Spectral Properties of Products of Independent Random Matrices

Natalie A. Coston
University of Colorado at Boulder, natalie.coston@colorado.edu

Follow this and additional works at: https://scholar.colorado.edu/math_gradetds
Part of the Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation
https://scholar.colorado.edu/math_gradetds/62

This Dissertation is brought to you for free and open access by Mathematics at CU Scholar. It has been accepted for inclusion in Mathematics Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Spectral Properties of Products of Independent Random Matrices

by

N. A. Coston

B.A., Northern Arizona University, 2011
M.S., University of Colorado, Boulder, 2015

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Mathematics

2018
This thesis entitled:
Spectral Properties of Products of Independent Random Matrices
written by N. A. Coston
has been approved for the Department of Mathematics

Dr. Sean O’Rourke

Dr. Philip Matchett Wood

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.
In the field of random matrix theory, there are many matrix models one may choose to study. This thesis focuses on independent and identically distributed (iid) random matrices. Given a random variable $\xi$, we say that a matrix is an iid random matrix if each entry is an iid copy of $\xi$, and we call $\xi$ the atom random variable. Given a positive constant integer $m$, consider $m$ random variables $\xi_1, \ldots, \xi_m$, and create an independent $n \times n$ iid random matrix for each of these random variables. The results presented in this thesis focus on the limiting behavior of the eigenvalues of the product of these $m$ independent iid random matrices. Call this product $P_n$ and note that $P_n$ is also an $n \times n$ random matrix, but the entries are no longer independent.

This thesis is comprised of two main results. First, we examine the locations of eigenvalues of matrix products as the size of the matrices tend to infinity. From the previous results by O’Rourke, Renfrew, Soshnikov, and Vu, we see that under certain moment assumptions on the atom random variables, as $n$ tends to infinity, the empirical spectral measure of the eigenvalues of the rescaled product $n^{-m/2}P_n$ converges to a measure supported on a disk centered at the origin in the complex plane with radius depending on the atom random variables. In this work, we study the asymptotic location of eigenvalues which can fall outside of this disk. These outlying eigenvalues may be present when the random matrix product $P_n$ is additively perturbed by a low rank, small norm deterministic matrix. We also consider multiplicative perturbations, and perturbations in any order. By studying these various perturbations, we characterize when a perturbed matrix product has outliers, and the asymptotic locations of these outlying eigenvalues.

The second result of this thesis studies the fluctuation of the eigenvalues of the rescaled product $n^{-m/2}P_n$. In particular, we define a linear statistic and use this to study the spectrum, or the collection of eigenvalues, of $n^{-m/2}P_n$. We see that the limiting distribution of the unnormalized
linear statistic is a mean-zero Gaussian distribution with variance depending on the the linear statistic. The explicit variance formula is computed as well.
Dedication

To my parents. Thank you for encouraging me to take the road less traveled, for teaching me that the climb is worth the view, and for your support every step of the way.
I would like to first and foremost thank my advisor, Sean O'Rourke, for his years of guidance and his endless patience. His expertise, persistence, and instruction made all of this possible. I would not be where I am today without him, and I am forever grateful.

I would also like to thank Jeff Rushall, for starting me on this path. Jeff saw potential in me, and because of him I better know myself. I am hugely grateful for his influence, and for his unwavering confidence in me.

I would like to thank my family and friends for their support and love, and for making every day better.

Finally, I would like to thank Will Daniels, who shared in every struggle and every victory. Thank you for being my adventure buddy, my breath of fresh air, and my best friend.
Contents

Chapter

1 Introduction and Background  
   1.1 Matrix Ensembles  
      1.1.1 iid Matrices  
      1.1.2 Wigner Matrices  
      1.1.3 Real Elliptic Random Matrices  
   1.2 Limiting Empirical Spectral Measure  
      1.2.1 iid Random Matrices  
      1.2.2 Products of iid Matrices  
      1.2.3 Other Matrix Ensembles  
   1.3 Central Limit Theorems  
      1.3.1 iid Matrices  
      1.3.2 Products of iid Matrices  
      1.3.3 Other Matrix Ensembles  

2 Main Results  
   2.1 Outliers  
      2.1.1 Related Results  
      2.1.2 Open Questions and Applications  
   2.2 Central Limit Theorem  
      2.2.1 Open Questions
# Tools and Notation

3.1 Linearization ............................................. 36
3.2 Matrix Notation ............................................. 37
3.3 Singular Value Inequalities .............................. 38

# Proofs of Theorems in Section 2.1

4.1 Overview and Preliminary Tools ........................ 41
4.2 Isotropic Limit Law and the Proof of Theorems in Section 2.1 ......................... 43
   4.2.1 Proof of Theorem 2.1.2 ............................ 44
   4.2.2 Proof of Theorem 2.1.5 ............................ 44
   4.2.3 Proof of Theorem 2.1.8 ............................ 49
   4.2.4 Proof of Theorems 2.1.6 and 2.1.7 ............... 51
4.3 Truncation and Useful Tools ............................. 55
4.4 Least Singular Value Bounds ............................. 58
4.5 Reductions to the Proof of Theorem 4.2.1 .............. 61
4.6 Concentration of Bilinear Forms Involving the Resolvent $\mathcal{G}_n$ ............... 63
   4.7.1 Neumann Series ..................................... 80
   4.7.2 Removing the Indicator Function .................... 80
   4.7.3 Combinatorial Arguments ........................... 81
4.7 Proof of Theorem 4.5.1 .................................. 79
4.8 Proofs of Results in Section 2.1.1 ....................... 96

# Proof of Theorem 2.2.2

5.1 Overview of the Proof .................................... 101
5.2 Reductions ................................................. 102
   5.2.1 Truncation of the Product ......................... 106
   5.2.2 Truncated Linearization .......................... 114
5.3 Proof of Theorem 5.2.11 ................................. 117
Figures

Figure

1.1 Circular Law and Perturbed Circular Law ............................................. 7
1.2 No Outliers in Matrix Product ................................................................. 9
1.3 Outliers for Elliptic Law .......................................................................... 13
1.4 Product of Elliptic Matrices ....................................................................... 15

2.1 Eigenvalues of a perturbed product of matrices ........................................ 25
2.2 Eigenvalues of a perturbed product of matrices, 2 ..................................... 27
2.3 Simulations of linear statistics ................................................................. 32
2.4 Simulations of linear statistics for complex atom variables ....................... 33

4.1 Path graphs, example 4.7.12 ................................................................. 85
4.2 Path graphs, example 4.7.13 ................................................................. 87
4.3 Path graphs, example 4.7.14 ................................................................. 88
4.4 Path graphs, example 4.7.15 ................................................................. 88
4.5 Path graphs, example 4.7.16 ................................................................. 89

C.1 Path graph proof example .................................................................. 190
Chapter 1

Introduction and Background

The study of random matrices can be traced back to at least 1928 when Wishart studied them with motivations in mathematical statistics [123]. In the 50s and 60s, Wigner and Dyson formalized random matrix ensembles, along with many other fundamental properties [49, 50, 120, 121, 122]. This laid the foundation on which the study of random matrices was built. The decades that followed brought huge developments in the field of random matrix theory by numerous authors.

In this thesis, we study the asymptotic behavior of the spectrum of products of large random matrices. We begin by introducing some classical ensembles of random matrices, and fixing some notation and definitions. Chapter 1 is devoted to previous results and background material. In Chapter 2 we present the main results of this thesis. Chapter 3 introduces some tools and notation used throughout the thesis, and the remainder of the thesis is devoted to the proofs of the main results from Chapter 2.

Throughout this thesis, we use asymptotic notation (such as $O$, $o$) under the assumption that $n \to \infty$; see Chapter 3 for a complete description of the asymptotic notation.

1.1 Matrix Ensembles

The results presented in this thesis focus on an ensemble of random matrices in which each entry is independent, but there are many other interesting random matrix ensembles. Some of the most notable ensembles are defined below.
1.1.1 iid Matrices

The focus of this thesis is the independent and identically distributed (iid) matrix ensemble, in which entries are independent and identically distributed. We define this ensemble formally.

**Definition 1.1.1** (iid random matrix). Let $\xi$ be a real-valued or 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 independent and identically distributed (iid) copies of $\xi$.

There are a few widely studied examples of iid random matrices to make note of. When $\xi$ is a standard complex Gaussian random variable, $X_n$ is known as the complex Ginibre ensemble. The real Ginibre ensemble is defined analogously. When each atom variable $\xi$ is a Bernoulli($-1, 1$) random variable, $X_n$ is known as the Bernoulli ensemble or the random sign ensemble.

1.1.2 Wigner Matrices

The Wigner matrix ensemble dates back to the 1950s when Wigner formalized the ensemble in [120, 121, 122]. This ensemble was foundational to the study of random matrices, and many of the first results about random matrices focused on the Wigner matrix ensemble.

**Definition 1.1.2** (Wigner real symmetric matrix). Let $\xi$ and $\zeta$ be real-valued random variables such that $\mathbb{E}[\xi] = 0, \mathbb{E}[\xi^2] = 1, \mathbb{E}[\zeta] = 0, \mathbb{E}[\zeta^2] < \infty$. We say that $X_n = (x_{ij})_{i,j=1}^n$ is an $n \times n$ Wigner real symmetric random matrix with atom variables $\xi$ and $\zeta$ if the following conditions hold.

- $\{x_{ij} : 1 \leq i < j \leq n\}$ is a collection of iid copies of $\xi$.
- $\{x_{ii} : 1 \leq i \leq n\}$ is a collection of iid copies of $\zeta$.
- $x_{ij} = x_{ji}$ for all $i, j$.

One of the classic examples of this ensemble is the symmetric Bernoulli ensemble, where $\xi$ and $\zeta$ both have the Bernoulli($-1, 1$) distribution. Another example is the Gaussian Orthogonal Ensemble (GOE), where $\xi$ has the distribution $N(0, 1)_\mathbb{R}$ and $\zeta$ has the distribution $N(0, 2)_\mathbb{R}$.

We can extend this to an analogous definition when the atom variables are complex.
Definition 1.1.3 (Wigner complex Hermitian matrix). Let $\xi$ be a complex-valued random variable and let $\zeta$ be a real-valued random variable such that $E[\xi] = 0, E[|\xi|^2] = 1, E[\zeta] = 0, E[\zeta^2] < \infty$. We say that $X_n = (x_{ij})_{i,j=1}^n$ is an $n \times n$ Wigner complex Hermitian random matrix with atom variables $\xi$ and $\zeta$ if the following conditions hold.

- $\{x_{ij} : 1 \leq i < j \leq n\}$ is a collection of iid copies of $\xi$.
- $\{x_{ii} : 1 \leq i \leq n\}$ is a collection of iid copies of $\zeta$.
- $x_{ij} = x_{ji}$ for all $i, j$.

A widely studied example of a Wigner complex Hermitian random matrix ensemble is the Gaussian Unitary Ensemble (GUE), where $\xi$ has the distribution of $\mathcal{N}(0,1)_\mathbb{C}$ and $\zeta$ has the distribution $\mathcal{N}(0,1)_\mathbb{R}$.

Note that, due to the fact that matrices in this ensemble are symmetric or Hermitian (in the real and complex settings respectively), all eigenvalues of a matrix from this ensemble are real. This allows for the use of techniques to study properties of the spectrum that do not generalize to the case when the eigenvalues are complex. For example, the moment method can be used to characterize properties of the spectrum for Wigner matrices, but the technique does not generalize when the spectrum is complex [115]. Due to the availability of techniques to study this ensemble, there is a rich collection of results characterizing local and global properties of the spectrum of Wigner matrices [16, 22, 42, 43, 44, 45, 49, 50, 52, 53, 54, 55, 56, 57, 58, 59, 60, 77, 79, 80, 81, 101, 102, 104, 107, 108, 109, 112, 116, 120, 121, 122].

1.1.3 Real Elliptic Random Matrices

Elliptic random matrices generalize Wigner matrices and iid matrices. In this ensemble, the entries in the lower triangular part of the matrix are correlated to those in the upper triangular part, but they are not equal. We formalize this ensemble with the following definition.

Definition 1.1.4 (Real elliptic random matrix). Let $(\xi_1, \xi_2)$ be a random vector in $\mathbb{R}^2$ and let $\zeta$
be a real-valued random variable. We say $X_n = (x_{ij})_{i,j=1}^n$ is an $n \times n$ real elliptic random matrix with atom variables $(\xi_1, \xi_2)$, and $\zeta$ if the following conditions hold.

- (independence) $\{x_{ii} : 1 \leq i \leq n\} \cup \{(x_{ij}, x_{ji}) : 1 \leq i < j \leq n\}$ is a collection of independent random elements.

- (off-diagonal entries) $\{(x_{ij}, x_{ji}) : 1 \leq i < j \leq n\}$ is a collection of iid copies of $(\xi_1, \xi_2)$.

- (diagonal entries) $\{x_{ii} : 1 \leq i \leq n\}$ is a collection of iid copies of $\zeta$.

We could define a complex elliptic random matrix analogously. In this ensemble, the off-diagonal entries are correlated and we let $\rho = \mathbb{E}[x_{ij}x_{ji}]$. Working with this ensemble brings new difficulties to studying the spectrum which were not present in either the iid matrix ensemble or the Wigner matrix ensemble. Indeed, difficulties arise due to the fact that the eigenvalues are complex, but the entries in the matrix are dependent across the diagonal.

### 1.2 Limiting Empirical Spectral Measure

The *eigenvalues* of an $n \times n$ matrix $M_n$ are the roots in $\mathbb{C}$ of the characteristic polynomial $\det(M_n - zI)$, where $I$ is the identity matrix. We let $\lambda_1(M_n), \ldots, \lambda_n(M_n)$ denote the eigenvalues of $M_n$ counted with (algebraic) multiplicity. Note that when a matrix $M_n$ is random, the eigenvalues are random as well. Next, we may define the *empirical spectral measure* $\mu_{M_n}$ of $M_n$ by

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

Again, if $M_n$ is a random matrix, then $\mu_{M_n}$ is also random. We say $\mu_{M_n}$ converges weakly in probability to another Borel probability measure $\mu$ on the complex plane $\mathbb{C}$ if, for every bounded and continuous function $f : \mathbb{C} \to \mathbb{C}$,

$$
\int_{\mathbb{C}} f d\mu_{M_n} \rightarrow \int_{\mathbb{C}} f d\mu \quad \text{(1.1)}
$$

in probability as $n \to \infty$. Observe that after integrating, $\int_{\mathbb{C}} f d\mu_{M_n}$ is a random variable and so we can characterize the convergence of this random variable using standard probability theory modes.
of convergence. We may also say $\mu_{M_n}$ converges weakly almost surely to $\mu$ if the convergence in (1.1) is replaced with almost sure convergence.

The distribution of the eigenvalues, and empirical spectral measure of a random matrix, depends on the matrix ensemble. One of the most fundamental problems in random matrix theory is to compute the limiting empirical spectral measure of an $n \times n$ random matrix $M_n$ as $n \to \infty$. Given a bounded continuous function $f$, we can write

$$\int_{C} f d\mu_{M_n} = \frac{1}{n} \sum_{i=1}^{n} f(\lambda_i(M_n)),$$

(1.2)

so the convergence of this random measure can be viewed as a law of large numbers for the eigenvalues of a random matrix.

We now discuss some previous results about the limiting empirical spectral measure of the ensembles defined in Section 1.1.

1.2.1 iid Random Matrices

Recall that when $\xi$ is a standard complex Gaussian random variable and $X_n$ is an iid random matrix with atom variable $\xi$, then $X_n$ is from the complex Ginibre ensemble. If $X_n$ is from the complex Ginibre ensemble, then we can view $X_n$ as a random matrix drawn from the probability density

$$\mathbb{P}(dM) = \frac{1}{\pi n^2} e^{-\text{tr}(M^*M)} dM$$

on the set of complex $n \times n$ matrices where, $dM$ denotes the Lebesgue measure on the $2n^2$ real entries of $M$. For this ensemble, the eigenvalues are well studied and it is in fact possible to compute the joint density of the eigenvalues of the random $n \times n$ matrix $X_n$. Ginibre did this in [63]. In [86, 87], Mehta used the results of Ginibre to prove that if $X_n$ is from the complex Ginibre ensemble, then the empirical spectral measure of $\frac{1}{\sqrt{n}}X_n$ converges to the uniform probability measure on the unit disk centered at the origin in the complex plane. In [51], Edelman proved that the same limiting distribution in the case when $X_n$ is from the real Ginibre ensemble. These results are examples of the circular law. In fact, similar results hold for general iid random matrices as well, but the
general case is more difficult to prove. This is in part due to the fact that there is no formula for the joint distribution of the eigenvalues. Bai [15, 19] made a breakthrough in the general case when he proved the result under a number of assumptions on the moments and smoothness of the atom variable $\xi$ by building on the earlier work of Girko [61, 65, 71]. Pan and Zhou [99] and Götze and Tikhomirov [73] made more advancements in this direction, and Tao and Vu [118] were able to prove the circular law under the assumption that $\mathbb{E}|\xi|^{2+\tau} < \infty$ for some $\tau > 0$. Finally, in 2009 Tao and Vu [117, 119] proved the circular law assuming only that $\xi$ has finite variance.

In general, the circular law describes the limiting empirical spectral measure of an iid random matrix. For any matrix $M$, we denote the Hilbert-Schmidt norm $\|M\|_2$ by the formula
\[
\|M\|_2 := \sqrt{\text{tr}(MM^*)} = \sqrt{\text{tr}(M^*M)}.
\] (1.3)

**Theorem 1.2.1** (Circular law; Corollary 1.12 from [119]). *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 $\text{rank}(A_n) = o(n)$ and $\sup_{n \geq 1} \frac{1}{n} \|A_n\|_2^2 < \infty$, then the empirical measure $\mu_{\frac{1}{\sqrt{n}}X_n+ A_n}$ of $\frac{1}{\sqrt{n}}X_n + A_n$ converges weakly almost surely to the uniform probability measure on the unit disk centered at the origin in the complex plane as $n \to \infty$.*

This result appears as [119 Corollary 1.12], but is the culmination of work by many authors. For a more complete overview of the circular law, we refer the interested reader to the excellent survey [37].

One of the main results in this thesis deals with outliers in the empirical spectral measure of iid matrices which have been perturbed by a low-rank deterministic matrix $A_n$. From Theorem 1.2.1 we see that the low-rank perturbation $A_n$ does not affect the limiting spectral measure. However, the perturbation $A_n$ may create one or more outliers. An example of this phenomenon is illustrated in Figure 1.1.

Theorem 1.2.1 also implies that with probability tending to one, the largest eigenvalue of $\frac{1}{\sqrt{n}}X_n$ in absolute value is at least $1 - o(1)$. The following result shows that when the atom variable
Figure 1.1: On the left, we have plotted the eigenvalues of a 500 × 500 random matrix from the Ginibre ensemble, scaled by \( \frac{1}{\sqrt{500}} \). Additionally, the unit circle is plotted for reference. The image on the right contains the eigenvalues of \( \frac{1}{\sqrt{500}} X + A \), where \( X \) is a 500 × 500 random matrix with iid symmetric ±1 Bernoulli entries and \( A = \text{diag}(1 + i, 2i - 1, 2, -i - 2, 0, \ldots, 0) \). For reference, we have also plotted each nonzero eigenvalue of \( A \) with a cross and have drawn a circle around each of these eigenvalues.

\( \xi \) has finite fourth moment, it is possible to improve this lower bound with a matching upper bound. Recall that the spectral radius of a matrix is the largest eigenvalue in absolute value.

**Theorem 1.2.2** (No outliers for iid matrices; Theorem 5.18 from [19]). Let \( \xi \) be a complex-valued random variable with mean zero, unit variance, and finite fourth moment. For each \( n \geq 1 \), let \( X_n \) be an iid random matrix with atom variable \( \xi \). Then the spectral radius of \( \frac{1}{\sqrt{n}} X_n \) converges to 1 almost surely as \( n \to \infty \).

**Remark 1.2.3.** In [36] it is conjectured that the spectral radius of \( \frac{1}{\sqrt{n}} X_n \) converges to 1 in probability as \( n \to \infty \) only assuming that \( \xi \) has mean zero and unit variance.

Theorem 1.2.2 asserts that almost surely all eigenvalues of \( \frac{1}{\sqrt{n}} X_n \) are contained in the disk of radius \( 1 + o(1) \) centered at the origin. However, as we saw in Figure 1.1 it is possible to have eigenvalues lying outside this disk when the random matrix is perturbed by a low-rank deterministic matrix, \( \frac{1}{\sqrt{n}} X_n + A_n \). In [114], Tao precisely describes the asymptotic locations of these outlying eigenvalues of \( \frac{1}{\sqrt{n}} X_n + A_n \).

**Theorem 1.2.4** (Outliers for small low-rank perturbations of iid matrices; Theorem 1.7 from [114]). Let \( \xi \) be a complex random variable with mean zero, unit variance, and finite fourth moment.
For each \( n \geq 1 \), let \( X_n \) be a \( n \times n \) random matrix whose entries are iid copies of \( \xi \), and let \( A_n \) be a deterministic matrix with rank \( O(1) \) and operator norm \( O(1) \). Let \( \varepsilon > 0 \), and suppose that for all sufficiently large \( n \), there are no eigenvalues of \( A_n \) in the band \( \{ z \in \mathbb{C} : 1 + \varepsilon < |z| < 1 + 3\varepsilon \} \), and there are \( j \) eigenvalues \( \lambda_1(A_n), \ldots, \lambda_j(A_n) \) for some \( j = O(1) \) in the region \( \{ z \in \mathbb{C} : |z| \geq 1 + 3\varepsilon \} \). Then, almost surely, for sufficiently large \( n \), there are precisely \( j \) eigenvalues 
\[
\lambda_i \left( \frac{1}{\sqrt{n}} X_n + A_n \right) = \lambda_i(A_n) + o(1)
\]
as \( n \to \infty \) for each \( 1 \leq i \leq j \).

In fact, the results in [103] extend Theorem 1.2.4 by also describing the joint fluctuations of the outlier eigenvalues about their asymptotic locations.

1.2.2 Products of iid Matrices

The results presented in this thesis focus on the product of several independent iid matrices. In this case, there is an analogue of the circular law (Theorem 1.2.1) which appears in [95]. We present the result due to O’Rourke, Renfrew, Soshnikov, and Vu.

**Theorem 1.2.5** (Theorem 2.4 from [95]). Fix an integer \( m \geq 1 \), and let \( \tau > 0 \). Let \( \xi_1, \ldots, \xi_m \) be real-valued random variables with mean zero, and assume, for each \( 1 \leq k \leq m \), that \( \xi_k \) 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 let \( A_{n,k} \) be a deterministic \( n \times n \) matrix. Assume \( X_{n,1}, \ldots, X_{n,m} \) are independent. If
\[
\max_{1 \leq k \leq m} \text{rank}(A_{n,k}) = O(n^{1-\varepsilon}) \quad \text{and} \quad \sup_{n \geq 1} \max_{1 \leq k \leq m} \frac{1}{n} \| A_{n,k} \|_2^2 < \infty
\]
for some fixed \( \varepsilon > 0 \), then the empirical spectral measure \( \mu_{P_n} \) of the product
\[
P_n := \prod_{k=1}^{m} \left( \frac{1}{\sqrt{n}} X_{n,k} + A_{n,k} \right)
\]

\[\text{1} \] Here and in the sequel, we use Pi (product) notation for products of matrices. To avoid any ambiguity, if
converges weakly almost surely to a (non-random) probability measure \( \mu \) as \( n \to \infty \). Here, the probability measure \( \mu \) is absolutely continuous with respect to Lebesgue measure on \( \mathbb{C} \) with density

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

where \( \sigma := \sigma_1 \cdots \sigma_m \).

Remark 1.2.6. When \( \sigma = 1 \), the density in (1.4) is easily related to the circular law (Theorem 1.2.1). Indeed, in this case, \( f \) 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 1.2.5 is a special case of [95, Theorem 2.4]. Figure 1.2.2 provides an illustration of this theorem.

![Figure 1.2](image-url)

Figure 1.2: This figure shows the eigenvalues, denoted by the small circles, of the product of four independent \( 500 \times 500 \) random matrices \( \frac{1}{\sqrt{500}} \mathbf{X}_{500} \) where the entries in each random matrix are iid Gaussian random variables. This figure illustrates Theorem 1.2.5.

Similar results have also been obtained in [34, 74, 96]. The Gaussian case was originally considered by Burda, Janik, and Waclaw [39]; see also [38]. We refer the reader to [3, 4, 6, 7, 8, M1, ..., Mm] are \( n \times n \) matrices, we define the order of the product

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

In many cases, such as in Theorem 1.2.5, the order of matrices in the product is irrelevant by simply relabeling indices.
and references therein for many other interesting results concerning products of random matrices with Gaussian entries.

In [89], Nemish extended Theorem 1.2.5 with a local result. In this result, he proved that if the entries in the matrices $X_{n,1}, \ldots, X_{n,m}$ are independent but not necessarily identically distributed and satisfy a subexponential decay condition, then in the bulk, the convergence described in Theorem 1.2.5 holds up to the scale $n^{-1/2+\varepsilon}$ for any fixed $\varepsilon > 0$.

**Theorem 1.2.7** (Theorem 1.2 from [89]). Let $m \geq 1$ be an integer and for each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ matrices with entries (real or complex valued, with independent real and imaginary parts) random variables with mean zero, variance $n^{-1}$, and satisfying the uniform subexponential decay condition

$$\exists \theta > 0, \text{ such that } \max_{1 \leq k \leq m} \max_{1 \leq i,j \leq n} P(|\sqrt{n}(X_{n,k})_{i,j}| > t) \leq \theta^{-1} e^{-t^\theta}.$$  

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the product $X_1 X_2 \cdots X_m$. Let $f: \mathbb{C} \to \mathbb{R}_+$ be a smooth non-negative function with compact support, such that $\|f\|_\infty \leq C$, $\|f'\|_\infty \leq n^C$ for constant some $C > 0$.

For any $d \in \mathbb{R}_+$ and $z_0 \in \mathbb{C}$ define

$$f_{z_0}(z) = n^{2d} f(n^d (z - z_0)).$$

Then for any $d \in (0, 1/2]$, any $\tau > 0$ small enough and $z_0$ such that $|z_0| \geq \tau$ and $1 - |z_0| \geq \tau$, and any $\varepsilon > 0$,

$$\left| \frac{1}{n} \sum_{j=1}^{n} f_{z_0}(\lambda_j) - \frac{1}{m\pi} \int_{\mathbb{C}} 1(|z|<1) f_{z_0}(z) |z|^{2/m-2} d\bar{z}dz \right| \leq n^{-1+2d+\varepsilon} \|\Delta f\|_1$$

with probability $1 - O_{C,\varepsilon,\tau}(n^{-C})$ for any constant $C > 0$.

In [72, Theorem 2.3], Götze, Naumov, and Tikhomirov prove another version of a local law under weaker moment conditions. Instead of the entries satisfying a subexponential decay condition, they prove a local law in which it is sufficient to require finite $4 + \tau$ moments for some $\tau > 0$.

For more results relating to the spectrum of products of independent matrices, we refer the reader to [5, 72, 89, 90, 95] and references therein.

\[\text{Here } \| \cdot \|_\infty \text{ denotes the infinity norm of a function and } \| \cdot \|_1 \text{ denotes the } L^1 \text{-norm of a function.}\]
1.2.3 Other Matrix Ensembles

Analogous results describing the limiting empirical spectral measure have been obtained for many ensembles of random matrices. Additionally, we refer the reader to [20, 22, 23, 25, 26, 27, 28, 29, 30, 35, 43, 44, 45, 47, 60, 66, 67, 68, 69, 70, 80, 81, 94, 97, 101, 102, 103, 104, 107, 114, 124] and references therein for results relating to locations and fluctuations of the outlier eigenvalues for various ensembles.

We state some results regarding the limiting empirical spectral density of other matrix ensembles now. First, we present the famous Wigner semicircle law.

**Theorem 1.2.8 (Semicircle Law).** Let $\xi$ be a complex-valued random variable with mean zero and $\text{Var}(\xi) = \sigma^2 > 0$, and let $\zeta$ be a real-valued random variable with finite mean and finite variance. Let $X_n$ be an $n \times n$ complex Hermitian Wigner random matrix with atom variables $\xi$ and $\zeta$. Then, as $n \to \infty$, the empirical spectral measure of $\frac{1}{\sqrt{n}}X_n$ converges almost surely to a measure $\mu_\sigma$ with density

$$f_\sigma(x) = \begin{cases} \frac{1}{2\pi\sigma}\sqrt{4\sigma^2 - x^2}, & |x| \leq 2\sigma \\ 0, & |x| > 2\sigma. \end{cases}$$

For a proof of this result, and other results relating to the semicircle law, see [11, 16, 19, 100, 121, 122]. The limiting behavior of eigenvalues is well explored in this case, partly due to the fact that all eigenvalues are real. In addition, we refer the reader to [42, 52, 53, 54, 55, 56, 57, 58, 59] for results regarding the local behavior of the spectrum.

Similar to the iid case, when a Wigner matrix is perturbed by a low rank deterministic matrix, it is possible to have outliers. This is characterized in the following theorem.

**Theorem 1.2.9 (Theorem 1.1 from [104]).** Let $\xi$ be a real random variable with mean zero, unit variance, and finite fourth moment, and let $\zeta$ be a real random variables with mean zero and finite variance. For each $n \geq 1$, let $X_n$ be an $n \times n$ Wigner matrix with atom variables $\xi, \zeta$. Let $k \geq 1$ and for each $n \geq k$, let $A_n$ be an $n \times n$ deterministic Hermitian matrix with rank $k$. 

\[ \]
and nonzero eigenvalues $\lambda_1(A_n), \ldots, \lambda_k(A_n)$ where $k, \lambda_1(A_n), \ldots, \lambda_k(A_n)$ are independent of $n$. Let \( S = \{1 \leq i \leq k : |\lambda_i(A_n)| > 1\} \). Then we have the following.

- For all $i \in S$, after labeling the eigenvalues of $\frac{1}{\sqrt{n}} X_n + A_n$ properly,
  \[
  \lambda_i \left( \frac{1}{\sqrt{n}} X_n + A_n \right) \longrightarrow \lambda_i(A_n) + \frac{1}{\lambda_i(A_n)}
  \]
  in probability as $n \to \infty$.

- For all $i \in \{1, \ldots, k\} \setminus S$, after labeling the eigenvalues of $\frac{1}{\sqrt{n}} X_n + A_n$ properly,
  \[
  \left| \lambda_i \left( \frac{1}{\sqrt{n}} X_n + A_n \right) \right| \longrightarrow 2
  \]
  in probability as $n \to \infty$.

In fact, in [44], Capitaine, Donati–Martin, and Féral strengthened the convergence of Theorem 1.2.9 to almost sure under additional assumptions on the atom variables.

Next, we discuss the limiting empirical spectral measure for the elliptic ensemble. The analogue to the Circular Law for elliptic matrices can be found in [66, 67, 68, 69, 70, 88, 91, 95].

**Theorem 1.2.10** (Theorem 1.14 from [91]). Let $(\xi_1, \xi_2, \zeta)$ be real valued random variables such that $\xi_1$ and $\xi_2$ have mean zero and variance one, and $\zeta$ has mean zero and has finite variance. Let \( \{X_n\}_{n \geq 1} \) be a sequence of elliptic random matrices with atom variables $(\xi_1, \xi_2, \zeta)$, let $\rho = \mathbb{E}[\xi_1 \xi_2]$, and assume that $-1 < \rho < 1$. For each $n \geq 1$, let $A_n$ be an $n \times n$ matrix, and assume the sequence \( \{A_n\}_{n \geq 1} \) satisfies $\sup_{n \geq 1} \text{rank}(A_n) = o(n)$ and $\sup_{n \geq 1} \frac{1}{n^2} \|A_n\|^2 < \infty$. Define the ellipsoid
\[
\mathcal{E}_\rho := \left\{ z \in \mathbb{C} : \frac{\text{Re}(z)^2}{(1 + \rho)^2} + \frac{\text{Im}(z)^2}{(1 - \rho)^2} \leq 1 \right\}
\]
and let $\mu_\rho$ denote the uniform probability measure on $\mathcal{E}_\rho$. Then the empirical spectral measure of $\frac{1}{\sqrt{n}} (X_n + A_n)$ converges almost surely to $\mu_\rho$ as $n \to \infty$.

We provide an illustration of Theorem 1.2.10 in Figure 1.3. Again, we can see that a small low rank perturbation does not effect the limiting spectral density. However, this perturbation may cause outliers. O’Rourke and Renfrew prove the following result which characterizes the asymptotic locations of outliers from a small low rank perturbation.
Figure 1.3: The plot shows the eigenvalues of a single $n \times n$ elliptic random matrix with Gaussian entries when $n = 1000$ and $\rho = 1/2$ after perturbing by a diagonal matrix with three nonzero eigenvalues: $2\sqrt{-1}$, $-3/2$, and $1+\sqrt{-1}$. The three circles are centered at $\frac{7}{4}\sqrt{-1}$, $-\frac{11}{6}$, and $\frac{5}{4}+\frac{3}{4}\sqrt{-1}$, respectively and each have radius $n^{-1/4}$. This illustrates Theorems 1.2.10 and 1.2.11. (Figure by O’Rourke and Renfrew, [94])

**Theorem 1.2.11** (Theorem 2.4 from [94]). Let $k \geq 1$ and $\delta > 0$. Let $\{X_n\}_{n \geq 1}$ be a sequence of elliptic random matrices with real-valued atom variables $(\xi_1, \xi_2), \zeta$ such that each of $\xi_1, \xi_2$ have mean zero and variance one, and $\zeta$ has mean zero and has finite variance. In addition, suppose $\xi_1, \xi_2,$ and $\zeta$ have finite fourth moments. Set $\rho = E[\xi_1 \xi_2]$, assume that $|\rho| < 1$, and define the ellipsoid

$$E_\rho := \left\{ z \in \mathbb{C} : \frac{\text{Re}(z)^2}{(1+\rho)^2} + \frac{\text{Im}(z)^2}{(1-\rho)^2} \leq 1 \right\}$$

and the neighborhood

$$E_{\rho, \delta} := \{ z \in \mathbb{C} : \text{dist}(z, E_\rho) \}.$$ 

For each $n \geq 1$, let $A_n$ be a deterministic $n \times n$ matrix, where $\sup_{n \geq 1} \| A_n \|_2 = O(1)$. Suppose for $n$ sufficiently large, there are no nonzero eigenvalues of $A_n$ which satisfy

$$\lambda_i(A_n) + \frac{\rho}{\lambda_i(A_n)} \in E_{\rho, 3\delta} \setminus E_{\rho, \delta} \text{ with } |\lambda_i(A_n)| > 1,$$

and there are $j$ eigenvalues $\lambda_1(A_n), \ldots, \lambda_j(A_n)$ for some $j \leq k$ which satisfy

$$\lambda_i(A_n) + \frac{\rho}{\lambda_i(A_n)} \in \mathbb{C} \setminus E_{\rho, 3\delta} \text{ with } |\lambda_i(A_n)| > 1.$$
Then, almost surely, for \( n \) sufficiently large, there are exactly \( j \) eigenvalues of \( \frac{1}{\sqrt{n}}X_n + A_n \) in the region \( \mathbb{C}\setminus\mathcal{E}_{\rho,2\delta} \), and after labeling the eigenvalues properly,

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

for each \( 1 \leq i \leq j \).

In addition, [95] proves the analogue of the Circular Law for the product of \( m \) independent elliptic random matrices. Interestingly, the limiting empirical spectral measure is the same as that of a product of independent iid matrices. We present this result below.

**Theorem 1.2.12** (Theorem 2.4 from [95]). Let \( m > 1 \) be an integer. For each \( 1 \leq k \leq m \), let \((\xi_{k,1},\xi_{k,2}),\xi_k\) be real random elements with mean zero such that \( \text{Var}(\xi_{k,1}) = \text{Var}(\xi_{k,2}) = \sigma_k^2 > 0 \), \( \text{Var}(\xi) < \infty \), there exists some \( \tau > 0 \) such that \( E|\xi_{k,1}|^{2+\tau} + E|\xi_{k,2}|^{2+\tau} < \infty \), and \( \rho_k := E[\xi_{k,1}\xi_{k,2}] \) satisfies \( |\rho_k| < 1 \). For each \( n \geq 1 \) and \( 1 \leq k \leq m \), let \( X_{n,k} \) be an \( n \times n \) real elliptic random matrix with atom variables \((\xi_{k,1},\xi_{k,2}),\xi_k\) and assume \( X_{n,1}, \ldots, X_{n,m} \) are independent. For each \( 1 \leq k \leq m \), let \( A_{n,k} \) be a \( n \times n \) deterministic matrix, and assume

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

for some \( \varepsilon > 0 \). Then the empirical spectral measure of the product

\[
P_n := n^{-m/2} \prod_{k=1}^{m} (X_{n,k} + A_{n,k})
\]

converges almost surely to a measure \( \mu_m \) which has density given by

\[
f_m(z) := \begin{cases} 
\frac{1}{m\pi} |z|^{-\frac{2}{m}-2}, & \text{if } |z| \leq \sigma, \\
0, & \text{if } |z| > \sigma
\end{cases}
\]

where \( \sigma := \sigma_1 \cdots \sigma_m \).

Figure 1.4 provides an illustration of Theorem 1.2.12 and compares products of elliptic matrices to products of other ensembles.
1.3 Central Limit Theorems

In addition to studying the limiting empirical spectral measure, one may also wonder how these eigenvalues fluctuate. While the results describing the limiting empirical spectral measure, such as Theorem 1.2.1, can be viewed as versions of the law of large numbers for eigenvalues of random matrices (see (1.2)), we may also ask about the fluctuations of the linear spectral statistics. For any $n \times n$ matrix $M$ and sufficiently smooth test function $f$, define the linear statistic

$$\text{tr } f(M) := \sum_{i=1}^{n} f(\lambda_i(M)).$$  

(1.5)
One main difference between central limit theorems for eigenvalues of random matrices and classical central limit theorems is that the variance of linear spectral statistics is $O(1)$ in many cases. Specifically, if $M_n$ is an $n \times n$ Wigner or an iid random matrix, then under certain moment assumptions on the atom variables and for any sufficiently smooth test function $f$,

$$\limsup_{n \to \infty} \text{Var} \left( \text{tr} f \left( \frac{1}{\sqrt{n}} M_n \right) \right) < \infty.$$ 

As a result of this, central limit theorems for the eigenvalues of random matrices often need not be rescaled in order to converge. There are many results regarding central limit theorems for eigenvalues of random matrices with various assumptions on the matrix ensembles and on the test function $f$. We discuss some of these results below.

### 1.3.1 iid Matrices

We present results regarding central limit theorems for eigenvalues of iid random matrices. First, observe the following result by Rider and Silverstein [105] for iid complex-valued matrices.

**Theorem 1.3.1 (Theorem 1.1 from [105]).** Let $\xi$ be a complex-valued random variable which satisfies the following conditions.

(i) $E[\xi] = 0$, and $E|\xi|^2 = 1$,

(ii) $E[\xi^2] = 0$,

(iii) $E|\xi|^k \leq k^{\alpha k}$ for every $k > 2$ and some $\alpha > 0$,

(iv) $\text{Re}(\xi)$ and $\text{Im}(\xi)$ 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 $f_1, f_2, \ldots, f_s$ analytic in a neighborhood of the disk $|z| \leq 4$ and otherwise bounded. Then as $n \to \infty$, the vector

$$\left( \text{tr} f_j \left( \frac{1}{\sqrt{n}} X_n \right) - E \left[ \text{tr} f_j \left( \frac{1}{\sqrt{n}} X_n \right) \right] \right)_{j=1}^s$$


converges in distribution to a mean-zero multivariate Gaussian vector
\[(F(f_1), F(f_2), \ldots, F(f_s))\]
with covariances
\[
E \left[ F(f_l)F(f_m) \right] = -\frac{1}{4\pi^2} \oint_C \oint_C f_l(\eta)f_m(\eta) \left\{ \frac{1}{\pi} \int_U \frac{d^2\eta}{(\eta - z)^2(\eta - \bar{w})^2} \right\} dzd\bar{w},
\]
in which \(U\) is the unit disk centered at the origin in the complex plane, \(C\) is any contour lying within the region of analyticity of both \(f_l\) and \(f_m\), and enclosing \(U\).

Other central limit theorem results for linear statistics of iid matrices include \([92, 106]\).

### 1.3.2 Products of iid Matrices

In \([83]\), Kopel, O’Rourke, and Vu study the fluctuation of the eigenvalues of a product of independent iid matrices. For this result, we must define what it means for random matrices to match moments. Two \(n \times n\) iid random matrices, \(M\) and \(M'\) with entries \(\xi(i,j)\) and \(\xi'(i,j)\) respectively, are said to match moments to order \(k\) if, for all \(1 \leq i, j \leq n\) and all \(a, b \geq 0\) such that \(a + b \leq k\),
\[
E \left[ \text{Re}(\xi(i,j))^a \text{Im}(\xi(i,j))^b \right] = E \left[ \text{Re}(\xi'(i,j))^a \text{Im}(\xi'(i,j))^b \right].
\]
In \([83]\), Kopel, O’Rourke, and Vu prove a central limit theorem for the eigenvalues of a product of independent iid random matrices which only requires test a function to have five continuous derivatives, but at the cost of requiring moment matching. The proof of this result uses a similar approach to \([106]\), where a central limit theorem is proved for a single Ginibre matrix.

**Theorem 1.3.2** (Theorem 1 from \([83]\)). *Let \(m \geq 1\) be an integer and \(1 > \tau_0 > 0\). Let \(\xi_1, \ldots, \xi_m\) be complex-valued random variables, and 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 \(\xi_k\) such that \(X_{n,1}, \ldots, X_{n,m}\) are independent and each \(X_{n,k}\) matches a complex Ginibre matrix to four moments. Let \(f : \mathbb{C} \to \mathbb{R}\) be a test function with at least five continuous derivatives, supported in the region \(\{z \in \mathbb{C} : \tau_0 < |z| < 1 - \tau_0\}\) and define the product
\[
P_n := n^{-m/2}X_{n,1} \cdots X_{n,m}.
\]
Then the centered linear statistic

\[ \text{tr} f(P_n) - E[\text{tr} f(P_n)] \]

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

\[
\frac{1}{4\pi} \int_{\mathbb{U}} |\nabla f(z)|^2 \, dz + \frac{1}{8\pi^2} \sum_{k \in \mathbb{Z}} |k| \left| \int_0^{2\pi} f(e^{\sqrt{-1} \theta}) e^{-\sqrt{-1} k \theta} \, d\theta \right|^2
\]

where \( \mathbb{U} \) is the unit disk centered at the origin in the complex plane. \(^3\)

### 1.3.3 Other Matrix Ensembles

In [108], Shcherbina proves a central limit theorem for linear eigenvalue statistics of Wigner matrices.

**Theorem 1.3.3** (Theorem 1 from [108]). Let \( \xi \) be a real random variable with mean zero, unit variance, and \( E|\xi|^4 < \infty \). Let \( \zeta \) be a real random variable with mean zero and variance \( \sigma^2 \). For each \( n \geq 1 \), let \( X_n \) be an \( n \times n \) real symmetric Wigner matrix with atom variables \( \xi, \zeta \). Let \( f \) be a real-valued test function which satisfies

\[
\int (1 + 2|l|)^{3/2+\varepsilon} \left| \hat{f}(l) \right|^2 \, dl < \infty
\]

for some \( \varepsilon > 0 \), where

\[
\hat{f}(l) := \frac{1}{\sqrt{2\pi}} \int f(x) e^{-\sqrt{-1} lx} \, dx
\]

is the Fourier transform of \( f \). Then

\[
\text{tr} f \left( \frac{1}{\sqrt{n}} X_n \right) - E \left[ \text{tr} f \left( \frac{1}{\sqrt{n}} X_n \right) \right]
\]

converges in distribution as \( n \to \infty \) to a mean-zero Gaussian random variable with variance

\[
\frac{1}{2\pi^2} \int_{-2}^2 \int_{-2}^2 \left( \frac{f(x) - f(y)}{x - y} \right)^2 \frac{4 - xy}{\sqrt{4 - x^2} \sqrt{4 - y^2}} \, dxdy
\]

\[
+ \frac{E|\xi|^4 - 3}{2\pi^2} \left( \int_{-2}^2 f(x)^2 \frac{2 - x^2}{\sqrt{4 - x^2}} \, dx \right)^2 + \frac{\sigma^2 - 2}{4\pi^2} \left( \int_{-2}^2 f(x) x \frac{2 - x^2}{\sqrt{4 - x^2}} \, dx \right)^2.
\]

\(^3\) Here, and throughout this result, we let \( \sqrt{-1} \) denote the imaginary unit and reserve \( i \) for an index.
A similar result was obtained by O’Rourke and Renfrew for real elliptic random matrices in [93].

**Theorem 1.3.4** (Theorem 2.2 from [93]). For each $n \geq 1$, let $X_n$ be an $n \times n$ real elliptic random matrix with atom variables $(\xi_1, \xi_2), \zeta$ which satisfy the following conditions.

(i) $\xi_1, \xi_2$ each have mean zero and unit variance,

(ii) $\zeta$ has mean zero and variance $\sigma^2$,

(iii) there exists a $\tau > 0$ such that

$$\mathbb{E}|\xi_1|^{6+\tau} + \mathbb{E}|\xi_2|^{6+\tau} + \mathbb{E} |\zeta|^{4+\tau} < \infty.$$

Set $\rho := \mathbb{E}[\xi_1 \xi_2]$ and assume that $|\rho| < 1$. Let $\delta > 0$. Define

$$\mathcal{E}_\rho := \left\{ z \in \mathbb{C} : \frac{(\text{Re} z)^2}{(1+\rho)^2} + \frac{(\text{Im} z)^2}{(1-\rho)^2} < 1 \right\}$$

and

$$\mathcal{E}_{\rho, \delta} := \{ z \in \mathbb{C} : \text{dist}(z, \mathcal{E}_\rho) \leq \delta \}$$

Let $f_1, \ldots, f_s$ be analytic in a neighborhood of $\mathcal{E}_{\rho, \delta}$ and bounded otherwise. In addition, assume

$$f_j(z) + f_j(\bar{z}) \in \mathbb{R}$$

for all $z \in \mathcal{E}_{\rho, \delta}$ and each $1 \leq j \leq k$. Then, as $n \to \infty$, the random vector

$$\left( \text{tr} f_j \left( \frac{1}{\sqrt{n}} X_n \right) - \mathbb{E} \left[ \text{tr} f_j \left( \frac{1}{\sqrt{n}} X_n \right) \right] \right)_{j=1}^{s}$$

converges in distribution to a mean-zero multivariate Gaussian vector $(F(f_1), \ldots, F(f_s))$ with co-variances

$$\mathbb{E}[F(f_i)F(f_j)] := -\frac{1}{4\pi^2} \oint_C \oint_C f_i(z)f_j(w)\nu(z,w)d\nu d\nu$$

$$= -\frac{1}{4\pi^2} \oint_C \oint_C f_i'(z)f_j'(w)m(z)m(w)\beta(z,w)d\nu d\nu$$
where $\mathcal{C}$ is the contour around the boundary of $\mathcal{E}_{\rho, \delta}$,

$$\nu(z, w) := \frac{\partial^2}{\partial z \partial w} m(z)m(w) \beta(z, w),$$

$$\beta(z, w) := \sigma^2 - \rho - 1 - \frac{\log(1 - \rho m(z)(m(w))}{m(z)m(w)} - \frac{\log(1 - m(z)m(w))}{m(z)m(w)}$$

$$+ \left( \frac{\mathbb{E}[\xi_1^2 \xi_2^2]}{2} - 2\rho^2 - 1 \right) m(z)m(w),$$

and

$$m(z) := \begin{cases} \frac{-z + \sqrt{z^2 - 4\rho}}{2\rho} & \text{for } \rho \neq 0 \\ \frac{-1}{z} & \text{for } \rho = 0 \end{cases}.$$
Chapter 2

Main Results

In this thesis, we study limiting properties of the spectrum of products of independent iid random matrices as the size of the matrices tend to $\infty$. We present two types of results. In the first collection of results, we study the asymptotic locations of outliers in the spectrum of perturbed products of independent iid random matrices. The results related to outliers in the spectrum of products of random matrices is joint work with Dr. Sean O’Rourke and Dr. Philip Matchett Wood.

The second result studies the limiting distribution of linear eigenvalue statistics for products of independent iid random matrices. This result is independent work, and was completed solely by the author of this thesis.

All of the main results of this thesis are presented below in Sections 2.1 and 2.2.

Remark 2.0.1. Throughout this thesis, we will take $P_n$ to denote many different matrix products. We will make it clear which product we are referencing whenever the notation $P_n$ is used. In addition, some notation is used in the proofs of both results with slightly different definitions. We make note of this when it occurs as well.

2.1 Outliers

We will first discuss the results about outliers of perturbed products. From Theorem 1.2.5, we see that the low-rank deterministic perturbations $A_{n,k}$ do not affect the limiting empirical spectral measure. However, as was the case in Theorem 1.2.1, the perturbations may create one or more outlier eigenvalues. The goal of this section is to study the asymptotic behavior of these outlier
eigenvalues. In view of Theorem 1.2.4 we will assume the atom variables $\xi_1, \ldots, \xi_m$ have finite fourth moments.

**Assumption 2.1.1.** The complex-valued random variables $\xi_1, \ldots, \xi_m$ are said to satisfy Assumption 2.1.1 if, for each $1 \leq k \leq m$,

- the real and imaginary parts of $\xi_k$ are independent,
- $\xi_k$ has mean zero and finite fourth moment, and
- $\xi_k$ has nonzero variance $\sigma_k^2$.

We begin with the analogue of Theorem 1.2.2 for the product of $m$ independent iid random matrices.

**Theorem 2.1.2** (No outliers for products of iid matrices). Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables which satisfy Assumption 2.1.1. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Define the products

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

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

**Remark 2.1.3.** A version of Theorem 2.1.2 was proven by Nemish in [90] under the additional assumption that the atom variables $\xi_1, \ldots, \xi_m$ satisfy a sub-exponential decay condition. In particular, this condition implies that all moments of $\xi_1, \ldots, \xi_m$ are finite. Theorem 2.1.2 only requires the fourth moments of the atom variables to be finite.

**Remark 2.1.4.** In view of Remark 1.2.3 it is natural to also conjecture that the spectral radius of $P_n$ is bounded above by $\sigma + o(1)$ in probability as $n \to \infty$ only assuming the atom variables
\(\xi_1, \ldots, \xi_m\) have mean zero and unit variance. Here, we need the result to hold almost surely, and hence require the atom variables have finite fourth moments.

