

Outliers in the spectrum for products of independent random matrices

Natalie Coston^a, Sean O'Rourke^a and Philip Matchett Wood^b

^a*Department of Mathematics, University of Colorado at Boulder, Boulder, CO 80309, USA.*

E-mail: natalie.coston@colorado.edu; sean.d.orourke@colorado.edu

^b*Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Dr., Madison, WI 53706, USA. E-mail: pmwood@math.wisc.edu*

Received 19 January 2018; revised 30 January 2019; accepted 7 May 2019

Abstract. For fixed $m \geq 1$, we consider the product of m independent $n \times n$ random matrices with iid entries as $n \rightarrow \infty$. 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 (*Probab. Theory Related Fields* **155** (2013) 231–263) for the case of a single iid matrix. Our methods also allow us to consider several other types of perturbations, including multiplicative perturbations.

Résumé. Pour un $m \geq 1$ fixé, nous considérons le produit de m matrices aléatoires indépendantes de taille $n \times n$, à coefficients i.i.d., lorsque $n \rightarrow \infty$. Sous certaines hypothèses sur les coefficients de chaque matrice, il est connu que la loi empirique limite des valeurs propres est décrite par la puissance m -ième de la loi circulaire. De plus, cette même loi limite apparaît toujours si chacune des matrices i.i.d. est perturbée additivement par une erreur déterministe de rang borné. Néanmoins, les perturbations de rang borné peuvent créer quelques valeurs propres atypiques (outliers). Nous décrivons la localisation asymptotique de ces valeurs propres atypiques, ce qui généralise un résultat de Tao (*Probab. Theory Related Fields* **155** (2013) 231–263) dans la cas d'une seule matrice i.i.d. Nos méthodes nous permettent également de considérer d'autres types de perturbations, dont des perturbations multiplicatives.

MSC: 60B20

Keywords: Product random matrices; Outlier eigenvalues; Perturbed products; Isotropic limit law

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 $\lambda_1(M_n), \dots, \lambda_n(M_n)$ 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 $f : \mathbb{C} \rightarrow \mathbb{C}$,

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

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

Throughout the paper, we use asymptotic notation (such as O, o) under the assumption that $n \rightarrow \infty$; see Section 3.1 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

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

Theorem 1.2 (Circular law; Corollary 1.12 from [58]). 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 $\text{rank}(A_n) = o(n)$ 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 $n \rightarrow \infty$.

This result appears as [58, Corollary 1.12], but is the culmination of work by many authors. We refer the interested reader to the excellent survey [26] 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.

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 known that the spectral radius of $\frac{1}{\sqrt{n}}X_n$ converges to 1 almost surely as $n \rightarrow \infty$ (see for example Theorem 5.18 from [12]). It is conjectured in [25] that the spectral radius of $\frac{1}{\sqrt{n}}X_n$ converges to 1 in probability as $n \rightarrow \infty$ only assuming that ξ has mean zero and unit variance. However, as we saw in Figure 1,

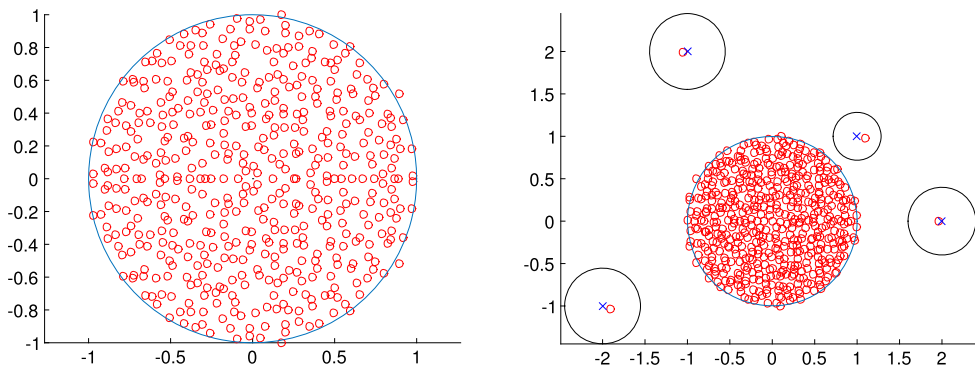


Fig. 1. On the left, we have plotted the eigenvalues of a 500×500 random matrix with iid standard Gaussian entries scaled by $\frac{1}{\sqrt{500}}$. Additionally, the unit circle is plotted for reference. The image on the right contains the eigenvalues of $\frac{1}{\sqrt{500}}X + A$, where X is a 500×500 random matrix with iid symmetric ± 1 Bernoulli entries and $A = \text{diag}(1 + i, 2i - 1, 2, -i - 2, 0, \dots, 0)$. For reference, we have also plotted each nonzero eigenvalue of A with a cross.

having spectral radius $1 + o(1)$ need not be the case if there is an additive perturbation of the form $\frac{1}{\sqrt{n}}X_n + A_n$. In this case, Tao [57] precisely describes the location of outlying eigenvalues.

Theorem 1.3 (Outliers for small low-rank perturbations of iid matrices; Theorem 1.7 from [57]). *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 $\lambda_1(A_n), \dots, \lambda_j(A_n)$ 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 $\lambda_1(\frac{1}{\sqrt{n}}X_n + A_n), \dots, \lambda_j(\frac{1}{\sqrt{n}}X_n + A_n)$ 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 $n \rightarrow \infty$ 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 [13–21,24,31–33,35,41,42,46,49,51–55,57] and references therein. In particular, the results in [53] extend Theorem 1.3 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 [47], due to Renfrew, Soshnikov, Vu, and the second author of the current paper.

Theorem 1.4 (Theorem 2.4 from [47]). *Fix an integer $m \geq 1$, and let $\tau > 0$. Let ξ_1, \dots, ξ_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 $X_{n,1}, \dots, X_{n,m}$ are independent. If*

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

for some fixed $\varepsilon > 0$, then the empirical spectral measure μ_{P_n} of the product¹

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

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

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

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

¹Here and in the sequel, we use Pi (product) notation for products of matrices. To avoid any ambiguity, if M_1, \dots, 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 1.4, the order of matrices in the product is irrelevant by simply relabeling indices.

Remark 1.5. 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.

Theorem 1.4 is a special case of [47, Theorem 2.4], which applies to so-called elliptic random matrices. Theorem 1.4 and the results in [47] are stated only for real random variables, but the proofs can be extended to the complex setting. Similar results have also been obtained in [23,38,48]. The Gaussian case was originally considered by Burda, Janik, and Waclaw [28]; see also [27]. We refer the reader to [1–6,29,36,37,43,56] and references therein for many other results concerning products of random matrices with Gaussian entries.

2. Main results

From Theorem 1.4, 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 1.3, we will assume the atom variables ξ_1, \dots, ξ_m have finite fourth moment.

Assumption 2.1. The complex-valued random variables ξ_1, \dots, ξ_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 by showing there are no outliers 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 ξ_1, \dots, ξ_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 ξ_1, \dots, ξ_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 $n \rightarrow \infty$. 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 [44] under the additional assumption that the atom variables ξ_1, \dots, ξ_m satisfy a sub-exponential decay condition. In particular, this condition implies that all moments of ξ_1, \dots, ξ_m are finite. Theorem 2.2 only requires the fourth moments of the atom variables to be finite.

Remark 2.4. It is natural to conjecture, analogously as in [25], that the spectral radius of P_n is bounded above by $\sigma + o(1)$ in probability as $n \rightarrow \infty$ only assuming the atom variables ξ_1, \dots, ξ_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 1.4, it is natural to consider perturbations of the form

$$P_n := \prod_{k=1}^m \left(\frac{1}{\sqrt{n}} X_{n,k} + A_{n,k} \right). \tag{3}$$

However, there are many other types of perturbations one might consider, such as multiplicative perturbations $P_n := \frac{1}{\sqrt{n}} X_{n,1} (I + A_{n,1}) \frac{1}{\sqrt{n}} X_{n,2} (I + A_{n,2}) \cdots \frac{1}{\sqrt{n}} X_{n,m} (I + A_{n,m})$ 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 show that only the deterministic term A_n determines the location of the outliers.

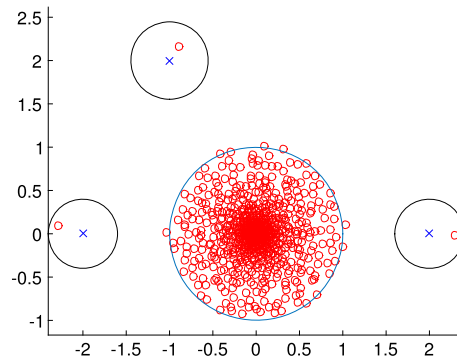


Fig. 2. In the above figure, we display the eigenvalues of products of random matrices of the form $\prod_{k=1}^5 \left(\frac{1}{\sqrt{1000}} X_k + A_k \right)$, where X_1, \dots, X_5 are 1000×1000 independent iid matrices with symmetric ± 1 Bernoulli entries, and the product of the deterministic matrices A_1, \dots, A_5 is $\text{diag}(-2, -1 + 2i, 2, 0, \dots, 0)$. Each nonzero eigenvalue of the product $A_1 \cdots A_5$ is marked with a cross.

The “mixed” terms M_n do not effect the asymptotic location of the outliers. While our methods apply to several types of perturbations, we will focus on perturbations of the sort in (3). Results for other types of perturbations appear in an earlier version of the current paper, see [34, Theorems 2.5, 2.6, and 2.7], and were removed from the current draft due to space limitations.

Theorem 2.5. *Let $m \geq 1$ be an integer, and assume ξ_1, \dots, ξ_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 ξ_1, \dots, ξ_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*

$$P_n := \prod_{k=1}^m \left(\frac{1}{\sqrt{n}} X_{n,k} + A_{n,k} \right), \quad A_n := \prod_{k=1}^m A_{n,k}, \quad (4)$$

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 $\lambda_1(A_n), \dots, \lambda_j(A_n)$ 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 the product 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 $n \rightarrow \infty$ for each $1 \leq i \leq j$.

Theorem 2.5 can be viewed as a generalization of Theorem 1.3. However, the most interesting cases occur when $m \geq 2$. Indeed, in these cases, Theorem 2.5 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 2 presents a numerical simulation of Theorem 2.5.

The methods of this paper can be generalized to the case where $X_{n,1} = \cdots = X_{n,m}$ almost surely; see [34, Section 3] for details. It is natural to ask if the same results can be extended to the elliptical random matrices considered in [47] (see Theorem 1.4). However, the techniques used in this paper (in particular, the combinatorial techniques in Section 10) 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. Beyond the location of the outliers, one can ask about fluctuations. For instance, in [53], the joint fluctuations of the outlier eigenvalues from Theorem 1.3 were studied.

2.1. Outline

In Sections 3 and 4, we present some preliminary results and tools. In particular, Section 3 presents our notation and some standard linearization results that will be used throughout the paper. Section 4 contains a key eigenvalue criterion lemma and a brief overview of the proofs of the main results. In Section 5, we state an isotropic limit law and use it to prove our main results. The majority of the paper (Sections 6–10) is devoted to the proof of this isotropic limit law. A few auxiliary results and proofs are presented in the appendices.

3. 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 4, we present a brief overview of our proofs and explain how these concepts will be used.

