Title
Local laws of random matrices and their applications

Author
Yang, Fan

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Local laws of random matrices and their applications

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mathematics

by

Fan Yang

2019

ABSTRACT OF THE DISSERTATION

Local laws of random matrices and their applications

by

Fan Yang
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2019
Professor Jun Yin, Chair

This thesis presents new results on spectral statistics of different families of large random matrices. Our main tool is certain types of *local estimates* of the resolvents (or the Green's functions) of the random matrices, which are generally referred to as *local laws*. Utilizing the standard approach developed over the last decade [36] combined with a comparison method developed recently in [59], we are able to prove (almost) optimal local laws for various random matrix ensembles with correlated and heavy-tailed entries. With these local laws, we establish the following three results.

We first study the largest eigenvalues for separable covariance matrices of the form $\mathcal{Q} := A^{1/2}XBX^*A^{1/2}$. Here $X = (x_{ij})$ is an $n \times N$ random matrix, whose entries are *i.i.d.* random variables with mean zero and variance N^{-1}; A and B are respectively $n \times n$ and $N \times N$ deterministic non-negative definite symmetric (or Hermitian) matrices. Under a sharp fourth moment tail condition, we prove that the limiting distribution of the largest eigenvalues of $\mathcal{Q}$ is universal under an $N^{2/3}$ scaling, as long as n/N converges to a finite $d \in (0, \infty)$ as $N \to \infty$. In particular, if $B = I$, then $\mathcal{Q}$ becomes the sample covariance matrix, which is one of the most fundamental objects of study in high-dimensional statistics. Our result provides the strongest edge universality result for large dimensional sample covariance matrices so far.

Then we study the *eigenvector empirical spectral distribution* (VESD)—an important tool in studying the limiting behavior of eigenvectors—for large separable covariance matrices. Under certain low moment assumptions, we prove an optimal convergence rate of the VESD to an anisotropic Marčenko-Pastur law in the metric of Kolmogorov distance. Our results improve the suboptimal convergence rate in [107] under much more relaxed assumptions.

Finally, we study the eigenvalue distribution of a deformed non-Hermitian random matrix ensemble of the form TX, where T is a deterministic $N \times M$ matrix and X is a random $M \times N$ matrix with independent entries, each of which has zero mean and variance $(N \wedge M)^{-1}$. We prove the empirical spectral distribution (ESD) of TX converges to an inhomogeneous local circular law, which is determined by the singular values of T. Moreover, the convergence holds up to the (almost) optimal local scale $(N \wedge M)^{-1/2+\varepsilon}$ for any $\varepsilon > 0$. Our proof depends on a lower tail estimate for the smallest singular value of $TX - z$ for any $z \in \mathbb{C}$. This is also provided in this thesis.

The dissertation of Fan Yang is approved.

Marek Biskup

Georg Menz

Terence Chi-Shen Tao

Jun Yin, Committee Chair

University of California, Los Angeles

2019

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

With regard to this thesis and my graduate study in mathematics, I am most indebted to my advisor, Jun Yin, for his guidance and continuous support. He led me into the exciting field of random matrix theory and raised a lot of interesting questions to me, which have been the starting points of this thesis. From Jun, I have learned a great deal about how to approach research problems and how to think creatively. I could not have imagined having a better mentor for my Ph.D. study in mathematics.

Besides my advisor, I would like to thank the other members of my committee—Marek Biskup, Georg Menz, and Terence Tao, for their insightful comments on this work and helpful suggestions on my future career. I also want to thank Thomas Liggett, who was in my committee for my oral exam at UCLA, and Timo Seppäläinen and Benedek Valkó, who were in my committee for my oral exam at UW-Madison. They unfortunately cannot attend my thesis defense, but their comments have helped me a lot in finishing this work.

I have been fortunate to be a part of the probability groups at UCLA and UW-Madison. I am really grateful to Jun and other probability faculty—Marek Biskup, Thomas Liggett, Georg Menz from UCLA, and David Anderson, Sebastien Roch, Timo Seppäläinen, Benedek Valkó, and Philip Matchett Wood—for making probability groups so active, with numerous seminars and visitors, and so delightful for me. In particular, I am thankful for the excellent course on stochastic analysis given by Timo Seppäläinen, which greatly arouse my interest in probability and finally led me to this direction during the first year of my Ph.D. study.

I would like to thank professors Paul Bourgade, Mark Rudelson, Roman Vershynin, and Horng-Tzer Yau for fruitful discussions and for generously sharing their understanding of many topics in random matrix theory and probability.

I also want to thank my friends and collaborators—Xiucai Ding and Haokai Xi—from whom I have learned a lot of ideas and methods both within and beyond mathematics. Among the friends I made during the Ph.D. years, I particularly want to thank my peers in probability—Nicholas Cook, Wai-Tong (Louis) Fan, Jiaoyang Huang, Wenpin Tang, Qiang Zeng and Yizhe Zhu—and my fellow students—Jingrui Cheng, Jiyuan Han, Hui Jin, Mao Li, Yun Li, Weiyi Liu, Zheng Lu, Chenchen Mou, Ziming Shi, Ruiwen Shu, Jiajun Tong, Bao Wang, Xiaoyao Wei, Tianqi Wu, Dongxi Ye, Penghang Yin, Yuming Zhang. I have learned a lot from the discussions with them, and their support has been important for my Ph.D. years.

Last but not least, my warmest thanks are given to my dear parents for bringing me up and for being such good role models in my life. None of this work can be done without their constant support for my interests and ambitions to purse science as my lifelong career.

VITA

2009	B.S. (Engineering Mechanics), Tsinghua University, Beijing, China.
2014	Ph.D. (Physics), The Chinese University of Hong Kong, Hong Kong. Thesis: Berry phases of quantum trajectories in semiconductors under strong terahertz fields.
2014–Present	Ph.D. Candidate (Mathematics), UCLA, Los Angeles, CA.

PUBLICATIONS

Edge universality of separable covariance matrices. Preprint, arXiv:1809.04572 (2018).

Convergence of eigenvector empirical spectral distribution of sample covariance matrices (with Haokai Xi and Jun Yin). To appear Annals of Statistics.

A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices (with Xiucai Ding). Annals of Applied Probability, Vol. 28 (2018): pp 1679-1738.

The smallest singular value of deformed random rectangular matrices. Preprint, arXiv:1702.04050 (2017).

Local circular law for the product of a deterministic matrix with a random matrix (with Haokai Xi and Jun Yin). Electronic Journal of Probability, Vol. 22 (2017): paper no. 64, 77pp.

CHAPTER 1

Introduction

The study of random matrices dates back to Wishart's study of the so-called Wishart matrices in multivariate statistics [103]. However, his work did not attract much attention at that time. The random matrix theory became influential in the 1950's due to the seminal work [102], where Wigner introduced random matrices to model the energy levels of heavy nuclei. In Wigner's vision, the local spectral statistics of strongly correlated quantum systems are *universal*, and should be given by the random matrix statistics of the same symmetry.

In the last two decades or so, there has been significant progress in understanding Wigner's conjecture regarding the universal behaviors of many different types of random matrix ensembles. One of the important methods is the so-called resolvent method. Given an Hermitian random matrix H, we define its *resolvent* (or *Green's function*) as

$$G(z) := (H - z)^{-1}, \quad z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}.$$

By taking the imaginary part, it is easy to see that a control of $G(E + \mathrm{i}\eta)$ yields a control of the eigenvalue density on a small scale of order η around E (which contains an order ηN eigenvalues). A local law is an estimate of $G(z)$ for all z with $\operatorname{Im} z \gg N^{-1}$. Such local laws has been very powerful tools in studying the local eigenvalue and eigenvector statistics of many random matrix ensembles [36, 90]. The overarching goal of this thesis is to prove certain local laws for random matrices with correlated entries or heavy tails, and use these local laws to derive some universal properties of these matrix ensembles.

This thesis is based on the following papers written by the authors with collaborators:

1. A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices.

2. Edge universality of separable covariance matrices.
3. Convergence of eigenvector empirical spectral distribution of sample covariance matrices.
4. Local circular law for the product of a deterministic matrix with a random matrix.
5. The smallest singular value of deformed random rectangular matrices.

These papers are roughly reproduced in this thesis, with some minor revisions and rearrangements to give a more unified presentation of the results. We now gives a brief review of the subsequent chapters.

1.1 Overview of the thesis

Edge universality of separable covariance matrices

The material in Chapter 2 is based on the author's work in [109], which is also an improvement of the author's previous work [26]. This chapter is concerned with the distribution of the largest eigenvalues of *sample covariance matrices*, or more generally, *separable covariance matrices*.

Sample covariance matrices are fundamental objects in modern multivariate statistics. Given an *i.i.d.* sequence of centered random vectors $\mathbf{y}_i \in \mathbb{R}^n$, $i = 1, \cdots, N$, the sample covariance matrix $\mathcal{Q} := N^{-1}\sum_i \mathbf{y}_i\mathbf{y}_i^*$ is the simplest estimator for the covariance matrix $A := \mathbb{E}\mathbf{y}_1\mathbf{y}_1^*$. In fact, if the dimension n of the data is fixed, then $\mathcal{Q}$ converges almost surely to A as $N \to \infty$. However, in many modern applications, high dimensional data, i.e. data with n being comparable to or even larger than N, is commonly collected in various fields, such as statistics [25, 53, 54, 55], economics [72] and population genetics [75], to name a few. In this setting, A cannot be estimated through Q directly due to the so-called *c*urse of dimensionality. Yet, some properties of A can be inferred from the eigenvalue and eigenvector statistics of $\mathcal{Q}$. Mathematically, the sample covariance matrices can be expressed in the form

$$\mathcal{Q} = A^{1/2}XX^*A^{1/2},$$

where the data matrix $X = (x_{ij})$ is an $n \times N$ random matrix with *i.i.d.* entries satisfying $\mathbb{E}x_{11} = 0$ and $\mathbb{E}|x_{11}|^2 = N^{-1}$.

The data model $A^{1/2}X$ corresponds to observing N independent samples, and hence is incompetent to model data with correlations between different samples. A more general model is the so-called *separable data model* of the form $Y = A^{1/2}XB^{1/2}$, where A and B are respectively $n \times n$ and $N \times N$ deterministic non-negative definite symmetric (or Hermitian) matrices. Without loss of generality, we may assume that the row indices of the data matrix correspond to the spatial locations and the column indices correspond to the observation times. Then A and B describe the *spatial* and *temporal* correlations within the data, and we shall refer to them as the *spatial* and *temporal* covariance matrices, respectively. Here the name "separable" is because the joint covariance of Y, viewed as an (Nn)-dimensional vector, is given by a separable form $A \otimes B$. In particular, if the entries of X are Gaussian, then the joint distribution of Y is $\mathcal{N}_{Nn}(0, A \otimes B)$. Note that the separable model describes a process where the time correlation does not depend on the spatial location and the spatial correlation does not depend on time, i.e. there is no space-time interaction. In fact, the spatio-temporal sampling data is commonly collected in environmental study [44, 60, 63, 68] and wireless communications [99].

In Chapter 2, we consider the separable covariance matrix defined as

$$\mathcal{Q} := YY^* = A^{1/2}XBX^*A^{1/2}.$$

The main result is Theorem 2.2.7, which gives the universality of the largest eigenvalue of $\mathcal{Q}$ under the optimal moment condition. More precisely, assuming $n/N \to d \in (0, \infty)$ as $N \to \infty$, $\mathbb{E}x_{ij}^3 = 0$, and some mild conditions on A and B, we prove that the limiting distribution of the largest eigenvalue of $\mathcal{Q}$ coincide with that of the corresponding Gaussian ensemble (i.e., $\mathcal{Q}$ with X being an *i.i.d.* Gaussian matrix) under a proper $N^{2/3}$ scaling, as long as the following moment condition holds:

$$\lim_{s \to \infty} s^4 \mathbb{P}(\sqrt{N}|x_{ij}| \geqslant s) = 0.$$

This result is commonly referred to as the *edge universality*, in the sense that the limiting distribution of the largest eigenvalue does not depend on the detailed distribution of the entries of X. If we take $B = I$, then $\mathcal{Q}$ becomes the normal sample covariance matrix and the edge universality holds true without the vanishing third moment condition. So far, this is the strongest edge universality result for sample covariance matrices with correlated data (i.e. non-diagonal A) and heavy tails, which improves the previous results in [12, 61] (assuming high moments and diagonal A), [59] (assuming high moments) and [26] (assuming diagonal A).

The proof of Theorem 2.2.7 is based on some (optimal) local laws on the resolvent of $\mathcal{Q}$:

$$\mathcal{G}(X, z) := (\mathcal{Q} - z)^{-1}, \quad z \in \mathbb{C}_+.$$

To prove the local laws, we use two comparison arguments: a self-consistent comparison approach developed in [59], and the Lindeberg replacement strategy with certain four moment matching [62, 93]. For a more detailed introduction, we refer the reader to Section 2.1.

Convergence of eigenvector empirical spectral distribution

The material in Chapter 3 is based on the author's work in [104]. In this chapter, we study the eigenvectors statistics of separable covariance matrix $\mathcal{Q} = A^{1/2}XBX^*A^{1/2}$ through the eigenvector empirical spectral distribution (VESD). Suppose $\mathcal{Q}$ has eigenvalue decomposition

$$\mathcal{Q} = \sum_{k=1}^{n} \lambda_k \xi_k \xi_k^*.$$

Then for any deterministic unit vector $\mathbf{u} \in \mathbb{C}^n$, we define a VESD of $\mathcal{Q}$ as

$$F^{(n)}_{\mathcal{Q},\mathbf{u}}(x) = \sum_{k=1}^{n} |\langle \xi_k, \mathbf{u}\rangle|^2 \mathbf{1}_{\{\lambda_k \leqslant x\}}.$$

The VESD is a useful tool in studying the limiting behavior of eigenvectors of large random matrices. For applications of VESD to separable covariance matrices and sample covariance matrices with spikes, we refer the reader to [7, 86, 88, 104, 106, 107] and Section 3.1.

The Chapter 3 is concerned with the convergence rate of the VESD of separable covariance matrices to certain deterministic distribution, which we shall refer to as the *anisotropic Marčenko-Pastur (MP) law*. Consider separable covariance matrices with diagonal covariance matrices A and B. We will prove that the Kolmogorov distance between the *expected*

VESD and the anisotropic MP distribution is bounded by $N^{-1+\varepsilon}$ for any fixed $\varepsilon > 0$, provided that the entries $\sqrt{N}x_{ij}$ have uniformly bounded 6th moments and $|n/N - 1| \geqslant \tau$ for some constant $\tau > 0$. This result improves the previous one obtained in [107], which gave the convergence rate $\mathrm{O}(N^{-1/2})$ assuming *i.i.d.* X entries, bounded 10th moment, $\Sigma = I$ and $n < N$. Moreover, we also prove that under the finite 8th moment assumption, the convergence rate of the VESD is $\mathrm{O}(N^{-1/2+\varepsilon})$ almost surely for any fixed $\varepsilon > 0$, which improves the previous bound $N^{-1/4+\varepsilon}$ in [107]. The more general cases with non-diagonal A and B are not considered in this work, and will be studied in the future.

Local circular law for deformed non-Hermitian random matrices

The material in Chapter 4 is based on the author's work in [105]. In this chapter, we study the eigenvalue distribution of deformed non-Hermitian random matrices. More precisely, we prove the convergence of the empirical spectral distribution (ESD) of the product TX of a deterministic $N \times M$ matrix T with a random $M \times N$ matrix X, where the entries of X are *i.i.d.* random variables with mean zero.

The study of ESD of non-Hermitian random matrices goes back to the celebrated paper [46], where Ginibre calculated the joint probability density for the eigenvalues of the *i.i.d.* random matrix X with independent complex Gaussian entries. In this case, the joint density distribution is integrable with an explicit kernel, which allowed him to derive the *circular law* for the eigenvalues, i.e. the eigenvalues of X are distributed almost uniformly on a circular disk for large N. However, for general *i.i.d.* random matrices with non-Gaussian entries, there is no explicit formula for the joint distribution of the eigenvalues. To deal with this difficulty, an *Hermitization technique* was developed in [47], where Girko was able to translate the convergence of complex empirical measures of a non-Hermitian matrix into the convergence of logarithmic transforms for a family of Hermitian matrices, or, to be more precise,

$$\mathrm{Tr}\log[(X-z)^*(X-z)] = \log\left[\det((X-z)^*(X-z))\right],$$

for a family of $z \in \mathbb{C}$. With this technique, the circular law was proved for general *i.i.d.*

random matrix ensembles in [6, 8] by assuming bounded density and bounded high moments for the entries of X. The moment and smoothness assumptions were relaxed subsequently in a series of papers by Tao and Vu [91], Pan and Zhou [74] and Götze and Tikhomirov [49]. The final result was presented in [96], where the circular law is proved under the optimal finite variance assumption on the X entries.

In this paper, we study the ESD of the deformed non-Hermitian random matrices of the form TX. We prove an inhomogeneous local circular law for the ESD of TX at any point z away from the boundary circle under the assumption that the matrix entries X_{ij} have sufficiently high moments. More precisely, suppose the boundary circle has radius r. Then if z satisfies $||z| - r| \geqslant \tau$ for some small constant $\tau > 0$, then the ESD of TX converges to $\tilde{\chi}_{\mathbb{D}}(z)\mathrm{d}A(z)$, where $\tilde{\chi}_{\mathbb{D}}$ is a rotation-invariant function determined by the singular values of T and $\mathrm{d}A$ denotes the Lebesgue measure on the disk $\{z : |z| \leqslant r\}$. The local circular law is valid around z up to the almost optimal scale $(N \wedge M)^{-1/2+\varepsilon}$ for any fixed $\varepsilon > 0$.

The main tool for our proof is an optimal local law for the resolvents of the family of Hermitian matrices $(TX - z)^*(TX - z)$. In the proof, we need some stability estimates on certain deterministic self-consistent equations, which are proved in Appendix A. Moreover, due to the singularity of the log function at 0, we also need to bound the smallest eigenvalue of $TX - z$ from below for any $z \in \mathbb{C}$. This was provided in Appendix B, which we shall introduce next.

The smallest singular value

The material in Appendix B is based on the author's work in [108]. The purpose of this appendix is to prove a lower tail estimate for the smallest singular value of $TX - z$, which is used in the proof in Section 4. However, motivated by potential applications in statistics, we shall consider a slightly more general deformed random matrix model of the form $TX - A$. Here X is an $M \times n$ random matrix with *i.i.d.* entries, which have zero mean, unit variance and arbitrarily high moments; T is an $N \times M$ deterministic matrix with comparable singular values $c \leqslant s_N(T) \leqslant s_1(T) \leqslant c^{-1}$ for some constant $c > 0$; A is an $N \times n$ deterministic matrix with $\|A\| = \mathrm{O}(\sqrt{N})$. Suppose $n \leqslant N \leqslant M = \mathrm{O}(N)$. Then we prove that for any constant $\varepsilon > 0$, the smallest singular value of $TX - A$ is larger than $N^{-\varepsilon}(\sqrt{N} - \sqrt{n-1})$ with high probability. If we assume further the entries of X have subgaussian decay, then the smallest singular value of $TX - A$ is at least of the order $\sqrt{N} - \sqrt{n-1}$ with high probability, which is an essentially optimal estimate. Our proof is based on an extension of the arguments in [82].

This appendix is relatively independent of the other parts of this thesis, and can be read separately.

1.2 Conventions

The fundamental large parameter in this thesis is N. All quantities that are not explicitly constant may depend on N, and we usually omit N from our notations. We use C to denote a generic large positive constant, which may depend on fixed parameters and whose value may change from one line to the next. Similarly, we use ε, τ, δ, ω and c to denote generic small positive constants. If a constant depend on a quantity a, we use $C(a)$ or C_a to indicate this dependence.

For two quantities a_N and b_N depending on N, the notation $a_N = \mathrm{O}(b_N)$ means that $|a_N| \leqslant C|b_r|$ for some constant $C > 0$, and $a_N = \mathrm{o}(b_N)$ means that $|a_N| \leqslant c_N|b_N|$ for some positive sequence $c_N \downarrow 0$ as $N \to \infty$. We also use the notations $a_N \lesssim b_N$ if $a_N = \mathrm{O}(b_N)$, and $a_N \sim b_N$ if $a_N = \mathrm{O}(b_N)$ and $b_N = \mathrm{O}(a_N)$.

For any (complex) matrix A, we use A^* to denote its conjugate transpose, A^T the transpose, $\|A\| := \|A\|_{l^2 \to l^2}$ the operator norm and $\|A\|_{HS}$ the Hilbert-Schmidt norm. In this thesis, we usually write the $N \times N$ identity matrix I_N as I or 1 without causing any confusions.

We use the notation $\mathbf{v} = (v_i)_{i=1}^N$ for a vector in $\mathbb{C}^N$, and denote its Euclidean norm by $|\mathbf{v}| \equiv \|\mathbf{v}\|_2$. The canonical inner product on $\mathbb{R}^n$ or $\mathbb{C}^n$ is denoted $\langle \cdot, \cdot \rangle$. The distance from a point x to a set D in $\mathbb{R}^n$ is denoted $\mathrm{dist}(x, D)$.

If two random variables X and Y have the same distribution, we write $X \overset{d}{=} Y$.