In view of Theorem 1.2.5, it is natural to consider perturbations of the form

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

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

\[
P_n := \frac{1}{\sqrt{n}} X_{n,1}(I + A_{n,1}) \frac{1}{\sqrt{n}} X_{n,2}(I + A_{n,2}) \cdots \frac{1}{\sqrt{n}} X_{n,m}(I + A_{n,m}), \tag{2.1}
\]

or perturbations of the form

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

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

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

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

This phenomenon is most easily observed in the case of multiplicative perturbations (2.1), for which there is no deterministic term (i.e., \(A_n = 0\) and the perturbation consists entirely of “mixed” terms). In this case, the heuristic above suggests that there should be no outliers, and this is the content of the following theorem.

**Theorem 2.1.5** (No outliers for products of iid matrices with multiplicative perturbations). Let \(m \geq 1\) be a fixed integer, and assume \(\xi_1, \ldots, \xi_m\) are complex-valued random variables which satisfy Assumption 2.1.1. For each \(n \geq 1\), let \(X_{n,1}, \ldots, X_{n,m}\) be independent \(n \times n\) iid random matrices with atom variables \(\xi_1, \ldots, \xi_m\), respectively. In addition for any fixed integer \(s \geq 1\), let
$A_{n,1}, A_{n,2}, \ldots, A_{n,s}$ be $n \times n$ deterministic matrices, each of which has rank $O(1)$ and operator norm $O(1)$. Define the product $P_n$ to be the product of the terms

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

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

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

**Theorem 2.1.6** (Outliers for small, low-rank perturbations of products of matrices). Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables which satisfy Assumption 2.1.1. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. In addition, let $A_n$ be an $n \times n$ deterministic matrix with rank $O(1)$ and operator norm $O(1)$. Define

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

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

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

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

Figure 2.1 presents a numerical simulation of Theorem 2.1.6. In the case that all the entries of $A_n$ take the same value, the product $P_n$ in (2.2) can be viewed as a product matrix whose entries have the same nonzero mean. Technically, Theorem 2.1.6 cannot be applied in this case, since such a matrix $A_n$ does not have operator norm $O(1)$. However, using a similar proof, we establish the following result.
Figure 2.1: In this figure, we have plotted the eigenvalues of $(500)^{-2}X_1 X_2 X_3 X_4 + A$, where $X_1, \ldots, X_4$ are independent $500 \times 500$ iid random matrices with symmetric $\pm 1$ Bernoulli entries and $A = \text{diag}(-1 + i, -2, 2, 0, \ldots, 0)$. The majority of the eigenvalues cluster inside the unit disk with the exception of three outliers. These outliers are close to the eigenvalues of $A$, each of which is marked with a cross, and a circle is drawn around each outlying eigenvalue.

**Theorem 2.1.7** (Outliers for a product matrix with nonzero mean). Let $m \geq 1$ be an integer, and let $\mu \in \mathbb{C}$ be nonzero. Assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables which satisfy Assumption 2.1.1. 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. Let $\phi_n := \frac{1}{\sqrt{n}}(1, \ldots, 1)^* \mathbf{x}$ and fix $\gamma > 0$. Define

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

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

Lastly, we consider the case of Theorem 1.2.5 where there are both “mixed” terms and a deterministic term.

**Theorem 2.1.8.** Let $m \geq 1$ be an integer, and assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables which satisfy Assumption 2.1.1. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. In addition, for each $1 \leq k \leq m$, let $A_{n,k}$ be a deterministic $n \times n$ matrix with rank $O(1)$ and operator norm $O(1)$. Define the products
\[ P_n := \prod_{k=1}^{m} \left( \frac{1}{\sqrt{n}} X_{n,k} + A_{n,k} \right), \quad A_n := \prod_{k=1}^{m} A_{n,k}, \] (2.3)

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

\[ \lambda_i(P_n) = \lambda_i(A_n) + o(1) \]

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

Theorem 2.1.8 can be viewed as a generalization of Theorem 1.2.4. In fact, when \( m = 1 \), Theorem 2.1.8 is just a restatement of Theorem 1.2.4. However, the most interesting cases occur when \( m \geq 2 \). Indeed, in these cases, Theorem 2.1.8 implies that the outliers of \( P_n \) are asymptotically close to the outliers of the product \( A_n \). Specifically, if even one of the deterministic matrices \( A_{n,k} \) is zero, asymptotically, there cannot be any outliers for the product \( P_n \). Figure 2.2 presents a numerical simulation of Theorem 2.1.8.

These results are proved in Chapter 4.

2.1.1 Related Results

While our main results of Section 2.1 have focused on independent iid matrices, it is also possible to consider the case when the random matrices \( X_{n,1}, \ldots, X_{n,m} \) are no longer independent. In particular, we consider the extreme case where \( X_{n,1} = \cdots = X_{n,m} \) almost surely. In this case, we obtain the following results, which are analogs of the results from Section 2.1.

**Theorem 2.1.9** (No outliers for multiplicative perturbations). Assume that \( \xi \) is a complex-valued random variable which satisfies Assumption 2.1.1 with \( \sigma^2 := \text{Var}(\xi) \). For each \( n \geq 1 \), let \( X_n \) be an \( n \times n \) iid random matrix with atom variable \( \xi \). In addition for any finite integer \( s \geq 1 \), let
Figure 2.2: In the above figure, we display the eigenvalues of products of random matrices of the form $\prod_{k=1}^{5} \left( \frac{1}{\sqrt{1000}} X_k + A_k \right)$, where $X_1, \ldots, X_5$ are $1000 \times 1000$ independent iid matrices with symmetric $\pm 1$ Bernoulli entries, and the product of the deterministic matrices $A_1, \ldots, A_5$ is diag($-2, -1 + 2i, 2, 0, \ldots, 0$). Each nonzero eigenvalue of the product $A_1 \cdots A_5$ is marked with a cross, and a circle is drawn around each outlying eigenvalue.

Let $A_{n,1}, A_{n,2}, \ldots, A_{n,s}$ be $n \times n$ deterministic matrices, each of which has rank $O(1)$ and operator norm $O(1)$. Define the product $P_n$ to be the product of $m$ copies of $\frac{1}{\sqrt{n}} X_n$ with the terms

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

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

For a single additive perturbation, we have the following analog of Theorem 2.1.6.

**Theorem 2.1.10** (Outliers for a single additive perturbation). Assume $\xi$ is a complex-valued random variable which satisfies Assumption 2.1.1 with $\sigma^2 := \text{Var}(\xi)$. For each $n \geq 1$, let $X_n$ be an $n \times n$ iid random matrix with atom variable $\xi$. In addition, let $A_n$ be an $n \times n$ deterministic matrix with rank $O(1)$ and operator norm $O(1)$. Define

$$P_n := n^{-m/2} X_n^m + A_n. \quad (2.4)$$

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

\[
\lambda_i(P_n) = \lambda_i(A_n) + o(1)
\]

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

Note that Theorem 2.1.7 can also be generalized in an analogous way to Theorem 2.1.10 above.

Finally, we have the following analog of Theorem 2.1.8.

**Theorem 2.1.11.** Assume \( \xi \) is a complex-valued random variable which satisfies Assumption 2.1.1 with \( \sigma^2 := \text{Var}(\xi) \). For each \( n \geq 1 \), let \( X_n \) be an \( n \times n \) iid random matrix with atom variable \( \xi \). In addition, let \( m \geq 1 \) be an integer and for each \( 1 \leq k \leq m \), let \( A_{n,k} \) be a deterministic \( n \times n \) matrix with rank \( O(1) \) and operator norm \( O(1) \). Define the products

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

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

\[
\lambda_i(P_n) = \lambda_i(A_n) + o(1)
\]

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

The proofs of these results are presented in Section 4.8 and use similar techniques to the proofs of the main results from Section 2.1.
2.1.2 Open Questions and Applications

Before proceeding, we discuss some open questions and applications which relate to these results.

2.1.2.1 Applications

Random matrices are useful tools in the study of many physically motivated systems, and we note here two potential applications for products of perturbed random matrices. First, iid Gaussian matrices can be used to model neural networks as in, for example, [1,2,14]. In the case of a linear version of the feed-forward networks in [2], the model becomes a perturbation of a product of iid random matrices, which, if the interactions in the model were fixed, could potentially be analyzed using approaches in the current thesis. Second, one can conceive of a dynamical system (see, for example, [78]) evolving according to a matrix equation, which, when iterated, would lead to a matrix product of the form discussed in Theorem 2.1.8.

2.1.2.2 Open Questions

While our main results have focused on iid matrices, it is natural to ask if the same results can be extended to other matrix models. For example, Theorem 1.2.5 and the results in [95] also hold for products of so-called elliptic random matrices. However, the techniques used in the proofs of these results (in particular, the combinatorial techniques in Section 4.7) rely heavily on the independence of the entries of each matrix. It is an interesting question whether an alternative proof can be found for the case when the entries of each matrix are allowed to be dependent.

In this result, we have focused on the asymptotic location of the outlier eigenvalues. One can also ask about the fluctuations of the outliers. For instance, in [103], the joint fluctuations of the outlier eigenvalues from Theorem 1.2.4 were studied.
2.2 Central Limit Theorem

Next, we will present the result regarding the fluctuations of linear spectral statistics of products of large, independent iid matrices. The main theorem of this section deals with the fluctuations of the eigenvalues of a product of $m = O(1)$ independent random matrices, each of which are iid. The theorem and proof are inspired by previous central limit theorems for eigenvalues of random matrices such as [93, Theorem 2.2] and [105, Theorem 1.1]. Some of the methods used in the proof of this result are generalized versions of those in [93, 105], but in order to employ these methods, it is necessary to adapt them to the case where the eigenvalues are from a product of $m$ independent iid matrices. In particular, many methods required iteration and generalization in order to be applicable to the setting of products of matrices.

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

**Assumption 2.2.1.** The real-valued random variables $\xi_1, \ldots, \xi_m$ are said to satisfy Assumption 2.2.1 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 this section. Chapter 5 is devoted to the proof of the below result.

**Theorem 2.2.2** (Central limit theorem 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 2.2.1. 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

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

(2.6)
and
\[
\sigma := \sigma_1 \cdots \sigma_m.
\]

Let \( \delta > 0 \), let \( s > 0 \) be a fixed integer, and let \( f_1, f_2, \ldots, f_s \) 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 exists an event \( E_n \) which holds with probability \( 1 - o(1) \) on which, as \( n \to \infty \), the random vector
\[
(tr f_i(P_n/\sigma)1_{E_n} - E[tr f_i(P_n/\sigma)1_{E_n}])_{i=1}^s
\]
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
\[
E[F(f_i)F(f_j)] = -\frac{1}{4\pi^2} \oint_C \oint_C f_i(z)f_j(w)(zw - 1)^{-2}dzdw \quad (2.8)
\]
and
\[
E[F(f_i)\overline{F(f_j)}] = \frac{1}{4\pi^2} \oint_C \oint_C f_i(z)\overline{f_j(w)}(z\bar{w} - 1)^{-2}dzd\bar{w} \quad (2.9)
\]
where \( C \) is the contour around the boundary of the disk \( D_\delta \).

The proof of this result can be found in Chapter 5. Figure 2.3 provides an illustration of this theorem with various assumptions on the entires, test function, and number of matrices in the product.

Remark 2.2.3. The event \( E_n \) will be formally defined in (5.3), and it may seem unintuitive to include this event now. However, the event is required in order to control the least singular value of the product, which is required by the methods used in this result. The use of this event could be avoided if instead of \( 4 + \tau \) finite moments, it was assumed that all atom random variables had \( 6 + \tau \) finite moments.

Remark 2.2.4. The covariance terms (2.8) and (2.9) of the limiting random vector is currently stated in terms of two contour integrals in \( \mathbb{C} \), but this could be stated in terms of an iterated integral over the real and imaginary parts of \( z \) as was done in [105, Theorem 1.1].
Figure 2.3: This figure provides an illustration of Theorem 2.2.2. All four plots contain 1000 observations of the linear statistic in (1.5) under various conditions on the atom variables, the number of matrices in the products, and the test function. The top left plot shows linear statistics computed with a product of 3 real $300 \times 300$ matrices, each with Bernoulli($-1, 1$) atom random variables, scaled by $300^{-3/2}$ and with test function $f(z) = z^2 + 2\sqrt{-1}z$. The top right plot contains linear statistics for a product of 10 real $300 \times 300$ matrices, each with Bernoulli($-1, 1$) atom random variables, scaled by $300^{-10/2}$ and with test function $f(z) = z^2 + 2\sqrt{-1}z$. The bottom left plot contains observations of the linear statistic for a product of 3 real $300 \times 300$ matrices, each with mean-zero Gaussian atom random variables, scaled by $300^{-3/2}$ and with test function $f(z) = \sqrt{-1}z^3 + z^2$. Finally, the bottom right plot contains observations for a product of 10 real $300 \times 300$ matrices, each with mean-zero Gaussian atom random variables, scaled by $300^{-10/2}$ and with test function $f(z) = \sqrt{-1}z^3 + z^2$. One may observe that in all four images, the distribution looks roughly Gaussian, and the variance does not depend on the number of matrices in the product, or on the distribution on the atom variables. Note that the top two plots appear to have a similar variance, and the bottom two also have a similar variance. This is due to the fact that the variance is dependent on the test function, which is the same for the top two, and the same for the bottom two.

Remark 2.2.5. This result can be extended to the case where each atom random variable is complex. In this case, Assumption 2.2.1 would need to be modified to require that the real and imaginary parts of each random variable are independent. Additionally, the variance and covariance terms in Theorem 2.2.2 would change. There are some areas of the proof where the details would need
to be modified in order to extend the result to the complex case. See Remarks 5.2.1, 5.2.5, 5.4.2, and 5.4.16. Figure 2.4 provides a simulation of Theorem 2.2.2 in the case when the atom random variables are complex.

![Figure 2.4: This figure provides an illustration of Theorem 2.2.2 using complex atom variables. Both plots contain 1000 observations of the linear statistic in (1.5). The left plot shows linear statistics computed with a product of 3 complex $300 \times 300$ matrices, each with mean-zero atom random variables, scaled by $300^{-3/2}$ and with test function $f(z) = z^2 + 2z$. The right plot contains linear statistics for a product of 10 complex $300 \times 300$ matrices, each with mean-zero Gaussian atom random variables, scaled by $300^{-10/2}$ and with test function $f(z) = z^2 + 2z$. Note that these simulations are consistent with the results of Theorem 2.2.2 even though the atom variables are complex. Theorem 2.2.2 can be extended to the case when the atom variables are complex. See Remark 2.2.5.]

2.2.1 Open Questions

We discuss some open questions related to the Theorem 2.2.2.

The matrices in question in Theorem 2.2.2 are assumed to be iid and independent of one another. The independence of entries and the independence of matrices are utilized in Sections 5.2 and 5.4. It is unclear what result would follow if the entries in each matrix were dependent in some way, or if the matrices themselves were dependent on one another. There may be an analogous result for products of elliptic random matrices based on a combination of the methods presented here and in [93].

Another key fact used in the proofs presented here is that the number of matrices in the product, $m$, is a fixed integer. It can be noted that the covariance structure of the product does
not depend on \( m \), and one might wonder if the same result holds if we let \( m \) grow sufficiently slowly with \( n \). Many asymptotic bounds depend on the finite constant \( m \), but the explicit dependence is never computed. However, it is possible that these proofs could be adapted to the case where \( m \) grows with \( n \) if the dependence on \( m \) is tracked throughout the calculations.

Methods used in this result, such as Cauchy’s integral formula, make it necessary for the test functions to be analytic. However, there may be other methods which would allow us to have less restrictive requirements for these test functions (see, for example, [83]). It is unclear what the necessary and sufficient conditions on test functions are in order to ensure that the centered linear statistics converge to a mean-zero Gaussian distribution.

Finally, it is unclear if there is a way to remove the event \( E_n \) from the statement of Theorem 2.2.2 without modifying the moment assumptions. With the proof methods used here, the event is necessary, but it is possible that another method may remove the necessity for the event to appear in the statement of Theorem 2.2.2.
Chapter 3

Tools and Notation

This chapter is devoted to introducing some additional concepts and notation required for the proofs of our main results from Chapter 2.

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 another constant, e.g. $C = C_k$, we indicate this with subscripts, e.g. $X = O_k(Y)$, $Y = \Omega_k(X)$, $X \ll_k Y$, and $Y \gg_k X$. We write $X = o(Y)$ if $|X| \leq c(n)Y$ for some sequence $c(n)$ that goes to zero as $n \to \infty$. Specifically, $o(1)$ denotes a term which tends to zero as $n \to \infty$. If we need the sequence $c(n)$ to depend on another constant, e.g. $c(n) = c_k(n)$, we indicate this with subscripts, e.g. $X = o_k(Y)$.

Throughout this thesis, 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 $1_E$ denote the indicator function of the event $E$, and we let $E^c$ denote the complement of the event $E$. We write a.s. for almost surely.

For a matrix $M$, we let $\|M\|$ denote the spectral norm of $M$, and we let $\|M\|_2$ denote the Hilbert-Schmidt norm of $M$ (defined in [1.3]). We denote the eigenvalues of an $n \times n$ matrix $M$ by $\lambda_1(M), \ldots, \lambda_n(M)$, and we let $\rho(M) := \max\{|\lambda_1(M)|, \ldots, |\lambda_n(M)|\}$ denote its spectral radius.
We let $I_n$ denote the $n \times n$ identity matrix and $0_n$ denote the $n \times n$ zero matrix. Often we will just write $I$ (or 0) for the identity matrix (alternatively, zero matrix) when the size can be deduced from context.

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

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

### 3.1 Linearization

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

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

(3.1)

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, [11, 39, 75, 95, 96] and references therein.

**Proposition 3.1.1.** Let $M_1, \ldots, M_m$ be $n \times n$ matrices. Let $P := M_1 \cdots M_m$, and assume $\mathcal{M}$ is the $mn \times mn$ block matrix defined in (3.1). Then

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


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

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

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

where \( Z_1 := P \) and

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

for \( 1 < k \leq m \). Since each product \( Z_2, \ldots, Z_m \) has the same characteristic polynomial\(^1\) as \( P \), it follows that

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

for all \( z \in \mathbb{C} \).

We will exploit Proposition 3.1.1 many times in the coming proofs.

### 3.2 Matrix Notation

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

If \( M \) is an \( n \times n \) matrix and \( 1 \leq i, j \leq n \), we let \( M_{ij} \) and \( M(i,j) \) denote the \((i, j)\)-entry of \( M \). Similarly, if \( \mathcal{M} \) is an \( N \times N \) matrix, we let \( \mathcal{M}_{ij} \) and \( \mathcal{M}(i,j) \) denote the \((i, j)\)-entry of \( \mathcal{M} \) for \( 1 \leq i, j \leq N \). However, in many instances, it is best to view \( N \times N \) matrices as block matrices with \( n \times n \) entries. To this end, we introduce the following notation. Let \( \mathcal{M} \) be an \( N \times N \) matrix. For \( 1 \leq a, b \leq m \), we let \( \mathcal{M}^{[a,b]} \) denote the \( n \times n \) matrix which is the \((a, b)\)-block of \( \mathcal{M} \). For convenience,

\(^1\) This fact can easily be deduced from Sylvester’s determinant theorem; see 4.1.
we extend this notation to include the cases where \( a = m + 1 \) or \( b = m + 1 \) by taking the value \( m + 1 \) to mean \( 1 \) (i.e., modulo \( m \)). For instance, \( \mathcal{M}^{[m+1,m]} = \mathcal{M}^{[1,m]} \). For \( 1 \leq i, j \leq n \), the notation \( \mathcal{M}^{[a,b]}_{ij} \) or \( \mathcal{M}^{[a,b]}_{(i,j)} \) denotes the \((i,j)\)-entry of \( \mathcal{M}^{[a,b]} \).

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

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

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

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

### 3.3 Singular Value Inequalities

For an \( n \times n \) matrix \( M \), recall that \( s_1(M) \geq \cdots \geq s_n(M) \) denote its ordered singular values.

We will need the following elementary bound concerning the largest and smallest singular values.

**Proposition 3.3.1.** Let \( M \) be an \( n \times n \) matrix and assume \( E \subseteq \mathbb{C} \) such that

\[
\inf_{z \in E} s_n(M - zI) \geq c
\]

for some constant \( c > 0 \). Then

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

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

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

\[
\sup_{z \in E} s_1(G(z)) = \sup_{z \in E} \frac{1}{s_n(M - zI)} = \frac{1}{\inf_{z \in E} s_n(M - zI)} \leq \frac{1}{c},
\]

as desired.

There are some other useful equalities we will use throughout this thesis.

**Lemma 3.3.2 (Eigenvalue criterion).** Let \(Y\) and \(A\) be \(n \times n\) matrices, and assume \(A = BC\), where \(B\) is an \(n \times k\) matrix and \(C\) is a \(k \times n\) matrix. Let \(z\) be a complex number which is not an eigenvalue of \(Y\). Then \(z\) is an eigenvalue of \(Y + A\) if and only if

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

**Remark 3.3.3.** The proof of Lemma 3.3.2 actually reveals that

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

provided the denominator does not vanish. Versions of this identity have appeared in previous publications including [13, 26, 27, 28, 43, 44, 80, 81, 94, 97, 102, 104, 114].

**Proof of Lemma 3.3.2.** Assume \(z\) is not an eigenvalue of \(Y\). Then \(\det(Y - zI) \neq 0\) and

\[
\det(Y + A - zI) = \det(Y - zI) \det(I + (Y - zI)^{-1}A)
\]

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

Thus, by (4.1), \(z\) is an eigenvalue of \(Y + A\) if and only if

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

as desired. \(\Box\)
We will make use of the Sherman–Morrison rank one perturbation formula (see \cite{76} 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

$$ \left(A + uv^*\right)^{-1} = A^{-1} - \frac{A^{-1}uv^*A^{-1}}{1 + v^*A^{-1}u} \quad (3.3) $$

and

$$ \left(A + uv^*\right)^{-1}u = \frac{A^{-1}u}{1 + v^*A^{-1}u}. \quad (3.4) $$

Also recall the Sherman–Morrison–Woodbury formula (for example, \cite{46} Theorem 1.1), which states that for an $N \times N$ matrix $A$ and $a \times N$ matrices $V, U$ for some fixed $a < N$,

$$ \left(A + UV^T\right)^{-1}U = A^{-1}U(I_a + V^TA^{-1}U)^{-1} \quad (3.5) $$

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

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

$$ A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1} \quad (3.6) $$

whenever $A$ and $B$ are invertible.

We also use Weyl’s inequality for the singular values (see, for example, \cite{31} Problem III.6.5)), which states that for $n \times n$ matrices $A$ and $B$,

$$ \max_{1 \leq i \leq n} |s_i(A) - s_i(B)| \leq \|A - B\|. \quad (3.7) $$
Chapter 4

Proofs of Theorems in Section 2.1

This chapter is devoted to the proofs of theorems listed in Section 2.1. In particular, in this chapter we prove Theorems 2.1.2, 2.1.5, 2.1.6, 2.1.7, and 2.1.8. Section 4.8 presents the proofs of Theorems 2.1.9, 2.1.10, and 2.1.11.

4.1 Overview and Preliminary Tools

Let us now briefly overview the proofs of our main results from Section 2.1. One of the key ingredients is the eigenvalue criterion lemma presented in Chapter 3 (Lemma 3.3.2), which is based on Sylvester’s determinant theorem:

$$\det(I + AB) = \det(I + BA)$$ (4.1)

whenever $A$ is an $n \times k$ matrix and $B$ is a $k \times n$ matrix. In particular, the left-hand side of (4.1) is an $n \times n$ determinant and the right-hand side is a $k \times k$ determinant.

In these proofs, we will exploit Proposition 3.1.1. Specifically, in order to study the product $X_{n,1} \cdots X_{n,m}$, we will consider the $mn \times mn$ block matrix

$$Y_n := \begin{bmatrix} 0 & X_{n,1} & 0 \\ 0 & 0 & X_{n,2} & 0 \\ & \ddots & \ddots & \ddots \\ 0 & 0 & X_{n,m-1} & 0 \\ X_{n,m} & 0 & & \end{bmatrix}$$ (4.2)
and its resolvent

\[ G_n(z) := \left( \frac{1}{\sqrt{n}} Y_n - zI \right)^{-1}, \]

(4.3)
defined for \( z \in \mathbb{C} \) provided \( z \) is not an eigenvalue of \( \frac{1}{\sqrt{n}} Y_n \). We study the location of the eigenvalues of \( \frac{1}{\sqrt{n}} Y_n \) in Theorem 4.2.1 below.

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

\[
A_n := \begin{bmatrix}
0 & A_{n,1} & 0 \\
0 & 0 & A_{n,2} & 0 \\
& & \ddots & \ddots \\
0 & 0 & A_{n,m-1} & 0 \\
A_{n,m} & & & 0
\end{bmatrix}.
\]

(4.4)

For concreteness, let us focus on the proof of Theorem 2.1.8. That is, we wish to study the eigenvalues of

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

outside the disk \( \{ z \in \mathbb{C} : |z| \leq \sigma + 2\varepsilon \} \). We first linearize the problem by invoking Proposition 3.1.1 with the matrix \( \frac{1}{\sqrt{n}} Y_n + A_n \), where \( Y_n \) and \( A_n \) are defined in (4.2) and (4.4). Indeed, by Proposition 3.1.1 it suffices to study the eigenvalues of \( \frac{1}{\sqrt{n}} Y_n + A_n \) outside of the disk \( \{ z \in \mathbb{C} : |z| \leq \sigma^{1/m} + \delta \} \) for some \( \delta > 0 \) (depending on \( \sigma, \varepsilon, \) and \( m \)). Let us suppose that \( A_n \) is rank one. In other words, assume \( A_n = vu^* \) for some \( u, v \in \mathbb{C}^{mn} \). In order to study the outlier eigenvalues, we will need to solve the equation

\[ \det \left( \frac{1}{\sqrt{n}} Y_n + A_n - zI \right) = 0 \]

(4.5)

for \( z \in \mathbb{C} \) with \( |z| > \sigma^{1/m} + \delta \). Assuming \( z \) is not eigenvalue of \( Y_n \), we can rewrite (4.5) as

\[ \det (I + G_n(z)A_n) = 0, \]

where the resolvent \( G_n(z) \) is defined in (4.3). From (4.1) and the fact that \( A_n = vu^* \), we find that this reduces to solving

\[ 1 + u^* G_n(z) v = 0. \]
Thus, the problem of locating the outlier eigenvalues reduces to studying the resolvent \( G_n(z) \). In particular, we develop an isotropic limit law in Section 4.2 to compute the limit of \( u^*G_n(z)v \). This limit law is inspired by the isotropic semicircle law developed by Knowles and Yin in \([80, 81]\) for Wigner random matrices as well as the isotropic law verified in \([94]\) for elliptic matrices.

The general case, when \( A_n \) is not necessarily rank one, is similar. In this case, we will exploit the following criterion to characterize the outlier eigenvalues.

### 4.2 Isotropic Limit Law and the Proof of Theorems in Section 2.1

This section is devoted to the proofs of Theorems 2.1.2, 2.1.5, 2.1.6, 2.1.7, and 2.1.8. The key ingredient is the following result concerning the properties of the resolvent \( G_n(z) \).

**Theorem 4.2.1** (Isotropic limit law). Let \( m \geq 1 \) be a fixed integer, and assume \( \xi_1, \ldots, \xi_m \) are complex-valued random variables with mean zero, unit variance, finite fourth moments, and independent real and imaginary parts. For each \( n \geq 1 \), let \( X_{n,1}, \ldots, X_{n,m} \) be independent \( n \times n \) iid random matrices with atom variables \( \xi_1, \ldots, \xi_m \), respectively. Recall that \( Y_n \) is defined in \((4.2)\) and its resolvent \( G_n(z) \) is defined in \((4.3)\). Then, for any fixed \( \delta > 0 \), the following statements hold.

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

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

\[
\sup_{z \in \mathbb{C} : |z| > 1 + \delta} \| G_n(z) \| \leq c.
\]

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

\[
\sup_{z \in \mathbb{C} : |z| > 1 + \delta} \left| u_n^* G_n(z)v_n + \frac{1}{z} u_n^* v_n \right| \to 0
\]

almost surely as \( n \to \infty \).
We conclude this section with the proofs of Theorems 2.1.2, 2.1.5, 2.1.6, 2.1.7, and 2.1.8 assuming Theorem 4.2.1. Sections 4.3–4.7 are devoted to the proof of Theorem 4.2.1.

4.2.1 Proof of Theorem 2.1.2

Consider

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

and note that by rescaling by \( \frac{1}{\sigma} \), it is sufficient to assume that \( \sigma_i = 1 \) for all \( 1 \leq i \leq m \). By Proposition 3.1.1, the eigenvalues of \( P_n \) are precisely the eigenvalues of \( n^{-m/2}Y^m_n \), each with multiplicity \( m \). Additionally, the eigenvalues of \( n^{-m/2}Y^m_n \) are exactly the \( m \)-th powers of the eigenvalues of \( n^{-1/2}Y_n \). Thus, it is sufficient to study the spectral radius of \( n^{-1/2}Y_n \). By part (i) of Theorem 4.2.1 we conclude that almost surely,

\[
\limsup_{n \to \infty} \rho \left( \frac{1}{\sqrt{n}}Y_n \right) \leq 1
\]

where \( \rho(M) \) denotes the spectral radius of the matrix \( M \). This completes the proof of Theorem 2.1.2.

4.2.2 Proof of Theorem 2.1.5

Recall that for this theorem, we define \( P_n \) to be the product of the terms

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

in some fixed order. By rescaling by \( \frac{1}{\sigma} \), it is sufficient to assume that \( \sigma_i = 1 \) for \( 1 \leq i \leq m \). Observe that if two deterministic terms

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

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

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

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

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

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

(4.6)

for the remainder of the proof. Next, define the $2mn \times 2mn$ matrix

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

(4.7)

where $0_{mn}$ denotes the $mn \times mn$ zero matrix, $Y_n$ is as defined in (4.2), and $\text{diag}(A_{n,1}, \ldots, A_{n,m})$ is defined so that

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

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

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

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

and

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

A simple computation reveals that

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

where

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

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

Define

\[
X_n := \begin{bmatrix} 0_{mn} & I_{mn} \\ \frac{1}{\sqrt{n}} \psi_n & 0_{mn} \end{bmatrix} \quad \text{and} \quad A_n^{\Box} := \begin{bmatrix} 0_{mn} & \text{diag}(A_{n,1}, \ldots, A_{n,m}) \\ 0_{mn} & 0_{mn} \end{bmatrix}.
\]

Then we can rewrite \( \mathcal{L}_n = X_n + A_n^{\Box} \).

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

\[
X_n^{2m} = n^{-m/2} \begin{bmatrix} Z_n & 0_{mn} \\ 0_{mn} & Z_n \end{bmatrix}
\]

where

\[
Z_n = \begin{bmatrix} Z_1 & \cdots & 0_n \\ \vdots & \ddots & \vdots \\ 0_n & \cdots & Z_m \end{bmatrix}.
\]
Thus, the eigenvalues of $X_{2m}^n$ are precisely the eigenvalues of $n^{-m/2}X_{n,1} \cdots X_{n,m}$, each with multiplicity $2m$.

Fix $\delta > 0$. By part [i] of Theorem 4.2.1 and Proposition 3.1.1, we find that almost surely, for $n$ sufficiently large, the eigenvalues of $n^{-m/2}X_{n,1} \cdots X_{n,m}$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. By (4.8), we conclude that almost surely, for $n$ sufficiently large, the eigenvalues of $X_n$ are contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \delta\}$. Therefore, almost surely, for $n$ sufficiently large $X_n - zI$ is invertible for all $|z| > 1 + \delta$. For all such values of $z$, we define

$$R_n(z) := (X_n - zI_{2mn})^{-1}.$$ 

Since $A_{n}^\square$ has rank $O(1)$ and operator norm $O(1)$, we can decompose (by the singular value decomposition) $A_{n}^\square = B_nC_n$, where $B_n$ is a $mn \times k$ matrix, $C_n$ is a $k \times mn$ matrix, $k = O(1)$, and both $B_n$ and $C_n$ have rank $O(1)$ and operator norm $O(1)$.

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

$$\det(L_n - zI_{2mn}) = \det(X_n + A_{n}^\square - zI_{2mn}) = 0$$

if and only if

$$\det(I_k + C_nR_n(z)B_n) = 0$$

by Lemma 3.3.2.

Using Schur’s Compliment to calculate the $mn \times mn$ blocks of $R_n(z)$, we can see that $R_n(z) = \left( \begin{array}{cc} zG_n(z^2) & G_n(z^2) \\ I + z^2G_n(z^2) & zG_n(z^2) \end{array} \right)$, where $G_n(z) := \left( \frac{1}{\sqrt{n}} \gamma_n - zI \right)^{-1}$ (which is defined for $|z| > 1 + \delta$ by Theorem 4.2.1); hence

$$R_n(z)_{[a,b]} = \begin{cases} 
  zG_n(z^2) & \text{if } a = b \\
  G_n(z^2) & \text{if } a = 1, b = 2 \\
  I + z^2G_n(z^2) & \text{if } a = 2, b = 1.
\end{cases}$$

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

$$u^*R_n(z)v = \sum_{1 \leq a,b \leq 2} (u^*)^{[a]} R_n(z)^{[a,b]} v^{[b]}$$
where $v^{[1]}, v^{[2]}$ denote the $mn \times 1$ sub-blocks of the vector $v$ and $(u^*)^{[1]}, (u^*)^{[2]}$ denote the $1 \times mn$ sub-blocks of the vector $u^*$. Additionally, if $u$ and $v$ have uniformly bounded norm for all $n$, then by Theorem 4.2.1, almost surely

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

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

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

and

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

We note that the off-diagonal blocks of $R_n(z)$ have a much different behavior than $G_n(z)$. In order to keep track of this behavior, define

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

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

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

Applying this to (4.9), we obtain that

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

almost surely. Hence, Lemma D.1.6 reveals that almost surely

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

By another application of Sylvester’s determinant formula (4.1) and by noticing that $\mathcal{H}_n A_n$ is the zero matrix, this can be rewritten as

$$\sup_{|z| > 1+\delta} \left| \det \left( I + R_n(z) A_n \right) - \det \left( I - \frac{1}{z} A_n \right) \right| = o(1).$$
Finally, since the eigenvalues of $A_n$ are all zero, $\det \left( I - \frac{1}{z}A_n \right) = 1$ so we can write the above statement as

$$\sup_{|z| > 1 + \delta} |\det (R_n) \det (L_n - zI) - 1| = o(1)$$

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

### 4.2.3 Proof of Theorem 2.1.8

Recall that for this result, we define

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

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

Let $\mathcal{Y}_n$ and $A_n$ be defined as in (4.2) and (4.4). Using Proposition 3.1.1, it will suffice to study the eigenvalues of $\frac{1}{\sqrt{n}} \mathcal{Y}_n + A_n$. In particular, let $\varepsilon' > 0$ such that $(1 + 2\varepsilon)^{1/m} = 1 + \varepsilon'$. Then we want to find solutions to

$$\det \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n + A_n - zI \right) = 0 \quad (4.10)$$

for $|z| \geq 1 + \varepsilon'$. By part (i) of Theorem 4.2.1 almost surely, for $n$ sufficiently large $\frac{1}{\sqrt{n}} \mathcal{Y}_n - zI$ is invertible for all $|z| \geq 1 + \varepsilon'$. By supposition, we can decompose (using the singular value decomposition) $A_n = B_nC_n$, where $B_n$ is $mn \times k$, $C_n$ is $k \times mn$, $k = O(1)$, and both $B_n$ and $C_n$ have rank $O(1)$ and operator norm $O(1)$. Thus, by Lemma 3.3.2 we need to investigate the values of $z \in \mathbb{C}$ with $|z| \geq 1 + \varepsilon'$ such that

$$\det(I_k + C_n G_n(z) B_n) = 0,$$
where $G_n(z)$ is defined in (4.3).

Since $k = O(1)$, Theorem 4.2.1 implies that

$$\sup_{|z| \geq 1+\varepsilon'} \left| C_n G_n(z) B_n - \left( -\frac{1}{z} \right) C_n B_n \right| \to 0$$

almost surely as $n \to \infty$. By Lemma D.0.6, this implies that

$$\sup_{|z| \geq 1+\varepsilon'} \left| \det \left( I_k + C_n G_n(z) B_n \right) - \det \left( I_n - \frac{1}{z} A_n \right) \right| \to 0$$

almost surely. By an application of Sylvester’s determinant theorem (4.1), this is equivalent to

$$\sup_{|z| \geq 1+\varepsilon'} \left| \det \left( I_k + C_n G_n B_n \right) - \det \left( I_n - \frac{1}{z} A_n \right) \right| \to 0$$

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

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

Since the eigenvalues of $A_n^m$ are precisely the eigenvalues of $A_n$, each with multiplicity $m$, it follows that $g$ has precisely $l := jm$ roots $\lambda_1(A_n), \ldots, \lambda_l(A_n)$ outside the disk $\{z \in \mathbb{C} : |z| < 1 + \varepsilon'\}$. Thus, by (4.12) and Rouché’s theorem, almost surely, for $n$ sufficiently large,

$$f(z) := \det \left( I_k + C_n G_n B_n \right)$$

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

Returning to (4.10), we conclude that almost surely for $n$ sufficiently large, $\frac{1}{\sqrt{n}} Y_n + A_n$ has exactly $l$ roots outside the disk $\{z \in \mathbb{C} : |z| < 1 + \varepsilon'\}$, and after possibly reordering the eigenvalues, these roots take the values

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

for $1 \leq i \leq l$.

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

$$(\lambda_i(A_n) + o(1))^m = \lambda_i(A_n^m) + o(1)$$
since $A_n$ has spectral norm $O(1)$, we conclude that almost surely, for $n$ sufficiently large, $P_n$ has exactly $j$ eigenvalues in the region $\{ z \in \mathbb{C} : |z| \geq 1 + 2\epsilon \}$, and after reordering the indices correctly

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

for $1 \leq i \leq j$. This completes the proof of Theorem 2.1.8.

4.2.4 Proof of Theorems 2.1.6 and 2.1.7

In the proofs of Theorems 2.1.6 and 2.1.7 we will make use of an alternative isotropic law which is a corollary of Theorem 4.2.1. We state and prove the result now.

**Corollary 4.2.2 (Alternative Isotropic Law).** Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \ldots, \xi_m$ are complex-valued random variables with mean zero, unit variance, finite fourth moments, and independent real and imaginary parts. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Then, for any fixed $\delta > 0$, the following statements hold.

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

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

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

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

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

almost surely as $n \to \infty$. 

Proof. Part (i) follows from Theorem 2.1.2. Let \( \mathcal{Y}_n \) be defined by (4.2), and let \( \mathcal{G}_n(z) \) be defined by (4.3). Then the last two parts of Corollary 4.2.2 follow from the last two parts of Theorem 4.2.1 due to the fact that
\[
\mathcal{G}_n^{[1,1]}(z) = z^{m-1} \left( n^{-m/2} X_{n,1} \cdots X_{n,m} - z^m I \right)^{-1}.
\]

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

Proof of Theorem 2.1.6. For this proof, we let \( P_n \) denote the product
\[
P_n := n^{-m/2} \prod_{k=1}^{m} X_{n,k} + A_n.
\]
By rescaling by \( \frac{1}{\sigma} \) it suffices to assume that \( \sigma_i = 1 \) for \( 1 \leq i \leq m \). Let \( \varepsilon > 0 \). Assume that for sufficiently large \( n \) there are no eigenvalues of \( A_n \) in the band \( \{ z \in \mathbb{C} : 1 + \varepsilon < |z| < 1 + 3\varepsilon \} \) and there are \( j \) eigenvalues, \( \lambda_1(A_n), \ldots, \lambda_j(A_n) \), in the region \( \{ z \in \mathbb{C} : |z| \geq 1 + 3\varepsilon \} \).

By Corollary 4.2.2, almost surely, for \( n \) sufficiently large, \( n^{-m/2} X_{n,1} \cdots X_{n,m} - zI \) is invertible for all \( |z| > 1 + \varepsilon \). We decompose (using the singular value decomposition) \( A_n = B_n C_n \), where \( B_n \) is \( n \times k \), \( C_n \) is \( k \times n \), \( k = O(1) \), and both \( B_n \) and \( C_n \) have rank \( O(1) \) and spectral norm \( O(1) \). By Lemma 3.3.2, the eigenvalues of \( P_n \) outside \( \{ z \in \mathbb{C} : |z| < 1 + 2\varepsilon \} \) are precisely the values of \( z \in \mathbb{C} \) with \( |z| > 1 + 2\varepsilon \) such that
\[
\det \left( I_k + C_n \left( n^{-m/2} X_{n,1} \cdots X_{n,m} - zI_n \right)^{-1} B_n \right) = 0.
\]
By Corollary 4.2.2
\[
\sup_{|z| > 1+\varepsilon} \left\| C_n \left( n^{-m/2} X_{n,1} \cdots X_{n,m} - zI_n \right)^{-1} B_n + \frac{1}{z} C_n B_n \right\| \to 0
\]
almost surely as \( n \to \infty \). By applying (4.1) and Lemma D.0.6 this gives
\[
\sup_{|z| > 1+\varepsilon} \left| \det \left( I_k + C_n \left( n^{-m/2} X_{n,1} \cdots X_{n,m} - zI_n \right)^{-1} B_n \right) - \det \left( I_n - \frac{1}{z} A_n \right) \right| \to 0
\]
almost surely as \( n \to \infty \). By recognizing the roots of \( \det \left( I_n - \frac{1}{z} A_n \right) \) as the eigenvalues of \( A_n \), Rouché’s Theorem implies that almost surely for \( n \) sufficiently large \( P_n \) has exactly \( j \) eigenvalues.
in the region \( \{ z \in \mathbb{C} : |z| > 1 + 2\varepsilon \} \), and after labeling the eigenvalues properly,

\[ \lambda_i(P_n) = \lambda_i(A_n) + o(1) \]

for \( 1 \leq i \leq j \). \( \square \)

The proof of Theorem 2.1.7 will require the following corollary of [114] Lemma 2.3.

**Lemma 4.2.3.** Let \( \varphi_n \) and \( X_{n,1}, \ldots, X_{n,m} \) be as in Theorem 2.1.7. Then almost surely,

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

**Proof.** Let \( u := n^{-(m-1)/2} \varphi_n^* X_{n,1} X_{n,2} \cdots X_{n,m-1} \). In view of [114], Theorem 1.4, it follows that \( \|u\| = O(1) \) almost surely. We now condition on \( X_{n,1}, \ldots, X_{n,m-1} \) so that \( \|u\| = O(1) \). As \( X_{n,m} \) is independent of \( u \), we apply [114], Lemma 2.3 to conclude that

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

almost surely, concluding the proof. \( \square \)

With this result, we may proceed to the proof of Theorem 2.1.7.

**Proof of Theorem 2.1.7.** For this proof, recall that we work with the product

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

By rescaling by \( \frac{1}{\sigma} \) it suffices to assume that \( \sigma_i = 1 \) for \( 1 \leq i \leq m \). Fix \( \gamma > 0 \) and let \( \varepsilon > 0 \). By Corollary 4.2.2 and Lemma 3.3.2 almost surely, for \( n \) sufficiently large, the only eigenvalues of \( P_n \) in the region \( \{ z \in \mathbb{C} : |z| > 1 + \varepsilon \} \) are the values of \( z \in \mathbb{C} \) with \( |z| > 1 + \varepsilon \) such that

\[ 1 + \mu n^\gamma \phi_n^* \left( n^{-m/2} X_{n,1} \cdots X_{n,m} - z I_n \right)^{-1} \phi_n = 0. \]  \( (4.13) \)

Define the functions

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

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

in the region \( \{ z \in \mathbb{C} : |z| > 1 + 2\varepsilon \} \), and after labeling the eigenvalues properly,
Observe that $g(z)$ has one zero located at $\mu n^\gamma$ which will be outside the disk $\{ z \in \mathbb{C} : |z| \leq 1 + \varepsilon \}$ for large enough $n$. By Corollary 4.2.2, it follows that almost surely

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

Thus, almost surely

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

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

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

For the next step of the proof, we will need to bound the spectral norm of $n^{-m/2}X_{n,1} \cdots X_{n,m}$. To do so, we apply [114, Theorem 1.4] and obtain that, almost surely, for $n$ sufficiently large,

$$n^{-m/2}||X_{n,1} \cdots X_{n,m}|| \leq (2.1)^m. \quad (4.14)$$

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

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

By Lemma 4.2.3, $\phi_n^* \left(n^{-m/2}X_{n,1} \cdots X_{n,m}\right) \varphi_n = o(1)$ almost surely, so one can see

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

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

$$|z||f(z) - g(z)| = o\left(\frac{n^\gamma}{|z|}\right) \quad (4.15)$$

almost surely. Since any root of $zf(z)$ outside $\{ z \in \mathbb{C} : |z| \leq 1 + \varepsilon \}$ must tend to infinity with $n$, it follows from Rouché’s theorem that almost surely, for $n$ sufficiently large, $zf(z)$ has precisely one root outside the disk $\{ z \in \mathbb{C} : |z| \leq 1 + \varepsilon \}$ and that root takes the value $\mu n^\gamma + o(n^\gamma)$. 
It remains to reduce the error from $o(n^γ)$ to $o(1)$. Fix $δ > 0$, and let $Γ$ be the circle around $µn^γ$ with radius $δ$. Then from (4.15) we see that almost surely

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

Hence, almost surely, for $n$ sufficiently large,

$$|z||f(z) - g(z)| < δ = |z||g(z)|$$

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

4.3 Truncation and Useful Tools

We now turn to the proof of Theorem 4.2.1. We will require the following standard truncation results for iid random matrices.

**Lemma 4.3.1.** Let $ξ$ be a complex-valued random variable with mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. Let $\text{Re}(ξ)$ and $\text{Im}(ξ)$ denote the real and imaginary parts of $ξ$ respectively, and let $\sqrt{-1}$ denote the imaginary unit. For $L > 0$, define

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

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

and

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

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

(i) $\text{Var}(\hat{ξ}) \geq \frac{1}{2}$

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

$\hat{\xi}$ has mean zero, unit variance, $E|\hat{\xi}|^4 \leq CE|\xi|^4$ for some absolute constant $C > 0$, and the real and imaginary parts of $\hat{\xi}$ are independent.

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

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

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

\[
+ \sqrt{-1} \left( \text{Im}(X_{(i,j)}) 1_{\{|\text{Im}(X_{(i,j)})| \leq L/\sqrt{2}\}} - E \left[ \text{Im}(X_{(i,j)}) 1_{\{|\text{Im}(X_{(i,j)})| \leq L/\sqrt{2}\}} \right] \right)
\]

(4.16)

and

\[
\hat{X}_{(i,j)} := \frac{\tilde{X}_{(i,j)}}{\sqrt{\text{Var}(\tilde{X}_{(i,j)})}}
\]

(4.17)

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

**Lemma 4.3.2.** Let $X$ be an iid random matrix with atom variable $\xi$ which has mean zero, unit variance, $m_4 := E|\xi|^4 < \infty$, and independent real and imaginary parts. Let $\hat{X}$ be as defined in (4.17). Then, there exist constants $C, L_0 > 0$ (depending only on $m_4$) such that for all $L > L_0$

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

almost surely.

**Proof.** Let $\tilde{X}$ be defined as in (4.16), and let $L_0$ be the value from Lemma 4.3.1. Begin by noting that

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

and thus it suffices to show that

\[
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left\| X - \tilde{X} \right\| \leq \frac{C}{L}
\]
and
\[
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left\| \tilde{X} - \hat{X} \right\| \leq \frac{C}{L}
\]
almost surely. First, by Lemma 4.3.1
\[
\begin{align*}
\frac{1}{\sqrt{n}} \left\| \tilde{X} - \hat{X} \right\| &= \frac{1}{\sqrt{n}} \left\| \hat{X} \right\| \left(1 - \frac{1}{\sqrt{\text{Var} (\xi)}}\right) \\
&\leq \frac{1}{\sqrt{n}} \left\| \hat{X} \right\| \sqrt{2} \left| \sqrt{\text{Var} (\xi)} - 1 \right| \\
&\leq \frac{1}{\sqrt{n}} \left\| \hat{X} \right\| \sqrt{2} \left| \text{Var} (\xi) - 1 \right| \\
&\leq \frac{1}{\sqrt{n}} \left\| \hat{X} \right\| \sqrt{2} \left(\frac{4}{L^2} E|\xi|^4 \right).
\end{align*}
\]
(4.18)

By [114, Theorem 1.4], we find that almost surely
\[
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left\| \tilde{X} \right\| \leq 2,
\]
and thus by (4.18)
\[
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left\| \tilde{X} - \hat{X} \right\| \leq \frac{C}{L}
\]
almost surely for all \(L \geq \max\{1, L_0\}\).

Next consider \(\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left\| X - \tilde{X} \right\|\). Note that \(X - \tilde{X}\) is an iid matrix with atom variable
\[
\begin{align*}
\text{Re}(X_{(i,j)}) 1_{\{|\text{Re}(X_{(i,j)})| > L/\sqrt{2}\}} &- \mathbb{E} \left[ \text{Re}(X_{(i,j)}) 1_{\{|\text{Re}(X_{(i,j)})| > L/\sqrt{2}\}} \right] \\
+ \sqrt{-1} \left( \text{Im}(X_{(i,j)}) 1_{\{|\text{Im}(X_{(i,j)})| > L/\sqrt{2}\}} - \mathbb{E} \left[ \text{Im}(X_{(i,j)}) 1_{\{|\text{Im}(X_{(i,j)})| > L/\sqrt{2}\}} \right] \right).
\end{align*}
\]
Thus, each entry has mean zero, variance
\[
\text{Var}((X - \tilde{X})_{i,j}) \leq \frac{8}{L^2} \mathbb{E}|\xi|^4,
\]
and finite fourth moment. Thus, again by [114, Theorem 1.4],
\[
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left\| X - \tilde{X} \right\| \leq \frac{C}{L}
\]
almost surely, and the proof is complete. \(\Box\)
We now consider the iid random matrices $X_{n,1}, \ldots, X_{n,m}$ from Theorem 4.2.1. For each \(1 \leq k \leq m\), define the truncation $\hat{X}_{n,k}$ as was done above for $\hat{X}$ in (4.17). Define $\hat{Y}_n$ by

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

(4.19)

Using Theorem 4.3.1, we have the following corollary.

**Corollary 4.3.3.** Let $X_{n,1}, \ldots, X_{n,m}$ be independent iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, each of which has mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. Let $\hat{X}_{n,1}, \ldots, \hat{X}_{n,m}$ be the truncations of $X_{n,1}, \ldots, X_{n,m}$ as was done in (4.17). In addition, let $Y_n$ be as defined in (4.2) and $\hat{Y}_n$ be as defined in (4.19). Then there exist constants $C, L_0 > 0$ (depending only on the atom variables $\xi_1, \ldots, \xi_m$) such that

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

(4.20)

almost surely for all $L > L_0$.