3.1. Notation

We use asymptotic notation (such as O, o, Ω) under the assumption that $n \rightarrow \infty$. In particular, $X = O(Y)$, $Y = \Omega(X)$, $X \ll Y$, and $Y \gg X$ denote the estimate $|X| \leq CY$, for some constant $C > 0$ independent of n and for all $n \geq C$. 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 $|X| \leq c(n)Y$ for some sequence $c(n)$ that goes to zero as $n \rightarrow \infty$. Specifically, $o(1)$ denotes a term which tends to zero as $n \rightarrow \infty$. 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 $\mathbb{P}(E) \geq 1 - O_C(n^{-C})$ 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 $\lambda_1(M), \dots, \lambda_n(M)$, and we let $\rho(M) := \max\{|\lambda_1(M)|, \dots, |\lambda_n(M)|\}$ 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 $s_1(M_n) \geq s_2(M_n) \geq \dots \geq s_n(M_n)$.

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 $\{1, \dots, N\}$. 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.

3.2. Linearization

Let M_1, \dots, 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

$$\mathcal{M} := \begin{bmatrix} 0 & M_1 & & 0 \\ 0 & 0 & M_2 & 0 \\ & & \ddots & \ddots \\ 0 & & & 0 & M_{m-1} \\ M_m & & & & 0 \end{bmatrix}. \tag{5}$$

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, [7,28,39,47,48] and references therein.

Proposition 3.1. *Let M_1, \dots, 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 (5). 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, \dots, Z_m has the same characteristic polynomial² as P , it follows that

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

for all $z \in \mathbb{C}$. □

We will exploit Proposition 3.1 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

$$\mathcal{Y}_n := \begin{bmatrix} 0 & X_{n,1} & & 0 \\ 0 & 0 & X_{n,2} & 0 \\ & & \ddots & \ddots \\ 0 & & & 0 & X_{n,m-1} \\ X_{n,m} & & & & 0 \end{bmatrix} \tag{6}$$

and its resolvent

$$\mathcal{G}_n(z) := \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right)^{-1}, \tag{7}$$

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 5.1 below.

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

$$\mathcal{A}_n := \begin{bmatrix} 0 & A_{n,1} & & 0 \\ 0 & 0 & A_{n,2} & 0 \\ & & \ddots & \ddots \\ 0 & & & 0 & A_{n,m-1} \\ A_{n,m} & & & & 0 \end{bmatrix}. \tag{8}$$

3.3. 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 $1 \leq i, j \leq n$, 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 $1 \leq i, j \leq N$. 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 $1 \leq a, b \leq m$, 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 $1 \leq i, j \leq n$, the notation $\mathcal{M}_{ij}^{[a,b]}$ or $\mathcal{M}_{(i,j)}^{[a,b]}$ 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 $1 \leq i, j \leq n$, 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 $\mathcal{M}_{n,(i,j)}^{[a,b]}$ to denote the (i, j) -entry of the block $\mathcal{M}_n^{[a,b]}$.

²This fact can easily be deduced from Sylvester's determinant theorem; see (9).

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 $v^{[1]}, \dots, v^{[m]}$ 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, $v_i^{[a]}$ denotes the i th entry in block a .

3.4. Singular value inequalities

For an $n \times n$ matrix M , recall that $s_1(M) \geq \dots \geq s_n(M)$ denote its ordered singular values. We will need the following elementary bound concerning the largest and smallest singular values.

Proposition 3.2. *Let M be an $n \times n$ matrix and assume $E \subseteq \mathbb{C}$ such that $\inf_{z \in E} s_n(M - zI) \geq c$ for some constant $c > 0$. Then $\sup_{z \in E} \|G(z)\| \leq \frac{1}{c}$ 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. \square

4. 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 4.1), which is based on Sylvester’s determinant theorem:

$$\det(I + AB) = \det(I + BA) \tag{9}$$

whenever A is an $n \times k$ matrix and B is a $k \times n$ matrix. In particular, the left-hand side of (9) 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 2.5. 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 3.1 with the matrix $\frac{1}{\sqrt{n}} \mathcal{Y}_n + \mathcal{A}_n$, where \mathcal{Y}_n and \mathcal{A}_n are defined in (6) and (8). Indeed, by Proposition 3.1, 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

$$\det\left(\frac{1}{\sqrt{n}} \mathcal{Y}_n + \mathcal{A}_n - zI\right) = 0 \tag{10}$$

for $z \in \mathbb{C}$ with $|z| > \sigma^{1/m} + \delta$. Assuming z is not eigenvalue of \mathcal{Y}_n , we can rewrite (10) as

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

where the resolvent $\mathcal{G}_n(z)$ is defined in (7). From (9) 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 5 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 [41,42] for Wigner random matrices as well as the isotropic law verified in [46] 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 (similar versions of the lemma below have appeared in [8,17–19,31,32,41,42,46,49,52,54,57]).

Lemma 4.1 (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 $\det(I_k + C(Y - zI_n)^{-1}B) = 0$.*

Proof of Lemma 4.1. Assume z is not an eigenvalue of Y . Then $\det(Y - zI) \neq 0$ and $\det(Y + A - zI) = \det(Y - zI) \det(I + (Y - zI)^{-1}BC)$. Thus, by (9), z is an eigenvalue of $Y + A$ if and only if $\det(I + C(Y - zI)^{-1}B) = 0$, as desired. \square

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

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1} \tag{11}$$

whenever A and B are invertible.

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

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

Theorem 5.1 (Isotropic limit law). *Let $m \geq 1$ be a fixed integer, and assume ξ_1, \dots, ξ_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 $X_{n,1}, \dots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables ξ_1, \dots, ξ_m , respectively. Recall that \mathcal{Y}_n is defined in (6) and its resolvent $\mathcal{G}_n(z)$ is defined in (7). 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 $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$.*
- (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*

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

almost surely as $n \rightarrow \infty$.

We conclude this section with the proofs of Theorems 2.2 and 2.5 assuming Theorem 5.1. Sections 6–10 are devoted to the proof of Theorem 5.1.

5.1. 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 3.1, 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 5.1, we conclude that almost surely, $\limsup_{n \rightarrow \infty} \rho(\frac{1}{\sqrt{n}} \mathcal{Y}_n) \leq 1$, where $\rho(M)$ denotes the spectral radius of the matrix M . This completes the proof of Theorem 2.2.

5.2. Proof of Theorem 2.5

Recall that

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

By rescaling P_n by $\frac{1}{\sigma}$, we may assume that $\sigma_i = 1$ for $1 \leq i \leq n$. 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 $\lambda_1(A_n), \dots, \lambda_j(A_n)$ that lie in the region $\{z \in \mathbb{C} : |z| \geq 1 + 3\varepsilon\}$.

Let \mathcal{Y}_n and \mathcal{A}_n be defined as in (6) and (8). Using Proposition 3.1, 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

$$\det\left(\frac{1}{\sqrt{n}}\mathcal{Y}_n + \mathcal{A}_n - zI\right) = 0 \tag{12}$$

for $|z| \geq 1 + \varepsilon'$. By part (i) of Theorem 5.1, almost surely, for n sufficiently large $\frac{1}{\sqrt{n}}\mathcal{Y}_n - zI$ is invertible for all $|z| \geq 1 + \varepsilon'$. 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 4.1, we need to investigate the values of $z \in \mathbb{C}$ with $|z| \geq 1 + \varepsilon'$ such that

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

where $\mathcal{G}_n(z)$ is defined in (7).

Since $k = O(1)$, Theorem 5.1 implies that

$$\sup_{|z| \geq 1 + \varepsilon'} \left\| \mathcal{C}_n\mathcal{G}_n(z)\mathcal{B}_n - \left(-\frac{1}{z}\right)\mathcal{C}_n\mathcal{B}_n \right\| \rightarrow 0 \tag{13}$$

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

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

almost surely. By an application of Sylvester’s determinant theorem (9), this is equivalent to

$$\sup_{|z| \geq 1 + \varepsilon'} \left| \det(I_k + \mathcal{C}_n\mathcal{G}_n\mathcal{B}_n) - \det\left(I_n - \frac{1}{z}\mathcal{A}_n\right) \right| \rightarrow 0 \tag{14}$$

almost surely as $n \rightarrow \infty$. 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 $\lambda_1(\mathcal{A}_n), \dots, \lambda_l(\mathcal{A}_n)$ outside the disk $\{z \in \mathbb{C} : |z| < 1 + \varepsilon'\}$. Thus, by (14) and Rouché’s theorem, almost surely, for n sufficiently large,

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

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 (12), 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 $(\frac{1}{\sqrt{n}}\mathcal{Y}_n + \mathcal{A}_n)^m$ 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

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

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 2.5.

6. Truncation and useful tools

We now turn to the proof of Theorem 5.1. We will require the following standard truncation results for iid random matrices.

Lemma 6.1. *Let ξ be a complex-valued random variable with mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. Let $\text{Re}(\xi)$ and $\text{Im}(\xi)$ denote the real and imaginary parts of ξ respectively, and let $\sqrt{-1}$ denote the imaginary unit. For $L > 0$, define*

$$\begin{aligned} \tilde{\xi} &:= \text{Re}(\xi) \mathbf{1}_{\{|\text{Re}(\xi)| \leq L/\sqrt{2}\}} - \mathbb{E}[\text{Re}(\xi) \mathbf{1}_{\{|\text{Re}(\xi)| \leq L/\sqrt{2}\}}] \\ &\quad + \sqrt{-1} (\text{Im}(\xi) \mathbf{1}_{\{|\text{Im}(\xi)| \leq L/\sqrt{2}\}} - \mathbb{E}[\text{Im}(\xi) \mathbf{1}_{\{|\text{Im}(\xi)| \leq L/\sqrt{2}\}}]) \end{aligned}$$

and

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

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) $\text{Var}(\hat{\xi}) \geq \frac{1}{2}$
- (ii) $|1 - \text{Var}(\hat{\xi})| \leq \frac{4}{L^2} \mathbb{E}|\xi|^4$
- (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

$$\begin{aligned} \tilde{X}_{(i,j)} &:= \text{Re}(X_{(i,j)}) \mathbf{1}_{\{|\text{Re}(X_{(i,j)})| \leq L/\sqrt{2}\}} - \mathbb{E}[\text{Re}(X_{(i,j)}) \mathbf{1}_{\{|\text{Re}(X_{(i,j)})| \leq L/\sqrt{2}\}}] \\ &\quad + \sqrt{-1} (\text{Im}(X_{(i,j)}) \mathbf{1}_{\{|\text{Im}(X_{(i,j)})| \leq L/\sqrt{2}\}} - \mathbb{E}[\text{Im}(X_{(i,j)}) \mathbf{1}_{\{|\text{Im}(X_{(i,j)})| \leq L/\sqrt{2}\}}]) \end{aligned} \tag{15}$$

and

$$\hat{X}_{(i,j)} := \frac{\tilde{X}_{(i,j)}}{\sqrt{\text{Var}(\tilde{X}_{(i,j)})}} \quad \text{for } 1 \leq i, j \leq n. \tag{16}$$

Lemma 6.2. *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 (16). Then, there exist constants $C, L_0 > 0$ (depending only on m_4) such that for all $L > L_0$, we have $\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|X - \hat{X}\| \leq \frac{C}{L}$ almost surely.*

Proof. Let \tilde{X} be defined as in (15), and let L_0 be the value from Lemma 6.1. Begin by noting that $\|X - \hat{X}\| \leq \|X - \tilde{X}\| + \|\tilde{X} - \hat{X}\|$ and thus it suffices to show that $\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|X - \tilde{X}\| \leq \frac{C}{L}$ and $\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|\tilde{X} - \hat{X}\| \leq \frac{C}{L}$ almost

surely. First, by Lemma 6.1,

$$\begin{aligned} \frac{1}{\sqrt{n}} \|\tilde{X} - \hat{X}\| &= \frac{1}{\sqrt{n}} \|\tilde{X}\| \left| 1 - \frac{1}{\sqrt{\text{Var}(\tilde{\xi})}} \right| \leq \frac{1}{\sqrt{n}} \|\tilde{X}\| \sqrt{2} |\sqrt{\text{Var}(\tilde{\xi})} - 1| \\ &\leq \frac{1}{\sqrt{n}} \|\tilde{X}\| \sqrt{2} |\text{Var}(\tilde{\xi}) - 1| \leq \frac{1}{\sqrt{n}} \|\tilde{X}\| \sqrt{2} \left(\frac{4}{L^2} \mathbb{E}|\xi|^4 \right). \end{aligned} \tag{17}$$

By [57, Theorem 1.4], we find that almost surely $\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|\tilde{X}\| \leq 2$, and thus by (17) we have $\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|\tilde{X} - \hat{X}\| \leq \frac{C}{L}$ almost surely for all $L \geq \max\{1, L_0\}$.

