvisited height coordinates, the edge which starts at vertex $(k/2, k/2)$ must terminate at $(k/2 + 1, i_{k/2+1})$ for some $1 \leq i_{k/2+1} \leq k/2$, resulting in the first type II edge. See Figure 9 for a visual representation of the graph up to this point.

We now claim that this first type II edge must in fact terminate at $(k/2 + 1, 1)$. By way of contradiction, suppose this edge terminates at vertex $(k/2 + 1, i)$ for any $1 < i \leq \frac{k}{2}$. Since this is the first type II edge, it cannot be parallel to any other previously drawn edge. Up to this point G has $k/2$ edges drawn and $k/2$ edges remaining to be drawn. Since all edges are by assumption to be parallel to exactly one other, a simple counting argument reveals that each edge drawn from this point on must be parallel to an existing edge. Since there is only one

edge which starts at height coordinate i, we must now draw the edge starting at $(k/2 + 1, i)$ and terminating at $(k/2 + 2, i + 1)$. By continuing this argument inductively, we must draw edges which start at $(k/2 + j + 1, i + j)$ and terminate at vertex $(k/2 + j + 2, i + j + 1)$ for $0 \leq j \leq k/2 - i$, until the height coordinate $k/2$ is reached again. Now, since $k - i + 1 < k$, we must draw at least one more edge, and this edge must start at vertex $(k - i, k/2)$. In order to draw an edge parallel to an existing edge, this edge must terminate at vertex $(k - i + 1, i)$. However, if we do this, we must now draw an edge parallel to an existing edge which starts a height coordinate i, a contradiction because the only previous edges which began at height coordinate i are parallel to one another and there cannot be three edges parallel. This concludes the proof of the claim.

By the previous claim, the first type II edge must terminate at $(k/2 + 1, 1)$. Again, since this is the first type II edge it cannot be parallel to any other edge. Up to this point G has $k/2$ edges drawn and $k/2$ edges remaining to be drawn. Since all edges are by assumption parallel to exactly one other, a simple counting argument reveals that each edge drawn from this point on must be parallel to an existing edge. Since there is only one edge which starts at height coordinate 1, we must now draw the edge starting at $(k/2+1, 1)$ and terminating at $(k/2+2, 2)$. By continuing this argument inductively, we must draw edges which start at $(k/2 + j + 1, j + 1)$ and terminate at vertex $(k/2 + j + 2, j + 2)$ for $0 \leq j \leq k/2 - 1$, until the height coordinate $k/2$ is reached again. Up to this point, $k - 1$ edges of G have been drawn and we must draw one more edge which starts at vertex $(k, k/2)$. Since there is only one previous edge in the graph which starts at height coordinate $k/2$, the final edge must terminate at height coordinate 1 and all edges are parallel to exactly one other edge. This results in the vertex set

$$V = \{(1, 1),\ (2, 2),\ \ldots, (k/2, k/2),\ (k/2+1, 1),\ (k/2+2, 2), \ldots (k, k/2),\ (k+1, 1)\}.$$

The corresponding canonical m-colored k-path graph would be

$$G^1(1, 2, \ldots, k/2, 1, 2, \ldots, k/2, 1),$$

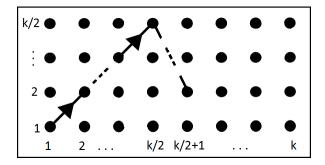and the proof is complete. $\qquad\square$

FIGURE 9. This is an example of a path graph that has type I edges drawn until a height of $k/2$ is reached, then the first type II edge is drawn.

Appendix D. Useful inequalities

Lemma D.1 (Lemma A.1 from [14]). *For $X = (x_1, x_2, \ldots, x_N)^T$ iid standardized complex entries, B an $N \times N$ Hermitian nonnegative definite matrix, we have, for any $p \geq 1$,*

$$\mathbb{E}\,|X^*BX|^p \leq K_p\left((\mathrm{tr}B)^p + \mathbb{E}|x_1|^{2p}\mathrm{tr}B^p\right).$$

where $K_p > 0$ depends only on p.

