# GAUSSIAN FLUCTUATIONS FOR LINEAR EIGENVALUE STATISTICS OF PRODUCTS OF INDEPENDENT IID RANDOM MATRICES

NATALIE COSTON AND SEAN O'ROURKE

Abstract. Consider the product $X = X_1 \cdots X_m$ of $m$ independent $n \times n$ iid random matrices. When $m$ is fixed and the dimension $n$ tends to infinity, we prove Gaussian limits for the centered linear spectral statistics of $X$ for analytic test functions. We show that the limiting variance is universal in the sense that it does not depend on $m$ (the number of factor matrices) or on the distribution of the entries of the matrices. The main result generalizes and improves upon previous limit statements for the linear spectral statistics of a single iid matrix by Rider and Silverstein as well as Renfrew and the second author.

Contents

## 1. Introduction and Background Material

This paper is concerned with fluctuations of linear eigenvalue statistics for products of random matrices with independent and identically distributed (iid) entries.

**Definition 1.1** (iid random matrix)**.** Let $\xi$ be a complex-valued random variable. We say $X_n$ is an $n \times n$ *iid random matrix* with atom variable $\xi$ if $X_n$ is an $n \times n$ matrix whose entries are iid copies of $\xi$.

Recall that the *eigenvalues* of an $n \times n$ matrix $M_n$ are the roots in $\mathbb{C}$ of the characteristic polynomial $\det(zI - M_n)$, where $I$ is the identity matrix. We let $\lambda_1(M_n), \ldots, \lambda_n(M_n)$ denote the eigenvalues of $M_n$ counted with (algebraic) multiplicity. The *empirical spectral measure* $\mu_{M_n}$ 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 $\mu_{M_n}$ is also random. In this case, we say $\mu_{M_n}$ converges *weakly in probability* (resp. *weakly almost surely*) to another Borel probability measure $\mu$ on the complex plane $\mathbb{C}$ if, for every bounded and continuous function $f : \mathbb{C} \to \mathbb{C}$,

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

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

For iid random matrices whose atom variable has finite variance, the limiting behavior of the empirical spectral measure is described by the circular law. Recall that the Hilbert-Schmidt norm $\|M\|_2$ of a matrix $M$ is defined by the formula

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

**Theorem 1.2** (Circular law; Corollary 1.12 from [65])**.** *Let $\xi$ 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 $\xi$, and let $A_n$ be a deterministic $n \times n$ matrix. If $\operatorname{rank}(A_n) = o(n)$ and $\sup_{n \geq 1} \frac{1}{n} \|A_n\|_2^2 < \infty$, then the empirical measure $\mu_{\frac{1}{\sqrt{n}} X_n + A_n}$ of $\frac{1}{\sqrt{n}} X_n + A_n$ converges weakly almost surely to the uniform probability measure on the unit disk centered at the origin in the complex plane as $n \to \infty$.*

This result appears as [65, Corollary 1.12], but is the culmination of work by many authors including [10, 13, 27, 30, 31, 32, 35, 46, 47, 55, 63, 64, 65]. We refer the interested reader to the survey [18] for further details.

### 1.1. Products of Independent iid Matrices.

The result presented in this paper focuses not on a single iid random matrix, but instead on the product of several independent iid matrices. The analogue of the circular law (Theorem 1.2) in this case has been derived by several authors [36, 53, 54] under various assumptions on the factor matrices; the version presented below is from [53]. Similar results are stated in [33].

**Theorem 1.3** (Theorem 2.4 from [53])**.** *Let $m \geq 1$ be an integer and $\tau > 0$. Let $\xi_1, \ldots, \xi_m$ be real-valued random variables with mean zero, and assume, for each $1 \leq k \leq m$, $\xi_k$ has nonzero variance $\sigma_k^2$ 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 $\xi_k$, and*

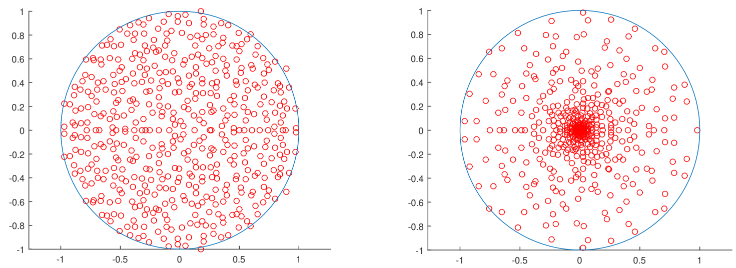


FIGURE 1. The leftmost figure shows the eigenvalues, denoted by the small circles, of a single $500 \times 500$ iid random matrix $\frac{1}{\sqrt{500}} X_{500}$ with Gaussian entries. The rightmost figure shows the eigenvalues, denoted by small circles, of a product of four independent $500 \times 500$ random matrices, each scaled by $\frac{1}{\sqrt{500}}$, where the entries in each random matrix are independent iid Gaussian random variables.

*let $A_{n,k}$ be a deterministic $n \times n$ matrix. Assume $X_{n,1}, \dots, X_{n,m}$ are independent. If*

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

*for some $\varepsilon > 0$, then the empirical spectral measure $\mu_{P_n}$ of the product*

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

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

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

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

*Remark* 1.4. When $\sigma = 1$, the density in (2) is easily related to the circular law (Theorem 1.2). Indeed, in this case, $\varphi_m$ is the density of $\psi^m$, where $\psi$ is a complex-valued random variable uniformly distributed on the unit disk centered at the origin in the complex plane.

Theorem 1.3 can be viewed as a generalization of the circular law (Theorem 1.2). Indeed, $\mu_1$ is simply the uniform measure on a disk of radius $\sigma$ centered at the origin. We emphasis here that the limiting empirical spectral measure $\mu_m$ depends on $m$ and is different for each integer $m$. Figure 1 provides a numerical illustration of Theorem 1.3.

The random matrix theory literature contains many papers concerning products of independent matrices with Gaussian entries; we refer the reader to [1, 2, 3, 4, 5, 6, 7, 19, 21, 22, 20, 28, 29, 39, 40, 44, 61] and references therein. Some other models of products and sums of random matrices have also been considered in [17].

Recently in [49], Nemish proved a local law version of Theorem 1.3 up to the optimal scale, under the assumption that the entries of each iid matrix have subexponential decay. Nemish's local law has also been extended by Götze, Naumov, and Tikhomirov [34] to include the case where the entries do not have subexponential decay but instead have finite $4+\tau$ moment for some fixed $\tau > 0$. The universality of the local correlation functions for the eigenvalues of such product matrices was recently established in [43].

1.2. **Fluctuations of Linear Eigenvalue Statistics.** The linear eigenvalue statistics of a random matrix describe the fluctuations of the spectrum about its limiting distribution. The uncentered linear eigenvalue statistics for an $n \times n$ matrix $M$ and sufficiently smooth test function $f$ (whose smoothness depends on the matrix ensemble under consideration) is defined by

$$\operatorname{tr} f(M) := \sum_{i=1}^{n} f(\lambda_i(M)) \tag{3}$$

where $\lambda_1(M), \ldots, \lambda_n(M)$ denote the eigenvalues of $M$.

In the classical central limit theorem, sums of $n$ iid random variables have variance on the order of $\sqrt{n}$. In contrast, the variance of linear spectral statistics for many ensembles of random matrices is often on the order of a constant. There are many results regarding the fluctuations of linear eigenvalue statistics for various ensembles of random matrices (and under various assumptions on the test functions $f$). Because the subject is so well studied, we do not give a full treatment here. We refer the reader to [9, 11, 25, 26, 41, 42, 43, 45, 50, 51, 56, 57, 58, 59, 60] and the references therein for further details. In the discussion below, we will only focus on linear statistics for iid random matrices and their products.

Rider and Silverstein, in the seminal paper [56], established Gaussian fluctuations for the linear eigenvalue statistics of iid random matrices with analytic test functions.

**Theorem 1.5** (Theorem 1.1 from [56])**.** *Let $\xi$ be a complex-valued random variable which satisfies the following conditions.*

(i) $\mathbb{E}[\xi] = 0$, *and* $\mathbb{E}[|\xi|^2] = 1$,
(ii) $\mathbb{E}[\xi^2] = 0$,
(iii) $\mathbb{E}[|\xi|^k] \leq k^{\alpha k}$ *for every* $k > 2$ *and some* $\alpha > 0$,
(iv) $\operatorname{Re}(\xi)$ *and* $\operatorname{Im}(\xi)$ *possesses a bounded joint density.*

*For each $n \geq 1$, let $X_n$ be an $n \times n$ iid random matrix with atom variable $\xi$. Consider test functions $f_1, f_2, \ldots, f_k$ analytic in a neighborhood of the disk $\{z \in \mathbb{C} : |z| \leq 4\}$ and otherwise bounded. Then as $n \to \infty$, the vector*

$$\left( \operatorname{tr} f_j \left( \frac{1}{\sqrt{n}} X_n \right) - n f_j(0) \right)_{j=1}^{k}$$

*converges in distribution to a mean-zero multivariate Gaussian vector $(F(f_1), F(f_2), \ldots, F(f_k))$ with covariances*

$$\mathbb{E}\left[ F(f_l)\overline{F(f_m)} \right] = \frac{1}{\pi} \int_{\mathbb{U}} \frac{d}{dz} f_l(z) \overline{\frac{d}{dz} f_m(z)} d^2 z,$$

*in which $\mathbb{U}$ is the unit disk centered at the origin and $d^2 z = d\operatorname{Re}(z) d\operatorname{Im}(z)$.*

Theorem 1.5 was later generalized and extended by Renfrew and the second author in [51]. The results in [51] remove several technical assumptions present in Theorem 1.5. Specifically, for the results in [51] to hold, conditions *(ii)* and *(iv)* from Theorem 1.5 are no longer required, and condition *(iii)* is replaced by the finiteness of $\mathbb{E}|\xi|^{6+\tau}$. In addition, the functions $f_1, \ldots, f_k$ are only required to be analytic in a neighborhood of the disk $\{z \in \mathbb{C} : |z| \leq 1\}$. More generally, the results in [51] also hold for an ensemble of elliptic random matrices which include iid random matrices as a special case.

For products of independent iid random matrices, much less in known. To the best of the authors' knowledge, the only result for fluctuations of linear eigenvalue statistics for products of iid random matrices is [43, Theorem 3], which requires the factor matrices to match moments with the complex Ginibre ensemble; we state this result from [43] below.

**Theorem 1.6** (Theorem 3 from [43]). *Let $f : \mathbb{C} \to \mathbb{R}$ be a test function with at least two continuous derivatives, supported in the region $\{z \in \mathbb{C} \,:\, \tau_0 < |z| < 1 - \tau_0\}$ for some fixed $\tau_0 > 0$. Let $m \geq 1$ be an integer and let*

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

*be a matrix product such that each $X_{n,i}$ is an $n \times n$ iid random matrix (which are all jointly independent) with an atom variable $\xi_i$ which satisfies the following:*

- *$\xi_i$ has mean zero and unit variance,*
- *$\xi_i$ has independent real and imaginary parts,*
- *$\xi_i$ satisfies the subgaussian decay condition that there exist constants $C, c > 0$ (independent of $n$) such that for each $t > 0$, $\mathbb{P}(|\xi_i| > t) \leq Ce^{-ct^2}$, and*
- *$\xi_i$ matches moments with a standard complex Gaussian random variable to four moments: for all $a, b \geq 0$ such that $a + b \leq 4$, $\mathbb{E}\left[\mathrm{Re}(\xi_i)^a \,\mathrm{Im}(\xi_i)^b\right] = \mathbb{E}\left[\mathrm{Re}(\zeta)^a \,\mathrm{Im}(\zeta)^b\right]$ where $\zeta$ is a standard complex Gaussian random variable.*

*Then the centered linear statistic*

$$\mathrm{tr}\, f(P_n) - \mathbb{E}\left[\mathrm{tr}\, f(P_n)\right]$$

*converges in distribution as $n \to \infty$ to the mean-zero Gaussian distribution with limiting variance*

$$\frac{1}{4\pi} \int_{\mathbb{U}} | \bigtriangledown f(z)|^2 d^2 z$$

*where $\mathbb{U}$ is the unit disk centered at the origin.*

**Acknowledgments.** The paper is based on a chapter from N. Coston's doctoral thesis, and she would like to thank her thesis committee for their feedback and support. The authors would also like to thank Philip Wood for providing useful feedback on an earlier draft of the manuscript. S. O'Rourke has been supported in part by NSF grants ECCS-1610003 and DMS-1810500.

## 2. Main Results

### 2.1. Fluctuations of Linear Eigenvalue Statistics for Product Matrices.

The main result of this paper is the analogue of Theorem 1.5 (and its generalization in [51]) for the product of independent iid random matrices.

For $1 \leq k \leq m$, let $\xi_k$ be a random variable which satisfies the following conditions.

**Assumption 2.1.** The real-valued random variables $\xi_1, \ldots, \xi_m$ (not necessarily defined on the same probability space) are said to satisfy Assumption 2.1 if, for each $1 \leq k \leq m$,

- $\xi_k$ has mean zero,
- $\xi_k$ has nonzero variance $\sigma_k^2$, and
- there exists $\tau > 0$ such that $\mathbb{E}|\xi_k|^{4+\tau} < \infty$.

The following theorem is the main result of the paper.

**Theorem 2.2** (Fluctuations of linear statistics for products of iid random matrices)**.** *Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \ldots, \xi_m$ are real-valued random variables which satisfy Assumption 2.1. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Define the products*

$$P_n := n^{-m/2} X_{n,1} \cdots X_{n,m} \tag{4}$$

*and*

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

*Let $\delta > 0$, $s \geq 1$ be a fixed integer, and $f_1, f_2, \ldots, f_s$ be test functions analytic in some neighborhood containing the disk $D_\delta := \{z \in \mathbb{C} : |z| \leq 1 + \delta\}$ and bounded otherwise. Then there exist deterministic sequences $\mathcal{A}_n(f_1), \ldots, \mathcal{A}_n(f_k)$ (with $\mathcal{A}_n(f_i)$ depending only $n, f_i$, and the distribution of $\xi_1, \ldots \xi_m$) such that the random vector*

$$\left(\operatorname{tr} f_i(P_n/\sigma) - \mathcal{A}_n(f_i)\right)_{i=1}^s \tag{5}$$

*converges in distribution to a mean-zero multivariate Gaussian random vector*

$$(F(f_1), \ldots, F(f_s))$$

*with variance and covariance terms defined by*

$$\mathbb{E}\left[F(f_i)F(f_j)\right] = -\frac{1}{4\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} f_i(z) f_j(w)(zw-1)^{-2} dz dw \tag{6}$$

*and*

$$\mathbb{E}\left[F(f_i)\overline{F(f_j)}\right] = \frac{1}{4\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} f_i(z) \overline{f_j(w)}(z\bar{w}-1)^{-2} dz d\bar{w} \tag{7}$$

*where $\mathcal{C}$ is the contour around the boundary of the disk $D_\delta$.*

A few remarks concerning Theorem 2.2 are in order. Heuristically, we may think of $\mathcal{A}_n(f_i)$ as the centering term, and subtracting this quantity is similar to subtracting the expectation (as was done in Theorem 1.6) or $nf_i(0)$ (as was done in Theorem 1.5). For technical reasons, defining this term requires some notation and concepts which will be introduced in the forthcoming sections; see (28) for details.

While the limiting empirical spectral measure for the product of $m$ iid matrices does depend on $m$ (Theorem 1.3), the variance and covariance terms, (6) and (7), for the fluctuations of the linear eigenvalue statistics do not depend on $m$. In other words, the variance and covariance terms are the same as in the case of a single iid matrix ($m = 1$); indeed, (6) and (7) match the analogous terms appearing in [51] for a single iid matrix. In this sense, the fluctuations of the linear statistics appear to be more universal than the global distribution of the eigenvalues. In certain cases, these covariance terms can be rewritten in terms of an iterated integral over the real and imaginary parts of $z$ as was done in [56, Theorem 1.1].

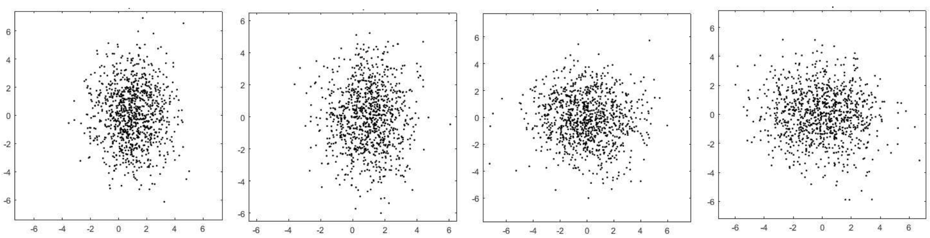


FIGURE 2. This figure provides an illustration of Theorem 2.2. All plots show 1000 observations of the linear statistic in (3). The leftmost plot shows linear statistics computed with a product of three Bernoulli$(-1,1)$ $300 \times 300$ matrices scaled by $300^{-3/2}$, and with $f(z) = z^2 + 2\sqrt{-1}z$. The second plot from the left shows linear statistics for a product of ten Bernoulli$(-1,1)$ $300 \times 300$ matrices scaled by $300^{-10/2}$, and with $f(z) = z^2 + 2\sqrt{-1}z$. The second plot from the right shows linear statistics for a product of three mean-zero Gaussian $300 \times 300$ matrices scaled by $300^{-3/2}$, and $g(z) = \sqrt{-1}z^3 + z^2$. Finally, the rightmost plot shows linear statistics for a product of ten mean-zero Gaussian $300 \times 300$ matrices scaled by $300^{-10/2}$, and with $g(z) = \sqrt{-1}z^3 + z^2$.

We also remark that Theorem 2.2 can be extended to the case where each atom variable is complex-valued under the assumption that the real and imaginary parts of each atom variable are independent. In this case, the covariance terms in Theorem 2.2 would change slightly. The changes required for the complex case can easily be found by inspecting the proof; we refer the reader to Remarks 4.2, 4.4, 6.2, and 6.16 for the details.

While Theorem 1.6 holds for a more general class of test functions than Theorem 2.2, it also requires subgaussian decay and a moment matching condition on the entries. In particular, Theorem 1.6 does not apply to iid matrices with real-valued entries. Thus, Theorem 2.2, while restricted to analytic functions, does apply to a much larger class of iid random matrices (such as the real Ginibre ensemble and Bernoulli matrices, whose entries take the values $\pm 1$ with equal probability).

Even for the case of a single iid matrix ($m = 1$), Theorem 2.2 improves upon the existing results in the literature. Compared to the main results of [51] (which were already an improvement over Theorem 1.5), Theorem 2.2 applies to a more general class of iid matrices while still applying to the same class of test functions.

Figure 2 provides a numerical illustration of Theorem 2.2 for various test functions and values of $m$.

Since the covariance formulas, (6) and (7), do not depend on $m$, it is an interesting open question to consider the case when the number of product matrices is allowed to depend on $n$, e.g., when $m$ grows (slowly) with $n$. The proof given below requires $m$ to be fixed, and there are several key bounds which depend on $m$. If this dependence could be tracked carefully, it may be possible that parts of the proof could be adapted to the case where $m$ grows with $n$.

2.2. **Outline and Overview.** The remainder of the paper is devoted to the proof of Theorem 2.2. Roughly speaking, the proof follows the main ideas from [56, 51] for studying the linear eigenvalue statistics of a single iid matrix. However, the product structure of the matrix $P_n$ introduces substantial new difficulties. For instance, when $m > 1$, the entries of $P_n$ are no longer independent. To get around this issue, we use a linearization technique. That is, instead of studying the product matrix $P_n$, we introduce the linearized matrix $\mathcal{Y}$, which is a $mn \times mn$ block matrix defined as follows:

$$\mathcal{Y} := \frac{1}{\sqrt{n}} \begin{bmatrix} 0 & X_{n,1} & 0 & & \cdots & 0 \\ 0 & 0 & X_{n,2} & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & \cdots & & 0 & X_{n,m-1} \\ X_{n,m} & 0 & \cdots & & 0 & 0 \end{bmatrix},$$

where any block not specified above is assumed to be zero. As has been observed previously [19, 23, 53, 54], the eigenvalues of $\mathcal{Y}^m$ are the same as the eigenvalues of $P_n$, up to some multiplicity factor. Thus, the problem can be reduced to studying the linear eigenvalue statistics of $\mathcal{Y}$.

The main challenge in studying the linear statistics of non-Hermitian random matrices is often computing the limiting variance. For example, in [56, 51], for a single iid matrix the variance is computed by deriving a recursive equation, which then must be solved in the limit as $n$ tends to infinity to obtain the limiting variance. In the case of analyzing the limiting variance for the linear statistics of $\mathcal{Y}$, the block structure of $\mathcal{Y}$ itself introduces new difficulties. In this case, it does not seem possible to derive a single recursive equation as was done in [56, 51] due to the prescience of so many deterministic zero blocks in $\mathcal{Y}$. To get around this issue, we instead derive a system of $m$ recursive equations and then solve the system of equations simultaneously. The derivation and solution of this system of recursive equations for the variance is the main technical advance of the present article and occupies the bulk of the proof. Interestingly, the limiting variance we derive for the linear statistics of $\mathcal{Y}$ *does* depend on $m$ and appears to have a form not encountered before in random matrix theory. When this limiting variance is translated back to the variance for the product matrix $P_n$, the dependence on $m$ vanishes.

The paper is organized as follows. In Section 3, we present some preliminary results, tools, and notation that will be used throughout the paper. In Section 4, we make some preliminary reductions and reduce the problem to the study of the linear statistics of the linearized matrix $\mathcal{Y}$. Section 5 begins the proof of the main result, which by Cauchy's integral formula, involves studying a sequence of stochastic processes involving the trace of the resolvent matrix. Section 6 proves the finite dimensional convergence of this sequence of stochastic processes, and Section 7 shows that this sequence of stochastic processes is tight, concluding the proof. These last two sections are based on [56, 51]. We warn the reader that the last two sections appear to inherit many of the technical challenges present in [56, 51] along with several additional challenges (based on the block structure of the linearized matrix $\mathcal{Y}$, as discussed above). As such, the material presented in these two sections is rather technical and some of the calculations are tedious. Some appendices follow with auxiliary results.

## 3. Preliminary Tools and Notation

This section is devoted to introducing some additional concepts and notation required for the proofs of our main results.

### 3.1. Notation.

We use asymptotic notation (such as $O, o, \Omega$) under the assumption that $n \to \infty$. In particular, $X = O(Y)$, $Y = \Omega(X)$, $X \ll Y$, and $Y \gg X$ denote the estimate $|X| \leq CY$, for some constant $C > 0$ independent of $n$ and for all $n \geq C$. If we need the constant $C$ to depend on a parameter $k$, 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 $n \to \infty$. Specifically, $o(1)$ denotes a term which tends to zero as $n \to \infty$. If we need the sequence $c(n)$ to depend on a parameter $k$, 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 $E$. $E^c$ denotes the complement of the event $E$. For $\delta > 0$, $D_\delta$ denotes the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$.

For a matrix $M$, we let $\|M\|$ denote the spectral norm of $M$. $\|M\|_2$ denotes the Hilbert-Schmidt norm of $M$ (defined in (1)). We let $I_n$ denote the $n \times n$ identity matrix. Often we will just write $I$ for the identity matrix when the size can be deduced from context.

The singular values of a matrix $M$ are the square roots of the eigenvalues of the matrix $M^*M$. For an $n \times n$ matrix, we will denote these $s_1(M_n), \ldots, s_n(M_n)$. Note that all singular values real and non-negative, so we let $s_1(M_n) \geq \cdots \geq s_n(M_n)$ by convention.

We write a.s., a.a., and a.e. for almost surely, Lebesgue almost all, and Lebesgue almost everywhere respectively. We use $\sqrt{-1}$ to denote the imaginary unit and reserve $i$ as an index.

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.

### 3.2. Linearization.

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

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

The following proposition relates the eigenvalues of $\mathcal{M}$ to the eigenvalues of the product $M_1 \cdots M_m$. We note that similar linearization tricks have been used previously; see, for example, [8, 19, 23, 53, 54] and references therein.

**Proposition 3.1** (Proposition 4.1 from [23])**.** *Let $M_1, \ldots, M_m$ be $n \times n$ matrices. Let $P := M_1 \cdots M_m$, and assume $\mathcal{M}$ is the $mn \times mn$ block matrix defined in* (8)*. Then*

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

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

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 $M, X$) and larger $N \times N$ matrices using calligraphic symbols (such as $\mathcal{M}$, $\mathcal{Y}$).

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 $1 \leq i, j \leq N$. 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 $M_n$.

3.4. **Singular Value Inequalities and Useful Identities.** Let $M$ denote an $n \times n$ matrix. We often want to know about the smallest and largest singular values of a matrix. Recall $s_1(M) \geq \cdots \geq s_n(M)$ denote the singular values of the matrix $M$.

**Proposition 3.2.** *Let $M$ be an $n \times n$ matrix, and assume that $E \subset \mathbb{C}$ and $c > 0$. If*

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

*then no eigenvalue of $M$ is contained in $E$ and*

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

*where $G(z) = (M - zI)^{-1}$ is the resolvent of $M$.*

The proof of Proposition 3.2 follows easily by observing that the operator norm of the resolvent can be bounded above by $1/s_n(M - zI)$; similar bounds were used in [51].

We will make use of the Sherman–Morrison rank one perturbation formula (see [37, Section 0.7.4]). Suppose $A$ is an invertible square matrix, and let $u$, $v$ be vectors. If $1 + v^* A^{-1} u \neq 0$, then

$$(A + uv^*)^{-1} = A^{-1} - \frac{A^{-1} uv^* A^{-1}}{1 + v^* A^{-1} u} \tag{9}$$

and

$$(A + uv^*)^{-1} u = \frac{A^{-1} u}{1 + v^* A^{-1} u}. \tag{10}$$

Also recall the Sherman–Morrison–Woodbury formula (for example, [24, Theorem 1.1]), which states that for an invertible $N \times N$ matrix $A$ and $a \times N$ matrices $V, U$ for some fixed $a < N$,

$$(A + UV^T)^{-1} U = A^{-1} U (I_a + V^T A^{-1} U)^{-1} \tag{11}$$

provided $I_a + V^T A^{-1} U$ is invertible.

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

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1} \tag{12}$$

whenever $A$ and $B$ are invertible.

We also use Weyl's inequality for the singular values (see, for example, [14, Problem III.6.5]), which states that for $n \times n$ matrices $A$ and $B$,

$$\max_{1 \leq i \leq n} |s_i(A) - s_i(B)| \leq \|A - B\| . \tag{13}$$

## 4. Reductions

In order to prove Theorem 2.2, we will make a series of reductions by truncating the entries in each factor matrix and applying the linearization techniques discussed above. Theorem 2.2 will follow from the following result.