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

$$\begin{aligned} &\text{Re}(X_{(i,j)}) \mathbf{1}_{\{|\text{Re}(X_{(i,j)})| > L/\sqrt{2}\}} - \mathbb{E}[\text{Re}(X_{(i,j)}) \mathbf{1}_{\{|\text{Re}(X_{(i,j)})| > L/\sqrt{2}\}}] \\ &+ \sqrt{-1} (\text{Im}(X_{(i,j)}) \mathbf{1}_{\{|\text{Im}(X_{(i,j)})| > L/\sqrt{2}\}} - \mathbb{E}[\text{Im}(X_{(i,j)}) \mathbf{1}_{\{|\text{Im}(X_{(i,j)})| > L/\sqrt{2}\}}]). \end{aligned}$$

Thus, each entry has mean zero, variance $\text{Var}((X - \tilde{X})_{i,j}) \leq \frac{8}{L^2} \mathbb{E}|\xi|^4$, and finite fourth moment. Thus, by [57, Theorem 1.4], we have $\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|X - \tilde{X}\| \leq \frac{C}{L}$ almost surely, and the proof is complete. \square

We now consider the iid random matrices $X_{n,1}, \dots, X_{n,m}$ from Theorem 5.1. For each $1 \leq k \leq m$, define the truncation $\hat{X}_{n,k}$ as was done above for \hat{X} in (16). Define $\hat{\mathcal{Y}}_n$ by

$$\hat{\mathcal{Y}}_n := \begin{bmatrix} 0 & \hat{X}_{n,1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{X}_{n,m-1} \\ \hat{X}_{n,m} & 0 & \dots & 0 \end{bmatrix}. \tag{18}$$

Using Theorem 6.1, we have the following corollary.

Corollary 6.3. *Let $X_{n,1}, \dots, X_{n,m}$ be independent iid random matrices with atom variables ξ_1, \dots, ξ_m , each of which has mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. Let $\hat{X}_{n,1}, \dots, \hat{X}_{n,m}$ be the truncations of $X_{n,1}, \dots, X_{n,m}$ as was done in (16). In addition, let \mathcal{Y}_n be as defined in (6) and $\hat{\mathcal{Y}}_n$ be as defined in (18). Then there exist constants $C, L_0 > 0$ (depending only on the atom variables ξ_1, \dots, ξ_m) such that*

$$\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \|\mathcal{Y}_n - \hat{\mathcal{Y}}_n\| \leq \frac{C}{L} \tag{19}$$

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

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

Therefore, the claim follows from Lemma 6.2. \square

7. 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, [22, Problem III.6.5]), which states that for $n \times n$ matrices A and B ,

$$\max_{1 \leq i \leq n} |s_i(A) - s_i(B)| \leq \|A - B\|. \tag{20}$$

We require the following theorem, which is based on [44, Theorem 2].

Theorem 7.1. Fix $L > 0$, and let ξ_1, \dots, ξ_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 $X_{n,1}, \dots, X_{n,m}$ be independent iid random matrices with atom variables ξ_1, \dots, ξ_m , respectively. Define \mathcal{Y}_n as in (6), 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 [44, Theorem 2], where the same conclusion was shown to hold almost surely rather than with overwhelming probability. However, many of the intermediate steps in [44] are proven to hold with overwhelming probability. We use these intermediate steps to prove Theorem 7.1 in Appendix B.

Lemma 7.2. Let $X_{n,1}, \dots, X_{n,m}$ satisfy the assumptions of Theorem 5.1, and let \mathcal{Y}_n be as defined in (6). 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 $\hat{X}_{n,1}, \dots, \hat{X}_{n,m}$ be defined as in (16), and let $\hat{\mathcal{Y}}_n$ be defined as in (18). By Theorem 7.1 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 6.3 and (20) we may choose L sufficiently large to ensure that almost surely, for n sufficiently large, $|s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) - s_{mn} \left(\frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n - zI \right)| \leq \frac{c'}{2}$, uniformly in z . We conclude that almost surely, for n sufficiently large, $\inf_{|z| \geq 1 + \delta} s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \geq \frac{c'}{2}$, and the proof is complete. \square

With this result we can prove the following lemma.

Lemma 7.3. Let $X_{n,1}, \dots, X_{n,m}$ satisfy the assumptions of Theorem 5.1, and fix $\delta > 0$. Then there exists a constant $c > 0$ such that almost surely, for n sufficiently large, $\inf_{|z| \geq 1 + \delta} s_{mn} (n^{-m/2} X_{n,1} \cdots X_{n,m} - z^m I) \geq c$.

Proof. Let \mathcal{Y}_n be defined as in (6). Then Lemma 7.2 implies that almost surely, for n sufficiently large, $\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI$ is invertible for all $|z| \geq 1 + \delta$. By computing the block inverse of this matrix, we find

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

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

$$\begin{aligned} \left\| (n^{-m/2} X_{n,1} \cdots X_{n,m} - z^m I)^{-1} \right\| &\leq |z|^{m-1} \left\| (n^{-m/2} X_{n,1} \cdots X_{n,m} - z^m I)^{-1} \right\| \\ &\leq \left\| \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right)^{-1} \right\|, \end{aligned}$$

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

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

and the claim follows from Lemma 7.2. \square

8. Reductions to the proof of Theorem 5.1

In this section, we will prove that it is sufficient to reduce the proof of Theorem 5.1 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 8.1. *Let $X_{n,1}, \dots, X_{n,m}$ be as in Theorem 5.1. Then there exists a constant L_0 such that the following holds for all $L > L_0$. Let $\hat{X}_{n,1}, \dots, \hat{X}_{n,m}$ be defined as in (16), let $\hat{\mathcal{Y}}_n$ be given by (18), and let $\hat{\mathcal{G}}_n(z) := (\frac{1}{\sqrt{n}}\hat{\mathcal{Y}}_n - zI)^{-1}$.*

- (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 $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$.*
- (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*

$$\sup_{z \in \mathbb{C}: 5 \leq |z| \leq 6} \left| u_n^* \hat{\mathcal{G}}_n(z) v_n + \frac{1}{z} u_n^* v_n \right| \rightarrow 0 \quad \text{almost surely as } n \rightarrow \infty.$$

Proof of Theorem 5.1 assuming Theorem 8.1. Part (i) of Theorem 5.1 follows from Lemma 7.2. In addition, part (ii) of Theorem 5.1 follows from an application of Lemma 7.2 and Proposition 3.2.

We now turn to the proof of part (iii). Fix $0 < \delta < 1$. Let \mathcal{Y}_n be given by (6), and let \mathcal{G}_n be given by (7). Let $\varepsilon, \varepsilon' > 0$, and observe that there exists a positive constant M_1 such that for $|z| \geq M_1$, we have $\|(-\frac{1}{z})u^*v\| \leq \frac{1}{|z|} \|u^*\| \|v\| \leq \frac{\varepsilon}{2}$. Also, by [57, 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,

$$\sup_{|z| \geq M} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \leq \varepsilon. \tag{21}$$

We now work on the region where $1 + \delta < |z| \leq M$. By the resolvent identity (11), we note that $\|\mathcal{G}_n(z) - \hat{\mathcal{G}}_n(z)\| \leq \|\mathcal{G}_n(z)\| \|\hat{\mathcal{G}}_n(z)\| \frac{1}{\sqrt{n}} \|\hat{\mathcal{Y}}_n - \mathcal{Y}_n\|$. Thus, by part (ii) of Theorem 5.1 (proven above), Theorem 8.1, and Corollary 6.3, there exist constants $C, c > 0$ such that

$$\limsup_{n \rightarrow \infty} \sup_{|z| > 1 + \delta} \|\mathcal{G}_n(z) - \hat{\mathcal{G}}_n(z)\| \leq \limsup_{n \rightarrow \infty} c^2 \frac{1}{\sqrt{n}} \|\hat{\mathcal{Y}}_n - \mathcal{Y}_n\| \leq c^2 \frac{C}{L} \leq \frac{\varepsilon'}{2} \tag{22}$$

almost surely for L sufficiently large.

From (22) and Theorem 8.1, almost surely, for n sufficiently large,

$$\begin{aligned} & \sup_{5 \leq |z| \leq 6} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \\ & \leq \sup_{5 \leq |z| \leq 6} |u^* \mathcal{G}_n(z) v - u^* \hat{\mathcal{G}}_n(z) v| + \sup_{5 \leq |z| \leq 6} \left| u^* \hat{\mathcal{G}}_n(z) v + \frac{1}{z} u^* v \right| \leq \varepsilon'. \end{aligned}$$

Since $\varepsilon' > 0$ was arbitrary, this implies that $\limsup_{n \rightarrow \infty} \sup_{5 \leq |z| \leq 6} |u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v| = 0$ 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 Theorem³ (see, for instance, [12, Lemma 2.14]) implies that we can extend this convergence to the larger region, and we obtain $\limsup_{n \rightarrow \infty} \sup_{1 + \delta \leq |z| \leq M} |u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v| = 0$ 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 (21), this completes the proof of Theorem 5.1 (since $\varepsilon > 0$ was arbitrary). □

³The hypothesis of Vitali's Convergence Theorem are satisfied almost surely, for n sufficiently large, by parts (i) and (ii) of Theorem 5.1. 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 (11).

It remains to prove Theorem 8.1. We prove parts (i) and (ii) of Theorem 8.1 now. The proof of part (iii) is lengthy and will be addressed in the forthcoming sections.

Proof of Theorem 8.1 (i) and (ii). Let $\delta > 0$, and observe that by Theorem 7.1 and the Borel–Cantelli lemma, there exists a constant $c > 0$ (depending only on δ) such that almost surely, for n sufficiently large,

$$\inf_{|z|>1+\delta} s_{mn} \left(\frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n - zI \right) \geq c. \tag{23}$$

This implies 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 (23) and Proposition 3.2, we conclude that almost surely, for n sufficiently large, $\sup_{|z|>1+\delta} \|\hat{\mathcal{G}}_n(z)\| \leq \frac{1}{c}$, proving (ii). \square

9. Concentration of bilinear forms involving the resolvent \mathcal{G}_n

Sections 9 and 10 are devoted to the proof of part (iii) of Theorem 8.1. Let $\hat{X}_{n,1}, \dots, \hat{X}_{n,m}$ be the truncated matrices from Theorem 8.1, and let $u_n, v_n \in \mathbb{C}^{mn}$ be deterministic unit vectors. For ease of notation, in Sections 9 and 10, we drop the decorations and write $X_{n,1}, \dots, X_{n,m}$ for $\hat{X}_{n,1}, \dots, \hat{X}_{n,m}$. 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:

$$\Omega_n := \left\{ \frac{1}{\sqrt{n}} \|\mathcal{Y}_n\| \leq 4.5 \right\}. \tag{24}$$

Lemma 9.1 (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 [12, Theorem 5.9] (alternatively, [57, 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. \square

By Lemma 9.1, Ω_n holds with overwhelming probability, i.e., for every $A > 0$,

$$\mathbb{P}(\Omega_n) = 1 - O_A(n^{-A}). \tag{25}$$

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

Lemma 9.2 (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*

$$\sup_{5 \leq |z| \leq 6} |u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n} - \mathbb{E}[u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n}]| < \varepsilon \tag{26}$$

for n sufficiently large.