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

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

Therefore, the claim follows from Lemma 4.3.2.

## 4.4 Least Singular Value Bounds

In this section, we study the least singular value of $\frac{1}{\sqrt{n}} Y_n - zI$. We begin by recalling Weyl's inequality for the singular values (see, for example, [31, Problem III.6.5]), which states that for $n \times n$ matrices $A$ and $B$,

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

(4.21)

We require the following theorem, which is based on [90, Theorem 2].
Theorem 4.4.1. Fix $L > 0$, and let $\xi_1, \ldots, \xi_m$ be complex-valued random variables, each having mean zero, unit variance, independent real and imaginary parts, and which satisfy
\[
\operatorname{sup}_{1 \leq k \leq m} |\xi_k| \leq L
\]
almost surely. Let $X_{n,1}, \ldots, X_{n,m}$ be independent iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Define $\mathcal{Y}_n$ as in (4.2), and fix $\delta > 0$. Then there exists a constant $c > 0$ (depending only on $\delta$) such that
\[
\inf_{|z| \geq 1+\delta} \frac{1}{s_{mn}} s_{mn} \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n - z I \right) \geq c
\]
with overwhelming probability.

A similar statement was proven in [90, Theorem 2], where the same conclusion was shown to hold almost surely rather than with overwhelming probability. However, many of the intermediate steps in [90] are proven to hold with overwhelming probability. We use these intermediate steps to prove Theorem 4.4.1 in Appendix B.

Lemma 4.4.2. Let $X_{n,1}, \ldots, X_{n,m}$ satisfy the assumptions of Theorem 4.2.1, and let $\mathcal{Y}_n$ be as defined in (4.2). Fix $\delta > 0$. Then there exists a constant $c > 0$ such that almost surely, for $n$ sufficiently large,
\[
\inf_{|z| \geq 1+\delta} \frac{1}{s_{mn}} s_{mn} \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n - z I \right) \geq c.
\]

Proof. Let $L > 0$ be a large constant to be chosen later. Let $\hat{X}_{n,1}, \ldots, \hat{X}_{n,m}$ be defined as in (4.17), and let $\hat{\mathcal{Y}}_n$ be defined as in (4.19). By Theorem 4.4.1 and the Borel–Cantelli lemma, there exists a constant $c' > 0$ such that almost surely, for $n$ sufficiently large,
\[
\inf_{|z| \geq 1+\delta} \frac{1}{s_{mn}} s_{mn} \left( \frac{1}{\sqrt{n}} \tilde{\mathcal{Y}}_n - z I \right) \geq c'.
\]

By Corollary 4.3.3 and (4.21) we may choose $L$ sufficiently large to ensure that almost surely, for $n$ sufficiently large,
\[
\left| s_{mn} \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n - z I \right) - s_{mn} \left( \frac{1}{\sqrt{n}} \hat{\mathcal{Y}}_n - z I \right) \right| \leq \frac{c'}{2},
\]
uniformly in \( z \). We conclude that almost surely, for \( n \) sufficiently large,

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

and the proof is complete.

With this result we can prove the following lemma.

**Lemma 4.4.3.** Let \( X_{n,1}, \ldots, X_{n,m} \) satisfy the assumptions of Theorem 4.2.1 and fix \( \delta > 0 \). Then there exists a constant \( c > 0 \) such that almost surely, for \( n \) sufficiently large,

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

**Proof.** Let \( \mathcal{Y}_n \) be defined as in (4.2). Then Lemma 4.4.2 implies that almost surely, for \( n \) sufficiently large, \( \frac{1}{\sqrt{n}} Y_n - zI \) is invertible for all \( |z| \geq 1 + \delta \). By computing the block inverse of this matrix, we find

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

Thus, for \( |z| \geq 1 + \delta \),

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

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

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

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

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

and the claim follows from Lemma 4.4.2.

**Remark 4.4.4.** Observe that \( z \) is an eigenvalue of \( \frac{1}{\sqrt{n}} \mathcal{Y}_n \) if and only if \( \det \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) = 0 \). Also, recall that \( \det \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) = \prod_i \det \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n - zI \right) \); this product is zero if and only if the smallest
singular value is zero. Since Lemma 4.4.2 and Lemma 4.4.3 bound the least singular values of $\frac{1}{\sqrt{n}}Y_n - zI$ and $n^{-m/2}X_{n,1} \cdots X_{n,m} - zI$ away from zero for $|z| \geq 1 + \delta$, we can conclude that such values of $z$ are almost surely, for $n$ sufficiently large, not eigenvalues of these matrices.

4.5 Reductions to the Proof of Theorem 4.2.1

In this section, we will prove that it is sufficient to reduce the proof of Theorem 4.2.1 to the case in which the entries of each matrix are truncated and where we restrict $z$ to the band $5 \leq |z| \leq 6$.

**Theorem 4.5.1.** Let $X_{n,1}, \ldots, X_{n,m}$ be as in Theorem 4.2.1. Then there exists a constant $L_0$ such that the following holds for all $L > L_0$. Let $\hat{X}_{n,1}, \ldots, \hat{X}_{n,m}$ be defined as in (4.17), let $\hat{Y}_n$ be given by (4.19), and let $\hat{G}_n(z) := \left(\frac{1}{\sqrt{n}}\hat{Y}_n - zI\right)^{-1}$.

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

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

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

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

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

almost surely as $n \rightarrow \infty$.

We now prove Theorem 4.2.1 assuming Theorem 4.5.1.

**Proof of Theorem 4.2.1.** Part (i) of Theorem 4.2.1 follows from Lemma 4.4.2 (see Remark 4.4.4). In addition, part (ii) of Theorem 4.2.1 follows from an application of Lemma 4.4.2 and Proposition 3.3.1.
We now turn to the proof of part (iii). Fix $0 < \delta < 1$. Let $\mathcal{Y}_n$ be given by (4.2), and let $\mathcal{G}_n$ be given by (4.3). Let $\varepsilon, \varepsilon' > 0$, and observe that there exists a positive constant $M_1$ such that for $|z| \geq M_1$,

$$\left\| \left( -\frac{1}{z} \right) u^* v \right\| \leq \frac{1}{z} \left\| u^* \right\| \left\| v \right\| \leq \varepsilon. \quad (4.22)$$

Also, by [114, Theorem 1.4] and Lemma D.0.10 there exists a constant $M_2 > 0$ such that almost surely, for $n$ sufficiently large,

$$\sup_{|z| \geq M_2} \left\| u^* \mathcal{G}_n(z) v \right\| \leq \sup_{|z| \geq M_2} \left\| \mathcal{G}_n(z) \right\| \leq \frac{\varepsilon}{2}. \quad (4.23)$$

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

$$\sup_{|z| \geq M} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \leq \varepsilon. \quad (4.22)$$

We now work on the region where $1 + \delta < |z| \leq M$. By the resolvent identity (3.6), we note that

$$\left\| \mathcal{G}_n(z) - \hat{\mathcal{G}}_n(z) \right\| \leq \left\| \mathcal{G}_n(z) \right\| \left\| \hat{\mathcal{G}}_n(z) \right\| \frac{1}{\sqrt{n}} \left\| \hat{\mathcal{Y}}_n - \mathcal{Y}_n \right\|. \quad (4.23)$$

Thus, by part (iii) of Theorem 4.2.1 (proven above), Theorem 4.5.1 and Corollary 4.3.3 there exist constants $C, c > 0$ such that

$$\limsup_{n \to \infty} \sup_{|z| > 1 + \delta} \left\| \mathcal{G}_n(z) - \hat{\mathcal{G}}_n(z) \right\| \leq \limsup_{n \to \infty} c^2 \frac{1}{\sqrt{n}} \left\| \hat{\mathcal{Y}}_n - \mathcal{Y}_n \right\| \leq \frac{c^2 C}{L} \leq \frac{\varepsilon'}{2} \quad (4.23)$$

almost surely for $L$ sufficiently large.

From (4.23) and Theorem 4.5.1 almost surely, for $n$ sufficiently large,

$$\sup_{5 \leq |z| \leq 6} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| \leq \sup_{5 \leq |z| \leq 6} \left| u^* \mathcal{G}_n(z) v - u^* \hat{\mathcal{G}}_n(z) v \right| + \sup_{5 \leq |z| \leq 6} \left| u^* \hat{\mathcal{G}}_n(z) v + \frac{1}{z} u^* v \right| \leq \varepsilon'. \quad (4.23)$$

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

$$\limsup_{n \to \infty} \sup_{5 \leq |z| \leq 6} \left| u^* \mathcal{G}_n(z) v + \frac{1}{z} u^* v \right| = 0$$
almost surely. Since the region \( \{ z \in \mathbb{C} : 1 + \delta \leq |z| \leq M \} \) is compact and contains the region \( \{ z \in \mathbb{C} : 5 \leq |z| \leq 6 \} \), Vitali’s Convergence Theorem\(^1\) (see, for instance, [19, Lemma 2.14]) implies that we can extend this convergence to the larger region, and we obtain

\[
\limsup_{n \to \infty} \sup_{1+\delta \leq |z| \leq M} \left| u^* G_n(z)v + \frac{1}{z} u^* v \right| = 0
\]

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

\[
\sup_{1+\delta \leq |z| \leq M} \left| u^* G_n(z)v + \frac{1}{z} u^* v \right| \leq \varepsilon.
\]

Combined with (4.22), this completes the proof of Theorem 4.2.1 (since \( \varepsilon > 0 \) was arbitrary). \( \square \)

It remains to prove Theorem 4.5.1. We prove parts (i) and (ii) of Theorem 4.5.1 now. The proof of part (iii) is lengthy and will be addressed in the forthcoming sections of Chapter 4.

**Proof of Theorem 4.5.1 (i) and (ii)** Let \( \delta > 0 \), and observe that by Theorem 4.4.1 and the Borel–Cantelli lemma, there exists a constant \( c > 0 \) (depending only on \( \delta \)) such that almost surely, for \( n \) sufficiently large,

\[
\inf_{|z| > 1+\delta} s_{mn} \left( \frac{1}{\sqrt{n}} \hat{Y}_n - zI \right) \geq c. \tag{4.24}
\]

This implies (see Remark 4.4.4) that almost surely, for \( n \) sufficiently large, the eigenvalues \( \frac{1}{\sqrt{n}} \hat{Y}_n \) are contained in the disk \( \{ z \in \mathbb{C} : |z| \leq 1 + \delta \} \), proving (i). From (4.24) and Proposition 3.3.1, we conclude that almost surely, for \( n \) sufficiently large,

\[
\sup_{|z| > 1+\delta} \left\| G_n(z) \right\| \leq \frac{1}{c},
\]

proving (ii) \( \square \)

### 4.6 Concentration of Bilinear Forms Involving the Resolvent \( G_n \)

Sections 4.6 and 4.7 are devoted to the proof of part (iii) of Theorem 4.5.1. Let \( \hat{X}_{n,1}, \ldots, \hat{X}_{n,m} \) be the truncated matrices from Theorem 4.5.1 and let \( u_n, v_n \in \mathbb{C}^{mn} \) be deterministic unit vectors.

\(^1\) The hypothesis of Vitali’s Convergence Theorem are satisfied almost surely, for \( n \) sufficiently large, by parts (i) and (ii) of Theorem 4.2.1. In addition, one can check that \( (u_n)^* G_n(z)v_n \) is holomorphic in the region \( \{ z \in \mathbb{C} : |z| > 1 + \delta \} \) almost surely, for \( n \) sufficiently large, using the resolvent identity (3.6).
For ease of notation, in Sections 4.6 and 4.7, we drop the decorations and write $X_{n,1}, \ldots, X_{n,m}$ for $\hat{X}_{n,1}, \ldots, \hat{X}_{n,m}$. Similarly, we write $\mathcal{Y}_n$ for $\hat{\mathcal{Y}}_n$ and $\mathcal{G}_n$ for $\hat{\mathcal{G}}_n$. Recall that all constants and asymptotic notation may depend on $m$ without explicitly showing the dependence.

Define the following event:

$$\Omega_n := \left\{ \frac{1}{\sqrt{n}} \|\mathcal{Y}_n\| \leq 4.5 \right\}. \tag{4.25}$$

**Lemma 4.6.1** (Spectral Norm Bound for $\mathcal{Y}_n$). Under the assumptions above, the event $\Omega_n$ holds with overwhelming probability.

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

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

Therefore, the claim follows from [19, Theorem 5.9] (alternatively, [114, Theorem 1.4]). In fact, the constant 4.5 can be replaced with any constant strictly larger than 2; 4.5 will suffice for what follows. □

By Lemma 4.6.1, $\Omega_n$ holds with overwhelming probability, i.e., for every $A > 0$,

$$\mathbb{P}(\Omega_n) = 1 - O_A(n^{-A}). \tag{4.26}$$

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

**Lemma 4.6.2** (Concentration). Let $u_n, v_n \in \mathbb{C}^{mn}$ be deterministic unit vectors. Then, under the assumptions above, for any $\varepsilon > 0$, almost surely

$$\sup_{5 \leq |i| \leq 6} \left| u_n^* \mathcal{G}_n(z)v_n 1_{\Omega_n} - \mathbb{E}[u_n^* \mathcal{G}_n(z)v_n 1_{\Omega_n}] \right| < \varepsilon \tag{4.27}$$

for $n$ sufficiently large.

The proof of Lemma 4.6.2 follows the arguments of [16, 94]. Before we begin the proof, we present some notation. Define $\mathcal{Y}_n^{(k)}$ to be the matrix $\mathcal{Y}_n$ with all entries in the $k$th row and the $k$th
column filled with zeros. Note that $\mathcal{Y}_n^{(k)}$ is still an $mn \times mn$ matrix. Define

$$G_n^{(k)} := \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n^{(k)} - zI \right)^{-1}, \quad (4.28)$$

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

$$\mathcal{F}_k := \sigma(r_1, \ldots, r_k, c_1, \ldots, c_k) \quad (4.29)$$

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

$$E_k[\cdot] := E[\cdot | \mathcal{F}_k] \quad (4.30)$$

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

$$\Omega_n^{(k)} := \left\{ \frac{1}{\sqrt{n}} \| \mathcal{Y}_n^{(k)} \| \leq 4.5 \right\}. \quad (4.31)$$

Observe that $\Omega_n \subseteq \Omega_n^{(k)}$ and therefore, by Lemma 4.6.1, $\Omega_n^{(k)}$ holds with overwhelming probability as well.

Remark 4.6.3. By Lemma [D.0.10] we have that

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

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

We will now collect some preliminary calculations and lemmas that will be required for the proof of Lemma 4.6.2.

Lemma 4.6.4 (Rosenthal’s Inequality; [11]). Let $\{x_k\}$ be a complex martingale difference sequence with respect to the filtration $\mathcal{F}_k$. Then for $p \geq 2$,

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

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

Lemma 4.6.5. Let $A$ be an $n \times n$ Hermitian positive semidefinite matrix, and let $S \subset [n]$. Then
\[ \sum_{i \in S} A_{ii} \leq \text{tr} A. \]

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

Lemma 4.6.6. Let $A$ be an $N \times N$ Hermitian positive semidefinite matrix with rank at most one. Suppose that $\xi$ is a complex-valued random variable with mean zero, unit variance, and which satisfies $|\xi| \leq L$ almost surely for some constant $L > 0$. Let $S \subset [N]$, 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 $i \not\in S$.

Then for any $p \geq 1$,
\[ \mathbb{E} |w^* Aw|^p \ll_{L, p} \|A\|^p. \] (4.32)

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

By Lemma [D.0.3] we get
\[ \mathbb{E} |w^* Aw|^p \ll_p (\text{tr} A_{S \times S})^p + \mathbb{E} |\xi|^{2p} \text{tr} A_{S \times S}^p \leq (\text{tr} A_{S \times S})^p + L^{2p} \text{tr} A_{S \times S}^p. \]

Since the rank of $A_{S \times S}$ is at most one, we find
\[ \text{tr} A_{S \times S} \leq \|A\| \]

and
\[ \text{tr} A_{S \times S}^p \leq \|A\|^p, \]
where we used the fact that the operator norm of a matrix bounds the operator norm of any sub-matrix. We conclude that

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

as desired. ∎

**Lemma 4.6.7.** Let $A$ be a deterministic complex $N \times N$ matrix. Suppose that $\xi$ is a complex-valued random variable with mean zero, unit variance, and which satisfies $|\xi| \leq L$ almost surely for some constant $L > 0$. Let $S, R \subseteq [N]$, and let $w = (w_i)_{i=1}^N$ and $t = (t_j)_{j=1}^N$ be independent vectors with the following properties:

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

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

Then for any $p \geq 1$,

$$\mathbb{E}|w^*At|^p \ll_{L,p} (\text{tr}(A^*A))^{p/2}.$$  \hspace{1cm} (4.33)

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

We first note that, by the Cauchy–Schwarz inequality, it suffices to assume $p$ is even. In this case, since $w$ is independent of $t$, Lemma D.0.3 implies that

$$\mathbb{E}|w^*At|^p = \mathbb{E}|w^*Att^*A^*w|^{p/2}$$

$$= \mathbb{E}|w_S^*(Att^*A^*)_{S \times S}w_S|^{p/2}$$

$$\ll_p \mathbb{E}\left[(\text{tr}(Att^*A^*))_{S \times S}^{p/2} + L^p \text{tr}(Att^*A^*)_{S \times S}^{p/2}\right].$$

Recall that for any matrix $B$, $\text{tr}(B^*B)^{p/2} \leq (\text{tr}(B^*B))^{p/2}$. By this fact and by Lemma 4.6.5, we observe that

$$\mathbb{E}\left[(\text{tr}(Att^*A^*))_{S \times S}^{p/2} + L^p \text{tr}(Att^*A^*)_{S \times S}^{p/2}\right] \ll_{L,p} \mathbb{E}\left[(\text{tr}(Att^*A^*))^{p/2}\right].$$
By a cyclic permutation of the trace, we have
\[
E \left[ (tr(Att^*A^*))^{p/2} \right] = E \left[ (t^*A^*At)^{p/2} \right] \leq E |t^*A^*At|^{p/2}.
\]

By Lemma D.0.3, Lemma 4.6.5, and a similar argument as above, we obtain
\[
E |t^*A^*At|^{p/2} = E |t_R^*(A^*A)_{R\times R}t_R|^{p/2} \leq L p \frac{tr(A^*A)}{R^2}.
\]

completing the proof.

Lemma 4.6.8. Let \( r_k \) be the \( k \)th row of \( Y_n \), \( c_k \) be the \( k \)th column of \( Y_n \), \( G_n^{(k)}(z) \) be the resolvent of \( Y_n^{(k)} \), and \( u_n \in \mathbb{C}^{mn} \) be a deterministic unit vector. Then, under the assumptions above,
\[
\sup_{5 \leq |z| \leq 6} \mathbb{E}_{k-1} \left| \frac{1}{n} r_k G_n^{(k)}(z) u_n u_n^* G_n^{(k)*}(z) 1_{\Omega_n^{(k)}} r_k^* \right|^p \ll_{L,p} n^{-p} \tag{4.34}
\]
and
\[
\sup_{5 \leq |z| \leq 6} \mathbb{E}_{k-1} \left| \frac{1}{n} c_k^* G_n^{(k)*}(z) u_n u_n^* G_n^{(k)*}(z) 1_{\Omega_n^{(k)}} c_k \right|^p \ll_{L,p} n^{-p} \tag{4.35}
\]
almost surely.

Proof. We will only prove the bound in (4.34) as the proof of (4.35) is identical. For each fixed \( z \)
in the band \( 5 \leq |z| \leq 6 \), we will show that
\[
\mathbb{E}_{k-1} \left| \frac{1}{n} r_k G_n^{(k)}(z) u_n u_n^* G_n^{(k)*}(z) 1_{\Omega_n^{(k)}} r_k^* \right|^p \ll_{L,p} n^{-p}
\]
surely, where the implicit constant does not depend on \( z \). The claim then follows by taking the supremum over all \( z \) in the band \( 5 \leq |z| \leq 6 \).

To this end, fix \( z \) with \( 5 \leq |z| \leq 6 \). Throughout the proof, we drop the dependence on \( z \) in the resolvent as it is clear from context. Note that
\[
\mathbb{E}_{k-1} \left| \frac{1}{n} r_k G_n^{(k)}(z) u_n u_n^* G_n^{(k)*}(z) 1_{\Omega_n^{(k)}} r_k^* \right|^p = \frac{1}{n^p} \mathbb{E}_{k-1} \left| r_k \left( G_n^{(k)}(z) u_n u_n^* G_n^{(k)*}(z) 1_{\Omega_n^{(k)}} \right) r_k^* \right|^p,
\]
and \(G_n^{(k)} u_n u_n^* G_n^{(k)*} 1_{\Omega_n^{(k)}}\) is independent of \(r_k\). In addition, observe that \(G_n^{(k)} u_n u_n^* G_n^{(k)*} 1_{\Omega_n^{(k)}}\) is Hermitian positive semidefinite matrix with rank at most one. Applying Lemma 4.6.6 and Remark 4.6.3 we obtain

\[
\mathbb{E}_{k-1} \left| \frac{1}{n} r_k G_n^{(k)} u_n u_n^* G_n^{(k)*} 1_{\Omega_n^{(k)}} \right|^p \ll_{L,p} \frac{1}{n^p} \mathbb{E}_{k-1} \left| G_n^{(k)} u_n u_n^* G_n^{(k)*} 1_{\Omega_n^{(k)}} \right|^p \ll_{L,p} n^{-p}
\]

surely, and the proof is complete. \(\square\)

**Lemma 4.6.9.** Let \(\zeta_1, \ldots, \zeta_{mn}\) be complex-valued random variables (not necessarily independent) which depend on \(Y_n\). Assume

\[
\sup_k |\zeta_k| 1_{\Omega_n^{(k)}} = O(1)
\]

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

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

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

\[
\mathbb{E} \left| \sum_{k=1}^{mn} (\zeta_k 1_{\Omega_n^{(k)}} - \zeta_k 1_{\Omega_n \cap \Omega_n^{(k)}}) \right|^p \leq \mathbb{E} \left| \sum_{k=1}^{mn} \zeta_k 1_{\Omega_n^{(k)} \cap \Omega_n^{(k)}} \right|^p \ll_p \mathbb{E} \left( \sum_{k=1}^{mn} 1_{\Omega_n^{(k)} \cap \Omega_n^{(k)}} \right)^p \ll_p n^{pP}(\Omega_n^{c}),
\]

and the claim follows from (4.26). \(\square\)

Now, we proceed to prove the main result of this section, Lemma 4.6.2.

**Proof of Lemma 4.6.2.** Define

\[
Y_n^{(k1)} := Y_n^{(k)} + e_k r_k, \quad Y_n^{(k2)} := Y_n^{(k)} + c_k e_k^*
\]

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

\[
G_n^{(k)} := \left( \frac{1}{\sqrt{n}} Y_n^{(k)} - z I \right)^{-1}, \quad j = 1, 2,
\]

(4.36)
and set

\[ \alpha_n^{(k)} := \frac{1}{1 + z^{-1} n^{-1} r_k G_n^{(k)} c_k \Omega_n}, \]

\[ \zeta_n^{(k)} := n^{-1} r_k G_n^{(k)} c_k, \]

\[ \eta_n^{(k)} := n^{-1} r_k G_n^{(k)} v_n u_n G_n^{(k)} c_k. \]

Using these definitions, we make the following observations:

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

\[ e_k^* G_n^{(k)} e_k = -z^{-1}, \quad e_k^* G_n^{(k)} v_n = -z^{-1} v_{n,k}, \quad u_n G_n^{(k)} e_k = -z^{-1} \bar{u}_{n,k} \]

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

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

\[ e_k^* G_n^{(k)} c_k = 0, \quad r_k G_n^{(k)} e_k = 0. \]

(iii) By \( \text{[3.3]}, \text{[i]}, \) and \( \text{[ii]} \)

\[ e_k^* G_n^{(k)} c_k = e_k^* G_n^{(k)} n^{-1/2} c_k - \frac{e_k^* G_n^{(k)} e_k n^{-1/2} r_k G_n^{(k)} c_k}{1 + n^{-1/2} r_k G_n^{(k)} e_k} = z^{-1} n^{-1} r_k G_n^{(k)} c_k, \]

so that

\[ \frac{1}{1 + e_k^* G_n^{(k)} \Omega_n n^{-1/2} c_k} = \alpha_n^{(k)}. \]

(iv) By the same argument as \( \text{[iii]} \)

\[ n^{-1/2} r_k G_n^{(k)} e_k = z^{-1} n^{-1} r_k G_n^{(k)} c_k, \]

so that

\[ \frac{1}{1 + n^{-1/2} r_k G_n^{(k)} e_k \Omega_n} = \alpha_n^{(k)}. \]
(v) By Schur’s compliment,
\[(G_n)_{(k,k)} = -\frac{1}{z + n^{-1}r_k G_n^{(k)} c_k}\]
provided the necessary inverses exist (which is the case on the event \(\Omega_n\)). Thus, on \(\Omega_n = \Omega_n \cap \Omega_n^{(k)}\) and uniformly for \(5 \leq |z| \leq 6\), by Remark 4.6.3,
\[
\left| \alpha_n^{(k)} \right| = \left| z \right| \left| (G_n)_{(k,k)} \right|
\leq 12.

On \(\Omega_n^c\), \(\alpha_n^{(k)} = 1\), so we have that, almost surely,
\[
\left| \alpha_n^{(k)} \right| \leq 12.
\]

(vi) By (3.4) and (iii),
\[
u_n^* G_n n^{-1/2} c_k 1_{\Omega_n} = \frac{u_n^* G_n^{(k1)} n^{-1/2} c_k 1_{\Omega_n}}{1 + e_k^* G_n^{(k1)} n^{-1/2} c_k} = \frac{u_n^* G_n^{(k1)} n^{-1/2} c_k 1_{\Omega_n}}{1 + e_k^* G_n^{(k1)} n^{-1/2} c_k 1_{\Omega_n}} = u_n^* G_n^{(k1)} n^{-1/2} c_k 1_{\Omega_n} \alpha_n^{(k)}.
\]

(vii) By (3.4) and (iv), a similar argument as above gives
\[
u_n^* G_n e_k 1_{\Omega_n} = u_n^* G_n^{(k1)} e_k 1_{\Omega_n} \alpha_n^{(k)}.
\]

(viii) By (3.3) and (iii),
\[
G_n^{(k1)} = G_n^{(k)} - G_n^{(k)} e_k n^{-1/2} r_k G_n^{(k)} e_k
= G_n^{(k)} - G_n^{(k)} e_k n^{-1/2} r_k G_n^{(k)}.
\]

(ix) By (3.3) and (ii), and by the same calculation as in (viii),
\[
G_n^{(k2)} = G_n^{(k)} - G_n^{(k)} e_k n^{-1/2} c_k e_k^* G_n^{(k)}.
\]
(x) By definition of $\alpha_n^{(k)}$,

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

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

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

(xii) The entries of $c_k$ and $r_k$ have mean zero, unit variance, and are bounded by $4L$ almost surely. In addition, $(r_k^T, c_k)$ and $G_n^{(k)} 1_{\Omega_n^{(k)}}$ are independent. By Remark 4.6.3,

$$\|G_n^{(k)} u_n v_n^* G_n^{(k)} v_n u_n^* c_n^{(k)}\| \leq 16$$
on $\Omega_n^{(k)}$. Thus, by Lemma 4.6.7 for any $p \geq 2$,

$$\sup_{1 \leq k \leq n} \mathbb{E}_{k-1}[|\eta_n^{(k)}|^p 1_{\Omega_n}] \leq \sup_{1 \leq k \leq n} \mathbb{E}_{k-1}[|\eta_n^{(k)}|^p 1_{\Omega_n^{(k)}}] \ll_{L,p} \sup_{1 \leq k \leq n} n^{-p} \mathbb{E}_{k-1} \left[ \left( \text{tr}(G_n^{(k)} u_n v_n^* G_n^{(k)} v_n u_n^* c_n^{(k)}) \right)^{p/2} 1_{\Omega_n^{(k)}} \right] \ll_{L,p} n^{-p}$$

since

$$G_n^{(k)} u_n v_n^* G_n^{(k)} v_n u_n^* c_n^{(k)}$$
is at most rank one. Similarly, Remark 4.6.3 give the almost sure bound $\|G_n^{(k)} G_n^{(k)}\| \leq 4$ on $\Omega_n^{(k)}$, which gives

$$\sup_{1 \leq k \leq n} \mathbb{E}_{k-1}[|\zeta_n^{(k)}|^p 1_{\Omega_n}] \leq \sup_{1 \leq k \leq n} \mathbb{E}_{k-1}[|\zeta_n^{(k)}|^p 1_{\Omega_n^{(k)}}] \ll_{L,p} \sup_{1 \leq k \leq n} n^{-p} \mathbb{E}_{k-1} \left[ \left( \text{tr}(G_n^{(k)} G_n^{(k)}) \right)^{p/2} 1_{\Omega_n^{(k)}} \right] \ll_{L,p} n^{-p/2}.$$
is Lipschitz continuous in the region \( \{ z \in \mathbb{C} : 5 \leq |z| \leq 6 \} \). Thus, by a standard net argument, it suffices to prove that almost surely, for \( n \) sufficiently large,

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

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

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

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

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

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

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

and so

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

for all \( p > 2 \). Define

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

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

From the resolvent identity (3.6), we observe that

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

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

\[
:= -\sum_{k=1}^{mn} (W_{k1} + W_{k2})
\]
where we define
\[
W_{k1} := (E_k - E_{k-1}) \left[ u_n^* n^{-1/2} c_k r_k G_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right],
\]
and
\[
W_{k2} := (E_k - E_{k-1}) \left[ u_n^* n^{-1/2} c_k e_k^* G_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right].
\]

By (i), (vii), and (ix) we can further decompose
\[
\sum_{k=1}^{mn} W_{k1} = \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ u_n^* n^{-1/2} c_k r_k G_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right]
\]
\[
= \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ u_n^* n^{-1/2} c_k r_k G_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]
\]
\[
= - \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ z^{-1} (\bar{u}_{n,k} - u_n^* n^{-1/2} c_k) n^{-1/2} r_k G_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]
\]
\[
= - \sum_{k=1}^{mn} (W_{k11} + W_{k12})
\]

where
\[
W_{k11} := (E_k - E_{k-1}) \left[ z^{-1} \bar{u}_{n,k} n^{-1/2} r_k G_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]
\]
and
\[
W_{k12} := - (E_k - E_{k-1}) \left[ z^{-1} u_n^* n^{-1} c_k r_k G_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right].
\]
Similarly, by (i), (vi), (viii), and (x),

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

\[
= \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n 1_{\Omega_n} \right]
\]

\[
= \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n 1_{\Omega_n} \right]
\]

\[
= \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n 1_{\Omega_n} \right]
\]

\[
= \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k e_k^* \mathcal{G}_n^{(k)} v_n 1_{\Omega_n} \right]
\]

\[
= - \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} 1_{\Omega_n} \right]
\]

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

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

\[
= - \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ z^{-1} v_{n,k} u_n^* \mathcal{G}_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} 1_{\Omega_n} \right]
\]

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

where

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

and

\[
W_{k22} := - (E_k - E_{k-1}) \left[ z^{-1} \bar{u}_{n,k} v_{n,k} \alpha_n^{(k)} 1_{\Omega_n} \right].
\]

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

\[
\| \sum_{k=1}^{mn} W_{k11} \|^p + \| \sum_{k=1}^{mn} W_{k12} \|^p + \| \sum_{k=1}^{mn} W_{k21} \|^p + \| \sum_{k=1}^{mn} W_{k22} \|^p = O_L (n^{-p/2}). \quad (4.40)
\]

We bound each term individually. To begin, observe that by Lemma 4.6.4, Lemma 4.6.8 and (v).
for any $p > 2$,

$$
E \left| \sum_{k=1}^{mn} W_{k1} \right|^p = E \left| \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ z^{-1} \bar{u}_{n,k} n^{-1/2} \tau_k G_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]^p \right|
$$

$$
\ll_p E \left[ \sum_{k=1}^{mn} \left| (E_k - E_{k-1}) \left[ z^{-1} \bar{u}_{n,k} n^{-1/2} \tau_k G_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right]^2 \right]^{p/2} \right]
$$

+ \left| (E_k - E_{k-1}) \left[ z^{-1} \bar{u}_{n,k} n^{-1/2} \tau_k G_n^{(k)} v_n \alpha_n^{(k)} \mathbf{1}_{\Omega_n} \right] \right|^p

$$
\ll_p E \left[ \sum_{k=1}^{mn} \left| \bar{u}_{n,k}^2 E_{k-1} \left[ n^{-1/2} \tau_k G_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right]^2 \right]^{p/2} \right]
$$

+ \sum_{k=1}^{mn} \left| \bar{u}_{n,k} \right|^p E \left| n^{-1/2} \tau_k G_n^{(k)} v_n \mathbf{1}_{\Omega_n} \right|^p

$$
\ll_p E \left[ \sum_{k=1}^{mn} \left| \bar{u}_{n,k}^2 E_{k-1} \left[ n^{-1/2} \tau_k G_n^{(k)} v_n \mathbf{1}_{\Omega_n}^{(k)} \right]^2 \right]^{p/2} \right]
$$

+ \sum_{k=1}^{mn} \left| \bar{u}_{n,k} \right|^p E \left| n^{-1/2} \tau_k G_n^{(k)} v_n \mathbf{1}_{\Omega_n}^{(k)} \right|^p

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

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

where we also used Jensen’s inequality and the fact that $|z| \geq 5$. Similarly, by Lemma 4.6.4, Lemma
4.6.8 and (v), for any $p > 2,$

\[
\begin{align*}
\mathbb{E} \left| \sum_{k=1}^{mn} W_{k21} \right|^p & = \mathbb{E} \left( \sum_{k=1}^{mn} (E_k - E_{k-1}) \left[ z^{-1} v_{n,k} u_n^* G_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} 1_{\Omega_n} \right] \right)^p \\
& \ll_p \mathbb{E} \left( \sum_{k=1}^{mn} (E_k - E_{k-1}) \left| z^{-1} v_{n,k} u_n^* G_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} 1_{\Omega_n} \right|^2 \right)^{p/2} \\
& \quad + \mathbb{E} \left( \sum_{k=1}^{mn} (E_k - E_{k-1}) \left| z^{-1} v_{n,k} u_n^* G_n^{(k)} n^{-1/2} c_k \alpha_n^{(k)} 1_{\Omega_n} \right|^p \right) \\
& \ll_p \mathbb{E} \left( \sum_{k=1}^{mn} \left| v_{n,k} \right|^2 E_{k-1} \left| u_n^* G_n^{(k)} n^{-1/2} c_k 1_{\Omega_n} \right|^2 \right)^{p/2} \\
& \quad + \sum_{k=1}^{mn} \left| v_{n,k} \right|^p E \left| u_n^* G_n^{(k)} n^{-1/2} c_k 1_{\Omega_n} \right|^p \\
& \ll_p \mathbb{E} \left( \sum_{k=1}^{mn} \left| v_{n,k} \right|^2 E_{k-1} \left| u_n^* G_n^{(k)} n^{-1/2} c_k 1_{\Omega_n^{(k)}} \right|^2 \right)^{p/2} \\
& \quad + \sum_{k=1}^{mn} \left| v_{n,k} \right|^p E \left| u_n^* G_n^{(k)} n^{-1/2} c_k 1_{\Omega_n^{(k)}} \right|^p \\
& \ll_{L,p} \left( \mathbb{E} \left( \sum_{k=1}^{mn} \left| v_{n,k} \right|^2 n^{-}\right)^{p/2} + \sum_{k=1}^{mn} \left| v_{n,k} \right|^p n^{-p/2} \right) \\
& \ll_{L,p} n^{-p/2}.
\end{align*}
\]
Next, by Lemma 4.6.4 (v) and (xii) for any $p > 2$,
\[
\mathbb{E} \left| \sum_{k=1}^{mn} W_{k12} \right|^p = \mathbb{E} \left| \sum_{k=1}^{mn} (E_k - E_{k-1}) [z^{-1} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}] \right|^p \\
= \mathbb{E} \left| \sum_{k=1}^{mn} (E_k - E_{k-1}) [z^{-1} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}] \right|^p \\
\ll_p \mathbb{E} \left[ \sum_{k=1}^{mn} (E_k - E_{k-1}) [z^{-1} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}]^2 \right]^{p/2} \\
+ \mathbb{E} \left[ \sum_{k=1}^{mn} (E_k - E_{k-1}) [z^{-1} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}] \right]^p \\
\ll_p \left( \mathbb{E} \left[ \sum_{k=1}^{mn} (E_k - E_{k-1}) [z^{-1} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}]^2 \right]^{p/2} + \mathbb{E} \left[ \sum_{k=1}^{mn} \eta_n^{(k)} 1_{\Omega_n} \right]^p \right) \\
\ll_{L,p} \left( \mathbb{E} \left[ \sum_{k=1}^{mn} n^{-2} \right]^{p/2} + \mathbb{E} \left[ \sum_{k=1}^{mn} n^{-p} \right] \right) \\
\ll_{L,p} n^{-p/2 + n^{-p+1}} \\
\ll_{L,p} n^{-p/2},
\]
where we also used Jensen’s inequality and the fact that $|z| \geq 5$; the last inequality follows from the fact that $p > 2$.

Finally, moving on to $W_{k22}$, by (xi) we can decompose this further as
\[
\sum_{k=1}^{mn} W_{k22} = \sum_{k=1}^{mn} (E_k - E_{k-1}) [z^{-1} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}] \\
= \sum_{i=1}^{mn} (E_k - E_{k-1}) [z^{-1} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}] \\
= \sum_{k=1}^{mn} (W_{k221} + W_{k222})
\]
where
\[
W_{k221} := (E_k - E_{k-1}) [z^{-1} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}] 
\]
and
\[
W_{k222} := -(E_k - E_{k-1}) [z^{-2} \bar{u}_n v_n \alpha_n^{(k)} 1_{\Omega_n}] 
\]
Since $z^{-1}u_{n,k}v_{n,k}$ is deterministic, it follows that

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

Thus, it suffices to show that $E|\sum_{k=1}^{mn} W_{k222}|^p = O_L(p(n^{-p/2})$ for $p > 2$. By Lemma 4.6.4 (v) and (xii), we have that for any $p > 2$,

$$E \left| \sum_{k=1}^{mn} W_{k222} \right|^p \ll_p E \left[ \sum_{k=1}^{mn} (E_k - E_{k-1}) \left| z^{-2} \bar{u}_{n,k} v_{n,k} c_n^{(k)} \alpha_n^{(k)} 1_{\Omega_n} \right|^2 \right]^{p/2}$$

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

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

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

where we again used Jensen’s inequality, the bound $|z| \geq 5$, and the fact that $u_n$ and $v_n$ are unit vectors. This completes the proof of (4.40), and hence the proof of Lemma 4.6.2 is complete. □

### 4.7 Proof of Theorem 4.5.1

In this section we complete the proof of Theorem 4.5.1. We continue to work under the assumptions and notation introduced in Section 4.6.

It remains to prove part (iii) of Theorem 4.5.1. In view of Lemma 4.6.2 and (4.26), it suffices to show that

$$\sup_{5 \leq |z| \leq 6} \left| E[u_n^* G_n(z) v_n 1_{\Omega_n}] + \frac{1}{z} u_n^* v_n \right| = o(1).$$  (4.41)
4.7.1 Neumann Series

We will rewrite the resolvent, \( G_n \), as a Neumann series. Indeed, for \(|z| \geq 5\),
\[
\frac{1}{\sqrt{n}} \left\| \frac{Y_n}{z} \right\| \leq \frac{9}{10} < 1
\] (4.42)
on the event \( \Omega_n \), so we may write
\[
G_n(z)1_{\Omega_n} = \left( I1_{\Omega_n} + \sum_{k=1}^{\infty} \left( \frac{1}{\sqrt{n}} \frac{Y_n1_{\Omega_n}}{z} \right)^{k} \right)
= -\frac{1}{z} I1_{\Omega_n} - \sum_{k=1}^{\infty} \left( \frac{\sqrt{n}1_{\Omega_n}}{z} \right)^{k}
\] almost surely. Therefore, using (4.26), we obtain
\[
E\left[ u_n^* G_n v_n 1_{\Omega_n} \right] = E \left[ -\frac{1}{z} u_n^* v_n 1_{\Omega_n} - u_n^* \sum_{k=1}^{\infty} \left( \frac{\sqrt{n}1_{\Omega_n}}{z} \right)^{k} v_n 1_{\Omega_n} \right]
= -\frac{1}{z} u_n^* v_n P(\Omega_n) - \sum_{k=1}^{\infty} \frac{1}{z^{k+1}} E \left[ u_n^* \left( \frac{\sqrt{n}1_{\Omega_n}}{z} \right)^{k} v_n 1_{\Omega_n} \right]
= -\frac{1}{z} u_n^* v_n + o(1) - \sum_{k=1}^{\infty} \frac{1}{z^{k+1}} E \left[ u_n^* \left( \frac{\sqrt{n}1_{\Omega_n}}{z} \right)^{k} v_n 1_{\Omega_n} \right].
\] (4.43)
Showing the sum in (4.43) converges to zero uniformly in the region \( \{ z \in \mathbb{C} : 5 \leq |z| \leq 6 \} \) will complete the proof of (4.41).

4.7.2 Removing the Indicator Function

In this subsection, we prove the following.

**Lemma 4.7.1.** Under the assumptions above, for any integer \( k \geq 1 \),
\[
\left| E \left[ u_n^* \left( \frac{1}{\sqrt{n}} Y_n \right)^{k} v_n \right] - E \left[ u_n^* \left( \frac{1}{\sqrt{n}} Y_n \right)^{k} v_n 1_{\Omega_n} \right] \right| = o_{k,L}(1).
\]
**Proof.** Since the entries of \( Y_n \) are truncated, it follows that
\[
\| Y_n \| \leq \| Y_n \|_2 \ll L n
\]
almost surely. Therefore, as $P(\Omega_c^n) = O(n^{-A})$ for any $A > 0$, we obtain

$$\left| \mathbb{E}\left[u_n^*\left(\frac{1}{\sqrt{n}}Y_n\right)^k v_n\right] - \mathbb{E}\left[u_n^*\left(\frac{1}{\sqrt{n}}Y_n\right)^k v_n\textbf{1}_{\Omega_n}\right] \right| \leq \mathbb{E}\left[\|Y_n\|^k \textbf{1}_{\Omega_n}\right] \leq L n^k P(\Omega_c^n)$$

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

4.7.3 Combinatorial Arguments

In this section, we will show that

$$\sup_{5 \leq |z| \leq 6} \left| \sum_{k=1}^{\infty} \frac{\mathbb{E}\left[u_n^*\left(\frac{1}{\sqrt{n}}Y_n\right)^k v_n\textbf{1}_{\Omega_n}\right]}{z^{k+1}} \right| = o_L(1). \quad (4.44)$$

In view of (4.42), the tail of the series can easily be controlled. Thus, it suffices to show that, for each integer $k \geq 1$,

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

By Lemma 4.7.1, it suffices to prove the statement above without the indicator function. In particular, the following lemma completes the proof of (4.44) (and hence completes the proof of Theorem 4.5.1).

**Lemma 4.7.2 (Moment Calculations).** Under the assumptions above, for any integer $k \geq 1$,

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

To prove Lemma 4.7.2, we will expand the above expression in terms of the entries of the random matrices $X_{n,1}, \ldots, X_{n,m}$. For brevity, in this section we will drop the subscript $n$ from our notation and just write $X_1, \ldots, X_m$ for $X_{n,1}, \ldots, X_{n,m}$. Similarly, we write the vectors $u_n$ and $v_n$ as $u$ and $v$, respectively.
To begin, we exploit the block structure of $Y_n$ and write
\[
E \left[ u^* \left( \frac{1}{\sqrt{n}} Y_n \right)^k v \right] = n^{-k/2} E \left[ \sum_{1 \leq a,b \leq m} (u^*)^{[a]} \left( Y_n^k \right)^{[a,b]} v^{[b]} \right]
\]
\[
= n^{-k/2} \sum_{1 \leq a,b \leq m} (u^*)^{[a]} E \left[ \left( Y_n^k \right)^{[a,b]} v^{[b]} \right].
\] (4.45)

Due to the block structure of $Y_n$, for each $1 \leq a \leq m$, there exists some $1 \leq b \leq m$ which depends on $a$ and $k$ such that the $[a,b]$ block of $Y_n^k$ is nonzero and all other blocks, $[a,c]$ for $c \neq b$, are zero. Additionally, the nonzero block entry $(Y_n^k)^{[a,b]}$ is $X_{a}X_{a+1} \cdots X_{a+k}$ where the subscripts are reduced modulo $m$ and we use modular class representatives $\{1, \ldots, m\}$ (as opposed to the usual $\{0, \ldots, m-1\}$). To show that the expectation of this sum is $o_{L,k}(1)$, we need to systematically count all terms which have a nonzero expectation. To do so, we use graphs to characterize the terms which appear in the sum. In particular we develop path graphs, each of which corresponds uniquely to a term in the expansion of (4.45). For each term, the corresponding path graph will record the matrix entries which appear, the order in which they appear, and the matrix from which they come. We begin with the following definitions.

**Definition 4.7.3.** We consider graphs where each vertex is specified by the ordered pair $(t, i_t)$ for integers $1 \leq t \leq k+1$ and $1 \leq i_t \leq n$. We call $t$ the *time* coordinate of the vertex and $i_t$ the *height* coordinate of the vertex. Let $(i_1, i_2, \ldots, i_{k+1})$ be a $k+1$ tuple of integers from the set $\{1, 2, \ldots, n\}$ and let $1 \leq a \leq m$. We define an *$m$-colored $k$-path graph* $G^a(i_1, i_2, \ldots, i_{k+1})$ to be the edge-colored directed graph with vertex set $V = \{(1, i_1), (2, i_2), \ldots, (k+1, i_{k+1})\}$ and directed edges from $(t, i_t)$ to $(t+1, i_{t+1})$ for $1 \leq t \leq k$, where the edge from $(t, i_t)$ to $(t+1, i_{t+1})$ is color $a + t - 1$, with the convention that colors are reduced modulo $m$, and the modulo class representatives are $\{1, \ldots, m\}$. The graph is said to *visit* a vertex $(t, i_t)$ if there exists an edge which begins or terminates at that vertex, so that $(t, i_t) \in V$. We say an edge of $G^a(i_1, \ldots, i_{k+1})$ is of *type I* if it terminates on a vertex $(t, i_t)$ such that $i_t \neq i_s$ for all $s < t$. We say an edge is of *type II* if it terminates on a vertex $(t, i_t)$ such that there exists some $s < t$ with $i_t = i_s$.

We make some observations about this definition.
Remark 4.7.4. We view $k$ and $n$ as specified parameters. Once these parameters are specified, the notation $G^a(i_1, \ldots, i_{k+1})$ completely determines the graph. The vertex set $V$ is a subset of the vertices of the $(k+1) \times n$ integer lattice and each graph has exactly $k$ directed edges. In each graph, there is an edge which begins on vertex $(1, i_1)$ for some $1 \leq i_1 \leq n$ and there is an edge which terminates on vertex $(k+1, i_{k+1})$ for some $1 \leq i_{k+1} \leq n$. Each edge begins at $(t, i_t)$ and terminates on $(t+1, i_{t+1})$ for some integers $1 \leq t \leq k$ and $1 \leq i_t, i_{t+1} \leq n$. Additionally, each edge is one of $m$ possible colors. Note that if we think about the edges as ordered by the time coordinate, then the order of the colors is a cyclic permutation of the coloring $1, 2, \ldots, m$, beginning with $a$. This cycle is repeated as many times as necessary in order to cover all edges.

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

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

**Definition 4.7.5.** For a given $m$-colored $k$-path graph $G^a(i_1, i_2, \ldots, i_{k+1})$, we say two edges $e_1$ and $e_2$ of $G^a(i_1, i_2, \ldots, i_{k+1})$ are time-translate parallel if $e_1$ begins at vertex $(t, i_t)$ and terminates at vertex $(t+1, i_{t+1})$ and edge $e_2$ begins at vertex $(t', i_{t'})$ and terminates at vertex $(t'+1, i_{t'+1})$ where $i_t = i_{t'}$ and $i_{t+1} = i_{t'+1}$ for some $1 \leq t, t' \leq k$ with $t \neq t'$.

**Remark 4.7.6.** Intuitively, two edges are time-translate parallel if they span the same two height coordinates at different times. Throughout this section, we shorten the term “time-translate parallel” and refer to edges with this property as “parallel” for brevity. We warn the reader that by this, we mean that the edges must span the same heights at two different times. For instance, the edge from $(1, 2)$ to $(2, 4)$ is not parallel to the edge from $(3, 1)$ to $(4, 3)$ since they don’t span the same heights, although these edges might appear parallel in the geometric interpretation of the
word. Also note that two parallel edges need not have the same color. See Figure 4.3 for examples of parallel and non parallel edges.

**Definition 4.7.7.** For a fixed \( k \), we say two \( m \)-colored \( k \)-path graphs \( G^a(i_1, \ldots, i_{k+1}) \) and \( G^a(i'_1, \ldots, i'_{k+1}) \) are equivalent, denoted \( G^a(i_1, \ldots, i_{k+1}) \sim G^a(i'_1, \ldots, i'_{k+1}) \), if there exists some permutation \( \sigma \) of \( \{1, \ldots, n\} \) such that \( (i_1, \ldots, i_{k+1}) = (\sigma(i'_1), \ldots, \sigma(i'_{k+1})) \). Note here that for two path graphs to be equivalent, the color of the first edge, and hence the color of all edges sequentially, must be the same in both.

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

**Definition 4.7.8.** For each equivalence class of graphs, the canonical \( m \)-colored \( k \)-path graph is the unique graph from an equivalence class which satisfies the following condition: If \( G^a(i_1, \ldots, i_{k+1}) \) visits vertex \((t, i_t)\), then for every \( 0 < i < i_t \) there exists \( s < t \) such that \( G^a(i_1, \ldots, i_{k+1}) \) visits \((s, i)\).

**Remark 4.7.9.** Observe that, intuitively, the canonical representation for \( G^a(i_1, \ldots, i_{k+1}) \) is the graph which does not “skip over” any height coordinates. Namely, the canonical graph necessarily begins at vertex \((1, 1)\) and at each time step the height of the next vertex can be at most one larger than the maximum height of all previous vertices.

Now, we return to the task at hand: the proof of Lemma 4.7.2. We fix a positive integer \( k \), and we expand as in (4.45).