**Theorem 4.1.** *Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \dots, \xi_m$ are real-valued random variables which satisfy Assumption 2.1. For each $n \geq 1$, let $X_{n,1}, \dots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \dots, \xi_m$, respectively. Define the products*

$$P_n := n^{-m/2} X_{n,1} \cdots X_{n,m} \quad \text{and} \quad \sigma = \sigma_1 \cdots \sigma_m.$$

*Let $\delta > 0$, and let $f$ be analytic in some neighborhood containing the disk $D_\delta := \{z \in \mathbb{C} : |z| \leq 1 + \delta\}$ and bounded otherwise. Then, there exists a constant $c > 0$ and a deterministic sequence $\mathcal{A}_n(f)$ (depending on $n$, $f$, and the distribution of $\xi_1, \dots, \xi_m$) such that the event*

$$E_n := \left\{ \inf_{|z| > 1 + \delta/2} s_n \left( P_n/\sigma - zI \right) \geq c \right\} \tag{14}$$

*holds with probability $1 - o(1)$ and as $n \to \infty$,*

$$\operatorname{tr} f(P_n/\sigma) \mathbf{1}_{E_n} - \mathcal{A}_n(f) \tag{15}$$

*converges in distribution to a mean zero Gaussian random variable $F(f)$ with covariance structure*

$$\mathbb{E}\left[ (F(f))^2 \right] = -\frac{1}{4\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} f(z) f(w) (zw - 1)^{-2} dz dw \tag{16}$$

*and*

$$\mathbb{E}\left[ F(f) \overline{F(f)} \right] = \frac{1}{4\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} f(z) \overline{f(w)} (z\bar{w} - 1)^{-2} dz d\bar{w} \tag{17}$$

*where $\mathcal{C}$ is the contour around the boundary of the disk $D_\delta$. In addition, the function $f \mapsto \mathcal{A}_n(f)$ is continuous and linear with the property that if $f(z) \in \mathbb{R}$ for all $z \in \mathbb{R} \cap D_\delta$, then $\mathcal{A}_n(f)$ is real-valued.*

We will now prove Theorem 2.2 assuming Theorem 4.1.

*Proof of Theorem 2.2.* Assume Theorem 4.1. It follows from standard least singular value bounds that there exists a constant $c > 0$ such that $E_n$ holds with probability $1 - o(1)$; the details are presented as Lemma B.3 from Appendix B. Thus, the prescience (or lack thereof) of the indicator function in (15) does not affect the limiting distribution.

To prove Theorem 2.2, we will invoke the Cramer–Wold device, but we will need to be careful as we are dealing with complex-valued random variables. If $f$ is analytic in a neighborhood containing the disk $D_\delta$, we can express $f$ as the power series

$$f(z) = \sum_{i=0}^{\infty} a_i z^i$$

in the same disk. We then define

$$\Re f(z) = \sum_{i=0}^{\infty} \mathrm{Re}(a_i) z^i \quad \text{and} \quad \Im f(z) = \sum_{i=0}^{\infty} \mathrm{Im}(a_i) z^i,$$

which are both analytic in $D_\delta$. (Notice that these are not the real and imaginary parts of $f$, which would not be analytic in $D_\delta$.)

In order to invoke the Cramer–Wold device and prove Theorem 2.2, we consider

$$f(z) = \sum_{l=0}^{s} \left( \alpha_l \Re f_l(z) + \beta_l \Im f_l(z) \right)$$

for some real-valued constants $\alpha_1, \beta_1, \dots, \alpha_s, \beta_s$.

As $f(z) \in \mathbb{R}$ for all $z \in \mathbb{R} \cap D_\delta$ and since the non-real eigenvalues of $P_n/\sigma$ come in complex conjugate pairs, it follows that $\operatorname{tr} f(P_n/\sigma) - \mathcal{A}_n(f)$ is real-valued. Applying the Cramer–Wold device and the convergence of (15), we conclude that

$$\left( \operatorname{tr} f_i(P_n/\sigma) - \mathcal{A}_n(f_i) \right)_{i=1}^{s}$$

converges to a multivariate Gaussian vector. The limiting covariances can now be extrapolated from (16) and (17). □

*Remark* 4.2. The proof above exploits the fact that the non-real eigenvalues come in complex conjugate pairs since the entries of the product matrix are real. In the case where the entries in each matrix are allowed to be complex-valued, this approach would no longer hold. In this case, one may apply a complex-valued version of the Cramer–Wold Theorem.

It remains to prove Theorem 4.1. By a simple rescaling, it is sufficient to prove Theorem 4.1 when $\sigma_i = 1$ for $1 \le i \le k$. For this reason, for the remainder of the paper we assume all atom variables have unit variance unless stated otherwise.

4.1. **Truncation of iid Matrices.** Since $\xi_k$ is assumed to have finite $4+\tau$ finite moments for $1 \le k \le m$, there exists $\varepsilon > 0$ such that for all $1 \le k \le m$,

$$\lim_{n\to\infty} n^{4\varepsilon} \mathbb{E}\left[ |\xi_k|^4 \mathbf{1}_{\{|\xi_k| > n^{1/2-\varepsilon}\}} \right] = 0. \tag{18}$$

Next, for a real-valued random variable $\xi$ with mean zero, variance one, and finite $(4+\tau)$th moment, define

$$\tilde{\xi} := \xi \mathbf{1}_{\{|\xi| \le n^{1/2-\varepsilon}\}} - \mathbb{E}\left[ \xi \mathbf{1}_{\{|\xi| \le n^{1/2-\varepsilon}\}} \right] \qquad \text{and} \qquad \hat{\xi} := \frac{\tilde{\xi}}{\sqrt{\mathrm{Var}(\tilde{\xi})}}. \tag{19}$$

Note that $\tilde{\xi}$ and $\hat{\xi}$ depend on $n$, but this dependence is not expressed in the notation.

**Lemma 4.3.** *Let $\xi$ be a real-valued random variable which satisfies Assumption 2.1 with unit variance and define $\tilde{\xi}$ and $\hat{\xi}$ as in (19). Then the following statements hold:*

(i) $|1 - \mathit{Var}(\tilde{\xi})| = o(n^{-1-2\varepsilon})$
(ii) *There exists an $N_0 > 0$ such that for any $n > N_0$, $\hat{\xi}$ has zero mean and unit variance, and almost surely*

$$|\hat{\xi}| \le 4n^{1/2-\varepsilon}.$$

(iii) *There exists* $N_0 > 0$ *such that for any* $n > N_0$*,*

$$\mathbb{E}|\hat{\xi}|^4 \leq 2^8 \mathbb{E}|\xi|^4.$$

The proof of this lemma is a standard truncation argument and can be found in Appendix A.

*Remark* 4.4. In the case where $\xi$ is complex-valued, this truncation will need to be modified in order to preserve independence between the real and imaginary parts of $\xi$ (see, for example, [23, Lemma 7.1]).

Next, we define various truncated matrices and prove a number of lemmas involving operator norms and Hilbert-Schmidt norms of these truncated matrices. The following lemmas will be used later in the proof.

Let $X$ be an $n \times n$ random matrix filled with iid copies of a random variable $\xi$ which satisfies Assumption 2.1. Define the $n \times n$ matrices $\dot{X}$, $\tilde{X}$, and $\hat{X}$ to be the matrices with entries given by

$$\dot{X}_{(i,j)} := X_{(i,j)} \mathbf{1}_{\{|X_{(i,j)}| \leq n^{1/2-\varepsilon}\}}, \tag{20}$$

$$\tilde{X}_{(i,j)} := X_{(i,j)} \mathbf{1}_{\{|X_{(i,j)}| \leq n^{1/2-\varepsilon}\}} - \mathbb{E}\left[ X_{(i,j)} \mathbf{1}_{\{|X_{(i,j)}| \leq n^{1/2-\varepsilon}\}} \right], \tag{21}$$

$$\hat{X}_{(i,j)} := \frac{\tilde{X}_{(i,j)}}{\sqrt{\operatorname{Var}(\tilde{X}_{(i,j)})}} \tag{22}$$

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

**Lemma 4.5.** *Let* $X_n$ *be an* $n \times n$ *iid random matrix with atom variable* $\xi$ *which satisfies Assumption 2.1 with unit variance and let* $\hat{X}_n$ *be the truncated matrix as defined in* (22)*. Then*

$$\mathbb{E}\left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2 = o\left(n^{-2\varepsilon}\right) \quad \textit{and} \quad \mathbb{P}\left( \left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\| > n^{-\varepsilon} \right) = o(1).$$

*Proof.* By Markov's inequality,

$$\begin{aligned}
\mathbb{P}\left( \left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\| > n^{-\varepsilon} \right) &\leq n^{2\varepsilon} \mathbb{E}\left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|^2 \\
&\leq n^{2\varepsilon} \mathbb{E}\left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2
\end{aligned}$$

so it is sufficient to prove $\mathbb{E}\left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2 = o\left(n^{-2\varepsilon}\right)$. By the triangle inequality,

$$\mathbb{E}\left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2 \ll \mathbb{E}\left[ \left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \tilde{X}_n \right\|_2^2 + \left\| \frac{1}{\sqrt{n}} \tilde{X}_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2 \right]$$

and we may deal with the two terms on the right hand side of the above expression separately. First, since $\left|\mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| > n^{1/2-\varepsilon}\}}\right]\right| = \left|\mathbb{E}[\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}}]\right|$ and by (18), we have

$$\mathbb{E}\left\|\frac{1}{\sqrt{n}}X_n - \frac{1}{\sqrt{n}}\tilde{X}_n\right\|_2^2 = \frac{1}{n}\mathbb{E}\left\|X_n - \tilde{X}_n\right\|_2^2$$
$$\leq \frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}\left|(X_n)_{(j,k)} - (\tilde{X}_n)_{(j,k)}\right|^2$$
$$\leq \frac{4}{n}\sum_{j,k=1}^{n}\mathbb{E}\left[|\xi|^2\frac{|\xi|^2}{(n^{1/2-\varepsilon})^2}\mathbf{1}_{\{|\xi|>n^{1/2-\varepsilon}\}}\right]$$
$$\leq \frac{4n^{2\varepsilon}}{n^{4\varepsilon}}n^{4\varepsilon}\mathbb{E}\left[|\xi|^4\mathbf{1}_{\{|\xi|>n^{1/2-\varepsilon}\}}\right]$$
$$= o(n^{-2\varepsilon}).$$

Observe that by Lemma 4.3, one has

$$\mathbb{E}\left\|\frac{1}{\sqrt{n}}\tilde{X}_n - \frac{1}{\sqrt{n}}\hat{X}_n\right\|_2^2 \leq \frac{1}{n}\sum_{j,k=1}^{n}\mathbb{E}\left|\hat{X}_{n(j,k)}\right|^2\left|\sqrt{\mathrm{Var}(\tilde{X}_{n,(j,k)})} - 1\right|^2$$
$$\leq n\left|\mathrm{Var}(\tilde{\xi}) - 1\right|^2$$
$$= o(n^{-1-4\varepsilon})$$

which concludes the proof. □

**Lemma 4.6.** *Let $X_n$ be an $n\times n$ iid random matrix with atom variable $\xi$ which satisfies Assumption 2.1 with unit variance. Let $\dot{X}_n$ and $\hat{X}_n$ be the truncated matrices as defined in* (20) *and* (22) *respectively. Then*

$$\mathbb{E}\left\|\hat{X}_n - \dot{X}_n\right\|_2^2 = o(1).$$

*Proof.* Let $\tilde{\xi}$ be as defined in (19) and observe that by the proof of Lemma 4.3 *(ii)*, $(\mathrm{Var}(\tilde{\xi}))^{-1/2} \leq 2$ for $n$ sufficiently large. Therefore

$$\mathbb{E}\left\|\hat{X}_n - \dot{X}_n\right\|_2^2 = \mathbb{E}\left[\sum_{i,j=1}^{n}\left|\hat{X}_{n,(i,j)} - \dot{X}_{n,(i,j)}\right|^2\right]$$
$$\leq n^2\mathbb{E}\left|\frac{\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}(1-(\mathrm{Var}(\tilde{\xi}))^{1/2}) - \mathbb{E}\left[\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}\right]}{(\mathrm{Var}(\tilde{\xi}))^{1/2}}\right|^2$$
$$\ll n^2\mathbb{E}\left|\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}(1-(\mathrm{Var}(\tilde{\xi}))^{1/2}) - \mathbb{E}\left[\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}\right]\right|^2$$
$$\ll n^2\left|1-\mathrm{Var}(\tilde{\xi})\right|^2\mathbb{E}\left[|\xi|^2\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}\right] + n^2\left|\mathbb{E}\left[\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}\right]\right|^2$$

for $n$ sufficiently large. By Lemma 4.3 *(i)*, we have

$$n^2\left|1-\mathrm{Var}(\tilde{\xi})\right|^2\mathbb{E}\left[|\xi|^2\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}\right] = o(n^{-4\varepsilon}).$$

Next, observe that since $\left|\mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}}\right]\right| = \left|\mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| > n^{1/2-\varepsilon}\}}\right]\right|$, we have

$$\begin{aligned} n^2 \left|\mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}}\right]\right|^2 &= n^2 \left|\mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| > n^{1/2-\varepsilon}\}}\right]\right|^2 \\ &\leq n^{-1+6\varepsilon} \left(\mathbb{E}\left[|\xi|^4 \mathbf{1}_{\{|\xi| > n^{1/2-\varepsilon}\}}\right]\right)^2 \\ &= o(1). \end{aligned}$$

□

**Lemma 4.7.** *Let $\hat{X}_n$ be an $n \times n$ iid random matrix with atom variable $\hat{\xi}$ which has mean zero, variance one, $\mathbb{E}|\hat{\xi}|^4 = O(1)$, and satisfies $|\hat{\xi}| \ll n^{1/2-\varepsilon}$ almost surely for some $\varepsilon > 0$. Then $\mathbb{E}\left\|\hat{X}_n\right\|^2 = O(n)$ where $\|\cdot\|$ denotes the operator norm.*

*Proof.* Observe that, for any constant $C > 0$

$$\begin{aligned} \mathbb{E}\left\|\hat{X}_n\right\|^2 &\ll \mathbb{E}\left\|\hat{X}_n \mathbf{1}_{\{\|\hat{X}_n\| \leq C\sqrt{n}\}}\right\|^2 + \mathbb{E}\left\|\hat{X}_n \mathbf{1}_{\{\|\hat{X}_n\| > C\sqrt{n}\}}\right\|^2 \\ &\ll n + n^{3-2\varepsilon} \mathbb{P}\left(\left\|\hat{X}_n\right\| > C\sqrt{n}\right) \end{aligned}$$

where the power of $n$ came from bounding the operator norm by the Frobenious norm. By [13, Theorem 5.9], there exists $C > 0$ sufficiently large so that $\mathbb{P}\left(\left\|\hat{X}_n\right\| > C\sqrt{n}\right) = O_\alpha(n^{-\alpha})$ for any $\alpha > 0$. By selecting $\alpha$ sufficiently large, we arrive at the desired result. □

**Lemma 4.8.** *Let $X_n$ be an $n \times n$ iid random matrix with atom variable $\xi$ which has mean zero, variance one, and finite $4 + \tau$ moment for some $\tau > 0$ and define $\dot{X}_n$ as in (20). Then $\mathbb{E}\left\|\dot{X}_n\right\|^2 = O(n)$ where $\|\cdot\|$ denotes the operator norm.*

*Proof.* Let $\hat{X}_n$ be the truncated $n \times n$ iid random matrix with entries as defined in (22) and observe that

$$\mathbb{E}\left\|\dot{X}_n\right\|^2 \ll \mathbb{E}\left\|\dot{X}_n - \hat{X}_n\right\|^2 + \mathbb{E}\left\|\hat{X}_n\right\|^2.$$

The proof follows by Lemmas 4.7 and 4.6. □

**Lemma 4.9.** *Let $X_n$ be an $n \times n$ iid random matrix with atom variable $\xi$ which has mean zero, variance one, and finite $4 + \tau$ moment for some $\tau > 0$. Then $\mathbb{E}\|X_n\|^2 = O(n)$ where $\|\cdot\|$ denotes the operator norm.*

*Proof.* Let $\hat{X}_n$ be the truncated $n \times n$ iid random matrix with entries defined by (22) and observe that by the triangle inequality we have

$$\mathbb{E}\|X_n\|^2 \ll \mathbb{E}\left\|X_n - \hat{X}_n\right\|^2 + \mathbb{E}\left\|\hat{X}_n\right\|^2. \tag{23}$$

Both terms in the right hand side of (23) are $O(n)$ by Lemmas 4.5 and 4.7 as desired. □

**Lemma 4.10.** *Let $X_{n,i}$ be as defined in Theorem 2.2 with $\sigma_i = 1$, and for each $1 \leq i \leq m$, define $\hat{X}_{n,i}$ as in (22). Define the product $P_n$ as in (4) and define the truncated product*

$$\hat{P}_n = n^{-m/2} \hat{X}_{n,1} \cdots \hat{X}_{n,m}. \tag{24}$$

*Then*

$$\mathbb{E}\left\|P_n - \hat{P}_n\right\|_2^2 = o(n^{-2\varepsilon}) \quad \text{and} \quad \mathbb{P}\left(\left\|P_n - \hat{P}_n\right\| > n^{-\varepsilon}\right) = o(1).$$

*Proof.* By Markov's inequality,

$$\mathbb{P}\left(\left\|P_n - \hat{P}_n\right\| > n^{-\varepsilon}\right) \leq n^{2\varepsilon}\mathbb{E}\left\|P_n - \hat{P}_n\right\|^2 \leq n^{2\varepsilon}\mathbb{E}\left\|P_n - \hat{P}_n\right\|_2^2$$

so it is sufficient to prove the first bound. To this end, note that by the triangle inequality, independence, and Lemma C.5, we have

$$\begin{aligned}\mathbb{E}\left\|P_n - \hat{P}_n\right\|_2^2 &= \mathbb{E}\left\|n^{-m/2}X_{n,1}\cdots X_{n,m} - n^{-m/2}\hat{X}_{n,1}\cdots\hat{X}_{n,m}\right\|_2^2 \\ &\ll n^{-m}\left(\mathbb{E}\left\|X_{n,1} - \hat{X}_{n,1}\right\|_2^2\mathbb{E}\left\|X_{n,2}\right\|^2\cdots\mathbb{E}\left\|X_{n,m-1}\right\|^2\mathbb{E}\left\|X_{n,m}\right\|^2 + \dots\right. \\ &\qquad \left. + \mathbb{E}\left\|\hat{X}_{n,1}\right\|^2\mathbb{E}\left\|\hat{X}_{n,2}\right\|^2\cdots\mathbb{E}\left\|\hat{X}_{n,m-1}\right\|^2\mathbb{E}\left\|X_{n,m} - \hat{X}_{n,m}\right\|_2^2\right).\end{aligned}$$

By Lemmas 4.7 and 4.9, $\mathbb{E}\left\|\hat{X}_{n,k}\right\|^2 = O(n)$ and $\mathbb{E}\left\|X_{n,k}\right\|^2 = O(n)$ for all $1 \leq k \leq m$. Therefore, by this observation and Lemma 4.5,

$$\begin{aligned}\mathbb{E}\left\|P_n - \hat{P}_n\right\|_2^2 &\ll n^{-1}\left(\mathbb{E}\left\|X_{n,1} - \hat{X}_{n,1}\right\|_2^2 + \dots + \mathbb{E}\left\|X_{n,m} - \hat{X}_{n,m}\right\|_2^2\right) \\ &= o(n^{-2\varepsilon}).\end{aligned}$$

□

**Lemma 4.11.** *Let $X_{n,i}$ be as defined in Theorem 2.2 with $\sigma_i = 1$, and for each $1 \leq i \leq m$, define $\dot{X}_{n,i}$ and $\hat{X}_{n,i}$ as in* (20) *and* (22) *respectively. Define $\hat{P}_n$ as in* (24) *and define the product*

$$\dot{P}_n = n^{-m/2}\dot{X}_{n,1}\cdots\dot{X}_{n,m}. \tag{25}$$

*Then*

$$\mathbb{E}\left\|\dot{P}_n - \hat{P}_n\right\|_2^2 = o(n^{-1}).$$

*Proof.* By the triangle inequality, independence, and Lemma C.5, we have

$$\begin{aligned}\mathbb{E}\left\|\dot{P}_n - \hat{P}_n\right\|_2^2 &= \mathbb{E}\left\|n^{-m/2}\dot{X}_{n,1}\cdots\dot{X}_{n,m} - n^{-m/2}\hat{X}_{n,1}\cdots\hat{X}_{n,m}\right\|_2^2 \\ &\ll n^{-m}\left(\mathbb{E}\left\|\dot{X}_{n,1} - \hat{X}_{n,1}\right\|_2^2\mathbb{E}\left\|\dot{X}_{n,2}\right\|^2\cdots\mathbb{E}\left\|\dot{X}_{n,m-1}\right\|^2\mathbb{E}\left\|\dot{X}_{n,m}\right\|^2 + \dots\right. \\ &\qquad \left. + \mathbb{E}\left\|\hat{X}_{n,1}\right\|^2\mathbb{E}\left\|\hat{X}_{n,2}\right\|^2\cdots\mathbb{E}\left\|\hat{X}_{n,m-1}\right\|^2\mathbb{E}\left\|\dot{X}_{n,m} - \hat{X}_{n,m}\right\|_2^2\right).\end{aligned}$$

By Lemmas 4.7 and 4.8, $\mathbb{E}\left\|\hat{X}_{n,k}\right\|^2 = O(n)$ and $\mathbb{E}\left\|\dot{X}_{n,k}\right\|^2 = O(n)$ for all $1 \leq k \leq m$. By this observation and Lemma 4.6,

$$\begin{aligned}\mathbb{E}\left\|\dot{P}_n - \hat{P}_n\right\|_2^2 &\ll n^{-1}\left(\mathbb{E}\left\|\dot{X}_{n,1} - \hat{X}_{n,1}\right\|_2^2 + \dots + \mathbb{E}\left\|\dot{X}_{n,m} - \hat{X}_{n,m}\right\|_2^2\right) \\ &= o(n^{-1}).\end{aligned}$$

□

With the preceding norm lemmas complete, we now show it is sufficient to consider a version of Theorem 4.1 in which all entries in the matrices are truncated. We now reduce to the case where we can consider the truncated product $\hat{P}_n$.

**Theorem 4.12.** *Let $X_{n,i}$ be as defined in Theorem 2.2 with $\sigma_i = 1$, $\hat{X}_{n,i}$ as defined in* (19)*, and $\hat{P}_n$ as in* (24)*. Let $\delta > 0$, and let $f$ be a function which is analytic in some neighborhood containing the disk $D_\delta$ and bounded otherwise. Then there exists a constant $c > 0$ such that the event*

$$\hat{E}_n := \left\{ \inf_{|z|>1+\delta/2} s_n(\hat{P}_n - zI) \geq c \right\} \tag{26}$$

*holds with overwhelming probability and*

$$\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n} - \mathbb{E}[\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n}] \tag{27}$$

*converges in distribution to a mean-zero Gaussian random variable $F(f)$ with covariance structure*

$$\mathbb{E}\left[(F(f))^2\right] = -\frac{1}{4\pi^2}\oint_{\mathcal{C}}\oint_{\mathcal{C}} f(z)f(w)(zw-1)^{-2}dzdw$$

*and*

$$\mathbb{E}\left[F(f)\overline{F(f)}\right] = \frac{1}{4\pi^2}\oint_{\mathcal{C}}\oint_{\mathcal{C}} f(z)\overline{f(w)}(z\bar{w}-1)^{-2}dzd\bar{w}.$$

We now prove Theorem 4.1 assuming Theorem 4.12.

*Proof of Theorem 4.1.* Suppose the conclusion of Theorem 4.12 holds. We define

$$\mathcal{A}_n(f) := \mathbb{E}\left[\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n}\right]. \tag{28}$$

There exists $c > 0$ such that $\hat{E}_n$ holds with overwhelming probability by Lemma B.2, and $E_n$ holds with probability $1 - o(1)$ by Lemma B.3. Thus, we may work on the intersection of these events, and in order to show that $\operatorname{tr} f(P_n)\mathbf{1}_{E_n} - \mathcal{A}_n(f)$ converges to a mean-zero Gaussian random variable with variance as in (16) and (17), it is sufficient to show that for any $\eta > 0$,

$$\mathbb{P}\left(\left|\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{E_n\cap\hat{E}_n} - \operatorname{tr} f(P_n)\mathbf{1}_{E_n\cap\hat{E}_n}\right| > \eta\right) = o(1).$$

To this end, define $\dot{X}_{n,k}$ as in (20) for each $1 \leq k \leq m$ and $\dot{P}_n$ as in (25). Observe that for any $1 \leq k \leq m$, by (18)

$$\begin{aligned}
\mathbb{P}(X_{n,k} \neq \dot{X}_{n,k}) &= \mathbb{P}\left(\bigcup_{i,j}\left\{|(X_{n,k})_{ij}| > n^{1/2-\varepsilon}\right\}\right) \\
&\leq n^2\mathbb{E}\left[\mathbf{1}_{\{|\xi_k|>n^{1/2-\varepsilon}\}}\right] \\
&\leq n^{4\varepsilon}\mathbb{E}\left[|\xi_k|^4\mathbf{1}_{\{|\xi|>n^{1/2-\varepsilon}\}}\right] \\
&= o(1).
\end{aligned}$$

By a union bound over all $1 \leq k \leq m$, we have that $\mathbb{P}\left(P_n \neq \dot{P}\right) = o(1)$ as well. Therefore, it is sufficient to show that

$$\mathbb{P}\left(\left|\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{E_n\cap\hat{E}_n} - \operatorname{tr} f(\dot{P}_n)\mathbf{1}_{E_n\cap\hat{E}_n}\right| > \eta\right) = o(1). \tag{29}$$

Define the event $\dot{E}_n := \{P_n = \dot{P}_n\} \cap E_n$ and observe that $\dot{E}_n$ holds with probability $1 - o(1)$. By this fact and (29), it is sufficient to prove

$$\mathbb{P}\left(\left|\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\dot{E}_n \cap \hat{E}_n} - \operatorname{tr} f(\dot{P}_n)\mathbf{1}_{\dot{E}_n \cap \hat{E}_n}\right| > \eta\right) = o(1). \tag{30}$$

On the events $\dot{E}_n$ and $\hat{E}_n$, the eigenvalues of $\hat{P}_n$ and $\dot{P}_n$ are contained in the interior of the contour $\mathcal{C}$, which is defined as the boundary of the disk $D_\delta$. Therefore, on $\dot{E}_n$, for any $z \in \mathcal{C}$, the least singular values of $\dot{P}_n - zI$ is bounded away from zero. Thus, by Cauchy's integral formula,

$$\begin{aligned}
&\mathbb{E}\left|(\operatorname{tr} f(\hat{P}_n) - \operatorname{tr} f(\dot{P}_n))\mathbf{1}_{\dot{E}_n \cap \hat{E}_n}\right|^2 \\
&\quad = \mathbb{E}\left|-\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)\left(\operatorname{tr}(\hat{P}_n - zI)^{-1} - \operatorname{tr}(\dot{P}_n - zI)^{-1}\right)\mathbf{1}_{\dot{E}_n \cap \hat{E}_n}\, dz\right|^2 \\
&\quad \ll_f \mathbb{E}\left[\sup_{z \in \mathcal{C}}\left|\operatorname{tr}(\hat{P}_n - zI)^{-1} - \operatorname{tr}(\dot{P}_n - zI)^{-1}\right|^2 \mathbf{1}_{\dot{E}_n \cap \hat{E}_n}\right].
\end{aligned}$$

Since $f$ is assumed to be analytic on the disk and bounded otherwise, by applying Markov's inequality to the left-hand side of (29), it is sufficient to show that

$$\mathbb{E}\left[\sup_{z \in \mathcal{C}}\left|\operatorname{tr}(\hat{P}_n - zI)^{-1} - \operatorname{tr}(\dot{P}_n - zI)^{-1}\right|^2 \mathbf{1}_{\dot{E}_n \cap \hat{E}_n}\right] = o(1).$$

By the resolvent identity, Lemma C.5, and Lemma 4.11, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{z \in \mathcal{C}}\left|(\operatorname{tr}(\hat{P}_n - zI)^{-1} - \operatorname{tr}(\dot{P}_n - zI)^{-1})\right|^2 \mathbf{1}_{\dot{E}_n \cap \hat{E}_n}\right] \\
&\quad = \mathbb{E}\left[\sup_{z \in \mathcal{C}}\left|\operatorname{tr}\left((\hat{P}_n - zI)^{-1}(\dot{P}_n - \hat{P}_n)(\dot{P}_n - zI)^{-1}\right)\right|^2 \mathbf{1}_{\dot{E}_n \cap \hat{E}_n}\right] \\
&\quad \ll n\mathbb{E}\left\|\dot{P}_n - \hat{P}_n\right\|_2^2 \\
&\quad = o(1)
\end{aligned}$$

since the spectral norms of the resolvents are bounded uniformly by a constant (Proposition 3.2) on their respective events. $\square$

4.2. **Linearization of the Product.** We now wish to linearize the product matrix $\hat{P}_n$ so that we can work with an $mn \times mn$ block matrix instead. Define the $mn \times mn$ matrix

$$\mathcal{Y}_n := n^{-1/2}\begin{bmatrix} 0 & \hat{X}_{n,1} & 0 & \cdots & 0 \\ 0 & 0 & \hat{X}_{n,2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \hat{X}_{n,m-1} \\ \hat{X}_{n,m} & 0 & 0 & \dots & 0 \end{bmatrix}. \tag{31}$$

Recall that by Proposition 3.1, $\mathcal{Y}_n^m$ has the same eigenvalues as $\hat{P}_n = n^{-m/2}\hat{X}_{n,1}\cdots\hat{X}_{n,m}$, each with multiplicity $m$. Therefore, the eigenvalues of the product $\hat{P}_n$ are completely determined by the eigenvalues of the linearized matrix $\mathcal{Y}_n$.