CHAPTER 2

Edge universality of separable covariance matrices

2.1 Introduction

In this chapter, we focus on the limiting distribution of the largest eigenvalues of high-dimensional sample covariance matrices, which is of great interest to the principal component analysis. The largest eigenvalue has been widely used in hypothesis testing problems on the structure of covariance matrices, see e.g. [13, 28, 54, 73]. Of course the list is very far from being complete, and we refer the reader to [53, 77, 110] for a comprehensive review. Precisely, we will consider sample covariance matrices of the form

$$\mathcal{Q} = A^{1/2} X X^* A^{1/2},$$

where the data matrix $X = (x_{ij})$ is an $n \times N$ random matrix with *i.i.d.* entries such that $\mathbb{E}x_{11} = 0$ and $\mathbb{E}|x_{11}|^2 = N^{-1}$, and A is an $n \times n$ deterministic non-negative definite symmetric (or Hermitian) matrix. On dimensionality, we assume that $n/N \to d \in (0, \infty)$ as $N \to \infty$. It is well-known that the empirical spectral distribution (ESD) of $\mathcal{Q}$ converges to the (deformed) Marchenko-Pastur (MP) law [66], whose rightmost edge λ_r gives the asymptotic location of the largest eigenvalue. Moreover, it was proved in a series of papers that under an $N^{2/3}$ scaling, the distribution of the largest eigenvalue $\lambda_1(\mathcal{Q})$ around λ_r converges to the Tracy-Widom distribution [97, 98], which arises as the limiting distribution of the rescaled largest eigenvalue of the Gaussian orthogonal (or unitary) ensemble. This result is commonly referred to as the *edge universality*, in the sense that it is independent of the detailed distribution of the entries of X. The limiting distribution of λ_1 was first obtained for $\mathcal{Q}$ with X consisting of *i.i.d.* centered Gaussian entries (i.e. X is a Wishart matrix) and with trivial covariance (i.e. $A = I$) [54]. The edge universality in the $A = I$ case was later proved for all random matrices X whose entries satisfy a sub-exponential decay [79]. When A is a non-scalar diagonal matrix, the Tracy-Widom distribution was first proved for Wishart matrix X in [28] (non-singular A case) and [71] (singular A case). Later the edge universality with general diagonal A was proved in [12, 61] for X with entries having arbitrarily high moments, and in [26] for X with entries satisfying the tail condition (2.1.1). The most general case with non-diagonal A is considered in [59], where the edge universality was proved under the arbitrarily high moments assumption.

A generalization of the sample covariance matrix model is the so-called *separable covariance matrix*, which are of the form $\mathcal{Q} := YY^* = A^{1/2}XBX^*A^{1/2}$. It has been proved to be very useful for various applications. For example, in wireless communications, it was shown in [100] that an estimate of the capacity is directly given by various informations of the largest eigenvalue. The spectral properties of separable covariance matrices have been investigated in some recent works, see e.g. [20, 29, 78, 101, 113]. However, the edge universality is much less known compared with sample covariance matrices. It is known that the edge universality generally follows from an optimal local law for the resolvent $G = (\mathcal{Q} - z)^{-1}$ near the spectral edge, where $z \in \mathbb{C}_+$ with $\operatorname{Im} z \gg N^{-1}$ [12, 26, 59, 61]. Consider an $n \times N$ matrix X consisting of independent centered entries with general variance profile $\mathbb{E}|x_{ij}|^2 = \sigma_{ij}/N$, then an optimal local law was prove in [1, 2] for the resolvent $(XX^* - z)^{-1}$ under the arbitrarily high moments assumption. Note that this gives the local law for G in the case where both A and B are diagonal. However, if A and B are not diagonal, no such local law is proved so far, let alone the edge universality.

In this thesis, we try to fill this gap. We shall prove that for general (non-diagonal) A and B satisfying some mild assumptions, the limiting distribution of the rescaled largest eigenvalue $N^{\frac{2}{3}}(\lambda_1(\mathcal{Q}) - \lambda_r)$ coincides with that of the corresponding Gaussian ensemble (i.e. $\mathcal{Q}^G = A^{1/2}X^GB(X^G)^*A^{1/2}$ with X^G being a Wishart matrix) as long as the following conditions hold:

$$\lim_{s\to\infty} s^4\mathbb{P}\left(|\sqrt{N}x_{11}| \geqslant s\right) = 0, \tag{2.1.1}$$

and

$$\mathbb{E}x_{11}^3 = 0. \tag{2.1.2}$$

For a precise statement, the reader can refer to Theorem 2.2.7. Note that the tail condition (2.1.1) is slightly weaker than the finite fourth moment condition for $\sqrt{N}x_{11}$, and in fact is sharp for the edge universality of the largest eigenvalue, see Remark 2.2.8. Historically, for sample covariance matrices, it was proved in [112] that $\lambda_1 \to \lambda_r$ almost surely in the null case with $A = I$ if the fourth moment exists. Later the finite fourth moment condition is proved to be also necessary for the almost sure convergence of λ_1 [4]. On the other hand, it was proved in [87] that $\lambda_1 \to \lambda_r$ in probability under the condition (2.1.1). If A is diagonal, it was proved in the atuhor's work [26] that the condition (2.1.1) is actually necessary and sufficient for the edge universality of sample covariance matrices to hold.

On the other hand, the condition (2.1.2) is more technical and should be considered to be removed in the future. We now discuss about it briefly. The main difficulty in studying $\mathcal{Q} = A^{1/2}XBX^*A^{1/2}$ and its resolvent is due to the fact that the entries of $A^{1/2}XB^{1/2}$ are not independent. We assume that A and B have eigendecompositions $A = U\Sigma U^*$ and $B = V\widetilde{\Sigma}V^*$. Then in the special case where $X \equiv X^G$ is Wishart, it is easy to see that

$$A^{1/2}X^G B(X^G)^* A^{1/2} \stackrel{d}{=} U\left(\Sigma^{1/2}X^G\widetilde{\Sigma}^{1/2}\right)U^* \sim \Sigma^{1/2}X^G\widetilde{\Sigma}^{1/2}, \tag{2.1.3}$$

which is reduced to a separable covariance matrix with diagonal Σ and $\widetilde{\Sigma}$. This case can be handled using the current method in [26]. To extend the result in the Gaussian case to the general X case, we use a self-consistent comparison argument developed in [59]. For this argument to work, we need to assume that the third moments of the X entries coincide with that of the Gaussian random variable, i.e. the condition (2.1.2). (Actually it is common that for a comparison argument to work for random matrices, some kind of four moment matching is needed; see e.g. [92, 93, 95].) If one of the A and B is diagonal, then a notable argument in [59, Section 8] can remove this requirement by exploring more detailed structures of the resolvents of $\mathcal{Q}$. However, their argument is quite specific and cannot be adapted to the general case with both A and B being non-diagonal. Nevertheless, this is still a welcome result, which shows that for sample covariance matrices, the condition (2.1.2) is not necessary and the edge universality holds as long as (2.1.1) holds.

Finally, we believe that the largest eigenvalue of the Gaussian separable covariance matrix

$\mathcal{Q}^G$ should converge to the Tracy-Widom distribution. However, to the best of our knowledge, so far there is no explicit proof for this fact. This will be studied in the future.

The rest of the chapter is organized as follows. In Section 2.2, we first define the limiting spectral distribution of the separable covariance matrix and its rightmost edge λ_r, which will depend only on the empirical spectral densities (ESD) of A and B. Then we will state the main theorem—Theorem 2.2.7— of this chapter. In Section 2.4, we introduce the notations and collect some tools including the *anisotropic local law* (Theorem 2.4.6), *rigidity of eigenvalues* (Theorem 2.4.8) and a comparison theorem (Theorem 2.4.10). In Section 2.5, we prove Theorem 2.2.7 with these tools. Then Section 2.6 and Section 2.7 are devoted to proving Theorem 2.4.6, Section 2.8.1 is devoted to proving Theorem 2.4.8, and Section 2.8.2 contains the proof for Theorem 2.4.10.

2.2 Definitions and Main Result

2.2.1 Separable covariance matrices

We consider a class of separable covariance matrices of the form $\mathcal{Q}_1 := A^{1/2}XBX^*A^{1/2}$, where A and B are deterministic non-negative definite symmetric (or Hermitian) matrices. Note that A and B are not necessarily diagonal. We assume that $X = (x_{ij})$ is an $n \times N$ random matrix with entries $x_{ij} = N^{-1/2}q_{ij}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant N$, where q_{ij} are *i.i.d.* random variables satisfying

$$\mathbb{E}q_{11} = 0, \qquad \mathbb{E}|q_{11}|^2 = 1. \tag{2.2.1}$$

For definiteness, in this chapter we focus on the real case, i.e. the random variable q_{11} is real. However, we remark that our proof can be applied to the complex case after minor modifications if we assume in addition that Re q_{11} and Im q_{11} are independent centered random variables with variance 1/2. We will also use the $N \times N$ matrix $\mathcal{Q}_2 := B^{1/2}X^*AXB^{1/2}$. We assume that the aspect ratio $d_N := n/N$ satisfies $\tau \leqslant d_N \leqslant \tau^{-1}$ for some constant $0 < \tau < 1$. Without loss of generality, by switching the roles of $\mathcal{Q}_1$ and $\mathcal{Q}_2$ if necessary, we can assume

that

$$\tau \leqslant d_N \leqslant 1 \quad \text{for all } N. \tag{2.2.2}$$

For simplicity of notations, we will often abbreviate d_N as d in this chapter. We denote the eigenvalues of $\mathcal{Q}_1$ and $\mathcal{Q}_2$ in descending order by $\lambda_1(\mathcal{Q}_1) \geqslant \ldots \geqslant \lambda_n(\mathcal{Q}_1)$ and $\lambda_1(\mathcal{Q}_2) \geqslant \ldots \geqslant \lambda_N(\mathcal{Q}_2)$. Since $\mathcal{Q}_1$ and $\mathcal{Q}_2$ share the same nonzero eigenvalues, we will for simplicity write λ_j, $1 \leqslant j \leqslant N \wedge n$, to denote the j-th eigenvalue of both $\mathcal{Q}_1$ and $\mathcal{Q}_2$ without causing any confusion.

We assume that A and B have eigendecompositions

$$A = U\Sigma U^*, \quad B = V\widetilde{\Sigma}V^*, \quad \Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n), \quad \widetilde{\Sigma} = \operatorname{diag}(\widetilde{\sigma}_1, \ldots, \widetilde{\sigma}_N), \tag{2.2.3}$$

where

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_n \geqslant 0, \quad \widetilde{\sigma}_1 \geqslant \widetilde{\sigma}_2 \geqslant \ldots \geqslant \widetilde{\sigma}_N \geqslant 0.$$

We denote the empirical spectral densities (ESD) of A and B by

$$\pi_A \equiv \pi_A^{(n)} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\sigma_i}, \quad \pi_B \equiv \pi_B^{(N)} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\widetilde{\sigma}_i}. \tag{2.2.4}$$

We assume that there exists a small constant $0 < \tau < 1$ such that for all N large enough,

$$\max\{\sigma_1, \widetilde{\sigma}_1\} \leqslant \tau^{-1}, \quad \max\left\{\pi_A^{(n)}([0,\tau]), \pi_B^{(N)}([0,\tau])\right\} \leqslant 1 - \tau. \tag{2.2.5}$$

Note the first condition means that the operator norms of A and B are bounded by τ^{-1}, and the second condition means that the spectrums of A and B cannot concentrate at zero.

We summarize our basic assumptions here for future reference.

Assumption 2.2.1. *We assume that X is an $n \times N$ random matrix with real i.i.d. entries satisfying (2.2.1), A and B are deterministic non-negative definite symmetric matrices satisfying* (2.2.3) *and (2.2.5), and d_N satisfies* (2.2.2).

2.2.2 Resolvents and limiting law

We will study the eigenvalue statistics of $\mathcal{Q}_1$ and $\mathcal{Q}_2$ through their *resolvents* (or *Green's functions*). It is equivalent to study the matrices

$$\widetilde{\mathcal{Q}}_1(X) := \Sigma^{1/2}U^*XBX^*U\Sigma^{1/2}, \quad \widetilde{\mathcal{Q}}_2(X) := \widetilde{\Sigma}^{1/2}V^*X^*AXV\widetilde{\Sigma}^{1/2}. \tag{2.2.6}$$

In this thesis, we shall denote the upper half complex plane and the right half real line by

$$\mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}, \quad \mathbb{R}_+ := [0, \infty).$$

Definition 2.2.2 (Resolvents). *For $z = E + \mathrm{i}\eta \in \mathbb{C}_+$, we define the resolvents for $\widetilde{\mathcal{Q}}_{1,2}$ as*

$$\mathcal{G}_1(X, z) := \left(\widetilde{\mathcal{Q}}_1(X) - z\right)^{-1}, \quad \mathcal{G}_2(X, z) := \left(\widetilde{\mathcal{Q}}_2(X) - z\right)^{-1}. \tag{2.2.7}$$

We denote the ESD $\rho^{(n)}$ of $\widetilde{\mathcal{Q}}_1$ and its Stieltjes transform as

$$\rho \equiv \rho^{(n)} := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(\widetilde{\mathcal{Q}}_1)}, \quad m(z) \equiv m^{(n)}(z) := \int \frac{1}{x - z}\rho_1^{(n)}(dx) = \frac{1}{n}\operatorname{Tr}\mathcal{G}_1(z). \tag{2.2.8}$$

We also introduce the following quantities:

$$m_1(z) \equiv m_1^{(n)}(z) := \frac{1}{N}\sum_{i=1}^{n} \sigma_i(\mathcal{G}_1)_{ii}(z), \quad m_2(z) \equiv m_2^{(N)}(x) := \frac{1}{N}\sum_{\mu=1}^{N} \widetilde{\sigma}_\mu(\mathcal{G}_2)_{\mu\mu}(z).$$

It was shown in [78] that if $d_N \to d \in (0, \infty)$ and $\pi_A^{(n)}$, $\pi_B^{(N)}$ converge to certain probability distributions, then almost surely $\rho^{(n)}$ converges to a deterministic distributions ρ_∞. We now describe it through the Stieltjes transform

$$m_\infty(z) := \int_{\mathbb{R}} \frac{\rho_\infty(dx)}{x - z}, \quad z \in \mathbb{C}_+.$$

For any finite N and $z \in \mathbb{C}_+$, we define $(m_{1c}^{(N)}(z), m_{2c}^{(N)}(z)) \in \mathbb{C}_+^2$ as the unique solution to the system of self-consistent equations

$$m_{1c}^{(n)}(z) = d_N \int \frac{x}{-z\left[1 + x m_{2c}^{(N)}(z)\right]}\pi_A^{(n)}(dx), \quad m_{2c}^{(N)}(z) = \int \frac{x}{-z\left[1 + x m_{1c}^{(N)}(z)\right]}\pi_B^{(N)}(dx). \tag{2.2.9}$$

Then we define

$$m_c(z) \equiv m_c^{(n)}(z) := \int \frac{1}{-z\left[1 + x m_{2c}^{(N)}(z)\right]}\pi_A^{(n)}(dx). \tag{2.2.10}$$

It is easy to verify that $m_c^{(n)}(z) \in \mathbb{C}_+$ for $z \in \mathbb{C}_+$. Letting $\eta \downarrow 0$, we can obtain a probability measure $\rho_c^{(n)}$ with the inverse formula

$$\rho_c^{(n)}(E) = \lim_{\eta \downarrow 0} \frac{1}{\pi}\operatorname{Im} m_c^{(n)}(E + \mathrm{i}\eta). \tag{2.2.11}$$

If $d_N \to d \in (0,\infty)$ and $\pi_A^{(n)}$, $\pi_B^{(N)}$ converge to certain probability distributions, then $m_c^{(n)}$ also converges and we define

$$m_\infty(z) := \lim_{N\to\infty} m_c^{(n)}(z), \quad z \in \mathbb{C}_+.$$

Letting $\eta \downarrow 0$, we can recover the asymptotic eigenvalue density ρ_∞ with

$$\rho_\infty(E) = \lim_{\eta\downarrow 0} \frac{1}{\pi} \operatorname{Im} m_\infty(E + \mathrm{i}\eta). \tag{2.2.12}$$

It is also easy to see that ρ_∞ is the weak limit of $\rho_c^{(n)}$.

The above definitions of $m_c^{(n)}$, $\rho_c^{(n)}$, m_∞ and ρ_∞ make sense due to the following theorem. Throughout the rest of this chapter, we often omit the super-indices (n) and (N) from our notations.

Theorem 2.2.3 (Existence, uniqueness, and continuous density)**.** *For any $z \in \mathbb{C}_+$, there exists a unique solution $(m_{1c}, m_{2c}) \in \mathbb{C}_+^2$ to the systems of equations in (2.2.9). The function m_c in (2.2.10) is the Stieltjes transform of a probability measure μ_c supported on $\mathbb{R}_+$. Moreover, μ_c has a continuous derivative $\rho_c(x)$ on $(0,\infty)$, which is defined by* (2.2.12)*.*

Proof. See [113, Theorem 1.2.1], [51, Theorem 2.4] and [22, Theorem 3.1]. □

We now make a small detour and discuss about another very enlightening way to understand the Stieltjes transforms $m_{1,2c}$ and m_c. Consider the vector solution $\mathbf{v} = (v_1, \cdots, v_n)$ to the following self-consistent vector equation [1, 2]:

$$\frac{1}{\mathbf{v}(z)} = -z + S\frac{1}{1 + S^T\mathbf{v}(z)}, \quad z \in \mathbb{C}_+, \tag{2.2.13}$$

where $1/\mathbf{v}$ denotes the entrywise reciprocal, and S is an $n \times N$ matrix with entries

$$S_{i\mu} = \frac{1}{N}\sigma_i\widetilde{\sigma}_\mu, \quad i \in [\![1,n]\!], \quad \mu \in [\![1,N]\!]. \tag{2.2.14}$$

In fact, if one regards $\mathcal{X}_1 := [\![1,n]\!]$ and $\mathcal{X}_2 := [\![1,N]\!]$ as measure spaces equipped with counting measures

$$\pi_1 = \sum_{i=1}^n \delta_i, \quad \pi_2 = \sum_{\mu=1}^N \delta_\mu,$$

then S defines a linear operator $S : l^\infty(\mathcal{X}_2) \to l^\infty(\mathcal{X}_1)$ such that

$$(S\mathbf{w})_i = \frac{\sigma_i}{N}\sum_{\mu=1}^{N} \widetilde{\sigma}_\mu w_\mu, \quad \mathbf{w} \in l^\infty(\mathcal{X}_2), \quad i \in \mathcal{X}_1.$$

Now we can regard (2.2.13) as a self-consistent equation of the function $\mathbf{v} : \mathbb{C}_+ \to l^\infty(\mathcal{X}_1)$. Suppose $\mathbf{v}$ is a solution to (2.2.13) with $\operatorname{Im} \mathbf{v}(z) > 0$, then it is easy to verify that

$$m_{1c} = \frac{1}{N}\sum_{i=1}^{n} \sigma_i v_i, \quad m_{2c} = \frac{1}{N}\sum_{\mu=1}^{N} \frac{\widetilde{\sigma}_\mu}{-z(1+\widetilde{\sigma}_\mu m_{1c})}, \quad m_c = \frac{1}{n}\sum_{i=1}^{n} v_i.$$

The structure of the solution $\mathbf{v}$ was well-studied in [1, 2]. In particular, one has the following preliminary result on the existence and uniqueness of the solution.

Theorem 2.2.4 (Proposition 2.1 of [1])**.** *There is a unique function $\mathbf{v} : \mathbb{C}_+ \to l^\infty(\mathcal{X}_1)$ satisfying* (2.2.13) *and $\operatorname{Im} \mathbf{v}(z) > 0$ for all $z \in \mathbb{C}_+$. Moreover, for each $k \in \mathcal{X}_1$, there is a unique probability measure μ_k on $\mathbb{R}$ such that v_k is the Stieltjes transform of μ_k, i.e.*

$$v_k(z) = \int_0^\infty \frac{1}{E-z}\mu_k(\mathrm{d}E), \quad z \in \mathbb{C}_+.$$

The measures μ_k, $k \in \mathcal{X}_1$, all have the same support contained in $[0, A]$, where

$$A := 4\max\left\{\|S\|_{l^\infty(\mathcal{X}_2)\to l^\infty(\mathcal{X}_1)}, \|S^*\|_{l^\infty(\mathcal{X}_1)\to l^\infty(\mathcal{X}_2)}\right\}.$$

Now we go back to study the equations in (2.2.9). If we define the function

$$f(z,m) := -m + \int \frac{x}{-z + xd_N\int \frac{t}{1+tm}\pi_A(dt)}\pi_B(dx), \tag{2.2.15}$$

then $m_{2c}(z)$ can be characterized as the unique solution to the equation $f(z,m) = 0$ with $\operatorname{Im} m > 0$, and $m_{1c}(z)$ is defined using the first equation in (2.2.9). Moreover, $m_{1,2c}(z)$ are the Stieltjes transforms of densities $\rho_{1,2c}$:

$$\rho_{1,2c}(E) = \lim_{\eta\downarrow 0}\frac{1}{\pi}\operatorname{Im} m_{1,2c}(E + \mathrm{i}\eta). \tag{2.2.16}$$

Then we have the following result.

Lemma 2.2.5. *The densities ρ_c and $\rho_{1,2c}$ all have the same support on $(0,\infty)$, which is a union of intervals:*

$$supp\,\rho_c \cap (0,\infty) = supp\,\rho_{1,2c} \cap (0,\infty) = \bigcup_{k=1}^{p}[a_{2k}, a_{2k-1}] \cap (0,\infty), \tag{2.2.17}$$

where $p \in \mathbb{N}$ depends only on $\pi_{A,B}$. Moreover, $(x,m) = (a_k, m_{2c}(a_k))$ are the real solutions to the equations

$$f(x,m) = 0, \quad \text{and} \quad \frac{\partial f}{\partial m}(x,m) = 0. \tag{2.2.18}$$

Moreover, we have $m_{1c}(a_1) \in (-\widetilde{\sigma}_1^{-1}, 0)$ and $m_{2c}(a_1) \in (-\sigma_1^{-1}, 0)$.

Proof. See Section 3 of [22]. □

We shall call a_k the spectral edges. In particular, we will focus on the rightmost edge $\lambda_r := a_1$. Now we make the following assumption: there exists a constant $\tau > 0$ such that

$$1 + m_{1c}(\lambda_r)\widetilde{\sigma}_1 \geqslant \tau, \quad 1 + m_{2c}(\lambda_r)\sigma_1 \geqslant \tau. \tag{2.2.19}$$

This assumption guarantees a regular square-root behavior of the spectral densities $\rho_{1,2c}$ near λ_r as shown by the following lemma.