For a fixed value \( 1 \leq a \leq m \), consider the nonzero block \( (\gamma^k_n)^{[a,b]} \). We further expand to see

\[
(u^*)^{[a]}E\left[\left(\gamma^k_n\right)^{[a,b]}\right]v^{[b]} = \sum_{i_1, \ldots, i_{k+1}} \bar{u}_{i_1}^{[a]}E[X_{a, (i_1, i_2)} X_{a+1, (i_2, i_3)} \cdots X_{a+k, (i_{k+1}, i_{k+1})}]v^{[b]}_{i_{k+1}},
\]

(4.46)

where the subscripts \( a, \ldots, a+k \) are reduced mod \( m \) with representatives \( \{1, \ldots, m\} \). Observe that by the structure of \( \gamma_n \), these subscripts must appear cyclically in the order \( a, a+1, \ldots, m, 1, \ldots, a-1 \), with the order repeating as many times as necessary before ending at \( b \). In particular, the subscripts are uniquely determined by the starting subscript \( a \) and the value of \( k \).
Figure 4.1: The graphs featured here are two possible 2-colored 3-path graphs, $G^1(4,1,4,3)$ and $G^1(1,2,1,3)$, which correspond to the terms $X_{1,4}X_{2,1}X_{1,4}$ and $X_{1,2}X_{2,1}X_{1,3}$, respectively. These two graphs are equivalent. The first graph is not a canonical graph while the second graph is a canonical graph. For ease of notation and clarity, we have drawn all vertices on the integer lattice as dots, with the time axis appearing horizontally and the height axis vertically. While each dot represents a vertex in the integer lattice, not all of the dots represent vertices in the path graphs. Only dots from which an edge begins or terminates are vertices of the path graph. We have also colored the edges with blue and green to represent the two colors. The colors of each edge are also represented by the number in parenthesis, e.g., (1) or (2). In either graph, the first and third edges are type I edges, while the second edge in each graph is type II. See Example 4.7.12 for further discussion.

We now consider the expectation on the right-hand side of (4.46). Since all entries of each matrix are independent, if an index appears only once in a product, that product will have expectation zero. Therefore, only terms in which every index appears more than once will have a nonzero expectation, and only such terms will contribute to the expected value of the sum. Note that for an entry to appear more than once, we not only need the index of the entries to match but also which of the $m$ matrices the entries came from. The following definition will assist in encoding each entry on the right-hand side of (4.46) as a unique $m$-colored $k$-path graph.

**Definition 4.7.10.** We say the term $X_{a,(i_1,i_2)}X_{a+1,(i_2,i_3)}\cdots X_{a+k,(i_k,i_{k+1})}$ from the expansion of (4.46) corresponds to the $m$-colored $k$-path graph $G^a(i_1,\ldots,i_{k+1})$. We use the notation

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

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

Each term in the expansion of (4.46) corresponds uniquely to an $m$-colored $k$-path graph. In
terms of the corresponding $m$-colored $k$-path graph, if an edge spans two vertices $(t, i_t)$ and $(t + 1, i_{t+1})$, and has color $a$ then the corresponding matrix product must contain the entry $X_{a,(i_t,i_{t+1})}$.

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

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

**Lemma 4.7.11.** If two path graphs $G^a(i_1,\ldots,i_{k+1})$ and $G^a(i'_1,\ldots,i'_{k+1})$ are equivalent, then

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

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

**Example 4.7.12.** Consider two 2-colored 3-path graphs: $G^1(4,1,4,3)$ and $G^1(1,2,1,3)$. $G^1(4,1,4,3)$ is the leftmost graph in Figure 4.1 and $G^1(1,2,1,3)$ is the rightmost graph in Figure 4.1. They correspond to the terms $X_{1,(4,1)}X_{2,(1,4)}X_{1,(4,3)}$ and $X_{1,(1,2)}X_{2,(2,1)}X_{1,(1,3)}$ respectively from the expansion of $u^*\left(\frac{1}{\sqrt{n}}Y_n\right)^3 v$ where $m = 2$. Note that these graphs are equivalent by the permutation which maps $4 \leftrightarrow 1 \leftrightarrow 2 \leftrightarrow 4$. $G^1(4,1,4,3)$ is not a canonical graph, while $G^1(1,2,1,3)$ is a canonical graph. Observe that since $X_{1,(4,1)}$ appeared first in the product $X_{1,(4,1)}X_{2,(1,4)}X_{1,(4,3)}$, the first edge in the corresponding path graph is an edge of color 1 spanning from height coordinate 4 to height coordinate 1.
**Example 4.7.13.** Consider a product of the form

\[ X_{1,(i_1,i_2)}X_{1,(i_2,i_1)}X_{1,(i_1,i_2)}X_{1,(i_2,i_1)}X_{1,(i_1,i_2)}X_{1,(i_2,i_1)}, X_{1,(i_1,i_2)} \]

where \( i_1, i_2 \) are distinct. This corresponds to a 1-colored 7-path graph whose canonical representative can be seen in Figure 4.2. Since all entries come from matrix \( X_1 \), we know \( m = 1 \). Since there are 7 terms in the product, \( k = 7 \). This term has expected value \( \mathbb{E} [(X_{1,(i_1,i_2)})^4] \mathbb{E} [(X_{1,(i_2,i_1)})^3] \). Since \( X_1 \) is an iid matrix, the particular choice of \( i_1 \neq i_2 \) is irrelevant to the expected value.

![Figure 4.2](image)

**Figure 4.2:** This 1-colored 7-path graph corresponds to the product in Example 4.7.13. Since all edges are the same color, they are not labeled with distinct colors. Observe that every edge in this graph is parallel to at least one other edge.

**Example 4.7.14.** Consider a product of the form

\[ X_{1,(i_1,i_2)}X_{1,(i_2,i_3)}X_{1,(i_3,i_1)}X_{1,(i_2,i_3)}X_{1,(i_3,i_4)}X_{1,(i_3,i_4)} \]

where \( i_1, i_2, i_3 \) and \( i_4 \) are distinct. This corresponds to a 1-colored 7-path graph whose canonical representative is featured in Figure 4.3. Since the only index pair which appears more than once is \( (i_3, i_4) \), the corresponding term will have zero expectation.

**Example 4.7.15.** Let \( i_1, i_2, i_3, i_4 \) be distinct, and consider the product

\[ X_{1,(i_1,i_2)}X_{2,(i_2,i_1)}X_{3,(i_3,i_2)}X_{4,(i_4,i_1)}X_{1,(i_1,i_2)}X_{2,(i_2,i_1)}X_{3,(i_3,i_2)}X_{4,(i_2,i_1)}. \]

Since there are 8 entries in this product, \( k = 8 \), and as there are entries from 4 matrices, \( m = 4 \). The corresponding canonical 4-colored 8-path graph representative is shown in Figure 4.4. This term has expectation

\[ \mathbb{E} \left[ (X_{1,(i_1,i_2)})^2(X_{2,(i_2,i_1)})^2(X_{3,(i_1,i_2)})^2(X_{4,(i_2,i_1)})^2 \right]. \]
Figure 4.3: This 1-colored 7-path graph is the canonical representative to the graph which corresponds to the product in Example 4.7.14. Note that the edge from (5, 3) to (6, 4) is parallel to the edge from (7, 3) to (8, 4), but no other edges in the graph are parallel. This implies that the corresponding term will have zero expectation.

Figure 4.4: This 4-colored 8-path graph corresponds to the product in Example 4.7.15. Since there are eight edges, this is one possible product from \(u^* \left( (n^{-1/2}Y_n)^8 v \right).\)

**Example 4.7.16.** Consider the product

\[ X_{3,(i_1,i_2)}X_{1,(i_2,i_3)}X_{2,(i_3,i_4)}X_{3,(i_4,i_1)}X_{1,(i_1,i_2)}X_{2,(i_2,i_4)}X_{3,(i_4,i_2)}X_{1,(i_2,i_1)}X_{2,(i_1,i_5)}. \]

Since there are 9 entries in this product, \(k = 9\) and we can see that there are 3 different matrices so that \(m = 3\). For any distinct indices \(i_1, i_2, i_3, i_4, i_5\), the corresponding canonical 3-colored 9-path graph representative is shown in Figure 4.5. In this product, every entry appears only once. Thus the expectation of this product factors, and the term will have expectation zero.

We now complete the proof of Lemma 4.7.2, which will occupy the remainder of the section.

Let

\[ \Delta^a_{n,k} := \{G^a(i_1, \ldots, i_{k+1}) : 1 \leq i_1, \ldots, i_{k+1} \leq n\}, \]

and let \(\tilde{\Delta}^a_{n,k}\) be the set of all canonical graphs in \(\Delta^a_{n,k}\). We now divide the proof into cases based on the value of \(k\).
Figure 4.5: This 3-colored 9-path graph is the canonical representative of the graph corresponding to the term in Example 4.7.16. There are nine edges that appear in this graph and 3 colors on those edges, indicating that $k = 9$ and $m = 3$. Since no edges are parallel to another edge of the same color, this path graph corresponds to a term which has zero expectation.

Case where $k$ is a multiple of $m$ If $k$ is a multiple of $m$, then by the block structure of $\mathcal{Y}_n$, it follows that $\mathcal{Y}_n^k$ is a block diagonal matrix. Since the diagonal blocks are the only nonzero blocks in this case, (4.45) simplifies to

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

$$= n^{-k/2} \sum_{1 \leq a \leq m} \sum_{G \in \Delta^a_{n,k}} u_G^* E[X_G] v_G$$

(4.47)

Recall that if $G \in \Delta^a_{n,k}$, then $G$ is an $m$-colored $k$-path graph which starts with color, $a$, i.e., $G = G^a(i_1, \ldots, i_{k+1})$ for some $i_1, \ldots, i_{k+1} \in [n]$. By Lemma 4.7.11 we can reduce the task of counting all terms with nonzero expectation to counting canonical graphs and the cardinality of each equivalence class.

Observe that if $k = m$, then any term in (4.47) will be of the form

$$\pi_{i_1} E[X_{a_1,i_2} \cdots X_{a_{k},i_{k+1}}] v_{i_{k+1}}$$

where each matrix contributes only one entry to the above expression. In this case, all terms are independent and the expectation in (4.47) is zero.

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

$$h(G) =: \max \{ i_t : (t, i_t) \in V_G \}$$

(4.48)
where $\tilde{G}$ is the canonical representative for the graph $G$ and $V_{\tilde{G}}$ is the vertex set for the graph $\tilde{G}$. We call $h(G)$ the maximal height (or sometimes just height) of a graph $G$. Intuitively, $h(G)$ is the number of distinct height coordinates $G$ visits. In terms of the canonical graph, $\tilde{G}$, this is the largest height coordinate visited by an edge in $\tilde{G}$. For each $a$, define

$$(\Delta_{n,k}^a)_1 := \{ G \in \Delta_{n,k}^a : h(G) > k/2 \} \quad (4.49)$$

$$(\Delta_{n,k}^a)_2 := \{ G \in \Delta_{n,k}^a : h(G) = k/2 \} \quad (4.50)$$

$$(\Delta_{n,k}^a)_3 := \{ G \in \Delta_{n,k}^a : h(G) < k/2 \}. \quad (4.51)$$

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

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

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

$$\sum_{G \in (\Delta_{n,k}^1)_1} u_G^c E[x_G] v_G = 0. \quad (4.52)$$

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

Note that if $k$ is odd, then this set will be empty; so assume $k$ is even. Now consider a graph $G \in (\Delta_{n,k}^1)_2$. By Lemma 4.7.11 we can assume that $G$ is canonical. If $G$ has any edges
which are not parallel to any other edges, then $E[x_G] = 0$ and it does not contribute to the expectation. Thus we can consider only graphs in which every edge is parallel to at least one other edge. Since any $G \in (\Delta_{n,k}^1)^2$ must visit exactly $k/2$ distinct height coordinates and since there must be precisely $k$ edges in $G$, a counting argument reveals that every edge in $G$ must be parallel to exactly one other edge in $G$. This gives way to the following lemma.

**Lemma 4.7.17.** Let $k \geq 2$ be any even integer (not necessarily a multiple of $m$). Then there is only one canonical $k$-path graph in $\Delta_{n,k}^1$ for which $h(G) = \frac{k}{2}$ and in which each edge is parallel to exactly one other edge.

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

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

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

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

Finally, if $c$ is even, then the edge from $(k/2, k/2)$ to $(k/2 + 1, 1)$ must have color $m$. Hence the edge from $(k/2 + 1, 1)$ to $(k/2 + 2, 2)$ will have color 1, which is the same color as the edge from $(1, 1)$ to $(2, 2)$. This means that when $k = cm$ and $c$ is even, every edge in $G^1(1, 2, \ldots, k/2, 1, 2, \ldots, k/2, 1)$ will be parallel to exactly one other edge of the same color.
In particular, note that for this graph
\[ |u_G^*E \left[ x_G \right] v_G| \leq \left| \prod_i^{[1]} \left| E \left[ X_{1,(1,2)} \cdots X_{m,(k/2,1)} \right] v_i^{[m]} \right| \right| \]
\[ \leq \left| \prod_i^{[1]} \left| E \left[ X_{1,(1,2)} \right]^2 \cdots E \left[ X_{m,(k/2,1)} \right]^2 v_i^{[m]} \right| \right| \]
\[ \leq \left| \prod_i^{[1]} \left| v_i^{[m]} \right| \right|. \tag{4.53} \]

For ease of notation, let \( \tilde{G} := G^1(1,2,\ldots,k/2,1,2,\ldots,k/2,1) \), and consider \( G^1(i_1,\ldots,i_{k+1}) \in (\Delta_{n,k}^1) \) such that \( G^1(i_1,\ldots,i_{k+1}) \sim \tilde{G} \). Observe that there are \( n \) options for the first coordinate \( i_1 \) of \( G^1(i_1,\ldots,i_{k+1}) \). If we fix the first coordinate, then there are at most \( (n-1)(n-2)\cdots(n-k/2-1) \leq n^{k/2-1} \) graphs with first coordinate \( i_1 \) which are equivalent to \( \tilde{G} \). If we repeated the computation of the expectation of any of these equivalent graphs, we would get a term similar to \( |u_i^{[1]}| |v_i^{[1]}| \) but with different starting and ending coordinates, yielding an upper bound of \( |\overline{u}_i^{[1]}| |v_i^{[1]}| \). Therefore, by the above argument and the Cauchy–Schwarz inequality, we obtain
\[ \left| \sum_{G \in (\Delta_{n,k}^1) \text{2}} u_G^*E \left[ x_G \right] v_G \right| \leq \sum_{1 \leq i_1 \leq n} n^{k/2-1} \left| \overline{u}_i^{[1]} \right| \left| v_i^{[1]} \right| \]
\[ \leq n^{k/2-1} \left\| u \right\| \left\| v \right\| \]
\[ \leq n^{k/2-1}. \tag{4.54} \]

Graphs in \( (\Delta_{n,k}^1) \text{3} \):

Consider an \( m \)-colored \( k \)-path graphs \( G \in (\Delta_{n,k}^1) \text{3} \), and assume that \( G \) is canonical. If \( G \) contains any edges which were not parallel to another edge, then the graph will correspond to a term with expectation zero. So consider a canonical graph \( G \in (\Delta_{n,k}^1) \text{3} \) such that all edges are parallel to at least one other edge. If \( h(G) = 1 \), then \( G = G^1(1,1,\ldots,1) \) and so
\[ |E \left[ x_G \right]| \leq E \left| X_{1,(1,1)} \cdots X_{1,(1,1)} \right| = E \left| X_{1,(1,1)} \right|^k \leq (4L)^k. \]

Note that this is the highest possible moment in a term. Let \( M := (4L)^k \). For any canonical
$m$-colored $k$-path graph $G \in (\tilde{\Delta}_{n,k}^1)_{3}$,

$$E[x_G] \leq M.$$  

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

Based on the above observations, we have

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

$$\leq n^{k-\frac{1}{2}} \sum_{G \in (\Delta_{n,k}^1)_{3}} |u^*_G| E[x_G] |v_G|$$

$$\leq n^{k-\frac{1}{2}} \sum_{G \in (\tilde{\Delta}_{n,k}^1)_{3}} M$$

$$\ll_{L,k} n^{k-\frac{1}{2}}.$$  \hspace{1cm} (4.55)

Combining the bounds:

By (4.52), (4.54), and (4.55) we conclude that

$$\left| \sum_{G \in \Delta_{n,k}^1} u^*_G E[x_G] v_G \right| \leq \sum_{i=1}^2 \left| \sum_{G \in (\tilde{\Delta}_{n,k}^1)_{i}} u^*_G E[x_G] v_G \right|$$

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

While the bounds above were calculated for $a = 1$, the same argument applies for any $a$ by simply permuting the colors. Therefore, in the case where $k$ is a multiple of $m$, from (4.47)
we have

$$\left| E\left[ u^{*}\left( \frac{1}{\sqrt{n}} \mathbb{Y}_{n} \right)^{k} v \right] \right| \leq n^{k/2} \sum_{1 \leq a \leq m} \left| \sum_{G \in \Delta_{a,n,k}^{1}} u_{G}^{*}E[x_{G}]v_{G} \right|$$

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

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

**Case where \( k \) is not a multiple of \( m \)** Now assume that \( k \) is not a multiple of \( m \). If \( k < m \), then each matrix has at most one entry in the product on the right-hand side of (4.46) and all terms will be independent. Hence the expectation will be zero. Therefore, consider the case when \( k > m \). Then there must exist some positive integer \( c \) such that

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

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

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

$$= n^{-k/2} \sum_{1 \leq a \leq m} \sum_{G \in \Delta_{a,n,k}^{a}} u_{G}^{*}E[x_{G}]v_{G}$$

(4.56)

Again, define \( h(G) \) as in (4.48) and define \( \Delta_{a,n,k}^{1} \), \( \Delta_{a,n,k}^{2} \), and \( \Delta_{a,n,k}^{3} \) as in (4.49), (4.50), and (4.51), respectively. Without loss of generality, assume that \( a = 1 \).

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

$$\sum_{G \in \Delta_{a,n,k}^{1}} u_{G}^{*}E[x_{G}]v_{G} = 0.$$  

(4.57)
If $G \in (\Delta_{n,k}^1)_3$ has height less than $k/2$, then we may still bound $\mathbb{E}[x_G] \leq M$. Therefore, we may use the same argument as in the previous case to conclude that

$$\left| \sum_{G \in (\Delta_{n,k}^1)_3} u_G^* \mathbb{E}[x_G] v_G \right| \ll_{L,k} n^{k/2-1}. \quad (4.58)$$

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

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

If $k$ is odd, then $(\Delta_{n,k}^1)_2$ is empty, so assume that $k$ is even. Consider a graph $G \in (\Delta_{n,k}^1)_2$ and by Lemma 4.7.11, we may assume that $G$ is canonical. If $G$ has any edges which are not parallel to any other edge, then $\mathbb{E}[x_G] = 0$, so assume each edge is parallel to at least one other edge. A counting argument reveals that in fact each edge must be parallel to exactly one other edge and by Lemma 4.7.17, we can conclude that in fact $G = G^1(1,2,\ldots,k/2,1,2,\ldots,k/2,1)$.

In order for this graph to correspond to a term with nonzero expectation, the colors on the pairs of parallel edges must match. In order for this to happen, we would need the edge from $(k/2+1,1)$ to $(k/2+2,2)$ to have color 1. This would force the edge from $(k/2,k/2)$ to $(k/2+1,1)$ to have color $m$. Note that if we think about drawing edges sequentially with the time coordinate, then this implies that the $k/2$th edge drawn from $(k/2,k/2)$ to $(k/2+1,1)$ is of color $m$, forcing $k/2$ to be a multiple of $m$. However, this would imply that $k$ is also a multiple of $m$, a contradiction. Hence in this case, if $G \in (\Delta_{n,k}^a)_2$, then $\mathbb{E}[x_G] = 0$. By Lemma 4.7.11 this gives

$$\sum_{G \in (\Delta_{n,k}^a)_2} u_G^* \mathbb{E}[x_G] v_G = 0. \quad (4.59)$$

Combining the bounds:
By (4.57), (4.58), and (4.59) we can see that
\[
\left| \sum_{G \in \Delta^1_{k,n}} u^*_G \mathbb{E}[x_G] v_G \right| \leq 3 \left| \sum_{i=1}^{\delta} \sum_{G \in (\Delta^1_{n,k})_i} u^*_G \mathbb{E}[x_G] v_G \right| \ll_{L,k} n^{k/2 - 1}.
\]

While the bounds above were calculated for \( a = 1 \), the same arguments apply for any \( a \) by simply permuting the colors. Thus, from (4.56) we have
\[
\left| \mathbb{E} \left[ u^* \left( \frac{1}{\sqrt{n}} \mathcal{Y}_n \right)^k v \right] \right| \leq n^{-k/2} \sum_{1 \leq a \leq m} \left| \sum_{G \in \Delta^a_{n,k}} u^*_G \mathbb{E}[x_G] v_G \right| \ll_{L,k} n^{-k/2} \sum_{1 \leq a \leq m} n^{k/2 - 1} \ll_{L,k} \frac{1}{n}
\]
in the case where \( k \) is not a multiple of \( m \).

Combining the cases above completes the proof of Lemma 4.7.2.

**Remark 4.7.18.** Note that if \( m = 1 \), then \( k \) is trivially a multiple of \( m \). Hence, the case where \( \mathcal{Y}_n \) is an \( n \times n \) matrix follows as a special case of the above argument.

### 4.8 Proofs of Results in Section 2.1.1

Before we prove the results from Section 2.1.1, we must prove an isotropic limit law for products of repeated matrices.

**Theorem 4.8.1** (Isotropic limit law for repeated products). Assume \( \xi \) is a complex-valued random variable with mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. For each \( n \geq 1 \), let \( X_n \) be an \( n \times n \) iid random matrix with atom variable \( \xi \). Define \( \mathcal{Y}_n \) as in (4.2) and define \( G_n(z) \) as in (4.3) but with \( X_{n,1} = X_{n,2} = \cdots = X_{n,m} = X_n \). Then, for any fixed \( \delta > 0 \), the following statements hold.
(i) Almost surely, for $n$ sufficiently large, the eigenvalues of $\frac{1}{\sqrt{n}}Y_n$ are contained in the disk 
\[ \{ z \in \mathbb{C} : |z| \leq 1 + \delta \} \]. In particular, this implies that almost surely, for $n$ sufficiently large, 
the matrix $\frac{1}{\sqrt{n}}Y_n - zI$ is invertible for every $z \in \mathbb{C}$ with $|z| > 1 + \delta$.

(ii) There exists a constant $c > 0$ (depending only on $\delta$ and $m$) such that almost surely, for $n$ sufficiently large,
\[
\sup_{z \in \mathbb{C} : |z| > 1 + \delta} \|G_n(z)\| \leq c.
\]

(iii) For each $n \geq 1$, let $u_n, v_n \in \mathbb{C}^{mn}$ be deterministic unit vectors. Then
\[
\sup_{z \in \mathbb{C} : |z| > 1 + \delta} \left| u_n^*G_n(z)v_n + \frac{1}{z}u_n^*v_n \right| \to 0
\]
almost surely as $n \to \infty$.

**Proof.** Fix $\delta > 0$. From [114, Theorem 1.4], the spectral radius of $\frac{1}{\sqrt{n}}X_n$ converges to 1 almost surely as $n \to \infty$. Thus, $n^{-m/2}(X_n)^m$ has spectral radius converging to 1 almost surely as well. It follows that the spectral radius of $n^{-m/2}(Y_n)^m$ converges to 1 almost surely, which in turn implies that the spectral radius of $\frac{1}{\sqrt{n}}Y_n$ converges to 1 almost surely as $n \to \infty$, proving claim (i).

To prove part (ii) we consider two events, both of which hold almost surely. By [114 Theorem 1.4], there exists a constant $K > 0$ such that almost surely, for $n$ sufficiently large, $n^{-1/2}\|X_n\| \leq K$, and hence, on the same event, $n^{-1/2}\|Y_n\| \leq K$. By Lemma D.0.10 this implies that almost surely, for $n$ sufficiently large,
\[
\sup_{z \in \mathbb{C} : |z| \geq K + 1} \left\| \left( \frac{1}{\sqrt{n}}Y_n - zI \right)^{-1} \right\| \leq 1. \tag{4.60}
\]
To deal with $1 + \delta \leq |z| \leq K + 1$, we observe that
\[
\left( \left( \frac{1}{\sqrt{n}}Y_n - zI \right)^{-1} \right)_{[a,b]} = z^{(m-1)-\alpha}n^{-\alpha/2}X_n^\alpha \left( n^{-m/2}X_n^m - z^mI \right)^{-1}. \tag{4.61}
\]


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

\[
\left\| \left( \frac{1}{\sqrt{n}} Y_n - z I \right)^{-1} \right\|_{[a,b]} \leq |z|^{(m-1)-\alpha} \left\| n^{-\alpha/2} X_n^\alpha \right\| \left\| \left( n^{-1/2} X_n - z e^{2\pi k \sqrt{-1}/m} I \right)^{-1} \right\|_m \]

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

We now bound

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

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

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

almost surely, for \( n \) sufficiently large. The bound for

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

follows from Lemma 4.4.2 (taking \( m = 1 \)). Returning to (4.61), we conclude that almost surely, for \( n \) sufficiently large

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

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

For (iii), Theorem 1.4 yields that almost surely, for \( n \) sufficiently large,

\[
\sup_{|z| \geq 5} \frac{1}{\sqrt{n}} \left\| \frac{Y_n}{z} \right\| \leq \frac{9}{10} < 1.
\]  

(4.62)

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

\[
G_n(z) = -\frac{1}{z} \left( I + \sum_{k=1}^{\infty} \left( \frac{1}{\sqrt{n}} \frac{Y_n}{z} \right)^k \right) = -\frac{1}{z} I - \sum_{k=1}^{\infty} \left( \frac{Y_n}{z} \right)^k.
\]
Thus, we have almost surely, for \( n \) sufficiently large,
\[
u^* G_n(z)v = \frac{1}{z} u^* v - \sum_{k=1}^{\infty} \frac{u^*(\frac{Y_n}{\sqrt{n}})^k v}{z^{k+1}}.
\]

We will show that the series on the right-hand side converges to zero almost surely uniformly in the region \( \{ z \in \mathbb{C} : 5 \leq |z| \leq 6 \} \). Indeed, from (4.62), the tail of the series is easily controlled. Thus, it suffices to show that, for each fixed integer \( k \geq 1 \),
\[
\left| u^*(\frac{1}{\sqrt{n}} Y_n)^k v_n \right| = o_k(1).
\]
But this follows from the block structure of \( Y_n \) and [114, Lemma 2.3].

We now extend this convergence to the region \( \{ z \in \mathbb{C} : |z| \geq 1 + \delta \} \). Let \( \varepsilon > 0 \). Let \( M \geq 6 \) be a constant to be chosen later. By Vitali’s convergence theorem (see, for instance [19, Lemma 2.14]), it follows that
\[
\sup_{1 + \delta \leq |z| \leq M} \left| u^*_n G_n(z) v_n + \frac{1}{z} u^*_n v_n \right| \rightarrow 0
\]
almost surely. In particular, almost surely, for \( n \) sufficiently large,
\[
\sup_{1 + \delta \leq |z| \leq M} \left| u^*_n G_n(z) v_n + \frac{1}{z} u^*_n v_n \right| \leq \varepsilon. \tag{4.63}
\]

Choose \( M_1 > 0 \) such that, for all \( |z| \geq M_1 \),
\[
\left\| \left( -\frac{1}{z} \right) u^* v \right\| \leq \frac{1}{z} \| u^* \| \| v \| \leq \frac{\varepsilon}{2}.
\]
Also there exists a constant \( M_2 > 0 \) such that
\[
\sup_{|z| \geq M_2} \| u^*_n G_n(z)v \| \leq \frac{\varepsilon}{2}
\]
almost surely, for \( n \) sufficiently large, by Lemma D.0.10 and [114, Theorem 1.4]. Take \( M := \max\{M_1, M_2, 6\} \). Then almost surely, for \( n \) sufficiently large,
\[
\sup_{|z| \geq M} \left| u^*_n G_n(z)v + \frac{1}{z} u^*_n v \right| \leq \varepsilon. \tag{4.64}
\]
Combining (4.63) and (4.64), we obtain almost surely, for \( n \) sufficiently large,
\[
\sup_{|z| \geq 1 + \delta} \left| u^*_n G_n(z)v + \frac{1}{z} u^*_n v \right| \leq \varepsilon.
\]
Since \( \varepsilon > 0 \) was arbitrary, the proof is complete. \( \square \)
We also need the following lemma in order to prove Theorem 2.1.10.

**Lemma 4.8.2.** Let \( \xi \) be a complex-valued random variable with mean zero, unit variance, finite fourth moment, and independent real and imaginary parts. For each \( n \geq 1 \), let \( X_n \) be an \( n \times n \) iid random matrix with atom variable \( \xi \). Let \( m \) be a positive integer. Then, for any fixed \( \delta > 0 \), the following statements hold.

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

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

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

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

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

almost surely as \( n \to \infty \).

The proof of this lemma is similar to the proof of Corollary 4.2.2; we omit the details.

With these results, we may proceed to the proofs of the results in Section 2.1.1. The proofs of Theorems 2.1.9, 2.1.10, and 2.1.11 follow the proofs of Theorems 2.1.5, 2.1.6, and 2.1.8, respectively, verbatim, except for the following changes:

- Take \( X_{n,1} = \cdots = X_{n,m} = X_n \),

- Replace all occurrences of Theorem 4.2.1 by Theorem 4.8.1,

- Replace all occurrences of Corollary 4.2.2 by Lemma 4.8.2,

- The scaling factor of \( \sigma \) needs to be replaced by \( \sigma^m \).
Chapter 5

Proof of Theorem 2.2.2

This chapter is devoted to the proof of Theorem 2.2.2. For the remainder of Chapter 5 we let $P_n$ denote the product

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

We warn the reader that notation in this chapter differs from that in Chapter 4. See Remarks 5.2.3, 5.2.6, 5.2.10, 5.2.12, 5.3.1, and 5.4.3.

We begin with an overview of the proof.

5.1 Overview of the Proof

The proof of Theorem 2.2.2 is organized as follows. In Section 5.2 some reductions are made to make the proof more manageable. In particular, we show it is sufficient to prove a version of Theorem 2.2.2 in the case where the matrices have been rescaled to have variance one, truncated, and linearized. Section 5.3 begins the proof of the main result. Using Cauchy’s integral formula and the continuous mapping theorem, we show it is sufficient to prove the convergence of a sequence of stochastic processes. Section 5.4 proves the finite dimensional convergence of the sequence of stochastic processes, and Section 5.5 shows that the sequence of stochastic processes is tight, concluding the proof.
5.2 Reductions

In order to prove Theorem 2.2.2, we will make some preliminary reductions. By the Cramer–Wold theorem, the above result will follow from proving the statement for just a single function $f$. Define

$$f := \gamma_1 f_1 + \gamma_2 f_2 + \cdots + \gamma_s f_s$$

for $\gamma_1, \gamma_2, \ldots, \gamma_s \in \mathbb{C}$. We show that it is sufficient to prove that $\text{tr} f(P_n/\sigma_1E_n) - \mathbb{E} [\text{tr} f(P_n/\sigma_1E_n)]$ converges to a mean-zero Gaussian random variable.

**Remark 5.2.1.** Note that while the Cramer–Wold theorem only holds for real random variables, we can apply it to our case since the entries in $X_{n,1}, \ldots, X_{n,m}$ are real. Since the eigenvalues of any real matrix $M$ come in complex conjugate pairs, $\text{tr}(f_i(M))$ separates easily into the real and imaginary terms, each of which is analytic and real. In the case where the entries in each matrix are complex random variables, this approach would no longer hold since the eigenvalues would no longer come in complex conjugate pairs. The result would certainly hold for a single function when the entries in the matrices are complex, but more care would need to be taken to extend the convergence to the case when we consider multiple test functions.

Theorem 2.2.2 will follow from the following result.

**Theorem 5.2.2.** Let $m \geq 1$ be a fixed integer, and assume $\xi_1, \ldots, \xi_m$ are real-valued random variables which satisfy Assumption 2.2.1. For each $n \geq 1$, let $X_{n,1}, \ldots, X_{n,m}$ be independent $n \times n$ iid random matrices with atom variables $\xi_1, \ldots, \xi_m$, respectively. Define the products

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

and

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

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 $c > 0$ such that the event

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

(5.3)
holds with probability $1 - o(1)$ and as $n \to \infty$,

$$
\text{tr} f(P_n/\sigma)1_{E_n} - \mathbb{E}[\text{tr} f(P_n/\sigma)1_{E_n}] \quad (5.4)
$$

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

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

and

$$
\mathbb{E}[F(f)\overline{F(f)}] = \frac{1}{4\pi^2} \oint_C \oint_C f(z)f(w)(zw - 1)^{-2}dzdw
$$

where $C$ is the contour around the boundary of the disk $D_\delta$.

We will now prove Theorem 2.2.2 assuming Theorem 5.2.2.

**Proof of Theorem 2.2.2.** Assume Theorem 5.2.2 and observe that by Lemma F.0.4, $E_n$ holds with probability $1 - o(1)$ as claimed. Since it is assumed that $f_i(z)$ is analytic in some neighborhood containing the contour, there exists an expansion $f_i(z) = \sum_{j=1}^{\infty} (a_i)_j z^j$ for coefficients $(a_i)_j \in \mathbb{C}$.

Then define the function $g_i(z) = \sum_{j=1}^{\infty} (a_i)_j z^j$ for each $1 \leq i \leq s$ and note that $g_i(z)$ is also analytic in the same neighborhood as $f_i(z)$. Thus $f_i(z) + g_i(z)$ and $f_i(z) - g_i(z)$ are analytic in the same neighborhood as well. Now, since $P_n/\sigma$ has real entries, the eigenvalues come in complex conjugate pairs. Therefore for all $1 \leq i \leq s$,

$$
\alpha_i \text{tr}(f_i(P_n/\sigma) + g_i(P_n/\sigma)) = \alpha_i \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} ((a_i)_j + \overline{(a_i)_j}) (\lambda_k(P_n/\sigma))^j \in \mathbb{R}
$$

and

$$
\beta_i \sqrt{-1} \text{tr}(f_i(P_n/\sigma) - g_i(P_n/\sigma)) = \beta_i \sqrt{-1} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} ((a_i)_j - \overline{(a_i)_j}) (\lambda_k(P_n/\sigma))^j \in \mathbb{R}
$$

are real and analytic in a neighborhood containing the contour for any arbitrary $\alpha_i, \beta_i \in \mathbb{R}$. In fact,

$$
\alpha_i \text{tr}(f_i(P_n/\sigma) + g_i(P_n/\sigma)) = 2\alpha_i \text{Re}(\text{tr}(f_i(P_n/\sigma)))
$$

and

$$
\beta_i \sqrt{-1} \text{tr}(f_i(P_n/\sigma) - g_i(P_n/\sigma)) = 2\beta_i \sqrt{-1} \text{Im}(\text{tr}(f_i(P_n/\sigma))).
$$
Therefore, it is sufficient to consider an arbitrary linear combination and in this case we may still use the Cramer–Wold theorem. Indeed, the convergence in distribution of (5.4) to a Gaussian implies that

\[
(\text{tr } f_i(P_n/\sigma)1_{E_n} - \mathbb{E}[\text{tr } f_i(P_n/\sigma)1_{E_n}])_{i=1}^s
\]

converges in distribution to a multivariate Gaussian vector as well. Now we are left with the task of verifying the covariance structure. This will follow from the expansion from the Cramer–Wold theorem. Specifically, observe that if we expand \( f \) in terms of the linear combination, we have

\[
\mathbb{E}\left[(F(f))^2\right] = -\frac{1}{4\pi^2} \oint_C \oint_C f(z)f(w)(zw - 1)^2 \, dz \, dw
\]

\[
= -\frac{1}{4\pi^2} \oint_C \oint_C (\gamma_1f_1(z) + \cdots + \gamma_sf_s(z)) (\gamma_1f_1(w) + \cdots + \gamma_sf_s(w)) (zw - 1)^2 \, dz \, dw
\]

and

\[
\mathbb{E}\left[F(f)F(f)\right] = \frac{1}{4\pi^2} \oint_C \oint_C f(z)f(w)(zw - 1)^{-2} \, dz \, d\bar{w}
\]

\[
= \frac{1}{4\pi^2} \oint_C \oint_C (\gamma_1f_1(z) + \cdots + \gamma_sf_s(z)) (\gamma_1f_1(w) + \cdots + \gamma_sf_s(w)) (zw - 1)^{-2} \, dz \, d\bar{w}
\]

where \( C \) is the contour around the boundary of the disk \( D_\delta \). Note that by selecting \( \gamma_i = 1 \) for some fixed \( 1 \leq i \leq L \), and \( \alpha_j = 0 \) for all \( j \neq i \), this implies

\[
\mathbb{E}\left[(F(f_i))^2\right] = -\frac{1}{4\pi^2} \oint_C \oint_C f_i(z)f_i(w)(zw - 1)^2 \, dz \, dw
\]

and

\[
\mathbb{E}\left[F(f_i)F(f_j)\right] = \frac{1}{4\pi^2} \oint_C \oint_C (f_i(z)f_j(w))(zw - 1)^{-2} \, dz \, d\bar{w}.
\]

Next, if we consider \( \gamma_i = \gamma_j = 1 \) and all other \( \gamma_k = 0 \) for \( k \neq i, j \), then we have the limiting variance of

\[
\text{tr } (f_i(P_n/\sigma) + f_j(P_n/\sigma))
\]
is given by

\[-\frac{1}{4\pi^2} \oint_{C} \oint_{C} (f_i(z) + f_j(z))(f_i(w) + f_j(w))(zw - 1)^2\,dz\,dw\]

\[= -\frac{1}{4\pi^2} \oint_{C} \oint_{C} (f_i(z)f_i(w) + f_i(z)f_j(w) + f_j(z)f_i(w) + f_j(z)f_j(w))(zw - 1)^2\,dz\,dw\]

\[= E[F(f_i)F(f_i)] - \frac{1}{4\pi^2} \oint_{C} \oint_{C} f_j(z)f_i(w)(zw - 1)^2 + E[F(f_j)F(f_j)]\]

\[= E[F(f_i)F(f_i)] + E[F(f_j)F(f_j)] - \frac{1}{2\pi^2} \oint_{C} \oint_{C} f_j(z)f_i(w)(zw - 1)^2\]

and therefore

\[E[F(f_i)F(f_j)] = -\frac{1}{4\pi^2} \oint_{C} \oint_{C} f_j(z)f_i(w)(zw - 1)^2\]

Additionally, if we select \(\gamma_i = \gamma_j = \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}}\) for some fixed \(1 \leq i, j \leq L\) and all other \(\gamma_k = 0\) for \(k \neq i, j\), then we show that the limiting distribution of

\[\text{tr} \left( \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_i(P_n/\sigma) + \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_j(P_n/\sigma) \right)\]

has variance

\[-\frac{1}{4\pi^2} \oint_{C} \oint_{C} \left( \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_i(z) + \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_j(z) \right) \times \left( \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_i(w) + \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_j(w) \right)(zw - 1)^2\,dz\,dw\]

\[= -\frac{1}{4\pi^2} \oint_{C} \oint_{C} \left( \sqrt{-1}f_i(z)f_i(w) + \sqrt{-1}f_i(z)f_j(w) + \sqrt{-1}f_j(z)f_i(w) + \sqrt{-1}f_j(z)f_j(w) \right)(zw - 1)^2\,dz\,dw\]
\[
\frac{1}{4\pi^2} \oint_C \oint_C \left( \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_i(z) + \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_j(z) \right)
\times \left( \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_i(w) + \left( \frac{1}{\sqrt{2}} + \frac{\sqrt{-1}}{\sqrt{2}} \right) f_j(w) \right) (z\bar{w} - 1)^2 dzd\bar{w} = \frac{1}{4\pi} \oint_C \oint_C (f_i(z)f_i(w) + f_i(z)f_j(w) + f_j(z)f_i(w) + f_j(z)f_j(w)) (z\bar{w} - 1)^2 dzd\bar{w}
\]

and therefore
\[
E \left[ F(f_i)\overline{F(f_i)} \right] = \frac{1}{4\pi^2} \oint_C \oint_C f_i(z)f_j(w)(z\bar{w} - 1)^2 dzd\bar{w}.
\]

Observe that by a simple rescaling, it is sufficient to prove Theorem 5.2.2 when \(\sigma_i = 1\) for \(1 \leq i \leq k\).

### 5.2.1 Truncation of the Product

In order to have greater control over the matrices, we truncate each one and consider the truncated product. Observe that since \(\xi_k\) is assumed to have finite \(4 + \tau\) finite moments for \(1 \leq k \leq m\), there exists an \(\varepsilon > 0\) such that

\[
\lim_{n \to \infty} n^{4\varepsilon} E \left[ |\xi_k|^4 1_{\{|\xi_k| > n^{1/2 - \varepsilon}\}} \right] = 0. \tag{5.5}
\]

Indeed, by the dominated convergence theorem, we can write

\[
E \left[ |\xi_k|^4 1_{\{|\xi_k| > n^{1/2 - \varepsilon}\}} \right] \leq n^{4\varepsilon} E \left[ \frac{|\xi_k|^{4+\tau}}{n^{(1/2 - \varepsilon)\tau}} 1_{\{|\xi_k| > n^{1/2 - \varepsilon}\}} \right]
= n^{\varepsilon(4+\tau) - \tau/2} E \left[ |\xi_k|^{4+\tau} 1_{\{|\xi_k| > n^{1/2 - \varepsilon}\}} \right]
= o(1)
\]
for \( \varepsilon \) sufficiently small.

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

\[
\tilde{\xi} := \xi 1_{\{|\xi| \leq n^{1/2-\varepsilon}\}} - \mathbb{E}\left[ \xi 1_{\{|\xi| \leq n^{1/2-\varepsilon}\}} \right] \quad \text{and} \quad \hat{\xi} := \frac{\tilde{\xi}}{\sqrt{\text{Var}(\tilde{\xi})}}.
\]  

(5.6)

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

**Remark 5.2.3.** We warn the reader that the truncation in Chapter 5 differs from that in Chapter 4, although the notation is the same. Indeed, in this chapter, the truncation (5.6) grows with \( n \) and in Chapter 4 the truncation (4.17) is at a constant level.

**Lemma 5.2.4.** Let \( \xi \) be a real-valued random variable which satisfies Assumption 2.2.1 and define \( \tilde{\xi} \) and \( \hat{\xi} \) as in (5.6). Then the following statements hold:

(i) \(|1 - \text{Var}(\tilde{\xi})| = o(n^{-1-2\varepsilon})\)

(ii) There exists an \( N_0 > 0 \) such that for any \( n > N_0 \), \( \tilde{\xi} \) has zero mean and unit variance, and almost surely

\[
|\hat{\xi}| \leq 4n^{1/2-\varepsilon}.
\]

(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 E.

**Remark 5.2.5.** In the case where \( \xi \) is complex, this truncation will need to be adjusted. In particular, the random variables will need to be truncated in a manner which preserves independence between
the real and imaginary parts of $\xi$. Similar bounds hold in the case when the complex random variables are truncated properly. See Lemma 4.3.1 from Outliers for an example.

Let $X$ be an $n \times n$ random matrix filled with iid copies of a random variable $\xi$ which satisfies Assumption 2.2.1. For each $1 \leq i, j \leq n$, define matrices $\tilde{X}$ and $\hat{X}$ to be the $n \times n$ matrices with entries defined by

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

(5.7)

$$
\hat{X}_{(i,j)} := \frac{\tilde{X}_{(i,j)}}{\sqrt{\text{Var}(\tilde{X}_{(i,j)})}}.
$$

(5.8)

**Remark 5.2.6.** Note that the definitions for $\tilde{X}$ and $\hat{X}$ differ here from Chapter 4. The definitions (5.7) and (5.8) are consistent with the truncation defined in (5.6).

**Lemma 5.2.7.** Let $X_n$ be an $n \times n$ iid random matrix with atom variable $\xi$ which satisfies Assumption 2.2.1 and let $\hat{X}_n$ be the truncated matrix as defined in (5.8). Then

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

and

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

**Proof.** By Markov’s inequality,

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

$$
\leq n^{2\varepsilon} \mathbb{E} \left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2
$$

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,

$$
\mathbb{E} \left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2 \leq \mathbb{E} \left[ \left\| \frac{1}{\sqrt{n}} X_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2 + \left\| \frac{1}{\sqrt{n}} \hat{X}_n - \frac{1}{\sqrt{n}} \hat{X}_n \right\|_2^2 \right]
$$
and we may deal with the two terms on the right hand side of the above expression separately. First, since $\left| E \left[ \mathbf{1}_{\{\xi > n^{1/2-\epsilon}\}} \right] \right| = \left| E[\mathbf{1}_{\{\xi \leq n^{1/2-\epsilon}\}}] \right|$ and by (5.5), we have

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

Next, we consider

$$
E \left[ \frac{1}{\sqrt{n}} \hat{X}_n - \frac{1}{\sqrt{n}} \hat{X}_n \right]_2.
$$

Observe that by Lemma 5.2.4 one has

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

which concludes the proof.

For each $1 \leq k \leq m$, define truncated matrices $\tilde{X}_{n,k}$ and $\hat{X}_{n,k}$ as in (5.7) and (5.8) respectively. Next define the truncated products

$$
\tilde{P}_n = n^{-m/2} \tilde{X}_{n,1} \cdots \tilde{X}_{n,m} \quad \text{and} \quad \hat{P}_n = n^{-m/2} \hat{X}_{n,1} \cdots \hat{X}_{n,m}
$$

and observe the following lemma.
Lemma 5.2.8. Let $X_{n,i}$ be as defined in Theorem 2.2.2, $P_n$ as defined in (5.2), $\tilde{X}_{n,i}$ and $\hat{X}_{n,i}$ as defined in (5.6), and define the truncated products $\tilde{P}_n$ and $\hat{P}_n$ as in (5.9). Then

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

and

$$
\mathbb{P} \left( \left\| P_n - \hat{P}_n \right\| > n^{-\varepsilon} \right) = o(1).
$$

Proof. By Markov’s inequality,

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

so it is sufficient to prove the first equality. To this end, note that by the triangle inequality, independence, and Lemma D.0.9, we have

$$
E \left\| P_n - \hat{P}_n \right\|_2^2 = E \left\| n^{-m/2} X_{n,1} \cdots X_{n,m} - n^{-m/2} \tilde{X}_{n,1} \cdots \tilde{X}_{n,m} \right\|_2^2
$$

$$
\ll n^{-m} \left( E \left\| X_{n,1} - \tilde{X}_{n,1} \right\|_2^2 E \left\| X_{n,2} \right\|^2 \cdots E \left\| X_{n,m-1} \right\|^2 E \left\| X_{n,m} \right\| + \cdots + E \left\| \tilde{X}_{n,1} \right\|^2 E \left\| \tilde{X}_{n,2} \right\|^2 \cdots E \left\| \tilde{X}_{n,m-1} \right\|^2 E \left\| X_{n,m} - \tilde{X}_{n,m} \right\|_2^2 \right).
$$

By Lemmas D.0.11 and D.0.12

$$
E \left\| \tilde{X}_{n,k} \right\|^2 = O(n) \quad \text{and} \quad E \left\| X_{n,k} \right\|^2 = O(n)
$$

for all $1 \leq k \leq m$. Therefore, by this observation and Lemma 5.2.7

$$
E \left\| P_n - \hat{P}_n \right\|_2^2
$$

$$
\ll n^{-1} E \left\| X_{n,1} - \tilde{X}_{n,1} \right\|_2^2 + \cdots + n^{-1} E \left\| X_{n,m} - \tilde{X}_{n,m} \right\|_2^2
$$

$$
= o(n^{-2\varepsilon}).
$$
This lemma lets us work with products of matrices with truncated entries, provided we can control the smallest singular value. Define

\[ \hat{E}_n := \left\{ \inf_{|z| > 1 + \delta/2} s_n(\hat{P}_n - zI) \geq c \right\} \quad (5.10) \]

where \( c > 0 \) is a constant. By Lemma F.0.3, \( \hat{E}_n \) holds with overwhelming probability. The next theorem shows it is sufficient to consider matrices with truncated entries.

**Theorem 5.2.9.** Let \( X_{n,i} \) be as defined in Theorem 2.2.2, \( \hat{X}_{n,i} \) as defined in (5.6), \( \hat{P}_n \) as in (5.9), and \( \hat{E}_n \) as defined in (5.10). Then \( \hat{E}_n \) holds with overwhelming probability and

\[ \text{tr} f(\hat{P}_n) - \text{E}[\text{tr} f(\hat{P}_n)] \]

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

\[ \text{E} \left[ (F(f))^2 \right] = -\frac{1}{4\pi^2} \oint_C \oint_C f(z)f(w)(zw - 1)^2 \, dz \, dw \]

and

\[ \text{E} \left[ F(f)\overline{F(f)} \right] = \frac{1}{4\pi^2} \oint_C \oint_C f(z)f(w)(z\overline{w} - 1)^{-2} \, dz \, d\overline{w}. \]

We now prove Theorem 5.2.2 assuming Theorem 5.2.9.