**Theorem 4.13.** *Let $\mathcal{Y}_n$ be the linearized matrix defined in* (31) *where $X_{n,i}$ are under the assumptions of Theorem 2.2 with $\sigma_i = 1$ and the entries of $\hat{X}_{n,i}$ are* *truncated as defined in* (22). *For every* $\delta > 0$, *there exists* $c > 0$ *such that the following holds. The event*

$$\Omega_n := \left\{ \inf_{|z|>1+\delta/2} s_{mn}(\mathcal{Y}_n - zI) \geq c \right\} \tag{32}$$

*holds with overwhelming probability, and for any function* $g$ *which is analytic in a neighborhood of the disk* $D_\delta$ *and bounded otherwise, the random variable*

$$\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n} - \mathbb{E}\left[\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n}\right]$$

*converges to a mean zero Gaussian random variable* $F(g)$ *with covariance structure*

$$\mathbb{E}\left[(F(g))^2\right] = -\frac{1}{4\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} g(z)g(w) \frac{m^2(zw)^{m-1}}{((zw)^m - 1)^2} dzdw \tag{33}$$

*and*

$$\mathbb{E}\left[F(g)\overline{F(g)}\right] = \frac{1}{4\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} g(z)\overline{g(w)} \frac{m^2(z\bar{w})^{m-1}}{((z\bar{w})^m - 1)^2} dzd\bar{w}, \tag{34}$$

*where* $\mathcal{C}$ *is the contour around the boundary of* $D_\delta$.

Note that the convergence in (33) and (34) depend on $m$. We prove Theorem 4.12 assuming Theorem 4.13.

*Proof of Theorem 4.12.* First begin by observing that by assumption, there exists a $c > 0$ such that $\Omega_n$ holds with overwhelming probability and by Lemma B.2, there exists another constant $c' > 0$ such that $\hat{E}_n$ holds with overwhelming probability as well. Let $f$ be any function which is analytic on the disk $D_\delta$ and bounded otherwise. Define the function $g(z) := \frac{1}{m} f(z^m)$ and note that this function $g$ is analytic on the disk $\{z \in \mathbb{C} : |z| \leq (1+\delta)^{1/m}\} = D_{\delta'}$ for some $\delta' > 0$ and bounded otherwise. By Proposition 3.1,

$$\operatorname{tr} f(\hat{P}_n) = \sum_{i=1}^{n} f(\lambda_i(\hat{P}_n)) = \sum_{i=1}^{mn} \frac{1}{m} f(\lambda_i(\mathcal{Y}_n^m)) = \sum_{i=1}^{mn} g(\lambda_i(\mathcal{Y}_n)) = \operatorname{tr} g(\mathcal{Y}_n).$$

By assumption, $\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n} - \mathbb{E}\left[\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n}\right]$ converges to a mean-zero Gaussian with covariance structure given in (33) and (34). We will show that

$$\mathbb{E}\left|\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n} - \mathbb{E}\left[\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n}\right] - \left(\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n} - \mathbb{E}\left[\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n}\right]\right)\right| = o(1).$$

To this end, observe that

$$\begin{aligned}
&\mathbb{E}\left|\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n} - \mathbb{E}\left[\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n}\right] - \left(\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n} - \mathbb{E}\left[\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n}\right]\right)\right| \\
&\quad = \mathbb{E}\left|\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n} - \operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n} - \mathbb{E}\left[\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n} - \operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n}\right]\right| \\
&\quad \leq 2\mathbb{E}\left|\left(\operatorname{tr} g(\mathcal{Y}_n) - \operatorname{tr} f(\hat{P}_n)\right)\mathbf{1}_{\hat{E}_n \cap \Omega_n}\right| \\
&\qquad\qquad + 2\mathbb{E}\left|\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n \cap \hat{E}_n^c}\right| + 2\mathbb{E}\left|\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n \cap \Omega_n^c}\right| \\
&\quad \ll_{f,g} 0 + n\mathbb{P}\left(\Omega_n \cap \hat{E}_n^c\right) + n\mathbb{P}\left(\hat{E}_n \cap \Omega_n^c\right) \\
&\quad = o(1)
\end{aligned}$$

since $\Omega_n$ and $\hat{E}_n$ both hold with overwhelming probability by assumption and Lemma B.2 respectively. To see that the variance follows as claimed, observe that by letting $z = re^{\theta_1\sqrt{-1}}$ and $w = re^{\theta_2\sqrt{-1}}$ where $r = 1+\delta$, we have

$$\begin{aligned}
&\frac{1}{4\pi^2}\oint_{\mathcal{C}}\oint_{\mathcal{C}}\frac{1}{m^2}f(z^m)\overline{f(w^m)}\frac{m^2(z\bar{w})^{m-1}}{((z\bar{w})^m-1)^2}dzd\bar{w}\\
&\quad=\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}f(r^me^{m\theta_1\sqrt{-1}})\overline{f(r^me^{m\theta_2\sqrt{-1}})}\frac{r^{2m}e^{m\theta_1\sqrt{-1}}e^{-m\theta_2\sqrt{-1}}}{(r^{2m}e^{m\theta_1\sqrt{-1}}e^{-m\theta_2\sqrt{-1}}-1)^2}d\theta_1d\theta_2.
\end{aligned}$$

Next by the substitution $m\theta_1 = \tau_1$ and $m\theta_2 = \tau_2$, and by noting that this substitution wraps around the contour $m$ times,

$$\begin{aligned}
&\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}f(r^me^{m\theta_1\sqrt{-1}})\overline{f(r^me^{m\theta_2\sqrt{-1}})}\frac{r^{2m}e^{m\theta_1\sqrt{-1}}e^{-m\theta_2\sqrt{-1}}}{(r^{2m}e^{m\theta_1\sqrt{-1}}e^{-m\theta_2\sqrt{-1}}-1)^2}d\theta_1d\theta_2\\
&\qquad=\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}f(r^me^{\tau_1\sqrt{-1}})\overline{f(r^me^{\tau_2\sqrt{-1}})}\frac{r^{2m}e^{\tau_1\sqrt{-1}}e^{-\tau_2\sqrt{-1}}}{(r^{2m}e^{\tau_1\sqrt{-1}}e^{-\tau_2\sqrt{-1}}-1)^2}d\tau_1d\tau_2
\end{aligned}$$

and finally, by letting $z' = r^me^{\tau_1\sqrt{-1}}$ and $w' = r^me^{\tau_2\sqrt{-1}}$, we have the claimed variance. □

## 5. Proof of Theorem 4.13

It remains to prove Theorem 4.13. Define the resolvent

$$\mathcal{G}_n(z) := (\mathcal{Y}_n - zI)^{-1} \tag{35}$$

where $\mathcal{Y}_n$ is defined in (31) and for $z$ not an eigenvalue of $\mathcal{Y}_n$. Also define

$$\Xi_n(z) := \operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n} - \mathbb{E}\left[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n}\right]. \tag{36}$$

By Lemma B.1, $\Omega_n$ holds with overwhelming probability. On the event $\Omega_n$, the eigenvalues of $\mathcal{Y}_n$ are contained in the interior of the disk $D_\delta$, and so, by Cauchy's integral formula, we have

$$\begin{aligned}
\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n} &- \mathbb{E}\left[\operatorname{tr} g(\mathcal{Y}_n)\mathbf{1}_{\Omega_n}\right]\\
&= \sum_{i=1}^{mn} g(\lambda_i(\mathcal{Y}_n))\mathbf{1}_{\Omega_n} - \mathbb{E}\left[\sum_{i=1}^{mn} g(\lambda_i(\mathcal{Y}_n))\mathbf{1}_{\Omega_n}\right]\\
&= \sum_{i=1}^{mn} -\frac{1}{2\pi i}\oint_{\mathcal{C}}\frac{g(z)\mathbf{1}_{\Omega_n}}{\lambda_i(\mathcal{Y}_n)-z}dz - \mathbb{E}\left[\sum_{i=1}^{mn} -\frac{1}{2\pi i}\oint_{\mathcal{C}}\frac{g(z)\mathbf{1}_{\Omega_n}}{\lambda_i(\mathcal{Y}_n)-z}dz\right]\\
&= -\frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\left(\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n} - \mathbb{E}\left[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n}\right]\right)dz\\
&= -\frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\Xi_n(z)dz
\end{aligned}$$

where for the remainder of the paper we let $\mathcal{C}$ denote the contour on the boundary of $D_\delta$.

We will reduce the proof of Theorem 4.13 to showing the convergence of the resolvent process $(\Xi_n(z))_{z\in\mathcal{C}}$ to the limiting Gaussian process defined in the following lemma.

**Lemma 5.1.** *Let $\{\Xi(z)\}_{z\in\mathcal{C}}$ denote the mean-zero Gaussian process with covariance structure defined by*

$$\mathbb{E}\left[\Xi(z)\overline{\Xi(w)}\right] = \frac{m^2(z\bar{w})^{m-1}}{((z\bar{w})^m - 1)^2} \tag{37}$$

*and with the property that $\overline{\Xi(z)} = \Xi(\bar{z})$ for all $z \in \mathcal{C}$. If $g$ be a function which is analytic on some neighborhood of the disk $D_\delta$ and bounded otherwise, then*

$$-\frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\Xi(z)dz \tag{38}$$

*is a mean-zero Gaussian random variable with covariance structure*

$$\mathbb{E}\left[\left(\frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\Xi(z)dz\right)^2\right] = -\frac{1}{4\pi^2}\oint_{\mathcal{C}}\oint_{\mathcal{C}} g(z)g(w)\frac{m^2(zw)^{m-1}}{((zw)^m-1)^2}dzdw \tag{39}$$

*and*

$$\mathbb{E}\left[\frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\Xi(z)dz \cdot \overline{\frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\Xi(z)dz}\right] = \frac{1}{4\pi^2}\oint_{\mathcal{C}}\oint_{\mathcal{C}} g(z)\overline{g(w)}\frac{m^2(z\bar{w})^{m-1}}{((z\bar{w})^m-1)^2}dzd\bar{w}. \tag{40}$$

*Proof.* The proof follows by a number of standard techniques. For instance, one can deduce the conclusion by computing moments (with an application of Fubini's theorem); we omit the details. □

The following result shows that $\{\Xi(z)\}_{z\in\mathcal{C}}$ is indeed the limiting distribution of the resolvent process $\{\Xi_n(z)\}_{z\in\mathcal{C}}$.

**Theorem 5.2.** *Let $\{\Xi_n(z)\}_{z\in\mathcal{C}}$ be the sequence of stochastic processes defied in* (36) *for $z$ on the contour $\mathcal{C}$ around the boundary of the disk $D_\delta$. Then $\{\Xi_n(z)\}_{z\in\mathcal{C}}$ converges in distribution to the mean-zero Gaussian process $\{\Xi(z)\}_{z\in\mathcal{C}}$ defined in Lemma 5.1.*

The bulk of the paper is devoted to the proof of Theorem 5.2. Before doing so, let us complete the proof of Theorem 4.13 assuming Theorem 5.2.

First note that by Lemma B.1, there exists $c > 0$ such that $\Omega_n$ holds with overwhelming probability. Next, observe that $\Xi_n(z)$ and $\Xi(z)$ are random elements in the space of continuous functions on the contour $\mathcal{C}$, which is a metric space with respect to the supremum norm. Since the map

$$\Xi_n(z) \mapsto \frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\Xi_n(z)dz \tag{41}$$

is continuous in this metric space, the continuous mapping theorem (see [15, Theorem 25.7]) and Lemma 5.1 show that Theorem 5.2 implies Theorem 4.13. Indeed, if $\{\Xi_n(z)\}_{z\in\mathcal{C}}$ converges in distribution to $\{\Xi(z)\}_{z\in\mathcal{C}}$, then $\frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\Xi_n(z)dz$ converges in distribution to $\frac{1}{2\pi i}\oint_{\mathcal{C}} g(z)\Xi(z)dz$ as desired.

In order to prove Theorem 5.2, we will use the following characterization of convergence, which is a result of Theorems 7.5 and 12.3 from [16].

**Theorem 5.3.** *Suppose that $\{\Xi(z)\}_{z\in\mathcal{C}}, \{\Xi_n(z)\}_{z\in\mathcal{C}}$ for $n \geq 1$ are stochastic processes on the contour $\mathcal{C} = \{z \in \mathbb{C} : |z| = 1+\delta\}$. Suppose $\Xi(z)$, $(\Xi_n(z))_{n=1}^{\infty}$ satisfy*

$$(\Xi_n(z_1), \Xi_n(z_2), \dots, \Xi_n(z_L)) \longrightarrow (\Xi(z_1), \Xi(z_2), \dots, \Xi(z_L)) \tag{42}$$

*in distribution as $n \to \infty$ for any fixed positive integer $L$ and any $z_1, \dots, z_L \in \mathcal{C}$, and suppose that there exists a constant $c > 0$ such that*

$$\sup_{z,w \in \mathcal{C}, z \neq w} \mathbb{E} \left| \frac{\Xi_n(z) - \Xi_n(w)}{z - w} \right|^2 \leq c \tag{43}$$

*for all $n$. Then $\{\Xi_n(z)\}_{z \in \mathcal{C}}$ converges in distribution to $\{\Xi(z)\}_{z \in \mathcal{C}}$ as $n \to \infty$.*

The proof of Theorem 5.2 reduces to showing that the two conditions from Theorem 5.3 are satisfied. In Section 6 we prove the convergence of the finite dimensional distributions (42). Section 7 contains the proof of the tightness of the sequence of stochastic processes (43).

## 6. Convergence of Finite Dimensional Distributions

This section is devoted to proving the convergence of the finite dimensional distributions of the stochastic process $\{\Xi_n(z)\}_{z \in \mathcal{C}}$. In particular, this section will be devoted to the proof of the following theorem.

**Theorem 6.1.** *For a fixed positive integer $L$ and any collection $(z_1, z_2, \dots, z_L)$ such that $|z_i| = 1 + \delta$ for $1 \leq i \leq L$, the random vector $(\Xi_n(z_1), \Xi_n(z_2), \dots, \Xi_n(z_L))$ converges in distribution to the random vector $(\Xi(z_1), \Xi(z_2), \dots, \Xi(z_L))$ where $\{\Xi(z)\}_{z \in \mathcal{C}}$ is defined in Lemma 5.1.*

To prove Theorem 6.1, we first make a sequence of reductions inspired by the proofs in [56, 51]. First, recall that by the Cramer–Wold theorem, it is sufficient to prove the convergence of an arbitrary linear combination of the components of the vector in question. Ergo, by the Cramer–Wold theorem, it is sufficient to show that

$$\sum_{l=1}^{L} (\alpha_l \Xi_n(z_l) + \beta_l \overline{\Xi_n(z_l)}) \tag{44}$$

converges in distribution to

$$\sum_{l=1}^{L} (\alpha_l \Xi(z_l) + \beta_l \overline{\Xi(z_l)}) \tag{45}$$

for $\alpha_l, \beta_l \in \mathbb{C}$ such that (44) is real. As stated in Lemma 5.1, since $\overline{\Xi(z)} = \Xi(\bar{z})$, it is sufficient to compute $\mathbb{E}[\Xi(z_i)\overline{\Xi(z_j)}]$ in order to characterize the covariance structure of $(\Xi(z_1), \Xi(z_2), \dots, \Xi(z_L))$.

*Remark* 6.2. In the case where the atom random variables are complex-valued, we would need to compute $\mathbb{E}[\Xi(z_i)\Xi(z_j)]$, and $\mathbb{E}[\Xi(z_i)\overline{\Xi(z_j)}]$ in order to characterize the covariance.

In order to prove Theorem 6.1, we will express the sum in (44) as a martingale difference sequence. Let $c_k$ denote the $k$th column of $\mathcal{Y}_n$ and define the $\sigma$-algebras

$$\mathcal{F}_k := \sigma(c_1, \dots, c_k, c_{n+1}, \dots, c_{n+k}, \dots, c_{(m-1)n+1}, \dots, c_{(m-1)n+k}) \tag{46}$$

for $1 \leq k \leq n$. Note that $\mathcal{F}_k$ is the $\sigma$-algebra generated by the first $k$ columns of each of the $n \times n$ blocks of $\mathcal{Y}_n$. Define $\mathcal{F}_0$ to be the trivial $\sigma$-algebra and note that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$. Then define the conditional expectation $\mathbb{E}_k[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathcal{F}_k]$ and observe that by definition of the $\sigma$-algebras, $\mathbb{E}_0[\,\cdot\,] = \mathbb{E}[\,\cdot\,]$ and $\mathbb{E}_n[\mathcal{Y}_n] = \mathcal{Y}_n$.

Also define $\mathcal{Y}_n^{(k)}$ to be the matrix $\mathcal{Y}_n$ with the columns $c_k, c_{n+k}, c_{2n+k}, \dots, c_{(m-1)n+k}$ replaced with zeros. Note that $\mathcal{Y}_n^{(k)}$ can be viewed as the matrix $\mathcal{Y}_n$ with the $k$th column in each block replaced by zeros. By Corollary B.5, for every $\delta > 0$, there exists some $c > 0$ such that the event

$$\Omega_{n,k} := \left\{ \inf_{|z|>1+\delta/2} s_{mn}\left(\mathcal{Y}_n^{(k)} - zI\right) \geq c \right\} \tag{47}$$

holds with overwhelming probability. Finally, define the resolvent

$$\mathcal{G}_n^{(k)}(z) := \left(\mathcal{Y}_n^{(k)} - zI\right)^{-1}. \tag{48}$$

The following lemma follows from an application of Proposition 3.2.

**Lemma 6.3.** *Define the events $\Omega_n$ and $\Omega_{n,k}$ as in* (32) *and* (47) *respectively. Then there exist a constant $C > 0$ such that, for all $z \in \mathcal{C}$, $\|\mathcal{G}_n(z)\| \leq C$ surely on $\Omega_n$ and $\left\|\mathcal{G}_n^{(k)}(z)\right\| \leq C$ surely on $\Omega_{n,k}$.*

With these definitions, we may write

$$\begin{aligned}
\Xi_n(z) &= \operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n} - \mathbb{E}\left[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n}\right] \\
&= \sum_{k=1}^{n} \left(\mathbb{E}_k[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n}] - \mathbb{E}_{k-1}[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n}]\right) \\
&= \sum_{k=1}^{n} Z_{n,k}(z)
\end{aligned}$$

where we define

$$Z_{n,k}(z) := (\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n}]. \tag{49}$$

With this notation, we can rewrite the linear combination from (44) as

$$\sum_{l=1}^{L}(\alpha_l \Xi_n(z_l) + \beta_l \overline{\Xi_n(z_l)}) = \sum_{k=1}^{n}\sum_{l=1}^{L}\left(\alpha_l Z_{n,k}(z_l) + \beta_l \overline{Z_{n,k}(z_l)}\right).$$

Let $M_{n,k} := \sum_{l=1}^{L}\left(\alpha_l Z_{n,k}(z_l) + \beta_l \overline{Z_{n,k}(z_l)}\right)$ for any fixed integer $L > 0$, and $z_i \in \mathcal{C}$, and any $\alpha_i, \beta_i \in \mathbb{C}$ such that $M_{n,k}$ is real and denote

$$M_n := \sum_{k=1}^{n} M_{n,k}. \tag{50}$$

In order to simplify computations, it will be beneficial to work with a slightly different expression in which some reductions are made.

**Lemma 6.4.** *Define $M_n$ as in* (50)*, define $U_k$ to be the $mn \times m$ matrix which contains as its columns $c_k, c_{n+k}, \dots, c_{(m-1)n+k}$, and define $V_k$ to be the $mn \times m$ matrix which contains as its columns $e_k, e_{n+k}, \dots, e_{(m-1)n+k}$ where $e_1, \dots, e_{mn}$ denote the standard basis elements of $\mathbb{C}^{mn}$. Define the martingale difference sequence*

$$\breve{M}_n := \sum_{k=1}^{n} \breve{M}_{n,k} = \sum_{k=1}^{n}\left(\sum_{l=1}^{L} \alpha_l \breve{Z}_{n,k}(z_l) + \beta_l \overline{\breve{Z}_{n,k}(z_l)}\right),$$

*and*

$$\breve{Z}_{n,k}(z) := -\mathbb{E}_k\left[\operatorname{tr}\left(V_k^T(\mathcal{G}_n^{(k)}(z))^2 U_k\right)\mathbf{1}_{\Omega_{n,k}}\right]. \tag{51}$$

*Then, as $n \to \infty$, if $\breve{M}_n$ converges in distribution, then $M_n$ also converges to the same distributional limit.*

We first develop some results we will need in the proof of Lemma 6.4. Define the event

$$Q_{n,k}(z) := \left\{ \left\| V_k^T \mathcal{G}_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k}} \right\| \leq 1/2 \right\}. \tag{52}$$

We will also need the following lemma.

**Lemma 6.5.** *Define the event $Q_{n,k}(z)$ as in (52). Then, uniformly for any $z \in \mathcal{C}$, $Q_{n,k}$ holds with overwhelming probability*

*Proof.* Let $\alpha > 0$ be arbitrary. We will prove that the complementary event holds with probability at most $O_\alpha(n^{-\alpha})$ uniformly for any $z \in \mathcal{C}$. By Markov's inequality and the forthcoming Lemma 6.9, uniformly for any $z \in \mathcal{C}$ and for any $p \geq 2$,

$$\begin{aligned}\mathbb{P}\left( \left\| V_k^T \mathcal{G}_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k}} \right\| \geq 1/2 \right) &\leq \frac{\mathbb{E} \left\| V_k^T \mathcal{G}_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k}} \right\|^{2p}}{(1/2)^{2p}} \\ &\ll_p n^{-2\varepsilon p + 4\varepsilon - 2}.\end{aligned}$$

Selecting $p$ sufficiently large completes the proof. □

**Lemma 6.6.** *Let $A$ be an $mn \times mn$ Hermitian positive semidefinite matrix with rank at most $d$ for some positive constant $d$. Suppose that $\xi$ is a complex-valued random variable with mean zero, unit variance, $\mathbb{E}|\xi|^4 = O(1)$, and which satisfies $|\xi| \leq n^{1/2-\varepsilon}$ almost surely for some constant $0 < \varepsilon < 1/2$. Let $S \subseteq [mn]$, and let $w = (w_i)_{i=1}^{mn}$ be a vector with the following properties:*

- *(i) $\{w_i : i \in S\}$ is a collection of iid copies of $\xi$,*
- *(ii) $w_i = 0$ for $i \notin S$.*

*Then for any $p \geq 2$,*

$$\mathbb{E} \left| w^* A w \right|^p \ll_{d,p} n^{(1-2\varepsilon)p + 4\varepsilon - 2} \left\| A \right\|^p. \tag{53}$$

*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 C.3, we get

$$\begin{aligned}\mathbb{E} \left| w^* A w \right|^p &\ll_p (\operatorname{tr} A_{S\times S})^p + \mathbb{E}|\xi|^{2p} \operatorname{tr} A_{S\times S}^p \\ &= (\operatorname{tr} A_{S\times S})^p + \mathbb{E}\left[ |\xi|^4 |\xi|^{2p-4} \right] \operatorname{tr} A_{S\times S}^p \\ &\leq (\operatorname{tr} A_{S\times S})^p + n^{(1-2\varepsilon)p+4\varepsilon-2} \mathbb{E}|\xi|^4 \operatorname{tr} A_{S\times S}^p.\end{aligned}$$

Since the rank of $A_{S\times S}$ is at most $d$, $\operatorname{tr} A_{S\times S} \ll_d \|A\|$ and $\operatorname{tr} A_{S\times S}^p \ll_d \|A\|^p$, where we used the fact that the operator norm of a matrix bounds the operator norm of any sub-matrix. We conclude that

$$\mathbb{E} \left| w^* A w \right|^p \ll_{d,p} \|A\|^p + n^{(1-2\varepsilon)p+4\varepsilon-2} \mathbb{E}|\xi|^4 \|A\|^p \ll_{d,p} n^{(1-2\varepsilon)p+4\varepsilon-2} \mathbb{E}|\xi|^4 \|A\|^p,$$

as desired. □

**Lemma 6.7.** *Let $A$ be a deterministic complex $mn \times mn$ matrix for some fixed $m > 0$. Suppose that $\xi$ is a complex-valued random variable with mean zero, unit variance, finite moments of all orders. Let $S, R \subseteq [mn]$, and let $w = (w_i)_{i=1}^{mn}$ and $t = (t_i)_{i=1}^{mn}$ be independent vectors with the following properties:*

*(i) $\{w_i : i \in S\}$ and $\{t_j : j \in R\}$ are independent collections of iid copies of $\xi$,*
*(ii) $w_i = 0$ for $i \notin S$, and $t_j = 0$ for $j \notin R$.*

*Then for any $p \geq 1$,*

$$\mathbb{E}\,|w^* A t|^{2p} \ll_p \mathbb{E}|\xi|^{4p}(\mathrm{tr}(A^*A))^p. \tag{54}$$

*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 $i, j \in R$.