The proof of Lemma 9.2 follows the arguments of [10,46]. 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 k th row and the k th column filled with zeros. Note that $\mathcal{Y}_n^{(k)}$ is still an $mn \times mn$ matrix. Define

$$\mathcal{G}_n^{(k)} := \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n^{(k)} - zI \right)^{-1}, \tag{27}$$

let r_k denote the k th row of \mathcal{Y}_n , and let c_k denote k th column of \mathcal{Y}_n . Also define the σ -algebra $\mathcal{F}_k := \sigma(r_1, \dots, r_k, c_1, \dots, c_k)$ 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

$$\mathbb{E}_k[\cdot] := \mathbb{E}[\cdot | \mathcal{F}_k] \tag{28}$$

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

$$\Omega_n^{(k)} := \left\{ \frac{1}{\sqrt{n}} \|\mathcal{Y}_n^{(k)}\| \leq 4.5 \right\}. \tag{29}$$

Observe that $\Omega_n \subseteq \Omega_n^{(k)}$ and therefore, by Lemma 9.1, $\Omega_n^{(k)}$ holds with overwhelming probability as well.

Remark 9.3. By Lemma D.3, we have that

$$\sup_{5 \leq |z| \leq 6} \|\mathcal{G}_n(z)\| \leq 2, \quad \sup_{5 \leq |z| \leq 6} \|\mathcal{G}_n^{(k)}(z)\| \leq 2$$

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

We will now collect some preliminary calculations and lemmata that will be required for the proof of Lemma 9.2.

Lemma 9.4 (Rosenthal’s Inequality; [30]). *Let $\{x_k\}$ be a complex martingale difference sequence with respect to the filtration \mathcal{F}_k . Then for $p \geq 2$,*

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

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

Lemma 9.5. *Let A be an $n \times n$ Hermitian positive semidefinite matrix, and let $S \subset [n]$. Then $\sum_{i \in S} A_{ii} \leq \text{tr } A$.*

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 9.6. *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 $|\xi| \leq L$ 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 $p \geq 1$,

$$\mathbb{E} |w^* A w|^p \ll_{L,p} \|A\|^p. \tag{30}$$

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

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

Since the rank of $A_{S \times S}$ is at most one, we find $\text{tr } A_{S \times S} \leq \|A\|$ and $\text{tr } A_{S \times S}^p \leq \|A\|^p$, where we used the fact that the operator norm of a matrix bounds the operator norm of any sub-matrix. We conclude that $\mathbb{E} |w^* A w|^p \ll_p \|A\|^p + L^{2p} \|A\|^p \ll_{L,p} \|A\|^p$, as desired. \square

Lemma 9.7. *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 $t = (t_j)_{j=1}^N$ 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 $p \geq 1$,

$$\mathbb{E} |w^* A t|^p \ll_{L,p} (\text{tr}(A^* A))^{p/2}. \tag{31}$$

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

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

Recall that for any matrix B , $\operatorname{tr}(B^* B)^{p/2} \leq (\operatorname{tr}(B^* B))^{p/2}$. By this fact and by Lemma 9.5, we observe that

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

By a cyclic permutation of the trace, we have

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

By Lemma D.1, Lemma 9.5, and a similar argument as above, we obtain

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

completing the proof. \square

Lemma 9.8. Let r_k be the k th row of \mathcal{Y}_n , c_k be the k th 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,

$$\sup_{5 \leq |z| \leq 6} \mathbb{E}_{k-1} \left| \frac{1}{n} r_k \mathcal{G}_n^{(k)}(z) u_n u_n^* \mathcal{G}_n^{(k)*}(z) \mathbf{1}_{\Omega_n^{(k)}} r_k^* \right|^p \ll_{L,p} n^{-p} \quad (32)$$

and

$$\sup_{5 \leq |z| \leq 6} \mathbb{E}_{k-1} \left| \frac{1}{n} c_k^* \mathcal{G}_n^{(k)*}(z) u_n u_n^* \mathcal{G}_n^{(k)}(z) \mathbf{1}_{\Omega_n^{(k)}} c_k \right|^p \ll_{L,p} n^{-p} \quad (33)$$

almost surely.

Proof. We will only prove the bound in (32) as the proof of (33) 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

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

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, we can observe that the matrix $\mathcal{G}_n^{(k)} u_n u_n^* \mathcal{G}_n^{(k)*} \mathbf{1}_{\Omega_n^{(k)}}$ is Hermitian positive semidefinite with rank at most one. Applying Lemma 9.6 and Remark 9.3, we obtain

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

surely, and the proof is complete. \square

Lemma 9.9. Let $\zeta_1, \dots, \zeta_{mn}$ 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 $p \geq 1$,

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

Proof. We will exploit the fact that $\Omega_n \subseteq \Omega_n^{(k)}$ for any $1 \leq k \leq mn$. Indeed, we have $\mathbb{E} |\sum_{k=1}^{mn} (\zeta_k \mathbf{1}_{\Omega_n^{(k)}} - \zeta_k \mathbf{1}_{\Omega_n \cap \Omega_n^{(k)}})|^p \leq \mathbb{E} |\sum_{k=1}^{mn} \zeta_k \mathbf{1}_{\Omega_n^{(k)} \cap \Omega_n^c}|^p \ll_p n^p \mathbb{P}(\Omega_n^c)$, and the claim follows from (25). \square

We will made use of the Sherman–Morrison rank one perturbation formula (see [40, 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

$$(A + uv^*)^{-1} = A^{-1} - \frac{A^{-1} u v^* A^{-1}}{1 + v^* A^{-1} u} \quad (34)$$

and

$$(A + uv^*)^{-1} u = \frac{A^{-1} u}{1 + v^* A^{-1} u}. \quad (35)$$

Now, we proceed to prove the main result of this section, Lemma 9.2.

Proof of Lemma 9.2. Define

$$\mathcal{Y}_n^{(k1)} := \mathcal{Y}_n^{(k)} + e_k r_k, \quad \mathcal{Y}_n^{(k2)} := \mathcal{Y}_n^{(k)} + c_k e_k^* \quad (36)$$

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

$$\mathcal{G}_n^{(kj)} := \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n^{(kj)} - zI \right)^{-1}, \quad j = 1, 2, \quad (37)$$

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,

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

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

(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 (34), (i), and (ii),

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

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

(iv) By the same argument as (iii), we have $n^{-1/2}r_k\mathcal{G}_n^{(k2)}e_k = z^{-1}n^{-1}r_k\mathcal{G}_n^{(k)}c_k$, 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 $5 \leq |z| \leq 6$, by Remark 9.3,

$$|\alpha_n^{(k)}| = \left| \frac{z}{z+n^{-1}r_k\mathcal{G}_n^{(k)}c_k} \right| = |z| |(\mathcal{G}_n)_{(k,k)}| \leq 12.$$

On Ω_n^c , $\alpha_n^{(k)} = 1$, so we have that, almost surely, $|\alpha_n^{(k)}| \leq 12$.

(vi) By (35) 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 (35) 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 (34) and (ii),

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

(ix) By (34) 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 $\zeta_n^{(k)}$,

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

(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 9.3, $\|\mathcal{G}_n^{(k)*}u_n v_n^*\mathcal{G}_n^{(k)*}\mathcal{G}_n^{(k)}v_n u_n^*\mathcal{G}_n^{(k)}\| \leq 16$ on $\Omega_n^{(k)}$. Thus, by Lemma 9.7, for any $p \geq 2$,

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

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 9.3 give the almost sure bound $\|\mathcal{G}_n^{(k)*}\mathcal{G}_n^{(k)}\| \leq 4$ on $\Omega_n^{(k)}$, which gives

$$\begin{aligned} \sup_{1 \leq k \leq n} \mathbb{E}_{k-1} [|\zeta_n^{(k)}|^p \mathbf{1}_{\Omega_n}] &\leq \sup_{1 \leq k \leq n} \mathbb{E}_{k-1} [|\zeta_n^{(k)}|^p \mathbf{1}_{\Omega_n^{(k)}}] \\ &\ll_{L,p} \sup_{1 \leq k \leq n} n^{-p} \mathbb{E}_{k-1} [(\text{tr}(\mathcal{G}_n^{(k)*}\mathcal{G}_n^{(k)}))^p \mathbf{1}_{\Omega_n^{(k)}}] \ll_{L,p} n^{-p/2}. \end{aligned}$$

With the above observations in hand, we now complete the proof. By the resolvent identity (11) and Remark 9.3, it follows that the function

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

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,

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

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

$$\mathbb{E}|u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} - \mathbb{E}[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n}]|^p = O_{L,p}(n^{-p/2}) \quad (38)$$

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

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

In view of Lemma 9.9 and since $\Omega_n \cap \Omega_n^{(k)} = \Omega_n$, it suffices to prove that

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

for all $p > 2$. Define

$$W_k := (\mathbb{E}_k - \mathbb{E}_{k-1})[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} - u_n^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n}] \quad (39)$$

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

From the resolvent identity (11), we observe that

$$\begin{aligned} \sum_{k=1}^{mn} W_k &= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n \frac{1}{\sqrt{n}} (\mathcal{Y}_n^{(k)} - \mathcal{Y}_n) \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right] \\ &= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \mathcal{G}_n \left(\frac{1}{\sqrt{n}} e_k r_k + \frac{1}{\sqrt{n}} c_k e_k^* \right) \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right] \\ &:= - \sum_{k=1}^{mn} (W_{k1} + W_{k2}) \end{aligned}$$

where we define $W_{k1} := (\mathbb{E}_k - \mathbb{E}_{k-1})[u_n^* \mathcal{G}_n n^{-1/2} e_k r_k \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n}]$, and $W_{k2} := (\mathbb{E}_k - \mathbb{E}_{k-1})[u_n^* \mathcal{G}_n n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n}]$.