**Proof of Theorem 5.2.2.** Begin by observing that by Lemmas F.0.3 and F.0.4, \( \hat{E}_n \) holds with overwhelming probability and \( E_n \) holds with probability \( 1 - o(1) \). Suppose that

\[ \text{tr} f(\hat{P}_n) - \text{E}[\text{tr} f(\hat{P}_n)] \]

converges in distribution to a mean-zero Gaussian random variable with variance

\[ \frac{1}{4\pi^2} \oint_C \oint_C f(z)f(w)(z\overline{w} - 1)^{-2} \, dz \, d\overline{w}. \]

By Lemma D.0.7 it is sufficient to show that

\[ \text{E} \left| \text{tr} f(\hat{P}_n) - \text{E}[\text{tr} f(\hat{P}_n)] - (\text{tr} f(P_n) - \text{E}[\text{tr} f(P_n)]) \right| \]

\[ \leq 2\text{E} \left| \text{tr} f(\hat{P}_n) - \text{tr} f(P_n) \right| \]

\[ = o(1). \]
Define the $\infty$-norm

$$\|f\|_\infty := \inf\{C \geq 0 : |f(z)| \leq C \text{ for almost every } z \in \mathbb{C}\}$$

and observe that by Cauchy’s integral formula,

$$\mathbb{E}\left|\operatorname{tr} f(\hat{P}_n)\mathbf{1}_{\hat{E}_n} - \operatorname{tr} f(P_n)\mathbf{1}_{E_n}\right|
= \mathbb{E}\left|\frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) \left(\operatorname{tr}(\hat{P}_n - zI)^{-1}\mathbf{1}_{\hat{E}_n} - \operatorname{tr}(P_n - zI)^{-1}\mathbf{1}_{E_n}\right) dz\right|
\leq \|f\|_\infty \cdot 2\pi (1 + \delta) \mathbb{E}\left[\sup_{z \in \mathcal{C}} \left|\operatorname{tr}(\hat{P}_n - zI)^{-1}\mathbf{1}_{\hat{E}_n} - \operatorname{tr}(P_n - zI)^{-1}\mathbf{1}_{E_n}\right|\right]
\leq \|f\|_\infty (1 + \delta) \sup_{z \in \mathcal{C}} \mathbb{E}\left|\operatorname{tr}(\hat{P}_n - zI)^{-1}\mathbf{1}_{\hat{E}_n} - \operatorname{tr}(P_n - zI)^{-1}\mathbf{1}_{E_n}\right|.
$$

Since $f$ is assumed to be analytic on the disk and bounded otherwise, $\|f\|_\infty < \infty$ and therefore it is sufficient to show that

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

To this end, observe that

$$\left|\operatorname{tr}(\hat{P}_n - zI)^{-1}\mathbf{1}_{\hat{E}_n} - \operatorname{tr}(P_n - zI)^{-1}\mathbf{1}_{E_n}\right|
= \left|\operatorname{tr}(\hat{P}_n - zI)^{-1}\mathbf{1}_{E_n \cap \hat{E}_n} + \operatorname{tr}(\hat{P}_n - zI)^{-1}\mathbf{1}_{E_n \cap \hat{E}_n}
- \left(\operatorname{tr}(P_n - zI)^{-1}\mathbf{1}_{E_n \cap \hat{E}_n} + \operatorname{tr}(P_n - zI)^{-1}\mathbf{1}_{E_n \cap \hat{E}_n}\right)\right|.$$
and therefore

\[
E \left| \operatorname{tr}(\hat{P}_n - zI)^{-1}1_{\hat{E}_n} - \operatorname{tr}(P_n - zI)^{-1}1_{E_n} \right|
\]

\[
= E \left| \operatorname{tr}(\hat{P}_n - zI)^{-1}1_{\hat{E}_n \cap E_n} + \operatorname{tr}(\hat{P}_n - zI)^{-1}1_{E_n \cap \hat{E}_n}
- \left( \operatorname{tr}(P_n - zI)^{-1}1_{E_n \cap \hat{E}_n} + \operatorname{tr}(P_n - zI)^{-1}1_{E_n \cap E_n} \right) \right|
\]

\[
\leq E \left| (\operatorname{tr}(\hat{P}_n - zI)^{-1} - \operatorname{tr}(P_n - zI)^{-1})1_{E_n \cap \hat{E}_n} \right|
+ E \left| \operatorname{tr}(\hat{P}_n - zI)^{-1}1_{E_n \cap E_n} \right|
+ E \left| \operatorname{tr}(P_n - zI)^{-1}1_{E_n \cap \hat{E}_n} \right|
\]

\[
\ll E \left| (\operatorname{tr}(\hat{P}_n - zI)^{-1} - \operatorname{tr}(P_n - zI)^{-1})1_{E_n \cap \hat{E}_n} \right|
+ \mathbb{P}(E_n^c) + n\mathbb{P}(E_n \cap \hat{E}_n^c)
\]

\[
\ll E \left| (\operatorname{tr}(\hat{P}_n - zI)^{-1} - \operatorname{tr}(P_n - zI)^{-1})1_{E_n \cap \hat{E}_n} \right| + o(1).
\]

Therefore, we have reduced the argument once again to showing that

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

By the linearity of the trace, the resolvent identity, Lemma \textbf{D.0.9} and Lemma \textbf{5.2.8} we have

\[
E \left| (\operatorname{tr}(\hat{P}_n - zI)^{-1} - \operatorname{tr}(P_n - zI)^{-1})1_{E_n \cap \hat{E}_n} \right|
\]

\[
= E \left| \operatorname{tr} \left( (\hat{P}_n - zI)^{-1}(P_n - \hat{P}_n)(P_n - zI)^{-1} \right) 1_{E_n \cap \hat{E}_n} \right|
\]

\[
\leq E \left[ \left\| (\hat{P}_n - zI)^{-1}1_{E_n \cap \hat{E}_n} \right\| \left\| (P_n - \hat{P}_n)(P_n - zI)^{-1}1_{E_n \cap \hat{E}_n} \right\|_2 \right]
\]

\[
\ll E \left[ \left\| (P_n - zI)^{-1}1_{E_n \cap \hat{E}_n} \right\| \left\| P_n - \hat{P}_n \right\|_2 \right]
\]

\[
\ll E \left\| P_n - \hat{P}_n \right\|_2
\]

\[
= o(1)
\]

since the spectral norms of the resolvents are bounded by a constant on their respective events almost surely. Since all of this was uniform in \( z \), the proof is complete. \( \square \)
5.2.2 Truncated Linearization

We now wish to linearize this product so that we can work with an $mn \times mn$ block matrix instead. Define the $mn \times mn$ matrix

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

(5.11)

Recall that by Proposition 3.1.1, $\mathcal{Y}_n$ has the same eigenvalues as $\hat{P}_n = n^{-m/2} \hat{X}_{n,1} \cdots \hat{X}_{n,m}$, each with multiplicity $m$. Also recall that if $\lambda$ is an eigenvalue of $M$, then $\lambda^m$ is an eigenvalue of $M^m$. Therefore, the eigenvalues of the product $\hat{P}_n$ are completely determined by the eigenvalues of the linearized matrix $\mathcal{Y}_n$.

Remark 5.2.10. Note that in this chapter, the normalization of $n^{-1/2}$ is built into the notation of $\mathcal{Y}_n$, and $\mathcal{Y}_n$ has truncated entries. This definition is slightly different than the definition (4.2) used in Chapter 4, in which the scaling of $n^{-1/2}$ was not built into the block matrix notation and the entries of $\mathcal{Y}_n$ are not truncated.

Theorem 5.2.11. Let $\mathcal{Y}_n$ be the linearized matrix defined in (5.11) where $X_{n,i}$ are under the assumptions of Theorem 2.2.2 and the entries of $\hat{X}_{n,i}$ are truncated as defined in (5.8). Define the event

$$
\Omega_n := \left\{ \inf_{|z|>1+\delta/2} \text{tr} \mathcal{Y}_n - zI \geq c \right\}
$$

(5.12)

for some $c > 0$. Then $\Omega_n$ holds with overwhelming probability and for any function $g$ which is analytic in a neighborhood of the disk $\{z \in \mathbb{C} : |z| \leq (1+\delta)^{1/m}\}$ and bounded otherwise, the random variable

$$
\text{tr} g(\mathcal{Y}_n) 1_{\Omega_n} - \mathbb{E} \left[ \text{tr} g(\mathcal{Y}_n) 1_{\Omega_n} \right]
$$
converges to a mean zero Gaussian random variable $F(g)$ with covariance structure

$$\mathbb{E}[(F(g))^2] = -\frac{1}{4\pi^2} \oint_C \oint_C g(z)g(w) \frac{m^2(zw)^{m-1}}{(zw - 1)^2} \, dz \, dw$$

and

$$\mathbb{E}[F(g)F(g)] = \frac{1}{4\pi^2} \oint_C \oint_C g(z)\bar{g}(w) \frac{m^2(z\bar{w})^{m-1}}{(z\bar{w} - 1)^2} \, dz \, d\bar{w}.$$ 

Remark 5.2.12. The event $\Omega_n$ defined in (5.12) differs from the event in Chapter 4 defined in equation (4.25).

We prove Theorem 5.2.9 assuming Theorem 5.2.11.

Proof of Theorem 5.2.9. Let $f$ be any function which is analytic on the disk $\{ z \in \mathbb{C} : |z| \leq 1 + \delta \}$ and define the function

$$g(z) := \frac{1}{m} f(z^m).$$

Note that this function $g$ is analytic on the disk $\{ z \in \mathbb{C} : |z| \leq (1 + \delta)^{1/m} \}$ and bounded otherwise and

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

By assumption,

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

converges to a mean-zero Gaussian with covariance structure

$$\mathbb{E}[(F(g))^2] = -\frac{1}{4\pi^2} \oint_C \oint_C g(z)g(w) \frac{m^2(zw)^{m-1}}{(zw - 1)^2} \, dz \, dw$$

and

$$\mathbb{E}[F(g)F(g)] = \frac{1}{4\pi^2} \oint_C \oint_C g(z)\bar{g}(w) \frac{m^2(z\bar{w})^{m-1}}{(z\bar{w} - 1)^2} \, dz \, d\bar{w},$$

so it is sufficient to show that

$$\mathbb{E}\left[ \text{tr} g(\mathcal{Y}_n) \mathbf{1}_{\Omega_n} - \mathbb{E}[\text{tr} g(\mathcal{Y}_n) \mathbf{1}_{\Omega_n}] - \left( \text{tr} f(\hat{P}_n) \mathbf{1}_{\hat{E}_n} - \mathbb{E}[\text{tr} f(\hat{P}_n) \mathbf{1}_{\hat{E}_n}] \right) \right] = o(1)$$
by Lemma D.0.7. To this end, observe that

\[ \mathbb{E} \left[ \text{tr} g(Y_n)1_{\Omega_n} - \mathbb{E}[\text{tr} g(Y_n)1_{\Omega_n}] - \left( \text{tr} f(\hat{P}_n)1_{\hat{E}_n} - \mathbb{E}\left[\text{tr} f(\hat{P}_n)1_{\hat{E}_n}\right] \right) \right] \]

\[ = \mathbb{E} \left[ \text{tr} g(Y_n)1_{\Omega_n} - \text{tr} f(\hat{P}_n)1_{\hat{E}_n} - \mathbb{E}\left[\text{tr} g(Y_n)1_{\Omega_n} - \text{tr} f(\hat{P}_n)1_{\hat{E}_n}\right] \right] \]

\[ \leq 2\mathbb{E} \left( \text{tr} g(Y_n) - \text{tr} f(\hat{P}_n) \right) 1_{\hat{E}_n \cap \Omega_n} \]

\[ + 2\mathbb{E} \left| \text{tr} g(Y_n)1_{\Omega_n \cap \hat{E}_n} \right| + 2\mathbb{E} \left| \text{tr} f(\hat{P}_n)1_{\hat{E}_n \cap \Omega_n} \right| \]

\[ \ll n\mathbb{P}(\Omega_n \cap \hat{E}_n^c) + n\mathbb{P}(\hat{E}_n^c \cap \Omega_n^c) \]

\[ = o(1) \]

since \( \Omega_n \) and \( \hat{E}_n \) both hold with overwhelming probability by lemmas F.0.2 and F.0.3 respectively.

To see that the variance follows as claimed, observe that by letting \( z = re^{\theta_1 i} \) and \( w = re^{\theta_2 i} \), we have

\[ \frac{1}{4\pi^2} \oint_C \oint_C \frac{1}{m^2} f(z^m)f(w^m) \frac{m^2(z\bar{w})^m-1}{((z\bar{w})^m-1)^2} dzd\bar{w} \]

\[ = \frac{1}{4\pi^2} \oint_C \oint_C f(r^m e^{m\theta_1 i})f(r^m e^{m\theta_2 i}) \frac{r^{2m}e^{m\theta_1 i}e^{-m\theta_2 i}}{(r^{2m}e^{m\theta_1 i}e^{-m\theta_2 i} - 1)^2} d\theta_1 d\theta_2. \]

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,

\[ \frac{1}{4\pi^2} \oint_C \oint_C f(r^m e^{m\theta_1 i})f(r^m e^{m\theta_2 i}) \frac{r^{2m}e^{m\theta_1 i}e^{-m\theta_2 i}}{(r^{2m}e^{m\theta_1 i}e^{-m\theta_2 i} - 1)^2} d\theta_1 d\theta_2 \]

\[ = \frac{m^2}{4\pi^2} \oint_C \oint_C f(r^m e^{\tau_1 i})f(r^m e^{\tau_2 i}) \frac{r^{2m}e^{\tau_1 i}e^{-\tau_2 i}}{(r^{2m}e^{\tau_1 i}e^{-\tau_2 i} - 1)^2} \frac{1}{m^2} d\tau_1 d\tau_2 \]

\[ = \frac{1}{4\pi^2} \oint_C \oint_C f(r^m e^{\tau_1 i})f(r^m e^{\tau_2 i}) \frac{r^{2m}e^{\tau_1 i}e^{-\tau_2 i}}{(r^{2m}e^{\tau_1 i}e^{-\tau_2 i} - 1)^2} d\tau_1 d\tau_2 \]
and finally, by letting \( z' = r^m e^{\tau_1 i} \) and \( w' = r^m e^{\tau_2 i} \), we have
\[
\frac{1}{4\pi^2} \oint_C \oint_C f(r^m e^{\tau_1 i}) f(r^m e^{\tau_2 i}) \frac{r^{2m} e^{\tau_1 i} e^{-\tau_2 i}}{(r^{2m} e^{\tau_1 i} e^{-\tau_2 i} - 1)^2} d\tau_1 d\tau_2 \\
= \frac{1}{4\pi^2} \oint_C \oint_C f(z') f(w') \frac{z' \bar{w}'}{(z' \bar{w'} - 1)^2} \frac{1}{z' \bar{w'}} dz' d\bar{w'} \\
= \frac{1}{4\pi^2} \oint_C \oint_C f(z') f(w')(z' \bar{w'} - 1)^{-2} dz' d\bar{w'}
\]
as claimed. The same calculation shows that
\[
-\frac{1}{4\pi^2} \oint_C \oint_C \frac{1}{m^2} f(z^m) f(w^m) m^2 (zw)^{m-1} \frac{dzdw}{((zw)^m - 1)^2} \\
= -\frac{1}{4\pi^2} \oint_C \oint_C f(z') f(w')(z' \bar{w'} - 1)^{-2} dz' dw',
\]
which concludes the variance calculation.

\[\square\]

### 5.3 Proof of Theorem 5.2.11

At this point, note that Theorem 2.2.2 will follow from proving Theorem 5.2.11, where we are now working with a linearized truncated \( mn \times mn \) matrix on the event \( \Omega_n \) where the least singular value is bounded away from zero.

Define the resolvent
\[
\mathcal{G}_n(z) := (\mathcal{Y}_n - zI)^{-1}
\]
and define
\[
\Xi_n(z) := \text{tr} \mathcal{G}_n(z) 1_{\Omega_n} - \mathbb{E} [\text{tr} \mathcal{G}_n(z) 1_{\Omega_n}].
\]

**Remark 5.3.1.** Note that the resolvent defined in (5.13) differs slightly from the resolvent defined in Chapter 4 in (4.3) due to the difference in scaling on \( \mathcal{Y}_n \) and the fact that in Chapter 5, the entries of \( \mathcal{Y}_n \) are assumed to be truncated.

By Cauchy’s integral formula, we have that for any contour \( \mathcal{C} \) inside the region of analyticity
containing the disk $D_{\delta}$,

$$
\text{tr} f(Y_n)1_{\Omega_n} - \mathbb{E}[\text{tr} f(Y_n)1_{\Omega_n}]
= \sum_{i=1}^{mn} f(\lambda_i(Y_n))1_{\Omega_n} - \mathbb{E}\left[\sum_{i=1}^{mn} f(\lambda_i(Y_n))1_{\Omega_n}\right]
= \sum_{i=1}^{mn} -\frac{1}{2\pi i} \oint_C f(z)1_{\Omega_n} dz - \mathbb{E}\left[\sum_{i=1}^{mn} -\frac{1}{2\pi i} \oint_C \lambda_i(Y_n) - z dz\right]
= -\frac{1}{2\pi i} \oint_C f(z) (\text{tr} G_n(z)1_{\Omega_n} - \mathbb{E}[\text{tr} G_n(z)1_{\Omega_n}]) dz
= -\frac{1}{2\pi i} \oint_C f(z)\Xi_n(z) dz.
$$

It is useful to note that since $\Xi_n(z)$ depends on the eigenvalues of $Y_n$ it is random variable. Since it is also a function of $z$ on the contour $C$, $\{\Xi_n(z)\}_{z \in C}$ is a sequence of random functions. From here on, we assume that $C$ is the contour around the boundary of the disk $D_{\delta}$.

Recall the continuous mapping theorem:

**Theorem 5.3.2** (Theorem 25.7 from [33]). Let $\{X_n\}$ and $X$ be random elements defined on a metric space $S$. Suppose that $S'$ is another metric space and $g : S \to S'$ is a function such that the set of discontinuity points $D_g$ of $g$ has $\mathbb{P}(X \in D_g) = 0$. If $X_n$ converges to $X_n$ in distribution to $X$ as $n \to \infty$, then $g(X_n)$ converges in distribution to $g(X)$ as $n \to \infty$ as well.

**Remark 5.3.3.** This theorem holds if the convergence is almost surely or in probability as well.

Let $\{\Xi(z)\}_{z \in C}$ denote a mean-zero Gaussian process with covariance structure defined by

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

(5.15)

Then observe that $\Xi_n(z)$ and $\Xi(z)$ are random elements of the space of continuous functions on the contour $C$, which is a metric space with respect to the supremum norm. Since the map

$$
\Xi_n(z) \mapsto \frac{1}{2\pi i} \oint_C f(z)\Xi_n(z) dz
$$

(5.16)

is continuous in this metric space, by Theorem 5.3.2 it is sufficient to prove that $\{\Xi_n(z)\}_{z \in C}$ converges in distribution to $\{\Xi(z)\}_{z \in C}$. Indeed, if the stochastic process $\{\Xi_n(z)\}_{z \in C}$ converges in
distribution to the stochastic process \( \{ \Xi(z) \}_{z \in \mathcal{C}} \), then \( \frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) \Xi_n(z) dz \) converges in distribution to \( \frac{1}{2\pi} \oint_{\mathcal{C}} f(z) \Xi(z) dz \) as desired.

It remains to be shown that if \( \{ \Xi(z) \}_{z \in \mathcal{C}} \) is a Gaussian random process with covariance structure (5.15), then the right hand side of (5.16) is Gaussian with appropriate variance.

**Lemma 5.3.4.** Let \( \{ \Xi(z) \}_{z \in \mathcal{C}} \) be a mean-zero Gaussian process with variance

\[
E \left[ \Xi(z) \Xi(w) \right] = \frac{m^2 (zw)^{m-1}}{(zw)^{m-1} - 1^2}.
\]

Then

\[
-\frac{1}{2\pi} \oint_{\mathcal{C}} f(z) \Xi(z) dz
\]

is a mean-zero Gaussian random variable with covariance structure

\[
E \left[ \left( \frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) \Xi(z) dz \right)^2 \right] = -\frac{1}{4\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} f(z) f(w) \frac{m^2 (zw)^{m-1}}{(zw)^{m-1} - 1^2} dz dw
\]

and

\[
E \left[ \frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) \Xi(z) dz \cdot \frac{1}{2\pi i} \oint_{\mathcal{C}} f(z) \Xi(z) dz \right] = \frac{1}{4\pi^2} \oint_{\mathcal{C}} \oint_{\mathcal{C}} f(z) f(w) \frac{m^2 (zw)^{m-1} \left( |z|^2 \right)^{m-1}}{(zw)^{m-1} - 1^2} dz dw.
\]

**Proof.** Begin by observing that since \( \Xi_n(z) \) is a Gaussian random variable for any \( z \in \mathcal{C} \), \( \Xi_n(z) \) has finite moments of all orders. Ergo, \( E |f(z) \Xi_n(z)|^p \) is finite for all \( p \geq 0 \) and all \( z \in \mathcal{C} \). Since the integral of a Gaussian random variable is Gaussian, we need only to verify the expectation and variance of (5.17). Observe that by Fubini’s theorem

\[
E \left[ \oint_{\mathcal{C}} f(z) \Xi(z) dz \right] = \oint_{\mathcal{C}} E \left[ f(z) \Xi(z) \right] dz = \oint_{\mathcal{C}} f(z) E \left[ \Xi(z) \right] dz = 0.
\]

Note that Fubini’s theorem is justified because the iterated integral is uniformly bounded in \( z \in \mathcal{C} \). Indeed, by the Cauchy–Schwarz inequality,

\[
\oint_{\mathcal{C}} E |f(z) \Xi(z)| dz \leq \left( \oint_{\mathcal{C}} \|f(z)\|_\infty E |\Xi(z)|^2 dz \right)^{1/2} \leq \left( \oint_{\mathcal{C}} \frac{m^2 |z|^{2(m-1)}}{(|z|^{2m} - 1)^2} dz \right)^{1/2} < \infty
\]
Similarly, since
\[ E \left[ \Xi(z)\Xi(w) \right] = \frac{m^2(z\bar{w})^{m-1}}{(zw)^m - 1} , \]
the Cauchy–Schwarz inequality justifies that the integral of \( E |f(z)f(w)\Xi(z)\Xi(w)| \) is finite, so again by standard methods using Fubini’s theorem, we have
\[
E \left[ \frac{1}{2\pi i} \oint_{C} f(z)\Xi(z)dz \cdot \frac{1}{2\pi i} \oint_{C} f(w)\Xi(w)dw \right] = -\frac{1}{4\pi^2} \oint_{C} \oint_{C} f(z)f(w)\frac{m^2(z\bar{w})^{m-1}}{(zw)^m - 1}dzdw
\]
and
\[
E \left[ \frac{1}{2\pi i} \oint_{C} f(z)\Xi(z)dz \cdot \frac{1}{2\pi i} \oint_{C} f(w)\Xi(w)dw \right] = \frac{1}{4\pi^2} \oint_{C} \oint_{C} f(z)f(w)\frac{m^2(z\bar{w})^{m-1}}{(zw)^m - 1}dzd\bar{w}
\]
as desired. \( \square \)

**Remark 5.3.5.** Note that since this is a complex random variable, we need to know both of the above quantities in order to fully characterize the distribution. Indeed, if \( X \) and \( Y \) are real random variables, and if we define \( Z = X + Yi \), then \( E[Z\bar{Z}] = E[X^2] + E[Y^2] \) and \( E[Z^2] = E[X^2] - E[Y^2] + 2iE[XY] \).

Theorem 5.2.11 and thus Theorem 2.2.2 will follow from proving the following theorem.

**Theorem 5.3.6.** Let \( \{\Xi_n(z)\}_{z \in C} \) be the sequence of stochastic processes defined in (5.14) for \( z \) on the contour \( C \) around the boundary of the disk \( D_\delta \). Then \( \{\Xi_n(z)\}_{z \in C} \) converges in distribution to a mean-zero Gaussian process \( \{\Xi(z)\}_{z \in C} \) with variance
\[
E \left[ \Xi(z)\Xi(w) \right] = \frac{m^2(z\bar{w})^{m-1}}{(zw)^m - 1} .
\]

In order to prove this, we will use the following characterization of convergence, which is a result of Theorems 7.5 and 12.3 from [32].

**Theorem 5.3.7.** Suppose that \( \{X(z)\}_{z \in C}, \{X_n(z)\}_{z \in C} \) for \( n \geq 1 \) are stochastic processes on the contour \( C = \{z \in \mathbb{C} : |z| = 1 + \delta\} \). Suppose \( X(z) \), \( (X_k(z))_{k=1}^\infty \) satisfy
\[
(X_n(z_1), X_n(z_2), ..., X_n(z_k)) \to (X(z_1), X(z_2), ..., X(z_k)) \quad (5.18)
\]
in distribution as \( n \to \infty \) for any fixed positive integer \( k \) and any \( z_1, \ldots, z_k \in \mathcal{C} \), and suppose that there exists a constant \( c > 0 \) such that

\[
\mathbb{E} \left\| \frac{X_n(z) - X_n(w)}{z-w} \right\|^2 \leq c \quad (5.19)
\]

holds for all \( z, w \in \mathcal{C} \) and \( n \). Then \( \{X_n(z)\}_{z \in \mathcal{C}} \) converges in distribution to \( \{X(z)\}_{z \in \mathcal{C}} \) as \( n \to \infty \).

With this theorem in hand, the proof of Theorem 5.3.7 will be complete once we prove convergence of finite dimensional distributions (5.18) and tightness (5.19). This is done in the following two sections. Section 5.4 proves the convergence of finite dimensional distribution, and then shows that this implies the variance claimed in Theorem 5.3.6. Section 5.5 proves the tightness of the sequence of stochastic processes.

5.4 Convergence of Finite Dimensional Distributions

This section is devoted to proving the convergence of 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 5.4.1.** For a fixed positive integer \( L \) and any collection \( (z_1, z_2, \ldots, z_L) \) such that \( |z_i| = 1 + \delta \) for \( 1 \leq i \leq L \), the random vector

\[
(\Xi_n(z_1), \Xi_n(z_2), \ldots, \Xi_n(z_L))
\]

converges in distribution to the random vector

\[
(\Xi(z_1), \Xi(z_2), \ldots, \Xi(z_L))
\]

where \( \Xi(z) \) is a mean-zero Gaussian process with variance and covariance structure

\[
\mathbb{E} \left[ \Xi(z_i)\Xi(z_j) \right] = \frac{m^2(z_i \bar{z}_j)^{m-1}}{((z_i \bar{z}_j)^{m} - 1)^2}.
\]
To prove Theorem 5.4.1, we first make a sequence of reductions inspired by the proofs in [105, 93]. 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

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

converges in distribution to

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

for \(\alpha_l, \beta_l \in \mathbb{C}\) such that (5.20) is real. Note that from the linear combination (5.21), we can determine the covariance structure of

\[
(\Xi(z_1), \Xi(z_2), \ldots, \Xi(z_L)).
\]

(5.22)

First, note that each component in (5.22) is complex. Typically, in order to fully characterize the limiting distribution of the vector, we need to calculate

$$E[\Xi(z_i)^2], E[\Xi(z_i)\Xi(z_i)], E[\Xi(z_i)\Xi(z_j)], \text{ and } E[\Xi(z_i)\Xi(z_j)]$$

(5.23)

for \(1 \leq i, j \leq L\). In this case, since the entries in the matrix \(Y_n\) are real, this computation can be simplified. We can see that

$$\Xi_n(z_i) = \Xi_n(z_i)$$

and since \(z_i \in \mathbb{C}\), we know \(\Xi_i \in \mathbb{C}\) as well. Therefore, in this case, it is sufficient to compute

$$E[\Xi(z_i)\Xi(z_j)]$$

in order to characterize the covariance structure of the vector.

Remark 5.4.2. In the case where the atom random variables are complex, the above statements would not be true. In particular, for complex atom variables, \(\Xi_n(z_i) \neq \Xi_n(z_i)\). Therefore, we would need to compute all four quantities in (5.23) to characterize the covariance.
In order to prove Theorem 5.4.1 we will express the sum in (5.20) as a martingale difference sequence. Recall that
\[ \Xi_n(z) = \text{tr} \mathcal{G}_n(z) 1_{\Omega_n} - \mathbb{E}[\text{tr} \mathcal{G}_n(z) 1_{\Omega_n}] \]

Let \( c_k \) denote the \( k \)th column of \( \mathcal{Y}_n \) and define the \( \sigma \)-algebras
\[ \mathcal{F}_k := \sigma(c_1, \ldots, c_k, c_{n+1}, \ldots, c_{n+k}, \ldots, c_{(m-1)n+1}, \ldots, c_{(m-1)n+k}) \] (5.24)
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 \( c_k, c_{n+k}, c_{2n+k}, \ldots, c_{(m-1)n+k} \) 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. Define the event
\[ \Omega_{n,k} := \left\{ \inf_{|z| > 1+\delta/2} s_{mn} \left( \mathcal{Y}_n^{(k)} - zI \right) \geq c \right\} \] (5.25)
for some \( c > 0 \) and note that by Corollary F.0.6, \( \Omega_{n,k} \) holds with overwhelming probability. Finally, define the resolvent
\[ \mathcal{G}_n^{(k)}(z) := \left( \mathcal{Y}_n^{(k)} - zI \right)^{-1} \] (5.26)

Remark 5.4.3. Note that the definitions of the \( \sigma \)-algebras \( \mathcal{F}_k \) differ between Chapter 4 and Chapter 5. Indeed, observe that (5.24) differs from (4.29). Due to the differing definitions of \( \mathcal{F}_k \), the conditional expectations \( \mathbb{E}_k \) differ as well. Additionally, the definitions of \( \mathcal{Y}_n^{(k)} \) differ between Chapters 4 and 5 as well. While they both involve setting rows and column equal to zero, in Chapter 4 we set rows and columns equal to zero while the definition of \( \mathcal{Y}_n^{(k)} \) above involves setting many columns equal to zero. Due to the differing definitions of \( \mathcal{Y}_n^{(k)} \), the definitions of \( \Omega_{n,k} \) and \( \mathcal{G}_n^{(k)}(z) \) change as well.
The following lemma follows from an application of Proposition 3.3.1.

**Lemma 5.4.4.** Define the events \( \Omega_n \) and \( \Omega_{n,k} \) as in (5.12) and (5.25) respectively. Then there exist a constant \( C > 0 \) such that, for all \( z \in \mathcal{C} \), \( \| G_n(z) \| \leq C \) surely on \( \Omega_n \) and \( \| G^{(k)}_n(z) \| \leq C \) surely on \( \Omega_{n,k} \).

With these definitions, we may write

\[
\Xi_n(z) = \text{tr} G_n(z) 1_{\Omega_n} - \mathbb{E} [\text{tr} G_n(z) 1_{\Omega_n}]
\]

\[
= \sum_{k=1}^{n} (\mathbb{E}_k[\text{tr} G_n(z) 1_{\Omega_n}] - \mathbb{E}_{k-1}[\text{tr} G_n(z) 1_{\Omega_n}])
\]

\[
= \sum_{k=1}^{n} Z_{n,k}(z)
\]

where we define

\[
Z_{n,k}(z) := (\mathbb{E}_k - \mathbb{E}_{k-1})[\text{tr} G_n(z) 1_{\Omega_n}].
\]

With this notation, we can rewrite the linear combination from (5.20) as

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

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

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

\[
M_n := \sum_{k=1}^{n} M_{n,k}.
\]

Notice that in fact \( M_n \) is the linear combination of the finite dimensional distributions, so the goal of this section is to prove that \( M_n \) converges to a mean-zero Gaussian with appropriate variance. In order to simplify computations, it will be beneficial to work with a slightly different expression in which some reductions are made.

**Lemma 5.4.5.** Define \( M_n \) as in (5.28), define \( U_k \) to be the \( mn \times m \) matrix which contains as its columns \( c_k, c_{n+k}, \ldots, c_{(m-1)n+k} \), and define \( V_k \) to be the \( mn \times m \) matrix which contains as its
columns $e_k, e_{n+k}, \ldots, e_{(m-1)n+k}$ where $e_1, \ldots e_{mn}$ denote the standard basis elements of $\mathbb{C}^{mn}$. Define

$$\tilde{M}_n := \sum_{k=1}^{n} \tilde{M}_{n,k}$$

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

and

$$\tilde{Z}_{n,k}(z) := -E_k \left[ \text{tr} \left( V_T^k (G_{n}^{(k)}(z))^2 U_k \right) 1_{\Omega_{n,k}} \right]. \quad (5.29)$$

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

Before proving this lemma, note that the difference between $M_n$ and $\tilde{M}_n$ relies completely on the difference between $Z_{n,k}(z)$ and $\tilde{Z}_{n,k}(z)$. If one can show that the difference of $M_n$ and $\tilde{M}_n$ converges to zero in probability, then these two random variables will converge in distribution to the same limit by Lemma D.0.7. The purpose of Lemma 5.4.5 is to introduce independence in each term. Namely, observe that $G_{n}^{(k)}(z) 1_{\Omega_{n,k}}$ is independent of the columns $c_k, c_{n+k}, \ldots, c_{(m-1)n+k}$. This independence will be helpful in proving Theorem 5.4.1.

We first develop some results we will need in the proof of Lemma 5.4.5. Define the event

$$Q_{n,k}(z) := \left\{ \left\| V_T^k G_{n}^{(k)}(z) U_k 1_{\Omega_{n,k}} \right\| \leq 1/2 \right\}. \quad (5.30)$$

We will also need the following lemma.

**Lemma 5.4.6.** Define the event $Q_{n,k}(z)$ as in (5.30). Then, uniformly for any $z \in \mathcal{C}$, $Q_{n,k}$ holds with overwhelming probability.

The proof of this lemma is presented in Appendix G.

**Lemma 5.4.7.** 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, and which satisfies $|\xi| \leq n^{1/2-\varepsilon}$ almost surely for some constant $\varepsilon > 0$. 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 \( i \not\in S \).

Then for any \( p \geq 2 \),

\[
E |w^* Aw|^p \ll_{d,p} n^{(1 - 2\varepsilon)p + 4\varepsilon - 2}E|\xi|^4 \|A\|^p.
\] (5.31)

Remark 5.4.8. Note that while Lemma 5.4.7 is similar to Lemma 4.6.6, the results differ slightly based on assumptions on the random variable \( \xi \).

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

\[
w^*Aw = \sum_{i,j} w_i A_{(i,j)} w_j = w^*_S A_{S \times S} w_S.
\]

By Lemma D.0.3 we get

\[
E |w^* Aw|^p \ll_p (\text{tr} A_{S \times S})^p + E|\xi|^{2p} \text{tr} A_{S \times S}^p
\]

\[
= (\text{tr} A_{S \times S})^p + E [|\xi|^4 |\xi|^{2p-4}] \text{tr} A_{S \times S}^p
\]

\[
\leq (\text{tr} A_{S \times S})^p + n^{(1 - 2\varepsilon)p + 4\varepsilon - 2}E|\xi|^4 \text{tr} A_{S \times S}^p
\]

Since the rank of \( A_{S \times S} \) is at most \( d \),

\[
\text{tr} A_{S \times S} \ll_d \|A\|
\]

and

\[
\text{tr} A_{S \times S}^p \ll_d \|A\|^p,
\]

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

\[
E |w^* Aw|^p \ll_{d,p} \|A\|^p + n^{(1 - 2\varepsilon)p + 4\varepsilon - 2}E|\xi|^4 \|A\|^p \ll_{d,p} n^{(1 - 2\varepsilon)p + 4\varepsilon - 2}E|\xi|^4 \|A\|^p,
\]

as desired. \( \square \)
Lemma 5.4.9. Let $A$ be a deterministic complex $mn \times mn$ matrix for some fixed $m > 0$. Suppose that $\xi$ and $\zeta$ are complex-valued random variables 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) $\{w_i: i \in S\}$ is an independent collection of iid copies of $\xi$,

(ii) $\{t_j: j \in R\}$ is an independent collection of iid copies of $\zeta$,

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

Then for any $p \geq 1$,

$$
\mathbb{E} |w^* At|^2p \ll_p \mathbb{E} |\xi|^{2p} \mathbb{E} |\zeta|^{2p} (\text{tr}(A^* A))^p .
$$

(5.32)

Remark 5.4.10. Note that while Lemma 5.4.9 is similar to Lemma 4.6.7, the results differ based on the assumptions of the random variables used.

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 D.0.3 implies that

$$
\mathbb{E} |w^* At|^2p = \mathbb{E} |w^* Att^* A^* w|^p
$$

$$
= \mathbb{E} |w_S^* (Att^* A^*)_{S \times S} w_S|^p
$$

$$
\ll_p \mathbb{E} [(\text{tr}(Att^* A^*)_{S \times S})^p + \mathbb{E} |\xi|^{2p} \text{tr}(Att^* A^*)_{S \times S}^p] .
$$

Recall that for any matrix $B$, $\text{tr}(B^* B)^p \leq (\text{tr}(B^* B))^p$. By this fact and by Lemma 4.6.5, we observe that

$$
\mathbb{E} [(\text{tr}(Att^* A^*)_{S \times S})^p + \mathbb{E} |\xi|^{2p} \text{tr}(Att^* A^*)_{S \times S}^p] \ll_p \mathbb{E} |\xi|^{2p} \mathbb{E} [(\text{tr}(Att^* A^*))^p] .
$$

By a cyclic permutation of the trace, we have

$$
\mathbb{E} [(\text{tr}(Att^* A^*))^p] = \mathbb{E} [(t^* A^* A t)^p] \leq \mathbb{E} |t^* A^* At|^p .
$$
By Lemma D.0.3, Lemma 4.6.5, and a similar argument as above, we have

\[ E|t^*A^*At|^p = E|t^*_R(A^*A)_{R \times R}|^p \]

\[ \ll_p (\text{tr}(A^*A)_{R \times R})^p + E|\zeta|^{2p} \text{tr}(A^*A)^p \]

\[ \ll_p E|\zeta|^{2p}(\text{tr}(A^*A))^p, \]

and thus we have

\[ E|w^*At|^{2p} \ll_p E|\xi|^{4p}(\text{tr}(A^*A))^p \]

completing the proof.

\[ \square \]

Remark 5.4.11. Note that if \( p \geq 1 \) and we also assume that \( |\xi| < n^{1/2-\varepsilon} \) surely for some \( \varepsilon > 0 \), then we may write

\[ E|w^*At|^{2p} \ll_p E|\xi|^{4p}(\text{tr}(A^*A))^p \]

\[ = E \left[ |\xi|^4|\xi|^{4p-4} \right] (\text{tr}(A^*A))^p \]

\[ \ll n^{(2-4\varepsilon)p+4\varepsilon-2}E|\xi|^4(\text{tr}(A^*A))^p. \]

Lemma 5.4.12. 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 \( G_n^{(k)}(z) \) be defined as in (5.26).

Then for any \( p \geq 2 \),

\[ E \left\| V_k^T G_n^{(k)}(z)U_k1_{\Omega_{n,k}} \right\|^{2p} \ll n^{-2ep+4\varepsilon-2}. \]

Proof. Begin by observing that

\[ E \left\| V_k^T G_n(z)U_k1_{\Omega_{n,k}} \right\|^{2p} \]

\[ \ll \max_{1 \leq i,j \leq m} E \left| (V_k^T G_n^{(k)}(z)U_k)_{(i,j)}1_{\Omega_{n,k}} \right|^{2p} \]

\[ = \max_{1 \leq i,j \leq m} E \left| e_{(i-1)n+k} G_n^{(k)}(z)c_{(j-1)n+k}1_{\Omega_{n,k}} \right|^{2p} \]

\[ = \max_{1 \leq i,j \leq m} E \left| e_{(j-1)n+k}^* G_n^{(k)}(z)^* e_{(i-1)n+k}1_{\Omega_{n,k}} \right|^p. \]
Note that the columns \(c(j-1)_{n+k}\) have a factor of \(n^{-1/2}\) built into the notation. By this observation and Lemma 5.4.7 and since the rank of \((G_n^{(k)}(z))^*e_{(i-1)n+k}e^T_{(i-1)n+k}G_n(z)\) is at most 1, for any \(1 \leq j \leq m\) we have

\[
\mathbb{E} \left| c_{(j-1)_{n+k}}(G_n^{(k)}(z))^*e_{(i-1)n+k}e^T_{(i-1)n+k}G_n^{(k)}(z)c_{(j-1)_{n+k}}1_{\Omega_{n,k}} \right|^p \ll_p n^{-p(n(1-2\varepsilon)p+4\varepsilon-2)} \mathbb{E} \left| \tilde{\xi}_{(j-2)_{n+k}} \right|^4 \left| (G_n^{(k)}(z))^*e_{(i-1)n+k}e^T_{(i-1)n+k}G_n^{(k)}(z)1_{\Omega_{n,k}} \right|^p \ll_p n^{-2\varepsilon p+4\varepsilon-2}.
\]

Thus,

\[
\mathbb{E} \left\| V_k^T G_n^{(k)}(z)U_k 1_{\Omega_{n,k}} \right\|^{2p} \ll_p n^{-2\varepsilon p+4\varepsilon-2}
\]

as advertised. \(\square\)

**Remark 5.4.13.** In the case when the exponent in Lemma 5.4.12 is 2, the above result can be modified since the truncated random variables defined in (5.9) have four finite moments. In this case, we have

\[
\mathbb{E} \left\| V_k^T G_n^{(k)}(z)U_k 1_{\Omega_{n,k}} \right\|^2 \ll n^{-1}.
\]

A similar argument also shows that

\[
\mathbb{E} \left\| V_k^T (G_n^{(k)}(z))^2U_k 1_{\Omega_{n,k}} \right\|^2 \ll n^{-1}
\]

and

\[
\mathbb{E} \left\| V_k^T (G_n^{(k)}(z))^2U_k 1_{\Omega_{n,k}} \right\|^{2p} \ll_p n^{-2\varepsilon p+4\varepsilon-2} \tag{5.33}
\]

for \(p \geq 2\). Full details of these arguments can be found in Lemmas H.0.1 and H.0.2 in Appendix H.

We now proceed with the proof of Lemma 5.4.5.

**Proof of Lemma 5.4.5.** To begin, note that the result will follow if we prove that

\[
\mathbb{E}|M_n - \hat{M}_n|^2 = o(1).
\]
Since the only difference between these two expressions is the difference between $Z_{n,k}(z)$ and $\tilde{Z}_{n,k}(z)$, it will be sufficient to prove that for any $z$ on the contour,

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

Since $Z_{n,k}$ and $\tilde{Z}_{n,k}$ are martingale difference sequences, by Lemma D.0.8, we can conclude that

$$\mathbb{E} \left| \sum_{k=1}^{n} \left( Z_{n,k}(z) - \tilde{Z}_{n,k}(z) \right) \right|^2 \ll \sum_{k=1}^{n} \mathbb{E} \left| Z_{n,k}(z) - \tilde{Z}_{n,k}(z) \right|^2.$$

Ergo, it is sufficient to prove that

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

uniformly for any $0 < k \leq n$ and any $z$ on the contour $C$. We will prove this by making 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 $Y_n^{(k)}$ has columns $k, n + k, \ldots, (m - 1)n + k$ replaced with zeros, we have

$$\mathbb{E}_k[\text{tr} \ G_n^{(k)}(z) 1_{\Omega_{n,k}}] = \mathbb{E}_{k-1}[\text{tr} \ G_n^{(k)}(z) 1_{\Omega_{n,k}}]$$

and thus

$$(\mathbb{E}_k - \mathbb{E}_{k-1})[\text{tr} \ G_n^{(k)}(z) 1_{\Omega_{n,k}}] = 0.$$

Thus, we can rewrite

$$Z_{n,k}(z) = (\mathbb{E}_k - \mathbb{E}_{k-1})[\text{tr} \ G_n(z) 1_{\Omega_n}]$$

$$= (\mathbb{E}_k - \mathbb{E}_{k-1})[\text{tr} \ G_n(z) 1_{\Omega_n} - \text{tr} \ G_n^{(k)}(z) 1_{\Omega_{n,k}}]$$

$$= (\mathbb{E}_k - \mathbb{E}_{k-1})[(\text{tr} \ G_n(z) - \text{tr} \ G_n^{(k)}(z)) 1_{\Omega_n \cap \Omega_{n,k}}]$$

$$+ (\mathbb{E}_k - \mathbb{E}_{k-1})[\text{tr} \ G_n(z) 1_{\Omega_n \cap \Omega_{n,k}^c}]$$

$$- (\mathbb{E}_k - \mathbb{E}_{k-1})[\text{tr} \ G_n^{(k)}(z) 1_{\Omega_{n,k} \cap \Omega_n^c}].$$

Note that, uniformly for $z$ with $|z| = 1 + \delta$, by Lemma 5.4.4

$$\mathbb{E} \left| (\mathbb{E}_k - \mathbb{E}_{k-1})[\text{tr} \ G_n(z) 1_{\Omega_n \cap \Omega_{n,k}^c}] \right|^2 \ll \mathbb{E} \left| \text{tr} \ G_n(z) 1_{\Omega_n \cap \Omega_{n,k}^c} \right|^2$$

$$\ll n^2 \mathbb{E} \left[ \|G_n(z)\|^2 1_{\Omega_n \cap \Omega_{n,k}^c} \right]$$

$$\ll_{\alpha} n^{2-\alpha}.$$
for any \( \alpha > 0 \). The same argument shows that \( \mathbb{E} \left[ |(E_k - E_{k-1})[\text{tr} \mathcal{G}_n^{(k)}(z)\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k}}]|^2 \right] \ll n^{2-\alpha} \) for any \( \alpha > 0 \). Ergo, we have reduced from working with \( Z_{n,k}(z) \) to working with \((E_k - E_{k-1})[\text{tr} \mathcal{G}_n(z) - \text{tr} \mathcal{G}_n^{(k)}(z))\mathbf{1}_{\Omega_{n} \cap \Omega_{n,k}}]\).

Next, observe that by linearity and cyclic permutation of the trace, and by the resolvent identity \((3.6)\),

\[
\text{tr} \mathcal{G}_n(z) - \text{tr} \mathcal{G}_n^{(k)}(z) = \text{tr} \left( \mathcal{G}_n(z) \left( \mathcal{Y}_n^{(k)} - \mathcal{Y}_n^{(k)} \right) \mathcal{G}_n^{(k)}(z) \right)
= -\text{tr} \left( \mathcal{G}_n(z) U_k V_k^T \mathcal{G}_n^{(k)}(z) \right)
= -\text{tr} \left( V_k^T \mathcal{G}_n^{(k)}(z) \mathcal{G}_n(z) U_k \right).
\]

To guarantee that \( I_m + V_k^T \mathcal{G}_n^{(k)} U_k \) is invertible, we may work on the event \( Q_{n,k} \) defined in \((5.30)\) which implies invertibility by the reverse triangle inequality. By Lemma \[5.4.6\] \( Q_{n,k} \) hold with overwhelming probability so that by Lemma \[5.4.4\] the Cauchy–Schwarz inequality, and bounding the spectral norm by the Frobenius norm, we have

\[
\mathbb{E} \left[ |(E_k - E_{k-1})[\text{tr} (V_k^T \mathcal{G}_n^{(k)}(z) \mathcal{G}_n(z) U_k) \mathbf{1}_{\Omega_{n} \cap \Omega_{n,k}}]|^2 \right] 
\ll \mathbb{E} \left[ \text{tr} (V_k^T \mathcal{G}_n^{(k)}(z) \mathcal{G}_n(z) U_k) \mathbf{1}_{\Omega_{n} \cap \Omega_{n,k} \cap Q_{n,k}} \right] 
\ll \mathbb{E} \left[ \left\| V_k^T \mathcal{G}_n^{(k)}(z) \mathcal{G}_n(z) U_k \mathbf{1}_{\Omega_{n} \cap \Omega_{n,k}} \right\|^2 \mathbf{1}_{Q_{n,k}} \right] 
\ll \mathbb{E} \left[ \left\| U_k \right\|^2 \mathbf{1}_{Q_{n,k}} \right] 
\ll n^2 \mathbb{E} \left[ \mathbf{1}_{Q_{n,k}} \right] 
\ll n^2 \alpha^{-\alpha}
\]

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

\[ -(E_k - E_{k-1})[\text{tr} (V_k^T \mathcal{G}_n^{(k)}(z) \mathcal{G}_n(z) U_k) \mathbf{1}_{\Omega_{n} \cap \Omega_{n,k} \cap Q_{n,k}}] \]

instead of \( Z_{n,k}(z) \). By the Sherman–Morrison–Woodbury formula \((3.5)\), we have

\[
- (E_k - E_{k-1})[\text{tr} (V_k^T \mathcal{G}_n^{(k)}(z) \mathcal{G}_n(z) U_k) \mathbf{1}_{\Omega_{n} \cap \Omega_{n,k} \cap Q_{n,k}}]
= - (E_k - E_{k-1})[\text{tr} (V_k^T \mathcal{G}_n^{(k)}(z))^2 U_k (I_m + V_k^T \mathcal{G}_n^{(k)} U_k)^{-1} \mathbf{1}_{\Omega_{n} \cap \Omega_{n,k} \cap Q_{n,k}}].
\]
Since \( G_n(z) \) is no longer present, we may drop the event \( \Omega_n \) gaining a sufficiently small error, and the same argument justifies working with

\[
-(\mathbb{E}_k - \mathbb{E}_{k-1}) \left[ \text{tr}(V_k^T (G_n^{(k)}(z))^2 U_k (I_m + V_k^T G_n^{(k)} U_k)^{-1}) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right]
\]

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

\[
\mathbb{E} \left[ (\mathbb{E}_k - \mathbb{E}_{k-1}) \left[ \text{tr}(V_k^T (G_n^{(k)}(z))^2 U_k (I_m + V_k^T G_n^{(k)} U_k)^{-1}) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right] \right] - \left( \mathbb{E}_k - \mathbb{E}_{k-1} \right) \left[ \text{tr}(V_k^T (G_n^{(k)}(z))^2 U_k I_m) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right]^2 \\
\ll \mathbb{E} \left[ \left\| V_k^T (G_n^{(k)}(z))^2 U_k ((I_m + V_k^T G_n^{(k)} U_k)^{-1} - I_m) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right].
\]

(5.35)

Note that by the resolvent identity (3.6),

\[
(I_m + V_k^T G_n^{(k)} U_k)^{-1} = I_m - (I_m + V_k^T G_n^{(k)} U_k)^{-1} V_k^T G_n^{(k)}(z) U_k.
\]

(5.36)

By iterating this twice, we get that

\[
(I_m + V_k^T G_n^{(k)} U_k)^{-1} - I_m = -V_k^T G_n^{(k)}(z) U_k + (I_m + V_k^T G_n^{(k)} U_k)^{-1} (V_k^T G_n^{(k)}(z) U_k)^2.
\]

Inserting this into the last line of (5.35), we get

\[
\mathbb{E} \left[ \left\| V_k^T (G_n^{(k)}(z))^2 U_k ((I_m + V_k^T G_n^{(k)} U_k)^{-1} - I_m) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right] \\
= \mathbb{E} \left[ \left\| V_k^T (G_n^{(k)}(z))^2 U_k (-V_k^T G_n^{(k)}(z) U_k) \\
+ (I_m + V_k^T G_n^{(k)} U_k)^{-1} (V_k^T G_n^{(k)}(z) U_k)^2 \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right] \\
\ll \mathbb{E} \left[ \left\| V_k^T (G_n^{(k)}(z))^2 U_k (V_k^T G_n^{(k)}(z) U_k) \mathbf{1}_{\Omega_{n,k}} \right\|^2 \right] \\
+ \mathbb{E} \left[ \left\| V_k^T (G_n^{(k)}(z))^2 U_k ((I_m + V_k^T G_n^{(k)} U_k)^{-1} (V_k^T G_n^{(k)}(z) U_k)) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right].
\]

(5.37)

(5.38)

We will bound each of the above terms separately. First, we begin with term (5.37). Note that by
Cauchy–Schwarz inequality, Lemma \textbf{5.4.12} and equation \textbf{(5.33)}, we have

\[
E \left\| V_k^T (G_n^{(k)}(z))^2 U_k (V_k^T G_n^{(k)}(z) U_k) 1_{\Omega_{n,k}} \right\|^2 \\
\ll \left( E \left\| V_k^T (G_n^{(k)}(z))^2 U_k 1_{\Omega_{n,k}} \right\|^4 \right)^{1/2} \left( E \left\| V_k^T G_n^{(k)}(z) U_k 1_{\Omega_{n,k}} \right\|^4 \right)^{1/2} \\
\ll (n^{-2} \cdot n^{-2})^{1/2} \\
= o(n^{-1}).
\]

It remains to show that term \textbf{(5.38)} is also \(o(n^{-1})\). To this end, observe that by the Cauchy–Schwarz inequality, Lemma \textbf{5.4.12} and equation \textbf{(5.33)},

\[
E \left\| V_k^T (G_n^{(k)}(z))^2 U_k (I_m + V_k^T G_n^{(k)} U_k)^{-1} (V_k^T G_n^{(k)}(z) U_k)^2 1_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \\
\ll \left( E \left\| (I_m + V_k^T G_n^{(k)} U_k)^{-1} 1_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right)^{1/2} \left( E \left\| (V_k^T G_n^{(k)}(z) U_k)^2 1_{\Omega_{n,k}} \right\|^4 \right)^{1/2} \\
\ll n^{-1} \left( E \left\| V_k^T G_n^{(k)}(z) U_k \right\|^8 1_{\Omega_{n,k}} \right)^{1/2} \\
\ll n^{-2\varepsilon-2}.
\]

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

\[
-(E_k - E_{k-1}) [\text{tr}(V_k^T (G_n^{(k)}(z))^2 U_k) 1_{\Omega_{n,k} \cap Q_{n,k}}].
\]

Next, we will justify removing the event \(Q_{n,k}\). Observe that by bounding the trace by the rank times the norm, and by Remark \textbf{5.4.13}

\[
E \left\| (E_k - E_{k-1}) [\text{tr}(V_k^T (G_n^{(k)}(z))^2 U_k) 1_{\Omega_{n,k}}] \right\|^2 \\
\ll E \left\| \text{tr}(V_k^T (G_n^{(k)}(z))^2 U_k) 1_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \\
\ll E \left\| V_k^T (G_n^{(k)}(z))^2 U_k 1_{\Omega_{n,k}} \right\|^2 1_{Q_{n,k}}^e \\
\ll n^{-1} E \left[ 1_{Q_{n,k}}^e \right] \\
\ll \alpha n^{-1-\alpha}.
\]
By selecting $\alpha$ sufficiently large in the above expression, we can proceed with

$$-(\mathbb{E}_k - \mathbb{E}_{k-1})[\text{tr}(V^T_k(G^{(k)}(z))^2U_k)1_{\Omega_{n,k}}].$$

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

$$\mathbb{E}_{k-1}[\text{tr}(V^T_k(G^{(k)}(z))^2U_k)1_{\Omega_{n,k}}]$$

$$= \mathbb{E}_{k-1}\left[\sum_{i=1}^{m} \sum_{a,b=1}^{m} (V^T_k)_{i,a}(G^{(k)}(z))_{a,b}^2(U_k)_{b,i}1_{\Omega_{n,k}}\right]$$

$$= \sum_{i=1}^{m} \sum_{a,b=1}^{m} (V^T_k)_{i,a}\mathbb{E}_{k-1}\left[(G^{(k)}_n(z))_{a,b}^21_{\Omega_{n,k}}\right]\mathbb{E}_{k-1}[(U_k)_{b,i}]$$

$$= 0.$$

Therefore, we can work with

$$-\mathbb{E}_k[\text{tr}(V^T_k(G^{(k)}(z))^2U_k)1_{\Omega_{n,k}}]$$

instead of $Z_{n,k}(z)$, completing the proof. \qed

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

**Theorem 5.4.14** (Theorem 35.12 of [33]). *For each $N$, suppose $Z_{N_1}, Z_{N_2}, \ldots, Z_{N_{r_N}}$ 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 $\nu^2$,*

$$\lim_{N \to \infty} \mathbb{P}\left(\left|\sum_{j=1}^{r_N} \mathbb{E}\left(Z^2_{N_j}|\mathcal{F}_{N_{j-1}}\right) - \nu^2\right| > \eta\right) = 0 \quad (5.39)$$

*and*

$$\lim_{N \to \infty} \sum_{j=1}^{r_N} \mathbb{E}\left(Z^2_{N_j} 1_{\{|Z_{N_j}| \geq \eta\}}\right) = 0. \quad (5.40)$$

*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 \(\{\tilde{M}_{n,k}\}_{k=1}^{n}\) and the corresponding \(\sigma\)-algebras are \(\{\mathcal{F}_k\}\). To utilize this theorem, first note that \(\{\tilde{M}_{n,k}\}\) is a real martingale difference sequence by choice of \(\alpha_l\) and \(\beta_l\), and it has finite second moments. For finite variance, let \(\kappa := \max_{1 \leq l \leq L}\{|\alpha_l|, |\beta_l|\}\) and observe that

\[
\mathbb{E} \left| \tilde{M}_{n,k} \right|^2 = \mathbb{E} \left| \sum_{l=1}^{L} \alpha_l \tilde{Z}_{n,k}(z_l) + \beta_l \tilde{Z}_{n,k}(z_l) \right|^2 \\
\ll_{L} \sum_{l=1}^{L} \mathbb{E} \left| \alpha_l \tilde{Z}_{n,k}(z_l) + \beta_l \tilde{Z}_{n,k}(z_l) \right|^2 \\
\ll_{\kappa,L} \sum_{l=1}^{L} \mathbb{E} \left| \tilde{Z}_{n,k}(z_l) \right|^2 \\
\ll_{\kappa,L} \sum_{l=1}^{L} \mathbb{E} \left\| V_k^T (G^{(k)}(z_l))^2 U_k \mathbf{1} \right\|_2^2 \\
\ll_{\kappa,L} \sum_{l=1}^{L} n^{-1}
\]

by independence of \(U_k\) from \(G^{(k)}(z)\) and by Remark 5.4.13. To apply Theorem 5.4.14, we must also verify that the hypotheses (5.39) and (5.40) hold. Verifying (5.39) holds for \(\{\tilde{M}_{n,k}\}_{k=1}^{n}\) is lengthy and will require new notation, so we begin with verifying (5.40) holds for \(\{\tilde{M}_{n,k}\}_{k=1}^{n}\).