Lemma D.2 (Follows from Lemma 10 in [53]). *Let A and B be $k \times k$ matrices with $\|A\|, \|B\| = O(1)$. Then*

$$|\det(A) - \det(B)| \ll_k \|A - B\|.$$

Lemma D.3 (Spectral norm bound for large $|z|$; Lemma 3.1 from [49]). *Let A be a square matrix that satisfies $\|A\| \leq \mathcal{K}$. Then*

$$\left\|(A - zI)^{-1}\right\| \leq \frac{1}{\varepsilon}$$

for all $z \in \mathbb{C}$ with $|z| \geq \mathcal{K} + \varepsilon$.

References

[1] L.F. Abbott, K. Rajan, *Eigenvalue Spectra of Random Matrices for Neural Networks* Phys. Rev. Lett. 97, 188104 (2006).

[2] L.F. Abbott, D. Sussillo, *Random Walk Initialization for Training Very Deep Feedforward Networks*, available at `arXiv:1412.6558v3 [cs.NE]`.

[3] G. Akemann, Z. Burda, *Universal microscopic correlation functions for products of independent Ginibre matrices*, J. Phys. A: Math. Theor. **45** (2012).

[4] G. Akemann, Z. Burda, M. Kieburg, *Universal distribution of Lyapunov exponents for products of Ginibre matrices*, J. Phys. A: Math. Theor. 47(39) (2014).

[5] G. Akemann, J. R. Ipsen, M. Kieburg, *Products of Rectangular Random Matrices: Singular Values and Progressive Scattering*, Phys. Rev. E **88**, (2013).

[6] G. Akemann, J. R. Ipsen, E. Strahov, *Permanental processes from products of complex and quaternionic induced Ginibre ensembles*, Random Matrices: Theory and Applications Vol. 3, No. 4 (2014) 1450014.

[7] G. Akemann, M. Kieburg, L. Wei, *Singular value correlation functions for products of Wishart random matrices*, J. Phys. A: Math. Theor. **46** (2013).

[8] G. Akemann, E. Strahov, *Hole probabilities and overcrowding estimates for products of complex Gaussian matrices*, J. Stat. Phys. (2013), Volume 151, Issue 6, pp 987–1003.

[9] G. Anderson, *Convergence of the largest singular value of a polynomial in independent Wigner matrices*, Ann. Probab., Volume 41, Number 3B (2013), 2103–2181.

[10] P. Arbenz, W. Gander, G. Golub, *Restricted rank modification of the symmetric eigenvalue problem: Theoretical considerations*, Linear Algebra and its Applications **104** (1988), 75–95.

[11] G.B. Arous, A. Choromanska, M. Henaff, Y. LeCun, M. Mathieu, *The Loss Surfaces of Multilayer Networks*, available at `arXiv:1412.0233v3 [cs.LG]`.

[12] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), 494–529.

[13] Z. D. Bai, G. M. Pan, *Limiting behavior of eigenvectors of large Wigner matrices*, J. Stat. Phys. 146 (2012), 519–549.

[14] Z. D. Bai, J. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab. Volume 26, Number 1 (1998), 316–345.

[15] Z. D. Bai, J. Silverstein, *Spectral analysis of large dimensional random matrices*, Mathematics Monograph Series **2**, Science Press, Beijing 2006.

[16] J. Baik, G. Ben Arous, S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab. Volume 33, Number 5 (2005), 1643–1697.

[17] S. Belinschi, H.Bercovivi, M. Capitaine, *On the outlying eigenvalues of a polynomial in large independent random matrices*, available at `arXiv:1703.08102 [math.PR]`.

[18] F. Benaych-Georges, G. Cébron, J. Rochet, *Fluctuation of matrix entries and application to outliers of elliptic matrices*, available at `arXiv:1602.02929 [math.PR]`.