Since $w$ is independent of $t$, Lemma C.3 implies that

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

Recall that for any matrix $B$, $\mathrm{tr}(B^*B)^p \leq (\mathrm{tr}(B^*B))^p$. By this and the fact that for a Hermitian positive semidefinite matrix, the partial trace is less than or equal to the full trace, we observe that

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

By a cyclic permutation of the trace, we have

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

By Lemma C.3, and a similar argument as above, we have

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

and thus by Jensen's inequality, we have

$$\mathbb{E}|w^* A t|^{2p} \ll_p \mathbb{E}|\xi|^{2p}\mathbb{E}\left[(\mathrm{tr}(A t t^* A^*))^p\right] \ll_p \mathbb{E}|\xi|^{4p}(\mathrm{tr}(A^*A))^p$$

completing the proof. ☐

*Remark* 6.8. Note that if $p \geq 1$ and we also assume that $\mathbb{E}|\xi|^4 = O(1)$ and $|\xi| < n^{1/2-\varepsilon}$ surely for some $\varepsilon > 0$, then we may write

$$\begin{aligned}\mathbb{E}|w^* A t|^{2p} &\ll_p \mathbb{E}|\xi|^{4p}(\mathrm{tr}(A^*A))^p \\ &= \mathbb{E}\left[|\xi|^4|\xi|^{4p-4}\right](\mathrm{tr}(A^*A))^p \\ &\ll n^{(2-4\varepsilon)p+4\varepsilon-2}\mathbb{E}|\xi|^4(\mathrm{tr}(A^*A))^p.\end{aligned}$$

**Lemma 6.9.** *Let $U_k$ be the $mn \times m$ matrix which contains as its columns the columns $c_k, c_{n+k}, \ldots, c_{(m-1)n+k}$ of $\mathcal{Y}_n$ and define $V_k$ to be the $mn \times m$ matrix which contains as its columns $e_k, e_{n+k}, \ldots, e_{(m-1)n+k}$ where $e_1, \ldots, e_{mn}$ denote the standard basis elements of $\mathbb{C}^{mn}$. Let $\mathcal{G}_n^{(k)}(z)$ be defined as in (48). Then*

$$\mathbb{E}\left\|V_k^T \mathcal{G}_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k}}\right\|^2 \ll n^{-1}$$

*and for any $p \geq 2$,*

$$\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^{2p} \ll_p n^{-2\varepsilon p+4\varepsilon-2}.$$

*Proof.* Begin by observing that

$$\begin{aligned}
&\mathbb{E}\left\|V_k^T\mathcal{G}_n(z)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^{2p} \\
&\quad\ll \max_{1\leq i,j\leq m}\mathbb{E}\left|(V_k^T\mathcal{G}_n^{(k)}(z)U_k)_{(i,j)}\mathbf{1}_{\Omega_{n,k}}\right|^{2p} \\
&\quad= \max_{1\leq i,j\leq m}\mathbb{E}\left|e_{(i-1)n+k}\mathcal{G}_n^{(k)}(z)c_{(j-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right|^{2p}
\end{aligned}$$

In the case when $p=1$, since the rank of $(\mathcal{G}_n^{(k)}(z))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n^{(k)}(z)$ is at most 1, for any $1\leq j\leq m$ we have

$$\begin{aligned}
&\max_{1\leq i,j\leq m}\mathbb{E}\left|e_{(i-1)n+k}\mathcal{G}_n^{(k)}(z)c_{(j-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right|^{2p} \\
&\quad\ll \mathbb{E}\left[c_{(j-1)n+k}^*(\mathcal{G}_n^{(k)}(z))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n(z)c_{(j-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right] \\
&\quad\ll n^{-1}\left\|(\mathcal{G}_n^{(k)}(z))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n(z)\mathbf{1}_{\Omega_{n,k}}\right\| \\
&\quad\ll n^{-1}
\end{aligned}$$

by Lemma 6.3. In the case where $p\geq 2$, we have

$$\begin{aligned}
&\max_{1\leq i,j\leq m}\mathbb{E}\left|e_{(i-1)n+k}\mathcal{G}_n^{(k)}(z)c_{(j-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right|^{2p} \\
&\quad= \max_{1\leq i,j\leq m}\mathbb{E}\left|c_{(j-1)n+k}^*(\mathcal{G}_n^{(k)}(z))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n^{(k)}(z)c_{(j-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right|^{p}.
\end{aligned}$$

Note that by definition of $\mathcal{Y}_n$ in (31), each entry in $c_{(j-1)n+k}$ has been scaled by $n^{-1/2}$. By this observation, Lemma 6.3, and Lemma 6.6,

$$\begin{aligned}
&\mathbb{E}\left|c_{(j-1)n+k}^*(\mathcal{G}_n^{(k)}(z))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n^{(k)}(z)c_{(j-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right|^{p} \\
&\quad\ll_p n^{-p}n^{(1-2\varepsilon)p+4\varepsilon-2}\left\|(\mathcal{G}_n^{(k)}(z))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right\|^{p} \\
&\quad\ll n^{-2\varepsilon p+4\varepsilon-2}
\end{aligned}$$

for any $1\leq j\leq m$ since the rank of $(\mathcal{G}_n^{(k)}(z))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n^{(k)}(z)$ is at most 1. □

*Remark* 6.10. The same argument as in Lemma 6.9 also shows that $\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k\mathbf{1}_{\Omega_{n,k}}\right\|^2 \ll n^{-1}$ and $\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k\mathbf{1}_{\Omega_{n,k}}\right\|^{2p} \ll_p n^{-2\varepsilon p+4\varepsilon-2}$.

We now proceed with the proof of Lemma 6.4.

*Proof of Lemma 6.4.* To begin, note that the result will follow if we prove that $\mathbb{E}|M_n-\check{M}_n|^2=o(1)$. Since the only difference between these two expressions is the difference between $Z_{n,k}(z)$ and $\check{Z}_{n,k}(z)$, it will be sufficient to prove that for any

$z$ on the contour, $\mathbb{E}\left|\sum_{k=1}^{n}\left(Z_{n,k}(z)-\breve{Z}_{n,k}(z)\right)\right|^2 = o(1)$. Since $Z_{n,k}$ and $\breve{Z}_{n,k}$ are martingale difference sequences, we will prove

$$\mathbb{E}\left|Z_{n,k}(z)-\breve{Z}_{n,k}(z)\right|^2 = o\left(n^{-1}\right)$$

uniformly for any $0 < k \leq n$ and any $z$ on the contour $\mathcal{C}$. To do so, we will make a sequence of comparisons, each of which differs from the previous expression by error terms which is $o(n^{-1})$. To begin, observe that since $\mathcal{Y}_n^{(k)}$ has columns $k, n+k, \ldots, (m-1)n+k$ replaced with zeros, we have $(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}] = 0$. Thus, we can rewrite

$$\begin{aligned}
Z_{n,k}(z) &= (\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n}] \\
&= (\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n} - \operatorname{tr}\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}] \\
&= (\mathbb{E}_k - \mathbb{E}_{k-1})[(\operatorname{tr}\mathcal{G}_n(z) - \operatorname{tr}\mathcal{G}_n^{(k)}(z))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}] \\
&\quad + (\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n\cap\Omega_{n,k}^c}] \\
&\quad - (\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega_n^c}].
\end{aligned}$$

Note that, uniformly for $z$ with $|z| = 1+\delta$, by Lemma 6.3 and since $\Omega_{n,k}$ holds with overwhelming probability,

$$\begin{aligned}
\mathbb{E}\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n\cap\Omega_{n,k}^c}]\right|^2 &\ll \mathbb{E}\left|\operatorname{tr}\mathcal{G}_n(z)\mathbf{1}_{\Omega_n\cap\Omega_{n,k}^c}\right|^2 \\
&\ll n^2\mathbb{E}\left[\|\mathcal{G}_n(z)\|^2\mathbf{1}_{\Omega_n\cap\Omega_{n,k}^c}\right] \\
&\ll_\alpha n^{2-\alpha}
\end{aligned}$$

for any $\alpha > 0$. Since $\Omega_{n,k}$ holds with overwhelming probability, the same argument shows that $\mathbb{E}\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega_n^c}]\right|^2 \ll_\alpha n^{2-\alpha}$ for any $\alpha > 0$. Ergo, we have reduced from working with $Z_{n,k}(z)$ to working with $(\mathbb{E}_k - \mathbb{E}_{k-1})[(\operatorname{tr}\mathcal{G}_n(z) - \operatorname{tr}\mathcal{G}_n^{(k)}(z))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}]$. Next, observe that by linearity and cyclic permutation of the trace, and by the resolvent identity (12),

$$\begin{aligned}
\operatorname{tr}\mathcal{G}_n(z) - \operatorname{tr}\mathcal{G}_n^{(k)}(z) &= \operatorname{tr}\left(\mathcal{G}_n(z)\left(\mathcal{Y}_n^{(k)} - \mathcal{Y}_n\right)\mathcal{G}_n^{(k)}(z)\right) \\
&= -\operatorname{tr}\left(\mathcal{G}_n(z)U_kV_k^T\mathcal{G}_n^{(k)}(z)\right) \\
&= -\operatorname{tr}\left(V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n(z)U_k\right). \qquad (55)
\end{aligned}$$

To guarantee that $I_m + V_k^T\mathcal{G}_n^{(k)}U_k$ is invertible, we wish to work on the event $Q_{n,k}$ defined in (52). By Lemma 6.5, $Q_{n,k}$ hold with overwhelming probability so that by Lemma 6.3, the Cauchy–Schwarz inequality, and bounding the spectral norm by the Frobenius norm, we have

$$\begin{aligned}
&\mathbb{E}\Big|(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n(z)U_k)\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}] \\
&\qquad\qquad -(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n(z)U_k)\mathbf{1}_{\Omega_n\cap\Omega_{n,k}\cap Q_{n,k}}]\Big|^2 \\
&\quad \ll \mathbb{E}\left[\left\|V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n(z)U_k\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}\right\|^2\mathbf{1}_{Q_{n,k}^c}\right] \\
&\quad \ll n^4\mathbb{E}\left[\|U_k\|^2\mathbf{1}_{Q_{n,k}^c}\right] \\
&\quad \ll_\alpha n^{6-\alpha}
\end{aligned}$$

for any $\alpha > 0$. By selecting $\alpha$ sufficiently large, we can justify working with

$$-(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n(z)U_k)\mathbf{1}_{\Omega_n\cap\Omega_{n,k}\cap Q_{n,k}}]$$

instead of $Z_{n,k}(z)$. By the Sherman–Morrison–Woodbury formula (11), we have

$$\begin{aligned}
&-(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n(z)U_k)\mathbf{1}_{\Omega_n\cap\Omega_{n,k}\cap Q_{n,k}}] \\
&\quad = -(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k(I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1})\mathbf{1}_{\Omega_n\cap\Omega_{n,k}\cap Q_{n,k}}].
\end{aligned}$$

Since $\mathcal{G}_n(z)$ is no longer present, we may drop the event $\Omega_n$ gaining a sufficiently small error, and the same argument justifies working with

$$-(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k(I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1})\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}]$$

instead of $Z_{n,k}(z)$. At this point, we wish to replace $(I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1}$ with $I_m$. To justify this, observe that

$$\begin{aligned}
&\mathbb{E}\Big|(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k(I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1})\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}] \\
&\qquad\qquad -(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_kI_m)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}]\Big|^2 \\
&\quad \ll \mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k((I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1} - I_m)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^2. \qquad (56)
\end{aligned}$$

Note that by the resolvent identity (12),

$$(I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1} = I_m - (I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1}V_k^T\mathcal{G}_n^{(k)}(z)U_k. \qquad (57)$$

By iterating this twice, we have

$$(I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1} - I_m = -V_k^T\mathcal{G}_n^{(k)}(z)U_k + (I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1}(V_k^T\mathcal{G}_n^{(k)}(z)U_k)^2.$$

Inserting this into the last line of (56), we get

$$\begin{aligned}
&\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k((I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1} - I_m)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^2 \\
&\quad \ll \mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k(V_k^T\mathcal{G}_n^{(k)}(z)U_k)\mathbf{1}_{\Omega_{n,k}}\right\|^2 \qquad (58) \\
&\qquad + \mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k(I_m + V_k^T\mathcal{G}_n^{(k)}U_k)^{-1}(V_k^T\mathcal{G}_n^{(k)}(z)U_k)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^2. \qquad (59)
\end{aligned}$$

We will bound each of the above terms separately. First, we begin with term (58). Note that by Cauchy–Schwarz inequality, Lemma 6.9, and Remark 6.10, we have

$$\begin{aligned}
&\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k(V_k^T\mathcal{G}_n^{(k)}(z)U_k)\mathbf{1}_{\Omega_{n,k}}\right\|^2 \\
&\quad \ll \left(\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k\mathbf{1}_{\Omega_{n,k}}\right\|^4 \mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^4\right)^{1/2} \\
&\quad = o(n^{-1}).
\end{aligned}$$

It remains to show that term (59) is also $o(n^{-1})$. To this end, observe that by the Cauchy–Schwarz inequality, Lemma 6.9, and Remark 6.10,

$$\begin{aligned}
&\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k(I_m+V_k^T\mathcal{G}_n^{(k)}U_k)^{-1}(V_k^T\mathcal{G}_n^{(k)}(z)U_k)^2\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^2 \\
&\quad \leq \Bigg(\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^8 \mathbb{E}\left\|(I_m+V_k^T\mathcal{G}_n^{(k)}U_k)^{-1}\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^8 \\
&\qquad\qquad \times \mathbb{E}\left\|(V_k^T\mathcal{G}_n^{(k)}(z)U_k)^2\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^8\Bigg)^{1/4} \\
&\quad \ll \left(\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k\mathbf{1}_{\Omega_{n,k}}\right\|^8 \mathbb{E}\left\|(V_k^T\mathcal{G}_n^{(k)}(z)U_k)^2\mathbf{1}_{\Omega_{n,k}}\right\|^8\right)^{1/4} \\
&\quad \ll \left(\mathbb{E}\left\|V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k\mathbf{1}_{\Omega_{n,k}}\right\|^8 \mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^{16}\right)^{1/4} \\
&\quad \ll \left(n^{-2\varepsilon\cdot 4+4\varepsilon-2}\cdot n^{-2\varepsilon\cdot 8+4\varepsilon-2}\right)^{1/4} \\
&\quad = \left(n^{-16\varepsilon-4}\right)^{1/4} \\
&\quad \ll n^{-4\varepsilon-1}.
\end{aligned}$$

Since the above term is also $o(n^{-1})$, we may proceed working with the term

$$-(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}].$$

Next, we will justify removing the event $Q_{n,k}$. Observe that by Remark 6.10 and repeating the same argument as above,

$$\begin{aligned}
&\mathbb{E}\Big|(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k)\mathbf{1}_{\Omega_{n,k}}] \\
&\qquad -(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}]\Big|^2 \ll_\alpha n^{-1-\alpha/2}.
\end{aligned}$$

By selecting $\alpha$ sufficiently large in the above expression, we can proceed with

$$-(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k)\mathbf{1}_{\Omega_{n,k}}].$$

Finally, note that $U_k$ is independent of $\mathcal{G}_n^{(k)}(z)$ and $\Omega_{n,k}$, so that

$$\begin{aligned}
&\mathbb{E}_{k-1}[\operatorname{tr}(V_k^T(\mathcal{G}_n^{(k)}(z))^2U_k)\mathbf{1}_{\Omega_{n,k}}] \\
&\quad = \sum_{i=1}^m\sum_{a,b=1}^{mn}(V_k^T)_{i,a}\mathbb{E}_{k-1}\left[(\mathcal{G}_n^{(k)}(z))^2_{a,b}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}_{k-1}[(U_k)_{b,i}] \\
&\quad = 0.
\end{aligned}$$

This completes the proof. ∎

To prove that $\breve{M}_n$ converges to a mean-zero Gaussian, we will use the following martingale difference sequence central limit theorem.

**Theorem 6.11** (Theorem 35.12 of [15])**.** *For each $N$, suppose $Z_{N_1}, Z_{N_2}, \dots, Z_{N_{r_N}}$ is a real martingale difference sequence with respect to the increasing $\sigma$-field $\{\mathcal{F}_{N_j}\}$ having second moments. Suppose, for any $\eta > 0$ and a positive constant $\nu^2$,*

$$\lim_{N\to\infty} \mathbb{P}\left( \left| \sum_{j=1}^{r_N} \mathbb{E}\left( Z_{N_j}^2 | \mathcal{F}_{N_{j-1}} \right) - \nu^2 \right| > \eta \right) = 0 \tag{60}$$

*and*

$$\lim_{N\to\infty} \sum_{j=1}^{r_N} \mathbb{E}\left( Z_{N_j}^2 \mathbf{1}_{\{|Z_{N_j}| \geq \eta\}} \right) = 0. \tag{61}$$

*Then as $N \to \infty$, the distribution of $\sum_{j=1}^{r_N} Z_{N_j}$ converges weakly to a Gaussian distribution with mean zero and variance $\nu^2$.*

We will apply this result to $\{\breve{M}_{n,k}\}_{k=1}^n$ and the corresponding $\sigma$-algebras are $\{\mathcal{F}_k\}$. Verifying (60) for $\{\breve{M}_{n,k}\}_{k=1}^n$ is lengthy and will require new notation, so we begin with verifying (61) for $\{\breve{M}_{n,k}\}_{k=1}^n$. Let $\eta > 0$ and observe that by Remark 6.10, we have

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}\left[ \breve{M}_{n,k}^2 \mathbf{1}_{\{|\breve{M}_{n,k}| > \eta\}} \right] &\ll \sum_{k=1}^n \mathbb{E}\left[ \frac{\breve{M}_{n,k}^4}{\eta^2} \mathbf{1}_{\{|\breve{M}_{n,k}| > \eta\}} \right] \\ &\ll_{\eta,L} \sum_{k=1}^n \sum_{l=1}^L \mathbb{E}\left\| V_k^T (\mathcal{G}_n^{(k)}(z_l))^2 U_k \mathbf{1}_{\Omega_{n,k}} \right\|^4 \\ &\ll_{\eta,L} n^{-1}.\end{aligned}$$

Condition (60) will follow from the following lemma.

**Lemma 6.12.** *The martingale difference sequence*

$$\{\breve{M}_{n,k}\} = \left\{ \sum_{l=1}^L \alpha_l \breve{Z}_{n,k}(z_l) + \beta_l \overline{\breve{Z}_{n,k}(z_l)} \right\}$$

*has finite second moments and satisfies*

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}_{k-1}[\breve{M}_{n,k}^2] \to \sum_{1 \leq i,j \leq L} &\alpha_i \alpha_j \frac{m^2 (z_i z_j)^{m-1}}{((z_i z_j)^m - 1)^2} + \alpha_i \beta_j \frac{m^2 (z_i \bar{z}_j)^{m-1}}{((z_i \bar{z}_j)^m - 1)^2} \\ &+ \beta_i \alpha_j \frac{m^2 (\bar{z}_i z_j)^{m-1}}{((\bar{z}_i z_j)^m - 1)^2} + \beta_i \beta_j \frac{m^2 (\bar{z}_i \bar{z}_j)^{m-1}}{((\bar{z}_i \bar{z}_j)^m - 1)^2}\end{aligned} \tag{62}$$

*in probability as $n \to \infty$.*

To prove Lemma 6.12, we will need some definitions and results. We develop these now before proceeding to the proof.

Define $\mathcal{Y}_n^{(k,s)}$ to be the matrix $\mathcal{Y}_n$ with columns $c_k, c_{n+k}, \dots, c_{(m-1)n+k}$, and $c_s$ filled with zeros and define the resolvent $\mathcal{G}_n^{(k,s)}(z) := \left( \mathcal{Y}_n^{(k,s)} - zI \right)^{-1}$. By Corollary

B.6, for any $\delta > 0$ there exists a constant $c > 0$ depending only on $\delta$ such that the event

$$\Omega_{n,k,s} := \left\{ \inf_{|z|>1+\delta/2} s_{mn}\left(\mathcal{Y}_n^{(k,s)} - zI\right) \geq c \right\} \tag{63}$$

holds with overwhelming probability. By the Sherman-Morrison formula (10), provided $1 + e_s^T \mathcal{G}_n^{(k,s)}(z)c_s$ is not zero, we may write

$$\mathcal{G}_n^{(k)}(z)c_s = \frac{\left(\mathcal{Y}_n^{(k,s)} - zI\right)^{-1} c_s}{1 + e_s^T \left(\mathcal{Y}_n^{(k,s)} - zI\right)^{-1} c_s} = \mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z) \tag{64}$$

where

$$\delta_{k,s}(z) := (1 + e_s^T \mathcal{G}_n^{(k,s)} c_s)^{-1}. \tag{65}$$

By the same formula,

$$c_s^*(\mathcal{G}_n^{(k)}(w))^* = (\delta_{k,s}(w))^* c_s^*(\mathcal{G}_n^{(k,s)}(w))^*. \tag{66}$$

To ensure that these quantities exist, we introduce the event

$$Q'_{n,k,s}(z) := \left\{ \left| e_s^T \mathcal{G}_n^{(k,s)}(z)c_s \mathbf{1}_{\Omega_{n,k,s}} \right| \leq 1/2 \right\}. \tag{67}$$

**Lemma 6.13.** *Define the event $Q'_{n,k,s}(z)$ as in (67). Then uniformly for any $z \in \mathcal{C}$, $Q'_{n,k,s}(z)$ holds with overwhelming probability*

*Proof.* Let $\alpha > 0$ be arbitrary. We will show the complement event holds with probability at most $O_\alpha(n^{-\alpha})$ uniformly for any $z \in \mathcal{C}$. Observe that by Markov's inequality and by Lemma 6.6, uniformly for any $z \in \mathcal{C}$ and for any $p \geq 2$,

$$\begin{aligned}
&\mathbb{P}\left( \left| e_s^T \mathcal{G}_n^{(k,s)}(z)c_s \mathbf{1}_{\Omega_{n,k,s}} \right| \geq 1/2 \right) \\
&\quad \ll_p \mathbb{E}\left| e_s^T \mathcal{G}_n^{(k,s)}(z)c_s \mathbf{1}_{\Omega_{n,k,s}} \right|^{2p} \\
&\quad = \mathbb{E}\left| c_s^*(\mathcal{G}_n^{(k,s)}(z))^* e_s e_s^T \mathcal{G}_n^{(k,s)}(z)c_s \mathbf{1}_{\Omega_{n,k,s}} \right|^{p} \\
&\quad \ll_{p,\alpha} n^{-2\varepsilon p + 4\varepsilon - 2}.
\end{aligned}$$

Selecting $p$ sufficiently large concludes the proof. □

The next Lemma follows by an application of Proposition 3.2.

**Lemma 6.14.** *On the event $\Omega_{n,k,s}$, there exists a constant $C > 0$ such that $\left\|\mathcal{G}_n^{(k,s)}(z)\right\| \leq C$ almost surely uniformly for any $z$ on the contour $\mathcal{C}$. There exists a constant $C > 0$ such that $\left|\delta_{k,s}(z)\mathbf{1}_{Q'_{n,k,s}}\right| \leq C$ almost surely uniformly for any $z$ on the contour $\mathcal{C}$.*

With these definitions and results in hand, we proceed with the proof of Lemma 6.12. In the proof of Lemma 6.12, we make some reductions, each of which produces error terms which are sufficiently small in $L^2$-norm. In particular, the proof of Lemma 6.12 uses techniques of expanding using a resolvent identity and invoking Vitali's Theorem to get a self consistent equation, allowing us to solve for the variance. Unlike the proof for a single matrix (see [56, Lemma 3.2] or [51, Theorem 5.2]), we iterate the process $m$ times before recovering a system of self consistent equations.

*Proof of Lemma 6.12.* We may begin by expanding

$$\begin{aligned}\mathbb{E}_{k-1}[\breve{M}_{n,k}^2] &= \mathbb{E}_{k-1}\left[\left(\sum_{l=1}^{L}\alpha_l\breve{Z}_{n,k}(z_l)+\beta_l\overline{\breve{Z}_{n,k}(z_l)}\right)^2\right]\\ &= \mathbb{E}_{k-1}\left[\sum_{i,j=1}^{L}\alpha_i\alpha_j\breve{Z}_{n,k}(z_i)\breve{Z}_{n,k}(z_j)\right]+\mathbb{E}_{k-1}\left[\sum_{i,j=1}^{L}\alpha_i\beta_j\breve{Z}_{n,k}(z_i)\overline{\breve{Z}_{n,k}(z_j)}\right]\\ &\quad+\mathbb{E}_{k-1}\left[\sum_{i,j=1}^{L}\beta_i\alpha_j\overline{\breve{Z}_{n,k}(z_i)}\breve{Z}_{n,k}(z_j)\right]+\mathbb{E}_{k-1}\left[\sum_{i,j=1}^{L}\beta_i\beta_j\overline{\breve{Z}_{n,k}(z_i)\breve{Z}_{n,k}(z_j)}\right]\end{aligned}$$

where $\breve{Z}_{n,k}(z)$ was defined in (51), and therefore

$$\begin{aligned}&\sum_{k=1}^{n}\mathbb{E}_{k-1}\left[\breve{M}_{n,k}^2\right]\\ &\quad=\sum_{k=1}^{n}\sum_{i,j=1}^{L}\alpha_i\alpha_j\mathbb{E}_{k-1}\left[\breve{Z}_{n,k}(z_i)\breve{Z}_{n,k}(z_j)\right] &(68)\\ &\qquad+\sum_{k=1}^{n}\sum_{i,j=1}^{L}\alpha_i\beta_j\mathbb{E}_{k-1}\left[\breve{Z}_{n,k}(z_i)\overline{\breve{Z}_{n,k}(z_j)}\right] &(69)\\ &\qquad+\sum_{k=1}^{n}\sum_{i,j=1}^{L}\beta_i\alpha_j\mathbb{E}_{k-1}\left[\overline{\breve{Z}_{n,k}(z_i)}\breve{Z}_{n,k}(z_j)\right] &(70)\\ &\qquad+\sum_{k=1}^{n}\sum_{i,j=1}^{L}\beta_i\beta_j\mathbb{E}_{k-1}\left[\overline{\breve{Z}_{n,k}(z_i)\breve{Z}_{n,k}(z_j)}\right]. &(71)\end{aligned}$$

We analyze each of these terms separately. Note that since the entries in the matrix $\mathcal{Y}_n$ are real, $\overline{\breve{Z}_{n,k}(z_j)}=\breve{Z}_{n,k}(\overline{z_j})$ so the calculations for all terms will be the same. Therefore it suffices to show that

$$\sum_{k=1}^{n}\sum_{i,j=1}^{L}\alpha_i\beta_j\mathbb{E}_{k-1}\left[\breve{Z}_{n,k}(z)\overline{\breve{Z}_{n,k}(w)}\right]\to\sum_{1\leq i,j,\leq L}\alpha_i\beta_j\frac{m^2(z\bar{w})^{m-1}}{((z\bar{w})^m-1)^2}$$

in probability for fixed $z, w \in \mathcal{C}$. For now, we focus on the sum over $k$. Observe that

$$\sum_{k=1}^{n} \mathbb{E}_{k-1}\left[\breve{Z}_{n,k}(z)\overline{\breve{Z}_{n,k}(w)}\right] \tag{72}$$

$$= \sum_{k=1}^{n} \mathbb{E}_{k-1}\left[\mathbb{E}_k\left[\sum_{i=1}^{m}\sum_{a,b=1}^{mn}(V_k^T)_{(i,a)}(\mathcal{G}_n^{(k)}(z))^2_{(a,b)}(U_k)_{(b,i)}\mathbf{1}_{\Omega_{n,k}}\right]\right.$$
$$\left.\times\, \mathbb{E}_k\left[\sum_{j=1}^{m}\sum_{c,d=1}^{mn}(U_k^*)_{(j,d)}(\mathcal{G}_n^{(k)}(w))^{2*}_{(d,c)}(V_k)_{(c,j)}\mathbf{1}_{\Omega_{n,k}}\right]\right]$$
$$= \sum_{k=1}^{n}\sum_{i,j=1}^{m}\sum_{a,b,c,d=1}^{mn}\mathbb{E}_{k-1}\left[(V_k^T)_{(i,a)}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(z))^2_{(a,b)}\mathbf{1}_{\Omega_{n,k}}\right](U_k)_{(b,i)}\right.$$
$$\left.\times\,(U_k^*)_{(j,d)}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{2*}_{(d,c)}\mathbf{1}_{\Omega_{n,k}}\right](V_k)_{(c,j)}\right]. \tag{73}$$

At this point, we may exploit the block structure of these matrices in order to reduce the number of terms in the above sums. First, since $V_k$ is a $mn \times m$ matrix which contains columns $e_k, e_{n+k} \dots e_{(m-1)n+k}$, $(V_k^T)_{(i,a)} = 0$ unless $a = (i-1)n+k$. The same argument shows that $(V_k)_{(c,j)} = 0$ unless $c = (j-1)n + k$. Since $U_k$ is independent of $\mathcal{G}_n^{(k)}(z)$, we can factor this out of the expectation and rewrite (73) as

$$\sum_{k=1}^{n}\sum_{i,j=1}^{m}\sum_{b,d=1}^{mn}\mathbb{E}_{k-1}\left[\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(z))^2_{((i-1)n+k,b)}\mathbf{1}_{\Omega_{n,k}}\right]\right.$$
$$\left.\times\,\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{2*}_{(d,(j-1)n+k)}\mathbf{1}_{\Omega_{n,k}}\right]\right]\mathbb{E}[(U_k)_{(b,i)}(U_k^*)_{(j,d)}].$$

Now, if $i \neq j$ or $b \neq d$, then $(U_k)_{(b,i)}$ and $(U_k^*)_{(j,d)}$ come from different columns or are different entries in the same column of $\mathcal{Y}_n$ and hence are independent. Therefore, the only non-zero terms are those in which $i = j$ and $b = d$. Thus the sum in (73) can be further reduced to

$$\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{b=1}^{mn}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(z))^2_{((i-1)n+k,b)}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{2*}_{(b,(i-1)n+k)}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}\left|(U_k)_{(b,i)}\right|^2.$$

Now, since $U_k$ is filled with columns $c_k, c_{n+k}, \dots c_{(m-1)n+k}$, we may analyze the structure of these columns to evaluate $\mathbb{E}\left|(U_k)_{(b,i)}\right|^2$. Since column $c_{(i-1)n+k}$ comes from the $i$th block of $\mathcal{Y}_n$, $(U_k)_{(b,i)} = 0$ unless $(i-2)n-1 \leq b \leq (i-1)n$ where these subscripts are reduced modulo $m$ and we use the convention that $-1n \equiv (m-1)n$ and $0n \equiv mn$. For $b$ in such a range, we have $\mathbb{E}|U_{(b,i)}|^2 = \frac{1}{n}$. Therefore we can simplify the sum in (73) further as

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{b=(i-2)n+1}^{(i-1)n}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(z))^2_{((i-1)n+k,b)}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{2*}_{(b,(i-1)n+k)}\mathbf{1}_{\Omega_{n,k}}\right].$$

Define the diagonal $mn \times mn$ matrix $\mathcal{D}_p$ with entries

$$(\mathcal{D}_p)_{(i,j)} := \begin{cases} 1 & \text{if } i = j,\ (p-1)n + 1 \leq i \leq pn \\ 0 & \text{otherwise} \end{cases} \tag{74}$$

for $1 \leq p \leq m$ and $1 \leq i, j \leq mn$. Note that $\mathcal{D}_p$ is nonzero only on the diagonal of the $p$th block. Then we have

$$\begin{aligned} &\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{b=(i-2)n+1}^{(i-1)n} \mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(z))^2_{((i-1)n+k,b)}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{2*}_{(b,(i-1)n+k)}\mathbf{1}_{\Omega_{n,k}}\right] \\ &= \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m} e^T_{(i-1)n+k}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(z))^2\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{2*}\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k} \end{aligned} \tag{75}$$

where the subscript on $\mathcal{D}_{i-1}$ is reduced modulo $m$ and in the range $\{1, \dots, m\}$. Next, if we can show that

$$\begin{aligned} &\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m} e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k} \\ &\qquad\longrightarrow -\ln\left(1 - \frac{1}{(z\bar{w})^m}\right), \end{aligned} \tag{76}$$

in probability as $n \to \infty$, then by Vitali's theorem (see for instance [13, Lemma 2.14]), it will follow that

$$\begin{aligned} &\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m} e^T_{(i-1)n+k}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(z))^2\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{2*}\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k} \\ &\qquad\longrightarrow \frac{m^2(z\bar{w})^{m-1}}{((z\bar{w})^m - 1)^2}. \end{aligned} \tag{77}$$

Vitali's theorem is justified because (76) is bounded and analytic in the region where $|z|, |w| > 1 + \delta/2$ and this region has an accumulation point. Note that here we apply Vitali's theorem twice, once in the variable $z$ and once in the variable $\bar{w}$.