Let \(\eta > 0\) and observe that by equation (5.33), we have

\[
\sum_{k=1}^{n} \mathbb{E} \left[ \tilde{M}_{n,k}^2 \mathbf{1}_{\{|\tilde{M}_{n,k}| > \eta\}} \right] \ll \sum_{k=1}^{n} \mathbb{E} \left| \tilde{M}_{n,k}^2 \mathbf{1}_{\{|\tilde{M}_{n,k}| > \eta\}} \right|^4 \\
\ll_{\eta} \sum_{k=1}^{n} \mathbb{E} \left( \sum_{l=1}^{L} \alpha_l \tilde{Z}_{n,k}(z_l) + \beta_l \tilde{Z}_{n,k}(z_l) \right)^4 \\
\ll_{\eta,\kappa,L} \sum_{k=1}^{n} \sum_{l=1}^{L} \mathbb{E} \left\| V_k^T (G^{(k)}(z_l))^2 U_k \mathbf{1} \right\|_2^4 \\
\ll_{\eta,\kappa,L} n^{-1}.
\]

This verifies condition (5.40) of Theorem 5.4.14. Now, to verify (5.39), we have the following lemma.

**Lemma 5.4.15.** The martingale difference sequence

\[
\{\tilde{M}_{n,k}\} = \left\{ \sum_{l=1}^{L} \alpha_l \tilde{Z}_{n,k}(z_l) + \beta_l \tilde{Z}_{n,k}(z_l) \right\}
\]
has finite second moments and satisfies
\[
\sum_{k=1}^{n} \mathbb{E}_{k-1}[\tilde{M}_{n,k}^2] \to \sum_{1 \leq i, j \leq L} \alpha_i \alpha_j (1 - z_i z_j)^{-2} + \alpha_i \beta_j (1 - z_i \bar{z}_j)^{-2} + \beta_i \alpha_j (1 - \bar{z}_i z_j)^{-2} + \beta_i \beta_j (1 - \bar{z}_i \bar{z}_j)^{-2}
\] (5.41)
in probability as \( n \to \infty \).

**Remark 5.4.16.** In the case where the atom variables are complex, the limit of \( \sum_{k=1}^{n} \mathbb{E}_{k-1}[\tilde{M}_{n,k}^2] \) would differ from that of Lemma 5.4.15.

In the proof of Lemma 5.4.15, we will need some definitions and results. We develop these definitions and results now before proceeding to the proof.

Define \( \mathcal{Y}_{n}^{(k,s)} \) to be the matrix \( \mathcal{Y}_n \) but with columns \( c_k, c_{n+k}, \ldots, c_{(m-1)n+k}, \) and \( c_s \) filled with zeros. Also define the event
\[
\Omega_{n,k,s} := \left\{ \inf_{|z| > 1 + \delta/2} s_{mn} \left( \mathcal{Y}_{n}^{(k,s)} - zI \right) \geq c \right\}
\] (5.42)
for some \( c > 0 \) and define
\[
\mathcal{G}_{n}^{(k,s)}(z) := \left( \mathcal{Y}_{n}^{(k,s)} - zI \right)^{-1}.
\] (5.43)
By Corollary F.0.7, there exists a constant \( c > 0 \) depending only on \( \delta \) such that \( \Omega_{n,k,s} \) holds with overwhelming probability. By the Sherman-Morrison formula (3.4), provided \( 1 + c_s \mathcal{G}_{n}^{(k,s)}(z)c_s \) is not zero, we may write
\[
\mathcal{G}_{n}^{(k)}(z)c_s = \left( \mathcal{Y}_{n}^{(k)} - zI \right)^{-1} c_s
\]
\[
= \left( \mathcal{Y}_{n}^{(k,s)} + c_s e_s^T - zI \right)^{-1} c_s
\]
\[
= \frac{\left( \mathcal{Y}_{n}^{(k,s)} - zI \right)^{-1} c_s}{1 + c_s^T \mathcal{G}_{n}^{(k,s)}(z)c_s}
\]
\[
= \mathcal{G}_{n}^{(k,s)}(z)c_s \delta_{k,s}(z)
\] (5.44)
where
\[
\delta_{k,s}(z) := (1 + c_s^T \mathcal{G}_{n}^{(k,s)}c_s)^{-1}.
\] (5.45)
By the same formula,
\[ c_n^*(G_k^{(k)}(w))^* = (\delta_k,s(w))^* c_n^*(G_n^{(k,s)}(w))^*. \] (5.46)

We will want to ensure that these quantities exist. To justify this, we introduce the event
\[ Q_{n,k,s}(z) := \left\{ \left| e_s^T G_n^{(k,s)}(z) c_s 1_{\Omega_{n,k,s}} \right| \leq 1/2 \right\}. \] (5.47)

**Lemma 5.4.17.** Define the event \( Q_{n,k,s}(z) \) as in (5.47). Then for any \( z \in C \), \( Q_{n,k,s}(z) \) holds with overwhelming probability.

The proof of this lemma is presented in Appendix G. The next Lemma follows by an application of Proposition 3.3.1.

**Lemma 5.4.18.** On the event \( \Omega_{n,k,s} \), almost surely \( \left\| G_n^{(k,s)}(z) \right\| \leq C \) for a constant \( C > 0 \) uniformly for any \( z \) on the contour \( C \). There exists a constant \( C > 0 \) such that \( |\delta_{k,s}(z) 1_{Q_{n,k,s}}| \leq C \) almost surely, for any \( z \) on the contour \( C \).

With these definitions and results in hand, we proceed with the proof of Lemma 5.4.15. In the proof of Lemma 5.4.15, we make some reductions, each of which produces error terms which are sufficiently small as to not effect the limiting distribution. The proof of Lemma 5.4.15 is similar to the proof of Lemma 3.2 in [105], and the proof of Theorem 5.2 in [93]. However, the following proof differs from previous results due to the block structure of the matrix in question, which must be taken into account. We iterate methods similar to those in [105] over each block.

**Proof of Lemma 5.4.15** We may begin by expanding
\[
E_{k-1}[\bar{M}_{n,k}^2] = E_{k-1} \left[ \left( \sum_{l=1}^{L} \alpha_l \bar{\zeta}_{n,k}(z_l) + \beta_l \bar{\zeta}_{n,k}(z_l) \right)^2 \right] = E_{k-1} \left[ \sum_{i,j=1}^{L} \alpha_i \alpha_j \bar{\zeta}_{n,k}(z_i) \bar{\zeta}_{n,k}(z_j) \right] + E_{k-1} \left[ \sum_{i,j=1}^{L} \alpha_i \beta_j \bar{\zeta}_{n,k}(z_i) \bar{\zeta}_{n,k}(z_j) \right]
\]
where $\tilde{Z}_{n,k}(z)$ was defined in (5.29), and therefore
\begin{equation}
\sum_{k=1}^{n} \mathbb{E}_{k-1} \left[ \hat{M}_{n,k}^2 \right] = \sum_{k=1}^{n} \sum_{i,j=1}^{L} \alpha_i \alpha_j \mathbb{E}_{k-1} \left[ \tilde{Z}_{n,k}(z_i) \tilde{Z}_{n,k}(z_j) \right] + \sum_{k=1}^{n} \sum_{i,j=1}^{L} \alpha_i \beta_j \mathbb{E}_{k-1} \left[ \tilde{Z}_{n,k}(z_i) \tilde{Z}_{n,k}(z_j) \right] + \sum_{k=1}^{n} \sum_{i,j=1}^{L} \beta_i \alpha_j \mathbb{E}_{k-1} \left[ \tilde{Z}_{n,k}(z_i) \tilde{Z}_{n,k}(z_j) \right] + \sum_{k=1}^{n} \sum_{i,j=1}^{L} \beta_i \beta_j \mathbb{E}_{k-1} \left[ \tilde{Z}_{n,k}(z_i) \tilde{Z}_{n,k}(z_j) \right].
\end{equation}

We analyze each of these terms separately. Note that since the entries in the matrix $\mathcal{Y}_n$ are real, $\overline{Z}_{n,k}(z_j) = \tilde{Z}_{n,k}(\overline{z}_j)$ so the calculations for all terms will be the same. Therefore it suffices to show that
\begin{equation}
\sum_{k=1}^{n} \sum_{i,j=1}^{L} \alpha_i \beta_j \mathbb{E}_{k-1} \left[ \tilde{Z}_{n,k}(z) \overline{Z}_{n,k}(w) \right] \rightarrow \sum_{1 \leq i,j \leq L} \alpha_i \beta_j (1-z\overline{w})^{-2}
\end{equation}
in distribution for fixed $z, w \in \mathbb{C}$. For now, we focus on the sum over $k$. Observe that
\begin{equation}
\sum_{k=1}^{n} \mathbb{E}_{k-1} \left[ \tilde{Z}_{n,k}(z) \overline{Z}_{n,k}(w) \right] = \sum_{k=1}^{n} \mathbb{E}_{k-1} \left[ \mathbb{E}_k \left[ \text{tr} \left( V_k^T (G_n^{(k)}(z))^2 U_k \right) \right] \mathbb{E}_k \left[ \text{tr} \left( V_k^T (G_n^{(k)}(w))^2 U_k \right) \right] \right]
\end{equation}
\begin{equation}
= \sum_{k=1}^{n} \mathbb{E}_{k-1} \left[ \mathbb{E}_k \left[ \sum_{i=1}^{m} \sum_{a,b=1}^{m} (V_k^T (G_n^{(k)}(z)))^2 (U_k(i,a))(G_n^{(k)}(z))_2 (U_k)(b,i) \right] \right]
\end{equation}
\begin{equation}
\times \mathbb{E}_k \left[ \sum_{j=1}^{m} \sum_{c,d=1}^{m} (U_k^*(j,d))(G_n^{(k)}(w))^{2*} (V_k)(c,j) \right] \right]
\end{equation}
\begin{equation}
= \sum_{k=1}^{n} \sum_{i,j=1}^{m} \sum_{a,b,c,d=1}^{m} \mathbb{E}_{k-1} \left[ (V_k^T (i,a) \mathbb{E}_k \left[ (G_n^{(k)}(z))_2 (U_k)(b,i) \right] \right]
\end{equation}
\begin{equation}
\times \mathbb{E}_k \left[ (U_k^*(j,d)) \mathbb{E}_k \left[ (G_n^{(k)}(w))^{2*} (V_k)(c,j) \right] \right]
\end{equation}
\begin{equation}
= \sum_{k=1}^{n} \sum_{i,j=1}^{m} \sum_{a,b,c,d=1}^{m} \mathbb{E}_{k-1} \left[ (V_k^T (i,a)) \mathbb{E}_k \left[ (G_n^{(k)}(z))_2 (U_k)(b,i) \right] \right]
\end{equation}
\begin{equation}
\times \mathbb{E}_k \left[ (U_k^*(j,d)) \mathbb{E}_k \left[ (G_n^{(k)}(w))^{2*} (V_k)(c,j) \right] \right].
\end{equation}
At this point, we may exploit the block structure in 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 $e_k, e_{n+k} \ldots e_{(m-1)n+k}$, $(V_k^T)_{(i,a)} = 0$ 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 $G^{(k)}_n(z)$, we can factor this out of the expectation and rewrite (5.53) as

$$\sum_{k=1}^{n} \sum_{i,j=1}^{m} \sum_{a,b,c,d=1}^{mn} E_{k-1} \left[ (V_k^T)_{(i,a)} E_k \left[ (G^{(k)}_n(z))^2_{(a,b)} 1_{\Omega_{n,k}} \right] (U_k)_{(b,i)} \right]$$

$$\times (U_k^*)_{(j,d)} E_k \left[ (G^{(k)}_n(w))^{2*}_{(d,c)} 1_{\Omega_{n,k}} \right] (V_k)_{(c,j)}$$

$$= \sum_{k=1}^{n} \sum_{i,j=1}^{m} \sum_{b,d=1}^{mn} E_{k-1} \left[ E_k \left[ (G^{(k)}_n(z))^2_{(i-1)n+k,b} 1_{\Omega_{n,k}} \right] (U_k)_{(b,i)} \right]$$

$$\times (U_k^*)_{(j,d)} E_k \left[ (G^{(k)}_n(w))^{2*}_{(d,(j-1)n+k)} 1_{\Omega_{n,k}} \right]$$

$$= \sum_{k=1}^{n} \sum_{i,j=1}^{m} \sum_{b,d=1}^{mn} E_{k-1} \left[ E_k \left[ (G^{(k)}_n(z))^2_{(i-1)n+k,b} 1_{\Omega_{n,k}} \right] \right]$$

$$\times E_k \left[ (G^{(k)}_n(w))^{2*}_{(d,(j-1)n+k)} 1_{\Omega_{n,k}} \right] E[(U_k)_{(b,i)}(U_k^*)_{(j,d)}].$$

Now, if $i \neq j$ or $b \neq d$, then $(U_k)_{(b,i)}$ and $(U_k^*)_{(j,d)}$ come from different columns in $Y_n$ and hence are independent and have expectation zero. Therefore, the only non-zero terms are those in which $i = j$ and $b = d$. Thus the sum in (5.53) can be further reduced to

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

Now, since $U_k$ is filled with columns $c_k, c_{n+k}, \ldots c_{(m-1)n+k}$, we may analyze the structure of these columns to evaluate $E \left| (U_k)_{(b,i)} \right|^2$. Since column $c_{(i-1)n+k}$ comes from the $i$th block of $Y_n$, the only nonzero elements of this column are the entries $(i-1)n + 1$ to $i$. Ergo $(U_k)_{(b,i)} = 0$ unless $(i-2)n - 1 \leq b \leq (i-1)n$. For $b$ in such a range, we have $E|U_{(b,i)}|^2 = \frac{1}{n}$. Therefore we can simplify the sum in (5.53) further as

$$\frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{b=-(i-2)n+1}^{(i-1)n} E_k \left[ (G^{(k)}_n(z))^2_{(i-1)n+k,b} 1_{\Omega_{n,k}} \right] E_k \left[ (G^{(k)}_n(w))^{2*}_{b,(i-1)n+k} 1_{\Omega_{n,k}} \right].$$
Note that in the above expression, the sum in terms of $b$ can be viewed as a trace of a partial product from the matrices. Define the diagonal $mn \times mn$ matrix $\mathcal{D}_p$ with entries

$$
(\mathcal{D}_p)_{(i,j)} := \begin{cases} 
1 & \text{if } i = j, \ (p-1)n + 1 \leq i \leq pn \\
0 & \text{otherwise}
\end{cases} \quad (5.54)
$$

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

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

Next, if we can show that

$$
\frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{m} e_{(i-1)n+k} \mathbb{E}_k \left[ (\mathcal{G}_n^{(k)}(z))^2 1_{\Omega_{n,k}} \right] D_i - 1 \mathbb{E}_k \left[ (\mathcal{G}_n^{(k)}(w))^{2*} 1_{\Omega_{n,k}} \right] e_{(i-1)n+k} \quad (5.56)
$$

in distribution as $n \to \infty$, then by Vitali’s theorem (see for instance [19, Lemma 2.14]), it will follow that

$$
\frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{m} e_{(i-1)n+k} \mathbb{E}_k \left[ (\mathcal{G}_n^{(k)}(z))^2 1_{\Omega_{n,k}} \right] D_i - 1 \mathbb{E}_k \left[ (\mathcal{G}_n^{(k)}(w))^{2*} 1_{\Omega_{n,k}} \right] e_{(i-1)n+k} \quad (5.57)
$$

Note that Vitali’s theorem is justified because (5.56) 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 (5.56), we will focus on a fixed term in the sum. Define

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

(5.58)
Provided the resolvent is defined, we have the matrix identity (see for example, [105, Equation (3.15)]),

\[ G_n^{(k)}(z) = -\frac{1}{z} I + \frac{1}{z} \sum_{t \neq n+k} G_n^{(k)}(z) c_t c_t^T \]  \hspace{1cm} (5.59)

where the notation \( t \neq *n + k \) indicates that the sum is over all \( 1 \leq t \leq mn \) such that \( t \neq k, n + k, \ldots, (m - 1)n + k \). We may use this to expand the term \( T_{n,k}(z, \bar{w}) \). Doing so, we have that

\[
\mathbb{E}_k [G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}}] D_{i-1} \mathbb{E}_k [(G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}}]
= \mathbb{E}_k \left[ \left( -\frac{1}{w} I + \frac{1}{w} \sum_{s \neq n+k} G_n^{(k)}(z) c_t c_t^T \right) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-1} \mathbb{E}_k \left[ e_s c_s^* (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right]
\]

\[
= \frac{1}{zw} D_{i-1} \mathbb{P}_k(\Omega_{n,k})^2 - \frac{1}{zw} \sum_{s \neq n+k} D_{i-1} \mathbb{E}_k \left[ e_s c_s^* (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] \mathbb{P}_k(\Omega_{n,k})
\]

\[
+ \frac{1}{zw} \sum_{s \neq n+k} \sum_{t \neq n+k} \mathbb{E}_k \left[ G_n^{(k)}(z) c_t c_t^T \mathbf{1}_{\Omega_{n,k}} \right] D_{i-1} \mathbb{E}_k \left[ e_s c_s^* (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right]
\]

where \( \mathbb{P}_k \) denotes the conditional probability with respect to \( \mathcal{F}_k \) and we assume that the subscript on \( D_{i-1} \) is reduced modulo \( m \). Therefore one can write

\[
T_{n,k}(z, \bar{w}) = \sum_{i=1}^{m} e_{(i-1)n+k}^T D_{i-1} \mathbb{E}_k \left[ e_s c_s^* (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] \mathbb{P}_k(\Omega_{n,k}) e_{(i-1)n+k}
\]

\[
- e_{(i-1)n+k}^T \frac{1}{zw} \sum_{s \neq n+k} D_{i-1} \mathbb{E}_k \left[ e_s c_s^* (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] \mathbb{P}_k(\Omega_{n,k}) e_{(i-1)n+k}
\]

\[
- e_{(i-1)n+k}^T \frac{1}{zw} \sum_{t \neq n+k} \mathbb{E}_k \left[ G_n^{(k)}(z) c_t c_t^T \mathbf{1}_{\Omega_{n,k}} \right] D_{i-1} \mathbb{P}_k(\Omega_{n,k}) e_{(i-1)n+k}
\]

\[
+ e_{(i-1)n+k}^T \frac{1}{zw} \sum_{s \neq n+k} \sum_{t \neq n+k} \mathbb{E}_k \left[ G_n^{(k)}(z) c_t c_t^T \mathbf{1}_{\Omega_{n,k}} \right] D_{i-1} \mathbb{E}_k \left[ e_s c_s^* (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] e_{(i-1)n+k}
\]

\hspace{1cm} (5.60)
We look at each term separately. For any \(1 \leq i \leq m\), term (5.60) simplifies to

\[
e_{(i-1)n+k}^T \mathcal{D}_{i-1} \frac{1}{z \bar{w}} (P_k(\Omega_{n,k}))^2 e_{(i-1)n+k} = 0.
\]

For terms (5.61) and (5.62), since \(s \neq jn + k\) for any \(1 \leq j \leq m\), we have the expression

\[
e_{(i-1)n+k}^T \mathcal{D}_{i-1} e_s,
\]

which results in an off diagonal element of \(\mathcal{D}_{i-1}\). Ergo,

\[
- e_{(i-1)n+k}^T \frac{1}{z \bar{w}} \sum_{s \neq s + n + k} \mathcal{D}_{i-1} E_k \left[ e_s c_s^* (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] P_k(\Omega_{n,k}) e_{(i-1)n+k}
\]

\[
= - \frac{1}{z \bar{w}} \sum_{s \neq s + n + k} e_{(i-1)n+k}^T \mathcal{D}_{i-1} c_s \mathbb{E}_k \left[ c_s^* (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] P_k(\Omega_{n,k}) e_{(i-1)n+k}
\]

\[
= 0
\]

since \((\mathcal{D}_{i-1})_{a,b} = 0\) unless \((i - 2)n + 1 \leq a = b \leq (i - 1)n\). Similarly, since \(t \neq jn + k\) for any \(1 \leq j \leq m\), we have \(e_{(i-1)n+k}^T \mathcal{D}_{i-1} e_s = 0\). Thus terms (5.61) and (5.62) are zero. Note that \(e_{(i-1)n+k}^T \mathcal{D}_{i-1} e_s = (\mathcal{D}_{i-1}(t, s), \mathcal{D}_{i-1}(t, s))\). This is zero unless \(t = s\) and \((i - 2)n + 1 \leq s \leq (i - 1)n\). Therefore term (5.63) can be simplified to

\[
\sum_{i=1}^{m} e_{(i-1)n+k}^T \frac{1}{z \bar{w}} \sum_{s \neq s + n + k} \sum_{t \neq n + k} \mathbb{E}_k \left[ G_n^{(k)}(z) c_t e_{i}^T 1_{\Omega_{n,k}} \right] \mathcal{D}_{i-1} E_k \left[ e_s c_s^* (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] e_{(i-1)n+k}
\]

\[
= \sum_{i=1}^{m} e_{(i-1)n+k}^T \frac{1}{z \bar{w}} \sum_{s=(i-2)n+1}^{(i-1)n} \sum_{s \neq s + n + k} \mathbb{E}_k \left[ G_n^{(k)}(z) c_s 1_{\Omega_{n,k}} \right] \mathbb{E}_k \left[ c_s^* (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] e_{(i-1)n+k},
\]

Ergo, we have

\[
\mathcal{T}_{n,k}(z, \bar{w})
\]

\[
= \frac{1}{z \bar{w}} \sum_{i=1}^{m} \sum_{s=(i-2)n+1}^{(i-1)n} e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k)}(z) c_s 1_{\Omega_{n,k}} \right] \mathbb{E}_k \left[ c_s^* (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] e_{(i-1)n+k}.
\]

(5.64)

We now remove the \(s\)th column from the resolvent. This produces independence between the resolvent and the column \(c_s\) in the right hand side of (5.64), which allows us to factor. In order to remove this column, we need to work on appropriate events. Since \(\Omega_{n,k,s}\) defined in (5.42)
and $Q'_{n,k,s}(z)$ defined in (5.47) hold with overwhelming probability by Corollary 5.0.7 and Lemma 5.4.17 respectively, we may insert them without effecting the limiting distribution. This is verified in Lemma 5.4.19. For ease of notation, we will drop the dependence on $z$ in $Q'_{n,k,s}(z)$. We proceed with

$$e^{T}((i - 1)n + k) \mathbb{E}_k \left[ G^{(k)}_n(z)c_1 \Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s} \right] \mathbb{E}_k \left[ c^*_s \left( G^{(k)}_n(w) \right)^* 1_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e^{(i - 1)n + k}.$$  

Then by (5.44) and (5.46), we have

$$e^{T}((i - 1)n + k) \mathbb{E}_k \left[ G^{(k)}_n(z)c_1 \Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s} \right] \mathbb{E}_k \left[ c^*_s \left( G^{(k)}_n(w) \right)^* 1_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e^{(i - 1)n + k} = e^{T}((i - 1)n + k) \mathbb{E}_k \left[ G^{(k,s)}_n(z) c_s \delta_{k,s}(z) 1_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] \times \mathbb{E}_k \left[ \delta_{k,s}(w)^* c^*_s \left( G^{(k,s)}_n(w) \right)^* 1_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e^{(i - 1)n + k}.$$ 

Since $G^{(k)}_n(z)$ 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

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

without effecting the limiting distribution. Next, we wish to replace $\delta_{k,s}(z)$ and $(\delta_{k,s}(w))^*$ with 1.
Observe that
\[
\mathbb{E}\left[ \frac{1}{n} \sum_{k=1}^{n} \frac{1}{z^k} \sum_{i=1}^{m} \sum_{s=(i-2)n+1, s \neq n+k}^{(i-1)n} e^{T(i-1)n+k} \left( \mathbb{E}_k \left[ G_n^{(k,s)}(z)c_s \delta_{k,s}(z) \mathbf{1}_{\Omega_{n,k,s} \cap Q'_n,k,s} \right] \times \mathbb{E}_k \left[ (\delta_{k,s}(w))^{(k,s)} c_s^{(k,s)}(G_n^{(k,s)}(w))^{(k,s)} \mathbf{1}_{\Omega_{n,k,s} \cap Q'_n,k,s} \right] e^{T(i-1)n+k} \right) \right]
\]
\[
\leq \max_{1 \leq k \leq n} \left\| \left( \sum_{s=(1)n+1, s \neq n+k}^{(1)n} e^{T(i-1)n+k} \left( \mathbb{E}_k \left[ G_n^{(k,s)}(z)c_s \delta_{k,s}(z) \mathbf{1}_{\Omega_{n,k,s} \cap Q'_n,k,s} \right] \times \mathbb{E}_k \left[ (\delta_{k,s}(w))^{(k,s)} c_s^{(k,s)}(G_n^{(k,s)}(w))^{(k,s)} \mathbf{1}_{\Omega_{n,k,s} \cap Q'_n,k,s} \right] e^{T(i-1)n+k} \right) \right\|^2.
\]

Therefore, it is sufficient to show that
\[
\mathbb{E}\left[ e^{T(i-1)n+k} \mathbb{E}_k \left[ G_n^{(k,s)}(z)c_s \delta_{k,s}(z) \mathbf{1}_{\Omega_{n,k,s} \cap Q'_n,k,s} \right] \times \mathbb{E}_k \left[ (\delta_{k,s}(w))^{(k,s)} c_s^{(k,s)}(G_n^{(k,s)}(w))^{(k,s)} \mathbf{1}_{\Omega_{n,k,s} \cap Q'_n,k,s} \right] e^{T(i-1)n+k} \right] = o(n^{-2}).
\]

This is done in Lemma \ref{5.4.20}. Since \( \delta_{n,k} \) is no longer present, we can justify dropping the event \( Q'_n,k,s \) by an argument similar to Lemma \ref{5.4.19}. Thus, we can continue from here working with
\[
\frac{1}{z^w} \sum_{i=1}^{m} \sum_{s=(i-2)n+1, s \neq n+k}^{(i-1)n} e^{T(i-1)n+k} \mathbb{E}_k \left[ G_n^{(k,s)}(z)c_s \mathbf{1}_{\Omega_{n,k,s}} \right] \mathbb{E}_k \left[ c_s^{(k,s)}(G_n^{(k,s)}(w))^{(k,s)} \mathbf{1}_{\Omega_{n,k,s}} \right] e^{T(i-1)n+k}.
\]
Next, for any $1 \leq i \leq m$, by independence of $c_s$ from $G^{(k,s)}_n(z)1_{\Omega_{n,k,s}}$, we can factor the above as

$$
\sum_{s=(i-2)n+1}^{(i-1)n} e^T_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k,s)}_n(z) c_s 1_{\Omega_{n,k,s}} \right] \mathbb{E}_k \left[ c_s^* (G^{(k,s)}_n(w))^* 1_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} 
= \sum_{s=(i-2)n+1}^{(i-1)n} e^T_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k,s)}_n(z) 1_{\Omega_{n,k,s}} \right] \mathbb{E}_k [c_s] 
\times \mathbb{E}_k [c^*_s] \mathbb{E}_k \left[ (G^{(k,s)}_n(w))^* 1_{\Omega_{n,k,s}} \right] e_{(i-1)n+k}.
$$

(5.65)

Observe that the value of $\mathbb{E}_k [c_s]$ depends on whether or not the column $c_s$ has been conditioned on. Consider a fixed $i$ between 1 and $m$. Since the sum in (5.65) is over values of $s$ in the $(i-1)$st block, and since the only columns that have been conditioned on are the first $k$ columns from each block, $\mathbb{E}_k [c_s] = 0$ for all $s > (i-2)n + k$. Since we exclude the $k$th columns from each block in the above sum, we can simplify (5.65) to

$$
\sum_{s=(i-2)n+1}^{(i-2)n+k-1} e^T_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k,s)}_n(z) 1_{\Omega_{n,k,s}} \right] c_s c^*_s \mathbb{E}_k \left[ (G^{(k,s)}_n(w))^* 1_{\Omega_{n,k,s}} \right] e_{(i-1)n+k}.
$$

Now, consider $c_s c^*_s$. Based on the block structure of $Y_n$, if $(i-2)n+1 \leq s < (i-2)n + k$, then we have

$$
\mathbb{E}[(c_s c^*_s)_{(a,b)}] = \begin{cases} 
\frac{1}{n} & \text{if } (i-3)n + 1 \leq a = b \leq (i-2)n \\
0 & \text{otherwise}
\end{cases}.
$$

Therefore, for a fixed term $i$ with $1 \leq i \leq m$ and since $(i-2)n+1 \leq s < (i-2)n + k$,

$$
\mathbb{E}[c_s c^*_s] = \frac{1}{n} \mathcal{D}_{i-2}.
$$

We wish now to replace $c_s c^*_s$ with its expectation. Observe that the terms

$$
e^T_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k,s)}_n(z) 1_{\Omega_{n,k,s}} \right] \left( c_s c^*_s - \frac{1}{n} \mathcal{D}_{i-2} \right) \times \mathbb{E}_k \left[ (G^{(k,s)}_n(w))^* 1_{\Omega_{n,k,s}} \right] e_{(i-1)n+k}
$$
satisfy the conditions of a martingale difference sequence in \( s \). Therefore we have

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

By Lemma 5.4.22

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

and therefore we may proceed with

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

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 5.4.19 shows that we can insert or drop these events without effecting the limiting distribution. Thus, it is sufficient to prove that

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

which is done in Lemma 5.4.23. Since

\[
\mathbb{E} \left[ \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{m} \frac{1}{z} \sum_{s=(i-2)n+1}^{(i-2)n+k-1} e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] \mathcal{D}_{i-2} \mathbb{E}_k \left[ \left( \mathcal{G}_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k}^2 \right] = o(1),
\]
the limiting distribution will be unaffected by replacing column \( c_s \). With this column replaced, in each term we now have

\[
\frac{1}{n} \sum_{s=(i-2)n+1}^{(i-2)n+k-1} e^{\frac{T}{(i-1)n+k}} E_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-2} \\
\times E_k \left[ (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] e^{\frac{T}{(i-1)n+k}} = \frac{k-1}{n} e^{\frac{T}{(i-1)n+k}} E_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-2} E_k \left[ (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] e^{\frac{T}{(i-1)n+k}} = \frac{k-1}{n} e^{\frac{T}{(i-1)n+k}} E_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-2} E_k \left[ (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] e^{\frac{T}{(i-1)n+k}}
\]

since there were \( k-1 \) terms in the above sum, and none of them depended on \( s \). Now, since this was true for every term in the sum from \( i = 1 \) up to \( i = m \) assuming that we reduce the indices modulo \( m \) with representatives \( 1, 2, \ldots m \), we have shown that

\[
E \left| \frac{1}{n} \sum_{k=1}^{n} \left( T_{n,k}(z, \overline{w}) - \frac{1}{z\overline{w}} \frac{k-1}{n} \sum_{i=1}^{m} e^{\frac{T}{(i-1)n+k}} E_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-2} \right)^2 = o(1) \right| = o(1).
\]

The goal is to iterate this process until \( T_{n,k}(z, \overline{w}) \) reappears. Lemma 5.4.24 verifies that, for any \( b \neq m \),

\[
E \left| \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{m} e^{\frac{T}{(i-1)n+k}} E_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-b} E_k \left[ (G_n^{(k)}(w))^* \mathbf{1}_{\Omega_{n,k}} \right] e^{\frac{T}{(i-1)n+k}} - \frac{1}{z\overline{w}} \frac{k-1}{n} e^{\frac{T}{(i-1)n+k}} E_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-(b+1)} \right|^2 = o(1)
\]

where \( i - b \) is reduced modulo \( m \). After iterating twice, we have

\[
E \left| \frac{1}{n} \sum_{k=1}^{n} \left( T_{n,k}(z, \overline{w}) - \frac{1}{z\overline{w}} \frac{k-1}{n} \right)^2 \sum_{i=1}^{m} e^{\frac{T}{(i-1)n+k}} E_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-3} \right|^2 = o(1).
\]
After iterating $m - 1$ times, we have

$$
\mathbb{E} \left[ \frac{1}{n} \sum_{k=1}^{n} \left( T_{n,k}(z, \overline{w}) - \left( \frac{1}{z} \frac{k-1}{n} \right)^{m-1} \sum_{i=1}^{m} e_{(i-1)n+k}^{T} \mathbb{E}_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-m} \right. \\
\times \mathbb{E}_k \left[ \left( G_n^{(k)}(w) \right)^* \mathbf{1}_{\Omega_{n,k}} \right] e_{(i-1)n+k} \right] \biggvert_{(5.59)}^{2} = o(1).
$$

To recover $T_{n,k}(z, \overline{w})$, we will iterate one final time. Due to the block structure of $D_{i-p}$, the $m$th iteration will result in less cancellation than in previous iterations. Consider the expansion due to (5.59),

$$
e_{(i-1)n+k}^{T} \mathbb{E}_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-m} \mathbb{E}_k \left[ \left( G_n^{(k)}(w) \right)^* \mathbf{1}_{\Omega_{n,k}} \right] e_{(i-1)n+k} = e_{(i-1)n+k}^{T} \mathbb{D}_{i-m} \frac{1}{z} \mathbb{E}_k \left[ G_n^{(k)}(z) c_t e_t^{T} \mathbf{1}_{\Omega_{n,k}} \right] D_{i-m} \mathbb{E}_k \left[ G_n^{(k)}(w) \right] e_{(i-1)n+k} \\
- e_{(i-1)n+k}^{T} \mathbb{E}_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-m} \mathbb{E}_k \left[ G_n^{(k)}(w) \right] e_{(i-1)n+k} \\
+ e_{(i-1)n+k}^{T} \mathbb{E}_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-m} \mathbb{E}_k \left[ G_n^{(k)}(w) \right] e_{(i-1)n+k},
$$

where the notation $t \neq s+n+k$ indicates that the sum is over all $1 \leq t \leq mn$ such that $t \neq k, n + k, \ldots, (m - 1)n + k$. Note that by the same argument as before,

$$
e_{(i-1)n+k}^{T} \mathbb{D}_{i-m} \frac{1}{z} \mathbb{E}_k \left[ G_n^{(k)}(z) \right] e_{(i-1)n+k} = (D_{i-m})_{((i-1)n+k, (i-1)n+k)} \frac{1}{z} \mathbb{E}_k \left[ G_n^{(k)}(z) \right] e_{(i-1)n+k} = 0.
$$

Now, note that $D_{i-m}$ is nonzero in diagonal entries $(i-m-1)n+1$ through $(i-m)n$. After reducing modulo $m$, we can see that this means that $D_{i-m}$ is nonzero in the $(i-1)^{st}$ block. Therefore

$$
e_{(i-1)n+k}^{T} \mathbb{D}_{i-m} \frac{1}{z} \mathbb{E}_k \left[ G_n^{(k)}(z) \right] e_{(i-1)n+k} = \frac{1}{z} \mathbb{E}_k \left[ G_n^{(k)}(z) \right] e_{(i-1)n+k} = 0.
$$

The same arguments as before show that

$$
e_{(i-1)n+k}^{T} \frac{1}{z} \sum_{s \neq s+n+k} \mathbb{D}_{i-m} \mathbb{E}_k \left[ e_s e_s^{*} G_n^{(k)}(w) \right] e_{(i-1)n+k} = 0.$$

Therefore, we have
\[ e^{T}_{(i-1)n+k} \frac{1}{z^{2w}} \sum_{t \neq n+k} \mathbb{E}_k \left[ G^{(k)}_n(z)c_t^T 1_{\Omega_{n,k}} \right] D_{i-m} \mathbb{P}_k(\Omega_{n,k}) e^{(i-1)n+k} = 0. \]

Therefore, we have
\[ e^{T}_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k)}_n(z) 1_{\Omega_{n,k}} \right] D_{i-m} \mathbb{E}_k \left[ (G^{(k)}_n(w))^* 1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \]
\[ = \frac{1}{z^w} - \frac{1}{z^w} (1 - (\mathbb{P}_k(\Omega_{n,k}))^2) \]
\[ + \frac{1}{z^w} e^{T}_{(i-1)n+k} \sum_{s \neq n+k  \text{ and } t \neq n+k} \mathbb{E}_k \left[ G^{(k)}_n(z)c_t^T 1_{\Omega_{n,k}} \right] D_{i-m} \]
\[ \times \mathbb{E}_k \left[ c_s^* (G^{(k)}_n(w))^* 1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \]
Since we can see that, by Lemma 5.4.21, \( E \left| (z \bar{w})^{-1} (1 - \mathbb{P}_k(\Omega_{n,k}))^2 \right|^2 = o_n(n^{-\alpha}) \) for any \( \alpha > 0 \), by rearranging we have
\[ \mathbb{E} \left| e^{T}_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k)}_n(z) 1_{\Omega_{n,k}} \right] D_{i-m} \mathbb{E}_k \left[ (G^{(k)}_n(w))^* 1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \right| \]
\[ = \frac{1}{z^w} - \frac{1}{z^w} e^{T}_{(i-1)n+k} \sum_{s = (i-1)n+1}^{(i-m)n} \mathbb{E}_k \left[ G^{(k)}_n(z)c_s^* 1_{\Omega_{n,k}} \right] \]
\[ \times \mathbb{E}_k \left[ c_s^* (G^{(k)}_n(w))^* 1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \right|^2 = o(n^{-\alpha}). \]
By the same argument as in the previous iterations, selecting \( \alpha \) sufficiently large, we have
\[ \mathbb{E} \left| \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{m} \left( e^{T}_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k)}_n(z) 1_{\Omega_{n,k}} \right] D_{i-m} \mathbb{E}_k \left[ (G^{(k)}_n(w))^* 1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \right) \]
\[ - \frac{1}{z^w} - \frac{k - 1}{n} \left( e^{T}_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k)}_n(z) 1_{\Omega_{n,k}} \right] D_{i-1} \right) \]
\[ \times \mathbb{E}_k \left[ (G^{(k)}_n(w))^* 1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \right|^2 = o(1) \]
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
\[ \mathbb{E} \left| \frac{1}{n} \sum_{k=1}^{n} \left( \mathcal{T}_{n,k}(z, \bar{w}) - \frac{m}{(z \bar{w})^m} \left( \frac{k - 1}{n} \right)^{m-1} - \left( \frac{1}{z \bar{w}} \right) \right) \right|^2 = o(1). \] (5.66)
The goal now is to regroup in order to compare the object of study, \( \frac{1}{n} \sum_{k=1}^{n} T_{n,k}(z, \bar{w}) \), to an appropriate Riemann sum. The sum we will compare to is

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

(5.67)

As \( n \to \infty \), (5.67) is the Riemann sum for

\[
\int_0^1 mx^{m-1} ((z \bar{w})^m - x^m)^{-1} dx
\]

which by a substitution of variables is equal to

\[-\ln ((z \bar{w})^m - 1) + \ln ((z \bar{w})^m) = -\ln \left( 1 - \frac{1}{(z \bar{w})^m} \right).\]

By regrouping the quantities inside the sum in (5.66), we can write

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

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

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

and thus

\[
T_{n,k}(z, \bar{w}) = \frac{mn(k-1)^{m-1}}{n^m(z \bar{w})^m - (k-1)^m} + \left( \frac{n^m(z \bar{w})^m}{n^m(z \bar{w})^m - (k-1)^m} \right) \cdot \mathcal{E}_{n,k}(z, \bar{w}).
\]

We are now ready to compare \( \frac{1}{n} \sum_{k=1}^{n} T_{n,k} \) to the Riemann sum in (5.67). By rearranging (5.66), we have

\[
\mathbb{E} \left| \frac{1}{n} \sum_{k=1}^{n} T_{n,k}(z \bar{w}) - \frac{1}{n} \sum_{k=1}^{n} \frac{m(k-1)^{m-1}}{n^{m-1}} \left( (z \bar{w})^m - \left( \frac{k-1}{n} \right)^m \right)^{-1} \right|^2
\]

\[\ll \mathbb{E} \left| \frac{1}{n} \sum_{k=1}^{n} \mathcal{E}_{n,k}(z, \bar{w}) \right|^2 \]

\[= o(1).\]
Therefore $\frac{1}{n} \sum_{k=1}^{n} T_{n,k}(z, \bar{w})$ converges to $-\ln \left(1 - \frac{1}{(z\bar{w})^m}\right)$ in probability as $n \to \infty$ as claimed in (5.56). By recalling that we invoked Vitali’s theorem, this implies

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

in probability as $n \to \infty$. This concludes the arguments for the convergence of terms (5.49) and (5.50). In fact, the same arguments show the convergence for terms (5.48) and (5.51) as well. This concludes the proof of Lemma 5.4.15.

By Theorem 5.4.14 this implies $\bar{M}_{n,k}(z) = \sum_{k=1}^{n} \bar{M}_{n,k}$ converges to a mean-zero Gaussian random variable with variance and covariance terms given in (5.41). Since $M_n$ and $\bar{M}_n$ converge in distribution as shown in Lemma 5.4.15, $M_n$ also converges to the mean zero Gaussian with variance and covariance terms determined in Lemma 5.4.15. Specifically,

$$\sum_{i=1}^{L} \left( \alpha_i \Xi_n(z_i) + \beta_i \bar{\Xi}_n(z_i) \right)$$

converges to a real mean-zero Gaussian with variance

$$\sum_{i,j=1}^{L} \left( \alpha_i \beta_j \frac{m^2 (z_i \bar{z}_j)^{m-1}}{((z_i \bar{z}_j)^m - 1)^2} + \beta_i \alpha_j \frac{m^2 (\bar{z}_i z_j)^{m-1}}{((\bar{z}_i z_j)^m - 1)^2} \right)$$

$$\alpha_i \alpha_j \frac{m^2 (z_i \bar{z}_j)^{m-1}}{((z_i \bar{z}_j)^m - 1)^2} + \beta_i \beta_j \frac{m^2 (\bar{z}_i z_j)^{m-1}}{((\bar{z}_i z_j)^m - 1)^2} \right).$$

From this, using the Cramer–Wold theorem, we can conclude that the finite dimensional distribution

$$(\Xi_n(z_1), \Xi_n(z_2), \ldots, \Xi_n(z_L))$$

converges to a mean-zero Gaussian vector with covariance terms

$$\mathbb{E} [\Xi(z_i) \Xi(z_j)] = \frac{m^2 (z_i \bar{z}_j)^{m-1}}{((z_i \bar{z}_j)^m - 1)^2}$$

and therefore it follows that

$$\mathbb{E} \left[ \Xi(z_i) \bar{\Xi}(z_j) \right] = \frac{m^2 (z_i \bar{z}_j)^{m-1}}{((z_i \bar{z}_j)^m - 1)^2}.$$
Lemma 5.4.19. Define all quantities as in Lemma 5.4.15. Then under the assumptions of Lemma 5.4.15, for any $\alpha > 0$, we have

\[
\mathbb{E}\left|e^{T}(i-1)n+k\mathbb{E}_{k}\left[G_{n}(k)(z)c_{s}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}_{k}\left[c_{s}^{*}(G_{n}(k)(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]e^{(i-1)n+k} - \mathbb{E}\left[c_{s}^{*}(G_{n}(k)(w))^{*}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]e^{(i-1)n+k}\right|^{2} = O_{\alpha}(n^{4-\alpha/4})
\]

uniformly in $k$.