[19] F. Benaych-Georges, A. Guionnet, M. Maida, *Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices*, Elec. J. Probab., **16** (2011), 1621–1662.

[20] F. Benaych-Georges, A. Guionnet, M. Maida, *Large deviations of the extreme eigenvalues of random deformations of matrices*, Probab. Theory Related Fields **154**:3-4 (2012), 703–751.

[21] F. Benaych-Georges, R. N. Rao, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Adv. Math., **227**, No. 1, (2011), 494–521.

[22] F. Benaych-Georges, R. N. Rao, *The singular values and vectors of low rank perturbations of large rectangular random matrices*, J. Multivariate Anal., Vol. 111, 120–135, 2012.

[23] F. Benaych-Georges, J. Rochet, *Outliers in the Single Ring Theorem*, Probab. Theory Relat. Fields (2016) 165:313–363.

[24] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, Springer-Verlag New York, 1997.

[25] C. Bordenave, *On the spectrum of sum and product of non-hermitian random matrices*, Elect. Comm. in Probab. **16** (2011), 104–113.

[26] C. Bordenave, M. Capitaine, *Outlier eigenvalues for deformed i.i.d. random matrices*, to appear in Communications on Pure and Applied Mathematics.

[27] C. Bordenave, P. Caputo, D. Chafaï, K. Tikhomirov *On the spectral radius of a random matrix*, available at `arXiv:1607.05484v1 [math.CO]`.

[28] C. Bordenave, D. Chafaï, *Around the circular law.* Probability Surveys **9** (2012), 1–89.

[29] S. Belinschi, M. Capitaine, *Spectral properties of polynomials in independent Wigner and deterministic matrices*, available at `arXiv:1611.07440v1 [math.PR]`.

[30] Z. Burda, R. A. Janik, B. Waclaw, *Spectrum of the product of independent random Gaussian matrices*, Phys. Rev. E **81** (2010).

[31] Z. Burda, A. Jarosz, G. Livan, M. A. Nowak, A. Swiech, *Eigenvalues and Singular Values of Products of Rectangular Gaussian Random Matrices*, Phys. Rev. E **82** (2010).

[32] Z. Burda, *Free products of large random matrices - a short review of recent developments*, J. Phys.: Conf. Ser. **473** 012002 (2013).

[33] D. L. Burkholder, *Distribution function inequalities for martingales*, Ann. Probab. **1** 19–42 (1973).

[34] M. Capitaine, C. Donati-Martin, D. Féral, *The largest eigenvalue of finite rank deformation of large Wigner matrices: convergence and non universality of the fluctuations*, Ann. Probab., **37**, (1), (2009), 1–47.

[35] M. Capitaine, C. Donati-Martin, D. Féral, *Central limit theorems for eigenvalues of deformations of Wigner matrices*, Ann. I.H.P.-Prob.et Stat., **48**, No. 1 (2012), 107–133.

[36] M. Capitaine, C. Donati-Martin, D. Féral, M. Février, *Free convolution with a semi-circular distribution and eigenvalues of spiked deformations of Wigner matrices*, Elec. J. Probab., Vol. 16, No. 64, 1750–1792 (2011).

[37] D. Féral, S. Péché, *The largest eigenvalue of rank one deformation of large Wigner matrices*, Comm. Math. Phys. 272:185–228, 2007.

[38] P. J. Forrester, *Lyapunov exponents for products of complex Gaussian random matrices*, J. Stat. Phys. **151**, 796–808 (2013).

[39] P. J. Forrester, *Probability of all eigenvalues real for products of standard Gaussian matrices*, J. Phys. A ,vol. 47, 065202 (2014).

[40] F. Götze, T. Tikhomirov, *On the Asymptotic Spectrum of Products of Independent Random Matrices*, available at `arXiv:1012.2710`.

[41] U. Haagerup, S. Thorbjørnsen, *A new application of random matrices: $Ext(C^*_{red}(F_2))$ is not a group*, Ann. of Math. **162**, 711–775 (2005).

[42] R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press (1991).