To analyze the limit of (76), we will focus on a fixed term in the sum. Define

$$\mathcal{T}_{n,k}(z, \overline{w}) := \sum_{i=1}^{m} e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}. \tag{78}$$

Provided the resolvent is defined, we have the matrix identity (see for example, [56, Equation (3.15)]),

$$\mathcal{G}_n^{(k)}(z) = -\frac{1}{z}I + \frac{1}{z}\sum_{t \neq *n+k} \mathcal{G}_n^{(k)}(z) c_t e_t^T \tag{79}$$

where the notation $t \neq *n + k$ indicates that the sum is over all $1 \leq t \leq mn$ such that $t \neq k, n+k, \dots, (m-1)n+k$. We may use this to expand the term $\mathcal{T}_{n,k}(z, \overline{w})$. If we expand a generic term in the sum (78), we can show the following lemma holds.

**Lemma 6.15.** *Define all quantities as in Lemma 6.12. Let $b$ be fixed with $1 \leq b < m$. Then under the assumptions of Lemma 6.12,*

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m}\Big(e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\right.$$
$$-\frac{1}{z\overline{w}}\frac{k-1}{n}e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}$$
$$\left.\times\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\Big)\right|^2 = o(1)$$

*where $i-b$ and $i-b-1$ are reduced modulo $m$ with representatives in $\{1,\dots,m\}$.*

The use of Lemma 6.15 is one of the main technical components of this paper. This lemma allows us to iterate the techniques used in previous results (see for example [51, 56]) which results in a system of self consistent equations. Due to the iterative process used in the proof, the two terms in the difference in Lemma 6.15 differ from $\mathcal{T}_{n,k}(z,\overline{w})$ defined in (78) because of the subscripts on $\mathcal{D}_p$. We prove Lemma 6.15 now, but several of the technical calculations are done separately for clarity. These calculations are presented in lemmas at the end of Section 6.

*Proof of Lemma 6.15.* To begin, we may use the matrix identity from equation (79) to expand $e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}$. Doing so, we have that

$$e_{(i-1)n+k}^T\mathbb{E}_k[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}]\mathcal{D}_{i-b}\mathbb{E}_k[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}]e_{(i-1)n+k}$$
$$= e_{(i-1)n+k}^T\mathcal{D}_{i-b}\frac{1}{z\overline{w}}(\mathbb{P}_k(\Omega_{n,k}))^2e_{(i-1)n+k} \tag{80}$$
$$-e_{(i-1)n+k}^T\frac{1}{z\overline{w}}\sum_{s\neq *n+k}\mathcal{D}_{i-b}\mathbb{E}_k\left[e_sc_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{P}_k(\Omega_{n,k})e_{(i-1)n+k} \tag{81}$$
$$-e_{(i-1)n+k}^T\frac{1}{z\overline{w}}\sum_{t\neq *n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_te_t^T\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b}\mathbb{P}_k(\Omega_{n,k})e_{(i-1)n+k} \tag{82}$$
$$+e_{(i-1)n+k}^T\frac{1}{z\overline{w}}\sum_{s\neq *n+k}\sum_{t\neq *n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_te_t^T\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b}$$
$$\times\mathbb{E}_k\left[e_sc_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}. \tag{83}$$

where $\mathbb{P}_k$ denotes the conditional probability with respect to $\mathcal{F}_k$ and we assume that the subscript on $\mathcal{D}_{i-b}$ is reduced modulo $m$. We simplify each term separately. For any $1 \leq i \leq m$, assuming $b \leq m$, term (80) equals zero since the only nonzero elements in $\mathcal{D}_{i-b}$ are in block $i-b$, and (80) selects an element from the $i$th block. For terms (81) and (82), since $s \neq jn+k$ for any $0 \leq j \leq m-1$, we have the expression $e_{(i-1)n+k}^T\mathcal{D}_{i-b}e_s$, which results in an off diagonal element of $\mathcal{D}_{i-b}$. Ergo,

$$-e_{(i-1)n+k}^T\frac{1}{z\overline{w}}\sum_{s\neq *n+k}\mathcal{D}_{i-b}\mathbb{E}_k\left[e_sc_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{P}_k(\Omega_{n,k})e_{(i-1)n+k}$$
$$=-\frac{1}{z\overline{w}}\sum_{s\neq *n+k}e_{(i-1)n+k}^T\mathcal{D}_{i-b}e_s\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{P}_k(\Omega_{n,k})e_{(i-1)n+k}$$
$$=0$$

since $(\mathcal{D}_{i-b})_{f,g} = 0$ unless $(i-b-1)n+1 \leq f, g \leq (i-b)n$ and $f = g$. Similarly, since $t \neq jn+k$ for any $1 \leq j \leq m$, we have $e_t^T \mathcal{D}_{i-b} e_{(i-1)n+k} = 0$. Thus terms (81) and (82) are zero. Note that $e_t^T \mathcal{D}_{i-b} e_s = (\mathcal{D}_{i-b})_{(t,s)}$. This is zero unless $t = s$ and $(i-b-1)n+1 \leq s \leq (i-b)n$. Therefore term (83) can be simplified to

$$e_{(i-1)n+k}^T \frac{1}{z\bar{w}} \sum_{s \neq *n+k} \sum_{t \neq *n+k} \mathbb{E}_k \left[ \mathcal{G}_n^{(k)}(z) c_t e_t^T \mathbf{1}_{\Omega_{n,k}} \right] \mathcal{D}_{i-b} \mathbb{E}_k \left[ e_s c_s^* (\mathcal{G}_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] e_{(i-1)n+k}$$

$$= e_{(i-1)n+k}^T \frac{1}{z\bar{w}} \sum_{\substack{s=(i-b-1)n+1 \\ s \neq *n+k}}^{(i-b)n} \mathbb{E}_k \left[ \mathcal{G}_n^{(k)}(z) c_s \mathbf{1}_{\Omega_{n,k}} \right] \mathbb{E}_k \left[ c_s^* (\mathcal{G}_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] e_{(i-1)n+k}.$$

Ergo, we have

$$e_{(i-1)n+k}^T \mathbb{E}_k[\mathcal{G}_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}}] \mathcal{D}_{i-b} \mathbb{E}_k[(\mathcal{G}_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}}] e_{(i-1)n+k}$$
$$= \frac{1}{z\bar{w}} \sum_{\substack{s=(i-b-1)n+1, \\ s \neq *n+k}}^{(i-b)n} e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k)}(z) c_s \mathbf{1}_{\Omega_{n,k}} \right] \mathbb{E}_k \left[ c_s^* (\mathcal{G}_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] e_{(i-1)n+k}. \tag{84}$$

We now remove the $s$th column from the resolvent. In order to remove this column, we need to work on the appropriate events. Since $\Omega_{n,k,s}$ defined in (63) holds with overwhelming probability by Corollary B.6, we may insert the event with as sufficiently small $L^2$ norm error. This is verified in Lemma 6.17. Since Lemma 6.13 proves that $Q'_{n,k,s}(z)$ (defined in (67)) also holds with overwhelming probability, a very similar argument shows this event can be inserted as well. For ease of notation, we will drop the dependence on $z$ in $Q'_{n,k,s}(z)$ and write $Q'_{n,k,s}$. We proceed with

$$e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k)}(z) c_s \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] \times \mathbb{E}_k \left[ c_s^* (\mathcal{G}_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e_{(i-1)n+k}.$$

Then by (64) and (66), we have

$$e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k)}(z) c_s \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] \times \mathbb{E}_k \left[ c_s^* (\mathcal{G}_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e_{(i-1)n+k}$$
$$= e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k,s)}(z) c_s \delta_{k,s}(z) \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] \times \mathbb{E}_k \left[ (\delta_{k,s}(w))^* c_s^* (\mathcal{G}_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e_{(i-1)n+k}.$$

Since $G_n^{(k)}(z)$ is no longer present in the expression, the same argument as above shows that we can now remove the event $\Omega_{n,k}$ and work with

$$e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k,s)}(z) c_s \delta_{k,s}(z) \mathbf{1}_{\Omega_{n,k,s} \cap Q'_{n,k,s}} \right] \times \mathbb{E}_k \left[ (\delta_{k,s}(w))^* c_s^* (\mathcal{G}_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e_{(i-1)n+k}$$

with a sufficiently small $L^2$-norm error. Next, we wish to replace $\delta_{k,s}(z)$ and $(\delta_{k,s}(w))^*$ with 1. Observe that

$$\begin{aligned}
&\mathbb{E}\Bigg|\frac{1}{n}\sum_{k=1}^{n}\frac{1}{z\bar{w}}\sum_{i=1}^{m}\sum_{\substack{s=(i-b-1)n+1,\\ s\neq *n+k}}^{(i-b)n} e_{(i-1)n+k}^{T}\Big(\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad \times\ \mathbb{E}_k\left[(\delta_{k,s}(w))^* c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\\
&\quad-\ \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\Big)e_{(i-1)n+k}\Bigg|^2\\
&\ll \max_{\substack{1\leq k\leq n\\ 1\leq i\leq m}}\mathbb{E}\Bigg|\sum_{\substack{s=(i-b-1)n+1,\\ s\neq *n+k}}^{(i-b)n} e_{(i-1)n+k}^{T}\Big(\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad \times\ \mathbb{E}_k\left[(\delta_{k,s}(w))^* c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\\
&\quad-\ \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\Big)e_{(i-1)n+k}\Bigg|^2.
\end{aligned}$$

Therefore, it is sufficient to show that

$$\begin{aligned}
&\mathbb{E}\Big|e_{(i-1)n+k}^{T}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad \times\mathbb{E}_k\left[(\delta_{k,s}(w))^* c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\\
&\quad -e_{(i-1)n+k}^{T}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad \times\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\Big|^2 = o(n^{-2})
\end{aligned}$$

uniformly in $i$, $k$, and $s$. This is done in Lemma 6.18. Since $\delta_{n,k}$ is no longer present, we can justify dropping the event $Q'_{n,k,s}$ by an argument similar to Lemma 6.17. Thus, we can continue from here working with

$$\frac{1}{z\bar{w}}\sum_{i=1}^{m}\sum_{\substack{s=(i-b-1)n+1,\\ s\neq *n+k}}^{(i-b)n} e_{(i-1)n+k}^{T}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\mathbf{1}_{\Omega_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}.$$

Next, for any $1\leq i\leq m$, by independence of $c_s$ from $\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}$, we can factor the above as

$$\begin{aligned}
&\sum_{\substack{s=(i-b-1)n+1,\\ s\neq *n+k}}^{(i-b)n} e_{(i-1)n+k}^{T}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\mathbf{1}_{\Omega_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\\
&\quad=\sum_{\substack{s=(i-b-1)n+1,\\ s\neq *n+k}}^{(i-b)n} e_{(i-1)n+k}^{T}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathbb{E}_k[c_s]\\
&\qquad\qquad \times\mathbb{E}_k[c_s^*]\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}. \qquad (85)
\end{aligned}$$

Observe that the value of $\mathbb{E}_k[c_s]$ depends on whether or not the column $c_s$ has been conditioned on. If the column has been conditioned on, then the expectation returns the column itself and otherwise the expectation is zero. Since $\mathbb{E}_k[\cdot]$ conditions on the first $k$ columns in each block, we can simplify (85) to

$$\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1} e_{(i-1)n+k}^T \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right] c_s c_s^* \mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right] e_{(i-1)n+k}.$$

Now, consider $c_s c_s^*$. Based on the block structure of $\mathcal{Y}_n$, if $(i-b-1)n+1 \leq s < (i-b-1)n+k$, then we have

$$\mathbb{E}[(c_s c_s^*)_{(f,g)}] = \begin{cases} \frac{1}{n} & \text{if } (i-b-2)n+1 \leq f = g \leq (i-b-1)n \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, for a fixed term $i$ with $1 \leq i \leq m$ and since $(i-b-1)n+1 \leq s < (i-b-1)n+k$, we have $\mathbb{E}\left[c_s c_s^*\right] = \frac{1}{n}\mathcal{D}_{i-b-1}$. We wish now to replace $c_s c_s^*$ with its expectation. Observe that the terms

$$e_{(i-1)n+k}^T \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\left(c_s c_s^* - \frac{1}{n}\mathcal{D}_{i-b-1}\right) \times \mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right] e_{(i-1)n+k}$$

satisfy the conditions of a martingale difference sequence in $s$. Therefore we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^{n}\frac{1}{z\overline{w}}\sum_{i=1}^{m}\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1} e_{(i-1)n+k}^T \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\right.\\
&\qquad\qquad\left.\times\left(c_s c_s^* - \frac{1}{n}\mathcal{D}_{i-b-1}\right)\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right] e_{(i-1)n+k}\right|^2\\
&\ll \max_{\substack{1\leq k\leq n\\ 1\leq i\leq m}} \sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1} \mathbb{E}\left|e_{(i-1)n+k}^T \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\right.\\
&\qquad\qquad\left.\times\left(c_s c_s^* - \frac{1}{n}\mathcal{D}_{i-b-1}\right)\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right] e_{(i-1)n+k}\right|^2.
\end{aligned}$$

By Lemma 6.20,

$$\begin{aligned}
&\mathbb{E}\left|e_{(i-1)n+k}^T \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\left(c_s c_s^* - \frac{1}{n}\mathcal{D}_{i-b-1}\right)\right.\\
&\qquad\qquad\left.\times\,\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right] e_{(i-1)n+k}\right|^2 = o(n^{-1})
\end{aligned}$$

uniformly in $i$, $k$, and $s$, and therefore we may proceed with

$$\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1} e_{(i-1)n+k}^T \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\frac{1}{n}\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right] e_{(i-1)n+k}.$$

Next, we wish to add back in column $c_s$ to $\mathcal{G}_n^{(k,s)}$ and $\Omega_{n,k,s}$. Since the events $\Omega_{n,k}$ and $\Omega_{n,k,s}$ both hold with overwhelming probability, an argument similar to Lemma 6.17 shows that we can insert or drop these events with a sufficiently small

$L^2$-norm error. This is achieved by showing

$$\mathbb{E}\left|\frac{1}{n}\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1} e_{(i-1)n+k}^T\left(\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]\right.\right.$$
$$\left.\left.-\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]\right)e_{(i-1)n+k}\right|^2=o(1),$$

which is done in Lemma 6.21. Since

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m}\frac{1}{n}\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1}\left(e_{(i-1)n+k}^T\left(\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]\right.\right.\right.$$
$$\left.\left.\left.-\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]\right)e_{(i-1)n+k}\right)\right|^2=o(1),$$

column $c_s$ may be reinserted. With this column replaced, in each term we now have

$$\begin{aligned}\frac{1}{n}\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1} & e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\\ &\times\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\\ =\frac{k-1}{n}& e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\end{aligned}$$

since there were $k-1$ terms in the above sum, and none of them depended on $s$. This concludes the proof of Lemma 6.15. □

With the proof of Lemma 6.15 complete, we continue with the proof of Lemma 6.12. Applying Lemma 6.15 in the base when $b=1$ gives

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^{n}\left(\mathcal{T}_{n,k}(z,\overline{w})-\frac{1}{z\bar{w}}\frac{k-1}{n}\sum_{i=1}^{m}e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-2}\right.\right.$$
$$\left.\left.\times\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\right)\right|^2=o(1).$$

The goal is to iterate this process until $\mathcal{T}_{n,k}(z,\overline{w})$ reappears. Lemma 6.15 verifies that, for any $b\neq m$,

$$\begin{aligned}\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{m}\right.&\left(e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\right.\\&-\frac{1}{z\bar{w}}\frac{k-1}{n}e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-(b+1)}\\&\left.\left.\times\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\right)\right|^2=o(1)\end{aligned}$$

where $i-b$ is reduced modulo $m$. After iterating twice, we have

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^{n}\left(\mathcal{T}_{n,k}(z,\overline{w})-\left(\frac{1}{z\bar{w}}\frac{k-1}{n}\right)^{2}\sum_{i=1}^{m}e_{(i-1)n+k}^{T}\mathbb{E}_{k}\left[\mathcal{G}_{n}^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-3}\right.\right.$$
$$\left.\left.\times\,\mathbb{E}_{k}\left[(\mathcal{G}_{n}^{(k)}(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\right)\right|^{2}=o(1).$$

After iterating $m-1$ times, we have

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^{n}\left(\mathcal{T}_{n,k}(z,\overline{w})-\left(\frac{1}{z\bar{w}}\frac{k-1}{n}\right)^{m-1}\sum_{i=1}^{m}e_{(i-1)n+k}^{T}\mathbb{E}_{k}\left[\mathcal{G}_{n}^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-m}\right.\right.$$
$$\left.\left.\times\,\mathbb{E}_{k}\left[(\mathcal{G}_{n}^{(k)}(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\right)\right|^{2}=o(1).$$

To recover $\mathcal{T}_{n,k}(z,\overline{w})$, we will iterate one final time. Due to the block structure of $\mathcal{D}_p$ for $1\le p\le m$, the $m$th iteration will result in less cancellation than in previous iterations. Consider the expansion due to (79),

$$\begin{aligned}
&e_{(i-1)n+k}^{T}\mathbb{E}_{k}\left[\mathcal{G}_{n}^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-m}\mathbb{E}_{k}\left[(\mathcal{G}_{n}^{(k)}(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\\
&=e_{(i-1)n+k}^{T}\mathcal{D}_{i-m}\frac{1}{z\bar{w}}(\mathbb{P}_{k}(\Omega_{n,k}))^{2}e_{(i-1)n+k}\\
&\quad-e_{(i-1)n+k}^{T}\frac{1}{z\overline{w}}\sum_{s\neq *n+k}\mathcal{D}_{i-m}\mathbb{E}_{k}\left[e_{s}c_{s}^{*}(\mathcal{G}_{n}^{(k)}(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{P}_{k}(\Omega_{n,k})e_{(i-1)n+k}\\
&\quad-e_{(i-1)n+k}^{T}\frac{1}{z\overline{w}}\sum_{t\neq *n+k}\mathbb{E}_{k}\left[\mathcal{G}_{n}^{(k)}(z)c_{t}e_{t}^{T}\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-m}\mathbb{P}_{k}(\Omega_{n,k})e_{(i-1)n+k}\\
&\quad+e_{(i-1)n+k}^{T}\frac{1}{z\overline{w}}\sum_{s\neq *n+k}\sum_{t\neq *n+k}\mathbb{E}_{k}\left[\mathcal{G}_{n}^{(k)}(z)c_{t}e_{t}^{T}\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-m}\\
&\qquad\qquad\times\,\mathbb{E}_{k}\left[e_{s}c_{s}^{*}(\mathcal{G}_{n}^{(k)}(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}
\end{aligned}$$

where the notation $t\neq *n+k$ indicates that the sum is over all $1\le t\le mn$ such that $t\neq k,n+k,\dots,(m-1)n+k$ and $\mathbb{P}_k$ denotes the conditional probability on the $\sigma$-algebra $\mathcal{F}_k$. We have

$$e_{(i-1)n+k}^{T}\mathcal{D}_{i-m}\frac{1}{z\bar{w}}(\mathbb{P}_{k}(\Omega_{n,k}))^{2}e_{(i-1)n+k}=(\mathcal{D}_{i-m})_{((i-1)n+k,(i-1)n+k)}\frac{1}{z\bar{w}}(\mathbb{P}_{k}(\Omega_{n,k}))^{2}.$$

Note that after reducing modulo $m$, $\mathcal{D}_{i-m}$ is nonzero in the $(i-1)$st block. Therefore

$$e_{(i-1)n+k}^{T}\mathcal{D}_{i-m}\frac{1}{z\bar{w}}(\mathbb{P}_{k}(\Omega_{n,k}))^{2}e_{(i-1)n+k}=\frac{1}{z\bar{w}}(\mathbb{P}_{k}(\Omega_{n,k}))^{2}.$$

By arguments similar to those used in the proof of Lemma 6.15, we can calculate that

$$e_{(i-1)n+k}^{T}\frac{1}{z\overline{w}}\sum_{s\neq *n+k}\mathcal{D}_{i-m}\mathbb{E}_{k}\left[e_{s}c_{s}^{*}(\mathcal{G}_{n}^{(k)}(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{P}_{k}(\Omega_{n,k})e_{(i-1)n+k}=0$$

and

$$e_{(i-1)n+k}^{T}\frac{1}{z\overline{w}}\sum_{t\neq *n+k}\mathbb{E}_{k}\left[\mathcal{G}_{n}^{(k)}(z)c_{t}e_{t}^{T}\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-m}\mathbb{P}_{k}(\Omega_{n,k})e_{(i-1)n+k}=0.$$

By the forthcoming Lemma 6.19, we can see that $\mathbb{E}\left|(z\bar{w})^{-1}(1-\mathbb{P}_k(\Omega_{n,k})^2)\right| = o_\alpha(n^{-\alpha})$ for any $\alpha > 0$, and therefore we have

$$\mathbb{E}\Bigg|e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-m}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k} - \frac{1}{z\bar{w}} - \frac{1}{z\bar{w}}e^T_{(i-1)n+k}\sum_{\substack{s=(i-1)n+1,\\ s\neq *n+k}}^{in}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}}\right] \times\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\Bigg|^2 = o_\alpha(n^{-\alpha}).$$

By the same argument as in Lemma 6.15, we have

$$\mathbb{E}\Bigg|\frac{1}{n}\sum_{k=1}^n\sum_{i=1}^m\Bigg(e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-m}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k} - \frac{1}{z\bar{w}} - \frac{1}{z\bar{w}}\frac{k-1}{n}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-1} \times\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\Bigg)\Bigg|^2 = o(1).$$

By recognizing that we have recovered $\mathcal{T}_{n,k}(z,\bar{w})$ in the previous expression, and by putting this together with the previous iterations of the process, we in total get

$$\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n\left(\mathcal{T}_{n,k}(z,\overline{w}) - \frac{m}{(z\bar{w})^m}\left(\frac{k-1}{n}\right)^{m-1} - \left(\frac{1}{z\bar{w}}\frac{k-1}{n}\right)^m\mathcal{T}_{n,k}(z,\bar{w})\right)\right|^2 = o(1). \tag{86}$$

The goal now is to regroup in order to compare the object of study, $\frac{1}{n}\sum_{k=1}^n\mathcal{T}_{n,k}(z,\bar{w})$, to an appropriate Riemann sum. The sum we will compare to is

$$\frac{1}{n}\sum_{k=1}^n\frac{m(k-1)^{m-1}}{n^{m-1}}\left((z\bar{w})^m - \left(\frac{k-1}{n}\right)^m\right)^{-1}. \tag{87}$$

As $n\to\infty$, (87) is the Riemann sum for $\int_0^1 mx^{m-1}\left((z\bar{w})^m - x^m\right)^{-1}dx$ which, by a substitution of variables, is equal to $-\ln\left(1-\frac{1}{(z\bar{w})^m}\right)$. By regrouping the quantities inside the sum in (86), we can write

$$\mathcal{T}_{n,k}(z,\bar{w})\left(1-\frac{(k-1)^m}{n^m(z\bar{w})^m}\right) = \frac{m}{(z\bar{w})^m}\left(\frac{k-1}{n}\right)^{m-1} + \mathcal{E}_{n,k}(z,\bar{w})$$

where $\mathcal{E}_{n,k}(z,\bar{w})$ is an error term which satisfies $\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n\mathcal{E}_{n,k}(z,\bar{w})\right|^2 = o(1)$. This implies

$$\mathcal{T}_{n,k}(z,\bar{w})\left(\frac{n^m(z\bar{w})^m-(k-1)^m}{n^m(z\bar{w})^m}\right) = \frac{m}{(z\bar{w})^m}\left(\frac{k-1}{n}\right)^{m-1} + \mathcal{E}_{n,k}(z,\bar{w})$$

and thus

$$\mathcal{T}_{n,k}(z,\bar{w}) = \frac{mn(k-1)^{m-1}}{n^m(z\bar{w})^m-(k-1)^m} + \left(\frac{n^m(z\bar{w})^m}{n^m(z\bar{w})^m-(k-1)^m}\right)\times\mathcal{E}_{n,k}(z,\bar{w}). \tag{88}$$

We are now ready to compare $\frac{1}{n}\sum_{k=1}^n \mathcal{T}_{n,k}$ to the Riemann sum in (87). By rearranging (88), we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n \mathcal{T}_{n,k}(z\bar{w}) - \frac{1}{n}\sum_{k=1}^n \frac{m(k-1)^{m-1}}{n^{m-1}}\left((z\bar{w})^m - \left(\frac{k-1}{n}\right)^m\right)^{-1}\right|^2 \\
&\ll \mathbb{E}\left|\frac{1}{n}\sum_{k=1}^n \mathcal{E}_{n,k}(z,\bar{w})\right|^2 \\
&= o(1).
\end{aligned}$$

Therefore $\frac{1}{n}\sum_{k=1}^n \mathcal{T}_{n,k}(z,\bar{w})$ converges to $-\ln\left(1-\frac{1}{(z\bar{w})^m}\right)$ in probability as $n\to\infty$ as claimed in (76). By recalling that we invoked Vitali's theorem, this implies

$$\sum_{k=1}^n \alpha_i\beta_j\mathbb{E}_k\left[\breve{Z}_{n,k}(z)\overline{\breve{Z}_{n,k}(w)}\right] \to \alpha_i\beta_j\frac{m^2(z\bar{w})^{m-1}}{((z\bar{w})^m-1)^2}$$

in probability as $n\to\infty$. This concludes the proof of Lemma 6.12. □

*Remark* 6.16. In the case where the atom variables are complex-valued, the limit of $\sum_{k=1}^n \mathbb{E}_{k-1}[\breve{M}_{n,k}^2]$ would differ from that of Lemma 6.12. In this case, the calculations for terms of the form $\beta_i\alpha_j\mathbb{E}_k\left[\overline{\breve{Z}_{n,k}(z_i)}\breve{Z}_{n,k}(z_j)\right]$ would differ from those of the form $\beta_i\alpha_j\mathbb{E}_k\left[\breve{Z}_{n,k}(z_i)\breve{Z}_{n,k}(z_j)\right]$ due to the fact that the conjugation would need to be carried throughout the entire calculation.