Proof. Observe that

\[
e^{T}(i-1)n+k\mathbb{E}_{k}\left[G_{n}(k)(z)c_{s}\mathbf{1}_{\Omega_{n,k}}\right]\mathbb{E}_{k}\left[c_{s}^{*}(G_{n}(k)(w))^{*}\mathbf{1}_{\Omega_{n,k}}\right]e^{(i-1)n+k} = e^{T}(i-1)n+k\mathbb{E}_{k}\left[G_{n}(k)(z)c_{s}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]\mathbb{E}_{k}\left[c_{s}^{*}(G_{n}(k)(w))^{*}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]e^{(i-1)n+k} + e^{T}(i-1)n+k\mathbb{E}_{k}\left[G_{n}(k)(z)c_{s}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]\mathbb{E}_{k}\left[c_{s}^{*}(G_{n}(k)(w))^{*}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]e^{(i-1)n+k} \tag{5.68}
\]

\[
e^{T}(i-1)n+k\mathbb{E}_{k}\left[G_{n}(k)(z)c_{s}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]\mathbb{E}_{k}\left[c_{s}^{*}(G_{n}(k)(w))^{*}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]e^{(i-1)n+k} + e^{T}(i-1)n+k\mathbb{E}_{k}\left[G_{n}(k)(z)c_{s}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]\mathbb{E}_{k}\left[c_{s}^{*}(G_{n}(k)(w))^{*}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]e^{(i-1)n+k} \tag{5.69}
\]

\[
e^{T}(i-1)n+k\mathbb{E}_{k}\left[G_{n}(k)(z)c_{s}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]\mathbb{E}_{k}\left[c_{s}^{*}(G_{n}(k)(w))^{*}\mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}}\right]e^{(i-1)n+k} \tag{5.70}
\]

Therefore, we must show terms (5.68), (5.69), and (5.70) are sufficiently small in the $L^{2}$-norm. The argument for all three terms is very similar, and we use Jensen’s inequality and the Cauchy–Schwarz inequality to separate the inner conditional expectations. In 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 $n^{-\alpha}$ since each event holds with overwhelming probability. We show the calculation for term
and the rest of the terms follow in a similar manner. Observe

\[
\begin{align*}
&= \mathbb{E} \left| e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k)}(z) c_s \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}} \right] \mathbb{E}_k \left[ c_s^* \left( \mathcal{G}_n^{(k)}(w) \right)^* \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}} \right] e_{(i-1)n+k} \right|^2 \\
&\leq \mathbb{E} \left[ E_k \left[ \| \mathcal{G}_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}} \|^2 \right] E_k \left[ \| c_s \|^2 \left( \| \mathcal{G}_n^{(k)}(w) \|^* \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}} \right)^2 \right] \right] \\
&\leq \left( \mathbb{E} \left[ \| \mathcal{G}_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}} \|^8 \mathbb{E} \| c_s \|^8 \mathbb{E} \| c_s^* \|^8 \mathbb{E} \left( \| \mathcal{G}_n^{(k)}(w) \|^* \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k,s}} \right) \right] \right)^{1/4} \\
&\leq n^4 \left( \mathbb{E} \left[ \left( \mathcal{G}_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k}} \right)^8 \mathbf{1}_{\Omega_{n,k,s}} \right] \mathbb{E} \left[ \left( \mathcal{G}_n^{(k)}(w) \right)^* \mathbf{1}_{\Omega_{n,k}} \left\| 8 \mathbf{1}_{\Omega_{n,k,s}} \right\| \right] \right)^{1/4} \\
&\ll n^4 \mathbb{P}(\Omega_{n,k,s})^{1/4} \\
&\ll \alpha n^4 n^{-\alpha/4}.
\end{align*}
\]

The arguments for terms (5.69) and (5.70) are similar. \qed

**Lemma 5.4.20.** Define all quantities as in Lemma 5.4.15. Then, for any \( i \) with \( 1 \leq i \leq m \) where subscripts are reduced modulo \( m \), under the assumptions of Lemma 5.4.15, we have

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

uniformly in \( k \).
Proof. We will show each of these terms are \( o(1) \). By the Cauchy–Schwarz inequality, Lemma 5.4.7 and Lemma 5.4.18, we have

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

\[
\times e_{(i-1)n+k} \mathbb{E}_k \left[ \left( \delta_{k,s}(w) \right)^* c_s^* \left( \mathcal{G}_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right] e^{(i-1)n+k} \]

\[
\leq \mathbb{E} \left[ e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k,s)}(z)c_s \delta_{k,s}(z) \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right] \times \mathbb{E}_k \left[ \left( \delta_{k,s}(w) \right)^* c_s^* \left( \mathcal{G}_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right] e^{(i-1)n+k} \right]^{1/2}
\]

\[
\leq \mathbb{E} \left[ \left( \delta_{k,s}(w) \right)^* c_s^* \left( \mathcal{G}_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right]^{1/2}
\]

We will show each of these terms are \( o(n^2) \). We begin with \( (5.71) \). By the Cauchy–Schwarz inequality, Lemma 5.4.7 and Lemma 5.4.18, we have

\[
\mathbb{E} \left[ e_{(i-1)n+k}^T \mathbb{E}_k \left[ \mathcal{G}_n^{(k,s)}(z)c_s \delta_{k,s}(z) \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right] \times \mathbb{E}_k \left[ \left( \delta_{k,s}(w) \right)^* c_s^* \left( \mathcal{G}_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right] e^{(i-1)n+k} \right]^{1/2}
\]

\[
\leq \mathbb{E} \left[ \left( \delta_{k,s}(w) \right)^* c_s^* \left( \mathcal{G}_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right]^{1/2}
\]

\[
\leq \mathbb{E} \left[ \left( \delta_{k,s}(w) \right)^* c_s^* \left( \mathcal{G}_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right]^{1/2}
\]

\[
\leq n^{-1} \left( n^{-2-4} \mathbb{E} \left[ \left( \delta_{k,s}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right] \right)^{1/4}
\]

\[
\leq n^{-3/2-\varepsilon} \mathbb{E} \left[ \left( \delta_{k,s}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s} \cap \mathcal{Q}_{n,k,s}'} \right]^{1/4}
\]

(5.73)
Now, recall $\delta_{k,s}(z)$ defined in (5.45). By the resolvent identity (3.6), we have

$$1 - \delta_{k,s}(z) = 1 - \left(1 + e_s^T G_n^{(k,s)}(z) c_s\right)^{-1} = \left(e_s^T G_n^{(k,s)}(z) c_s\right) \delta_{k,s}(z).$$

Thus

$$\delta_{k,s}(z) = 1 - \left(e_s^T G_n^{(k,s)}(z) c_s\right) \delta_{k,s}(z)$$

and by iterating this equality we have

$$\delta_{k,s}(z) = 1 - (e_s^T G_n^{(k,s)}(z) c_s) \delta_{k,s}(z)$$

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

and thus

$$((\delta_{k,s}(w))^* - 1) = - (e_s^T G_n^{(k,s)}(w) c_s)^* + (e_s^T G_n^{(k,s)}(w) c_s)^2 (\delta_{k,s}(w))^*.$$  (5.74)

We replace $((\delta_{k,s}(w))^* - 1)$ in (5.73) with the expression on the right hand side of (5.74) and use Lemma 5.4.7 to see

$$n^{-3/2-\varepsilon} \left( \mathbb{E} \left[\left((\delta_{k,s}(w))^* - 1\right)1_{\Omega_{n,k,s} \cap Q_{n,k,s}^c} \right] \right)^{1/4}$$

$$\ll n^{-3/2-\varepsilon} \left( \mathbb{E} \left[\left(- (e_s^T G_n^{(k,s)}(w) c_s)^* + (\delta_{k,s}(w))^* (e_s^T G_n^{(k,s)}(w) c_s)^2\right)1_{\Omega_{n,k,s} \cap Q_{n,k,s}^c} \right] \right)^{1/4}$$

$$\ll n^{-3/2-\varepsilon} \left( \mathbb{E} \left[|c_s^* (G_n^{(k,s)}(w))^* e_s 1_{\Omega_{n,k,s} \cap Q_{n,k,s}^c}|^8 \right] + \mathbb{E} \left[|c_s^* (G_n^{(k,s)}(w))^* (e_s^T G_n^{(k,s)}(w) c_s)^2 1_{\Omega_{n,k,s} \cap Q_{n,k,s}^c}|^8 \right] \right)^{1/4}$$

$$\ll n^{-3/2-\varepsilon} \left( \mathbb{E} \left[|c_s^* (G_n^{(k,s)}(w))^* e_s e_s^T G_n^{(k,s)}(w) c_s 1_{\Omega_{n,k,s} \cap Q_{n,k,s}^c}|^4 \right] + \mathbb{E} \left[|c_s^* (G_n^{(k,s)}(w))^* e_s e_s^T G_n^{(k,s)}(w) c_s 1_{\Omega_{n,k,s} \cap Q_{n,k,s}^c}|^8 \right] \right)^{1/4}$$

$$\ll n^{-3/2-\varepsilon} \left( n^{-4\varepsilon-2} \mathbb{E} \left[|G_n^{(k,s)}(w)|^4 e_s e_s^T G_n^{(k,s)}(w) 1_{\Omega_{n,k,s} \cap Q_{n,k,s}^c} \right]^4 \right)^{1/4}$$

$$+ n^{-12\varepsilon-2} \mathbb{E} \left[|G_n^{(k,s)}(w)|^8 e_s e_s^T G_n^{(k,s)}(w) 1_{\Omega_{n,k,s} \cap Q_{n,k,s}^c} \right]^8 \right)^{1/4}$$

$$\ll n^{-3/2-\varepsilon} n^{-1/2-\varepsilon}$$

$$= o(n^{-2}).$$
This shows term (5.71) is $o(n^{-2})$. A very similar argument shows that term (5.72) is $o(n^{-2})$ as well. We omit the details. This completes the proof.

**Lemma 5.4.21.** Define all quantities as in Lemma 5.4.15. Then under the assumptions of Lemma 5.4.15
\[
\mathbb{E} \left| \frac{1}{zw} (1 - P_k(\Omega_{n,k})^2) \right|^2 = o(n^{-\alpha})
\]
for any $\alpha > 0$ and uniformly in $k$.

**Proof.** Observe that since $z, \bar{w} \in C$ are fixed with $|z| = |\bar{w}| = 1 + \delta$, we know that
\[
\left| \frac{1}{zw} (1 - P_k(\Omega_{n,k})^2) \right|^2 \ll |1 - P_k(\Omega_{n,k})^2| \ll |1 - P_k(\Omega_{n,k})|^2.
\]
Since $\Omega_{n,k}$ holds with overwhelming probability by Corollary F.0.6
\[
\mathbb{E} |1 - P_k(\Omega_{n,k})|^2 = \mathbb{E} |P_k(\Omega_{n,k})|^2 \leq \mathbb{P}(\Omega_{n,k}^c) = o(n^{-\alpha})
\]
for any $\alpha > 0$.

**Lemma 5.4.22.** Define all quantities as in Lemma 5.4.15. Then under the assumptions of Lemma 5.4.15
\[
\mathbb{E} \left| e^T_{(i-1)n+k} e_{(i-1)n+k} \right|^2 = o(n^{-1})
\]
uniformly in $k$.

**Proof.** To begin, observe that we can rewrite
\[
e^T_{(i-1)n+k} e_{(i-1)n+k} = \frac{1}{n} \left( \sqrt{n} c_s^T \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} \right).
\]
Consider a $mn \times 1$ vector $v_{i-2}$ defined

$$(v_{i-2})_a = \begin{cases} 
1 & \text{if } (i-3)n + 1 \leq a \leq (i-2)n \\
0 & \text{otherwise}
\end{cases}.$$ 

Then observe by cyclic permutation of the trace, we have

$$\frac{1}{n} e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k}$$

$$= \frac{1}{n} \text{tr} \left( \frac{1}{n} e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} \right)^2$$

$$= \frac{1}{n} \text{tr} \left( D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] v_{i-2} \right)^2$$

By this observation and Lemma [D.0.2] we have

$$\mathbb{E} \left[ e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] c_s e_k^* \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} \right]^2$$

$$= \frac{1}{n^2} \mathbb{E} \left[ \sqrt{n} c_s^* \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] \sqrt{n} c_s \right]^2$$

$$\leq \frac{1}{n^2} \mathbb{E} \left[ \xi_1 \right]^4 \mathbb{E} \left[ \text{tr} \left( D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] \right)^2 \right]$$

$$\leq \frac{1}{n^2} \mathbb{E} \left[ \text{tr} \left( D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] v_{i-2} \right)^2 \right]$$

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 (5.75) by $O(n^{-2})$ which completes the proof.

**Lemma 5.4.23.** Define all quantities as in Lemma [5.4.15]. Then under the assumptions of Lemma
5.4.15 we have

\[
\begin{align*}
\mathbb{E} \left[ \frac{1}{n} \sum_{s=(i-2)n+1}^{(i-2)n+k-1} e^{T} (i-1)n+k \mathbb{E}_k \left[ G_n^{(k,s)}(z)1_{\Omega_{n,k,s}} \right] D_i - 2 \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^*1_{\Omega_{n,k,s}} \right] e^{(i-1)n+k} \right] \\
- \frac{1}{n} \sum_{s=(i-2)n+1}^{(i-2)n+k-1} e^{T} (i-1)n+k \mathbb{E}_k \left[ G_n^{(k,s)}(z)1_{\Omega_{n,k}} \right] D_i - 2 \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^*1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \right] ^{2} = o(1)
\end{align*}
\]

uniformly in \( k \).

Proof. To begin, observe that we can write

\[
\begin{align*}
\mathbb{E} \left[ \frac{1}{n} \sum_{s=(i-2)n+1}^{(i-2)n+k-1} e^{T} (i-1)n+k \mathbb{E}_k \left[ G_n^{(k,s)}(z)1_{\Omega_{n,k,s}} \right] D_i - 2 \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^*1_{\Omega_{n,k,s}} \right] e^{(i-1)n+k} \right] \\
- \frac{1}{n} \sum_{s=(i-2)n+1}^{(i-2)n+k-1} e^{T} (i-1)n+k \mathbb{E}_k \left[ G_n^{(k,s)}(z)1_{\Omega_{n,k}} \right] D_i - 2 \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^*1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \right] ^{2} \leq \frac{1}{n} \sum_{s=(i-2)n+1}^{(i-2)n+k-1} \mathbb{E} [ e^{T} (i-1)n+k G_n^{(k,s)}(z)1_{\Omega_{n,k,s}} ] D_i - 2 \mathbb{E}_k [ (G_n^{(k,s)}(w))^* e^{(i-1)n+k} 1_{\Omega_{n,k,s}} ] \\
- \mathbb{E}_k [ e^{T} (i-1)n+k G_n^{(k,s)}(z)1_{\Omega_{n,k}} ] D_i - 2 \mathbb{E}_k [ (G_n^{(k,s)}(w))^* e^{(i-1)n+k} 1_{\Omega_{n,k}} ] ^{2}
\end{align*}
\]

so it suffices to prove that

\[
\begin{align*}
\mathbb{E} \left[ e^{T} (i-1)n+k G_n^{(k,s)}(z)1_{\Omega_{n,k,s}} ] D_i - 2 \mathbb{E}_k [ (G_n^{(k,s)}(w))^* e^{(i-1)n+k} 1_{\Omega_{n,k,s}} ] \\
- \mathbb{E}_k [ e^{T} (i-1)n+k G_n^{(k,s)}(z)1_{\Omega_{n,k}} ] D_i - 2 \mathbb{E}_k [ (G_n^{(k,s)}(w))^* e^{(i-1)n+k} 1_{\Omega_{n,k}} ] ^{2} = o(1)
\end{align*}
\]
uniformly in $k$ and $s$. To this end, we can expand this term as

\[
\mathbb{E} \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right] 
\]

\[
\ll \mathbb{E} \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right] 
\]

\[\leq \mathbb{E} \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right]^2 \quad (5.76)\]

\[\leq \mathbb{E} \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right]^2 \quad (5.77)\]

\[\leq \mathbb{E} \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right]^2 \quad (5.78)\]

\[\leq \mathbb{E} \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right]^2 \quad (5.79)\]

\[\leq \mathbb{E} \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right]^2 \quad (5.80)\]

\[\leq \mathbb{E} \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right]^2 \quad (5.81)\]

We will bound each of these terms separately. We begin with (5.76). Observe that by the Cauchy–
Schwarz inequality and Lemma 5.4.18 for any $\alpha > 0$,

$$
\mathbb{E} \left| e_{(i-1)n+k}^T (z)^{\Omega_{n,k,s}} \right| \mathcal{D}_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k}^{\Omega_{n,k,s}} \right] \\
- \mathbb{E}_k \left[ e_{(i-1)n+k}^T G_n^{(k,s)}(z)^{\Omega_{n,k,s}} \right] \mathcal{D}_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k}^{\Omega_{n,k,s}\cap\Omega_{n,k}} \right]
$$

$$\ll \mathbb{E} \left\| G_n^{(k,s)}(z)^{\Omega_{n,k,s}} \right\| \mathcal{D}_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* 1_{\Omega_{n,k,s}} \right] \\
- \mathbb{E}_k \left[ G_n^{(k,s)}(z)^{\Omega_{n,k,s}} \right] \mathcal{D}_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* 1_{\Omega_{n,k,s}\cap\Omega_{n,k}} \right]
$$

$$= \mathbb{E} \left\| G_n^{(k,s)}(z)^{\Omega_{n,k,s}} \right\|^4 \mathcal{D}_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* (1_{\Omega_{n,k,s}} 1_{\Omega_{n,k}}) \right]^2
$$

$$\leq \left( \mathbb{E} \left[ \left\| G_n^{(k,s)}(z)^{\Omega_{n,k,s}} \right\|^4 \right]\mathcal{D}_{i-2} \mathbb{E}_k \left[ (G_n^{(k,s)}(w))^* 1_{\Omega_{n,k,s}} \right] \right)^{1/2} \mathbb{E} \left[ \left\| G_n^{(k,s)}(w)^* 1_{\Omega_{n,k,s}} \right\|^{\frac{4}{1}} \mathcal{D}_{i-2} \mathbb{E}_k \left[ 1_{\Omega_{n,k}} \right] \right]^{1/2}
$$

$$\ll \alpha n^{-\alpha/2}.$$

Selecting $\alpha$ sufficiently large shows that (5.76) is $o(1)$. Very similar arguments show that (5.78), (5.79), and (5.81) are all $o(1)$. This leaves us with only the two terms which differ by resolvents.
We will first deal with (5.77). Observe that by the resolvent identity (3.6) and Lemma 5.4.7,

\begin{align*}
  \mathbb{E} \left| e^{T (i-1)n+k} G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right| D_{i-2} \mathbb{E} \left[ (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s} \cap \Omega_{n,k}} \right] \\
  - \mathbb{E} \left[ e^{T (i-1)n+k} G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \mathbb{E} \left[ (G_n^{(k)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s} \cap \Omega_{n,k}} \right]^2 \\
  = \mathbb{E} \left[ e^{T (i-1)n+k} G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right] D_{i-2} \\
  \times \mathbb{E} \left[ (G_n^{(k,s)}(w)(\gamma_n^{(k)} - \gamma_n^{(k,s)}) G_n^{(k)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s} \cap \Omega_{n,k}} \right]^2 \\
  \leq \left( \mathbb{E} \left| e^{T (i-1)n+k} G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right| D_{i-2} \right)^4 \\
  \times \mathbb{E} \left[ \left| (G_n^{(k,s)}(w)(c_s e_s^T G_n^{(k)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s} \cap \Omega_{n,k}} \right| \right]^4 \\
  \leq \left( \mathbb{E} \left| e_{(i-1)n+k} \right|^4 \left| G_n^{(k,s)}(z) 1_{\Omega_{n,k,s}} \right|^4 \left| D_{i-2} \right|^4 \right) \\
  \times \mathbb{E} \left[ \left| (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right|^4 \left| e_s \right|^4 \left| c_s^* (G_n^{(k,s)}(w))^* e_{(i-1)n+k} 1_{\Omega_{n,k,s}} \right|^4 \right]^{1/2} \\
  \ll \left( \mathbb{E} \left| c_s^* (G_n^{(k,s)}(w))^* e_{(i-1)n+k} e^{T (i-1)n+k} G_n^{(k,s)}(w) c_s 1_{\Omega_{n,k,s}} \right|^2 \right)^{1/2} \\
  \ll \left( n^{-2} \mathbb{E} \left| (G_n^{(k,s)}(w))^* e_{(i-1)n+k} e^{T (i-1)n+k} G_n^{(k,s)}(w) 1_{\Omega_{n,k,s}} \right|^2 \right)^{1/2} \\
  \ll n^{-1}.
\end{align*}

A very similar argument also shows that (5.80) is $o(1)$ as well, concluding the proof.

\textbf{Lemma 5.4.24.} Define all quantities as in Lemma 5.4.15. Let $b$ be fixed with $1 \leq b < m$, and reduce $i - b$ modulo $m$. Then, under the assumptions of Lemma 5.4.15

\begin{align*}
  \mathbb{E} \left| \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{m} \left( e^{T (i-1)n+k} G_n^{(k)}(z) 1_{\Omega_{n,k}} \right) D_{i-b} \mathbb{E} \left[ (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] e_{(i-1)n+k} \\
  - \frac{1}{n} \frac{k-1}{n} e^{T (i-1)n+k} G_n^{(k)}(z) 1_{\Omega_{n,k}} \right] D_{i-(b+1)} \\
  \times \mathbb{E} \left[ (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] e_{(i-1)n+k} \right|^2 = o(1)
\end{align*}

uniformly in $k$. 

Proof. This proof follows by the same arguments as the first expansion in the proof of Lemma \ref{lem:5.4.15}. However, now the index on $D_{i-b}$ is arbitrary, assuming $b \neq m$. We leave out details which are identical to the arguments in Lemma \ref{lem:5.4.15}. Begin by observing that provided the resolvent is defined, we can expand using the resolvent expansion (5.59) to get

$$e^T_{(i-1)n+k}E_k \left[ G_n^{(k)}(z) 1_{\Omega_{n,k}} \right] \bar{D}_{i-b} E_k \left[ (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] e_{(i-1)n+k}$$

$$= e^T_{(i-1)n+k} \left( \frac{1}{z-w} (P_k(\Omega_{n,k}))^2 e_{(i-1)n+k} \right) \bar{D}_{i-b} E_k \left[ e_s e^*_s (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] P_k(\Omega_{n,k}) e_{(i-1)n+k} \quad (5.82)$$

$$- e^T_{(i-1)n+k} \frac{1}{z-w} \sum_{s \neq n+k} \bar{D}_{i-b} E_k \left[ e_s e^*_s (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] P_k(\Omega_{n,k}) e_{(i-1)n+k} \quad (5.83)$$

$$- e^T_{(i-1)n+k} \frac{1}{z-w} \sum_{t \neq n+k} E_k \left[ G_n^{(k)}(z) c_t e^*_t 1_{\Omega_{n,k}} \right] \bar{D}_{i-b} P_k(\Omega_{n,k}) e_{(i-1)n+k} \quad (5.84)$$

$$+ e^T_{(i-1)n+k} \frac{1}{z-w} \sum_{s \neq n+k} \sum_{t \neq n+k} E_k \left[ G_n^{(k)}(z) c_t e^*_t 1_{\Omega_{n,k}} \right] \bar{D}_{i-b}$$

$$\times E_k \left[ e_s e^*_s (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] e_{(i-1)n+k} \quad (5.85)$$

where the notation $t \neq n+k$ indicates that the sum is over all $1 \leq t \leq mn$ such that $t \neq k, n+k, \ldots, (m-1)n+k$. We will handle each term individually. For term (5.82), observe that

$$e^T_{(i-1)n+k} \left( \frac{1}{z-w} (P_k(\Omega_{n,k}))^2 e_{(i-1)n+k} \right) = 0$$

since the only nonzero elements in $D_{i-b}$ are in block $i-b$, and the above product selects an element from the $i$th block. This is zero unless $b = m$ since we are reducing mod $m$.

For terms (5.83) and (5.84), note that since any off diagonal element of $D_{i-b}$ will be zero,

$$e^T_{(i-1)n+k} \frac{1}{z-w} \sum_{s \neq n+k} \bar{D}_{i-b} E_k \left[ e_s e^*_s (G_n^{(k)}(w))^* 1_{\Omega_{n,k}} \right] P_k(\Omega_{n,k}) e_{(i-1)n+k} = 0$$

and

$$e^T_{(i-1)n+k} \frac{1}{z-w} \sum_{t \neq n+k} E_k \left[ G_n^{(k)}(z) c_t e^*_t 1_{\Omega_{n,k}} \right] \bar{D}_{i-b} P_k(\Omega_{n,k}) e_{(i-1)n+k} = 0.$$
Therefore,

\[
\begin{align*}
& e^{T_{i-1}}_{(i-1)n+k} \mathbb{E}_k \left[ G^{(k)}_n(z) 1_{\Omega_{n,k}} \right] D_{i-1} \mathbb{E}_k \left[ (G^{(k)}_n(w))^* 1_{\Omega_{n,k}} \right] e^{(i-1)n+k} \\
& = e^{T_{i-1}}_{(i-1)n+k} \frac{1}{z^W} \sum_{s=(i-b-1)n+1}^{(i-b)n} \mathbb{E}_k \left[ G^{(k)}_n(z) c_s 1_{\Omega_{n,k}} \right] \\
& \quad \times \mathbb{E}_k \left[ c_s^* (G^{(k)}_n(w))^* 1_{\Omega_{n,k}} \right] e^{(i-1)n+k}.
\end{align*}
\]

Again, from here we wish to remove column \( c_s \) from \( G^{(k)}_n(z) \) and \( (G^{(k)}_n(w))^* \). By arguments similar to Lemma 5.4.19, we may insert events \( \Omega_{n,k,s} \) and \( Q'_{n,k,s} \) since they both hold with overwhelming probability by Corollary F.0.7 and Lemma 5.4.17. By equations (5.44) and (5.46), we can write

\[
\begin{align*}
& e^{T_{i-1}}_{(i-1)n+k} \frac{1}{z^W} \sum_{s=(i-b-1)n+1}^{(i-b)n} \mathbb{E}_k \left[ G^{(k)}_n(z) c_s 1_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] \\
& \quad \times \mathbb{E}_k \left[ c_s^* (G^{(k)}_n(w))^* 1_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e^{(i-1)n+k}.
\end{align*}
\]

By Lemma 5.4.20, we can replace \( \delta_{k,s}(z) \) and \( (\delta_{k,s}(w))^* \) with 1 with a sufficiently small error, so we can proceed with

\[
\begin{align*}
& e^{T_{i-1}}_{(i-1)n+k} \frac{1}{z^W} \sum_{s=(i-b-1)n+1}^{(i-b)n} \mathbb{E}_k \left[ G^{(k,s)}_n(z) c_s 1_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] \\
& \quad \times \mathbb{E}_k \left[ c_s^* (G^{(k,s)}_n(w))^* 1_{\Omega_{n,k} \cap \Omega_{n,k,s} \cap Q'_{n,k,s}} \right] e^{(i-1)n+k}.
\end{align*}
\]

We can now drop the events \( \Omega_{n,k} \) and \( Q'_{n,k,s} \) since they both hold with overwhelming probability.
Now, by independence we have

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

and since the terms in the above sum are a martingale difference sequence in $s$, by Lemma 5.4.22 we can replace column $c_s$ and we have

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

$$\times \mathbb{E}_k \left[ e_{(i-1)n+k}^T \left( G_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k}$$

and since the terms in the above sum are a martingale difference sequence in $s$, by Lemma 5.4.22 we have

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

$$\times \mathbb{E}_k \left[ \left( G_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k}$$

$$- \frac{1}{n} e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k,s}} \right] D_{i-(b+1)}$$

$$\times \mathbb{E}_k \left[ \left( G_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k,s}} \right] e_{(i-1)n+k} \right) = o(1).$$

We may reinsert the event $\Omega_{n,k}$ since it holds with overwhelming probability, and thus by Lemma 5.4.23 we can replace column $c_s$ and we have

$$\mathbb{E} \left[ \frac{1}{n^2} \frac{1}{z \bar{w}} \sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{s=(i-b-1)n+1}^{(i-b)n} \left( e_{(i-1)n+k}^T \right) \mathbb{E}_k \left[ G_n^{(k,s)}(z) \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k}} \right] D_{i-(b+1)}$$

$$\times \mathbb{E}_k \left[ \left( G_n^{(k,s)}(w) \right)^* \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k}} \right] e_{(i-1)n+k}$$

$$- e_{(i-1)n+k}^T \mathbb{E}_k \left[ G_n^{(k)}(z) \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k}} \right] D_{i-(b+1)}$$

$$\times \mathbb{E}_k \left[ \left( G_n^{(k)}(w) \right)^* \mathbf{1}_{\Omega_{n,k} \cap \Omega_{n,k}} \right] e_{(i-1)n+k} \right) ^2 = o(1).$$
Dropping the event $\Omega_{n,k,s}$ is justified since it holds with overwhelming probability, and after doing so, we see that the terms

$$e_{(i-1)n+k}^T \mathbb{E}_k \left[ G^{(k)}_n(z) \mathbf{1}_{\Omega_{n,k}} \right] D_{i-(b+1)} \mathbb{E}_k \left[ (G^{(k)}_n(w))^T \mathbf{1}_{\Omega_{n,k}} \right] e_{(i-1)n+k}$$

no longer depend on $s$. Summing over $s$ results in a factor of $k - 1$, which concludes the proof of the lemma.

5.5 Tightness

In order to extend the finite dimensional convergence proved in Section 5.4 to convergence of the stochastic process $\{\Xi_n(z)\}_{z \in \mathbb{C}}$, we must check that the sequence of stochastic processes $\{\Xi_n(z)\}_{z \in \mathbb{C}}$ is tight. Namely, recall that we must verify condition (5.19) in Theorem 5.3.7. To check this condition, it will be helpful to recenter $\Xi_n(z)$.

Define the modified sequence

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

which differs from $\Xi_n(z)$ by the fact that we have subtracted the trace of $G^{(k)}_n(z)$ 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 5.5.1. Let $\Xi_n(z)$ be as defined in (5.14) and $\tilde{\Xi}_n(z)$ as in (5.86). Then under the assumptions of Theorem 5.2.11

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

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

Proof. We can see that

$$\mathbb{E} \left| \frac{\Xi_n(z) - \Xi_n(w)}{z - w} \right|^2 \ll \mathbb{E} \left| \frac{\Xi_n(z) - \Xi_n(w)}{z - w} - \frac{\tilde{\Xi}_n(z) - \tilde{\Xi}_n(w)}{z - w} \right|^2 + \mathbb{E} \left| \frac{\tilde{\Xi}_n(z) - \tilde{\Xi}_n(w)}{z - w} \right|^2.$$
Now note that by the resolvent identity \([3.6]\)

\[
\Xi_n(z) - \Xi_n(w) = \sum_{k=1}^{n} (E_k - E_{k-1}) [\text{tr}(G_n(z)(w - z)G_n(w))1_{\Omega_n}]
\]

and

\[
\tilde{\Xi}_n(z) - \tilde{\Xi}_n(w) = \sum_{k=1}^{n} (E_k - E_{k-1}) [\text{tr}(G_n(z)(w - z)G_n(w))1_{\Omega_n \cap \Omega_{n,k}}
\]

\[- \text{tr}(G_n^{(k)}(z)(w - z)G_n^{(k)}(w))1_{\Omega_n \cap \Omega_{n,k}}].
\]

Therefore, by cyclic permutation of the trace and Lemma \([D.0.8]\)

\[
E \left| \frac{\Xi_n(z) - \Xi_n(w)}{z - w} - \frac{\tilde{\Xi}_n(z) - \tilde{\Xi}_n(w)}{z - w} \right|^2
\]

\[= E \left| \sum_{k=1}^{n} ((E_k - E_{k-1}) [\text{tr}(G_n(w)G_n(z))(1_{\Omega_n} - 1_{\Omega_n \cap \Omega_{n,k}})]
\]

\[+ (E_k - E_{k-1}) [\text{tr}(G_n^{(k)}(z)G_n^{(k)}(w))1_{\Omega_n \cap \Omega_{n,k}}]) \right|^2
\]

\[\ll \sum_{k=1}^{n} \left( E \left| \text{tr}(G_n(w)G_n(z))1_{\Omega_n} 1_{\Omega_{n,k}} \right|^2
\]

\[+ E \left| (E_k - E_{k-1}) [\text{tr}(G_n^{(k)}(z)G_n^{(k)}(w))1_{\Omega_n \cap \Omega_{n,k}}]\right|^2 \right)
\]

\[\ll \alpha n^{3-\alpha} + \sum_{k=1}^{n} E \left| (E_k - E_{k-1}) [\text{tr}(G_n^{(k)}(z)G_n^{(k)}(w))1_{\Omega_n \cap \Omega_{n,k}}]\right|^2
\]

for any \(\alpha > 0\). Since \((E_k - E_{k-1}) [\text{tr}(G_n^{(k)}(z)G_n^{(k)}(w))1_{\Omega_{n,k}}] = 0\), we have

\[
\sum_{k=1}^{n} E \left| (E_k - E_{k-1}) [\text{tr}(G_n^{(k)}(z)G_n^{(k)}(w))1_{\Omega_n \cap \Omega_{n,k}}]\right|^2
\]

\[\ll \sum_{k=1}^{n} \left( E \left| (E_k - E_{k-1}) [\text{tr}(G_n^{(k)}(z)G_n^{(k)}(w))1_{\Omega_{n,k}}]\right|^2
\]

\[+ E \left| (E_k - E_{k-1}) [\text{tr}(G_n^{(k)}(z)G_n^{(k)}(w))1_{\Omega_n \cap \Omega_{n,k}}]\right|^2 \right)
\]

\[\ll \alpha n^{3-\alpha}
\]
for any $\alpha > 0$. Ergo, we have
\[
\mathbb{E} \left| \frac{\Xi_n(z) - \Xi_n(w)}{z - w} - \frac{\tilde{\Xi}_n(z) - \tilde{\Xi}_n(w)}{z - w} \right|^2 \ll_\alpha n^{3-\alpha}
\]
for any $\alpha > 0$. 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.

\textbf{Lemma 5.5.2.} Let $\{\tilde{\Xi}_n(z)\}$ be the sequence of stochastic processes defined in (5.86). 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 $\mathcal{C}$.

\textit{Proof.} The idea behind this proof is similar to what was done in the proof of Lemma 5.4.15 where we remove columns to achieve independence. First, observe that by definition of $\tilde{\Xi}_n(z)$, linearity of trace, and the resolvent identity (3.6),
\[
\frac{\tilde{\Xi}_n(z) - \tilde{\Xi}_n(w)}{z - w} = \sum_{k=1}^n (E_k - E_{k-1}) \left[ \text{tr}(\mathcal{G}_n(z) - \mathcal{G}_n(w) - (\mathcal{G}_n^{(k)}(z) - \mathcal{G}_n^{(k)}(w))) \mathbf{1}_{\Omega_n \cap \Omega_{n,k}} \right] \left( z - w \right)
\]
\[
= - \sum_{k=1}^n (E_k - E_{k-1}) \left[ \text{tr}(\mathcal{G}_n(z)\mathcal{G}_n(w) - \mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)) \mathbf{1}_{\Omega_n \cap \Omega_{n,k}} \right].
\] (5.87)

Now note that
\[
(\mathcal{G}_n(z) - \mathcal{G}_n^{(k)}(z))(\mathcal{G}_n(w) - \mathcal{G}_n^{(k)}(w))
\]
\[
= \mathcal{G}_n(z)\mathcal{G}_n(w) - \mathcal{G}_n(z)\mathcal{G}_n^{(k)}(w) - \mathcal{G}_n^{(k)}(z)\mathcal{G}_n(w) + \mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)
\]
which implies
\[
\mathcal{G}_n(z)\mathcal{G}_n(w) + \mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)
\]
\[
= (\mathcal{G}_n(z) - \mathcal{G}_n^{(k)}(z))(\mathcal{G}_n(w) - \mathcal{G}_n^{(k)}(w)) + \mathcal{G}_n(z)\mathcal{G}_n^{(k)}(w) + \mathcal{G}_n^{(k)}(z)\mathcal{G}_n(w).
\]
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 (3.6), we have

$$\mathcal{G}_n(z)\mathcal{G}_n(w) - \mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)$$

$$= (\mathcal{G}_n(z)(U_k V_k^T)\mathcal{G}_n^{(k)}(z)) (\mathcal{G}_n(w)(U_k V_k^T)\mathcal{G}_n^{(k)}(w))$$

$$+ (\mathcal{G}_n(z)(U_k V_k^T)\mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)$$

$$+ \mathcal{G}_n^{(k)}(z)(\mathcal{G}_n(w)(U_k V_k^T)\mathcal{G}_n^{(k)}(w))$$

where $U_k$ is again the $mn \times m$ matrix containing columns $c_k, c_{n+k}, \ldots, c_{(m-1)n+k}$ and $V_k$ is the $mn \times m$ matrix containing columns $e_k, e_{n+k}, \ldots, e_{(m-1)n+k}$. By the Sherman–Morrison–Woodbury formula (3.5), 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 5.4, we can guarantee that this matrix is invertible by working on the event $Q_{n,k}$ defined in (5.30). Since $Q_{n,k}$ holds with overwhelming probability by Lemma 5.6.6, the same argument as in Section 5.4 shows that we can work on this event with error $o(n^{-\alpha})$ 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

$$\mathcal{G}_n(z)\mathcal{G}_n(w) - \mathcal{G}_n^{(k)}(z)\mathcal{G}_n^{(k)}(w)$$

$$= (\mathcal{G}_n^{(k)}(z)U_k \Delta_{n,k}(z)V_k^T \mathcal{G}_n^{(k)}(z)) (\mathcal{G}_n^{(k)}(w)U_k \Delta_{n,k}(w)V_k^T \mathcal{G}_n^{(k)}(w))$$

$$+ (\mathcal{G}_n^{(k)}(z)U_k \Delta_{n,k}(z)V_k^T \mathcal{G}_n^{(k)}(z))\mathcal{G}_n^{(k)}(w)$$

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

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

\[
\begin{align*}
\text{tr}(G_n(z)G_n(w) - G_n^{(k)}(z)G_n^{(k)}(w)) & = \text{tr} \left( (G_n^{(k)}(z)U_k \Delta_{n,k}(z)V_k^T G_n^{(k)}(z)) (G_n^{(k)}(w)U_k \Delta_{n,k}(w)V_k^T G_n^{(k)}(w)) \right) \\
 & + \text{tr} \left( (G_n^{(k)}(z)U_k \Delta_{n,k}(z)V_k^T G_n^{(k)}(z)) G_n^{(k)}(w) \right) \\
 & + \text{tr} \left( G_n^{(k)}(z)(G_n^{(k)}(w)U_k \Delta_{n,k}(w)V_k^T G_n^{(k)}(w)) \right).
\end{align*}
\]

Putting all of these observations together, we have shown that

\[
\begin{align*}
E \left| \hat{\Xi}_n(z) - \hat{\Xi}_n(w) \right|^2 & \leq \sum_{k=1}^n E \left| (E_k - E_{k-1}) \text{tr}(G_n(z)G_n(w) - G_n^{(k)}(z)G_n^{(k)}(w)) 1_{\Omega_n \cap \Omega_{n,k}} \right|^2 \\
& \leq \sum_{k=1}^n E \left| \text{tr} \left( (G_n^{(k)}(z)U_k \Delta_{n,k}(z)V_k^T G_n^{(k)}(z)) \times (G_n^{(k)}(w)U_k \Delta_{n,k}(w)V_k^T G_n^{(k)}(w)) \right) 1_{\Omega_{n,k}} \right|^2 \\
& + \sum_{k=1}^n E \left| \text{tr} \left( (G_n^{(k)}(z)U_k \Delta_{n,k}(z)V_k^T G_n^{(k)}(z)) G_n^{(k)}(w) \right) 1_{\Omega_{n,k}} \right|^2 \\
& + \sum_{k=1}^n E \left| \text{tr} \left( G_n^{(k)}(z)(G_n^{(k)}(w)U_k \Delta_{n,k}(w)V_k^T G_n^{(k)}(w)) \right) 1_{\Omega_{n,k}} \right|^2 + O(1).
\end{align*}
\]

Note that since \(G_n(z)\) is no longer present in (5.88), (5.89), and (5.90), we are justified dropping the event \(\Omega_n\) as well. The \(O(1)\) is due to the error we get when introducing the event \(Q_{n,k}\) and dropping the event \(\Omega_n\). Since the terms of the sum over \(k\) form a martingale difference sequence, Lemma D.0.8 allowed us to bring the squares inside of the sum.

Next, we show that we can replace \(\Delta_{n,k}(z)\) and \(\Delta_{n,k}(w)\) with \(I_m\) in (5.88), (5.89), and (5.90). We begin by showing the calculation for term 5.88. Observe that by cyclic permutation of the trace, on the event \(\Omega_{n,k} \cap Q_{n,k}\), we have
We bound each expectation in \((5.91), (5.92), (5.93), \) and \((5.94)\). We start by bounding the expectation.

We now show that this is bounded in the \(L^2\)-norm. We use the Cauchy–Schwarz inequality to break the above expression into pieces which have bounded expectation. Observe that

\[
\mathbb{E} \left| \text{tr} \left( \left( G_n^{(k)}(z) U_k \Delta_{n,k}(z) V_k^T G_n^{(k)}(z) \right) \left( G_n^{(k)}(w) U_k \Delta_{n,k}(w) V_k^T G_n^{(k)}(w) \right) \right) \right| \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}}
\]

\[
- \text{tr} \left( \left( G_n^{(k)}(z) U_k V_k^T G_n^{(k)}(z) \right) \left( G_n^{(k)}(w) U_k V_k^T G_n^{(k)}(w) \right) \right)
\]

\[
= \text{tr} \left( \left( \Delta_{n,k}(z) - I_m \right) V_k^T G_n^{(k)}(z) G_n^{(k)}(w) U_k \Delta_{n,k}(w) V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \right)
\]

\[
+ \text{tr} \left( V_k^T G_n^{(k)}(z) G_n^{(k)}(w) U_k (\Delta_{n,k}(w) - I_m) V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \right).
\]

We now show that this is bounded in the \(L^2\)-norm. We use the Cauchy–Schwarz inequality to break the above expression into pieces which have bounded expectation. Observe that

\[
\mathbb{E} \left| \text{tr} \left( \left( G_n^{(k)}(z) U_k \Delta_{n,k}(z) V_k^T G_n^{(k)}(z) \right) \left( G_n^{(k)}(w) U_k \Delta_{n,k}(w) V_k^T G_n^{(k)}(w) \right) \right) \right| \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}}
\]

\[
- \text{tr} \left( \left( G_n^{(k)}(z) U_k V_k^T G_n^{(k)}(z) \right) \left( G_n^{(k)}(w) U_k V_k^T G_n^{(k)}(w) \right) \right)
\]

\[
\ll \mathbb{E} \left[ \left\| (\Delta_{n,k}(z) - I_m) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \left\| V_k^T G_n^{(k)}(z) G_n^{(k)}(w) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right]
\]

\[
\times \left\| \Delta_{n,k}(w) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \left\| V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right]
\]

\[
+ \mathbb{E} \left[ \left\| V_k^T G_n^{(k)}(z) G_n^{(k)}(w) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right]
\]

\[
\times \left\| (\Delta_{n,k}(w) - I_m) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \left\| V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^2 \right]
\]

\[
\ll \left( \mathbb{E} \left\| (\Delta_{n,k}(z) - I_m) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^4 \right)^{1/4} \left( \mathbb{E} \left\| V_k^T G_n^{(k)}(z) G_n^{(k)}(w) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^8 \right)^{1/4}
\]

\[
\times \left( \mathbb{E} \left\| \Delta_{n,k}(w) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^{16} \mathbb{E} \left\| V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^{16} \right)^{1/8}
\]

\[
+ \left( \mathbb{E} \left\| V_k^T G_n^{(k)}(z) G_n^{(k)}(w) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^4 \right)^{1/2}
\]

\[
\times \left( \mathbb{E} \left\| (\Delta_{n,k}(w) - I_m) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^8 \mathbb{E} \left\| V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^8 \right)^{1/4}
\]

We bound each expectation in \((5.91), (5.92), (5.93), \) and \((5.94)\). We start by bounding the expectation of terms involving resolvents. By the same argument as in Lemma \([5.4.12] \) we can show that for \(p \geq 4\) and \(p\) even,

\[
\mathbb{E} \left\| V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^p \ll_p n^{-(p-4)-2}
\]
which shows

\[
\left( \mathbb{E}\left\| V_k^T G_n^{(k)}(z) G_n^{(k)}(w) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^8 \right)^{1/4} \ll n^{-\varepsilon - 1/2},
\]

\[
\left( \mathbb{E}\left\| V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^{16} \right)^{1/8} \ll n^{-3/2\varepsilon - 1/4},
\]

\[
\left( \mathbb{E}\left\| V_k^T G_n^{(k)}(z) G_n^{(k)}(w) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^4 \right)^{1/2} \ll n^{-1},
\]

and

\[
\left( \mathbb{E}\left\| V_k^T G_n^{(k)}(w) G_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^8 \right)^{1/4} \ll n^{-\varepsilon - 1/2}.
\]

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 (5.91) and (5.94).

Recall the expansion from (5.36) can be iterated to get

\[
\Delta_{n,k}(z) = I_m - (V_k^T G_n^{(k)}(z) U_k) + (V_k^T G_n^{(k)}(z) U_k)^2 \Delta_{n,k}(z).
\]

Using this fact, Lemma 5.4.12 and the Cauchy–Schwarz inequality, we get

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

and

\[
\left( \mathbb{E}\left\| (\Delta_{n,k}(w) - I_m) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^8 \right)^{1/4} \ll \left( \mathbb{E}\left\| V_k^T G_n^{(k)}(z) U_k \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^8 \right)^{1/4} + \mathbb{E}\left\| (V_k^T G_n^{(k)}(z) U_k)^2 \Delta_{n,k}(z) \mathbf{1}_{\Omega_{n,k} \cap Q_{n,k}} \right\|^8 \ll (n^{-4\varepsilon - 2} + n^{-12\varepsilon - 2})^{1/4} \ll n^{-\varepsilon - 1/2}.
\]

Therefore, combining these bounds, we can bound (5.91), (5.92), (5.93) and (5.94) by a constant multiple of \(n^{-7/4}\). This concludes the argument to show that we may replace all \(\Delta_{n,k}(z)\) and
\(\Delta_{n,k}(w)\) in term (5.88) with \(I_m\). A very similar argument shows that, for terms (5.89) and (5.90), we have

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

and

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

Ergo, we need to show only that the expression

\[
\mathbb{E}\left|\left( \operatorname{tr}\left( \left( \mathcal{G}^{(k)}_n(z)U_kV_k^T\mathcal{G}^{(k)}_n(z)\right)\mathcal{G}^{(k)}_n(w)U_kV_k^T\mathcal{G}^{(k)}_n(w) \right) \right)^2\right| \ll n^{-2}
\]

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 (5.95), and observe that

\[
\mathbb{E}\left|\left( \operatorname{tr}\left( \left( \mathcal{G}^{(k)}_n(z)U_kV_k^T\mathcal{G}^{(k)}_n(z)\right)\mathcal{G}^{(k)}_n(w) \right)\right)^2\right| \ll n^{-2}
\]

is bounded by \(O(n^{-1})\), where we cyclically permuted the trace and reordered the product of resolvents again.
Since this matrix is \( m \times m \), if we can bound an arbitrary entry uniformly, then we can bound the norm. We then wish to bound

\[
\max_{1 \leq i, j \leq m} E \left| c^T_{(i-1)n+k} (G^{(k)}_n (w))^* (G^{(k)}_n (z))^* V_k V_k^T G^{(k)}_n (z) G^{(k)}_n (w) c_{(j-1)n+k} \mathbf{1}_{\Omega_{n,k}} \right|^2.
\]

Note that

\[
(G^{(k)}_n (w))^* (G^{(k)}_n (z))^* V_k V_k^T G^{(k)}_n (z) G^{(k)}_n (w)
\]

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 5.4.9, we have

\[
E \left| c^*_{(i-1)n+k} (G^{(k)}_n (w))^* (G^{(k)}_n (z))^* V_k V_k^T G^{(k)}_n (z) G^{(k)}_n (w) c_{(j-1)n+k} \mathbf{1}_{\Omega_{n,k}} \right|^2 \ll n^{-2} E \left[ \text{tr} \left( (G^{(k)}_n (w))^* (G^{(k)}_n (z))^* V_k V_k^T G^{(k)}_n (z) G^{(k)}_n (w) \mathbf{1}_{\Omega_{n,k}} \right)^2 \right]
\]

\[
\ll n^{-2} E \left\| V_k G^{(k)}_n (z) G^{(k)}_n (w) (G^{(k)}_n (w))^* (G^{(k)}_n (z))^* V_k \mathbf{1}_{\Omega_{n,k}} \right\|^2.
\]

From here, we bound an arbitrary element of \( V_k G^{(k)}_n (z) G^{(k)}_n (w) (G^{(k)}_n (w))^* (G^{(k)}_n (z))^* V_k \). For any \( 1 \leq i, j \leq m \), it follows easily that

\[
E \left| e^T_{(i-1)n+k} G^{(k)}_n (z) G^{(k)}_n (w) (G^{(k)}_n (w))^* (G^{(k)}_n (z))^* e_{(j-1)n+k} \mathbf{1}_{\Omega_{n,k}} \right|^2 \ll 1.
\]

Ergo, we have

\[
E \left[ \text{tr} \left( (V_k^T G^{(k)}_n (z) G^{(k)}_n (w) U_k)^2 \right) \mathbf{1}_{\Omega_{n,k}} \right]^2 \ll n^{-2}.
\]

This concludes the argument for term (5.95). Since terms (5.96) and (5.97) are symmetric in \( z \) and
$w$, the argument will be the same for both terms. We show the argument for (5.96). Observe that

$$\mathbb{E} \left| \text{tr} \left( V_k^T G_n^{(k)}(w) (G_n^{(k)}(z))^2 U_k \right) \right|^2_{1 \Omega_{n,k}} \ll \max_{1 \leq i \leq m} \mathbb{E} \left| e_{(i-1)n+k}^T G_n^{(k)}(w) (G_n^{(k)}(z))^2 c_{(i-1)n+k} 1_{\Omega_{n,k}} \right|^2$$

$$\ll \max_{1 \leq i \leq m} \mathbb{E} \left[ c_{(i-1)n+k}^* (G_n^{(k)}(z))^{2*} (G_n^{(k)}(w))^* e_{(i-1)n+k} \right. $$

$$\times e_{(i-1)n+k}^T G_n^{(k)}(w) (G_n^{(k)}(z))^2 c_{(i-1)n+k} 1_{\Omega_{n,k}} \left. \right]$$

$$\ll n^{-1}.$$ 

This concludes the argument for (5.96) and the proof of Lemma 5.5.2. \qed
Bibliography


Appendix A

Proof of Lemma 4.3.1

In this section, we present the proof of Lemma 4.3.1. Notation used in Appendix A is consistent with Chapter 4.