Lemma 2.2.6. *Under the assumptions* (2.2.2)*,* (2.2.5) *and* (2.2.19)*, there exist constants $a_{1,2} > 0$ such that*

$$\rho_{1,2c}(\lambda_r - x) = a_{1,2}x^{1/2} + \mathrm{O}(x), \quad x \downarrow 0, \tag{2.2.20}$$

and

$$m_{1,2c}(z) = m_{1,2c}(\lambda_r) + \pi a_{1,2}(z-\lambda_r)^{1/2} + \mathrm{O}(|z-\lambda_r|), \quad z \to \lambda_r, \quad \operatorname{Im} z \geqslant 0. \tag{2.2.21}$$

The estimates (2.2.20) *and* (2.2.21) *also hold for ρ_c and m_c with a different constant.*

Proof. Differentiating the equation $f(z,m) = 0$ with respect to m, we can get that $z'(m_r) = 0$ and $z''(m_r) = -\partial_m^2 f(\lambda_r, m_r)/\partial_z f(\lambda_r, m_r)$, where $m_r := m_{2c}(\lambda_r)$. After a straightforward calculation, we have

$$\partial_z f(z,m) = \int \frac{x}{z^2\left[1 + xg(z,m)\right]^2}\pi_B(dx), \quad g(z,m) := d_N \int \frac{t}{-z\left(1+tm\right)}\pi_A(dt),$$

and

$$\partial_m^2 f(z,m) = -2\int \frac{x^3}{z\left[1+xg(z,m)\right]^3}\left(\partial_m g(z,m)\right)^2 \pi_B(dx) + \int \frac{x^2}{z\left[1+xg(z,m)\right]^2}\partial_m^2 g(z,m)\pi_B(dx),$$

where

$$\partial_m g(z,m) = d_N \int \frac{t^2}{z\,(1+tm)^2}\pi_A(dt), \quad \partial_m^2 g(z,m) = -2d_N \int \frac{t^3}{z\,(1+tm)^3}\pi_A(dt).$$

Using (2.2.5) and (2.2.19), it is easy to show that

$$|\partial_z f(\lambda_r, m_r)| \sim 1, \quad \left|\partial_m^2 f(\lambda_r, m_r)\right| \sim 1.$$

Thus we have $|z''(m_r)| \sim 1$, which by Theorem 3.3 of [22], implies (2.2.20) and (2.2.21) for ρ_{2c} and m_{2c}. The estimates for ρ_{1c}, m_{1c}, ρ_c, and m_c then follow from simple applications of (2.2.9) and (2.2.10). □

2.2.3 Main result

The main result of this chapter is the following theorem.

Theorem 2.2.7. *Let $\mathcal{Q}_1 := A^{1/2}XBX^*A^{1/2}$ be an $n \times n$ separable covariance matrix with A, B and X satisfying Assumption 2.2.1 and* (2.2.19). *Let λ_1 be the largest eigenvalue of $\mathcal{Q}_1$. If the conditions (2.1.1) and* (2.1.2) *hold, then we have*

$$\lim_{N\to\infty} \mathbb{P}(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s) = \lim_{N\to\infty} \mathbb{P}^G(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s) \tag{2.2.22}$$

for all $s \in \mathbb{R}$, where $\mathbb{P}^G$ denotes the law for X with i.i.d. Gaussian entries. The condition (2.1.2) *is not necessary if A or B is diagonal.*

Remark 2.2.8. The moment condition is actually sharp in the following sense. If the condition (2.1.1) does not hold for X, then one can show that (see e.g. [26, Section 4]) for any fixed $a > \lambda_r$,

$$\limsup_{N\to\infty} \mathbb{P}\left(\lambda_1(XX^*) \geqslant a\right) > 0,$$

where $\lambda_1(XX^*)$ denotes the largest eigenvalue of XX^*. Thus if $\min\{\sigma_n, \widetilde{\sigma}_N\} \geqslant \tau$ for some constant $\tau > 0$, we then have

$$\limsup_{N\to\infty} \mathbb{P}\left(\lambda_1(\mathcal{Q}_1) \geqslant a\right) > 0$$

for any fixed $a > \lambda_r$, and the edge universality (2.2.22) cannot hold.

Remark 2.2.9. It is clear that (2.2.22) gives the edge universality of the largest eigenvalues of separable covariance matrices. However, to the best of our knowledge, so far there is no explicit proof for the limiting distribution of the largest eigenvalue of $\mathcal{Q}_1$ when X is Gaussian. We will handle this problem in the future, i.e. we will show that there exists $\gamma_0 \equiv \gamma_0(N)$ depending only on $\pi_{A,B}$ and the aspect ratio d_N such that

$$\lim_{N\to\infty} \mathbb{P}^G \left(\gamma_0 N^{2/3}(\lambda_1 - \lambda_r) \leqslant s\right) = F_1(s), \quad s \in \mathbb{R},$$

where F_1 is the type-1 Tracy-Widom distribution. Thus (2.2.22) in fact shows that the distribution of the rescaled largest eigenvalue of $\mathcal{Q}_1$ converges to the Tracy-Widom distribution if the conditions (2.1.1) and (2.1.2) hold. In particular, in the case of sample covariance matrices, the condition (2.1.2) is not necessary. Hence we conclude that the rescaled largest eigenvalue of a sample covariance matrix with correlated rows converges to the Tracy-Widom distribution if the tail condition (2.1.1) holds.

Remark 2.2.10. The universality result (2.2.22) can be extended to the joint distribution of the k largest eigenvalues for any fixed k:

$$\lim_{N\to\infty} \mathbb{P}\left(\left(N^{2/3}(\lambda_i - \lambda_r) \leqslant s_i\right)_{1\leqslant i\leqslant k}\right) = \lim_{N\to\infty} \mathbb{P}^G \left(\left(N^{2/3}(\lambda_i - \lambda_r) \leqslant s_i\right)_{1\leqslant i\leqslant k}\right), \tag{2.2.23}$$

for all $s_1, s_2, \ldots, s_k \in \mathbb{R}$. Let H^{GOE} be an $N \times N$ random matrix belonging to the Gaussian orthogonal ensemble. The joint distribution of the k largest eigenvalues of H^{GOE}, $\mu_1^{GOE} \geqslant \ldots \geqslant \mu_k^{GOE}$, can be written in terms of the Airy kernel for any fixed k [42]. In the future, we will prove that

$$\lim_{N\to\infty} \mathbb{P}^G \left(\left(\gamma_0 N^{2/3}(\lambda_i - \lambda_r) \leqslant s_i\right)_{1\leqslant i\leqslant k}\right) = \lim_{N\to\infty} \mathbb{P}\left(\left(N^{2/3}(\mu_i^{GOE} - 2) \leqslant s_i\right)_{1\leqslant i\leqslant k}\right),$$

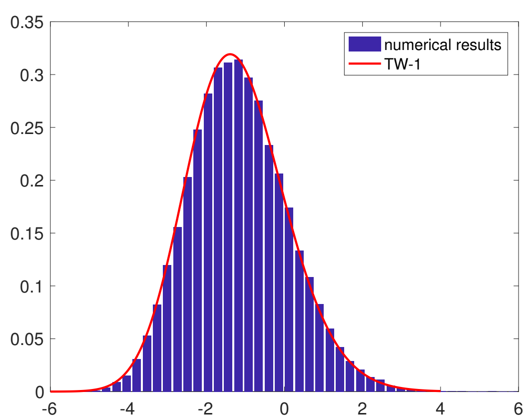
for all $s_1, s_2, \ldots, s_k \in \mathbb{R}$. Hence (2.2.23) gives a complete description of the finite-dimensional correlation functions of the largest eigenvalues of $\mathcal{Q}_1$.

Remark 2.2.11. A key input for the proof of (2.2.22) is the anisotropic local law for the resolvents in (2.2.7). Our basic strategy is first to prove the anisotropic local law for $\mathcal{G}_{1,2}$ when X is Gaussian, and then to obtain the anisotropic local law for a general X through a comparison with the Gaussian case. Without (2.1.2), the comparison argument cannot give the anisotropic local law up to the optimal scale. However, in the case where A or B is diagonal, the condition (2.1.2) is not needed for the comparison argument in [59] to work. We will try to remove the assumption (2.1.2) completely in future works.

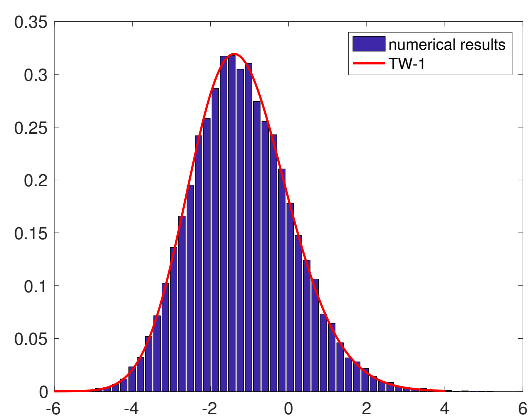
Finally, we illustrate the edge universality result with some numerical simulations. Consider the following setting: (1) $N = 2n$, i.e. $d_N = 0.5$; (2) we take

$$\Sigma = \text{diag}(\underbrace{1, \cdots, 1}_{n/2}, \underbrace{4, \cdots, 4}_{n/2}), \quad \widetilde{\Sigma} = \text{diag}(\underbrace{1, \cdots, 1}_{N/2}, \underbrace{4, \cdots, 4}_{N/2});$$

(3) U and V are orthogonal matrices uniformly chosen from orthogonal groups $\mathbf{O}(n)$ and $\mathbf{O}(N)$. Then we take $n = 1000$ and calculate the largest eigenvalues for 20000 independently chosen matrices. The histograms are plotted in Fig. 2.1. In case (a), the entries $\sqrt{N}x_{ij}$ are drawn independently from a symmetric distribution with mean zero, variance 1 and satisfying (2.1.1); in case (b), the entries $\sqrt{N}x_{ij}$ are *i.i.d.* Gaussian with mean zero and variance 1. We translate and rescale the numerical results properly, and one can observe that they fit the type-1 Tracy-Widom distribution very well.



(a) For X satisfying (2.1.1).



(b) For Gaussian X.

Figure 2.1: Histograms for the largest eigenvalues of 20000 ensembles.

2.3 Statistical applications

In this section, we briefly discuss some applications of our result to high-dimensional statistics. We remark that heavy-tailed data, correlated data, and data with time correlations are commonly collected in finance, environmental study and telecommunications. For this type of data, many high-dimensional statistical hypothesis tests that rely on some strong moment or independence assumptions cannot be employed, and our edge universality result then serves as a valuable tool for many statistical applications.

Sample covariance matrices

If we take $B = I$, then $\mathcal{Q}_1$ becomes the normal sample covariance matrix and Theorem 2.2.7 indicates that the edge universality of the largest eigenvalue of $\mathcal{Q}_1$ holds true for correlated data (i.e. non-diagonal A) with heavy tails as in (2.1.1). So far, this is the strongest edge universality for sample covariance matrices compared with [12, 61] (assuming high moments and diagonal A), [59] (assuming high moments) and [26] (assuming diagonal A). The sample covariance matrices are widely used in various applied fields: multivariate statistics, empirical finance, signal processing, population genetics, and machine learning, to name a few. We now give a few concrete examples of applications of our edge universality result.

Consider the following signal plus noise model

$$\mathbf{y} = \Gamma \mathbf{s} + A^{1/2}\mathbf{x}, \tag{2.3.1}$$

where Γ is an $n \times k$ deterministic matrix, $\mathbf{s}$ is a k-dimensional centered signal vector, A is an $n \times n$ deterministic positive definite matrix, and $\mathbf{x}$ is an n-dimensional noise vector with *i.i.d.* mean zero and variance one entries. Moreover, the signal vector and the noise vector are assumed to be independent. In practice, suppose we observe N such *i.i.d.* samples and set the matrices

$$Y = \Gamma S + A^{1/2}X, \quad S := (\mathbf{s}_1, \cdots, \mathbf{s}_N), \quad X := (\mathbf{x}_1, \cdots, \mathbf{x}_N).$$

The above model is a standard model in classic signal processing [56]. A fundamental task is to detect the signals via observed samples, and the very first step is to know whether there exists any such signal, i.e.,

$$\mathbf{H}_0: \ k = 0 \quad \text{vs.} \quad \mathbf{H}_1: \ k \geqslant 1. \tag{2.3.2}$$

The model (2.3.1) is also widely used in various other fields. For example, in multivariate statistics, one wants to determine whether there exists any relation between two sets of variables. To test the independence, we can adopt the multivariate multiple regression model (2.3.1), where $\mathbf{x}$ and $\mathbf{y}$ are the two sets of variables for testing [52]. Then we wish to test the null hypothesis that these regression coefficients are all zero:

$$\mathbf{H}_0: \ \Gamma = 0 \ \text{ vs. } \ \mathbf{H}_1: \ \Gamma \neq 0. \tag{2.3.3}$$

Another example is from financial studies [39, 40, 41]. In the empirical research of finance, (2.3.1) is the factor model, where $\mathbf{s}$ is the common factor, Γ is the factor loading matrix and $\mathbf{x}$ is the idiosyncratic component. In order to analyze the stock return $\mathbf{y}$, we first need to know if the factor $\mathbf{s}$ is significant for the prediction. Then a statistical test can be also constructed as (2.3.3).

For the above hypothesis testing problems (2.3.2) and (2.3.3), the largest eigenvalue of the observed samples serves as a natural choice for the tests. In high-dimensional setting, this problem was considered in [13, 69] under the assumptions that $\mathbf{z}$ is Gaussian and $A = I$. Nadakuditi and Silverstein [70] also considered this problem with correlated Gaussian noise (i.e. A is not a multiple of I). For general diagonal A, the problem beyond Gaussian was considered in [12, 61] under the assumption that the entries of X have arbitrarily high moments, and in [26] under the condition (2.1.1). Our result shows that, for *h*eavy-tailed correlated data satisfying (2.1.1), one can still use the largest eigenvalue as our test static in the above high-dimensional statistical inference problems.

Remark 2.3.1. A small issue in choosing the largest eigenvalue as our test static is that the covariance matrix A is usually unknown in practice. Hence our result cannot be applied directly since the parameters λ_r and γ_0 in Remark 2.2.9 depend on (the singular values of) A. However, we can adopt the strategy in [72] and use the following statistics $\mathbf{T}_1 :=$

$(\lambda_1 - \lambda_2)/(\lambda_2 - \lambda_3)$ to eliminate the unknown parameters γ_0 and λ_r. According to Remark 2.2.10, the limiting distribution of $\mathbf{T}_1$ is uniquely determined by the Tracy-Widom law. (Although the explicit formula is unavailable currently, one can approximate the limiting distribution of $\mathbf{T}_1$ via numerical simulations using GOE or GUE.) The main advantage of $\mathbf{T}_1$ is that its limiting distribution is independent of A under $\mathbf{H}_0$, which makes it asymptotically pivotal.

Separable covariance matrices

The data model $Y = A^{1/2}XB^{1/2}$ is widely used in spatio-temporal data modeling, where the rows indices correspond to the spatial locations and the column indices correspond to the observation times. The spectral properties of $\mathcal{Q}_1 = YY^*$ have been investigated in some recent works [20, 29, 78, 101, 113]. If the entries of X are symmetrically distributed and the singular values of A, B are such that (2.2.19) holds, then Theorem 2.2.7 shows that the largest eigenvalue of $\mathcal{Q}_1$ satisfies the edge universality as long as the tail condition (2.1.1) holds. We now give some examples of the applications of this result. Without loss of generality, we shall call A the spatial covariance matrix and B the temporal covariance matrix.

Again we consider the model (2.3.1). Instead of observing *i.i.d.* samples, we assume that the observations at different times are correlated and the correlations are independent of the spatial locations. Denoting the temporal covariance matrix by B, we then have the spatio-temporal sampling data

$$Y = \Gamma S B^{1/2} + A^{1/2}XB^{1/2}, \quad S := (\mathbf{s}_1, \cdots, \mathbf{s}_N), \quad X := (\mathbf{x}_1, \cdots, \mathbf{x}_N).$$

We can again form the hypothesis testing problem (2.3.2) or (2.3.3). In high-dimensional setting, the largest singular value of Y is a natural choice for the test static: under $\mathbf{H}_0$, the largest singular value of Y satisfies the Tracy-Widom distribution asymptotically. We can also use the Onatski's statistics $\mathbf{T}_1 := (\lambda_1 - \lambda_2)/(\lambda_2 - \lambda_3)$ if no information on A and B is known a priori.

The spatio-temporal data model $Y = A^{1/2}XB^{1/2}$ is widely used in modeling environmental data [44, 60, 63, 68] and wireless communications [99]. We can consider to test whether the space-time data follows a specific separable covariance model with spatial and time covariance matrices $\widetilde{A}$ and $\widetilde{B}$. Then we can use the largest singular value of $\widetilde{A}^{-1/2} Y \widetilde{B}^{-1/2}$ as a test static. (Another interesting test static for this hypothesis testing problem is the eigenvector empirical spectral distribution (VESD); see Chapter 3 below.) In wireless communications, the importance of obtaining more detailed information on the largest singular values is becoming more transparent. For example, it was shown in [100] that an estimate of the capacity is directly given by various informations of the largest singular value, which is described by Theorem 2.2.7.

Finally, we remark that one can also perform principal component analysis for separable covariance matrices, and study the phase transition phenomena caused by a few large isolated eigenvalues of A or B as in the case of spiked covariance matrices [10, 11, 15, 76]. We expect that our edge universality result will serve as an important input for the study of the eigenvalues and eigenvectors for the principal components (the outliers) and the bulk components (the non-outliers). Moreover, as byproducts of the proof of Theorem 2.2.7, we obtain the isotropic delocalization of eigenvectors (Lemma 2.4.9) and the rigidity of eigenvalues (Theorem 2.4.8), which can also be valuable tools for statistical studies and applications.

2.4 Basic notations and tools

In this section, we state the main tools for our proof—the local laws for separable covariance matrices and some important corollaries of them. Their proofs constitute the main part of this chapter, and will be postponed to Sections 2.6-2.8.2.

2.4.1 Notations

We first introduce some notations for our proof. We will use the following notion of stochastic domination, which was first introduced in [30] and subsequently used in many works on random matrix theory, such as [14, 15, 18, 32, 33, 59]. It simplifies the presentation of the results and their proofs by systematizing statements of the form "ξ is bounded by ζ with high probability up to a small power of N".

Definition 2.4.1 (Stochastic domination)**.** *(i) Let*

$$\xi = \left(\xi^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\right), \quad \zeta = \left(\zeta^{(N)}(u) : N \in \mathbb{N}, u \in U^{(N)}\right)$$

be two families of nonnegative random variables, where $U^{(N)}$ is a possibly N-dependent parameter set. We say ξ is stochastically dominated by ζ, uniformly in u, if for any fixed (small) $\varepsilon > 0$ and (large) $D > 0$,

$$\sup_{u \in U^{(N)}} \mathbb{P}\left[\xi^{(N)}(u) > N^{\varepsilon}\zeta^{(N)}(u)\right] \leqslant N^{-D}$$

for large enough $N \geqslant N_0(\varepsilon, D)$, and we shall use the notation $\xi \prec \zeta$. Throughout this thesis, the stochastic domination will always be uniform in all parameters that are not explicitly fixed (such as matrix indices, and z that takes values in some compact set). Note that $N_0(\varepsilon, D)$ may depend on quantities that are explicitly constant, such as τ in Assumption 2.2.1 and (2.2.19)*. If for some complex family ξ we have $|\xi| \prec \zeta$, then we will also write $\xi \prec \zeta$ or $\xi = O_{\prec}(\zeta)$.*

(ii) We extend the definition of $O_{\prec}(\cdot)$ to matrices in the weak operator sense as follows. Let A be a family of random matrices and ζ be a family of nonnegative random variables. Then $A = O_{\prec}(\zeta)$ means that $|\langle \mathbf{v}, A\mathbf{w}\rangle| \prec \zeta \|\mathbf{v}\|_2 \|\mathbf{w}\|_2$ uniformly in any deterministic vectors $\mathbf{v}$ *and* $\mathbf{w}$*. Here and throughout the following, whenever we say "uniformly in any deterministic vectors", we mean that "uniformly in any deterministic vectors belonging to a set of cardinality $N^{\mathrm{O}(1)}$".*

(iii) We say an event Ξ holds with high probability if for any constant $D > 0$, $\mathbb{P}(\Xi) \geqslant 1 - N^{-D}$ for large enough N.

The following lemma collects basic properties of stochastic domination $\prec$, which will be used tacitly in the proof.

Lemma 2.4.2 (Lemma 3.2 in [14])**.** *Let ξ and ζ be families of nonnegative random variables.*

(i) Suppose that $\xi(u, v) \prec \zeta(u, v)$ uniformly in $u \in U$ and $v \in V$. If $|V| \leqslant N^C$ for some constant C, then $\sum_{v \in V} \xi(u, v) \prec \sum_{v \in V} \zeta(u, v)$ uniformly in u.

(ii) If $\xi_1(u) \prec \zeta_1(u)$ and $\xi_2(u) \prec \zeta_2(u)$ uniformly in $u \in U$, then $\xi_1(u)\xi_2(u) \prec \zeta_1(u)\zeta_2(u)$ uniformly in u.

(iii) Suppose that $\Psi(u) \geqslant N^{-C}$ is deterministic and $\xi(u)$ satisfies $\mathbb{E}\xi(u)^2 \leqslant N^C$ for all u. Then if $\xi(u) \prec \Psi(u)$ uniformly in u, we have $\mathbb{E}\xi(u) \prec \Psi(u)$ uniformly in u.

Definition 2.4.3 (Bounded support condition)**.** *We say a random matrix X satisfies the bounded support condition with q, if*

$$\max_{i,j} |x_{ij}| \prec q. \tag{2.4.1}$$

Here $q \equiv q(N)$ is a deterministic parameter and usually satisfies $N^{-1/2} \leqslant q \leqslant N^{-\phi}$ for some (small) constant $\phi > 0$. Whenever (2.4.1) holds, we say that X has support q.