The Cramer–Wold theorem implies the convergence of finite dimensional distributions which completes the proof of Theorem 6.1. The remainder of this section is devoted to proving the technical lemmas needed for Lemma 6.15.

**Lemma 6.17.** *Define all quantities as in Lemma 6.12. Then under the assumptions of Lemma 6.12, for any $\alpha>0$, we have*

$$\begin{aligned}
\mathbb{E}\Big| &e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k} \\
&-e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]e_{(i-1)n+k}\Big|^2 \\
&= o_\alpha(n^{4-\alpha/4})
\end{aligned}$$

*uniformly in $i$ and $k$.*

*Proof.* Observe that

$$\begin{aligned}
&e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\\
&\quad= e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]e_{(i-1)n+k}\\
&\qquad+ e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}\cap\Omega^c_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]e_{(i-1)n+k} \qquad (89)\\
&\qquad+ e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega^c_{n,k,s}}\right]e_{(i-1)n+k} \qquad (90)\\
&\qquad+ e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}\cap\Omega^c_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega^c_{n,k,s}}\right]e_{(i-1)n+k}. \qquad (91)
\end{aligned}$$

Therefore, we must show terms (89), (90), and (91) are sufficiently small in the $L^2$-norm. The argument for all three terms is very similar. We use Jensen's inequality and the Cauchy–Schwarz inequality to separate the inner conditional expectations. For resolvent terms where a complement event is not present, we bound by a constant and in terms where a complement event is present, we bound by $O(n^{-\alpha})$ since each event holds with overwhelming probability. We show the calculation for term (89), and the other terms follow in a similar manner. Observe

$$\begin{aligned}
&\mathbb{E}\left|e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)c_s\mathbf{1}_{\Omega_{n,k}\cap\Omega^c_{n,k,s}}\right]\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]e_{(i-1)n+k}\right|^2\\
&\quad\le \mathbb{E}\left[\mathbb{E}_k\left[\left\|\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega^c_{n,k,s}}\right\|^2\|c_s\|^2\right]\mathbb{E}_k\left[\|c_s^*\|^2\left\|(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right\|^2\right]\right]\\
&\quad\le \left(\mathbb{E}\left\|\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega^c_{n,k,s}}\right\|^4\mathbb{E}\|c_s\|^4\,\mathbb{E}\|c_s^*\|^4\,\mathbb{E}\left\|(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right\|^4\right)^{1/2}\\
&\quad\ll \left(\mathbb{E}\left\|\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega^c_{n,k,s}}\right\|^4\mathbb{E}\left\|(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right\|^4\right)^{1/2}\\
&\quad\ll \mathbb{P}(\Omega^c_{n,k,s})^{1/2}\\
&\quad\ll n^{-\alpha/2}.
\end{aligned}$$

□

**Lemma 6.18.** *Define all quantities as in Lemma 6.12. Then, for any $i$ with $1 \le i \le m$ where subscripts are reduced modulo $m$, under the assumptions of Lemma 6.12, we have*

$$\begin{aligned}
\mathbb{E}\Big|&e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\times\mathbb{E}_k\left[(\delta_{k,s}(w))^*c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\\
&-e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\times\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}]\Big|^2 = o(n^{-2})
\end{aligned}$$

*uniformly in $i$, $s$, and $k$.*

*Proof.* Observe that by the triangle inequality,

$$\begin{aligned}
&\mathbb{E}\Big|e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad\times\mathbb{E}_k\left[(\delta_{k,s}(w))^*c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\\
&\qquad-e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad\qquad\times\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\Big|^2\\
&\ll\mathbb{E}\Big|e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad\times\mathbb{E}_k\left[((\delta_{k,s}(w))^*-1)c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\Big|^2 &(92)\\
&\quad+\mathbb{E}\Big|e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s(\delta_{k,s}(z)-1)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad\times\mathbb{E}_k\left[c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\Big|^2. &(93)
\end{aligned}$$

We will show each of these terms are $o(n^{-2})$. We begin with (92). By the generalized Hölders inequality, Lemma 6.6, and Lemma 6.14, we have

$$\begin{aligned}
&\mathbb{E}\Big|e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\\
&\qquad\qquad\times\mathbb{E}_k\left[((\delta_{k,s}(w))^*-1)c_s^*(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]e_{(i-1)n+k}\Big|^2\\
&\leq\left(\mathbb{E}\left|e_{(i-1)n+k}^T\mathcal{G}_n^{(k,s)}(z)c_s\delta_{k,s}(z)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right|^4\right)^{1/2}\\
&\quad\times\left(\mathbb{E}\left|((\delta_{k,s}(w))^*-1)\mathbf{1}_{Q'_{n,k,s}}\right|^8\right)^{1/4}\left(\mathbb{E}\left|c_s^*(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k,s}}\right|^8\right)^{1/4}\\
&\ll\left(\mathbb{E}\left[\left|c_s^*(\mathcal{G}_n^{(k,s)}(z))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n^{(k,s)}(z)c_s\right|^2\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right]\right)^{1/2}\\
&\quad\times\left(\mathbb{E}\left|((\delta_{k,s}(w))^*-1)\mathbf{1}_{Q'_{n,k,s}}\right|^8\right)^{1/4}\\
&\quad\times\left(\mathbb{E}\left|c_s^*(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}e_{(i-1)n+k}^T\mathcal{G}_n^{(k,s)}(w)c_s\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right|^4\right)^{1/4}\\
&\ll n^{-3/2-\varepsilon}\left(\mathbb{E}\left|((\delta_{k,s}(w))^*-1)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right|^8\right)^{1/4}. &(94)
\end{aligned}$$

Now, recall $\delta_{k,s}(z)$ defined in (65). By the resolvent identity (12), we have $1-\delta_{k,s}(z)=1-\left(1+e_s^T\mathcal{G}_n^{(k,s)}(z)c_s\right)^{-1}=\left(e_s^T\mathcal{G}_n^{(k,s)}(z)c_s\right)\delta_{k,s}(z)$. This gives

$$\delta_{k,s}(z)=1-(e_s^T\mathcal{G}_n^{(k,s)}(z)c_s)+(e_s^T\mathcal{G}_n^{(k,s)}(z)c_s)^2\delta_{k,s}(z).$$

Ergo,

$$((\delta_{k,s}(w))^*-1=-(e_s^T\mathcal{G}_n^{(k,s)}(w)c_s)^*+(e_s^T\mathcal{G}_n^{(k,s)}(w)c_s)^{2*}(\delta_{k,s}(w))^*. \tag{95}$$

We replace $((\delta_{k,s}(w))^* - 1)$ in (94) with the expression on the right hand side of (95) and use Lemma 6.6 to see

$$
\begin{aligned}
&n^{-3/2-\varepsilon}\left(\mathbb{E}\left|((\delta_{k,s}(w))^* - 1)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right|^8\right)^{1/4} \\
&\ll n^{-3/2-\varepsilon}\left(\mathbb{E}\left|(-(e_s^T\mathcal{G}_n^{(k,s)}(w)c_s)^* + (\delta_{k,s}(w))^*(e_s^T\mathcal{G}_n^{(k,s)}(w)c_s)^{2*})\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right|^8\right)^{1/4} \\
&\ll n^{-3/2-\varepsilon}\left(\mathbb{E}\left|c_s^*(\mathcal{G}_n^{(k,s)}(w))^*e_se_s^T\mathcal{G}_n^{(k,s)}(w)c_s\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right|^4\right. \\
&\qquad\left. + \mathbb{E}\left[\left|(\delta_{k,s}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right|^8\left|c_s^*(\mathcal{G}_n^{(k,s)}(w))^*e_se_s^T\mathcal{G}_n^{(k,s)}(w)c_s\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right|^8\right]\right)^{1/4} \\
&\ll n^{-3/2-\varepsilon}\left(n^{-4\varepsilon-2}\mathbb{E}\left\|(\mathcal{G}_n^{(k,s)}(w))^*e_se_s^T\mathcal{G}_n^{(k,s)}(w)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right\|^4\right. \\
&\qquad\left. + n^{-12\varepsilon-2}\mathbb{E}\left\|(\mathcal{G}_n^{(k,s)}(w))^*e_se_s^T\mathcal{G}_n^{(k,s)}(w)\mathbf{1}_{\Omega_{n,k,s}\cap Q'_{n,k,s}}\right\|^8\right)^{1/4} \\
&\ll n^{-3/2-\varepsilon}n^{-1/2-\varepsilon}.
\end{aligned}
$$

This shows term (92) is $o(n^{-2})$. A very similar argument shows that term (93) is $o(n^{-2})$ as well. We omit the details. This completes the proof. $\square$

**Lemma 6.19.** *Define all quantities as in Lemma 6.12. Then under the assumptions of Lemma 6.12, for any $\alpha > 0$*

$$\mathbb{E}\left|\frac{1}{z\bar{w}}(1 - \mathbb{P}_k(\Omega_{n,k})^2)\right|^2 = o_\alpha(n^{-\alpha})$$

*uniformly in $k$.*

*Proof.* Observe that since $z, \bar{w} \in \mathcal{C}$ are fixed with $|z| = |\bar{w}| = 1 + \delta$, we know that

$$\left|\frac{1}{z\bar{w}}(1 - \mathbb{P}_k(\Omega_{n,k})^2)\right|^2 \ll \left|1 - \mathbb{P}_k(\Omega_{n,k})^2\right|^2 \ll \left|1 - \mathbb{P}_k(\Omega_{n,k})\right|^2.$$

Since $\Omega_{n,k}$ holds with overwhelming probability by Corollary B.5,

$$\mathbb{E}\left|1 - \mathbb{P}_k(\Omega_{n,k})\right|^2 = \mathbb{E}\left|\mathbb{P}_k(\Omega_{n,k}^c)\right|^2 \le \mathbb{P}(\Omega_{n,k}^c) = o_\alpha(n^{-\alpha})$$

for any $\alpha > 0$. $\square$

**Lemma 6.20.** *Define all quantities as in Lemma 6.12 and assume $(i-b-1)n+1 \le s \le (i-b)n$. Then under the assumptions of Lemma 6.12,*

$$
\begin{aligned}
\mathbb{E}\Big|&e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]c_sc_s^*\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k} \\
&- \frac{1}{n}e_{(i-1)n+k}^T\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\Big|^2 = o(n^{-1})
\end{aligned}
$$

*uniformly in $k$.*

*Proof.* To begin, observe that by viewing the expression as a trace, and by cyclic permutation, we can rewrite

$$\begin{aligned}&e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]c_sc_s^*\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\\&\quad=\frac{1}{n}\left(\sqrt{n}c_s^*\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\sqrt{n}c_s\right).\end{aligned}$$

For any complex-valued $N\times N$ matrix $A$ and any subset $S\subseteq[N]$, let $A_{S\times S}$ denote the $|S|\times|S|$ matrix which has entries $A_{(i,j)}$ for $i,j\in S$. Let $S_b=\{(i-b-2)n+1,(i-b-2)n+2,\dots,(i-b-1)n\}$. Then observe by cyclic permutation of the trace, we have

$$\begin{aligned}&\frac{1}{n}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\\&\quad=\operatorname{tr}\left(\frac{1}{n}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\right)\\&\quad=\frac{1}{n}\operatorname{tr}\left(\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\right)\\&\quad=\frac{1}{n}\operatorname{tr}\left(\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\right)_{S_b\times S_b}.\end{aligned}$$

By this observation and Lemma C.2 we have

$$\begin{aligned}&\mathbb{E}\Bigg|e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]c_sc_s^*\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\\&\qquad-\frac{1}{n}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\Bigg|^2\\&\quad=\frac{1}{n^2}\mathbb{E}\Bigg|\sqrt{n}c_s^*\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\sqrt{n}c_s\\&\qquad-\operatorname{tr}\left(\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\right)_{S_b\times S_b}\Bigg|^2\\&\quad\ll\frac{1}{n^2}\mathbb{E}\Bigg[\operatorname{tr}\Bigg(\left(\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\right)^*\\&\qquad\times\left(\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\right)\Bigg)\Bigg].\qquad(96)\end{aligned}$$

Observe that the rank is at most 1 so by bounding the trace by the rank times the norm, and by the fact that each term in the above expression can be bounded in norm by a constant, we can bound (96) by $O(n^{-2})$ which completes the proof. □

**Lemma 6.21.** *Define all quantities as in Lemma 6.12. Then under the assumptions of Lemma 6.12, we have*

$$\begin{aligned}&\mathbb{E}\Bigg|\frac{1}{n}\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\\&\qquad-\frac{1}{n}\sum_{s=(i-2)n+1}^{(i-2)n+k-1}e^T_{(i-1)n+k}\mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\Bigg|^2=o(1)\end{aligned}$$

*uniformly in $k$.*

*Proof.* To begin, observe that

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1} e_{(i-1)n+k}^T \mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]e_{(i-1)n+k}\right.\\
&\qquad\left.-\frac{1}{n}\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1} e_{(i-1)n+k}^T \mathbb{E}_k\left[\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right]e_{(i-1)n+k}\right|^2\\
&\ll \frac{1}{n}\sum_{s=(i-b-1)n+1}^{(i-b-1)n+k-1}\mathbb{E}\left|\mathbb{E}_k\left[e_{(i-1)n+k}^T\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k,s}}\right]\right.\\
&\qquad\qquad\left.-\mathbb{E}_k\left[e_{(i-1)n+k}^T\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right]\right|^2
\end{aligned}$$

so it suffices to prove that

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}_k\left[e_{(i-1)n+k}^T\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k,s}}\right]\right.\\
&\qquad\left.-\mathbb{E}_k\left[e_{(i-1)n+k}^T\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right]\right|^2 = o(1)
\end{aligned}$$

uniformly in $i$, $k$ and $s$. We can expand this difference using the triangle inequality to get differences which only vary by an event or a resolvent. We begin by observing that by the Cauchy–Schwarz inequality and Lemma 6.14, for any $\alpha > 0$,

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}_k\left[e_{(i-1)n+k}^T\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k,s}}\right]\right.\\
&\qquad\left.-\mathbb{E}_k\left[e_{(i-1)n+k}^T\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k,s}\cap\Omega_{n,k}}\right]\right|^2\\
&\ll \mathbb{E}\left\|\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right]\right.\\
&\qquad\left.-\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}\cap\Omega_{n,k}}\right]\right\|^2\\
&= \mathbb{E}\left\|\mathbb{E}_k\left[\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*(\mathbf{1}_{\Omega_{n,k,s}}\mathbf{1}_{\Omega_{n,k}^c})\right]\right\|^2\\
&\le \left(\mathbb{E}\left[\left\|\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k,s}}\right\|^4\|\mathcal{D}_{i-b-1}\|^4\right]\mathbb{E}\left[\left\|(\mathcal{G}_n^{(k,s)}(w))^*\mathbf{1}_{\Omega_{n,k,s}}\right\|^4\mathbf{1}_{\Omega_{n,k}^c}\right]\right)^{1/2}\\
&\ll_\alpha n^{-\alpha/2}.
\end{aligned}$$

The same argument shows that each indicator can be replaced with $\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}$. We also bound terms that differ by a resolvent. To this end, observe that by the resolvent identity (12) and Lemma 6.6,

$$
\begin{aligned}
&\mathbb{E}\Big|\mathbb{E}_k\left[e^T_{(i-1)n+k}\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\\
&\quad-\mathbb{E}_k\left[e^T_{(i-1)n+k}\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\Big|^2\\
&=\mathbb{E}\Big|\mathbb{E}_k\left[e^T_{(i-1)n+k}\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\\
&\qquad\times\mathbb{E}_k\left[(\mathcal{G}_n^{(k,s)}(w)(c_se_s^T)\mathcal{G}_n^{(k)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\Big|^2\\
&\leq\Bigg(\mathbb{E}\left[\left\|e_{(i-1)n+k}\right\|^4\left\|\mathcal{G}_n^{(k,s)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right\|^4\left\|\mathcal{D}_{i-b-1}\right\|^4\right]\\
&\qquad\quad\times\mathbb{E}\left[\left\|(\mathcal{G}_n^{(k)}(w))^*\mathbf{1}_{\Omega_{n,k}}\right\|^4\left\|e_s\right\|^4\left|c_s^*(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k,s}}\right|^4\right]\Bigg)^{1/2}\\
&\ll\left(\mathbb{E}\left|c_s^*(\mathcal{G}_n^{(k,s)}(w))^*e_{(i-1)n+k}e^T_{(i-1)n+k}\mathcal{G}_n^{(k,s)}(w)c_s\mathbf{1}_{\Omega_{n,k,s}}\right|^2\right)^{1/2}\\
&\ll n^{-1}.
\end{aligned}
$$

The same argument shows that all instances of $\mathcal{G}_n^{(k,s)}(z)$ or $(\mathcal{G}_n^{(k,s)}(w))^*$ can be replaced with $\mathcal{G}_n^{(k)}(z)$ or $(\mathcal{G}_n^{(k)}(w))^*$ respectively gaining an error that is $o(1)$ in $L^2$-norm. Finally, the same argument as before shows that

$$
\begin{aligned}
&\mathbb{E}\Big|\mathbb{E}_k\left[e^T_{(i-1)n+k}\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}\right]\\
&-\mathbb{E}_k\left[e^T_{(i-1)n+k}\mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k}}\right]\mathcal{D}_{i-b-1}\mathbb{E}_k\left[(\mathcal{G}_n^{(k)}(w))^*e_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right]\Big|^2=o(1).
\end{aligned}
$$

Replacing all instances of $\mathbf{1}_{\Omega_{n,k}\cap\Omega_{n,k,s}}$ with $\mathbf{1}_{\Omega_{n,k}}$ completes the proof. □

## 7. Tightness

In order to extend the finite dimensional convergence proved in Section 6 to convergence of the stochastic process $\{\Xi_n(z)\}_{z\in\mathcal{C}}$, we must check that the sequence of stochastic processes $\{\Xi_n(z)\}_{z\in\mathcal{C}}$ is tight. Namely, recall that we must verify condition (43) in Theorem 5.3. To check this condition, it will be helpful to recenter $\Xi_n(z)$. Define the modified sequence

$$\tilde{\Xi}_n(z)=\sum_{k=1}^n(\mathbb{E}_k-\mathbb{E}_{k-1})[(\operatorname{tr}(\mathcal{G}_n(z))-\operatorname{tr}(\mathcal{G}_n^{(k)}(z)))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}] \tag{97}$$

which differs from $\Xi_n(z)$ by the fact that we have subtracted the trace of $\mathcal{G}_n^{(k)}(z)$ and multiplied by $\mathbf{1}_{\Omega_{n,k}}$. We wish to proceed from here working with $\tilde{\Xi}_n(z)$ instead of $\Xi_n(z)$. We will be justified in doing so after proving the following lemma.

**Lemma 7.1.** *Let $\Xi_n(z)$ be as defined in* (36) *and $\hat{\Xi}_n(z)$ as in* (97). *Then under the assumptions of Theorem 4.13,*

$$\mathbb{E}\left|\frac{\Xi_n(z)-\Xi_n(w)}{z-w}\right|^2\leq c+\mathbb{E}\left|\frac{\tilde{\Xi}_n(z)-\tilde{\Xi}_n(w)}{z-w}\right|^2$$

*for some constant $c>0$ independent of $n$ and of any choice of $z,w$ on the contour $\mathcal{C}$.*

*Proof.* We can see that

$$\begin{aligned}
&\mathbb{E}\left|\frac{\Xi_n(z)-\Xi_n(w)}{z-w}\right|^2\\
&\ll \mathbb{E}\left|\frac{\Xi_n(z)-\Xi_n(w)}{z-w}-\frac{\tilde{\Xi}_n(z)-\tilde{\Xi}_n(w)}{z-w}\right|^2+\mathbb{E}\left|\frac{\tilde{\Xi}_n(z)-\tilde{\Xi}_n(w)}{z-w}\right|^2.
\end{aligned}$$

Now note that by the resolvent identity (12),

$$\Xi_n(z)-\Xi_n(w)=\sum_{k=1}^{n}(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n(z)(w-z)\mathcal{G}_n(w))\mathbf{1}_{\Omega_n}]$$

and

$$\begin{aligned}
&\tilde{\Xi}_n(z)-\tilde{\Xi}_n(w)\\
&\quad=\sum_{k=1}^{n}(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n(z)(w-z)\mathcal{G}_n(w))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}\\
&\qquad\qquad-\operatorname{tr}(\mathcal{G}_n^{(k)}(z)(w-z)\mathcal{G}_n^{(k)}(w))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}].
\end{aligned}$$

Therefore, by cyclic permutation of the trace and since the covariance terms in a martingale difference sequence are zero, we have

$$\begin{aligned}
&\mathbb{E}\left|\frac{\Xi_n(z)-\Xi_n(w)}{z-w}-\frac{\tilde{\Xi}_n(z)-\tilde{\Xi}_n(w)}{z-w}\right|^2\\
&=\mathbb{E}\left|\sum_{k=1}^{n}\big((\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n(w)\mathcal{G}_n(z))(\mathbf{1}_{\Omega_n}-\mathbf{1}_{\Omega_n\cap\Omega_{n,k}})]\right.\\
&\qquad\left.-(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}]\big)\right|^2\\
&\ll\sum_{k=1}^{n}\left(\mathbb{E}\left|\operatorname{tr}(\mathcal{G}_n(w)\mathcal{G}_n(z))\mathbf{1}_{\Omega_n}\mathbf{1}_{\Omega_{n,k}^c}\right|^2\right.\\
&\qquad\left.+\mathbb{E}\left|(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w))\mathbf{1}_{\Omega_{n,k}}]\right|^2\right)\\
&\ll_\alpha n^{3-\alpha}
\end{aligned}$$

for any $\alpha>0$ since $(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w))\mathbf{1}_{\Omega_{n,k}}]=0$. Note that any choice of $\alpha\geq 3$ suffices to show this term is bounded by a constant, concluding the proof. □

The tightness of $\{\Xi_n(z)\}_{z\in\mathcal{C}}$ will follow from the following lemma.