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

$$\begin{aligned} \sum_{k=1}^{mn} W_{k1} &= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [u_n^* \mathcal{G}_n n^{-1/2} e_k r_k \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n}] \\ &= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [u_n^* \mathcal{G}_n^{(k2)} n^{-1/2} e_k r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \\ &= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} (\bar{u}_{n,k} - u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k) n^{-1/2} r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \\ &= - \sum_{k=1}^{mn} (W_{k11} + W_{k12}) \end{aligned}$$

where $W_{k11} := (\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}\bar{u}_{n,k}n^{-1/2}r_k\mathcal{G}_n^{(k)}v_n\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]$ and $W_{k12} := -(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}u_n^*\mathcal{G}_n^{(k)}n^{-1}c_kr_k\mathcal{G}_n^{(k)} \times v_n\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]$. Similarly, by (i), (vi), (viii), and (x),

$$\begin{aligned} \sum_{k=1}^{mn} W_{k2} &= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [u_n^* \mathcal{G}_n n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n}] \\ &= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \frac{\mathcal{G}_n^{(k1)} n^{-1/2} c_k}{1 + r_k^* \mathcal{G}_n^{(k1)} c_k} e_k^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right] \\ &= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[u_n^* \frac{\mathcal{G}_n^{(k1)} n^{-1/2} c_k}{1 + r_k^* \mathcal{G}_n^{(k1)} c_k \mathbf{1}_{\Omega_n}} e_k^* \mathcal{G}_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right] \\ &= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [u_n^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \\ &= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k1)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \\ &= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} v_{n,k} u_n^* (\mathcal{G}_n^{(k)} - \mathcal{G}_n^{(k1)}) e_k n^{-1/2} r_k \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \end{aligned}$$

and continuing the computation we have

$$\begin{aligned} &= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} v_{n,k} (u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} - u_n^* \mathcal{G}_n^{(k)} e_k n^{-1} r_k \mathcal{G}_n^{(k)} c_k \alpha_n^{(k)}) \mathbf{1}_{\Omega_n}] \\ &= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} v_{n,k} (u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} + \bar{u}_{n,k} z^{-1} n^{-1} r_k \mathcal{G}_n^{(k)} c_k \alpha_n^{(k)}) \mathbf{1}_{\Omega_n}] \\ &= - \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} \mathbf{1}_{\Omega_n} - z^{-1} \bar{u}_{n,k} v_{n,k} \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \\ &= - \sum_{k=1}^{mn} (W_{k21} + W_{k22}) \end{aligned}$$

where $W_{k21} := (\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}v_{n,k}u_n^*\mathcal{G}_n^{(k)}n^{-1/2}c_k\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]$ and $W_{k22} := -(\mathbb{E}_k - \mathbb{E}_{k-1})[z^{-1}\bar{u}_{n,k}v_{n,k}\alpha_n^{(k)}\mathbf{1}_{\Omega_n}]$. Thus, in order to complete the proof, it suffices to show that for all $p > 2$,

$$\mathbb{E} \left| \sum_{k=1}^{mn} W_{k11} \right|^p + \mathbb{E} \left| \sum_{k=1}^{mn} W_{k12} \right|^p + \mathbb{E} \left| \sum_{k=1}^{mn} W_{k21} \right|^p + \mathbb{E} \left| \sum_{k=1}^{mn} W_{k22} \right|^p = O_{L,p}(n^{-p/2}). \tag{40}$$

We bound each term individually. To begin, observe that by Lemma 9.4, Lemma 9.8, and (v), for any $p > 2$,

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

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

where we also used Jensen's inequality and the fact that $|z| \geq 5$. Similarly, by Lemma 9.4, Lemma 9.8, and (v), for any $p > 2$,

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

Next, by Lemma 9.4, (v), and (xii), for any $p > 2$,

$$\begin{aligned}
\mathbb{E} \left| \sum_{k=1}^{mn} W_{k12} \right|^p &= \mathbb{E} \left| \sum_{k=1}^{mn} (E_k - \mathbb{E}_{k-1}) [z^{-1} u_n^* \mathcal{G}_n^{(k)} n^{-1} c_k r_k \mathcal{G}_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \right|^p \\
&= \mathbb{E} \left| \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} \alpha_n^{(k)} \eta_n^{(k)} \mathbf{1}_{\Omega_n}] \right|^p \\
&\ll_p \mathbb{E} \left[\sum_{k=1}^{mn} \mathbb{E}_{k-1} |(\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} \alpha_n^{(k)} \eta_n^{(k)} \mathbf{1}_{\Omega_n}]|^2 \right]^{p/2} \\
&\quad + \mathbb{E} \left[|(\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} \alpha_n^{(k)} \eta_n^{(k)} \mathbf{1}_{\Omega_n}]|^p \right]
\end{aligned}$$

$$\begin{aligned} & \ll_p \left(\mathbb{E} \left[\sum_{k=1}^{mn} \mathbb{E}_{k-1} |\eta_n^{(k)} \mathbf{1}_{\Omega_n}|^2 \right]^{p/2} + \sum_{k=1}^{mn} \mathbb{E} |\eta_n^{(k)} \mathbf{1}_{\Omega_n}|^p \right) \\ & \ll_{L,p} \left(\mathbb{E} \left[\sum_{k=1}^{mn} n^{-2} \right]^{p/2} + \sum_{k=1}^{mn} n^{-p} \right) \ll_{L,p} n^{-p/2} + n^{-p+1} \ll_{L,p} n^{-p/2}, \end{aligned}$$

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

$$\begin{aligned} \sum_{k=1}^{mn} W_{k22} &= \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} \bar{u}_{n,k} v_{n,k} \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \\ &= \sum_{i=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-1} \bar{u}_{n,k} v_{n,k} (1 - z^{-1} \zeta_n^{(k)} \alpha_n^{(k)}) \mathbf{1}_{\Omega_n}] \\ &= \sum_{k=1}^{mn} (W_{k221} + W_{k222}) \end{aligned}$$

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 $\mathbb{E} |\sum_{k=1}^{mn} W_{k222}|^p = O_L(p(n^{-p/2}))$ for $p > 2$. By Lemma 9.4, (v), and (xii), we have that for any $p > 2$,

$$\begin{aligned} \mathbb{E} \left| \sum_{k=1}^{mn} W_{k222} \right|^p &= \mathbb{E} \left| \sum_{k=1}^{mn} (\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-2} \bar{u}_{n,k} v_{n,k} \zeta_n^{(k)} \alpha_n^{(k)} \mathbf{1}_{\Omega_n}] \right|^p \\ &\ll_p \mathbb{E} \left[\sum_{k=1}^{mn} \mathbb{E}_{k-1} |(\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-2} \bar{u}_{n,k} v_{n,k} \zeta_n^{(k)} \alpha_n^{(k)} \mathbf{1}_{\Omega_n}]|^2 \right]^{p/2} \\ &\quad + \mathbb{E} \left[\sum_{k=1}^{mn} |(\mathbb{E}_k - \mathbb{E}_{k-1}) [z^{-2} \bar{u}_{n,k} v_{n,k} \zeta_n^{(k)} \alpha_n^{(k)} \mathbf{1}_{\Omega_n}]|^p \right] \\ &\ll_p \mathbb{E} \left[\sum_{k=1}^{mn} |\bar{u}_{n,k}|^2 |v_{n,k}|^2 \mathbb{E}_{k-1} |\zeta_n^{(k)} \mathbf{1}_{\Omega_n}|^2 \right]^{p/2} + \sum_{k=1}^{mn} |\bar{u}_{n,k}|^p |v_{n,k}|^p \mathbb{E} |\zeta_n^{(k)} \mathbf{1}_{\Omega_n}|^p \\ &\ll_{L,p} \left[\sum_{k=1}^{mn} |\bar{u}_{n,k}|^2 |v_{n,k}|^2 n^{-1} \right]^{p/2} + \sum_{k=1}^{mn} |\bar{u}_{n,k}|^p |v_{n,k}|^p n^{-p/2} \ll_{L,p} n^{-p/2}, \end{aligned}$$

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 (40), and hence the proof of Lemma 9.2 is complete. \square

10. Proof of Theorem 8.1

In this section we complete the proof of Theorem 8.1. We continue to work under the assumptions and notation introduced in Section 9. It remains to prove part (iii) of Theorem 8.1. In view of Lemma 9.2 and (25), it suffices to show that

$$\sup_{5 \leq |z| \leq 6} \left| \mathbb{E} \left[u_n^* \mathcal{G}_n(z) v_n \mathbf{1}_{\Omega_n} \right] + \frac{1}{z} u_n^* v_n \right| = o(1). \tag{41}$$

10.1. *Neumann series*

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

$$\frac{1}{\sqrt{n}} \left\| \frac{\mathcal{Y}_n}{z} \right\| \leq \frac{9}{10} < 1 \tag{42}$$

on the event Ω_n , so we may write

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

almost surely. Therefore, using (25), we obtain

$$\begin{aligned} \mathbb{E} \left[u_n^* \mathcal{G}_n v_n \mathbf{1}_{\Omega_n} \right] &= \mathbb{E} \left[-\frac{1}{z} u_n^* v_n \mathbf{1}_{\Omega_n} - u_n^* \sum_{k=1}^{\infty} \frac{\left(\frac{\mathcal{Y}_n}{\sqrt{n}} \right)^k}{z^{k+1}} v_n \mathbf{1}_{\Omega_n} \right] \\ &= -\frac{1}{z} u_n^* v_n \mathbb{P}(\Omega_n) - \sum_{k=1}^{\infty} \frac{1}{z^{k+1}} \mathbb{E} \left[u_n^* \left(\frac{\mathcal{Y}_n}{\sqrt{n}} \right)^k v_n \mathbf{1}_{\Omega_n} \right] \\ &= -\frac{1}{z} u_n^* v_n + o(1) - \sum_{k=1}^{\infty} \frac{1}{z^{k+1}} \mathbb{E} \left[u_n^* \left(\frac{\mathcal{Y}_n}{\sqrt{n}} \right)^k v_n \mathbf{1}_{\Omega_n} \right]. \end{aligned} \tag{43}$$

Showing the sum in (43) converges to zero uniformly in the region $\{z \in \mathbb{C} : 5 \leq |z| \leq 6\}$ will complete the proof of (41).

10.2. *Removing the indicator function*

In this subsection, we prove the following.

Lemma 10.1. *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

$$\begin{aligned} \left| \mathbb{E} \left[u_n^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v_n \right] - \mathbb{E} \left[u_n^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v_n \mathbf{1}_{\Omega_n} \right] \right| &\leq \mathbb{E} \left[\|\mathcal{Y}_n\|^k \mathbf{1}_{\Omega_n^c} \right] \\ &\ll_L n^k \mathbb{P}(\Omega_n^c) \ll_{L,A} n^{k-A}. \end{aligned}$$

Choosing A sufficiently large (in terms of k), completes the proof. □

10.3. Combinatorial arguments

In this section, we will show that

$$\sup_{5 \leq |z| \leq 6} \left| \sum_{k=1}^{\infty} \frac{\mathbb{E}[u_n^* (\frac{1}{\sqrt{n}} \mathcal{Y}_n)^k v_n \mathbf{1}_{\Omega_n}]}{z^{k+1}} \right| = o_L(1). \tag{44}$$

In view of (42), the tail of the series can easily be controlled. Thus, it suffices to show that, for each integer $k \geq 1$, we have $\mathbb{E}[u_n^* (\frac{1}{\sqrt{n}} \mathcal{Y}_n)^k v_n \mathbf{1}_{\Omega_n}] = o_{L,k}(1)$. By Lemma 10.1, it suffices to prove the statement above without the indicator function. In particular, the following lemma completes the proof of (41) (and hence completes the proof of Theorem 8.1).

Lemma 10.2 (Moment Calculations). *Under the assumptions above, for any integer $k \geq 1$, we have $\mathbb{E}[u_n^* (\frac{1}{\sqrt{n}} \mathcal{Y}_n)^k \times v_n] = o_{L,k}(1)$.*

To prove Lemma 10.2, we will expand the above expression in terms of the entries of the random matrices $X_{n,1}, \dots, X_{n,m}$. For brevity, in this section we will drop the subscript n from our notation and just write X_1, \dots, X_m for $X_{n,1}, \dots, X_{n,m}$. 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

$$\begin{aligned} \mathbb{E} \left[u^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v \right] &= n^{-k/2} \mathbb{E} \left[\sum_{1 \leq a, b \leq m} (u^*)^{[a]} (\mathcal{Y}_n^k)^{[a,b]} v^{[b]} \right] \\ &= n^{-k/2} \sum_{1 \leq a, b \leq m} (u^*)^{[a]} \mathbb{E} [(\mathcal{Y}_n^k)^{[a,b]}] v^{[b]}. \end{aligned} \tag{45}$$

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 $\{1, \dots, m\}$ (as opposed to the usual $\{0, \dots, m-1\}$). 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 (45). 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 10.3. 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 $(i_1, i_2, \dots, i_{k+1})$ be a $k+1$ tuple of integers from the set $\{1, 2, \dots, n\}$ and let $1 \leq a \leq m$. We define an *m-colored k-path graph* $G^a(i_1, i_2, \dots, i_{k+1})$ to be the edge-colored directed graph with vertex set $V = \{(1, i_1), (2, i_2), \dots, (k+1, i_{k+1})\}$ 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 $\{1, \dots, m\}$. 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 $G^a(i_1, \dots, i_{k+1})$ 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 10.4. We view k and n as specified parameters. Once these parameters are specified, the notation $G^a(i_1, \dots, i_{k+1})$ 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 $1, 2, \dots, m$, beginning with a . This cycle is repeated as many times as necessary in order to cover all edges.