Next we introduce a convenient self-adjoint linearization trick, which has been proved to be useful in studying the local laws of random matrices of the Gram type [1, 2, 59, 105]. We define the following $(n+N) \times (n+N)$ self-adjoint block matrix, which is a linear function of X:

$$H \equiv H(X) := \begin{pmatrix} 0 & \Sigma^{1/2} U^* X V \widetilde{\Sigma}^{1/2} \\ \widetilde{\Sigma}^{1/2} V^* X^* U \Sigma^{1/2} & 0 \end{pmatrix}, \tag{2.4.2}$$

Then we define its resolvent (Green's function) as

$$G \equiv G(X,z) := \left[H(X) - \begin{pmatrix} I_{n\times n} & 0 \\ 0 & z I_{N\times N} \end{pmatrix} \right]^{-1}, \quad z \in \mathbb{C}_+. \tag{2.4.3}$$

By Schur complement formula, we can verify that (recall (2.2.7))

$$\begin{aligned} G &= \begin{pmatrix} z\mathcal{G}_1 & \mathcal{G}_1 \Sigma^{1/2} U^* X V \widetilde{\Sigma}^{1/2} \\ \widetilde{\Sigma}^{1/2} V^* X^* U \Sigma^{1/2} \mathcal{G}_1 & \mathcal{G}_2 \end{pmatrix} \\ &= \begin{pmatrix} z\mathcal{G}_1 & \Sigma^{1/2} U^* X V \widetilde{\Sigma}^{1/2} \mathcal{G}_2 \\ \mathcal{G}_2 \widetilde{\Sigma}^{1/2} V^* X^* U \Sigma^{1/2} & \mathcal{G}_2 \end{pmatrix}. \end{aligned} \tag{2.4.4}$$

Thus a control of G yields directly a control of the resolvents $\mathcal{G}_{1,2}$. For simplicity of notations, we define the index sets

$$\mathcal{I}_1 := \{1, ..., n\}, \quad \mathcal{I}_2 := \{n+1, ..., n+N\}, \quad \mathcal{I} := \mathcal{I}_1 \cup \mathcal{I}_2.$$

Then we label the indices of the matrices according to

$$X = (X_{i\mu} : i \in \mathcal{I}_1, \mu \in \mathcal{I}_2), \quad A = (A_{ij} : i, j \in \mathcal{I}_1), \quad B = (B_{\mu\nu} : \mu, \nu \in \mathcal{I}_2).$$

In the rest of this chapter, we will consistently use the latin letters $i, j \in \mathcal{I}_1$, greek letters $\mu, \nu \in \mathcal{I}_2$, and $a, b \in \mathcal{I}$.

Next we introduce the spectral decomposition of G. Let

$$\Sigma^{1/2} U^* X V \widetilde{\Sigma}^{1/2} = \sum_{k=1}^{n \wedge N} \sqrt{\lambda_k} \xi_k \zeta_k^*,$$

be a singular value decomposition of $\Sigma^{1/2} U^* X V \widetilde{\Sigma}^{1/2}$, where

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_{n \wedge N} \geqslant 0 = \lambda_{n \wedge N + 1} = \ldots = \lambda_{n \vee N},$$

$\{\xi_k\}_{k=1}^n$ are the left-singular vectors, and $\{\zeta_k\}_{k=1}^N$ are the right-singular vectors. Then using (2.4.4), we can get that for $i, j \in \mathcal{I}_1$ and $\mu, \nu \in \mathcal{I}_2$,

$$G_{ij} = \sum_{k=1}^{n} \frac{z \xi_k(i) \xi_k^*(j)}{\lambda_k - z}, \qquad G_{\mu\nu} = \sum_{k=1}^{N} \frac{\zeta_k(\mu) \zeta_k^*(\nu)}{\lambda_k - z}, \tag{2.4.5}$$

$$G_{i\mu} = \sum_{k=1}^{n \wedge N} \frac{\sqrt{\lambda_k} \xi_k(i) \zeta_k^*(\mu)}{\lambda_k - z}, \qquad G_{\mu i} = \sum_{k=1}^{n \wedge N} \frac{\sqrt{\lambda_k} \zeta_k(\mu) \xi_k^*(i)}{\lambda_k - z}. \tag{2.4.6}$$

2.4.2 Local laws

For any constants $c_0, C_0 > 0$ and $a \leqslant 1$, we define a domain of the spectral parameter z as

$$S(c_0, C_0, a) := \left\{ z = E + \mathrm{i}\eta : \lambda_r - c_0 \leqslant E \leqslant C_0 \lambda_r, N^{-1+a} \leqslant \eta \leqslant 1 \right\}. \tag{2.4.7}$$

In particular, we shall denote

$$S(c_0, C_0, -\infty) := \{ z = E + \mathrm{i}\eta : \lambda_r - c_0 \leqslant E \leqslant C_0 \lambda_r, 0 \leqslant \eta \leqslant 1 \}. \tag{2.4.8}$$

We define the distance to the rightmost edge as

$$\kappa \equiv \kappa_E := |E - \lambda_r|, \quad \text{for } z = E + \mathrm{i}\eta. \tag{2.4.9}$$

Then we have the following lemma, which summarizes some basic properties of m_{2c} and ρ_{2c}.

Lemma 2.4.4. *Suppose the assumptions* (2.2.2), (2.2.5) *and* (2.2.19) *hold. Then there exists sufficiently small constant* $\widetilde{c} > 0$ *such that the following estimates hold:*

(1)

$$\rho_{1,2c}(x) \sim \sqrt{\lambda_r - x}, \qquad \text{for } x \in [\lambda_r - 2\widetilde{c}, \lambda_r]; \tag{2.4.10}$$

(2) for $z = E + \mathrm{i}\eta \in S(\widetilde{c}, C_0, -\infty)$,

$$|m_{1,2c}(z)| \sim 1, \quad \operatorname{Im} m_{1,2c}(z) \sim \begin{cases} \eta/\sqrt{\kappa+\eta}, & \text{if } E \geqslant \lambda_r \\ \sqrt{\kappa+\eta}, & \text{if } E \leqslant \lambda_r \end{cases}; \tag{2.4.11}$$

(3) there exists constant $\tau' > 0$ *such that*

$$\min_{\mu \in \mathcal{I}_2} |1 + m_{1c}(z)\widetilde{\sigma}_\mu| \geqslant \tau', \quad \min_{i \in \mathcal{I}_1} |1 + m_{2c}(z)\sigma_i| \geqslant \tau', \tag{2.4.12}$$

for any $z \in S(\widetilde{c}, C_0, -\infty)$.

The estimates (2.4.10) *and* (2.4.11) *also hold for* ρ_c *and* m_c.

Proof. The estimate (2.4.10) is already given by Lemma 2.2.6. The estimate (2.4.11) can be proved easily with (2.2.21). It remains to prove (2.4.12). By assumption (2.2.19) and the fact $m_{2c}(\lambda_r) \in (-\sigma_1^{-1}, 0)$, we have

$$|1 + m_{2c}(\lambda_r)\sigma_i| \geqslant \tau, \quad i \in \mathcal{I}_1.$$

With (2.2.21), we see that if $\kappa + \eta \leqslant 2c_0$ for some sufficiently small constant $c_0 > 0$, then

$$|1 + m_{2c}(z)\sigma_k| \geqslant \tau/2.$$

Then we consider the case with $E \geqslant \lambda_r + c_0$ and $\eta \leqslant c_1$ for some constant $c_1 > 0$. In fact, for $\eta = 0$ and $E > \lambda_r$, $m_{2c}(E)$ is real and it is easy to verify that $m_{2c}'(E) \geqslant 0$ using the Stieltjes transform formula

$$m_{2c}(z) := \int_{\mathbb{R}} \frac{\rho_{2c}(dx)}{x - z}, \tag{2.4.13}$$

Hence we have

$$1 + \sigma_i m_{2c}(E) \geqslant 1 + \sigma_i m_{2c}(\lambda_r) \geqslant \tau, \quad \text{for } E \geqslant \lambda_r + c_0.$$

Using (2.4.13) again, we can get that

$$\left|\frac{\mathrm{d}m_{2c}(z)}{\mathrm{d}z}\right| \leqslant c_0^{-2}, \quad \text{for } E \geqslant \lambda_r + c_0.$$

Thus if c_1 is sufficiently small, we have

$$|1+\sigma_k m_{2c}(E+\mathrm{i}\eta)| \geqslant \tau/2$$

for $E \geqslant \lambda_r + c_0$ and $\eta \leqslant c_1$. Finally, it remains to consider the case with $\eta \geqslant c_1$. Note that we have $|m_{2c}(z)| \sim \operatorname{Im} m_{2c}(z) \sim 1$ by (2.4.11). If $\sigma_k \leqslant |2m_{2c}(z)|^{-1}$, then $|1+\sigma_k m_{2c}(z)| \geqslant 1/2$. Otherwise, we have

$$|1+\sigma_k m_{2c}(z)| \geqslant \sigma_k \operatorname{Im} m_{2c}(z) \geqslant \frac{\operatorname{Im} m_{2c}(z)}{2|m_{2c}(z)|} \gtrsim 1.$$

In sum, we have proved the second estimate in (2.4.12). The first estimate can be proved in a similar way. □

Definition 2.4.5 (Classical locations of eigenvalues)**.** *The classical location γ_j of the j-th eigenvalue of $\mathcal{Q}_1$ is defined as*

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} \rho_c(x)dx > \frac{j-1}{n} \right\}. \tag{2.4.14}$$

In particular, we have $\gamma_1 = \lambda_r$.

In the rest of this section, we present some results that will be used in the proof of Theorem 2.2.7. Their proofs will be given in subsequent sections. For any matrix X satisfying Assumption 2.2.1 and the tail condition (2.1.1), we can construct a matrix X^s that approximates X with probability $1-\mathrm{o}(1)$, and satisfies Assumption 2.2.1, the bounded support condition (2.4.1) with $q \leqslant N^{-\phi}$ for some small constant $\phi > 0$, and

$$\mathbb{E}|x_{ij}^s|^3 = \mathrm{O}(N^{-3/2}), \quad \mathbb{E}|x_{ij}^s|^4 = \mathrm{O}_\prec(N^{-2}); \tag{2.4.15}$$

see Section 2.5 for the details. We will need the following local laws, eigenvalues rigidity, eigenvector delocalization, and edge universality results for separable covariance matrices with X^s.

We define the deterministic limit Π of the resolvent G in (2.4.3) as

$$\Pi(z) := \begin{pmatrix} -\left(1+m_{2c}(z)\Sigma\right)^{-1} & 0 \\ 0 & -z^{-1}(1+m_{1c}(z)\widetilde{\Sigma})^{-1} \end{pmatrix}. \tag{2.4.16}$$

Note that we have

$$\frac{1}{nz}\sum_{i\in\mathcal{I}_1}\Pi_{ii} = m_c. \tag{2.4.17}$$

Define the control parameters

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m_{2c}(z)}{N\eta}} + \frac{1}{N\eta}. \tag{2.4.18}$$

Note that by (2.4.11) and (2.4.12), we have

$$\|\Pi\| = \mathrm{O}(1), \quad \Psi \gtrsim N^{-1/2}, \quad \Psi^2 \lesssim (N\eta)^{-1}, \quad \Psi(z) \sim \sqrt{\frac{\operatorname{Im} m_{1c}(z)}{N\eta}} + \frac{1}{N\eta}, \tag{2.4.19}$$

for $z \in S(\widetilde{c}, C_0, -\infty)$. Now we are ready to state the local laws for $G(X,z)$. For the purpose of proving Theorem 2.2.7, we shall relax the condition (2.1.2) a little bit.

Theorem 2.4.6 (Local laws)**.** *Suppose Assumption 2.2.1 and* (2.2.19) *hold. Suppose X satisfies the bounded support condition (2.4.1) with $q \leqslant N^{-\phi}$ for some constant $\phi > 0$. Furthermore, suppose X satisfies* (2.4.15) *and*

$$\left|\mathbb{E}x_{ij}^3\right| \leqslant b_N N^{-2}, \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant N, \tag{2.4.20}$$

where b_N is an N-dependent deterministic parameter satisfying $1 \leqslant b_N \leqslant N^{1/2}$. Fix $C_0 > 1$ and let $c_0 > 0$ be a sufficiently small constant. Given any $\varepsilon > 0$, we define the domain

$$\widetilde{S}(c_0, C_0, \varepsilon) := S(c_0, C_0, \varepsilon) \cap \left\{ z = E + \mathrm{i}\eta : b_N\left(\Psi^2(z) + \frac{q}{N\eta}\right) \leqslant N^{-\varepsilon} \right\}. \tag{2.4.21}$$

Then for any fixed $\varepsilon > 0$, the following estimates hold.

(1) ***Anisotropic local law****: For any $z \in \widetilde{S}(c_0, C_0, \varepsilon)$ and deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$,*

$$|\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec q + \Psi(z). \tag{2.4.22}$$

(2) ***Averaged local law****: For any* $z \in \widetilde{S}(c_0, C_0, \varepsilon)$*, we have*

$$|m(z) - m_c(z)| \prec q^2 + (N\eta)^{-1}. \tag{2.4.23}$$

where m *is defined in* (2.2.8)*. Moreover, outside of the spectrum we have the following stronger estimate*

$$|m(z) - m_c(z)| \prec q^2 + \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}, \tag{2.4.24}$$

uniformly in $z \in \widetilde{S}(c_0, C_0, \varepsilon) \cap \{z = E + i\eta : E \geqslant \lambda_r, N\eta\sqrt{\kappa+\eta} \geqslant N^{\varepsilon}\}$*, where* κ *is defined in* (2.4.9)*.*

The above estimates are uniform in the spectral parameter z *and any set of deterministic vectors of cardinality* $N^{O(1)}$*. If* A *or* B *is diagonal, then* (2.4.22)*-*(2.4.24) *hold for* $z \in S(c_0, C_0, \varepsilon)$*.*

The main difficulty for the proof of Theorem 2.4.6 is due to the fact that the entries of $A^{1/2}XB^{1/2}$ are not independent anymore. However, notice that if $X \equiv X^{Gauss}$ is a Wishart matrix, we have

$$\Sigma^{1/2}U^*X^{Gauss}V\widetilde{\Sigma}^{1/2} \stackrel{d}{=} \Sigma^{1/2}X^{Gauss}\widetilde{\Sigma}^{1/2}.$$

In this case, the problem is reduced to proving the anisotropic local law for separable covariance matrices with diagonal spatial and temporal covariance matrices, which can be handled using the standard resolvent methods as in e.g. [14, 79]. To go from the Gaussian case to the general X case, we adopt a continuous self-consistent comparison argument developed in [59]. In order for this argument to work, we need to assume (2.1.2). Under the weaker condition (2.4.20), we cannot prove the local laws up to the optimal scale $\eta \gg N^{-1}$, but only up to the scale $\eta \gg \max\{\frac{qb_N}{N}, \frac{\sqrt{b_N}}{N}\}$ near the edge. However, to prove the edge universality, we only need to have a good local law up to the scale $\eta \leqslant N^{-2/3-\varepsilon}$, hence b_N can take values up to $b_N \ll N^{1/3}$. (In the proof of Theorem 2.2.7 in Section 2.5, we will take $b_N = N^{-\varepsilon}$ for some small constant $\varepsilon > 0$.) Finally, if A or B is diagonal, one can prove the local laws up to the optimal scale for all $b_N = O(N^{1/2})$ by using an improved comparison argument in [59].

Following the above discussions, we divide the proof of Theorem 2.4.6 into two steps. In Section 2.6, we give the proof for separable covariance matrices of the form $\Sigma^{1/2}X\widetilde{\Sigma}X^*\Sigma^{1/2}$, which implies the local laws in the Gaussian X case. In Section 2.7, we apply the self-consistent comparison argument in [59] to extend the result to the general X case. Compared with [59], there are two differences in our setting: (1) the support of X in Theorem 2.4.6 is $q = \mathrm{O}(N^{-\phi})$ for some constant $0 < \phi \leqslant 1/2$, while [59] dealt with X with smaller support $q = \mathrm{O}(N^{-1/2})$; (2) one has $B = I$ in [59], which simplifies the proof a little bit.

The second moment of the error $\langle \mathbf{u}, (G - \Pi)\mathbf{v}\rangle$ in fact satisfies a stronger bound.

Lemma 2.4.7. *Suppose the assumptions in Theorem 2.4.6 hold. Then for any fixed $\varepsilon > 0$ and $z \in \widetilde{S}(c_0, C_0, \varepsilon)$, we have the following bound*

$$\mathbb{E}|\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle|^2 \prec \Psi^2(z), \tag{2.4.25}$$

for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$.

With Theorem 2.4.6 as a key input, we can prove a stronger estimate on $m(z)$ that is independent of q. This averaged local law implies the rigidity of eigenvalues for $\mathcal{Q}_1$. Note that for any fixed E, $\Psi^2(E + \mathrm{i}\eta) + q/(N\eta)$ is monotonically decreasing with respect to η, hence there is a unique $\eta_1(E)$ such that

$$b_N \left(\Psi^2(E + \mathrm{i}\eta_1(E)) + \frac{q}{N\eta_1(E)} \right) = 1.$$

Then we define $\eta_l(E) := \max_{E \leqslant x \leqslant \lambda_r} \eta_1(x)$ ("l" for lower bound) for $E \leqslant \lambda_r$, and $\eta_l(E) := \eta_l(\lambda_r)$ for $E > \lambda_r$. Note that by (2.4.18), we always have $\eta_l(E) = \mathrm{O}(b_N/N)$.

Theorem 2.4.8 (Rigidity of eigenvalues)**.** *Suppose the assumptions in Theorem 2.4.6 hold. Fix the constants c_0 and C_0 as given in Theorem 2.4.6. Then for any fixed $\varepsilon > 0$, we have*

$$|m(z) - m_c(z)| \prec (N\eta)^{-1}, \tag{2.4.26}$$

uniformly in $z \in \widetilde{S}(c_0, C_0, \varepsilon)$. Moreover, outside of the spectrum we have the following stronger estimate

$$|m(z) - m_c(z)| \prec \frac{1}{N(\kappa + \eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa + \eta}}, \tag{2.4.27}$$

uniformly in $z \in \widetilde{S}(c_0, C_0, \varepsilon) \cap \{z = E + i\eta : E \geqslant \lambda_r, N\eta\sqrt{\kappa+\eta} \geqslant N^{\varepsilon}\}$ *for any fixed* $\varepsilon > 0$*. If* A *or* B *is diagonal, then* (2.4.26) *and* (2.4.27) *hold for* $z \in S(c_0, C_0, \varepsilon)$*. The bounds (2.4.26) and* (2.4.27) *imply that for any constant* $0 < c_1 < c_0$*, the following estimates hold.*

(1) For any $E \geqslant \lambda_r - c_1$*, we have*

$$|n(E) - n_c(E)| \prec N^{-1} + (\eta_l(E))^{3/2} + \eta_l(E)\sqrt{\kappa_E}, \tag{2.4.28}$$

where κ_E *is defined in* (2.4.9)*, and*

$$n(E) := \frac{1}{N}\#\{\lambda_j \geqslant E\}, \quad n_c(E) := \int_E^{+\infty} \rho_{2c}(x)dx. \tag{2.4.29}$$

(2) If $b_N \leqslant N^{1/3-c}$ *for some constant* $c > 0$*, then for any* j *such that* $\lambda_r - c_1 \leqslant \gamma_j \leqslant \lambda_r$*, we have*

$$|\lambda_j - \gamma_j| \prec j^{-1/3}N^{-2/3} + \eta_0, \tag{2.4.30}$$

where $\eta_0 := \eta_l(\lambda_r - c_1) = \mathrm{O}(b_N/N)$*.*

The estimates (2.4.28) and (2.4.30) follow from the estimates (2.4.26) and (2.4.27) combined with a standard argument using Helffer-Sjöstrand calculus. The details are already given in [34], [38] and [79]. Hence to prove Theorem 2.4.8, we only need to show that (2.4.26) and (2.4.27) hold.

The anisotropic local law (2.4.22) implies the following delocalization properties of eigenvectors.

Lemma 2.4.9 (Isotropic delocalization of eigenvectors)**.** *Suppose* (2.4.22) *and* (2.4.30) *hold. Then for any deterministic unit vectors* $\mathbf{u} \in \mathbb{C}^{\mathcal{I}_1}$*,* $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}$ *and constant* $0 < c_1 < c_0$*, we have*

$$\max_{k:\lambda_r - c_1 \leqslant \gamma_k \leqslant \lambda_r} \left\{|\langle \mathbf{u}, \xi_k\rangle|^2 + |\langle \mathbf{v}, \zeta_k\rangle|^2\right\} \prec \eta_0, \tag{2.4.31}$$

where η_0 *is defined below* (2.4.30)*.*

Proof. Choose $z_0 = E + iN^{\varepsilon}\eta_0 \in \widetilde{S}(c_0, C_0, \varepsilon)$. By (2.4.22) and (2.4.19), we have $\mathrm{Im}\langle \mathbf{v}, G(z_0)\mathbf{v}\rangle = \mathrm{O}(1)$ with high probability. Then using the spectral decomposition (2.4.5), we get

$$\sum_{k=1}^{N} \frac{N^{\varepsilon}\eta_0|\langle \mathbf{v}, \zeta_k\rangle|^2}{(\lambda_k - E)^2 + N^{2\varepsilon}\eta_0^2} = \mathrm{Im}\,\langle \mathbf{v}, G(z_0)\mathbf{v}\rangle = \mathrm{O}(1) \quad \text{with high probability.} \tag{2.4.32}$$

By (2.4.30), we have that $\lambda_k + \mathrm{i}N^{\varepsilon}\eta_0 \in \widetilde{S}(c_0, C_0, \varepsilon)$ with high probability for every k such that $\lambda_r - c_1 \leqslant \gamma_k \leqslant \lambda_r$. Then choosing $E = \lambda_k$ in (2.4.32) yields that

$$|\langle \mathbf{v}, \zeta_k\rangle|^2 \lesssim N^{\varepsilon}\eta_0 \quad \text{with high probability.}$$

Since ε is arbitrary, we get $|\langle \mathbf{v}, \zeta_k\rangle|^2 \prec \eta_0$. In a similar way, we can prove $|\langle \mathbf{u}, \xi_k\rangle|^2 \prec \eta_0$. $\square$

Finally, we have the following edge universality result for separable covariance matrices with support $q \leqslant N^{-\phi}$ and satisfying the condition (2.4.15).