**Lemma 7.2.** *Let $\{\tilde{\Xi}_n(z)\}$ be the sequence of stochastic processes defined in* (97). *It holds that*

$$\mathbb{E}\left|\frac{\tilde{\Xi}_n(z)-\tilde{\Xi}_n(w)}{z-w}\right|^2\leq c$$

*for a constant $c>0$ independent of $n$ and of any choice of $z,w$ on the contour $\mathcal{C}$.*

*Proof.* The idea behind this proof is similar to what was done in the proof of Lemma 6.12 where we remove columns to achieve independence. First, observe that by definition of $\tilde{\Xi}_n(z)$, linearity of trace, and the resolvent identity (12),

$$\begin{aligned}&\frac{\tilde{\Xi}_n(z)-\tilde{\Xi}_n(w)}{z-w}\\&=\sum_{k=1}^{n}\frac{(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n(z)-\mathcal{G}_n(w)-(\mathcal{G}_n^{(k)}(z)-\mathcal{G}_n^{(k)}(w)))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}]}{z-w}\\&=-\sum_{k=1}^{n}(\mathbb{E}_k-\mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n(z)\mathcal{G}_n(w)-\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}].\end{aligned}\tag{98}$$

Now note that

$$\begin{aligned}&(\mathcal{G}_n(z)-\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n(w)-\mathcal{G}_n^{(k)}(w))\\&\quad=\mathcal{G}_n(z)\mathcal{G}_n(w)-\mathcal{G}_n(z)\mathcal{G}_n^{(k)}(w)-\mathcal{G}_n^{(k)}(z)\mathcal{G}_n(w)+\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)\end{aligned}$$

which implies

$$\begin{aligned}&\mathcal{G}_n(z)\mathcal{G}_n(w)+\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)\\&\quad=(\mathcal{G}_n(z)-\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n(w)-\mathcal{G}_n^{(k)}(w))+\mathcal{G}_n(z)\mathcal{G}_n^{(k)}(w)+\mathcal{G}_n^{(k)}(z)\mathcal{G}_n(w).\end{aligned}$$

By subtracting $2\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)$ from each side of the previous equality, regrouping, and applying the resolvent identity (12), we have

$$\begin{aligned}&\mathcal{G}_n(z)\mathcal{G}_n(w)-\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)\\&=(\mathcal{G}_n(z)(U_kV_k^T)\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n(w)(U_kV_k^T\mathcal{G}_n^{(k)}(w))\\&\quad+(\mathcal{G}_n(z)(U_kV_k^T)\mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)\\&\quad+\mathcal{G}_n^{(k)}(z)(\mathcal{G}_n(w)(U_kV_k^T)\mathcal{G}_n^{(k)}(w))\end{aligned}$$

where we recall that $U_k$ is the $mn\times n$ matrix which contains as its columns $c_k, c_{n+k}, \dots, c_{(m-1)n+k}$ and $V_k$ is the $mn\times m$ matrix which contains as its columns $e_k, e_{n+k}, \dots, e_{(m-1)n+k}$. By the Sherman–Morrison–Woodbury formula (11), we know $\mathcal{G}_n(z)U_k=\mathcal{G}_n^{(k)}(z)U_k(I_m+V_k^T\mathcal{G}_n^{(k)}(z)U_k)^{-1}=\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)$ where $\Delta_{n,k}(z):=(I_m+V_k^T\mathcal{G}_n^{(k)}(z)U_k)^{-1}$ provided $I_m+V_k^T\mathcal{G}_n^{(k)}(z)U_k$ is invertible. Recall, as was done in Section 6, we can guarantee that this matrix is invertible by working on the event $Q_{n,k}$ defined in (52). Since $Q_{n,k}$ holds with overwhelming probability by Lemma 6.5, the same argument as in Section 6 shows that we can work on this event with error $o_\alpha(n^{-\alpha})$ for any $\alpha>0$, so we are justified doing so. Therefore we can continue on the event $\Omega_{n,k}\cap Q_{n,k}$, with

$$\begin{aligned}&\mathcal{G}_n(z)\mathcal{G}_n(w)-\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)\\&=(\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w))\\&\quad+(\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)\\&\quad+\mathcal{G}_n^{(k)}(z)(\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w)).\end{aligned}$$

Since $(\mathcal{Y}_n^{(k)}-zI)$ and $(\mathcal{Y}_n^{(k)}-wI)$ commute, we can interchange the order in which we multiply $\mathcal{G}_n^{(k)}(z)$ and $\mathcal{G}_n^{(k)}(w)$. By this observation and by cyclic permutation of the trace, we have

$$\begin{aligned}
&\operatorname{tr}(\mathcal{G}_n(z)\mathcal{G}_n(w) - \mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)) \\
&= \operatorname{tr}\Big((\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w))\Big) \\
&\quad + \operatorname{tr}\Big((\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)\Big) \\
&\quad + \operatorname{tr}\Big(\mathcal{G}_n^{(k)}(z)(\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w))\Big).
\end{aligned}$$

Putting all of these observations together, we have shown that

$$\begin{aligned}
&\mathbb{E}\left|\frac{\tilde{\Xi}_n(z) - \tilde{\Xi}_n(w)}{z-w}\right|^2 \\
&\quad \ll \sum_{k=1}^n \mathbb{E}\left|(\mathbb{E}_k - \mathbb{E}_{k-1})[\operatorname{tr}(\mathcal{G}_n(z)\mathcal{G}_n(w) - \mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w))\mathbf{1}_{\Omega_n\cap\Omega_{n,k}}]\right|^2 \\
&\quad \le \sum_{k=1}^n \mathbb{E}\Big|\operatorname{tr}\Big((\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z)) \\
&\qquad\qquad \times (\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w))\Big)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\Big|^2 \qquad (99) \\
&\qquad + \sum_{k=1}^n \mathbb{E}\Big|\operatorname{tr}\Big((\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)\Big)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\Big|^2 \qquad (100) \\
&\qquad + \sum_{k=1}^n \mathbb{E}\Big|\operatorname{tr}\Big(\mathcal{G}_n^{(k)}(z)(\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w))\Big)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\Big|^2 + O(1). \qquad (101)
\end{aligned}$$

Note that since $\mathcal{G}_n(z)$ is no longer present in (99), (100), and (101), we are justified dropping the event $\Omega_n$ as well. The $O(1)$ is due to the error from introducing the event $Q_{n,k}$ and dropping the event $\Omega_n$. Next, we show that we can replace $\Delta_{n,k}(z)$ and $\Delta_{n,k}(w)$ with $I_m$ in (99), (100), and (101). We begin by showing the calculation for term (99). Observe that by cyclic permutation of the trace, on the event $\Omega_{n,k} \cap Q_{n,k}$, we have

$$\begin{aligned}
&\operatorname{tr}\Big((\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w))\Big) \\
&\quad - \operatorname{tr}\Big((\mathcal{G}_n^{(k)}(z)U_kV_k^T\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n^{(k)}(w)U_kV_k^T\mathcal{G}_n^{(k)}(w))\Big) \\
&= \operatorname{tr}\Big((\Delta_{n,k}(z) - I_m)V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w)\mathcal{G}_n^{(k)}(z)U_k\Big) \\
&\quad + \operatorname{tr}\Big(V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k(\Delta_{n,k}(w) - I_m)V_k^T\mathcal{G}_n^{(k)}(w)\mathcal{G}_n^{(k)}(z)U_k\Big).
\end{aligned}$$

We use the generalized Hölders inequality to break the above expression into pieces which have bounded expectation. By bounding the trace by the rank times the norm, we have

$$\mathbb{E}\Big|\mathrm{tr}\left((\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w))\right)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}$$
$$-\ \mathrm{tr}\left((\mathcal{G}_n^{(k)}(z)U_kV_k^T\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n^{(k)}(w)U_kV_k^T\mathcal{G}_n^{(k)}(w))\right)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\Big|^2$$
$$\ll \left(\mathbb{E}\left\|(\Delta_{n,k}(z)-I_m)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^4\right)^{1/2}\left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^8\right)^{1/4} \tag{102}$$
$$\times\left(\mathbb{E}\left\|\Delta_{n,k}(w)\mathbf{1}_{Q_{n,k}}\right\|^{16}\right)^{1/8}\left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(w)\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^{16}\right)^{1/8} \tag{103}$$
$$+\left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^4\right)^{1/2} \tag{104}$$
$$\times\left(\mathbb{E}\left\|(\Delta_{n,k}(w)-I_m)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^8\right)^{1/4}\left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(w)\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^8\right)^{1/4} \tag{105}$$

We bound each expectation in (102), (103), (104), and (105). We start by bounding the expectation of terms in which two resolvents appear. By the same argument as in Lemma 6.9, we can show that for $p \geq 4$

$$\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(w)\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^p \ll_p n^{-\varepsilon(p-4)-2}$$

which shows

$$\left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(w)\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^{16}\right)^{1/8} \ll n^{-3/2\varepsilon-1/4},$$

$$\left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^4\right)^{1/2} \ll n^{-1},$$

and

$$\left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^8\right)^{1/4} \ll n^{-\varepsilon-1/2}.$$

Since the second term in (105) differs from the last line above only by the order in which we multiply resolvents, the same bound holds for the second term in (105). Now we bound terms involving $\Delta_{n,k}$. Recall that $\Delta_{n,k}$ is bounded by a constant almost surely on $Q_{n,k}$ so we need only to bound the expectations involving $\Delta_{n,k}(z) - I_m$ in terms (102) and (105). Recall the expansion from (57) can be iterated to get

$$\Delta_{n,k}(z) = I_m - (V_k^T\mathcal{G}_n^{(k)}(z)U_k) + (V_k^T\mathcal{G}_n^{(k)}(z)U_k)^2\Delta_{n,k}(z).$$

Using this fact, Lemma 6.9, and the Cauchy–Schwarz inequality, we get

$$\left(\mathbb{E}\left\|(\Delta_{n,k}(z)-I_m)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^4\right)^{1/2}$$
$$\ll\left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^4+\mathbb{E}\left\|(V_k^T\mathcal{G}_n^{(k)}(z)U_k)^2\Delta_{n,k}(z)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^4\right)^{1/2}$$
$$\ll n^{-2\varepsilon-1}+n^{-1}$$

and

$$\begin{aligned}
&\left(\mathbb{E}\left\|(\Delta_{n,k}(w) - I_m)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^8\right)^{1/4} \\
&\ll \left(\mathbb{E}\left\|V_k^T\mathcal{G}_n^{(k)}(z)U_k\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^8 + \mathbb{E}\left\|(V_k^T\mathcal{G}_n^{(k)}(z)U_k)^2\Delta_{n,k}(z)\mathbf{1}_{\Omega_{n,k}\cap Q_{n,k}}\right\|^8\right)^{1/4} \\
&\ll n^{-\varepsilon-1/2} + n^{-3\varepsilon-1/2.}
\end{aligned}$$

Therefore, combining these bounds, we can bound (102), (103), (104) and (105) by $O(n^{-7/4})$. This concludes the argument to show that we may replace all $\Delta_{n,k}(z)$ and $\Delta_{n,k}(w)$ in term (99) with $I_m$. A very similar argument shows that, for terms (100) and (101),

$$\mathbb{E}\left|\operatorname{tr}\left((\mathcal{G}_n^{(k)}(z)U_k\Delta_{n,k}(z)V_k^T\mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)\right) - \operatorname{tr}\left((\mathcal{G}_n^{(k)}(z)U_kV_k^T\mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)\right)\right|^2$$

and

$$\mathbb{E}\left|\operatorname{tr}\left(\mathcal{G}_n^{(k)}(z)(\mathcal{G}_n^{(k)}(w)U_k\Delta_{n,k}(w)V_k^T\mathcal{G}_n^{(k)}(w))\right) - \operatorname{tr}\left(\mathcal{G}_n^{(k)}(z)(\mathcal{G}_n^{(k)}(w)U_kV_k^T\mathcal{G}_n^{(k)}(w))\right)\right|^2$$

are both $O(n^{-2})$. Ergo, we need to show only that the expression

$$\begin{aligned}
&\mathbb{E}\left|\operatorname{tr}\left((\mathcal{G}_n^{(k)}(z)U_kV_k^T\mathcal{G}_n^{(k)}(z))(\mathcal{G}_n^{(k)}(w)U_kV_k^T\mathcal{G}_n^{(k)}(w))\right)\right. \\
&\qquad\left. + \operatorname{tr}\left((\mathcal{G}_n^{(k)}(z)U_kV_k^T\mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)\right) + \operatorname{tr}\left(\mathcal{G}_n^{(k)}(z)(\mathcal{G}_n^{(k)}(w)U_kV_k^T\mathcal{G}_n^{(k)}(w))\right)\right|^2 \\
&\ll \mathbb{E}\left|\operatorname{tr}\left(V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k\right)^2\right|^2 \qquad (106) \\
&\qquad + \mathbb{E}\left|\operatorname{tr}\left(V_k^T\mathcal{G}_n^{(k)}(w)(\mathcal{G}_n^{(k)}(z))^2U_k\right)\right|^2 \qquad (107) \\
&\qquad + \mathbb{E}\left|\operatorname{tr}\left(V_k^T\mathcal{G}_n^{(k)}(z)(\mathcal{G}_n^{(k)}(w))^2U_k\right)\right|^2 \qquad (108)
\end{aligned}$$

is bounded by $O(n^{-1})$, where we cyclically permuted the trace and reordered the product of resolvents again. We will bound each term separately. First consider term (106), and observe that

$$\begin{aligned}
&\mathbb{E}\left|\operatorname{tr}\left((V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k)^2\right)\mathbf{1}_{\Omega_{n,k}}\right|^2 \\
&\qquad \ll \mathbb{E}\left\|U_k^*(\mathcal{G}_n^{(k)}(w))^*(\mathcal{G}_n^{(k)}(z))^*V_kV_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)U_k\mathbf{1}_{\Omega_{n,k}}\right\|^2
\end{aligned}$$

Since this matrix is $m \times m$, if we can bound an arbitrary entry uniformly, then we can bound the norm. We now wish to bound

$$\max_{1\le i,j\le m}\mathbb{E}\left|c_{(i-1)n+k}^T(\mathcal{G}_n^{(k)}(w))^*(\mathcal{G}_n^{(k)}(z))^*V_kV_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)c_{(j-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right|^2.$$

Note that $(\mathcal{G}_n^{(k)}(w))^*(\mathcal{G}_n^{(k)}(z))^*V_kV_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)$ is independent of the $k$th column of each block, and hence is independent from $c_{(i-1)n+k}$ and $c_{(j-1)n+k}$. It is also rank at most $m$ and it is Hermitian positive definite. Then by Lemma 6.6

(when $i = j$) or Lemma 6.7 (when $i \neq j$), we have

$$\begin{aligned}&\mathbb{E}\left|c^*_{(i-1)n+k}(\mathcal{G}_n^{(k)}(w))^*(\mathcal{G}_n^{(k)}(z))^*V_kV_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)c_{(j-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right|^2\\&\quad\ll n^{-2}\mathbb{E}\left[\operatorname{tr}\left(((\mathcal{G}_n^{(k)}(w))^*(\mathcal{G}_n^{(k)}(z))^*V_kV_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)\mathbf{1}_{\Omega_{n,k}})^2\right)\right]\\&\quad\ll n^{-2}\end{aligned}$$

where we bounded an arbitrary element of $V_k^T\mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)(\mathcal{G}_n^{(k)}(w))^*(\mathcal{G}_n^{(k)}(z))^*V_k$ by a constant on the event $\Omega_{n,k}$. This concludes the argument for term (106). Since terms (107) and (108) are symmetric in $z$ and $w$, the argument will be the same for both terms. We show the argument for (107). Observe that

$$\begin{aligned}&\mathbb{E}\left|\operatorname{tr}\left(V_k^T\mathcal{G}_n^{(k)}(w)(\mathcal{G}_n^{(k)}(z))^2U_k\right)\mathbf{1}_{\Omega_{n,k}}\right|^2\\&\quad\ll \max_{1\leq i\leq m}\mathbb{E}\left|e^T_{(i-1)n+k}\mathcal{G}_n^{(k)}(w)(\mathcal{G}_n^{(k)}(z))^2c_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}}\right|^2\\&\quad\ll \max_{1\leq i\leq m}\mathbb{E}\Big[c^*_{(i-1)n+k}(\mathcal{G}_n^{(k)}(z))^{2*}(\mathcal{G}_n^{(k)}(w))^*e_{(i-1)n+k}\\&\qquad\qquad\times e^T_{(i-1)n+k}\mathcal{G}_n^{(k)}(w)(\mathcal{G}_n^{(k)}(z))^2c_{(i-1)n+k}\mathbf{1}_{\Omega_{n,k}}\Big]\\&\quad\ll n^{-1}.\end{aligned}$$

This concludes the argument for (107) and the proof of Lemma 7.2. □

## Appendix A. Truncation Arguments

This section is devoted to the proof of Lemma 4.3.

*Proof of Lemma 4.3.* First, we prove property *(i)*. Observe that

$$\begin{aligned}1 &= \operatorname{Var}(\xi)\\&= \mathbb{E}[\xi^2\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}] + \mathbb{E}[\xi^2\mathbf{1}_{\{|\xi|> n^{1/2-\varepsilon}\}}]\\&= \operatorname{Var}(\tilde{\xi}) + \left(\mathbb{E}[\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}]\right)^2 + \mathbb{E}[\xi^2\mathbf{1}_{\{|\xi|> n^{1/2-\varepsilon}\}}].\end{aligned}$$

Also observe that

$$0 = \mathbb{E}[\xi] = \mathbb{E}[\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}] + \mathbb{E}[\xi\mathbf{1}_{\{|\xi|> n^{1/2-\varepsilon}\}}]$$

which implies $\left|\mathbb{E}[\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}]\right| = \left|\mathbb{E}[\xi\mathbf{1}_{\{|\xi|> n^{1/2-\varepsilon}\}}]\right|$. Hence

$$\begin{aligned}|1-\operatorname{Var}(\tilde{\xi})| &= \left(\mathbb{E}[\xi\mathbf{1}_{\{|\xi|\leq n^{1/2-\varepsilon}\}}]\right)^2 + \mathbb{E}[\xi^2\mathbf{1}_{\{|\xi|> n^{1/2-\varepsilon}\}}]\\&= \left|\mathbb{E}[\xi\mathbf{1}_{\{|\xi|> n^{1/2-\varepsilon}\}}]\right|^2 + \mathbb{E}[\xi^2\mathbf{1}_{\{|\xi|> n^{1/2-\varepsilon}\}}]\\&\leq 2\mathbb{E}\left[\frac{|\xi|^4}{n^{1-2\varepsilon}}\mathbf{1}_{\{|\xi|> n^{1/2-\varepsilon}\}}\right]\\&= o(n^{-1-2\varepsilon}).\end{aligned}$$

Next we move onto *(ii)*. By construction, $\mathbb{E}[\hat{\xi}] = 0$ and $\operatorname{Var}(\hat{\xi}) = 1$ provided $n$ is sufficiently large. By part *(i)*,

$$1 - \frac{C}{n^{1+2\varepsilon}} \leq \operatorname{Var}(\tilde{\xi})$$

for some constant $C > 0$ so choosing $N_0 > \left(\frac{4C}{3}\right)^{1/(1+2\varepsilon)}$ ensures that $\frac{1}{4} \leq \mathrm{Var}(\tilde{\xi})$, which gives $2 \geq \left(\mathrm{Var}(\tilde{\xi})\right)^{-1/2}$ for $n > N_0$. With such an $n > N_0$,

$$\begin{aligned}\left|\hat{\xi}\right| &= \left|\frac{\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}} - \mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}}\right]}{\sqrt{\mathrm{Var}(\tilde{\xi})}}\right| \\ &\leq 2\left|\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}}\right| + 2\left|\mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}}\right]\right| \\ &\leq 4n^{1/2-\varepsilon}\end{aligned}$$

almost surely. For part *(iii)*, we have

$$\begin{aligned}\mathbb{E}|\hat{\xi}|^4 &= \mathbb{E}\left|\frac{\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}} - \mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}}\right]}{\sqrt{\mathrm{Var}(\tilde{\xi})}}\right|^4 \\ &\leq 2^4 \mathbb{E}\left|\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}} - \mathbb{E}\left[\xi \mathbf{1}_{\{|\xi| \leq n^{1/2-\varepsilon}\}}\right]\right|^4 \\ &\leq 2^8 \mathbb{E}\left|\xi\right|^4\end{aligned}$$

completing the proof of the claim. □

## Appendix B. Largest and Smallest Singular Values

In this section, we consider events concerning the largest and smallest singular values for the random matrices appearing in this paper. These results are included as an appendix because the methods used to prove them are slight modifications of those in [23, 48, 52]. In order to prove these results, we need to introduce an intermediate truncation of the matrices. Specifically, let $\xi_1, \xi_2, \ldots \xi_m$ be real-valued random variables each having mean zero, variance one, and finite $4+\tau$ moment for some $\tau > 0$. Let $X_{n,1} X_{n,2}, \ldots X_{n,m}$ be independent iid $n \times n$ random matrices with atom random variables $\xi_1, \xi_2, \ldots \xi_m$ respectively. For a fixed $\varepsilon > 0$, and for each $1 \leq k \leq m$, define truncated random variables (at $n^{1/2-\varepsilon}$) $\tilde{\xi}_k$ and $\hat{\xi}_k$ as in (19). Also define truncated matrices $\tilde{X}_{n,k}$ and $\hat{X}_{n,k}$ as in (21) and (22) respectively. Define the linearized truncated matrix $\mathcal{Y}_n$ as in (31). Also recall that $P_n = n^{-m/2} X_{n,1} X_{n,2} \cdots X_{n,m}$ and $\hat{P}_n = n^{-m/2} \hat{X}_{n,1} \hat{X}_{n,2} \cdots \hat{X}_{n,m}$.

Let $X$ be an $n \times n$ random matrix filled with iid copies of a random variable $\xi$ which has mean zero, unit variance, and finite $4+\tau$ moment. For a fixed constant $L > 0$, define matrices $\mathring{X}$ and $\check{X}$ to be the $n \times n$ matrices with entries defined by

$$\mathring{X}_{(i,j)} := X_{(i,j)} \mathbf{1}_{\{|X_{(i,j)}| \leq L/\sqrt{2}\}} - \mathbb{E}\left[X_{(i,j)} \mathbf{1}_{\{|X_{(i,j)}| \leq L/\sqrt{2}\}}\right] \tag{109}$$

and

$$\check{X}_{(i,j)} := \frac{\mathring{X}_{(i,j)}}{\sqrt{\mathrm{Var}(\mathring{X}_{(i,j)})}} \tag{110}$$

for $1 \leq i, j \leq n$. Define $\mathring{X}_{n,1}, \mathring{X}_{n,2}, \dots \mathring{X}_{n,m}$ and $\check{X}_{n,1}, \check{X}_{n,2}, \dots \check{X}_{n,m}$ as in (109) and (110) respectively. Finally, define the linearized truncated matrix

$$\check{\mathcal{Y}}_n := n^{-1/2} \begin{bmatrix} 0 & \check{X}_{n,1} & 0 & \cdots & 0 \\ 0 & 0 & \check{X}_{n,2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \check{X}_{n,m-1} \\ \check{X}_{n,m} & 0 & 0 & \dots & 0 \end{bmatrix}. \tag{111}$$

**Lemma B.1.** *Fix $\varepsilon > 0$. For a fixed integer $m > 0$, let $\xi_1, \xi_2, \dots \xi_m$ be real-valued random variables each mean zero, variance one, and finite $4 + \tau$ moment for some $\tau > 0$. Let $\hat{X}_{n,1}, \hat{X}_{n,2}, \dots, \hat{X}_{n,m}$ be independent iid random matrices with atom variables as defined in (22), and define $\mathcal{Y}_n$ as in (31). For every $\delta > 0$, there exists a constant $c > 0$ depending only on $\delta$ such that*

$$\inf_{|z|>1+\delta/2} s_{mn}\left(\mathcal{Y}_n - zI\right) \geq c$$

*with overwhelming probability.*

*Proof.* Fix $\delta > 0$ and define $\check{\mathcal{Y}}_n$ as in (111). By [23, Lemma 8.1], which is based on techniques in [48, 49], we know that there exists a constant $c' > 0$ which depends only on $\delta$ such that $\inf_{|z|>1+\delta/2} s_{mn}\left(\check{\mathcal{Y}}_n - zI\right) \geq c'$ with overwhelming probability. Note that by Weyl's inequality (13),

$$\sup_{z \in \mathcal{C}} \left| s_{mn}\left(\check{\mathcal{Y}}_n - zI\right) - s_{mn}\left(\mathcal{Y}_n - zI\right) \right| \leq \left\| \check{\mathcal{Y}}_n - \mathcal{Y}_n \right\| \leq \max_{1 \leq k \leq m} \frac{1}{\sqrt{n}} \left\| \check{X}_{n,k} - \hat{X}_{n,k} \right\|. \tag{112}$$

Focusing on an arbitrary value of $k$, we have

$$\frac{1}{\sqrt{n}} \left\| \check{X}_{n,k} - \hat{X}_{n,k} \right\| \leq \frac{1}{\sqrt{n}} \left\| \frac{\mathring{X}_{n,k}}{\sqrt{\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})}} - \frac{\tilde{X}_{n,k}}{\sqrt{\operatorname{Var}((\tilde{X}_{n,k})_{(i,j)})}} \right\|$$

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

$$\frac{1}{\sqrt{n}} \left\| \frac{\mathring{X}_{n,k}}{\sqrt{\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})}} - \mathring{X}_{n,k} \right\| = \frac{1}{\sqrt{n}} \left\| \frac{\mathring{X}_{n,k}\left(1 - \sqrt{\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})}\right)}{\sqrt{\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})}} \right\|.$$

By [23, Lemma 7.1], $\left(\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})\right)^{-1/2} \leq 2$ for $L$ sufficiently large. Additionally, an argument similar to that of [23, Lemma 7.1] shows that $\left|1 - \sqrt{\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})}\right| \leq \frac{C}{L^2}$ for any $1 \leq i, j \leq n$ and some constant $C > 0$. Therefore by [62, Theorem 1.4], for $L$ sufficiently large,

$$\frac{1}{\sqrt{n}} \left\| \frac{\mathring{X}_{n,k}}{\sqrt{\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})}} - \mathring{X}_{n,k} \right\| \leq \frac{C}{L^2\sqrt{n}} \left\| \frac{\mathring{X}_{n,k}}{\sqrt{\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})}} \right\| \leq \frac{c'}{16}$$

with overwhelming probability. Similarly,

$$\frac{1}{\sqrt{n}}\left\|\frac{\tilde{X}_{n,k}}{\sqrt{\operatorname{Var}((\tilde{X}_{n,k})_{(i,j)})}}-\tilde{X}_{n,k}\right\| = \frac{1}{\sqrt{n}}\left\|\frac{\tilde{X}_{n,k}\left(1-\sqrt{\operatorname{Var}((\tilde{X}_{n,k})_{(i,j)})}\right)}{\sqrt{\operatorname{Var}((\tilde{X}_{n,k})_{(i,j)})}}\right\|.$$

By the arguments to prove part *(ii)* of Lemma 4.3, $\left(\operatorname{Var}((\tilde{X}_{n,k})_{(i,j)})\right)^{-1/2} \leq 2$ for $n$ sufficiently large. Also, by part *(i)* of Lemma 4.3, we can show that $\left|1-\sqrt{\operatorname{Var}((\tilde{X}_{n,k})_{(i,j)})}\right| = o(n^{-1+2\varepsilon})$. Therefore by [13, Theorem 5.9],

$$\frac{1}{\sqrt{n}}\left\|\frac{\tilde{X}_{n,k}}{\sqrt{\operatorname{Var}((\tilde{X}_{n,k})_{(i,j)})}}-\tilde{X}_{n,k}\right\| = o(n^{-1-2\varepsilon})\frac{1}{\sqrt{n}}\left\|\tilde{X}_{n,k}\right\| \leq \frac{c'}{16}$$

with overwhelming probability. Ergo, by the triangle inequality, for $L$ sufficiently large,

$$\begin{aligned}\frac{1}{\sqrt{n}}\left\|\check{X}_{n,k}-\hat{X}_{n,k}\right\| &\leq \frac{1}{\sqrt{n}}\left\|\frac{\mathring{X}_{n,k}}{\sqrt{\operatorname{Var}((\mathring{X}_{n,k})_{(i,j)})}}-\frac{\tilde{X}_{n,k}}{\sqrt{\operatorname{Var}((\tilde{X}_{n,k})_{(i,j)})}}\right\| \\ &\leq \frac{c'}{8}+\frac{1}{\sqrt{n}}\left\|\mathring{X}_{n,k}-\tilde{X}_{n,k}\right\| \qquad (113)\end{aligned}$$

with overwhelming probability.