Notice that we call $G^a(i_1, i_2, \dots, i_{k+1})$ 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 10.5. For a given m -colored k -path graph $G^a(i_1, i_2, \dots, i_{k+1})$, we say two edges e_1 and e_2 of $G^a(i_1, i_2, \dots, i_{k+1})$ 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 10.6. 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 4 for examples of parallel and non parallel edges.

Definition 10.7. For a fixed k , we say two m -colored k -path graphs $G^a(i_1, \dots, i_{k+1})$ and $G^a(i'_1, \dots, i'_{k+1})$ are *equivalent*, denoted $G^a(i_1, \dots, i_{k+1}) \sim G^a(i'_1, \dots, i'_{k+1})$, if there exists a permutation σ of $\{1, \dots, n\}$ with $(i_1, \dots, i_{k+1}) = (\sigma(i'_1), \dots, \sigma(i'_{k+1}))$. 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 10.8. 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 $G^a(i_1, \dots, i_{k+1})$ visits vertex (t, i_t) , then for every $0 < i < i_t$ there exists $s < t$ such that $G^a(i_1, \dots, i_{k+1})$ visits (s, i) .

Remark 10.9. Intuitively, the canonical representation for $G^a(i_1, \dots, i_{k+1})$ 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.

Now, we return to the task at hand: the proof of Lemma 10.2. We fix a positive integer k , and we expand as in (45). For a fixed value $1 \leq a \leq m$, consider the nonzero block $(\mathcal{Y}_n^k)^{[a,b]}$. We further expand to see

$$(u^*)^{[a]} \mathbb{E}[(\mathcal{Y}_n^k)^{[a,b]}] v^{[b]} = \sum_{i_1, \dots, i_{k+1}} \bar{u}_{i_1}^{[a]} \mathbb{E}[X_{a,(i_1,i_2)} X_{a+1,(i_2,i_3)} \cdots X_{a+k,(i_k,i_{k+1})}] v_{i_{k+1}}^{[b]}, \tag{46}$$

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 (46). 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 (46) as a unique m -colored k -path graph.

Definition 10.10. 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 (46) *corresponds* to the m -colored k -path graph $G^a(i_1, \dots, i_{k+1})$. 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 we have that $X_{a,(i_1,i_2)} X_{a+1,(i_2,i_3)} \cdots X_{a+k,(i_k,i_{k+1})}$ 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 (46) 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 10.11. *If two path graphs $G^a(i_1, \dots, i_{k+1})$ and $G^a(i'_1, \dots, i'_{k+1})$ are equivalent, then*

$$\mathbb{E}[x_{G^a(i_1, \dots, i_{k+1})}] = \mathbb{E}[x_{G^a(i'_1, \dots, i'_{k+1})}].$$

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 and figures, and note may be found in [34] (see [34, Examples 11.13, 11.14, and 11.16] and [34, Figures 5, 7, and 8]).

Example 10.12. Consider two 2-colored 3-path graphs, namely $G^1(4, 1, 4, 3)$ and $G^1(1, 2, 1, 3)$. The graph $G^1(4, 1, 4, 3)$ is the leftmost in Figure 3 and $G^1(1, 2, 1, 3)$ is the rightmost graph in Figure 3. 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 $u^*(\frac{1}{\sqrt{n}}\mathcal{Y}_n)^3v$ 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. In Figure 3 as in all figures in this paper, we denote the color of the edge by the numerical value adjacent to that edge.

Example 10.13. 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 4. Since the only index pair which is appears more than once is (i_3, i_4) , the corresponding term will have zero expectation.

We now complete the proof of Lemma 10.2, which will occupy the remainder of the section. Let $\Delta_{n,k}^a := \{G^a(i_1, \dots, i_{k+1}) : 1 \leq i_1, \dots, i_{k+1} \leq n\}$, 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

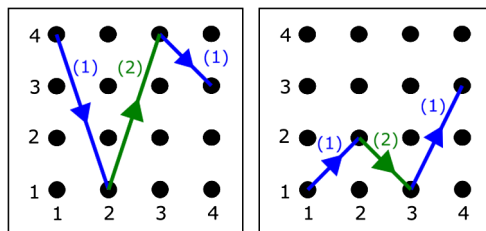


Fig. 3. The graphs featured here are two possible 2-colored 3-path graphs, $G^1(4, 1, 4, 3)$ and $G^1(1, 2, 1, 3)$, which correspond to the terms $X_{1,(4,1)}X_{2,(1,4)}X_{1,(4,3)}$ and $X_{1,(1,2)}X_{2,(2,1)}X_{1,(1,3)}$, respectively. These two graphs are equivalent. The first graph is not a canonical 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 lattice, 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.

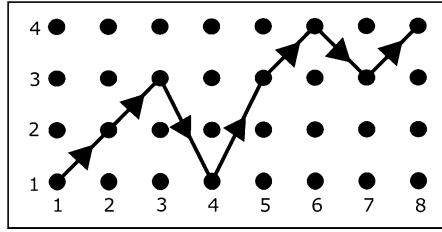


Fig. 4. This 1-colored 7-path graph is the canonical representative to the graph which corresponds to the product in Example 10.13. Note that the edge from (5, 3) to (6, 4) is parallel to the edge from (7, 3) to (8, 4), but no other edges in the graph are parallel. This implies that the corresponding term will have zero expectation.

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, (45) simplifies to

$$\begin{aligned} \mathbb{E} \left[u^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v \right] &= n^{-k/2} \sum_{1 \leq a \leq m} (u^*)^{[a]} \mathbb{E} [(\mathcal{Y}_n^k)^{[a,a]}] v^{[a]} \\ &= n^{-k/2} \sum_{1 \leq a \leq m} \sum_{G \in \Delta_{n,k}^a} u_G^* \mathbb{E} [x_G] v_G \end{aligned} \tag{47}$$

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., $G = G^a(i_1, \dots, i_{k+1})$ for some $i_1, \dots, i_{k+1} \in [n]$. By Lemma 10.11, 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 (47) will be of the form

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

where each matrix contributes only one entry to the above expression. In this case, all terms are independent and the expectation in (47) is zero.

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

$$h(G) =: \max \{ i_t : (t, i_t) \in V_{\tilde{G}} \}, \tag{48}$$

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

$$(\Delta_{n,k}^a)_1 := \{ G \in \Delta_{n,k}^a : h(G) > k/2 \} \tag{49}$$

$$(\Delta_{n,k}^a)_2 := \{ G \in \Delta_{n,k}^a : h(G) = k/2 \} \tag{50}$$

$$(\Delta_{n,k}^a)_3 := \{ G \in \Delta_{n,k}^a : h(G) < k/2 \}. \tag{51}$$

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 $1, 2, \dots, m$ 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

$$\sum_{G \in (\Delta_{n,k}^1)_1} u_G^* \mathbb{E}[x_G] v_G = 0. \tag{52}$$

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 $G \in (\Delta_{n,k}^1)_2$. By Lemma 10.11, 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 $G \in (\Delta_{n,k}^1)_2$ 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 10.14. *Let $k \geq 2$ be any even integer (not necessarily a multiple of m). Then there is only one canonical k -path graph in $\Delta_{n,k}^1$ 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

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

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 $G^1(1, 2, \dots, k/2, 1, 2, \dots, k/2, 1)$ will be parallel to exactly one other edge of the same color. In particular, note that for this graph

$$\begin{aligned} |u_G^* \mathbb{E}[x_G] v_G| &\leq |\bar{u}_1^{[1]}| |\mathbb{E}|X_{1,(1,2)} \cdots X_{m,(k/2,1)}|| |v_1^{[m]}| \\ &\leq |\bar{u}_1^{[1]}| |\mathbb{E}|X_{1,(1,2)}|^2 \cdots |\mathbb{E}|X_{m,(k/2,1)}|^2 |v_1^{[m]}| \\ &\leq |\bar{u}_1^{[1]}| |v_1^{[m]}|. \end{aligned} \tag{53}$$

For ease of notation, let $\tilde{G} := G^1(1, 2, \dots, k/2, 1, 2, \dots, k/2, 1)$, and consider $G^1(i_1, \dots, i_{k+1}) \in (\Delta_{n,k}^1)_2$ such that $G^1(i_1, \dots, i_{k+1}) \sim \tilde{G}$. Observe that there are n options for the first coordinate i_1 of $G^1(i_1, \dots, i_{k+1})$. 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 (53) 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

$$\begin{aligned} \left| \sum_{G \in (\Delta_{n,k}^1)_2} u_G^* \mathbb{E}[x_G] v_G \right| &\leq \sum_{1 \leq i_1 \leq n} n^{k/2-1} |\bar{u}_{i_1}| |v_{i_1}| \\ &\leq n^{k/2-1} \|u\| \|v\| \\ &\leq n^{k/2-1}. \end{aligned} \tag{54}$$

Graphs in $(\Delta_{n,k}^1)_3$: Consider an m -colored k -path graphs $G \in (\Delta_{n,k}^1)_3$, 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 $G \in (\Delta_{n,k}^1)_3$ such that all edges are parallel to at least one other edge. If $h(G) = 1$, then $G = G^1(1, 1, \dots, 1)$ and so

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

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

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

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 at most k options for an ending coordinate.

Based on the above observations, we have

$$\begin{aligned}
 \left| \sum_{G \in (\Delta_{n,k}^1)_3} u_G^* \mathbb{E}[x_G] v_G \right| &\leq \sum_{G \in (\tilde{\Delta}_{n,k}^1)_3} |n^{h(G)} u_G^* \mathbb{E}[x_G] v_G| \\
 &\leq n^{\frac{k}{2} - \frac{1}{2}} \sum_{G \in (\tilde{\Delta}_{n,k}^1)_3} |u_G^* \mathbb{E}[x_G] v_G| \\
 &\leq n^{\frac{k}{2} - \frac{1}{2}} \sum_{G \in (\tilde{\Delta}_{n,k}^1)_3} M \\
 &\ll_{L,k} n^{\frac{k}{2} - \frac{1}{2}}.
 \end{aligned} \tag{55}$$

Combining the bounds: By (52), (54), and (55) we conclude that

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

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 (47) we have

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

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 (46) 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 $\{1, 2, \dots, m\}$. In this case we can write

$$\begin{aligned}
 \mathbb{E} \left[u^* \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v \right] &= n^{-k/2} \sum_{1 \leq a \leq m} (u^*)^{[a]} \mathbb{E}[(\mathcal{Y}_n^k)^{[a, a+r]}] v^{[a+r]} \\
 &= n^{-k/2} \sum_{1 \leq a \leq m} \sum_{G \in \Delta_{n,k}^a} u_G^* \mathbb{E}[x_G] v_G
 \end{aligned} \tag{56}$$

Again, define $h(G)$ as in (48) and define $(\Delta_{n,k}^a)_1$, $(\Delta_{n,k}^a)_2$, and $(\Delta_{n,k}^a)_3$ as in (49), (50), and (51), 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

$$\sum_{G \in (\Delta_{n,k}^1)_1} u_G^* \mathbb{E}[x_G] v_G = 0. \tag{57}$$

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

$$\left| \sum_{G \in (\Delta_{n,k}^1)_3} u_G^* \mathbb{E}[x_G] v_G \right| \ll_{L,k} n^{k/2-1}. \tag{58}$$

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 $(\Delta_{n,k}^1)_2$ is empty, so assume that k is even. Consider a graph $G \in (\Delta_{n,k}^1)_2$ and by Lemma 10.11, 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 10.14, we can conclude that in fact $G = G^1(1, 2, \dots, k/2, 1, 2, \dots, k/2, 1)$.