Theorem 2.4.10. *Let $X^{(1)}$ and $X^{(2)}$ be two separable covariance matrices satisfying the assumptions in Theorem 2.4.6. Suppose $b_N \leqslant N^{1/3-c}$ for some constant $c > 0$. Then there exist constants $\varepsilon, \delta > 0$ such that for any $s \in \mathbb{R}$,*

$$\begin{aligned}\mathbb{P}^{(1)}\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s - N^{-\varepsilon}\right) - N^{-\delta} &\leqslant \mathbb{P}^{(2)}\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s\right) \\ &\leqslant \mathbb{P}^{(1)}\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s + N^{-\varepsilon}\right) + N^{-\delta},\end{aligned} \tag{2.4.33}$$

where $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$ denote the laws of $X^{(1)}$ and $X^{(2)}$, respectively.

Remark 2.4.11. As in [31, 38, 62], Theorem 2.4.10 can be can be generalized to finite correlation functions of the k largest eigenvalues for any fixed k:

$$\begin{aligned}\mathbb{P}^{(1)}\left(\left(N^{2/3}(\lambda_i - \lambda_r) \leqslant s_i - N^{-\varepsilon}\right)_{1\leqslant i\leqslant k}\right) - N^{-\delta} &\leqslant \mathbb{P}^{(2)}\left(\left(N^{2/3}(\lambda_i - \lambda_r) \leqslant s_i\right)_{1\leqslant i\leqslant k}\right) \\ &\leqslant \mathbb{P}^{(1)}\left(\left(N^{2/3}(\lambda_i - \lambda_r) \leqslant s_i + N^{-\varepsilon}\right)_{1\leqslant i\leqslant k}\right) + N^{-\delta}.\end{aligned} \tag{2.4.34}$$

The proof of (2.4.34) is similar to that of (2.4.33) except that it uses a general form of the Green function comparison theorem; see e.g. [38, Theorem 6.4]. As a corollary, we can get the stronger edge universality result (2.2.23).

The proofs for Lemma 2.4.7, Theorem 2.4.8 and Theorem 2.4.10 follow essentially the same path as discussed below. First, for random matrix $\widetilde{X}$ with small suppoort $q = \mathrm{O}(N^{-1/2})$, we have the averaged local laws (2.4.26)-(2.4.27) and the following anisotropic local law

$$\left|\langle \mathbf{u}, G(\widetilde{X}, z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle\right| \prec \Psi(z).$$

With these estimates, one can prove that Lemma 2.4.7, Theorem 2.4.8 and Theorem 2.4.10 hold in the small support case using the methods in e.g. [31, 38, 79]. Then it suffices to use a comparison argument to show that the large support case is "sufficiently close" to the small support case. In fact, given any matrix X satisfying the assumptions in Theorem 2.4.6, we can construct a matrix $\widetilde{X}$ having the same first four moments as X but with smaller support $q = \mathrm{O}(N^{-1/2})$, which is the content of the next lemma.

Lemma 2.4.12 (Lemma 5.1 in [62]). *Suppose X satisfies the assumptions in Theorem 2.4.6. Then there exists another matrix $\widetilde{X} = (\widetilde{x}_{ij})$, such that $\widetilde{X}$ satisfies the bounded support condition (2.4.1) with $q = N^{-1/2}$, and the first four moments of the X entries and $\widetilde{X}$ entries match, i.e.*

$$\mathbb{E}x_{ij}^k = \mathbb{E}\widetilde{x}_{ij}^k, \quad k = 1, 2, 3, 4. \tag{2.4.35}$$

It is known that the Lindeberg replacement strategy combined with the four moment matching usually implies some universality results in random matrix theory, see e.g. [92, 93, 95]. This is actually also true in our case. We shall extend the Green function comparison method developed in [62] (which is essentially an iterative application of the Lindeberg strategy together with the four moment matching), and prove that Lemma 2.4.7, Theorem 2.4.8 and Theorem 2.4.10 also hold for the large support case. The proof of Lemma 2.4.7 and Theorem 2.4.8 will be given in Section 2.8.1, and Theorem 2.4.10 will be proved in Section 2.8.2.

2.5 Proof of the main result

In this section, we prove Theorem 2.2.7 with the results in Section 2.4.2. Given the matrix X satisfying Assumption 2.2.1 and the tail condition (2.1.1), we introduce a cutoff on its matrix entries at the level $N^{-\varepsilon}$. For any fixed $\varepsilon > 0$, define

$$\alpha_N := \mathbb{P}\left(|q_{11}| > N^{1/2-\varepsilon}\right), \quad \beta_N := \mathbb{E}\left[\mathbf{1}\left(|q_{11}| > N^{1/2-\varepsilon}\right)q_{11}\right].$$

By (2.1.1) and integration by parts, we get that for any fixed $\delta > 0$ and large enough N,

$$\alpha_N \leqslant \delta N^{-2+4\varepsilon}, \quad |\beta_N| \leqslant \delta N^{-3/2+3\varepsilon}. \tag{2.5.1}$$

Let $\rho(x)$ be the distribution density of q_{11}. Then we define independent random variables q^s_{ij}, q^l_{ij}, c_{ij}, $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant N$, in the following ways:

- q^s_{ij} has distribution density $\rho_s(x)$, where

$$\rho_s(x) = \mathbf{1}\left(\left|x - \frac{\beta_N}{1-\alpha_N}\right| \leqslant N^{1/2-\varepsilon}\right) \frac{\rho\left(x - \frac{\beta_N}{1-\alpha_N}\right)}{1-\alpha_N};$$

- q^l_{ij} has distribution density $\rho_l(x)$, where

$$\rho_l(x) = \mathbf{1}\left(\left|x - \frac{\beta_N}{1-\alpha_N}\right| > N^{1/2-\varepsilon}\right) \frac{\rho\left(x - \frac{\beta_N}{1-\alpha_N}\right)}{\alpha_N};$$

- c_{ij} is a Bernoulli 0-1 random variable with $\mathbb{P}(c_{ij} = 1) = \alpha_N$ and $\mathbb{P}(c_{ij} = 0) = 1 - \alpha_N$.

Let X^s, X^l and X^c be random matrices such that $X^s_{ij} = N^{-1/2} q^s_{ij}$, $X^l_{ij} = N^{-1/2} q^l_{ij}$ and $X^c_{ij} = c_{ij}$. It is easy to check that for independent X^s, X^l and X^c,

$$X_{ij} \stackrel{d}{=} X^s_{ij}\left(1 - X^c_{ij}\right) + X^l_{ij} X^c_{ij} - \frac{1}{\sqrt{N}} \frac{\beta_N}{1-\alpha_N}. \tag{2.5.2}$$

If we define the $n \times N$ matrix $Y = (Y_{ij})$ by

$$Y_{ij} = \frac{1}{\sqrt{N}} \frac{\beta_N}{1-\alpha_N} = \mathrm{O}(\delta N^{-2+3\varepsilon}), \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant N,$$

then we have $\|Y\| = \mathrm{O}(N^{-1+3\varepsilon})$. In the proof below, one will see that (recall (2.2.6))

$$\left\|\Sigma^{1/2} U^*(X+Y) V \widetilde{\Sigma}^{1/2}\right\| = \lambda_1^{1/2}\left(\widetilde{\mathcal{Q}}_1(X+Y)\right) = \mathrm{O}(1)$$

with probability $1 - \mathrm{o}(1)$, where λ_1 denotes the largest eigenvalue of $\widetilde{\mathcal{Q}}_1$. Thus with probability $1 - \mathrm{o}(1)$, we have

$$\left|\lambda_1\left(\widetilde{\mathcal{Q}}_1(X+Y)\right) - \lambda_1\left(\widetilde{\mathcal{Q}}_1(X)\right)\right| = O\left(N^{-1+3\varepsilon}\right). \tag{2.5.3}$$

Hence the deterministic part in (2.5.2) is negligible under the scaling $N^{2/3}$.

By (2.1.1), (2.1.2) and integration by parts, it is easy to check that

$$\begin{gathered} \mathbb{E} q^s_{11} = 0, \quad \mathbb{E}|q^s_{11}|^2 = 1 - \mathrm{O}(N^{-1+2\varepsilon}), \\ \mathbb{E}|q^s_{11}|^3 = \mathrm{O}(1), \quad \mathbb{E}(q^s_{11})^3 = \mathrm{O}(N^{-1/2+\varepsilon}), \quad \mathbb{E}|q^s_{11}|^4 = \mathrm{O}(\log N). \end{gathered} \tag{2.5.4}$$

Thus $X_1 := (\mathbb{E}|q_{11}^s|^2)^{-1/2} X^s$ is a matrix that satisfies the assumptions for X in Theorem 2.4.6 with $b_N = N^{\varepsilon}$ and $q = N^{-\varepsilon}$. Then by Theorem 2.4.10, there exist constants $\varepsilon, \delta > 0$ such that for any $s \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}^G\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s - N^{-\varepsilon}\right) - N^{-\delta} \leqslant \mathbb{P}^s\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s\right) \\ \leqslant \mathbb{P}^G(N^{2/3}\left(\lambda_1 - \lambda_r\right) \leqslant s + N^{-\varepsilon}\right) + N^{-\delta}, \end{aligned} \tag{2.5.5}$$

where $\mathbb{P}^s$ denotes the law for X^s and $\mathbb{P}^G$ denotes the law for Wishart matrix. Now we write the first two terms on the right-hand side of (2.5.2) as

$$X_{ij}^s(1 - X_{ij}^c) + X_{ij}^l X_{ij}^c = X_{ij}^s + R_{ij} X_{ij}^c, \quad R_{ij} := X_{ij}^l - X_{ij}^s.$$

We define the matrix $R^c := (R_{ij} X_{ij}^c)$. It remains to show that the effect of R^c on λ_1 is negligible. Note that X_{ij}^c is independent of X_{ij}^s and R_{ij}.

We first introduce a cutoff on matrix X^c as $\widetilde{X}^c := \mathbf{1}_A X^c$, where

$$A := \left\{\#\{(i,j) : X_{ij}^c = 1\} \leqslant N^{5\varepsilon}\right\} \cap \left\{X_{ij}^c = X_{kl}^c = 1 \Rightarrow \{i,j\} = \{k,l\} \text{ or } \{i,j\} \cap \{k,l\} = \varnothing\right\}.$$

If we regard the matrix X^c as a sequence $\mathbf{X}^c$ of nN *i.i.d.* Bernoulli random variables, it is easy to obtain from the large deviation formula that

$$\mathbb{P}\left(\sum_{i=1}^{nN} \mathbf{X}_i^c \leqslant N^{5\varepsilon}\right) \geqslant 1 - \exp(-N^{\varepsilon}), \tag{2.5.6}$$

for sufficiently large N. Suppose the number m of the nonzero elements in X^c is given with $m \leqslant N^{5\varepsilon}$. Then it is easy to check that

$$\mathbb{P}\left(\exists\, i = k, j \neq l \text{ or } i \neq k, j = l \text{ such that } X_{ij}^c = X_{kl}^c = 1 \,\middle|\, \sum_{i=1}^{nN} \mathbf{X}_i^c = m\right) = \mathrm{O}(m^2 N^{-1}). \tag{2.5.7}$$

Combining the estimates (2.5.6) and (2.5.7), we get that

$$\mathbb{P}(A) \geqslant 1 - \mathrm{O}(N^{-1+10\varepsilon}). \tag{2.5.8}$$

On the other hand, by condition (2.1.1), we have

$$\mathbb{P}\left(|R_{ij}| \geqslant \omega\right) \leqslant \mathbb{P}\left(|q_{ij}| \geqslant \frac{\omega}{2} N^{1/2}\right) = \mathrm{o}(N^{-2}), \tag{2.5.9}$$

for any fixed constant $\omega > 0$. Hence if we introduce the matrix

$$E = \mathbf{1}\left(A \cap \left\{\max_{i,j} |R_{ij}| \leqslant \omega\right\}\right) R^c,$$

then we have

$$\mathbb{P}(E = R^c) = 1 - \mathrm{o}(1) \tag{2.5.10}$$

by (2.5.8) and (2.5.9). Thus we only need to study the largest eigenvalue of $\widetilde{\mathcal{Q}}_1(X^s + E)$, where $\max_{i,j} |E_{ij}| \leqslant \omega$ and $\mathrm{rank}(E) \leqslant N^{5\varepsilon}$. In fact, it suffices to prove that

$$\mathbb{P}\left(\left|\lambda_1^s - \lambda_1^E\right| \leqslant N^{-3/4}\right) = 1 - \mathrm{o}(1), \quad \lambda_1^s := \lambda_1\left(\widetilde{\mathcal{Q}}_1(X^s)\right), \quad \lambda_1^E := \lambda_1\left(\widetilde{\mathcal{Q}}_1(X^s + E)\right). \tag{2.5.11}$$

The estimate (2.5.11), combined with (2.5.3), (2.5.5) and (2.5.10), concludes (2.2.22).

Now we prove (2.5.11). Since $\widetilde{X}^c$ is independent of X^s, the positions of the nonzero elements of E are independent of X^s. Without loss of generality, we assume the m nonzero entries of E are

$$e_{11},\ e_{22},\ \cdots,\ e_{mm}, \quad m \leqslant N^{5\varepsilon}. \tag{2.5.12}$$

For other choices of the positions of nonzero entries, the proof is exactly the same, but we make this assumption to simplify the notations. By the definition of E, we have $|e_{ii}| \leqslant \omega$, $1 \leqslant i \leqslant m$. We define the matrices

$$H^s := \begin{pmatrix} 0 & \Sigma^{1/2} U^* X^s V \widetilde{\Sigma}^{1/2} \\ (\Sigma^{1/2} U^* X^s V \widetilde{\Sigma}^{1/2})^* & 0 \end{pmatrix}$$

and $H^E := H^s + P$, where

$$\begin{aligned} P := & \begin{pmatrix} 0 & \Sigma^{1/2} U^* E V \widetilde{\Sigma}^{1/2} \\ (\Sigma^{1/2} U^* E V \widetilde{\Sigma}^{1/2})^* & 0 \end{pmatrix} \\ = & \begin{pmatrix} \Sigma^{1/2} U^* & 0 \\ 0 & \widetilde{\Sigma}^{1/2} V^* \end{pmatrix} \begin{pmatrix} 0 & E \\ E^* & 0 \end{pmatrix} \begin{pmatrix} U \Sigma^{1/2} & 0 \\ 0 & V \widetilde{\Sigma}^{1/2} \end{pmatrix} \\ = & \begin{pmatrix} \Sigma^{1/2} U^* & 0 \\ 0 & \widetilde{\Sigma}^{1/2} V^* \end{pmatrix} W P_D W^* \begin{pmatrix} U \Sigma^{1/2} & 0 \\ 0 & V \widetilde{\Sigma}^{1/2} \end{pmatrix}, \end{aligned}$$

where P_D is a $2m \times 2m$ diagonal matrix

$$P_D = \operatorname{diag}\left(e_{11}, \ldots, e_{mm}, -e_{11}, \ldots, -e_{mm}\right),$$

and W is an $(n+N) \times 2m$ matrix such that

$$W_{ab} = \begin{cases} \delta_{a,i}/\sqrt{2} + \delta_{a,(n+i)}/\sqrt{2}, & b = i,\ i \leqslant m \\ \delta_{a,i}/\sqrt{2} - \delta_{a,(n+i)}/\sqrt{2}, & b = i+m,\ i \leqslant m \end{cases}.$$

With the identity

$$\det \begin{pmatrix} -I_{n\times n} & \Sigma^{1/2}U^* X V \widetilde{\Sigma}^{1/2} \\ (\Sigma^{1/2}U^* X V \widetilde{\Sigma}^{1/2})^* & -z I_{N\times N} \end{pmatrix} = z^{N-n} \det(-I_{N\times N}) \det\left(\widetilde{\mathcal{Q}}_1(X) - z I_{n\times n}\right),$$

and Lemma 6.1 of [58], we find that if $\mu \notin \sigma(\widetilde{\mathcal{Q}}_1(X^s))$, then μ is an eigenvalue of $\widetilde{\mathcal{Q}}_1(X^s + \gamma E)$ if and only if

$$\det\left(O^* G^s(\mu) O + (\gamma P_D)^{-1}\right) = 0, \tag{2.5.13}$$

where

$$G^s(\mu) := \left(H^s - \begin{pmatrix} I_{n\times n} & 0 \\ 0 & \mu I_{N\times N} \end{pmatrix}\right)^{-1}, \quad O := \begin{pmatrix} \Sigma^{1/2}U^* & 0 \\ 0 & \widetilde{\Sigma}^{1/2} V^* \end{pmatrix} W.$$

Define $R^\gamma := O^* G^s O + (\gamma P_D)^{-1}$ for $0 < \gamma < 1$. Now let $\mu := \lambda_1^s \pm N^{-3/4}$. We claim that

$$\mathbb{P}\left(\det R^\gamma(\mu) \neq 0 \text{ for all } 0 < \gamma \leqslant 1\right) = 1 - \mathrm{o}(1). \tag{2.5.14}$$

If (2.5.14) holds, then μ is not an eigenvalue of $\widetilde{\mathcal{Q}}_1(X + \gamma E)$ with probability $1 - \mathrm{o}(1)$. Denote the largest eigenvalue of $\widetilde{\mathcal{Q}}_1(X + \gamma E)$ by λ_1^γ, $0 < \gamma \leqslant 1$, and define $\lambda_1^0 := \lim_{\gamma \downarrow 0} \lambda_1^\gamma$. Then we have $\lambda_1^0 = \lambda_1^s$ and $\lambda_1^1 = \lambda_1^E$. With the continuity of λ_1^γ with respect to γ and the fact that $\lambda_1^0 \in (\lambda_1^s - N^{-3/4}, \lambda_1^s + N^{-3/4})$, we find that

$$\lambda_1^E = \lambda_1^1 \in (\lambda_1^s - N^{-3/4}, \lambda_1^s + N^{-3/4}),$$

with probability $1 - \mathrm{o}(1)$, which proves (2.5.11).

Finally, we prove (2.5.14). Note that $\eta_0 = \mathrm{O}(b_N/N) = \mathrm{O}(N^{-1+\varepsilon})$, hence $z = \lambda_r + \mathrm{i}N^{-2/3}$ is in $\widetilde{S}(c_0, C_0, \delta)$ for any small constant $\delta > 0$. Now we write

$$R^\gamma(\mu) = O^*\left(G^s(\mu) - G^s(z)\right)O + O^*\left(G^s(z) - \Pi(z)\right)O + O^*\Pi(z)O + (\gamma P_D)^{-1}. \tag{2.5.15}$$

With (2.4.19), we have

$$\|O^*\Pi(z)O\| = \mathrm{O}(1) \tag{2.5.16}$$

By Lemma 2.4.7, we have

$$\mathbb{E}\left|\left[O^*\left(G^s(z)-\Pi(z)\right)O\right]_{ab}\right|^2 \prec \Psi^2 = \mathrm{O}(N^{-2/3}), \quad 1 \leqslant a,b \leqslant 2m,$$

where we used (2.4.11) and (2.4.18) in the second step. Then with Markov's inequality and a union bound, we can get that

$$\max_{1\leqslant a,b\leqslant 2m}\left|\left[O^*\left(G^s(z)-\Pi(z)\right)O\right]_{ab}\right| \leqslant N^{-1/6} \tag{2.5.17}$$

holds with probability $1-\mathrm{O}(mN^{-1/3})$. Thus we have

$$\left\|O^*\left(G^s(z)-\Pi(z)\right)O\right\| = \mathrm{O}(mN^{-1/6}) = \mathrm{O}(1) \quad \text{with probability } 1-\mathrm{O}(mN^{-1/3}). \tag{2.5.18}$$

It remains to bound the first term in (2.5.15). As pointed out in Remark 2.4.11, we can extend (2.5.5) to finite correlation functions of the largest eigenvalues. Since the largest eigenvalues in the Gaussian case are separated in the scale $N^{-2/3}$, we conclude that

$$\mathbb{P}\left(\min_i |\lambda_i(\widetilde{Q}_1(X^s))-\mu| \geqslant N^{-3/4}\right) = 1-\mathrm{o}(1). \tag{2.5.19}$$

On the other hand, the rigidity result (2.4.30) gives that

$$|\mu-\lambda_r| \prec N^{-2/3}+N^{-3/4}. \tag{2.5.20}$$

Using (2.4.31), (2.5.19), (2.5.20) and the rigidity estimate (2.4.30), we can get that for any set Ω of deterministic unit vectors of cardinality $N^{\mathrm{O}(1)}$,

$$\sup_{\mathbf{u},\mathbf{v}\in\Omega}\left|\langle \mathbf{u}, \left(G^s(z)-G^s(\mu)\right)\mathbf{v}\rangle\right| \leqslant N^{-1/4+3\varepsilon} \tag{2.5.21}$$

with probability $1-\mathrm{o}(1)$. For instance, for deterministic unit vectors $\mathbf{u},\mathbf{v}\in\mathbb{C}^{\mathcal{I}_2}$ and any small constant $c>0$, we have with probability $1-\mathrm{o}(1)$ that

$$
\begin{aligned}
&|\langle \mathbf{u}, (G^s(z)-G^s(\mu))\,\mathbf{v}\rangle| \leqslant \sum_k |\langle \mathbf{u},\zeta_k\rangle\langle \mathbf{v},\zeta_k\rangle| \left|\frac{1}{\lambda_k-z}-\frac{1}{\lambda_k-\mu}\right| \\
&\prec \frac{1}{N^{2/3}}\sum_{\gamma_k\leqslant\lambda_r-c_1}|\langle \mathbf{u},\zeta_k\rangle\langle \mathbf{v},\zeta_k\rangle| + \frac{N^\varepsilon}{N^{5/3}}\sum_{\gamma_k>\lambda_r-c_1}\frac{1}{|\lambda_k-z||\lambda_k-\mu|} \\
&\leqslant \frac{1}{N^{2/3}} + \frac{N^\varepsilon}{N^{5/3}}\sum_{1\leqslant k\leqslant N^\varepsilon}\frac{1}{|\lambda_k-z||\lambda_k-\mu|} + \frac{N^\varepsilon}{N^{5/3}}\sum_{k>N^c,\gamma_k>\lambda_r-c_1}\frac{1}{|\lambda_k-z||\lambda_k-\mu|} \\
&\prec \frac{1}{N^{2/3}} + \frac{N^{2\varepsilon}}{N^{1/4}} + \frac{N^\varepsilon}{N^{2/3}}\left(\frac{1}{N}\sum_{k>N^\varepsilon,\gamma_k>\lambda_r-c}\frac{1}{|\lambda_k-z||\lambda_k-\mu|}\right) \prec N^{-1/4+2\varepsilon},
\end{aligned}
$$

where in the first step we used (2.4.5), in the second step (2.4.31) and $|\lambda_k-z||\lambda_k-\mu|\gtrsim 1$ for $\gamma_k\leqslant\lambda_r-c_1$ due to (2.4.30), in the third step the Cauchy-Schwarz inequality, in the fourth step (2.5.19), and in the last step $|\lambda_k-z||\lambda_k-\mu|\sim(k/N)^{-4/3}$ for $k>N^\varepsilon$ by the rigidity estimate (2.4.30). For the other choices of deterministic unit vectors $\mathbf{u},\mathbf{v}\in\mathbb{C}^{\mathcal{I}_{1,2}}$, we can prove (2.5.21) in a similar way. Now with (2.5.21), we can get that

$$
\|O^*\left(G^s(\mu)-G^s(z)\right)O\| = \mathrm{O}(mN^{-1/4+3\varepsilon}) \quad \text{with probability } 1-\mathrm{o}(1). \tag{2.5.22}
$$

With (2.5.16), (2.5.18) and (2.5.22), we see that as long as ω is chosen to be sufficiently small, we have

$$
\|O^*\left(G^s(\mu)-G^s(z)\right)O + O^*\left(G^s(z)-\Pi(z)\right)O + O^*\Pi(z)O\| < (\gamma\omega)^{-1}
$$

for all $0<\gamma\leqslant 1$ with probability $1-\mathrm{o}(1)$. This proves the claim (2.5.14), which further gives (2.5.11) and completes the proof.