Now, recall that the entries of $\mathring{X}_{n,k}$ are truncated at level $L$ for a fixed $L > 0$ so for sufficiently large $n$, $L \leq n^{1/2-\varepsilon}$. Note that if all entries are less than $L$ in absolute value, then the entries in $\mathring{X}_{n,k}$ and $\tilde{X}_n$ agree. Similarly, if all entries are greater than $n^{1/2-\varepsilon}$ then the entries in $\mathring{X}_{n,k}$ and $\tilde{X}_n$ agree. Ergo, we need only consider the case when there exists some entries $1 \leq i, j \leq n$ such that $L \leq |(\tilde{X}_{n,k})_{i,j}| \leq n^{1/2-\varepsilon}$. For each $1 \leq k \leq m$, define the random variables

$$\dot{\xi}_k := \xi_k \mathbf{1}_{\{L\leq|\xi_k|\leq n^{1/2-\varepsilon}\}} - \mathbb{E}\left[\xi_k \mathbf{1}_{\{L\leq|\xi_k|\leq n^{1/2-\varepsilon}\}}\right]$$

and define $\dot{X}_{n,k}$ to be the matrix with entries

$$(\dot{X}_{n,k})_{(i,j)} := (X_{n,k})_{(i,j)}\mathbf{1}_{\{L\leq|(X_{n,k})_{(i,j)}|\leq n^{1/2-\varepsilon}\}} - \mathbb{E}\left[(X_{n,k})_{(i,j)}\mathbf{1}_{\{L\leq|(X_{n,k})_{(i,j)}|\leq n^{1/2-\varepsilon}\}}\right].$$

for $1 \leq i, j \leq n$. Note that the definitions of $\dot{\xi}$ and $\dot{X}_{n,k}$ differs from the definitions in Section 4. We will use the definition given in this appendix for the remainder of this proof. We can write

$$\frac{1}{\sqrt{n}}\left\|\mathring{X}_{n,k}-\tilde{X}_{n,k}\right\| = \frac{1}{\sqrt{n}}\left\|\dot{X}_{n,k}\right\|.$$

By [13, Lemma 5.9], for $L$ sufficiently large

$$\frac{1}{\sqrt{n}}\left\|\dot{X}_{n,k}\right\| \leq \frac{c'}{8} \qquad (114)$$

with overwhelming probability. Thus, by choosing $L$ large enough to satisfy both conditions, by (113) and (114),

$$\max_{1\leq k\leq m}\frac{1}{\sqrt{n}}\left\|\check{X}_{n,k}-\hat{X}_{n,k}\right\| < \frac{c'}{4}$$

with overwhelming probability. By recalling (112), this implies that, for $L$ sufficiently large,

$$\inf_{|z|>1+\delta/2} s_{mn}\left(\mathcal{Y}_n - zI\right) \geq c$$

with overwhelming probability where $c = \frac{c'}{2}$. ☐

**Lemma B.2.** *Fix $\varepsilon > 0$. For a fixed integer $m > 0$, let $\xi_1, \xi_2, \dots \xi_m$ be real-valued random variables each mean zero, variance one, and finite $4+\tau$ moment for some $\tau > 0$. Let $X_{n,1}, X_{n,2}, \dots, X_{n,m}$ be independent iid random matrices with atom variables $\xi_1, \xi_2, \dots, \xi_m$ respectively. Define $\hat{X}_{n,1}, \hat{X}_{n,2}, \dots \hat{X}_{n,m}$ as in (22), and define $\hat{P}_n$ as in (24). For any $\delta > 0$, there exists a constant $c > 0$ depending only on $\delta$ such that*

$$\inf_{|z|>1+\delta/2} s_{mn}\left(\hat{P}_n - zI\right) \geq c$$

*with overwhelming probability.*

*Proof.* Fix $\delta > 0$. By Lemma B.1, we know that there exists some $c' > 0$ such that $\inf_{|z|>1+\delta/2} s_{mn}\left(\mathcal{Y}_n - zI\right) \geq c'$ with overwhelming probability as well. Recall that $s_{mn}\left(\mathcal{Y}_n - zI\right) = s_1\left(\left(\mathcal{Y}_n - zI\right)^{-1}\right)$ provided $z$ is not an eigenvalue of $\mathcal{Y}_n$. A block inverse matrix calculation reveals that

$$\left(\left(\mathcal{Y}_n - zI\right)^{-1}\right)^{[1,1]} = z^{m-1}\left(\hat{P}_n - z^m I\right)^{-1}$$

where the notation $A^{[1,1]}$ denotes the upper left $n \times n$ block of $A$. Therefore,

$$\frac{1}{c'} \geq \sup_{|z|>1+\delta/2} s_1\left(\left(\mathcal{Y}_n - zI\right)^{-1}\right) \geq \sup_{|z|>1+\delta/2} |z|^{m-1}\left\|\left(\hat{P}_n - z^m I\right)^{-1}\right\|.$$

This implies that there exists a constant $c > 0$ such that

$$\frac{1}{c} \geq \sup_{|z|>1+\delta/2} s_1\left(\left(\hat{P}_n - zI\right)^{-1}\right)$$

with overwhelming probability. This gives $\inf_{|z|>1+\delta/2} s_n\left(\hat{P}_n - zI\right) \geq c$ with overwhelming probability. ☐

**Lemma B.3.** *For a fixed integer $m > 0$, let $\xi_1, \xi_2, \dots \xi_m$ be real-valued random variables each satisfying Assumption 2.1. Fix $\delta > 0$ and let $X_{n,1}, X_{n,2}, \dots X_{n,m}$ be independent iid random matrices with atom variables $\xi_1, \xi_2, \dots \xi_m$ respectively. Then there exists a constant $c > 0$ depending only on $\delta$ such that*

$$\inf_{|z|>1+\delta/2} s_n\left(P_n/\sigma - zI\right) \geq c$$

*with probability $1 - o(1)$ where $\sigma = \sigma_1 \cdots \sigma_m$.*

*Proof.* By a simple rescaling, it is sufficient to assume that the variance of each random variable is 1 so that $\sigma = 1$. Let $\delta > 0$ and recall by Lemma B.2 there exists a $c' > 0$ depending only on $\delta$ such that $\inf_{|z|>1+\delta/2} s_n\left(\hat{P}_n - zI\right) \geq c'$ with overwhelming probability. Then by Lemma 4.10,

$$\begin{aligned}
&\mathbb{P}\left(\inf_{|z|>1+\delta/2} s_n\left(P_n - zI\right) < \frac{c'}{2}\right) \\
&\quad = \mathbb{P}\left(\inf_{|z|>1+\delta/2} s_n\left(P_n - zI\right) < \frac{c'}{2} \quad \text{and} \quad \left\|P_n - \hat{P}_n\right\| \le n^{-\varepsilon}\right) \\
&\qquad + \mathbb{P}\left(\inf_{|z|>1+\delta/2} s_n\left(P_n - zI\right) < \frac{c'}{2} \quad \text{and} \quad \left\|P_n - \hat{P}_n\right\| > n^{-\varepsilon}\right) \\
&\quad \le \mathbb{P}\left(\inf_{|z|>1+\delta/2} s_n\left(P_n - zI\right) < \frac{c'}{2} \quad \text{and} \quad \left\|P_n - \hat{P}_n\right\| \le n^{-\varepsilon}\right) \\
&\qquad + \mathbb{P}\left(\left\|P_n - \hat{P}_n\right\| > n^{-\varepsilon}\right) \\
&\quad \le \mathbb{P}\left(\inf_{|z|>1+\delta/2} s_n\left(P_n - zI\right) < \frac{c'}{2} \quad \text{and} \quad \left\|P_n - \hat{P}_n\right\| \le n^{-\varepsilon}\right) + o(1).
\end{aligned}$$

Suppose that there exists a $z_0 \in \mathbb{C}$ with $|z_0| \ge 1 + \delta/2$ such that $s_n\left(P_n - z_0 I\right) < \frac{c'}{2}$ and $\left\|P_n - \hat{P}_n\right\| < n^{-\varepsilon} < \frac{c'}{2}$. Then, by Weyl's inequality (13), $\left|s_n(P_n - z_0 I) - s_n(\hat{P}_n - z_0 I)\right| < \frac{c'}{2}$ which implies $s_n(\hat{P}_n - z_0 I) < c'$. Thus, for $n$ sufficiently large to ensure that $n^{-\varepsilon} < \frac{c'}{2}$, by Lemma 4.10

$$\mathbb{P}\left(\inf_{|z|>1+\delta/2} s_n\left(P_n - zI\right) < \frac{c'}{2}\right) \le \mathbb{P}\left(\inf_{|z|>1+\delta/2} s_n\left(\hat{P}_n - zI\right) < c'\right) + o(1).$$

Thus, selecting $c = \frac{c'}{2}$, we have $\inf_{|z|>1+\delta/2} s_n\left(P_n - zI\right) \ge c$ with probability $1 - o(1)$. □

**Lemma B.4.** *Let $A$ be an $n \times n$ matrix. Let $R$ be a subset of the integer set $\{1, 2, \dots n\}$. Let $A^{(R)}$ denote the matrix $A$, but with the $r$th column replaced with zero for each $r \in R$. Then*

$$s_n\left(A^{(R)} - zI\right) \ge \min\{s_n(A - zI), |z|\}.$$

*Proof.* Let $A^{((R))}$ denote the matrix $A$ with column $r$ removed for all $r \in R$. Note that $A^{((R))}$ is an $n \times (n - |R|)$ matrix, which is distinct from the $n \times n$ matrix $A^{(R)}$. Also, let $I^{((R))}$ denote the $n \times n$ identity matrix with column $r$ removed for all $r \in R$. In order to bound the least singular value of $(A^{(R)} - zI)$, we will consider the eigenvalues of $(A - zI)^* (A - zI)$, $\left(A^{(R)} - zI\right)^* \left(A^{(R)} - zI\right)$, and $\left(A^{((R))} - zI^{((R))}\right)^* \left(A^{((R))} - zI^{((R))}\right)$.

Now, observe that $\left(A^{((R))} - zI^{((R))}\right)^* \left(A^{((R))} - zI^{((R))}\right)$ is an $(n-|R|) \times (n-|R|)$ matrix, and is a principle sub-matrix of the Hermitian matrix $(A - zI)^*(A - zI)$. Therefore, the eigenvalues of $\left(A^{((R))} - zI^{((R))}\right)^* \left(A^{((R))} - zI^{((R))}\right)$ must interlace with the eigenvalues of $(A - zI)^* (A - zI)$ by Cauchy's interlacing theorem [38, Theorem 1]. This implies

$$s_n\left(A^{((R))} - zI^{((R))}\right)^2 \ge s_n\left(A - zI\right)^2.$$

Next, we compare the eigenvalues of $\left(A^{(R)} - zI\right)^* \left(A^{(R)} - zI\right)$ to the eigenvalues of $\left(A^{((R))} - zI^{((R))}\right)^* \left(A^{((R))} - zI^{((R))}\right)$. Note that, after a possible permutation of columns to move all zero columns of $A^{(R)}$ to be in the last $|R|$ columns, the product $\left(A^{(R)} - zI\right)^* \left(A^{(R)} - zI\right)$ becomes

$$\begin{bmatrix} \left(A^{((R))} - zI^{((R))}\right)^* \left(A^{((R))} - zI^{((R))}\right) & 0 \cdot I_{|R| \times (n-|R|)} \\ 0 \cdot I_{(n-|R|) \times |R|} & |z|^2 \cdot I_{|R| \times |R|} \end{bmatrix}.$$

Due to the block structure of the matrix above, if $w$ is an eigenvalue of $\left(A^{(R)} - zI\right)^* \left(A^{(R)} - zI\right)$, then either $w$ is an eigenvalue of $\left(A^{((R))} - zI^{((R))}\right)^* \left(A^{((R))} - zI^{((R))}\right)$ or $w$ is $|z|^2$. Ergo,

$$\begin{aligned} s_n \left(A^{(R)} - zI\right)^2 &= \min \left\{ s_n \left(A^{((R))} - zI^{((R))}\right)^2, \ |z|^2 \right\} \\ &\geq \min \left\{ s_n \left(A - zI\right)^2, \ |z|^2 \right\} \end{aligned}$$

which implies $s_n \left(A^{(R)} - zI\right) \geq \min \left\{ s_n \left(A - zI\right), \ |z| \right\}$ concluding the proof. $\square$

This lemma gives way to the following two corollaries.

**Corollary B.5.** *Fix $\varepsilon > 0$. For a fixed integer $m > 0$, let $\xi_1, \xi_2, \dots \xi_m$ be real-valued random variables each mean zero, variance one, and finite $4 + \tau$ moment for some $\tau > 0$. Let $X_{n,1}, X_{n,2}, \dots, X_{n,m}$ be independent iid random matrices with atom variables $\xi_1, \xi_2, \dots, \xi_m$ respectively, and define $\hat{X}_{n,1}, \hat{X}_{n,2}, \dots \hat{X}_{n,m}$ as in (22). Define $\mathcal{Y}_n$ as in (31) and $\mathcal{Y}_n^{(k)}$ as $\mathcal{Y}_n$ with the columns $c_k, c_{n+k}, c_{2n+k}, \dots, c_{(m-1)n+k}$ replaced with zeros. For any $\delta > 0$, there exists a constant $c > 0$ depending only on $\delta$ such that*

$$\inf_{|z| > 1 + \delta/2} s_{mn} \left( \mathcal{Y}_n^{(k)} - zI \right) \geq c$$

*with overwhelming probability.*

*Proof.* Note that by Lemmas B.1 and B.4,

$$\begin{aligned} \inf_{|z| > 1 + \delta/2} s_{mn} \left( \mathcal{Y}_n^{(k)} - zI \right) &\geq \inf_{|z| > 1 + \delta/2} \min \left\{ s_{mn} \left( \mathcal{Y}_n - zI \right), \ |z| \right\} \\ &\geq \inf_{|z| > 1 + \delta/2} \min \left\{ s_{mn} \left( \mathcal{Y}_n - zI \right), \ 1 \right\} \\ &\geq \min \left\{ c', \ 1 \right\} \end{aligned}$$

with overwhelming probability for some constant $c' > 0$ depending only on $\delta$. The result follows by setting $c = \min \{c', \ 1\}$. $\square$

**Corollary B.6.** *Fix $\varepsilon > 0$. For a fixed integer $m > 0$, let $\xi_1, \xi_2, \dots \xi_m$ be real-valued random variables each mean zero, variance one, and finite $4 + \tau$ moment for some $\tau > 0$. Let $\hat{X}_{n,1}, \hat{X}_{n,2}, \dots, \hat{X}_{n,m}$ be independent iid random matrices with atom variables as defined in (22). Define $\mathcal{Y}_n$ as in (31) and $\mathcal{Y}_n^{(k,s)}$ as $\mathcal{Y}_n$ with the columns $c_k, c_{n+k}, c_{2n+k}, \dots, c_{(m-1)n+k}$ and $c_s$ replaced with zeros. For any $\delta > 0$, there exists a constant $c > 0$ depending only on $\delta$ such that*

$$\inf_{|z| > 1 + \delta/2} s_{mn} \left( \mathcal{Y}_n^{(k,s)} - zI \right) \geq c$$

*with overwhelming probability.*

The proof of Corollary B.6 follows in exactly the same way as the proof of Corollary B.5.

## Appendix C. Useful Lemmas

**Lemma C.1** (Lemma 2.7 from [12]). *For $X = (x_1, x_2, \ldots, x_N)^T$ iid standardized complex entries, $B$ an $N \times N$ complex matrix, we have, for any $p \geq 2$,*

$$\mathbb{E}\,|X^*BX - \operatorname{tr}(B)|^p \leq K_p \left( \left( \mathbb{E}\,|x_1|^4 \operatorname{tr} B^*B \right)^{p/2} + \mathbb{E}|x_1|^{2p} \operatorname{tr}(B^*B)^{p/2} \right)$$

*where the constant $K_p > 0$ depends only on $p$.*

**Lemma C.2.** *Let $A$ be an $N \times N$ complex-valued matrix. Suppose that $\xi$ is a complex-valued random variable with mean zero and unit variance. 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 $\xi$,*
*(ii) $w_i = 0$ for $i \notin S$.*

*Additionally, $A_{S\times S}$ denote the $|S| \times |S|$ matrix which has entries $A_{(i,j)}$ for $i, j \in S$. Then for any even $p \geq 2$,*

$$\mathbb{E}\,|w^*Aw - \operatorname{tr}(A_{S\times S})|^p \ll_p \mathbb{E}\,|\xi|^{2p} \left(\operatorname{tr}(A^*A)\right)^{p/2}.$$

*Proof.* Let $w_S$ denote the $|S|$-vector which contains entries $w_i$ for $i \in S$ and observe

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

Therefore, by Lemma C.1, for any even $p \geq 2$,

$$\begin{aligned}\mathbb{E}\,|w^*Aw - \operatorname{tr}(A_{S\times S})|^p &= \mathbb{E}\,|w_S^* A_{S\times S} w_S - \operatorname{tr}(A_{S\times S})|^p \\ &\ll_p \left( \mathbb{E}\,|\xi|^4 \operatorname{tr}(A_{S\times S}^* A_{S\times S}) \right)^{p/2} + \mathbb{E}\,|\xi|^{2p} \operatorname{tr}(A_{S\times S}^* A_{S\times S})^{p/2} \\ &\ll_p \mathbb{E}\,|\xi|^{2p} \left( \operatorname{tr}(A_{S\times S}^* A_{S\times S}) \right)^{p/2}.\end{aligned}$$

Now observe that

$$\operatorname{tr}(A_{S\times S}^* A_{S\times S}) = \sum_{i,j\in S} A_{i,j}^* A_{j,i} \leq \sum_{i,j=1}^N A_{i,j}^* A_{j,i} = \operatorname{tr}(A^*A).$$

Therefore

$$\mathbb{E}\,|w^*Aw - \operatorname{tr}(A_{S\times S})|^p \ll_p \mathbb{E}\,|\xi|^{2p} \left( \operatorname{tr}(A_{S\times S}^* A_{S\times S}) \right)^{p/2} \leq \mathbb{E}\,|\xi|^{2p} \left( \operatorname{tr}(A^*A) \right)^{p/2}.$$

□

**Lemma C.3** (Lemma A.1 from [12]). *For $X = (x_1, x_2, \ldots, x_N)^T$ iid standardized complex entries, $B$ an $N\times N$ complex-valued Hermitian nonnegative definite matrix, we have, for any $p \geq 1$,*

$$\mathbb{E}\,|X^*BX|^p \leq K_p \left( (\operatorname{tr} B)^p + \mathbb{E}|x_1|^{2p} \operatorname{tr} B^p \right).$$

*where $K_p > 0$ depends only on $p$.*

**Lemma C.4.** *Let $A$ be an $N \times N$ Hermitian positive semidefinite matrix. Suppose that $\xi$ is a complex-valued random variable with mean zero and unit variance. 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 $\xi$,*
*(ii) $w_i = 0$ for $i \notin S$.*

*Then for any $p \geq 2$,*

$$\mathbb{E}\left|w^* A w\right|^p \ll_p \mathbb{E}|\xi|^{2p} (\operatorname{tr} A)^p . \tag{115}$$

*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 have

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

By Lemma C.3, we get

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

Since $A$ is non-negative definite, the diagonal elements are non-negative so that $\operatorname{tr}(A_{S\times S}^p) \leq (\operatorname{tr}(A_{A\times A}))^p$. By this and the fact that for a Hermitian positive semi-definite matrix, the partial trace is less than or equal to the full trace, we observe that

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

□

**Lemma C.5.** *Let $A$ and $B$ be $n \times n$ matrices. Then*

$$|\operatorname{tr}(AB)| \leq \sqrt{n} \left\|AB\right\|_2 \leq \sqrt{n} \left\|A\right\| \cdot \left\|B\right\|_2 .$$

*Proof.* This follows by an application of the Cauchy–Schwarz inequality and an application of [13, Theorem A.10]. □

## References


[1] K. Adhikari, N. Kishore Reddy, T. Ram Reddy, K. Saha, *Determinantal point processes in the plane from products of random matrices*, Ann. Inst. Henri Poincaré Probab. Stat., 52(1):1646, 2016.

[2] G. Akemann, Z. Burda, *Universal microscopic correlation functions for products of independent Ginibre matrices*, J. Phys. A: Math. Theor. **45** (2012).

[3] G. Akemann, Z. Burda, M. Kieburg, *Universal distribution of Lyapunov exponents for products of Ginibre matrices*, J. Phys. A: Math. Theor. 47(39) (2014).

[4] G. Akemann, J. R. Ipsen, M. Kieburg, *Products of rectangular random matrices: singular values and progressive scattering*, Phys. Rev. E **88**, (2013).

[5] G. Akemann, J. R. Ipsen, E. Strahov, *Permanental processes from products of complex and quaternionic induced Ginibre ensembles*, Random Matrices: Theory and Applications Vol. 3, No. 4 (2014) 1450014.

[6] G. Akemann, M. Kieburg, L. Wei, *Singular value correlation functions for products of Wishart random matrices*, J. Phys. A: Math. Theor. **46** (2013).

[7] G. Akemann, E. Strahov, *Hole probabilities and overcrowding estimates for products of complex Gaussian matrices*, J. Stat. Phys. (2013), Volume 151, Issue 6, pp 987–1003.

[8] G. Anderson, *Convergence of the largest singular value of a polynomial in independent Wigner matrices*, Ann. Probab., Volume 41, Number 3B (2013), 2103–2181.

[9] G. Anderson, O. Zeitouni, *CLT for a band matrix model*, Probab. Theory and Related Fields, vol. 134 (2006), 283–338.

[10] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), 494–529.

[11] Z. D. Bai, J.W. Silverstein, *CLT for linear spectral statistic of large-dimensional sample covariance matrix*, Ann. Probab. Volume 32 (2004), 553-605.

[12] Z. D. Bai, J. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab. Volume 26, Number 1 (1998), 316-345.

[13] Z. D. Bai, J. Silverstein, *Spectral analysis of large dimensional random matrices*, Mathematics Monograph Series **2**, Science Press, Beijing 2006.
[14] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics, Springer-Verlag New York, 1997.
[15] P. Billingsley, *Probability and Measure*, 3rd Edition, Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1995.
[16] P. Billingsley, *Convergence of Probability Measures*, 1st Edition, John Wiley and Sons, Inc. 1968.
[17] C. Bordenave, *On the spectrum of sum and product of non-hermitian random matrices*, Elect. Comm. in Probab. **16** (2011), 104–113.
[18] C. Bordenave, D. Chafaï, *Around the circular law*, Probability Surveys **9** 1–89, (2012).
[19] Z. Burda, R. A. Janik, B. Waclaw, *Spectrum of the product of independent random Gaussian matrices*, Phys. Rev. E **81** (2010).
[20] Z. Burda, A. Jarosz, G. Livan, M. A. Nowak, A. Swiech, *Eigenvalues and singular values of products of rectangular Gaussian random matrices*, Phys. Rev. E **82** (2010).
[21] Z. Burda, M. A. Nowak, A. Swiech, *Spectral relations between products and powers of isotropic random matrices*, Phys. Rev. E **86**, 061137 (2012)
[22] Z. Burda, *Free products of large random matrices - a short review of recent developments*, J. Phys.: Conf. Ser. **473** 012002 (2013).
[23] N. Coston, S. O'Rourke, P. Wood, *Outliers in the spectrum for products of independent random matrices*, available at `arXiv:1711.07420`.
[24] C. Y. Deng, *A generalization of the Sherman–Morrison–Woodbury formula*, Applied Mathematics Letters, Volume 24, Issue 9, Sept. 2011, 1561–1564.
[25] P. Diaconis, M. Shahshahani, *On the eigenvalues of random matrices*, J. Appl. Probab. 31A (1994), 49-62.
[26] P. Diaconis, S.N. Evans, *Linear functionals of eigenvalues of random matrices*, Trans. Amer. Math. Soc. vol. 353, no. 7 (2001), 2615–2633.
[27] A. Edelman, *The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law*, J. Multivariate Anal. **60**, 203–232 (1997).
[28] P. J. Forrester, *Lyapunov exponents for products of complex Gaussian random matrices*, J. Stat. Phys. **151**, 796-808 (2013).
[29] P. J. Forrester, *Probability of all eigenvalues real for products of standard Gaussian matrices*, J. Phys. A, vol. 47, 065202 (2014).
[30] J. Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, J. Math. Phys. **6** (1965), 440–449.
[31] V. L. Girko, *Circular law*, Theory Probab. Appl. **29** (1984), 694–706.
[32] V. L. Girko, *The circular law*, Teor. Veroyatnost. i Primenen. **29** (4): 669–679 (1984).
[33] V. L. Girko, A. Vladimirova, *L.I.F.E.: and Halloween Law*, Random Operators and Stochastic Equations, 18(4), pp. 327-353 (2010).
[34] F. Götze, A. Naumov, T. Tikhomirov, *On local laws for non-Hermitian random matrices and their products*, Doklady Mathematics. 96. 10.1134/S1064562417060072.
[35] F. Götze, T. Tikhomirov, *The circular law for random matrices*, Ann. Probab. Volume 38, Number 4 1444–1491 (2010).
[36] F. Götze, T. Tikhomirov, *On the asymptotic spectrum of products of independent random matrices*, available at `arXiv:1012.2710`.
[37] R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press (1991).
[38] S. Hwang, *Cauchy's interlace theorem for eigenvalues of Hermitian matrices*, The American Mathematical Monthly, vol. 111, no. 2, 2004, pp. 157-159.
[39] J.R. Ipsen, *Products of independent Gaussian random matrices*, Bielefeld: Bielefeld University; 2015.
[40] J.R. Ipsen, M. Kieburg, *Weak commutation relations and eigenvalue statistics for products of rectangular random matrices*, Phys. Rev. E, 89:032106, Mar 2014.
[41] K. Johansson, *On fluctuations of eigenvalues of random Hermitian matrices*, Duke Math. J., vol. 91 (1998), 151–204.
[42] P. Kopel, *Linear Statistics of Non-Hermitian Matrices Matching the Real or Complex Ginibre Ensemble to Four Moments*, available at `arXiv:1510.02987 [math.PR]`.
[43] P. Kopel, S. O'Rourke, V. Vu *Random matrix products: Universality and least singular values*, available at `arXiv 1802.03004`.

[44] A. B. J. Kuijlaars, L. Zhang, *Singular values of products of Ginibre random matrices, multiple orthogonal polynomials and hard edge scaling limits*, Communications in Mathematical Physics, Volume 332, Issue 2, pp 759-781 (2014).

[45] A. Lytova, L. Pastur, *Central limit theorem for linear eigenvalue statistics of random matrices with independent entries*, Annals of Probability, vol. 37 (2009), 1778–1840.

[46] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Acad. Press (1967).

[47] M. L. Mehta, *Random Matrices*, third edition. Elsevier/Academic Press, Amsterdam (2004).

[48] Y. Nemish, *No outliers in the spectrum of the product of independent non-Hermitian random matrices with independent entries*, J. Theor. Probab. (2018) 31: 402.

[49] Y. Nemish, *Local law for the product of independent non-Hermitian random matrices with independent entries*, Electron. J. Probab. 22 (2017), no. 22, 1–35.

[50] I. Nourdin, G. Peccati, *Universal Gaussian fluctuations of non-Hermitian matrix ensembles: from weak convergence to almost sure CLTs*, Latin American journal of probability and mathematical statistics 7, 341–375 (2010).

[51] S. O'Rourke, D. Renfrew, *Central limit theorem for linear eigenvalue statistics of elliptic random matrices*, Journal of Theoretical Probability, Volume 29, Issue 3 (2016), pp. 1121–1191.

[52] S. O'Rourke, D. Renfrew, *Low rank perturbations of large elliptic random matrices*, Electronic Journal of Probability Vol. 19, No. 43, 1–65 (2014).

[53] S. O'Rourke, D. Renfrew, A. Soshnikov, V. Vu, *Products of independent elliptic random matrices*, J. Stat. Phys., Vol. 160, No. 1 (2015), 89–119.

[54] S. O'Rourke, A. Soshnikov, *Products of independent non-Hermitian random matrices*, Electronic Journal of Probability, Vol. 16, Art. 81, 2219–2245, (2011).

[55] G. Pan, W. Zhou, *Circular law, extreme singular values and potential theory*, Journal of Multivariate Analysis **101**, 645–656 (2010).

[56] B. Rider, J.W. Silverstein, *Gaussian fluctuations for non-Hermitian random matrix ensembles*, Ann. Probab. 34 (2006), 21182143.

[57] M. Shcherbina, *Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices*, Zh. Mat. Fiz. Anal. Geom., 7:2 (2011), 176–192.

[58] Y. Sinai, A. Soshnikov, *Central limit theorem for traces of large random symmetric matrices with independent matrix elements*, Bol. Soc. Brasil. Mat. (N.S.), vol. 29 (1998), 1–24.

[59] A. Soshnikov, *The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities*, Ann. Probab., vol. 28 (2000), 1353–1370.

[60] P. Sosoe, P. Wong, *Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices*, Advances in Mathematics, Vol. 249, 20, Dec. 2013, pp. 37–87.

[61] E. Strahov, *Differential equations for singular values of products of Ginibre random matrices*, J. Phys. A: Math. Theor. **47** 325203, 2014.

[62] T. Tao, *Outliers in the spectrum of iid matrices with bounded rank perturbations*, Probab. Theory Related Fields **155** (2013), 231–263.

[63] T. Tao, V. Vu, *Random matrices: The circular law*, Communication in Contemporary Mathematics **10** (2008), 261–307.

[64] T. Tao, V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 3, 377–396.

[65] T. Tao, V. Vu, *Random matrices: Universality of ESDs and the circular law*, Ann. Probab. Volume 38, Number **5** (2010), 2023–2065.

Department of Mathematics, University of Colorado at Boulder, Boulder, CO 80309
*E-mail address*: `natalie.coston@colorado.edu`

Department of Mathematics, University of Colorado at Boulder, Boulder, CO 80309
*E-mail address*: `sean.d.orourke@colorado.edu`