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 $G \in (\tilde{\Delta}_{n,k}^a)_2$, then $\mathbb{E}[x_G] = 0$. By Lemma 10.11, this gives

$$\sum_{G \in (\Delta_{n,k}^a)_2} u_G^* \mathbb{E}[x_G] v_G = 0. \tag{59}$$

Combining the bounds: By (57), (58), and (59) we can see that

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

While the bounds above were calculated for $a = 1$, the same arguments apply for any a by simply permuting the colors. Thus, from (56) we have

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

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

Combining the cases above completes the proof of Lemma 10.2.

Remark 10.15. 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.

Appendix A: Proof of Lemma 6.1

In this section, we present the proof of Lemma 6.1.

Proof of Lemma 6.1. Take $L_0 := \sqrt{8\mathbb{E}|\xi|^4}$. We begin by proving (ii). Observe that

$$\begin{aligned}
1 &= \mathbb{E}|\xi|^2 \\
&= \mathbb{E}|\operatorname{Re}(\xi)|^2 + \mathbb{E}|\operatorname{Im}(\xi)|^2 \\
&= \mathbb{E}[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)| \leq L/\sqrt{2}\}}] + \mathbb{E}[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)| > L/\sqrt{2}\}}] \\
&\quad + \mathbb{E}[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)| \leq L/\sqrt{2}\}}] + \mathbb{E}[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)| > L/\sqrt{2}\}}] \\
&= \operatorname{Var}(\operatorname{Re}(\tilde{\xi})) + \operatorname{Var}(\operatorname{Im}(\tilde{\xi})) \\
&\quad + |\mathbb{E}[\operatorname{Re}(\xi) \mathbf{1}_{\{|\operatorname{Re}(\xi)| \leq L/\sqrt{2}\}}]|^2 + \mathbb{E}[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)| > L/\sqrt{2}\}}] \\
&\quad + |\mathbb{E}[\operatorname{Im}(\xi) \mathbf{1}_{\{|\operatorname{Im}(\xi)| \leq L/\sqrt{2}\}}]|^2 + \mathbb{E}[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)| > L/\sqrt{2}\}}],
\end{aligned}$$

which implies

$$\begin{aligned}
1 - \operatorname{Var}(\tilde{\xi}) &= |\mathbb{E}[\operatorname{Re}(\xi) \mathbf{1}_{\{|\operatorname{Re}(\xi)| \leq L/\sqrt{2}\}}]|^2 + \mathbb{E}[|\operatorname{Re}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Re}(\xi)| > L/\sqrt{2}\}}] \\
&\quad + |\mathbb{E}[\operatorname{Im}(\xi) \mathbf{1}_{\{|\operatorname{Im}(\xi)| \leq L/\sqrt{2}\}}]|^2 + \mathbb{E}[|\operatorname{Im}(\xi)|^2 \mathbf{1}_{\{|\operatorname{Im}(\xi)| > L/\sqrt{2}\}}].
\end{aligned}$$

Thus, using the fact that $\operatorname{Re}(\xi)$ and $\operatorname{Im}(\xi)$ both have mean zero (so, for example, $\mathbb{E}[\operatorname{Re}(\xi) \mathbf{1}_{\{|\operatorname{Re}(\xi)| \leq L/\sqrt{2}\}}] = -\mathbb{E}[\operatorname{Re}(\xi) \mathbf{1}_{\{|\operatorname{Re}(\xi)| > L/\sqrt{2}\}}]$) 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 $\operatorname{Var}(\tilde{\xi}) \geq \frac{1}{2}$,

$$\begin{aligned}
|\hat{\xi}| &\leq \frac{|\operatorname{Re}(\xi)| \mathbf{1}_{\{|\operatorname{Re}(\xi)| \leq L/\sqrt{2}\}} + \mathbb{E}[|\operatorname{Re}(\xi)| \mathbf{1}_{\{|\operatorname{Re}(\xi)| \leq L/\sqrt{2}\}}]}{\sqrt{\operatorname{Var}(\tilde{\xi})}} \\
&\quad + \frac{|\operatorname{Im}(\xi)| \mathbf{1}_{\{|\operatorname{Im}(\xi)| \leq L/\sqrt{2}\}} + \mathbb{E}[|\operatorname{Im}(\xi)| \mathbf{1}_{\{|\operatorname{Im}(\xi)| \leq L/\sqrt{2}\}}]}{\sqrt{\operatorname{Var}(\tilde{\xi})}} \\
&\leq 4L
\end{aligned}$$

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 $\operatorname{Var}(\tilde{\xi}) \geq \frac{1}{2}$ and Jensen's inequality to obtain

$$\begin{aligned}
\mathbb{E}|\hat{\xi}|^4 &\ll \frac{1}{\operatorname{Var}(\tilde{\xi})^2} (\mathbb{E}[|\operatorname{Re}(\xi)|^4 \mathbf{1}_{\{|\operatorname{Re}(\xi)| \leq L/\sqrt{2}\}}] + \mathbb{E}[|\operatorname{Im}(\xi)|^4 \mathbf{1}_{\{|\operatorname{Im}(\xi)| \leq L/\sqrt{2}\}}]) \\
&\ll \mathbb{E}|\xi|^4,
\end{aligned}$$

as desired. \square

Appendix B: Proof of Theorem 7.1

This section is devoted to the proof of Theorem 7.1. We begin with Lemma B.1 below, which is based on [44, 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 be 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 $\text{Im}(z) > 0$. Assume

- (i) $\mu_n \rightarrow \mu$ as $n \rightarrow \infty$,
- (ii) there exists a constant $c > 0$ such that $\mu([0, c]) = 0$,
- (iii) $\sup_{E \in [0, c]} |m_n(E + \sqrt{-1}n^{-1/2}) - m(E + \sqrt{-1}n^{-1/2})| = o(n^{-1/2})$.

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

$$\text{Im} m_n(E + \sqrt{-1}n^{-1/2}) - \text{Im} m(E + \sqrt{-1}n^{-1/2}) = \int \frac{n^{-1/2} d(\mu_n - \mu)(x)}{(E-x)^2 + n^{-1}}.$$

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

$$\int_c^\infty \frac{d(\mu_n - \mu)(x)}{(E-x)^2 + n^{-1}} \rightarrow 0$$

uniformly for any $E \in [0, c/2]$ by the assumption that $\mu_n \rightarrow \mu$. Therefore, it must be the case that

$$\sup_{E \in [0, c/2]} \int_0^c \frac{d\mu_n(x)}{(E-x)^2 + n^{-1}} \rightarrow 0.$$

Take $n_0 \geq 1$ such that

$$\sup_{E \in [0, c/2]} \int_0^c \frac{d\mu_n(x)}{(E-x)^2 + n^{-1}} \leq 1/2 \tag{60}$$

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

$$\sup_{E \in [0, c/2]} \int_0^c \frac{d\mu_n(x)}{(E-x)^2 + n^{-1}} = \sup_{E \in [0, c/2]} \frac{1}{n} \sum_{j=1}^n \frac{1}{(E - \lambda_{n,j})^2 + n^{-1}}$$

$$\begin{aligned} &\geq \sup_{E \in [0, c/2]} \frac{1}{n} \frac{1}{(E - \lambda_{n,i})^2 + n^{-1}} \\ &\geq 1, \end{aligned}$$

a contradiction of (60). We conclude that $\mu_n([0, c/2]) = 0$ for all $n > n_0$. □

With Lemma B.1 in hand, we are now prepared to prove Theorem 7.1. The proof below is based on a slight modification to the arguments from [44,45]. As such, in some places we will omit technical computations and only provide appropriate references and necessary changes to results from [44,45].

Fix $\delta > 0$. It suffices to prove that

$$\inf_{1+\delta \leq |z| \leq 6} s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \geq c \tag{61}$$

and

$$\inf_{|z| > 6} s_{mn} \left(\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \geq c' \tag{62}$$

with overwhelming probability for some constants $c, c' > 0$ depending only on δ .

The second bound (62) 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, [57, Theorem 1.4]) gives

$$\|\mathcal{Y}_n\| \leq 3\sqrt{n}$$

with overwhelming probability. The bound in (62) then follows by applying Lemma D.3.

We now turn to the bound in (61). 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 [45, Theorem 2.6], for all $|z| \geq 1 + \delta$, there exists a probability measure μ_z supported on $[0, \infty)$ such that $\mu_{n,z} \rightarrow \mu_z$ with overwhelming probability. Moreover, from [9, 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 [44, 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 (20). The proof of Theorem 7.1 is complete.

Appendix C: Proof of Lemma 10.14

This section is devoted to the proof of Lemma 10.14.

Proof of Lemma 10.14. 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 high 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.

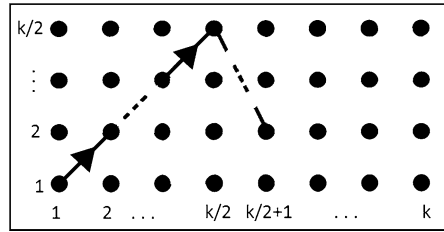


Fig. 5. This is an example of a path graph that has type I edges drawn until a height of $k/2$ is reached, then the first type II edge is drawn.

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 be 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 height 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 height coordinate $k/2$ is reached. Namely, we have vertices $(1, 1), \dots, (k/2, k/2)$.

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 5 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 at 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 $V = \{(1, 1), (2, 2), \dots, (k/2, k/2), (k/2 + 1, 1), (k/2 + 2, 2), \dots, (k, k/2), (k + 1, 1)\}$. The corresponding canonical m -colored k -path graph would be $G^1(1, 2, \dots, k/2, 1, 2, \dots, k/2, 1)$, and the proof is complete. \square

Appendix D: Useful inequalities

Lemma D.1 (Lemma A.1 from [11]). For $X = (x_1, x_2, \dots, x_N)^T$ iid standardized complex entries, for B an $N \times N$ Hermitian nonnegative definite matrix, and for any $p \geq 1$, we have $\mathbb{E}|X^*BX|^p \leq K_p((\text{tr } B)^p + \mathbb{E}|x_1|^{2p} \text{tr } B^p)$, where $K_p > 0$ depends only on p .

Lemma D.2 (Follows from Lemma 10 in [50]). 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 [46]). Let A be a square matrix that satisfies $\|A\| \leq \mathcal{K}$. Then $\|(A - zI)^{-1}\| \leq \frac{1}{\varepsilon}$ for all $z \in \mathbb{C}$ with $|z| \geq \mathcal{K} + \varepsilon$.