2.6 Local laws for Gaussian ensembles

As discussed below Theorem 2.4.6, in this step we prove Theorem 2.4.6 for separable covariance matrices of the form $\Sigma^{1/2}X\widetilde{\Sigma}X^*\Sigma^{1/2}$, which will imply the local laws in the Gaussian X case. Thus in this section, we deal with the following resolvent:

$$
G(X,z)=\left[\begin{pmatrix} 0 & \Sigma^{1/2}X\widetilde{\Sigma}^{1/2} \\ \widetilde{\Sigma}^{1/2}X^*\Sigma^{1/2} & 0\end{pmatrix} - \begin{pmatrix} I_{n\times n} & 0 \\ 0 & zI_{N\times N}\end{pmatrix}\right]^{-1} \tag{2.6.1}
$$

with X satisfying (2.4.1) with $q=N^{-1/2}$. More precisely, we will prove the following result.

Proposition 2.6.1. *Suppose Assumption 2.2.1 and* (2.2.19) *hold. Suppose X satisfies the bounded support condition (2.4.1) with $q = N^{-1/2}$. Suppose A and B are diagonal, i.e. $U = I_{n\times n}$ and $V = I_{N\times N}$. Fix $C_0 > 1$ and let $c_0 > 0$ be a sufficiently small constant. Then for any fixed $\varepsilon > 0$, the following estimates hold.*

(1) ***Anisotropic local law****: For any $z \in S(c_0, C_0, \varepsilon)$ and deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$,*

$$|\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec \Psi(z). \tag{2.6.2}$$

(2) ***Averaged local law****: We have*

$$|m(z) - m_c(z)| \prec (N\eta)^{-1} \tag{2.6.3}$$

for any $z \in S(c_0, C_0, \varepsilon)$, and

$$|m(z) - m_c(z)| \prec \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}, \tag{2.6.4}$$

for any $z \in S(c_0, C_0, \varepsilon) \cap \{z = E + \mathrm{i}\eta : E \geqslant \lambda_r, N\eta\sqrt{\kappa+\eta} \geqslant N^{\varepsilon}\}$.

Both of the above estimates are uniform in the spectral parameter z and the deterministic vectors $\mathbf{u}, \mathbf{v}$.

The proof Proposition 2.6.1 is similar to the previous proof of the local laws, such as [14, 26, 59, 105]. Thus instead of giving all the details, we only describe briefly the proof. In particular, we shall focus on the key self-consistent equation argument, which is (almost) the only part that departs significantly from the previous proof in e.g. [14]. In the proof, we always denote the spectral parameter by $z = E + \mathrm{i}\eta$.

2.6.1 Basic resolvent estimates

In this subsection, we collect some basic tools that will be used. For simplicity, we denote $Y := \Sigma^{1/2} X \widetilde{\Sigma}^{1/2}$.

Definition 2.6.2 (Minors)**.** *For any $(n+N)\times(n+N)$ matrix $\mathcal{A}$ and $\mathbb{T} \subseteq \mathcal{I}$, we define the minor $\mathcal{A}^{(\mathbb{T})} := (\mathcal{A}_{ab} : a, b \in \mathcal{I}\backslash\mathbb{T})$ as the $(n+N-|\mathbb{T}|)\times(n+N-|\mathbb{T}|)$ matrix obtained* *by removing all rows and columns indexed by* $\mathbb{T}$. *Note that we keep the names of indices when defining* $\mathcal{A}^{(\mathbb{T})}$, *i.e.* $(\mathcal{A}^{(\mathbb{T})})_{ab} = \mathcal{A}_{ab}$ *for* $a, b \notin \mathbb{T}$. *Correspondingly, we define the resolvent minor as*

$$G^{(\mathbb{T})} := \left[\left(H - \begin{pmatrix} I_{n\times n} & 0 \\ 0 & zI_{N\times N} \end{pmatrix} \right)^{(\mathbb{T})} \right]^{-1} = \begin{pmatrix} z\mathcal{G}_1^{(\mathbb{T})} & \mathcal{G}_1^{(\mathbb{T})} Y^{(\mathbb{T})} \\ \left(Y^{(\mathbb{T})}\right)^* \mathcal{G}_1^{(\mathbb{T})} & \mathcal{G}_2^{(\mathbb{T})} \end{pmatrix}$$
$$= \begin{pmatrix} z\mathcal{G}_1^{(\mathbb{T})} & Y^{(\mathbb{T})} \mathcal{G}_2^{(\mathbb{T})} \\ \mathcal{G}_2^{(\mathbb{T})} \left(Y^{(\mathbb{T})}\right)^* & \mathcal{G}_2^{(\mathbb{T})} \end{pmatrix},$$

and the partial traces

$$m_1^{(\mathbb{T})} := \frac{1}{Nz} \sum_{i \notin \mathbb{T}} \sigma_i G_{ii}^{(\mathbb{T})}, \quad m_2^{(\mathbb{T})} := \frac{1}{N} \sum_{\mu \notin \mathbb{T}} \widetilde{\sigma}_\mu G_{\mu\mu}^{(\mathbb{T})}.$$

For convenience, we will adopt the convention that for any minor $\mathcal{A}^{(T)}$ *defined as above,* $\mathcal{A}_{ab}^{(T)} = 0$ *if* $a \in \mathbb{T}$ *or* $b \in \mathbb{T}$. *We will abbreviate* $(\{a\}) \equiv (a)$, $(\{a, b\}) \equiv (ab)$, *and* $\sum_a^{(\mathbb{T})} := \sum_{a \notin \mathbb{T}}$.

Lemma 2.6.3. *(Resolvent identities).*

(i) For $i \in \mathcal{I}_1$ *and* $\mu \in \mathcal{I}_2$, *we have*

$$\frac{1}{G_{ii}} = -1 - \left(YG^{(i)}Y^*\right)_{ii}, \quad \frac{1}{G_{\mu\mu}} = -z - \left(Y^*G^{(\mu)}Y\right)_{\mu\mu}. \tag{2.6.5}$$

(ii) For $i \neq j \in \mathcal{I}_1$ *and* $\mu \neq \nu \in \mathcal{I}_2$, *we have*

$$G_{ij} = G_{ii}G_{jj}^{(i)} \left(YG^{(ij)}Y^*\right)_{ij}, \quad G_{\mu\nu} = G_{\mu\mu}G_{\nu\nu}^{(\mu)} \left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu}. \tag{2.6.6}$$

For $i \in \mathcal{I}_1$ *and* $\mu \in \mathcal{I}_2$, *we have*

$$G_{i\mu} = G_{ii}G_{\mu\mu}^{(i)} \left(-Y_{i\mu} + \left(YG^{(i\mu)}Y\right)_{i\mu}\right), \quad G_{\mu i} = G_{\mu\mu}G_{ii}^{(\mu)} \left(-Y_{\mu i}^* + \left(Y^*G^{(\mu i)}Y^*\right)_{\mu i}\right). \tag{2.6.7}$$

(iii) For $a \in \mathcal{I}$ *and* $b, c \in \mathcal{I}\backslash\{a\}$,

$$G_{bc}^{(a)} = G_{bc} - \frac{G_{ba}G_{ac}}{G_{aa}}, \quad \frac{1}{G_{bb}} = \frac{1}{G_{bb}^{(a)}} - \frac{G_{ba}G_{ab}}{G_{bb}G_{bb}^{(a)}G_{aa}}. \tag{2.6.8}$$

(iv) All of the above identities hold for $G^{(\mathbb{T})}$ *instead of* G *for* $\mathbb{T} \subset \mathcal{I}$, *and in the case where* A *and* B *are not diagonal.*

Proof. All these identities can be proved using Schur's complement formula. The reader can refer to, for example, [59, Lemma 4.4]. □

Lemma 2.6.4. *Fix constants $c_0, C_0 > 0$. The following estimates hold uniformly for all $z \in S(c_0, C_0, a)$ for any $a \in \mathbb{R}$:*

$$\|G\| \leqslant C\eta^{-1}, \quad \|\partial_z G\| \leqslant C\eta^{-2}. \tag{2.6.9}$$

Furthermore, we have the following identities:

$$\sum_{i\in\mathcal{I}_1} |G_{ji}|^2 = \sum_{i\in\mathcal{I}_1} |G_{ij}|^2 = \frac{|z|^2}{\eta}\operatorname{Im}\left(\frac{G_{jj}}{z}\right), \tag{2.6.10}$$

$$\sum_{\mu\in\mathcal{I}_2} |G_{\nu\mu}|^2 = \sum_{\mu\in\mathcal{I}_2} |G_{\mu\nu}|^2 = \frac{\operatorname{Im} G_{\nu\nu}}{\eta}, \tag{2.6.11}$$

$$\sum_{i\in\mathcal{I}_1} |G_{\mu i}|^2 = \sum_{i\in\mathcal{I}_1} |G_{i\mu}|^2 = G_{\mu\mu} + \frac{\bar{z}}{\eta}\operatorname{Im} G_{\mu\mu}, \tag{2.6.12}$$

$$\sum_{\mu\in\mathcal{I}_2} |G_{i\mu}|^2 = \sum_{\mu\in\mathcal{I}_2} |G_{\mu i}|^2 = \frac{G_{ii}}{z} + \frac{\bar{z}}{\eta}\operatorname{Im}\left(\frac{G_{ii}}{z}\right). \tag{2.6.13}$$

All of the above estimates remain true for $G^{(\mathbb{T})}$ instead of G for any $\mathbb{T} \subseteq \mathcal{I}$, and in the case where A and B are not diagonal.

Proof. These estimates and identities can be proved through simple calculations with (2.4.4), (2.4.5) and (2.4.6). We refer the reader to [59, Lemma 4.6] and [105, Lemma 3.5]. □

Lemma 2.6.5. *Fix constants $c_0, C_0 > 0$. For any $\mathbb{T} \subseteq \mathcal{I}$ and $a \in \mathbb{R}$, the following bounds hold uniformly in $z \in S(c_0, C_0, a)$:*

$$\left|m_1 - m_1^{(\mathbb{T})}\right| + \left|m_2 - m_2^{(\mathbb{T})}\right| \leqslant \frac{C\,|\mathbb{T}|}{N\eta}, \tag{2.6.14}$$

where $C > 0$ is a constant depending only on τ.

Proof. For $\mu \in \mathcal{I}_2$, we have

$$\left|m_2 - m_2^{(\mu)}\right| = \frac{1}{N}\left|\sum_{\nu\in\mathcal{I}_2} \widetilde{\sigma}_\nu \frac{G_{\nu\mu}G_{\mu\nu}}{G_{\mu\mu}}\right| \leqslant \frac{C}{N|G_{\mu\mu}|}\sum_{\nu\in\mathcal{I}_2} |G_{\nu\mu}|^2 = \frac{C\operatorname{Im} G_{\mu\mu}}{N\eta|G_{\mu\mu}|} \leqslant \frac{C}{N\eta},$$

where in the first step we used (2.6.8), and in the second and third steps we used (2.6.11). Similarly, using (2.6.8) and (2.6.12) we get

$$\left|m_2 - m_2^{(i)}\right| = \frac{1}{N}\left|\sum_{\nu\in\mathcal{I}_2}\widetilde{\sigma}_\nu \frac{G_{\nu i}G_{i\nu}}{G_{ii}}\right| \leqslant \frac{C}{N|G_{ii}|}\left(\frac{G_{ii}}{z} + \frac{\bar{z}}{\eta}\mathrm{Im}\left(\frac{G_{ii}}{z}\right)\right) \leqslant \frac{C}{N\eta}.$$

Similarly, we can prove the same bounds for the m_1 case. Then (2.6.14) can be proved by induction on the indices in $\mathbb{T}$. □

The following lemma gives large deviation bounds for bounded supported random variables.

Lemma 2.6.6 (Lemma 3.8 of [34]). *Let (x_i), (y_j) be independent families of centered and independent random variables, and (A_i), (B_{ij}) be families of deterministic complex numbers. Suppose the entries x_i, y_j have variance at most N^{-1} and satisfy the bounded support condition (2.4.1) with $q \leqslant N^{-\varepsilon}$ for some constant $\varepsilon > 0$. Then we have the following bounds:*

$$\Big|\sum_i A_i x_i\Big| \prec q\max_i |A_i| + \frac{1}{\sqrt{N}}\Big(\sum_i |A_i|^2\Big)^{1/2}, \tag{2.6.15}$$

$$\Big|\sum_{i,j} x_i B_{ij} y_j\Big| \prec q^2 B_d + qB_o + \frac{1}{N}\Big(\sum_{i\neq j}|B_{ij}|^2\Big)^{1/2}, \tag{2.6.16}$$

$$\Big|\sum_i \bar{x}_i B_{ii} x_i - \sum_i (\mathbb{E}|x_i|^2) B_{ii}\Big| \prec qB_d, \tag{2.6.17}$$

$$\Big|\sum_{i\neq j} \bar{x}_i B_{ij} x_j\Big| \prec qB_o + \frac{1}{N}\left(\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}, \tag{2.6.18}$$

where $B_d := \max_i |B_{ii}|$ and $B_o := \max_{i\neq j}|B_{ij}|$.

For the proof of Proposition 2.6.1, it is convenient to introduce the following random control parameters.

Definition 2.6.7 (Control parameters). *We define the random errors*

$$\Lambda := \max_{a,b\in\mathcal{I}}\left|(G-\Pi)_{ab}\right|, \quad \Lambda_o := \max_{a\neq b\in\mathcal{I}}|G_{ab}|, \quad \theta := |m_1 - m_{1c}| + |m_2 - m_{2c}|, \tag{2.6.19}$$

and the random control parameter (recall Ψ defined in (2.4.18))

$$\Psi_\theta := \sqrt{\frac{\mathrm{Im}\, m_{2c} + \theta}{N\eta}} + \frac{1}{N\eta}. \tag{2.6.20}$$

2.6.2 Entrywise local law

The main goal of this subsection is to prove the following entrywise local law. The anisotropic local law (2.6.2) then follows from the entrywise local law combined with a polynomialization method as we will explain at the end of this subsection.

Proposition 2.6.8. *Suppose the assumptions in Proposition 2.6.1 hold. Fix $C_0 > 0$ and let $c_0 > 0$ be a sufficiently small constant. Then for any fixed $\varepsilon > 0$, the following estimate holds uniformly for $z \in S(c_0, C_0, \varepsilon)$:*

$$\max_{a,b} |G_{ab}(X,z) - \Pi_{ab}(z)| \prec \Psi(z). \tag{2.6.21}$$

In analogy to [34, Section 3] and [59, Section 5], we introduce the Z variables

$$Z_a^{(\mathbb{T})} := (1 - \mathbb{E}_a)\big(G_{aa}^{(\mathbb{T})}\big)^{-1}, \quad a \notin \mathbb{T}, \tag{2.6.22}$$

where $\mathbb{E}_a[\cdot] := \mathbb{E}[\cdot \mid H^{(a)}]$, i.e. it is the partial expectation over the randomness of the a-th row and column of H. By (2.6.5), we have

$$Z_i = (\mathbb{E}_i - 1)\left(YG^{(i)}Y^*\right)_{ii} = \sigma_i \sum_{\mu,\nu \in \mathcal{I}_2} \sqrt{\widetilde{\sigma}_\mu \widetilde{\sigma}_\nu} G_{\mu\nu}^{(i)} \left(\frac{1}{N}\delta_{\mu\nu} - X_{i\mu}X_{i\nu}\right), \tag{2.6.23}$$

$$Z_\mu = (\mathbb{E}_\mu - 1)\left(Y^*G^{(\mu)}Y\right)_{\mu\mu} = \widetilde{\sigma}_\mu \sum_{i,j \in \mathcal{I}_1} \sqrt{\sigma_i \sigma_j} G_{ij}^{(\mu)} \left(\frac{1}{N}\delta_{ij} - X_{i\mu}X_{j\mu}\right). \tag{2.6.24}$$

The following lemma plays a key role in the proof of local laws.

Lemma 2.6.9. *Suppose the assumptions in Proposition 2.6.1 hold. Let $c_0 > 0$ be a sufficiently small constant and fix $C_0, \varepsilon > 0$. Define the z-dependent event $\Xi(z) := \{\Lambda(z) \leqslant (\log N)^{-1}\}$. Then there exists constant $C > 0$ such that the following estimates hold uniformly for all $a \in \mathcal{I}$ and $z \in S(c_0, C_0, \varepsilon)$:*

$$\mathbf{1}(\Xi)\left(\Lambda_o + |Z_a|\right) \prec \Psi_\theta, \tag{2.6.25}$$

and

$$\mathbf{1}\left(\eta \geqslant 1\right)\left(\Lambda_o + |Z_a|\right) \prec \Psi_\theta. \tag{2.6.26}$$

Proof. Applying Lemma 2.6.6 to Z_i in (2.6.23), we get that on Ξ,

$$|Z_i| \prec q + \frac{1}{N}\left(\sum_{\mu,\nu} \widetilde{\sigma}_\mu \left|G^{(i)}_{\mu\nu}\right|^2\right)^{1/2} = q + \frac{1}{N}\left(\sum_{\mu} \frac{\widetilde{\sigma}_\mu \operatorname{Im} G^{(i)}_{\mu\mu}}{\eta}\right)^{1/2} = q + \sqrt{\frac{\operatorname{Im} m_2^{(i)}}{N\eta}}, \quad (2.6.27)$$

where we used (2.2.5), (2.6.11) and the fact that $\max_{a,b}|G_{ab}| = \mathrm{O}(1)$ on event Ξ. Now by (2.6.19), (2.6.20) and the bound (2.6.14), we have that

$$\sqrt{\frac{\operatorname{Im} m_2^{(i)}}{N\eta}} = \sqrt{\frac{\operatorname{Im} m_{2c} + \operatorname{Im}(m_2^{(i)} - m_2) + \operatorname{Im}(m_2 - m_{2c})}{N\eta}} \leqslant C\Psi_\theta. \quad (2.6.28)$$

Together with the fact that $q = N^{-1/2} \lesssim \Psi_\theta$ by (2.4.19), we get (2.6.25) for $\mathbf{1}(\Xi)|Z_i|$. Similarly, we can prove the same estimate for $\mathbf{1}(\Xi)|Z_\mu|$, where in the proof one need to use (2.6.10) and (2.4.19). If $\eta \geqslant 1$, we also have $\max_{a,b}|G_{ab}| = \mathrm{O}(1)$ by (2.6.9). Then repeating the above proof, we obtain (2.6.26) for $\mathbf{1}(\eta \geqslant 1)|Z_a|$. Similarly, using (2.6.6) and Lemmas 2.6.4-2.6.6, we can prove that

$$\mathbf{1}(\Xi)\left(|G_{ij}| + |G_{\mu\nu}|\right) + \mathbf{1}(\eta \geqslant 1)\left(|G_{ij}| + |G_{\mu\nu}|\right) \prec \Psi_\theta. \quad (2.6.29)$$

It remains to prove the bounds for $G_{i\mu}$ and $G_{\mu i}$ entries. Using (2.6.7), (2.4.1), the bound $\max_{a,b}|G_{ab}| = \mathrm{O}(1)$ on Ξ, Lemma 2.6.4 and Lemma 2.6.6, we get that

$$\begin{aligned}|G_{i\mu}| &\prec q + \frac{1}{N}\left(\sum_{j,\nu}^{(i\mu)} \widetilde{\sigma}_\nu \left|G^{(i\mu)}_{\nu j}\right|^2\right)^{1/2} = q + \frac{1}{N}\left(\sum_{\nu}^{(\mu)} \widetilde{\sigma}_\nu \left(G^{(i\mu)}_{\nu\nu} + \frac{\bar z}{\eta}\operatorname{Im} G^{(i\mu)}_{\nu\nu}\right)\right)^{1/2} \\ &\lesssim q + \sqrt{\frac{|m_2^{(i\mu)}|}{N}} + \sqrt{\frac{\operatorname{Im} m_2^{(i\mu)}}{N\eta}}.\end{aligned}$$

As in (2.6.28), we can show that

$$\sqrt{\frac{\operatorname{Im} m_2^{(i\mu)}}{N\eta}} = \mathrm{O}(\Psi_\theta). \quad (2.6.30)$$

For the other term, we have

$$\sqrt{\frac{|m_2^{(i\mu)}|}{N}} \leqslant \sqrt{\frac{|m_{2c}| + |m_2^{(i\mu)} - m_2| + |m_2 - m_{2c}|}{N}} \lesssim \frac{1}{N\sqrt{\eta}} + \sqrt{\frac{\theta}{N}} + \sqrt{\frac{|m_{2c}|}{N}} \lesssim \Psi_\theta, \quad (2.6.31)$$

where we used (2.6.14) and $|m_{2c}|N^{-1} = O(\Psi^2)$ by (2.4.19). With (2.6.30) and (2.6.31), we obtain that $\mathbf{1}(\Xi)|G_{i\mu}| \prec \Psi_\theta$. Together with (2.6.29), we get the estimate (2.6.25) for $\mathbf{1}(\Xi)\Lambda_o$. Finally, the estimate (2.6.26) for $\mathbf{1}(\eta \geqslant 1)\Lambda_o$ can be proved in a similar way with the bound $\mathbf{1}(\eta \geqslant 1)\max_{a,b}|G_{ab}| = \mathrm{O}(1)$. $\square$

A key component of the proof for Proposition 2.6.8 is an analysis of the self-consistent equation. Recall the equations in (2.2.9) and the function $f(z,m)$ in (2.2.15).