Proof of Lemma 4.3.1. Take $L_0 := \sqrt{8E|\xi|^4}$. We begin by proving (ii). Observe that

$$1 = E|\xi|^2 = E|\text{Re}(\xi)|^2 + E|\text{Im}(\xi)|^2$$

$$= E\left[|\text{Re}(\xi)|^2\mathbf{1}_{\{|\text{Re}(\xi)| \leq L/\sqrt{2}\}}\right] + E\left[|\text{Re}(\xi)|^2\mathbf{1}_{\{|\text{Re}(\xi)| > L/\sqrt{2}\}}\right]$$

$$+ E\left[|\text{Im}(\xi)|^2\mathbf{1}_{\{|\text{Im}(\xi)| \leq L/\sqrt{2}\}}\right] + E\left[|\text{Im}(\xi)|^2\mathbf{1}_{\{|\text{Im}(\xi)| > L/\sqrt{2}\}}\right]$$

$$= \text{Var}(\text{Re}(\tilde{\xi})) + \text{Var}(\text{Im}(\tilde{\xi}))$$

$$+ \left|E\left[\text{Re}(\xi)\mathbf{1}_{\{|\text{Re}(\xi)| \leq L/\sqrt{2}\}}\right]\right|^2 + E\left[|\text{Re}(\xi)|^2\mathbf{1}_{\{|\text{Re}(\xi)| > L/\sqrt{2}\}}\right]$$

$$+ \left|E\left[\text{Im}(\xi)\mathbf{1}_{\{|\text{Im}(\xi)| \leq L/\sqrt{2}\}}\right]\right|^2 + E\left[|\text{Im}(\xi)|^2\mathbf{1}_{\{|\text{Im}(\xi)| > L/\sqrt{2}\}}\right] ,$$

which implies

$$1 - \text{Var}(\tilde{\xi}) = \left|E\left[\text{Re}(\xi)\mathbf{1}_{\{|\text{Re}(\xi)| \leq L/\sqrt{2}\}}\right]\right|^2 + E\left[|\text{Re}(\xi)|^2\mathbf{1}_{\{|\text{Re}(\xi)| > L/\sqrt{2}\}}\right]$$

$$+ \left|E\left[\text{Im}(\xi)\mathbf{1}_{\{|\text{Im}(\xi)| \leq L/\sqrt{2}\}}\right]\right|^2 + E\left[|\text{Im}(\xi)|^2\mathbf{1}_{\{|\text{Im}(\xi)| > L/\sqrt{2}\}}\right] .$$

Thus, using the fact that $\text{Re}(\xi)$ and $\text{Im}(\xi)$ both have mean zero (so, for example,

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

and then applying Jensen’s inequality, we
obtain

$$
|1 - \text{Var}(\hat{\xi})| \leq 2E[|\text{Re}(\xi)|^2 1_{\{|\text{Re}(\xi)| > L/\sqrt{2}\}}] + 2E[|\text{Im}(\xi)|^2 1_{\{|\text{Im}(\xi)| > L/\sqrt{2}\}}]
$$

$$
\leq 2E[|\xi|^2 1_{\{|\xi| > L/\sqrt{2}\}}]
$$

$$
\leq \frac{4}{L^2} E|\xi|^4.
$$

This concludes the proof of (ii).

Property (i) follows easily from (ii) by the choice of $L_0$. Next we move onto the proof of (iii).

One can see that since $\text{Var}(\hat{\xi}) \geq \frac{1}{2}$,

$$
|\hat{\xi}| \leq \frac{|\text{Re}(\xi)| 1_{\{|\text{Re}(\xi)| \leq L/\sqrt{2}\}} + E \left[|\text{Re}(\xi)| 1_{\{|\text{Re}(\xi)| \leq L/\sqrt{2}\}}\right]}{\sqrt{\text{Var}(\hat{\xi})}}
$$

$$
+ \frac{|\text{Im}(\xi)| 1_{\{|\text{Im}(\xi)| \leq L/\sqrt{2}\}} + E \left[|\text{Im}(\xi)| 1_{\{|\text{Im}(\xi)| \leq L/\sqrt{2}\}}\right]}{\sqrt{\text{Var}(\hat{\xi})}}
$$

$$
\leq 4L
$$

almost surely.

For (iv), we observe that $\hat{\xi}$ has mean zero and unit variance by construction. Additionally, since the real and imaginary parts of $\hat{\xi}$ depend only on the real and imaginary parts of $\xi$ respectively, they are independent by construction. For the fourth moment, we use that $\text{Var}(\hat{\xi}) \geq \frac{1}{2}$ and Jensen's inequality to obtain

$$
E|\hat{\xi}|^4 \ll \frac{1}{\text{Var}(\hat{\xi})^2} \left(E \left[|\text{Re}(\xi)|^4 1_{\{|\text{Re}(\xi)| \leq L/\sqrt{2}\}}\right] + E \left[|\text{Im}(\xi)|^4 1_{\{|\text{Im}(\xi)| \leq L/\sqrt{2}\}}\right]\right)
$$

$$
\ll E|\xi|^4,
$$

as desired. □
Appendix B

Proof of Theorem 4.4.1

This section is devoted to the proof of Theorem 4.4.1. We begin with Lemma B.0.1 below, which is based on [90, Theorem 4]. Throughout this section, we use $\sqrt{-1}$ for the imaginary unit and reserve $i$ as an index.

**Lemma B.0.1.** Let $\mu$ be a probability measure on $[0, \infty)$, and for each $n \geq 1$, let

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_{n,i}}$$

for some triangular array $\{\lambda_{n,i}\}_{i \leq n}$ of nonnegative real numbers. Let $m_n$ by the Stieltjes transform of $\mu_n$ and $m$ be the Stieltjes transform of $\mu$, i.e.,

$$m_n(z) := \int \frac{d\mu_n(x)}{x - z}, \quad m(z) := \int \frac{d\mu(x)}{x - z}$$

for all $z \in \mathbb{C}$ with $\text{Im}(z) > 0$. Assume

(i) $\mu_n \rightarrow \mu$ as $n \rightarrow \infty$,

(ii) there exists a constant $c > 0$ such that $\mu([0,c]) = 0$,

(iii) $\sup_{E \in [0,c]} |m_n(E + \sqrt{-1}n^{-1/2}) - m(E + \sqrt{-1}n^{-1/2})| = o(n^{-1/2})$.

Then there exists a constant $n_0 \geq 1$ such that $\mu_n([0,c/2]) = 0$ for all $n > n_0$.

**Proof.** Observe that

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

\[ \sup_{E \in [0,c]} \left| \int \frac{d(\mu_n - \mu)(x)}{(E-x)^2 + n^{-1}} \right| = o(1). \]

We decompose this integral into two parts

\[ \int \frac{d(\mu_n - \mu)(x)}{(E-x)^2 + n^{-1}} = \int_0^c \frac{d\mu_n(x)}{(E-x)^2 + n^{-1}} + \int_c^{\infty} \frac{d(\mu_n - \mu)(x)}{(E-x)^2 + n^{-1}}, \]

where we used the assumption that \( \mu([0,c]) = 0 \).

Observe that

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

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

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

Take \( n_0 \geq 1 \) such that

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

for all \( n \geq n_0 \).

In order to reach a contradiction, assume there exists \( n > n_0 \) and \( i \in [n] \) such that \( \lambda_{n,i} \in [0,c/2] \). Then

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

a contradiction of (B.1). We conclude that \( \mu_n([0,c/2]) = 0 \) for all \( n > n_0 \). \( \square \)

With Lemma B.0.1 in hand, we are now prepared to prove Theorem 4.4.1 The proof below is based on a slight modification to the arguments from [90] [89]. As such, in some places we will omit technical computations and only provide appropriate references and necessary changes to results from [90] [89].
Fix $\delta > 0$. It suffices to prove that

$$\inf_{1 + \delta \leq |z| \leq 6} s_{mn} \left( \frac{1}{\sqrt{n}} Y_n - zI \right) \geq c$$  \hspace{1cm} (B.2)

and

$$\inf_{|z| > 6} s_{mn} \left( \frac{1}{\sqrt{n}} Y_n - zI \right) \geq c'$$  \hspace{1cm} (B.3)

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

The second bound (B.3) follows by Lemma D.0.10. Indeed, a bound on the spectral norm of $Y_n$ (which follows from standard bounds on the spectral norms of $X_{n,k}$; see, for example, [114, Theorem 1.4]) gives

$$\|Y_n\| \leq 3\sqrt{n}$$

with overwhelming probability. The bound in (B.3) then follows by applying Lemma D.0.10.

We now turn to the bound in (B.2). To prove this bound, we will use Lemma B.0.1. Let $\mu_{n,z}$ be the empirical spectral measure constructed from the eigenvalues of

$$\left( \frac{1}{\sqrt{n}} Y_n - zI \right) \left( \frac{1}{\sqrt{n}} Y_n - zI \right)^*.$$  

From [89, Theorem 2.6], for all $|z| \geq 1 + \delta$, there exists a probability measure $\mu_z$ supported on $[0, \infty)$ such that $\mu_{n,z} \to \mu_z$ with overwhelming probability. Moreover, from [15, Lemma 4.2] there exists a constant $c > 0$ (depending only on $\delta$) such that $\mu_z([0, c]) = 0$ for all $|z| \geq 1 + \delta$. Lastly, condition (iii) in Lemma B.0.1 follows for all $1 + \delta \leq |z| \leq 6$ with overwhelming probability from [90, Theorem 5]. Applying Lemma B.0.1, we conclude that

$$s_{mn} \left( \frac{1}{\sqrt{n}} Y_n - zI \right) \geq c/2$$

with overwhelming probability uniformly for all $1 + \delta \leq |z| \leq 6$. The bound for the infimum can now be obtained by a simple net argument and Weyl’s inequality (4.21). The proof of Theorem 4.4.1 is complete.
Appendix C

Proof of Lemma 4.7.17

This section is devoted to the proof of Lemma 4.7.17. The proof proceeds inductively. We begin with a graph only containing the vertex (1, 1) and then add vertices and edges sequentially with time. First, edge 1 is added; it will span from (1, 1) to (2, i_2). Next, edge 2 is added and will span from (2, i_2) to (3, i_3), and so on. We use induction to prove that at each time step \( t \), there is only one possible choice for \( i_{t+1} \), resulting in a unique canonical graph with maximal height \( k/2 \) and in which each edge is parallel to exactly one other edge.

The edge starting at vertex (1, 1) can either be of type II (terminating on (2, 1)) or of type I (terminating on (2, 2)). By way of contradiction, assume the edge is type II. Since \( G \) still has \( k/2 - 1 \) more height coordinates left to reach, it would require at least \( k/2 - 1 \) type I edges to reach height coordinate \( k/2 \). Since each edge must be parallel to exactly one other edge, at some point there must be a type II edge, returning to a height coordinate previously visited. This edge will also need to be parallel to another edge. Counting all pairs of parallel edges shows that \( G \) must have at least \( k + 1 \) more edges, a contradiction. Hence the edge starting at vertex (1, 1) must be of type I.

Assume that all edges up to time coordinate \( t \), where \( 1 \leq t < k/2 - 1 \), are type I edges. Then \( G \) must have an edge starting at vertex \((t + 1, t + 1)\). This edge can either of type I or type II. In order to reach a contradiction, assume that the edge is type II. Then \( G \) must have at least \( k/2 - t - 1 \) more type I edges in order to reach the height \( k/2 \), and \( G \) has exactly \( k - t - 1 \) more
edges to be added. Visiting each unvisited height coordinate would require at least \( k/2 - t - 1 \) more type I edges, and at some point after visiting new height coordinates, \( G \) must return to a smaller height coordinate, resulting in a type II edge. None of these edges could be parallel to any previous edges. Thus, overall \( G \) would need to have at least \( k - t + 1 \) more edges, a contradiction to the fact that \( G \) must have exactly \( k - t - 1 \) more edges.

We conclude that each edge of \( G \) must be type I until the height coordinate \( k/2 \) is reached. Namely, we have vertices \((1,1), \ldots, (k/2, k/2)\).

At this point \( G \) must have an edge starting at vertex \((k/2, k/2)\). Note that \( G \) has \( k/2 - 1 \) edges up to this point, none of which are parallel to any other edge. \( G \) must have edges parallel to the edges previously introduced and \( G \) has exactly \( k/2 + 1 \) edges remaining to do so. Since there are no remaining unvisited height coordinates, the edge which starts at vertex \((k/2, k/2)\) must terminate at \((k/2 + 1, i)\) for some \( 1 \leq i \leq k/2 \), resulting in the first type II edge. See Figure C.1 for a visual representation of the graph up to this point.

We now claim that this first type II edge must in fact terminate at \((k/2 + 1, 1)\). By way of contradiction, suppose this edge terminates at vertex \((k/2 + 1, i)\) for any \( 1 < i \leq k/2 \). Since this is the first type II edge, it cannot be parallel to any other previously drawn edge. Up to this point \( G \) has \( k/2 \) edges drawn and \( k/2 \) edges remaining to be drawn. Since all edges are by assumption to be parallel to exactly one other, a simple counting argument reveals that each edge drawn from this point on must be parallel to an existing edge. Since there is only one edge which starts at height coordinate \( i \), we must now draw the edge starting at \((k/2 + 1, i)\) and terminating at \((k/2 + 2, i + 1)\). By continuing this argument inductively, we must draw edges which start at \((k/2 + j + 1, i + j)\) and terminate at vertex \((k/2 + j + 2, i + j + 1)\) for \( 0 \leq j \leq k/2 - i \), until the height coordinate \( k/2 \) is reached again. Now, since \( k - i + 1 < k \), we must draw at least one more edge, and this edge must start at vertex \((k - i, k/2)\). In order to draw an edge parallel to an existing edge, this edge must terminate at vertex \((k - i + 1, i)\). However, if we do this, we must now draw an edge parallel to an existing edge which starts a height coordinate \( i \), a contradiction because the only previous edges which began at height coordinate \( i \) are parallel to one another and there cannot be three
edges parallel. This concludes the proof of the claim.

By the previous claim, the first type II edge must terminate at \((k/2 + 1, 1)\). Again, since this is the first type II edge it cannot be parallel to any other edge. Up to this point \(G\) has \(k/2\) edges drawn and \(k/2\) edges remaining to be drawn. Since all edges are by assumption parallel to exactly one other, a simple counting argument reveals that each edge drawn from this point on must be parallel to an existing edge. Since there is only one edge which starts at height coordinate 1, we must now draw the edge starting at \((k/2 + 1, 1)\) and terminating at \((k/2 + 2, 2)\). By continuing this argument inductively, we must draw edges which start at \((k/2 + j + 1, j + 1)\) and terminate at vertex \((k/2 + j + 2, j + 2)\) for \(0 \leq j \leq k/2 - 1\), until the height coordinate \(k/2\) is reached again. Up to this point, \(k - 1\) edges of \(G\) have been drawn and we must draw one more edge which starts at vertex \((k, k/2)\). Since there is only one previous edge in the graph which starts at height coordinate \(k/2\), the final edge must terminate at height coordinate 1 and all edges are parallel to exactly one other edge. This results in the vertex set

\[
V = \{(1,1), (2,2), \ldots, (k/2,k/2), (k/2 + 1,1), (k/2 + 2,2), \ldots (k,k/2), (k+1,1)\}.
\]

The corresponding canonical \(m\)-colored \(k\)-path graph would be

\[
G^1(1,2,\ldots,k/2,1,2,\ldots,k/2,1),
\]

and the proof is complete.

\[\square\]

Figure C.1: This is an example of a path graph that has type I edges drawn until a height of \(k/2\) is reached, then the first type II edge is drawn.
Appendix D

Useful Inequalities and Lemmas

Lemma D.0.1 (Lemma 2.7 from [18]). 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 $p \geq 2$,

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

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

Lemma D.0.2. 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, let $v = (v_i)_{i=1}^N$ be a vector with the following properties:

(i) $v_i = 1$ for $i \in S$,

(ii) $v_i = 0$ for $i \notin S$.

Then for any $p \geq 2$,

$$
\mathbb{E} |w^*Aw - \text{tr}(v^TAv)|^p \ll_p \left( \mathbb{E} |w_i|^4 \text{tr}(A^*A) \right)^{p/2} + \mathbb{E} |w_i|^{2p} (\text{tr}(A^*A))^{p/2}.
$$
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

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

and

\[
\text{tr}(v^TAv) = \sum_i v_i A_{(i,i)} v_i = \sum_{i \in S} A_{(i,i)} = \text{tr}(A_{S \times S}).
\]

Therefore, by Lemma D.0.1

\[
\mathbb{E}|w^*Aw - \text{tr}(v^TAv)|^p = \mathbb{E}|w_S^* A_{S \times S} w_S - \text{tr}(A_{S \times S})|^p
\]

\[
\leq_p \left( \mathbb{E}|w_i|^4 \text{tr}(A_{S \times S}^* A_{S \times S}) \right)^{p/2} + \mathbb{E}|w_i|^{2p} \text{tr}(A_{S \times S}^* A_{S \times S})^{p/2}
\]

\[
\leq_p \left( \mathbb{E}|w_i|^4 \text{tr}(A_{S \times S}^* A_{S \times S}) \right)^{p/2} + \mathbb{E}|w_i|^{2p} \left( \text{tr}(A_{S \times S}^* A_{S \times S}) \right)^{p/2}
\]

Now observe that

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

Therefore, we have

\[
\mathbb{E}|w^*Aw - \text{tr}(v^TAv)|^p
\]

\[
\leq_p \left( \mathbb{E}|w_i|^4 \text{tr}(A_{S \times S}^* A_{S \times S}) \right)^{p/2} + \mathbb{E}|w_i|^{2p} \left( \text{tr}(A_{S \times S}^* A_{S \times S}) \right)^{p/2}
\]

\[
\leq \left( \mathbb{E}|w_i|^4 \text{tr}(A^* A) \right)^{p/2} + \mathbb{E}|w_i|^{2p} \left( \text{tr}(A^* A) \right)^{p/2}
\]

concluding the proof. \( \square \)

**Lemma D.0.3** (Lemma A.1 from [18]). For \( X = (x_1, x_2, \ldots, x_N)^T \) iid standardized complex entries, \( B \) an \( N \times N \) Hermitian nonnegative definite matrix, we have, for any \( p \geq 1 \),

\[
\mathbb{E}|X^* BX|^p \leq K_p \left( \langle tr B \rangle^p + \mathbb{E}|x_1|^{2p} \text{tr} B^p \right)
\]

where \( K_p > 0 \) depends only on \( p \).
Lemma D.0.4. 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 \( p \geq 2 \),

\[
\mathbb{E} |X^* B X|^p \leq K_p \mathbb{E}|x_1|^{2p} \left( (\text{tr } B^* B)^{p/2} + |\text{tr } B|^p \right)
\]

where \( K_p > 0 \) depends only on \( p \).

**Proof.** This is a corollary from Lemma D.0.3 since \( \text{tr}(B^* B)^{p/2} \leq (\text{tr } B^* B)^{p/2} \) for any matrix \( B \).

Lemma D.0.5. 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 \( p \geq 2 \),

\[
\mathbb{E} |w^* A w|^p \ll_p \mathbb{E}|w_i|^{2p} (\text{tr } A)^p .
\]

**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 D.0.4 we get

\[
\mathbb{E} |w^* A w|^p \ll_p \mathbb{E}|w_i|^{2p} \left( (\text{tr } A_{S \times S}^* A_{S \times S})^{p/2} + |\text{tr } A_{S \times S}|^p \right).
\]

Since \( A \) is Hermitian, the sub-matrix \( A_{S \times S} \) is also Hermitian and so \( A_{S \times S}^* A_{S \times S} = (A_{S \times S})^2 \). Also recall that if a matrix \( B \) is non-negative definite, the diagonal elements are non-negative so that \( \text{tr}(B)^p \leq (\text{tr}(B))^p \). By this fact and by Lemma 4.6.5 we observe that

\[
\mathbb{E}|w_i|^{2p} \left( (\text{tr } A_{S \times S}^* A_{S \times S})^{p/2} + |\text{tr } A_{S \times S}|^p \right) \ll_p \mathbb{E}|w_i|^{2p} \left( (\text{tr}(A_{S \times S})^p)^{p/2} + (\text{tr } A)^p \right)
\]
\[
\ll_p \mathbb{E}|w_i|^{2p} ((\text{tr } A_{S \times S})^p + (\text{tr } A)^p)
\]
\[
\ll_p \mathbb{E}|w_i|^{2p} (\text{tr } A)^p
\]
concluding the proof.

**Lemma D.0.6** (Follows from Lemma 10 in [98]). Let $A$ and $B$ be $k \times k$ matrices with $\|A\|, \|B\| = O(1)$. Then

$$|\det(A) - \det(B)| \ll_k \|A - B\|.$$

**Lemma D.0.7.** Suppose that $X_n$ and $Y_n$ are two sequences of random variables and assume that $Y_n$ converges in distribution to a random variable $Y$ as $n \to \infty$. If any of the following hold:

1. For any $\varepsilon > 0$, $\mathbb{P}(|X_n - Y_n| > \varepsilon) = o(1),$
2. $\mathbb{P}(X_n = Y_n) = 1 - o(1),$
3. $\mathbb{E}|X_n - Y_n|^p = o(1)$ for $p \geq 1,$

then $X_n$ converges in distribution to $Y$ as $n \to \infty$ as well.

**Proof.** These all imply $|X_n - Y_n| \to 0$ in distribution as $n \to \infty$. Thus the claim follows by [33, Theorem 25.4].

**Lemma D.0.8** (Theorem 3.2 from [41]). Let $\{X_i\}_{i=1}^N$ be a complex martingale difference sequence with respect to the filtration $\{\mathcal{F}_i\}_{i=1}^N$. Then

$$\mathbb{E}\left|\sum_{i=1}^N X_i\right|^2 \ll \sum_{i=1}^N \mathbb{E}|X_i|^2.$$

**Lemma D.0.9.** Let $A$ and $B$ be $n \times n$ matrices. Then

$$|\text{tr}(AB)| \leq \|AB\|_2 \leq \|A\| \cdot \|B\|_2.$$

**Proof.** This follows from the definition of the Hilbert–Schmidt norm and an application of [19, Theorem A.10].
Lemma D.0.10 (Spectral norm bound for large $|z|$; Lemma 3.1 from [94]). Let $A$ be a square matrix that satisfies $\|A\| \leq K$. Then
\[
\left\|(A - zI)^{-1}\right\| \leq \frac{1}{\varepsilon}
\]
for all $z \in \mathbb{C}$ with $|z| \geq K + \varepsilon$.

Lemma D.0.11. Let $\hat{X}_n$ be an $n \times n$ iid random matrix with atom variable $\hat{\xi}$ which has mean zero, variance one, finite fourth moment, and satisfies $|\hat{\xi}| < n^{1/2 - \varepsilon}$ for some $\varepsilon > 0$. Then
\[
E \left\| \hat{X}_n \right\|^2 = O(n)
\]
where $\|\cdot\|$ denotes the operator norm.

Proof. Observe that, for a constant $C > 0$
\[
E \left\| \hat{X}_n \right\|^2 \leq E \left\| \hat{X}_n 1_{\{\|\hat{X}_n\| \leq C\sqrt{n}\}} \right\|^2 + E \left\| \hat{X}_n 1_{\{\|\hat{X}_n\| > C\sqrt{n}\}} \right\|^2 \\
\leq O(n) + n^{3 - 2\varepsilon} P \left( \| \hat{X}_n \| > C\sqrt{n} \right)
\]
where the power of $n$ came from bounding the operator norm by the Frobenius norm. By [19] Theorem 5.9, $P \left( \| \hat{X}_n \| > C_n \sqrt{n} \right) = O(n^{-\alpha})$ for any $\alpha > 0$, so by selecting $\alpha$ sufficiently large, arrive at the desired result.

Lemma D.0.12. 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
\[
E \|X_n\|^2 = O(n)
\]
where $\|\cdot\|$ denotes the operator norm.

Proof. Let $\hat{X}_n$ be the truncated $n \times n$ iid random matrix with entries defined by
\[
\hat{X}_{(i,j)} := X_{(i,j)} 1_{\{|X_{(i,j)}| \leq n^{1/2 - \varepsilon}\}} - E \left[ X_{(i,j)} 1_{\{|X_{(i,j)}| \leq n^{1/2 - \varepsilon}\}} \right], \\
\hat{X}_{(i,j)} := \frac{\hat{X}_{(i,j)}}{\sqrt{\text{Var}(\hat{X}_{(i,j)})}}.
\]
By the triangle inequality we have

\[ E \|X_n\|^2 \leq E \|X_n - \hat{X}_n\|^2 + E \|\hat{X}_n\|^2. \]

By Lemma 5.2.7

\[ E \|X_n - \hat{X}_n\|^2 = o(n^{-4\epsilon}) \]

and by Lemma D.0.11

\[ E \|\hat{X}_n\|^2 = O(n), \]

thus \( E \|X_n\|^2 = O(n) \) as desired. \qed
Appendix E

Proof of Lemma 5.2.4

We prove Lemma 5.2.4. Notation used in Appendix E is consistent with Chapter 5.

Proof of Lemma 5.2.4. First, we prove property (i). Observe that

\[ 1 = \text{Var}(\xi) = \mathbb{E}[\xi^2 1_{\{|\xi| \leq n^{1/2-\epsilon}\}}] + \mathbb{E}[\xi^2 1_{\{|\xi| > n^{1/2-\epsilon}\}}] \]

\[ = \text{Var}(\tilde{\xi}) + \left( \mathbb{E}[\xi 1_{\{|\xi| \leq n^{1/2-\epsilon}\}}] \right)^2 + \mathbb{E}[\xi^2 1_{\{|\xi| > n^{1/2-\epsilon}\}}]. \]

Ergo

\[ |1 - \text{Var}(\tilde{\xi})| = \left( \mathbb{E}[\xi 1_{\{|\xi| \leq n^{1/2-\epsilon}\}}] \right)^2 + \mathbb{E}[\xi^2 1_{\{|\xi| > n^{1/2-\epsilon}\}}]. \]

Also observe that

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

which implies

\[ -\mathbb{E}[\xi 1_{\{|\xi| \leq n^{1/2-\epsilon}\}}] = \mathbb{E}[\xi 1_{\{|\xi| > n^{1/2-\epsilon}\}}] \]

and thus

\[ \left| \mathbb{E}[\xi 1_{\{|\xi| \leq n^{1/2-\epsilon}\}}] \right| = \left| \mathbb{E}[\xi 1_{\{|\xi| > n^{1/2-\epsilon}\}}] \right|. \]
Hence

\[ |1 - \text{Var}(\tilde{\xi})| = \left( \mathbb{E}[\xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}}] \right)^2 + \mathbb{E}[\xi^2 \mathbf{1}_{\{\xi > n^{1/2-\varepsilon}\}}] \]

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

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

\[ \leq 2 \mathbb{E} \left[ \frac{|\xi|^4}{n^{1-2\varepsilon}} \mathbf{1}_{\{\xi > n^{1/2-\varepsilon}\}} \right] \]

\[ = o(n^{-1-2\varepsilon}). \]

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

\[ 1 - \frac{C}{n^{1+2\varepsilon}} \leq \text{Var}(\hat{\xi}) \]

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

\[ |\xi| = \left| \frac{\xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} - \mathbb{E} \left[ \xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} \right]}{\sqrt{\text{Var}(\hat{\xi})}} \right| \]

\[ \leq 2 \left| \xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} \right| + 2 \left| \mathbb{E} \left[ \xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} \right] \right| \]

\[ \leq 4n^{1/2-\varepsilon} \]

almost surely. For part (iii) we have

\[ \mathbb{E}[|\xi|^4] = \mathbb{E} \left[ \left| \frac{\xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} - \mathbb{E} \left[ \xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} \right]}{\sqrt{\text{Var}(\hat{\xi})}} \right]^4 \right] \]

\[ \leq 2^{4} \mathbb{E} \left| \xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} - \mathbb{E} \left[ \xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} \right] \right|^4 \]

\[ \leq 2^{8} \mathbb{E} \left| \xi \mathbf{1}_{\{\xi \leq n^{1/2-\varepsilon}\}} \right|^4 \]

\[ \leq 2^{8} \mathbb{E} |\xi|^4 \]

completing the proof of the claim.
Appendix F

Chapter 5 Events

In this section, we will prove that all the necessary events used in Chapter 5 hold with the necessary probabilities. All notation used in Appendix F is consistent with Chapter 5. 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 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 (5.6). Also define truncated matrices $\tilde{X}_{n,k}$ and $\hat{X}_{n,k}$ as in (5.7) and (5.8) respectively. Define the linearized truncated matrix $Y_n$ as in (5.11). 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 $\hat{X}$ and $\hat{X}$ to be the $n \times n$ matrices with entries defined by

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

and

$$\hat{X}_{(i,j)} := \frac{\hat{X}_{(i,j)}}{\sqrt{\text{Var}(\hat{X}_{(i,j)})}}.$$
for \(1 \leq i, j \leq n\). Define \(\check{X}_{n,1}, \check{X}_{n,2}, \ldots, \check{X}_{n,m}\) and \(\hat{X}_{n,1}, \hat{X}_{n,2}, \ldots, \hat{X}_{n,m}\) as in (F.1) and (F.2) respectively. Finally, define the linearized truncated matrix

\[
\check{Y}_n := n^{-1/2} \begin{bmatrix}
0 & \check{X}_{n,1} & 0 & \cdots & 0 \\
0 & 0 & \check{X}_{n,2} & \cdots & 0 \\
0 & 0 & 0 & \ddots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \hat{X}_{n,m-1} \\
\hat{X}_{n,m} & 0 & 0 & \cdots & 0
\end{bmatrix}.
\]  

(F.3)

These will be useful in the upcoming proofs.

**Remark F.0.1.** In the case where the atom variables are complex valued, these theorem statements and proofs would need to be adjusted. They would follow in the same manner with appropriate adjustments throughout. In particular, the truncation would need to preserve independence between the real and imaginary parts of the random variables.

**Lemma F.0.2.** Fix \(\varepsilon > 0\). For a fixed integer \(m > 0\), let \(\xi_1, \xi_2, \ldots, \xi_m\) be real-valued random variables each mean zero, variance one, and finite \(4 + \tau\) moment for some \(\tau > 0\). Let \(\check{X}_{n,1}, \check{X}_{n,2}, \ldots, \check{X}_{n,m}\) be independent iid random matrices with atom variables as defined in (5.8), and define \(Y_n\) as in (5.11). Fix \(\delta > 0\). Then there exists a constant \(c > 0\) depending only on \(\delta\) such that

\[
\inf_{|z| > 1 + \delta/2} s_{mn} (Y_n - zI) \geq c
\]

with overwhelming probability.

**Proof.** Fix \(\delta > 0\) and define \(\check{Y}_n\) as in (F.3). By Lemma 4.4.1, we know that there exists a constant \(c' > 0\) which depends only on \(\delta\) such that

\[
\inf_{|z| > 1 + \delta/2} s_{mn} (\check{Y}_n - zI) \geq c'
\]
with overwhelming probability. Note that by Weyl’s inequality \(3.7\),
\[
\sup_{z \in \mathbb{C}} |s_{mn}(\tilde{Y}_n - zI) - s_{mn}(Y_n - zI)| \leq \|\tilde{Y}_n - Y_n\| \\
\leq \max_{1 \leq k \leq m} \frac{1}{\sqrt{n}} \|\tilde{X}_{n,k} - \hat{X}_{n,k}\|.
\]

Focusing on an arbitrary value of \(k\), we have
\[
\frac{1}{\sqrt{n}} \left\|\tilde{X}_{n,k} - \hat{X}_{n,k}\right\| \leq \frac{1}{\sqrt{n}} \left\|\hat{X}_{n,k} \left(1 - \sqrt{\text{Var}(\hat{X}_{n,k}(1,1))}\right)\right\|.
\]

Observe that
\[
\frac{1}{\sqrt{n}} \left\|\frac{\tilde{X}_{n,k}}{\sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}} - \hat{X}_{n,k}\right\| = \frac{1}{\sqrt{n}} \left\|\frac{\tilde{X}_{n,k} \left(1 - \sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}\right)}{\sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}}\right\|.
\]

By the proof of item (ii) in Lemma 4.3.1, \(\left\{\text{Var}(\hat{X}_{n,k}(1,1))\right\}^{-1/2} \leq 2\) for \(L\) sufficiently large. Additionally, an argument similar to that of the proof of item (i) of Lemma 4.3.1 shows that
\[
\left|1 - \sqrt{\text{Var}(\hat{X}_{n,k}(1,1))}\right| \leq C^L \text{ for some constant } C > 0.
\]

Therefore by [114, Theorem 1.4], for \(L\) sufficiently large,
\[
\frac{1}{\sqrt{n}} \left\|\frac{\tilde{X}_{n,k}}{\sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}} - \hat{X}_{n,k}\right\| \leq \frac{C}{L^2 \sqrt{n}} \left\|\frac{\tilde{X}_{n,k} \left(1 - \sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}\right)}{\sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}}\right\| \leq \frac{c'}{16}
\]

with overwhelming probability. Similarly,
\[
\frac{1}{\sqrt{n}} \left\|\frac{\tilde{X}_{n,k}}{\sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}} - \hat{X}_{n,k}\right\| = \frac{1}{\sqrt{n}} \left\|\frac{\tilde{X}_{n,k} \left(1 - \sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}\right)}{\sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}}\right\|.
\]

By the arguments to prove part (ii) of Lemma 5.2.4, \(\left\{\text{Var}(\tilde{X}_{n,k}(1,1))\right\}^{-1/2} \leq 2\). Also, by very similar arguments to the proof of part (i) of Lemma 5.2.4, we can show that
\[
\left|1 - \sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}\right| = o(n^{-1+2\varepsilon}).
\]

Therefore by [19, Theorem 5.9],
\[
\frac{1}{\sqrt{n}} \left\|\frac{\tilde{X}_{n,k}}{\sqrt{\text{Var}(\tilde{X}_{n,k}(1,1))}} - \hat{X}_{n,k}\right\| = o(n^{-1+2\varepsilon}) \frac{1}{\sqrt{n}} \left\|\tilde{X}_{n,k}\right\| \leq \frac{c'}{16}.
\]
with overwhelming probability. Ergo, by the triangle inequality, for \( L \) sufficiently large,

\[
\frac{1}{\sqrt{n}} \left\| \hat{X}_{n,k} - \tilde{X}_{n,k} \right\| \\
\leq \frac{1}{\sqrt{n}} \left\| \frac{\hat{X}_{n,k}}{\sqrt{\text{Var}((\hat{X}_{n,k})_{(1,1)})}} - \frac{\tilde{X}_{n,k}}{\sqrt{\text{Var}((\tilde{X}_{n,k})_{(1,1)})}} \right\| \\
\leq \frac{c'}{8} + \frac{1}{\sqrt{n}} \left\| \hat{X}_{n,k} - \tilde{X}_{n,k} \right\|.
\]

\[(F.4)\]

with overwhelming probability.

Now, recall that the entries of \( \hat{X}_{n,k} \) 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 \( L \), then the entries in \( \hat{X}_{n,k} \) and \( \hat{X}_n \) agree. Similarly, if all entries are greater than \( n^{1/2 - \varepsilon} \) then the entries in \( \hat{X}_{n,k} \) and \( \hat{X}_n \) agree. Ergo, we need only consider the case when there exists some entries \( 1 \leq i, j \leq n \) such that \( L \leq |(\hat{X}_{n,k})_{i,j}| \leq n^{1/2 - \varepsilon} \).

For each \( 1 \leq k \leq m \), define the random variables

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

and define \( \hat{X}_{n,k} \) to be the matrix with entries

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

for \( 1 \leq i, j \leq n \). Then we can write

\[
\frac{1}{\sqrt{n}} \left\| \hat{X}_{n,k} - \tilde{X}_{n,k} \right\| = \frac{1}{\sqrt{n}} \left\| \hat{X}_{n,k} \right\|.
\]

By [19] Lemma 5.9], for \( L \) sufficiently large,

\[
\frac{1}{\sqrt{n}} \left\| \hat{X}_{n,k} \right\| \leq \frac{c'}{8}
\]

\[(F.5)\]

with overwhelming probability. Thus, by choosing \( L \) large enough to satisfy both conditions, by \((F.4)\) and \((F.5)\),

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

with overwhelming probability, which implies that

\[
\inf_{|z| > 1 + \delta/2} s_{mn} (\mathcal{Y}_n - zI) \geq c
\]
with overwhelming probability where $c = \frac{c'}{2}$.

**Lemma F.0.3.** Fix $\varepsilon > 0$. For a fixed integer $m > 0$, let $\xi_1, \xi_2, \ldots, \xi_m$ be real-valued random variables each mean zero, variance one, and finite $4 + \tau$ moment for some $\tau > 0$. Let $X_{n,1}, X_{n,2}, \ldots, X_{n,m}$ be independent iid random matrices with atom variables $\xi_1, \xi_2, \ldots, \xi_m$ respectively. Define $\hat{X}_{n,1}, \hat{X}_{n,2}, \ldots \hat{X}_{n,m}$ as in (5.8), and define $\hat{P}_n$ as in (5.9). Fix $\delta > 0$. Then 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 F.0.2 we know that there exists some $c' > 0$ such that

$$
\inf_{|z| > 1 + \delta/2} s_{mn} (Y_n - zI) \geq c'
$$

with overwhelming probability. Recall that

$$
s_n (Y_n' - zI) = s_1 \left( (Y_n - zI)^{-1} \right)
$$

provided $z$ is not an eigenvalue of $Y_n$. A block inverse matrix calculation reveals that

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

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

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

$$= \sup_{|z| > 1 + \delta/2} \left\| (Y_n - zI)^{-1} \right\|$$

$$\geq \sup_{|z| > 1 + \delta/2} |z|^{m-1} \left\| (\hat{P}_n - z^m I)^{-1} \right\|$$

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

$$\frac{1}{c} \geq \sup_{|z| > 1 + \delta/2} s_1 \left( (\hat{P}_n - zI)^{-1} \right)$$
which overwhelming probability. This gives
\[
\inf_{|z| > 1 + \delta/2} s_n \left( \hat{P}_n - zI \right) \geq c
\]
with overwhelming probability.

**Lemma F.0.4.** For a fixed integer \( m > 0 \), let \( \xi_1, \xi_2, \ldots, \xi_m \) be real-valued random variables each satisfying Assumption 2.2.1. Fix \( \delta > 0 \) and let \( X_{n,1}, X_{n,2}, \ldots, X_{n,m} \) be independent iid random matrices with atom variables \( \xi_1, \xi_2, \ldots, \xi_m \) 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 \( 1 - o(1) \).

**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 F.0.3 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 5.2.8
\[
\mathbb{P} \left( \inf_{|z| > 1 + \delta/2} s_n \left( P_n - zI \right) < \frac{c'}{2} \right)
\]
\[
= \mathbb{P} \left( \inf_{|z| > 1 + \delta/2} s_n \left( P_n - zI \right) < \frac{c'}{2} \text{ and } \left\| P_n - \hat{P}_n \right\| \leq n^{-\varepsilon} \right) + \mathbb{P} \left( \inf_{|z| > 1 + \delta/2} s_n \left( P_n - zI \right) < \frac{c'}{2} \text{ and } \left\| P_n - \hat{P}_n \right\| > n^{-\varepsilon} \right)
\]
\[
\leq \mathbb{P} \left( \inf_{|z| > 1 + \delta/2} s_n \left( P_n - zI \right) < \frac{c'}{2} \text{ and } \left\| P_n - \hat{P}_n \right\| \leq n^{-\varepsilon} \right) + \mathbb{P} \left( \left\| P_n - \hat{P}_n \right\| > n^{-\varepsilon} \right)
\]
\[
\leq \mathbb{P} \left( \inf_{|z| > 1 + \delta/2} s_n \left( P_n - zI \right) < \frac{c'}{2} \text{ and } \left\| P_n - \hat{P}_n \right\| \leq n^{-\varepsilon} \right) + o(1).
\]
Suppose that there exists a \( z_0 \in \mathbb{C} \) with \( |z_0| \geq 1+\delta/2 \) such that \( s_n(P_n - z_0 I) < \frac{c'}{2} \) and \( \|P_n - \hat{P}_n\| < n^{-\varepsilon} < \frac{c'}{2} \). Then, by Weyl's inequality (3.7)

\[
\left| s_n(P_n - z_0 I) - s_n(\hat{P}_n - z_0 I) \right| < \frac{c'}{2}
\]

which implies \( s_n(\hat{P}_n - z_0 I) < c' \). Thus, for \( n \) sufficiently large to ensure that \( n^{-\varepsilon} < \frac{c'}{2} \),

\[
\mathbb{P} \left( \inf_{|z| > 1+\delta/2} s_n(P_n - z I) < \frac{c'}{2} \right)
\leq \mathbb{P} \left( \inf_{|z| > 1+\delta/2} s_n(P_n - z I) < \frac{c'}{2} \quad \text{and} \quad \|P_n - \hat{P}_n\| \leq n^{-\varepsilon} \right) + o(1)
\leq \mathbb{P} \left( \inf_{|z| > 1+\delta/2} s_n(\hat{P}_n - z I) < c' \right) + o(1)
= o(n^{-\alpha}) + o(1)
\]

for any \( \alpha > 0 \). Thus, selecting \( c = \frac{c'}{2} \), we have

\[
\inf_{|z| > 1+\delta/2} s_n(P_n - z I) \geq c
\]

with probability \( 1 - o(1) \). \( \square \)

**Lemma F.0.5.** Let \( A \) be an \( n \times n \) matrix. Let \( R \) be a subset of the integer set \( \{1, 2, \ldots, n\} \). Let \( A^{(R)} \) denote the matrix \( A \), but with the \( r \)th column replaced with zero for each \( r \in R \). Then

\[
s_n \left( A^{(R)} - z I \right) \geq \min \{ s_n(A - z I), |z| \}.
\]

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

\[
(A - z I)^* \ (A - z I), \ (A^{(R)} - z I)^* \ (A^{(R)} - z I),
\]

and \( (A^{(\overline{R})} - z I^{(\overline{R})})^* \ (A^{(\overline{R})} - z I^{(\overline{R})}) \).
Now, observe that \((A^{(R)} - zI^{(R)})^* (A^{(R)} - zI^{(R)})\) is an \((n - |R|) \times (n - |R|)\) matrix, and is a sub-matrix of the Hermitian matrix \((A - zI)^* (A - zI)\). Therefore, the eigenvalues of \((A^{(R)} - zI^{(R)})^* (A^{(R)} - zI^{(R)})\) must interlace with the eigenvalues of \((A - zI)^* (A - zI)\) by Cauchy’s interlacing theorem [77, Theorem 1]. This implies
\[
s_n \left( A^{(R)} - zI^{(R)} \right)^2 \geq s_n (A - zI)^2.
\]

Next, we compare the eigenvalues of \((A^{(R)} - zI)^* (A^{(R)} - zI)\) to the eigenvalues of \((A^{(R)} - zI^{(R)})^* (A^{(R)} - zI^{(R)})\). 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 becomes
\[
\left( A^{(R)} - zI \right)^* \left( A^{(R)} - zI \right) = \begin{bmatrix}
(A^{(R)} - zI^{(R)})^* (A^{(R)} - zI^{(R)}) & 0 \cdot I_{|R| \times (n - |R|)} \\
0 \cdot I_{(n - |R|) \times |R|} & |z|^2 \cdot I_{|R| \times |R|}
\end{bmatrix}.
\]
Due to the block structure of the matrix above, if \(w\) is an eigenvalue of \((A^{(R)} - zI)^* (A^{(R)} - zI)\), then either \(w\) is an eigenvalue of \((A^{(R)} - zI^{(R)})^* (A^{(R)} - zI^{(R)})\) or \(w\) is \(|z|^2\). Ergo,
\[
s_n \left( A^{(R)} - zI \right)^2 = \min \left\{ s_n \left( A^{(R)} - zI^{(R)} \right)^2, |z|^2 \right\}
\geq \min \left\{ s_n (A - zI)^2, |z|^2 \right\}
\]
which implies
\[
s_n \left( A^{(R)} - zI \right) \geq \min \{ s_n (A - zI), |z| \}
\]
concluding the proof.

This lemma gives way to the following two corollaries.

**Corollary F.0.6.** Fix \(\varepsilon > 0\). For a fixed integer \(m > 0\), let \(\xi_1, \xi_2, \ldots, \xi_m\) be real-valued random variables each mean zero, variance one, and finite \(4 + \tau\) moment for some \(\tau > 0\). Let \(X_{n,1}, X_{n,2}, \ldots, X_{n,m}\) be independent iid random matrices with atom variables \(\xi_1, \xi_2, \ldots, \xi_m\) respectively, and define \(\hat{X}_{n,1}, \hat{X}_{n,2}, \ldots, \hat{X}_{n,m}\) as in (5.8). Define \(Y_n\) as in (5.11) and \(Y_n^{(k)}\) as \(Y_n\) with the
columns $c_k, c_{n+k}, c_{2n+k}, \ldots, c_{(m-1)n+k}$ replaced with zeros. Fix $\delta > 0$. Then there exists a constant $c > 0$ depending only on $\delta$ such that

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

with overwhelming probability.

**Proof.** Note that by Lemma F.0.5

$$\inf_{|z| > 1+\delta/2} s_{mn} \left( Y_n^{(k)} - zI \right) \geq \inf_{|z| > 1+\delta/2} \min \{ s_{mn} (Y_n - zI), |z| \} \geq \inf_{|z| > 1+\delta/2} \min \{ s_{mn} (Y_n - zI), 1 \}.$$

By Lemma F.0.2 there exists a constant $c' > 0$ which depends only on $\delta$ such that

$$\inf_{|z| > 1+\delta/2} s_{mn} (Y_n - zI) > c'$$

with overwhelming probability. Therefore,

$$\inf_{|z| > 1+\delta/2} s_{mn} \left( Y_n^{(k)} - zI \right) \geq \min \{ c', 1 \}$$

with overwhelming probability. The result follows by setting $c = \min \{ c', 1 \}$.

**Corollary F.0.7.** Fix $\varepsilon > 0$. For a fixed integer $m > 0$, let $\xi_1, \xi_2, \ldots, \xi_m$ be real-valued random variables each mean zero, variance one, and finite $4 + \tau$ moment for some $\tau > 0$. Let $\hat{X}_{1n}, \hat{X}_{2n}, \ldots, \hat{X}_{mn}$ be independent iid random matrices with atom variables as defined in (5.8).

Define $Y_n$ as in (5.11) and $Y_n^{(k,s)}$ as $Y_n$ with the columns $c_k, c_{n+k}, c_{2n+k}, \ldots, c_{(m-1)n+k}$ and $c_s$ replaced with zeros. Fix a $\delta > 0$. Then there exists a constant $c > 0$ depending only on $\delta$ such that

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

with overwhelming probability.

The proof of Corollary F.0.7 follows in exactly the same way as the proof of Corollary F.0.6.
Appendix G

Proofs of Lemmas 5.4.6 and 5.4.17

Notation used in Appendix G is consistent with Chapter 5.

*Proof of Lemma 5.4.6.* Let $\alpha > 0$ be arbitrary. We will prove that the complementary event holds with probability at most $O(\alpha(n^{-\alpha}))$ uniformly in $z \in \mathcal{C}$. Observe that uniformly in $z \in \mathcal{C}$, by Markov’s inequality and Lemma 5.4.12, for any $p \geq 2$, we have

$$
\mathbb{P}\left( \left\| V^T_k G_n^{(k)}(z) U_k 1_{\Omega_{n,k}} \right\| \geq 1/2 \right) \leq \frac{\mathbb{E} \left\| V^T_k G_n^{(k)}(z) U_k 1_{\Omega_{n,k}} \right\|^{2p}}{(1/2)^{2p}} \ll_p n^{-2p+4\varepsilon-2}.
$$

Since $p$ was arbitrary, selecting $p$ sufficiently large concludes the proof.

*Proof of Lemma 5.4.17.* Let $\alpha > 0$ be arbitrary. We will show the complement event holds with probability at most $O(\alpha(n^{-\alpha}))$ uniformly in $z \in \mathcal{C}$. Observe that uniformly for any $z \in \mathcal{C}$, by Markov’s inequality and Lemma 5.4.17, for any $p \geq 2$, we have

$$
\mathbb{P}\left( \left\| e^T_s G_n^{(k,s)}(z) c_s 1_{\Omega_{n,k,s}} \right\| \geq 1/2 \right) \ll \frac{\mathbb{E} \left\| e^T_s G_n^{(k,s)}(z) c_s 1_{\Omega_{n,k,s}} \right\|^{2p}}{(1/2)^{2p}} \ll_p \mathbb{E} \left\| e^* s (G_n^{(k,s)}(z))^* e_s c_s^T G_n^{(k,s)}(z) c_s 1_{\Omega_{n,k,s}} \right\|^{p} \ll_p n^{-2p+4\varepsilon-2}.
$$

Since $p$ was arbitrary, selecting $p$ sufficiently large concludes the proof.
Appendix H

Proofs of Remark 5.4.13

Notation used in Appendix H is consistent with Chapter 5.

Lemma H.0.1. Let $U_k$ be the $mn \times m$ matrix which contains as its columns $c_k, c_{n+k}, \ldots, c_{(m-1)n+k}$, and define $V_k$ to be the $mn \times m$ matrix which contains as its columns $e_k, e_{n+k}, \ldots, e_{(m-1)n+k}$ where $e_1, \ldots, e_{mn}$ denote the standard basis elements of $\mathbb{C}^{mn}$. Let $G_n^{(k)}(z)$ be defined as in (5.26). Then

$$
\mathbb{E} \left\| V_k^T G_n^{(k)}(z) U_k 1_{\Omega_{n,k}} \right\|_2^2 \ll n^{-1}.
$$

Proof. Observe that

$$
\mathbb{E} \left\| V_k^T G_n^{(k)}(z) U_k 1_{\Omega_{n,k}} \right\|_2^2
\ll \max_{1 \leq i,j \leq m} \mathbb{E} \left| (V_k^T G_n^{(k)}(z) U_k)_{(i,j)} 1_{\Omega_{n,k}} \right|^2
= \max_{1 \leq i,j \leq m} \mathbb{E} \left| e_{(i-1)n+k} G_n^{(k)}(z) c_{(j-1)n+k} 1_{\Omega_{n,k}} \right|^2
= \max_{1 \leq i,j \leq m} \mathbb{E} \left[ c_{(j-1)n+k}^* (G_n^{(k)}(z))^* e_{(i-1)n+k} e_{(i-1)n+k}^T G_n(z) c_{(j-1)n+k} 1_{\Omega_{n,k}} \right]
\ll n^{-1}
$$

as advertised. \qed

Lemma H.0.2. Let $U_k$ be the $mn \times m$ matrix which contains as its columns $c_k, c_{n+k}, \ldots, c_{(m-1)n+k}$, and define $V_k$ to be the $mn \times m$ matrix which contains as its columns $e_k, e_{n+k}, \ldots, e_{(m-1)n+k}$ where
\(e_1, \ldots, e_{mn}\) denote the standard basis elements of \(\mathbb{C}^{mn}\). Let \(G^{(k)}_n(z)\) be defined as in (5.26). Then

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

**Proof.** By the same argument as in the proof of Lemma 1.0.1 above, we have

\[
\mathbb{E} \left\| V_k^T (G_n(z))^2 U_k \mathbf{1}_{\Omega_{n,k}} \right\|^2 \\
\ll \max_{1 \leq i, j \leq m} \mathbb{E} \left[ c_{(j-1)n+k}^* (G^{(k)}_n(z))^{2s} e_{(i-1)n+k} e_{(i-1)n+k}^T (G_n(z))^2 c_{(j-1)n+k} \mathbf{1}_{\Omega_{n,k}} \right] \\
\ll n^{-1}
\]

as advertised. \(\square\)