[43] J.R. Ipsen *Products of Independent Gaussian Random Matrices*, PhD Thesis, Bielefeld University (2015), available at `arXiv:1510.06128v1 [math-ph]`.

[44] A. Knowles, J. Yin, *The isotropic semicircle law and deformation of Wigner matrices*, Communications on Pure and Applied Mathematics, Vol. LXVI, 1663–1749 (2013).

[45] A. Knowles, J. Yin, *The outliers of a deformed Wigner matrix*, Annals of Probability 2014, Vol. 42, No. 5, 1980–2031.

[46] A. B. J. Kuijlaars, L. Zhang, *Singular Values of Products of Ginibre Random Matrices, Multiple Orthogonal Polynomials and Hard Edge Scaling Limits*, Communications in Mathematical Physics, Volume 332, Issue 2, pp 759–781 (2014).

[47] Y. Nemish, *No outliers in the spectrum of the product of independent non-Hermitian random matrices with independent entries*, Journal of Theoretical Probability (2016) 1–43.

[48] Y. Nemish, *Local law for the product of independent non-Hermitian random matrices with independent entries*, Electron. J. Probab. Volume 22 (2017), paper no. 22, 35 .

[49] S. O'Rourke, D. Renfrew, *Low rank perturbations of large elliptic random matrices*, Electronic Journal of Probability Vol. 19, No. 43, 1–65 (2014).

[50] S. O'Rourke, D. Renfrew, A. Soshnikov, V. Vu, *Products of independent elliptic random matrices*, J. Stat. Phys., Vol. 160, No. 1 (2015), 89–119.

[51] S. O'Rourke, A. Soshnikov, *Products of Independent Non-Hermitian Random Matrices*, Electronic Journal of Probability, Vol. **16**, Art. 81, 2219–2245, (2011).

[52] S. O'Rourke, P. Wood, *Spectra of nearly Hermitian random matrices*, Annales de l'Institut Henri Poincaré - Probabilité et Statistiques, Vol. **53**, No. 3 (2017), 1241–1279.

[53] P. Pagacz, M. Wojtylak, *Random Perturbations of Matrix Polynomials*, available at `arXiv:1703.01858 [math.PR]`.

[54] S. Péché, *The largest eigenvalue of small rank perturbations of Hermitian random matrices*, Prob. Theory Relat. Fields, Vol. **134**, No. 1, 127–173, 2006.

[55] A. Pizzo, D. Renfrew, A. Soshnikov, *On Finite Rank Deformations of Wigner Matrices*, Ann. Inst. Henri Poincaré Probab. Stat., 49(1):64–94, 2013.

[56] A. B. Rajagopalan, *Outlier eigenvalue fluctuations of perturbed iid matrices*, available at `arXiv:1507.01441 [math.PR]`.

[57] D. Renfrew, A. Soshnikov, *On Finite Rank Deformations of Wigner Matrices II: Delocalized Perturbations*, Random Matrices Theory Appl. **02**, 1250015, 2013.

[58] J. Rochet, *Complex Outliers of Hermitian Random Matrices*, Journal of Theoretical Probability, Vol **30**, Issue 4 (2017) 1624–1654

[59] E. Strahov, *Differential equations for singular values of products of Ginibre random matrices*, J. Phys. A: Math. Theor. **47** 325203, 2014.

[60] T. Tao, *Outliers in the spectrum of iid matrices with bounded rank perturbations*, Probab. Theory Related Fields **155** (2013), 231–263.

[61] T. Tao, V. Vu, *Random matrices: Universality of ESDs and the circular law*, Ann. Probab. Volume 38, Number **5** (2010), 2023–2065.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO AT BOULDER, BOULDER, CO 80309
E-mail address: `natalie.coston@colorado.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF COLORADO AT BOULDER, BOULDER, CO 80309
E-mail address: `sean.d.orourke@colorado.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WISCONSIN-MADISON, 480 LINCOLN DR., MADISON, WI 53706
E-mail address: `pmwood@math.wisc.edu`