Lemma 2.6.10. *Let $c_0 > 0$ be a sufficiently small constant and fix $C_0, \varepsilon > 0$. Then the following estimates hold uniformly in $z \in S(c_0, C_0, \varepsilon)$:*

$$\mathbf{1}(\eta \geqslant 1)\,|f(z,m_2)| \prec N^{-1/2}, \quad \mathbf{1}(\eta \geqslant 1)\left| m_1(z) - d_N \int \frac{x}{-z\,[1+xm_2(z)]}\pi_A^{(n)}(dx)\right| \prec N^{-1/2}, \tag{2.6.32}$$

and

$$\mathbf{1}(\Xi)\,|f(z,m_2)| \prec \Psi_\theta, \quad \mathbf{1}(\Xi)\left| m_1(z) - d_N \int \frac{x}{-z\,[1+xm_2(z)]}\pi_A^{(n)}(dx)\right| \prec \Psi_\theta, \tag{2.6.33}$$

where Ξ is as given in Lemma 2.6.9. Moreover, we have the finer estimates

$$\mathbf{1}(\Xi)\,|f(z,m_2)| \prec \mathbf{1}(\Xi)\left(|[Z]_1| + |[Z]_2|\right) + \Psi_\theta^2, \tag{2.6.34}$$

and

$$\mathbf{1}(\Xi)\left| m_1(z) - d_N \int \frac{x}{-z\,[1+xm_2(z)]}\pi_A^{(n)}(dx)\right| \prec \mathbf{1}(\Xi)\,|[Z]_1| + \Psi_\theta^2, \tag{2.6.35}$$

where

$$[Z]_1 := \frac{1}{N}\sum_{i\in\mathcal{I}_1}\frac{\sigma_i}{(1+\sigma_i m_2)^2}Z_i, \quad [Z]_2 := \frac{1}{N}\sum_{\mu\in\mathcal{I}_2}\frac{\widetilde{\sigma}_\mu}{(1+\widetilde{\sigma}_\mu m_1)^2}Z_\mu. \tag{2.6.36}$$

Proof. We first prove (2.6.34) and (2.6.35), from which (2.6.33) follows due to (2.6.25) and (2.4.12). By (2.6.5), (2.6.23) and (2.6.24), we have

$$\frac{1}{G_{ii}} = -1 - \frac{\sigma_i}{N}\sum_{\mu\in\mathcal{I}_2}\widetilde{\sigma}_\mu G_{\mu\mu}^{(i)} + Z_i = -1 - \sigma_i m_2 + \varepsilon_i, \tag{2.6.37}$$

and

$$\frac{1}{G_{\mu\mu}} = -z - \frac{\widetilde{\sigma}_\mu}{N}\sum_{i\in\mathcal{I}_1}\sigma_i G_{ii}^{(\mu)} + Z_\mu = -z - z\widetilde{\sigma}_\mu m_1 + \varepsilon_\mu, \tag{2.6.38}$$

where

$$\varepsilon_i := Z_i + \sigma_i\left(m_2 - m_2^{(i)}\right) \quad \text{and} \quad \varepsilon_\mu := Z_\mu + z\widetilde{\sigma}_\mu\left(m_1 - m_1^{(\mu)}\right).$$

By (2.6.14) and (2.6.25), we have for all i and μ,

$$\mathbf{1}(\Xi)\left(|\varepsilon_i| + |\varepsilon_\mu|\right) \prec \Psi_\theta. \tag{2.6.39}$$

Moreover, by (2.6.8) we have

$$\mathbf{1}(\Xi)\left(|m_2-m_2^{(i)}|+|m_1-m_1^{(\mu)}|\right)\leqslant \mathbf{1}(\Xi)\frac{1}{N}\left(\sum_{\nu\in\mathcal{I}_2}\widetilde{\sigma}_\nu\left|\frac{G_{\nu i}G_{i\nu}}{G_{ii}}\right|+\sum_{j\in\mathcal{I}_1}\sigma_j\left|\frac{G_{j\mu}G_{\mu j}}{G_{\mu\mu}}\right|\right)\prec \Psi_\theta^2, \tag{2.6.40}$$

where we used (2.6.25) and $|G_{ii}|\sim|G_{\mu\mu}|\sim 1$ on Ξ in the second step. Now using (2.6.37), (2.6.39), (2.6.40), (2.6.25), (2.4.12) and the definition of Ξ, we can obtain that

$$\mathbf{1}(\Xi)G_{ii}=\mathbf{1}(\Xi)\left[\frac{1}{-(1+\sigma_i m_2)}-\frac{Z_i}{(1+\sigma_i m_2)^2}+O_\prec\left(\Psi_\theta^2\right)\right]. \tag{2.6.41}$$

Taking average $\frac{1}{Nz}\sum_i\sigma_i$, we get

$$\mathbf{1}(\Xi)m_1=\mathbf{1}(\Xi)\left[\frac{1}{N}\sum_i\frac{\sigma_i}{-z(1+\sigma_i m_2)}-z^{-1}[Z]_1+O_\prec\left(\Psi_\theta^2\right)\right], \tag{2.6.42}$$

which proves (2.6.35). On the other hand, using (2.6.38), (2.6.39), (2.6.40), (2.6.25), (2.4.12) and the definition of Ξ, we obtain that

$$\mathbf{1}(\Xi)G_{\mu\mu}=\mathbf{1}(\Xi)\left[\frac{1}{-z(1+\widetilde{\sigma}_\mu m_1)}-\frac{Z_\mu}{z^2\left(1+\widetilde{\sigma}_\mu m_1\right)^2}+O_\prec\left(\Psi_\theta^2\right)\right]. \tag{2.6.43}$$

Taking average $N^{-1}\sum_\mu\widetilde{\sigma}_\mu$, we get

$$\mathbf{1}(\Xi)m_2=\mathbf{1}(\Xi)\left[\frac{1}{N}\sum_\mu\frac{\widetilde{\sigma}_\mu}{-z(1+\widetilde{\sigma}_\mu m_1)}-z^{-2}[Z]_2+O_\prec\left(\Psi_\theta^2\right)\right]. \tag{2.6.44}$$

Plugging (2.6.42) into (2.6.44), and using (2.4.12) and the definition of Ξ, we can obtain that

$$\mathbf{1}(\Xi)m_2=\mathbf{1}(\Xi)\left[\frac{1}{N}\sum_\mu\frac{\widetilde{\sigma}_\mu}{-z+\frac{\widetilde{\sigma}_\mu}{N}\sum_i\frac{\sigma_i}{1+\sigma_i m_2}}+O_\prec\left(|[Z]_1|+|[Z]_2|+\Psi_\theta^2\right)\right]. \tag{2.6.45}$$

Comparing with (2.2.15), we have proved (2.6.34).

Then we prove (2.6.32). Using the bound $\mathbf{1}(\eta\geqslant 1)\max_{a,b}|G_{ab}|=\mathrm{O}(1)$, we trivially have $|m_1|+|m_2|+\theta=\mathrm{O}(1)$. Thus we have $\mathbf{1}(\eta\geqslant 1)\Psi_\theta=\mathrm{O}(N^{-1/2})$. Then (2.6.14) and (2.6.26) together give that

$$\mathbf{1}(\eta\geqslant 1)(|\varepsilon_i|+|\varepsilon_\mu|)\prec N^{-1/2}. \tag{2.6.46}$$

First we claim that in the case $\eta\geqslant 1$, with high probability,

$$|m_1|\geqslant \operatorname{Im} m_1\geqslant c,\quad |m_2|\geqslant \operatorname{Im} m_2\geqslant c, \tag{2.6.47}$$

for some constant $c > 0$. By the spectral decomposition (2.4.5), we have

$$\operatorname{Im} G_{ii} = \operatorname{Im} \sum_{k=1}^{M} \frac{z|\xi_k(i)|^2}{\lambda_k - z} = \sum_{k=1}^{M} |\xi_k(i)|^2 \operatorname{Im} \left(-1 + \frac{\lambda_k}{\lambda_k - z} \right) \geqslant 0.$$

Then by (2.6.38), $G_{\mu\mu}^{-1}$ is of order $O(1)$ and has imaginary part $\leqslant -\eta + O_{\prec}\left(N^{-1/2}\right)$. This implies $\operatorname{Im} G_{\mu\mu} \gtrsim \eta$ with high probability, which gives the second estimate of (2.6.47) by (2.2.5). Moreover, with (2.2.5) we also get that $\operatorname{Im}(1 + \sigma_i m_2) \gtrsim 1$ for $i \leqslant \tau n$. Then with (2.6.37) and a similar argument as above, we obtain the first estimate of (2.6.47). Next, we claim that in the case $\eta \geqslant 1$, with high probability,

$$|1 + \widetilde{\sigma}_\mu m_1| \geqslant c', \quad |1 + \sigma_i m_2| \geqslant c', \tag{2.6.48}$$

for some constant $c' > 0$. In fact, if $\sigma_i \leqslant 2|m_2|^{-1}$, we trivially have $|1 + \sigma_i m_2| \geqslant 1/2$. Otherwise, we have

$$|1 + \sigma_i m_2| \geqslant \frac{\operatorname{Im} m_2}{2|m_2|} \geqslant c'$$

by (2.6.47). The first estimate in (2.6.48) can be proved in the same way. Finally, with (2.6.46), (2.6.47) and (2.6.48), we can repeat the previous arguments between (2.6.37) and (2.6.45) to get (2.6.32). □

The following lemma gives the stability of the equation $f(z, m) = 0$. Roughly speaking, it states that if $f(z, m_2(z))$ is small and $m_2(\widetilde{z}) - m_{2c}(\widetilde{z})$ is small for $\operatorname{Im} \widetilde{z} \geqslant \operatorname{Im} z$, then $m_2(z) - m_{2c}(z)$ is small. For an arbitrary $z \in S(c_0, C_0, \varepsilon)$, we define the discrete set

$$L(w) := \{z\} \cup \{z' \in S(c_0, C_0, \varepsilon) : \operatorname{Re} z' = \operatorname{Re} z, \operatorname{Im} z' \in [\operatorname{Im} z, 1] \cap (N^{-10}\mathbb{N})\}.$$

Thus, if $\operatorname{Im} z \geqslant 1$, then $L(z) = \{z\}$; if $\operatorname{Im} z < 1$, then $L(z)$ is a 1-dimensional lattice with spacing N^{-10} plus the point z. Obviously, we have $|L(z)| \leqslant N^{10}$.

Lemma 2.6.11. *Let $c_0 > 0$ be a sufficiently small constant and fix $C_0, \varepsilon > 0$. The self-consistent equation $f(z, m) = 0$ is stable on $S(c_0, C_0, \varepsilon)$ in the following sense. Suppose the z-dependent function δ satisfies $N^{-2} \leqslant \delta(z) \leqslant (\log N)^{-1}$ for $z \in S(c_0, C_0, \varepsilon)$ and that δ is Lipschitz continuous with Lipschitz constant $\leqslant N^2$. Suppose moreover that for each fixed E, the function $\eta \mapsto \delta(E + \mathrm{i}\eta)$ is non-increasing for $\eta > 0$. Suppose that $u_2 : S(c_0, C_0, \varepsilon) \to \mathbb{C}$*

is the Stieltjes transform of a probability measure. Let $z \in S(c_0, C_0, \varepsilon)$ and suppose that for all $z' \in L(z)$ we have

$$|f(z, u_2)| \leqslant \delta(z). \tag{2.6.49}$$

Then we have

$$|u_2(z) - m_{2c}(z)| \leqslant \frac{C\delta}{\sqrt{\kappa + \eta + \delta}}, \tag{2.6.50}$$

for some constant $C > 0$ independent of z and N, where κ is defined in (2.4.9).

Proof. This lemma can proved with the same method as in e.g. [14, Lemma 4.5] and [59, Appendix A.2]. The only input is Lemma 2.2.6. □

Note that by Lemma 2.6.11 and (2.6.32), we immediately get that

$$\mathbf{1}(\eta \geqslant 1)\theta(z) \prec N^{-1/2}. \tag{2.6.51}$$

From (2.6.26), we obtain the off-diagonal estimate

$$\mathbf{1}(\eta \geqslant 1)\Lambda_o(z) \prec N^{-1/2}. \tag{2.6.52}$$

Using (2.6.37), (2.6.38) and (2.6.51), we get that

$$\mathbf{1}(\eta \geqslant 1)\left(|G_{ii} - \Pi_{ii}| + |G_{\mu\mu} - \Pi_{\mu\mu}|\right) \prec N^{-1/2}, \tag{2.6.53}$$

which gives the diagonal estimate. These bounds can be easily generalized to the case $\eta \geqslant c$ for any fixed $c > 0$. Compared with (2.6.21), one can see that the bounds (2.6.52) and (2.6.53) are optimal for the $\eta \geqslant c$ case. Now it remains to deal with the small η case (in particular, the local case with $\eta \ll 1$). We first prove the following weak bound.

Lemma 2.6.12 (Weak entrywise local law)**.** *Let $c_0 > 0$ be a sufficiently small constant and fix $C_0, \varepsilon > 0$. Then we have*

$$\Lambda(z) \prec (N\eta)^{-1/4}, \tag{2.6.54}$$

uniformly in $z \in S(c_0, C_0, \varepsilon)$.

Proof. One can prove this lemma using a continuity argument as in e.g. [14, Section 4.1], [33, Section 5.3] or [34, Section 3.6]. The key inputs are Lemmas 2.6.9-2.6.11, and the estimates (2.6.51)-(2.6.53) in the $\eta \geqslant 1$ case. All the other parts of the proof are essentially the same. □

To get the strong entrywise local law as in (2.6.21), we need stronger bounds on $[Z]_1$ and $[Z]_2$ in (2.6.34) and (2.6.35). They follow from the following *fluctuation averaging lemma*.

Lemma 2.6.13 (Fluctuation averaging). *Suppose Φ and Φ_o are positive, N-dependent deterministic functions on $S(c_0, C_0, \varepsilon)$ satisfying $N^{-1/2} \leqslant \Phi, \Phi_o \leqslant N^{-c}$ for some constant $c > 0$. Suppose moreover that $\Lambda \prec \Phi$ and $\Lambda_o \prec \Phi_o$. Then for all $z \in S(c_0, C_0, \varepsilon)$ we have*

$$|[Z]_1| + |[Z]_2| \prec \Phi_o^2. \tag{2.6.55}$$

Proof. We suppose that the event Ξ holds. The bound (2.6.55) can be proved in a similar way as [14, Lemma 4.9] and [33, Theorem 4.7]. Take $[Z]_1$ as an example. The only complication of the proof is that the coefficients $\sigma_i/(1+\sigma_i m_2)^2$ are random and depend on i. This can be dealt with by writing, for any $i \in \mathcal{I}_1$,

$$m_2 = m_2^{(i)} + \frac{1}{N}\sum_{\mu\in\mathcal{I}_2} \widetilde{\sigma}_\mu \frac{G_{\mu i}G_{i\mu}}{G_{ii}} = m_2^{(i)} + \mathrm{O}(\Lambda_o^2).$$

Then we write

$$\begin{aligned}
[Z]_1 &= \frac{1}{N}\sum_{i\in\mathcal{I}_1} \frac{\sigma_i}{\left(1+m_2^{(i)}\sigma_i\right)^2} Z_i + \mathrm{O}(\Lambda_o^2) = \frac{1}{N}\sum_{i\in\mathcal{I}_1}(1-\mathbb{E}_i)\left[\frac{\sigma_i}{\left(1+m_2^{(i)}\sigma_i\right)^2} G_{ii}^{-1}\right] + \mathrm{O}(\Lambda_o^2) \\
&= \frac{1}{N}\sum_{i\in\mathcal{I}_1}(1-\mathbb{E}_i)\left[\frac{\sigma_i}{\left(1+m_2\sigma_i\right)^2} G_{ii}^{-1}\right] + \mathrm{O}(\Lambda_o^2).
\end{aligned} \tag{2.6.56}$$

Now the method to bound the first term in the line (2.6.56) is only a slight modification of the one in [14] or [33]. For the proof of an even more complicated fluctuation averaging lemma, one can also refer to the Proof of Lemma 4.4.10 in Section **??**. Finally, we use that Ξ holds with high probability by Lemma 2.6.12 to conclude the proof. □

Now we give the proof of Proposition 2.6.8.

Proof of Proposition 2.6.8. .By Lemma 2.6.12, the event Ξ holds with high probability. Then by Lemma 2.6.12 and Lemma 2.6.9, we can take

$$\Phi_o = \sqrt{\frac{\operatorname{Im} m_{2c} + (N\eta)^{-1/4}}{N\eta}} + \frac{1}{N\eta}, \quad \Phi = \frac{1}{(N\eta)^{1/4}}, \tag{2.6.57}$$

in Lemma 2.6.13. Then (2.6.34) gives

$$|f(z, m_2)| \prec \frac{\operatorname{Im} m_{2c} + (N\eta)^{-1/4}}{N\eta}.$$

Using Lemma 2.6.11, we get

$$|m_2 - m_{2c}| \prec \frac{\operatorname{Im} m_{2c}}{N\eta\sqrt{\kappa+\eta}} + \frac{1}{(N\eta)^{5/8}} \prec \frac{1}{(N\eta)^{5/8}}, \tag{2.6.58}$$

where we used $\operatorname{Im} m_{2c} = \mathrm{O}(\sqrt{\kappa+\eta})$ by (2.4.11) in the second step. With (2.6.35) and (2.6.58), we get the same bound for m_1, which gives

$$\theta \prec (N\eta)^{-5/8}, \tag{2.6.59}$$

Then using Lemma 2.6.9 and (2.6.59), we obtain that

$$\Lambda_o \prec \sqrt{\frac{\operatorname{Im} m_{2c} + (N\eta)^{-5/8}}{N\eta}} + \frac{1}{N\eta} \tag{2.6.60}$$

uniformly in $z \in S(c_0, C_0, \varepsilon)$, which is a better bound than the one in (2.6.57). Taking the RHS of (2.6.60) as the new Φ_o, we can obtain an even better bound for Λ_o. Iterating the above arguments, we get the bound

$$\theta \prec (N\eta)^{-\sum_{k=1}^{l} 2^{-k} - 2^{-l-2}}$$

after l iterations. This implies

$$\theta \prec (N\eta)^{-1} \tag{2.6.61}$$

since l can be arbitrarily large. Now with (2.6.61), Lemma 2.6.9, (2.6.41) and (2.6.43), we can obtain (2.6.21). □

Proof of Proposition 2.6.1. We now can finish the proof of Proposition 2.6.1 using Proposition 2.6.8. By (2.6.41) and (2.6.61), we have

$$m = \frac{1}{n}\sum_i \frac{1}{-z(1+\sigma_i m_2)} - \frac{1}{n}\sum_i \frac{Z_i}{z\,(1+\sigma_i m_2)^2} + O_\prec\left(\Psi^2\right).$$

Using the same method as in Lemma 2.6.13, we can obtain that

$$\left|\frac{1}{n}\sum_i \frac{Z_i}{(1+\sigma_i m_2)^2}\right| \prec \Psi^2.$$

Together with (2.2.10), (2.4.12) and (2.6.61), we get that

$$|m - m_c| \prec (N\eta)^{-1} + \Psi^2 \prec (N\eta)^{-1},$$

where we used (2.4.19) in the second step. This proves (2.6.3).

For $z \in S_{out}(c_0, C_0, \varepsilon) := S(c_0, C_0, \varepsilon) \cap \{z = E + \mathrm{i}\eta : E \geqslant \lambda_r, N\eta\sqrt{\kappa+\eta} \geqslant N^{\varepsilon}\}$, we have

$$\Psi^2 \leqslant 2\left[\frac{\operatorname{Im} m_{2c}(z)}{N\eta} + \frac{1}{(N\eta)^2}\right] \lesssim \frac{1}{N\sqrt{\kappa+\eta}} + \frac{1}{(N\eta)^2} \lesssim \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}},$$

where we used (2.4.11) in the second step. Thus to prove (2.6.4), it suffices to prove that

$$|m_2 - m_{2c}| \prec \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}, \quad z \in S_{out}(c_0, C_0, \varepsilon). \tag{2.6.62}$$

In fact, taking $\Phi_o = \Phi = \Psi$ in Lemma 2.6.13 and then using Lemma 2.6.11, we get that

$$|m_2 - m_{2c}| \prec \frac{\Psi^2}{\sqrt{\kappa+\eta}} \lesssim \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}.$$

This finishes the proof of (2.6.62), and hence (2.6.4).