References

- [1] G. Akemann and Z. Burda. Universal microscopic correlation functions for products of independent Ginibre matrices. *J. Phys. A: Math. Theor.* **45** (2012). MR2993423 <https://doi.org/10.1088/1751-8113/45/46/465201>
- [2] G. Akemann, Z. Burda and M. Kieburg. Universal distribution of Lyapunov exponents for products of Ginibre matrices. *J. Phys. A: Math. Theor.* **47** (39) (2014). MR3262164 <https://doi.org/10.1088/1751-8113/47/39/395202>
- [3] G. Akemann, J. R. Ipsen and M. Kieburg. Products of rectangular random matrices: Singular values and progressive scattering. *Phys. Rev. E* **88** 052118 (2013). <https://doi.org/10.1103/PhysRevE.88.052118>
- [4] G. Akemann, J. R. Ipsen and E. Strahov. Permanent processes from products of complex and quaternionic induced Ginibre ensembles. *Random Matrices Theory Appl.* **3** (4) (2014) 1450014. MR3279619 <https://doi.org/10.1142/S2010326314500142>
- [5] G. Akemann, M. Kieburg and L. Wei. Singular value correlation functions for products of Wishart random matrices. *J. Phys. A: Math. Theor.* **46** (2013). MR3081917 <https://doi.org/10.1088/1751-8113/46/27/275205>
- [6] G. Akemann and E. Strahov. Hole probabilities and overcrowding estimates for products of complex Gaussian matrices. *J. Stat. Phys.* **151** (6) (2013) 987–1003. MR3063493 <https://doi.org/10.1007/s10955-013-0750-8>
- [7] G. Anderson. Convergence of the largest singular value of a polynomial in independent Wigner matrices. *Ann. Probab.* **41** (3B) (2013) 2103–2181. MR3098069 <https://doi.org/10.1214/11-AOP739>
- [8] P. Arbenz, W. Gander and G. Golub. Restricted rank modification of the symmetric eigenvalue problem: Theoretical considerations. *Linear Algebra Appl.* **104** (1988) 75–95. MR0944010 [https://doi.org/10.1016/0024-3795\(88\)90309-6](https://doi.org/10.1016/0024-3795(88)90309-6)
- [9] Z. D. Bai. Circular law. *Ann. Probab.* **25** (1997) 494–529. MR1428519 <https://doi.org/10.1214/aop/1024404298>
- [10] Z. D. Bai and G. M. Pan. Limiting behavior of eigenvectors of large Wigner matrices. *J. Stat. Phys.* **146** (2012) 519–549. MR2880031 <https://doi.org/10.1007/s10955-011-0407-4>
- [11] Z. D. Bai and J. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** (1) (1998) 316–345. MR1617051 <https://doi.org/10.1214/aop/1022855421>
- [12] Z. D. Bai and J. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices. Mathematics Monograph Series 2.* Science Press, Beijing, 2006. MR2567175 <https://doi.org/10.1007/978-1-4419-0661-8>
- [13] J. Baik, G. Ben Arous and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** (5) (2005) 1643–1697. MR2165575 <https://doi.org/10.1214/009117905000000233>
- [14] S. Belinschi, H. Bercovivi and M. Capitaine. On the outlying eigenvalues of a polynomial in large independent random matrices. Available at [arXiv:1703.08102](https://arxiv.org/abs/1703.08102) [math.PR].
- [15] S. T. Belinschi and M. Capitaine. Spectral properties of polynomials in independent Wigner and deterministic matrices. *J. Funct. Anal.* **273** (12) (2017) 3901–3963. MR3711884 <https://doi.org/10.1016/j.jfa.2017.07.010>
- [16] F. Benaych-Georges, G. Cébron and J. Rochet. Fluctuation of matrix entries and application to outliers of elliptic matrices. *Canad. J. Math.* **70** (1) (2018) 3–25. MR3744883 <https://doi.org/10.4153/CJM-2017-024-8>
- [17] F. Benaych-Georges, A. Guionnet and M. Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** (2011) 1621–1662. MR2835249 <https://doi.org/10.1214/EJP.v16-929>
- [18] F. Benaych-Georges, A. Guionnet and M. Maida. Large deviations of the extreme eigenvalues of random deformations of matrices. *Probab. Theory Related Fields* **154** (3–4) (2012) 703–751. MR3000560 <https://doi.org/10.1007/s00440-011-0382-3>
- [19] F. Benaych-Georges and R. N. Rao. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** (1) (2011) 494–521. MR2782201 <https://doi.org/10.1016/j.aim.2011.02.007>
- [20] F. Benaych-Georges and R. N. Rao. The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** (2012) 120–135. MR2944410 <https://doi.org/10.1016/j.jmva.2012.04.019>
- [21] F. Benaych-Georges and J. Rochet. Outliers in the single ring theorem. *Probab. Theory Related Fields* **165** (2016) 313–363. MR3500273 <https://doi.org/10.1007/s00440-015-0632-x>
- [22] R. Bhatia. *Matrix Analysis. Graduate Texts in Mathematics.* Springer-Verlag, New York, 1997. MR1477662 <https://doi.org/10.1007/978-1-4612-0653-8>
- [23] C. Bordenave. On the spectrum of sum and product of non-Hermitian random matrices. *Electron. Commun. Probab.* **16** (2011) 104–113. MR2772389 <https://doi.org/10.1214/ECP.v16-1606>
- [24] C. Bordenave and M. Capitaine. Outlier eigenvalues for deformed i.i.d. random matrices. *Comm. Pure Appl. Math.* **69** (11) (2016) 2131–2194. MR3552011 <https://doi.org/10.1002/cpa.21629>
- [25] C. Bordenave, P. Caputo, D. Chafaï and K. Tikhomirov. On the spectral radius of a random matrix: an upper bound without fourth moment. *Ann. Probab.* **46** (4) (2018) 2268–2286. MR3813992 <https://doi.org/10.1214/17-AOP1228>
- [26] C. Bordenave and D. Chafaï. Around the circular law. *Probab. Surv.* **9** (2012) 1–89. MR2908617 <https://doi.org/10.1214/11-PS183>
- [27] Z. Burda. Free products of large random matrices – a short review of recent developments. *J. Phys., Conf. Ser.* **473** (2013) 012002. <https://doi.org/10.1088/1742-6596/473/1/012002>
- [28] Z. Burda, R. A. Janik and B. Waclaw. Spectrum of the product of independent random Gaussian matrices. *Phys. Rev. E* **81** (2010). MR2736204 <https://doi.org/10.1103/PhysRevE.81.041132>
- [29] Z. Burda, A. Jaroš, G. Livan, M. A. Nowak and A. Swiech. Eigenvalues and singular values of products of rectangular Gaussian random matrices. *Phys. Rev. E* **82** (2010). MR2787469 <https://doi.org/10.1103/PhysRevE.82.061114>
- [30] D. L. Burkholder. Distribution function inequalities for martingales. *Ann. Probab.* **1** (1973) 19–42. MR0365692 <https://doi.org/10.1214/aop/1176997023>

- [31] M. Capitaine, C. Donati-Martin and 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. MR2489158 <https://doi.org/10.1214/08-AOP394>
- [32] M. Capitaine, C. Donati-Martin and D. Féral. Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** (1) (2012) 107–133. MR2919200 <https://doi.org/10.1214/10-AIHP410>
- [33] M. Capitaine, C. Donati-Martin, D. Féral and M. Février. Free convolution with a semi-circular distribution and eigenvalues of spiked deformations of Wigner matrices. *Electron. J. Probab.* **16** (64) (2011) 1750–1792. MR2835253 <https://doi.org/10.1214/EJP.v16-934>
- [34] N. Coston, S. O'Rourke and P. Wood. Outliers in the spectrum for products of independent random matrices. Available at [arXiv:1711.07420](https://arxiv.org/abs/1711.07420) [math.PR].
- [35] D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** (2007) 185–228. MR2291807 <https://doi.org/10.1007/s00220-007-0209-3>
- [36] P. J. Forrester. Lyapunov exponents for products of complex Gaussian random matrices. *J. Stat. Phys.* **151** (2013) 796–808. MR3055376 <https://doi.org/10.1007/s10955-013-0735-7>
- [37] P. J. Forrester. Probability of all eigenvalues real for products of standard Gaussian matrices. *J. Phys. A* **47** (2014) 065202. MR3159513 <https://doi.org/10.1088/1751-8113/47/6/065202>
- [38] F. Götze and T. Tikhomirov. On the asymptotic spectrum of products of independent random matrices. Available at [arXiv:1012.2710](https://arxiv.org/abs/1012.2710).
- [39] U. Haagerup and S. Thorbjørnsen. A new application of random matrices: $\text{Ext}(C_{\text{red}}^*(F_2))$ is not a group. *Ann. of Math.* **162** (2005) 711–775. MR2183281 <https://doi.org/10.4007/annals.2005.162.711>
- [40] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, 1991. MR1084815
- [41] A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **LXVI** (2013) 1663–1749. MR3103909 <https://doi.org/10.1002/cpa.21450>
- [42] A. Knowles and J. Yin. The outliers of a deformed Wigner matrix. *Ann. Probab.* **42** (5) (2014) 1980–2031. MR3262497 <https://doi.org/10.1214/13-AOP855>
- [43] A. B. J. Kuijlaars and L. Zhang. Singular values of products of Ginibre random matrices, multiple orthogonal polynomials and hard edge scaling limits. *Comm. Math. Phys.* **332** (2) (2014) 759–781. MR3257662 <https://doi.org/10.1007/s00220-014-2064-3>
- [44] Y. Nemish. No outliers in the spectrum of the product of independent non-Hermitian random matrices with independent entries. *J. Theoret. Probab.* (2016) 1–43. MR3769819
- [45] Y. Nemish. Local law for the product of independent non-Hermitian random matrices with independent entries. *Electron. J. Probab.* **22** (2017), paper no. 22, 35. MR3622892 <https://doi.org/10.1214/17-EJP38>
- [46] S. O'Rourke and D. Renfrew. Low rank perturbations of large elliptic random matrices. *Electron. J. Probab.* **19** (43) (2014) 1–65. MR3210544 <https://doi.org/10.1214/EJP.v19-3057>
- [47] S. O'Rourke, D. Renfrew, A. Soshnikov and V. Vu. Products of independent elliptic random matrices. *J. Stat. Phys.* **160** (1) (2015) 89–119. MR3357969 <https://doi.org/10.1007/s10955-015-1246-5>
- [48] S. O'Rourke and A. Soshnikov. Products of independent non-Hermitian random matrices. *Electron. J. Probab.* **16**, Art. **81** (2011) 2219–2245. MR2861673 <https://doi.org/10.1214/EJP.v16-954>
- [49] S. O'Rourke and P. Wood. Spectra of nearly Hermitian random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** (3) (2017) 1241–1279. MR3689967 <https://doi.org/10.1214/16-AIHP754>
- [50] P. Pagacz and M. Wojtylak. Random perturbations of matrix polynomials. Available at [arXiv:1703.01858](https://arxiv.org/abs/1703.01858) [math.PR].
- [51] S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** (1) (2006) 127–173. MR2221787 <https://doi.org/10.1007/s00440-005-0466-z>
- [52] A. Pizzo, D. Renfrew and A. Soshnikov. On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** (1) (2013) 64–94. MR3060148 <https://doi.org/10.1214/11-AIHP459>
- [53] A. B. Rajagopalan. Outlier eigenvalue fluctuations of perturbed iid matrices. Available at [arXiv:1507.01441](https://arxiv.org/abs/1507.01441) [math.PR]. MR3389123
- [54] D. Renfrew and A. Soshnikov. On finite rank deformations of Wigner matrices II: Delocalized perturbations. *Random Matrices Theory Appl.* **02** (2013) 1250015. MR3039820 <https://doi.org/10.1142/S2010326312500153>
- [55] J. Rochet. Complex outliers of Hermitian random matrices. *J. Theoret. Probab.* **30** (4) (2017) 1624–1654. MR3736185 <https://doi.org/10.1007/s10959-016-0686-4>
- [56] E. Strahov. Differential equations for singular values of products of Ginibre random matrices. *J. Phys. A: Math. Theor.* **47** (2014) 325203. MR3240458 <https://doi.org/10.1088/1751-8113/47/32/325203>
- [57] T. Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields* **155** (2013) 231–263. MR3010398 <https://doi.org/10.1007/s00440-011-0397-9>
- [58] T. Tao and V. Vu. Random matrices: Universality of ESDs and the circular law. *Ann. Probab.* **38** (5) (2010) 2023–2065. MR2722794 <https://doi.org/10.1214/10-AOP534>