Finally, with (2.6.21), one can repeat the polynomialization method in [14, Section 5] to get the anisotropic local law (2.6.2). The only difference is that one need to use the first bound in (2.2.5). □

2.7 A self-consistent comparison approach

In this section, we finish the proof of Theorem 2.4.6 for a general X satisfying (2.4.15), (2.4.20) and the bounded support condition (2.4.1) with $q \leqslant N^{-\phi}$ for some constant $\phi > 0$. The proposition 2.6.1 implies that (2.4.22) holds for Gaussian X^{Gauss}. Thus the basic idea is to prove that for X satisfying the assumptions in Theorem 2.4.6,

$$\left\langle \mathbf{u}, \left(G(X,w) - G(X^{Gauss},w)\right)\mathbf{v}\right\rangle \prec q + \Psi(z)$$

uniformly for deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$ and $z \in \widetilde{S}(c_0, C_0, \varepsilon)$.

For simplicity of notations, we introduce the following generalized entries. For $\mathbf{v}, \mathbf{w} \in \mathbb{C}^{\mathcal{I}}$ and $a \in \mathcal{I}$, we shall denote

$$G_{\mathbf{vw}} := \langle \mathbf{v}, G\mathbf{w} \rangle, \quad G_{\mathbf{v}a} := \langle \mathbf{v}, G\mathbf{e}_a \rangle, \quad G_{a\mathbf{w}} := \langle \mathbf{e}_a, G\mathbf{w} \rangle, \tag{2.7.1}$$

where $\mathbf{e}_a$ is the standard unit vector along a-th axis. Given vectors $\mathbf{x} \in \mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{y} \in \mathbb{C}^{\mathcal{I}_2}$, we always identify them with their natural embeddings $\begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ \mathbf{y} \end{pmatrix}$ in $\mathbb{C}^{\mathcal{I}}$. The exact meanings will be clear from the context. Now similar to Lemma 2.6.4, we can prove the following estimates for $\mathcal{G}$.

Lemma 2.7.1. *For $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_2$, we define $\mathbf{u}_i = U^* \mathbf{e}_i \in \mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{v}_\mu = V^* \mathbf{e}_\mu \in \mathbb{C}^{\mathcal{I}_2}$, i.e. $\mathbf{u}_i$ is the i-th row vector of U and $\mathbf{v}_\mu$ is the μ-th row vector of V. Let $\mathbf{x} \in \mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{y} \in \mathbb{C}^{\mathcal{I}_2}$. Then we have*

$$\sum_{i \in \mathcal{I}_1} |G_{\mathbf{x}\mathbf{u}_i}|^2 = \sum_{i \in \mathcal{I}_1} |G_{\mathbf{u}_i\mathbf{x}}|^2 = \frac{|z|^2}{\eta} \operatorname{Im}\left(\frac{G_{\mathbf{xx}}}{z} \right), \tag{2.7.2}$$

$$\sum_{\mu \in \mathcal{I}_2} \left|G_{\mathbf{y}\mathbf{v}_\mu}\right|^2 = \sum_{\mu \in \mathcal{I}_2} \left|G_{\mathbf{v}_\mu\mathbf{y}}\right|^2 = \frac{\operatorname{Im} G_{\mathbf{yy}}}{\eta}, \tag{2.7.3}$$

$$\sum_{i \in \mathcal{I}_1} |G_{\mathbf{y}\mathbf{u}_i}|^2 = \sum_{i \in \mathcal{I}_1} |G_{\mathbf{u}_i\mathbf{y}}|^2 = G_{\mathbf{yy}} + \frac{\bar{z}}{\eta} \operatorname{Im} G_{\mathbf{yy}}, \tag{2.7.4}$$

$$\sum_{\mu \in \mathcal{I}_2} \left|G_{\mathbf{x}\mathbf{v}_\mu}\right|^2 = \sum_{\mu \in \mathcal{I}_2} \left|G_{\mathbf{v}_\mu\mathbf{x}}\right|^2 = \frac{G_{\mathbf{xx}}}{z} + \frac{\bar{z}}{\eta} \operatorname{Im}\left(\frac{G_{\mathbf{xx}}}{z} \right). \tag{2.7.5}$$

All of the above estimates remain true for $G^{(\mathbb{T})}$ instead of G for any $\mathbb{T} \subseteq \mathcal{I}$.

Proof. We only prove (2.7.3) and (2.7.4). The proof for (2.7.2) and (2.7.5) is very similar. With (2.4.5), we get that

$$\sum_{\mu \in \mathcal{I}_2} \left|G_{\mathbf{y}\mathbf{v}_\mu}\right|^2 = \sum_{\mu \in \mathcal{I}_2} \langle \mathbf{y}, G\mathbf{v}_\mu \rangle \langle \mathbf{v}_\mu, G^\dagger \mathbf{y} \rangle = \sum_{k=1}^{N} \frac{|\langle \mathbf{y}, \zeta_k \rangle|^2}{(\lambda_k - E)^2 + \eta^2} = \frac{\operatorname{Im} G_{\mathbf{yy}}}{\eta}. \tag{2.7.6}$$

For simplicity, we denote $Y := \Sigma^{1/2} U^* X V \widetilde{\Sigma}^{1/2}$. Then with (2.4.4) and (2.4.6), we get that

$$\sum_{i \in \mathcal{I}_1} |G_{\mathbf{y}\mathbf{u}_i}|^2 = \left(\mathcal{G}_2 Y^\dagger Y \mathcal{G}_2^\dagger \right)_{\mathbf{yy}} = \left(\mathcal{G}_2 \left(Y^\dagger Y - \bar{z} \right) \mathcal{G}_2^\dagger \right)_{\mathbf{yy}} + \bar{z} \left(\mathcal{G}_2 \mathcal{G}_2^\dagger \right)_{\mathbf{yy}} = G_{\mathbf{yy}} + \frac{\bar{z}}{\eta} \operatorname{Im} G_{\mathbf{yy}},$$

where we used $\mathcal{G}_2^\dagger = \left(Y^\dagger Y - \bar{z} \right)^{-1}$ and (2.7.6) in the last step. □

Our proof basically follows the arguments in [59, Section 7] with some modifications. Thus we will not give all the details. We first focus on proving the anisotropic local law (2.4.22), and the proof of (2.4.23)-(2.4.24) will be given at the end of this section. By polarization, to prove (2.4.22) it suffices to prove that

$$\langle \mathbf{v}, (G(X,z) - \Pi(z))\, \mathbf{v} \rangle \prec q + \Psi(z) \tag{2.7.7}$$

uniformly in $z \in \widetilde{S}(c_0, C_0, \varepsilon)$ and any deterministic unit vector $\mathbf{v} \in \mathbb{C}^{\mathcal{I}}$. In fact, we can obtain the more general bound (2.4.22) by applying (2.7.7) to the vectors $\mathbf{u}+\mathbf{v}$ and $\mathbf{u}+i\mathbf{v}$, respectively.

The proof consists of a bootstrap argument from larger scales to smaller scales in multiplicative increments of $N^{-\delta}$, where

$$\delta \in \left(0, \frac{\min\{\varepsilon, \phi\}}{2C_a}\right). \tag{2.7.8}$$

Here $\varepsilon > 0$ is the constant in $\widetilde{S}(c_0, C_0, \varepsilon)$, $\phi > 0$ is a constant such that $q \leqslant N^{-\phi}$, $C_a > 0$ is an absolute constant that will be chosen large enough in the proof. For any $\eta \geqslant N^{-1+\varepsilon}$, we define

$$\eta_l := \eta N^{\delta l} \text{ for } \ l = 0, ..., L-1, \quad \eta_L := 1. \tag{2.7.9}$$

where $L \equiv L(\eta) := \max\left\{l \in \mathbb{N} |\ \eta N^{\delta(l-1)} < 1\right\}$. Note that $L \leqslant \delta^{-1}$.

By (2.6.9), the function $z \mapsto G(z) - \Pi(z)$ is Lipschitz continuous in $\widetilde{S}(c_0, C_0, \varepsilon)$ with Lipschitz constant bounded by N^2. Thus to prove (2.7.7) for all $z \in \widetilde{S}(c_0, C_0, \varepsilon)$, it suffices to show that (2.7.7) holds for all z in some discrete but sufficiently dense subset $\mathbf{S} \subset \widetilde{S}(c_0, C_0, \varepsilon)$. We will use the following discretized domain $\mathbf{S}$.

Definition 2.7.2. *Let* $\mathbf{S}$ *be an* N^{-10}*-net of* $\widetilde{S}(c_0, C_0, \varepsilon)$ *such that* $|\mathbf{S}| \leqslant N^{20}$ *and*

$$E + \mathrm{i}\eta \in \mathbf{S} \Rightarrow E + \mathrm{i}\eta_l \in \mathbf{S} \text{ for } l = 1, ..., L(\eta).$$

The bootstrapping is formulated in terms of two scale-dependent properties ($\mathbf{A}_m$) and ($\mathbf{C}_m$) defined on the subsets

$$\mathbf{S}_m := \left\{ z \in \mathbf{S} \mid \operatorname{Im} z \geqslant N^{-\delta m} \right\}.$$

($\mathbf{A}_m$) For all $z \in \mathbf{S}_m$, all deterministic unit vectors $\mathbf{x} \in \mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{y} \in \mathbb{C}^{\mathcal{I}_2}$, and all X satisfying the assumptions in Theorem 2.4.6, we have

$$\operatorname{Im}\left(\frac{G_{\mathbf{xx}}(z)}{z}\right) + \operatorname{Im} G_{\mathbf{yy}}(z) \prec \operatorname{Im} m_{2c}(z) + N^{C_a\delta}(q + \Psi(z)). \tag{2.7.10}$$

($\mathbf{C}_m$) For all $z \in \mathbf{S}_m$, all deterministic unit vector $\mathbf{v} \in \mathbb{C}^{\mathcal{I}}$, and all X satisfying the assumptions in Theorem 2.4.6, we have

$$|G_{\mathbf{vv}}(z) - \Pi_{\mathbf{vv}}(z)| \prec N^{C_a\delta}(q + \Psi(z)). \tag{2.7.11}$$

It is trivial to see that ($\mathbf{A}_0$) holds by (2.6.9) and (2.4.11). Moreover, it is easy to observe the following result.

Lemma 2.7.3. *For any m, property ($\mathbf{C}_m$) implies property ($\mathbf{A}_m$).*

Proof. By (2.4.11), (2.4.12) and the definition of Π in (2.4.16), it is easy to get that

$$\operatorname{Im}\left(\frac{\Pi_{\mathbf{xx}}(z)}{z}\right) + \operatorname{Im} \Pi_{\mathbf{yy}}(z) \lesssim \operatorname{Im} m_{2c}(z),$$

which finishes the proof. □

The key step is the following induction result.

Lemma 2.7.4. *For any $1 \leqslant m \leqslant \delta^{-1}$, property ($\mathbf{A}_{m-1}$) implies property ($\mathbf{C}_m$).*

Combining Lemmas 2.7.3 and 2.7.4, we conclude that (2.7.11) holds for all $w \in \mathbf{S}$. Since δ can be chosen arbitrarily small under the condition (2.7.8), we conclude that (2.7.7) holds for all $w \in \mathbf{S}$, and (2.4.22) follows for all $z \in \widetilde{S}(c_0, C_0, \varepsilon)$. What remains now is the proof of Lemma 2.7.4. Denote

$$F_{\mathbf{v}}(X, z) := |G_{\mathbf{vv}}(X, z) - \Pi_{\mathbf{vv}}(z)|. \tag{2.7.12}$$

By Markov's inequality, it suffices to prove the following lemma.

Lemma 2.7.5. *Fix $p \in \mathbb{N}$ and $m \leqslant \delta^{-1}$. Suppose that the assumptions of Theorem 2.4.6 and property ($\mathbf{A}_{m-1}$) hold. Then we have*

$$\mathbb{E} F_{\mathbf{v}}^p(X, z) \leqslant \left[N^{C_a\delta}(q + \Psi(z))\right]^p \tag{2.7.13}$$

for all $z \in \mathbf{S}_m$ and any deterministic unit vector $\mathbf{v}$.

In the rest of this section, we focus on proving Lemma 2.7.5. First, in order to make use of the assumption $(\mathbf{A}_{m-1})$, which has spectral parameters in $\mathbf{S}_{m-1}$, to get some estimates for G with spectral parameters in $\mathbf{S}_m$, we shall use the following rough bounds for $G_{\mathbf{xy}}$.

Lemma 2.7.6. *For any $z = E + \mathrm{i}\eta \in \mathbf{S}$ and unit vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{\mathcal{I}}$, we have*

$$|G_{\mathbf{xy}}(z) - \Pi_{\mathbf{xy}}(z)| \prec N^{2\delta} \sum_{l=1}^{L(\eta)} \left[\mathrm{Im}\left(\frac{G_{\mathbf{x}_1\mathbf{x}_1}(E+\mathrm{i}\eta_l)}{E+\mathrm{i}\eta_l}\right) + \mathrm{Im}\, G_{\mathbf{x}_2\mathbf{x}_2}(E+\mathrm{i}\eta_l) \right.$$
$$\left. + \mathrm{Im}\left(\frac{G_{\mathbf{y}_1\mathbf{y}_1}(E+\mathrm{i}\eta_l)}{E+\mathrm{i}\eta_l}\right) + \mathrm{Im}\, G_{\mathbf{y}_2\mathbf{y}_2}(E+\mathrm{i}\eta_l)\right] + 1,$$

where $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}$ for $\mathbf{x}_1, \mathbf{y}_1 \in \mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{x}_2, \mathbf{y}_2 \in \mathbb{C}^{\mathcal{I}_2}$, and η_l is defined in (2.7.9).

Proof. The proof is the same as the one for [59, Lemma 7.12]. □

Recall that for a given family of random matrices A, we use $A = O_\prec(\zeta)$ to mean $|\langle \mathbf{v}, A\mathbf{w}\rangle| \prec \zeta \|\mathbf{v}\|_2 \|\mathbf{w}\|_2$ uniformly in any deterministic vectors $\mathbf{v}$ and $\mathbf{w}$ (see Definition 2.4.1 (ii)).

Lemma 2.7.7. *Suppose $(\mathbf{A}_{m-1})$ holds, then*

$$G(z) - \Pi(z) = \mathrm{O}_\prec(N^{2\delta}), \tag{2.7.14}$$

and

$$\mathrm{Im}\left(\frac{G_{\mathbf{xx}}(z)}{z}\right) + \mathrm{Im}\, G_{\mathbf{yy}}(z) \prec N^{2\delta}\left[\mathrm{Im}\, m_{2c}(z) + N^{C_a\delta}(q + \Psi(z))\right], \tag{2.7.15}$$

for all $z \in \mathbf{S}_m$ and any deterministic unit vectors $\mathbf{x} \in \mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{y} \in \mathbb{C}^{\mathcal{I}_2}$.

Proof. The proof is the same as the one for [59, Lemma 7.13]. □

Now we are ready to perform the self-consistent comparison. We divide the proof into three subsections. In Sections 2.7.1-2.7.2, we prove Lemma 2.7.5 under the condition

$$\mathbb{E}x_{ij}^3 = 0, \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant N, \tag{2.7.16}$$

for $z \in S(c_0, C_0, \varepsilon)$. Then in Section 2.7.3, we show how to relax (2.7.16) to (2.4.20) for $z \in \widetilde{S}(c_0, C_0, \varepsilon)$.

2.7.1 Interpolation and expansion

Definition 2.7.8 (Interpolating matrices)**.** *Introduce the notations* $X^0 := X^{Gauss}$ *and* $X^1 := X$*. Let* $\rho^0_{i\mu}$ *and* $\rho^1_{i\mu}$ *be the laws of* $X^0_{i\mu}$ *and* $X^1_{i\mu}$*, respectively. For* $\theta \in [0,1]$*, we define the interpolated law*

$$\rho^\theta_{i\mu} := (1-\theta)\rho^0_{i\mu} + \theta\rho^1_{i\mu}.$$

We shall work on the probability space consisting of triples (X^0, X^θ, X^1) *of independent* $\mathcal{I}_1 \times \mathcal{I}_2$ *random matrices, where the matrix* $X^\theta = (X^\theta_{i\mu})$ *has law*

$$\prod_{i\in\mathcal{I}_1}\prod_{\mu\in\mathcal{I}_2}\rho^\theta_{i\mu}(dX^\theta_{i\mu}). \tag{2.7.17}$$

For $\lambda \in \mathbb{R}$*,* $i \in \mathcal{I}_1$ *and* $\mu \in \mathcal{I}_2$*, we define the matrix* $X^{\theta,\lambda}_{(i\mu)}$ *through*

$$\left(X^{\theta,\lambda}_{(i\mu)}\right)_{j\nu} := \begin{cases} X^\theta_{i\mu}, & \text{if } (j,\nu) \neq (i,\mu) \\ \lambda, & \text{if } (j,\nu) = (i,\mu) \end{cases}.$$

We also introduce the matrices

$$G^\theta(z) := G\left(X^\theta, z\right), \quad G^{\theta,\lambda}_{(i\mu)}(z) := G\left(X^{\theta,\lambda}_{(i\mu)}, z\right).$$

We shall prove Lemma 2.7.5 through interpolation matrices X^θ between X^0 and X^1. It holds for X^0 by Proposition 2.6.1.

Lemma 2.7.9. *Lemma 2.7.5 holds if* $X = X^0$*.*

Using (2.7.17) and fundamental calculus, we get the following basic interpolation formula.

Lemma 2.7.10. *For* $F : \mathbb{R}^{\mathcal{I}_1\times\mathcal{I}_2} \to \mathbb{C}$ *we have*

$$\frac{d}{d\theta}\mathbb{E}F(X^\theta) = \sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left[\mathbb{E}F\left(X^{\theta,X^1_{i\mu}}_{(i\mu)}\right) - \mathbb{E}F\left(X^{\theta,X^0_{i\mu}}_{(i\mu)}\right)\right] \tag{2.7.18}$$

provided all the expectations exist.

We shall apply Lemma 2.7.10 with $F(X) = F^p_{\mathbf{v}}(X,z)$ for $F_{\mathbf{v}}(X,z)$ defined in (2.7.12). The main work is devoted to proving the following self-consistent estimate for the right-hand side of (2.7.18).

Lemma 2.7.11. *Fix $p \in 2\mathbb{N}$ and $m \leqslant \delta^{-1}$. Suppose (2.7.16) and $(\mathbf{A_{m-1}})$ hold, then we have*

$$\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left[\mathbb{E}F_{\mathbf{v}}^p\left(X_{(i\mu)}^{\theta,X_{i\mu}^1},z\right)-\mathbb{E}F_{\mathbf{v}}^p\left(X_{(i\mu)}^{\theta,X_{i\mu}^0},z\right)\right]=\mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi(z))\right]^p+\mathbb{E}F_{\mathbf{v}}^p(X^\theta,z)\right) \tag{2.7.19}$$

for all $\theta \in [0,1]$, $z \in \mathbf{S}_m$ and any deterministic unit vector $\mathbf{v}$.

Combining Lemmas 2.7.9-2.7.11 with a Grönwall's argument, we can conclude Lemma 2.7.5 and hence (2.7.7). In order to prove Lemma 2.7.11, we compare $X_{(i\mu)}^{\theta,X_{i\mu}^0}$ and $X_{(i\mu)}^{\theta,X_{i\mu}^1}$ via a common $X_{(i\mu)}^{\theta,0}$, i.e. we will prove that

$$\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left[\mathbb{E}F_{\mathbf{v}}^p\left(X_{(i\mu)}^{\theta,X_{i\mu}^u},z\right)-\mathbb{E}F_{\mathbf{v}}^p\left(X_{(i\mu)}^{\theta,0},z\right)\right]=\mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi(z))\right]^p+\mathbb{E}F_{\mathbf{v}}^p(X^\theta,z)\right) \tag{2.7.20}$$

for all $u \in \{0,1\}$, $\theta \in [0,1]$, $w \in \mathbf{S}_m$, and any deterministic unit vector $\mathbf{v}$.

Underlying the proof of (2.7.20) is an expansion approach which we will describe below. During the proof, we always assume that $(\mathbf{A}_{m-1})$ holds. Also the rest of the proof is performed at a fixed $z \in \mathbf{S}_m$. We define the $\mathcal{I}\times\mathcal{I}$ matrix $\Delta_{(i\mu)}^\lambda$ as

$$\Delta_{(i\mu)}^\lambda := \lambda\begin{pmatrix} 0 & \Sigma^{1/2}\mathbf{u}_i\mathbf{v}_\mu^*\widetilde{\Sigma}^{1/2} \\ \widetilde{\Sigma}^{1/2}\mathbf{v}_\mu\mathbf{u}_i^*\Sigma^{1/2} & 0\end{pmatrix}, \tag{2.7.21}$$

where we recall the definitions of $\mathbf{u}_i$ and $\mathbf{v}_\mu$ in Lemma 2.7.1. Then we have for any $\lambda,\lambda' \in \mathbb{R}$ and $K \in \mathbb{N}$,

$$\mathrm{G}_{(i\mu)}^{\theta,\lambda'} = G_{(i\mu)}^{\theta,\lambda}+\sum_{k=1}^K G_{(i\mu)}^{\theta,\lambda}\left(\Delta_{(i\mu)}^{\lambda-\lambda'}G_{(i\mu)}^{\theta,\lambda}\right)^k+G_{(i\mu)}^{\theta,\lambda'}\left(\Delta_{(i\mu)}^{\lambda-\lambda'}G_{(i\mu)}^{\theta,\lambda}\right)^{K+1}. \tag{2.7.22}$$

The following result provides a priori bounds for the entries of $G_{(i\mu)}^{\theta,\lambda}$.

Lemma 2.7.12. *Suppose that y is a random variable satisfying $|y| \prec q$. Then*

$$G_{(i\mu)}^{\theta,y}-\Pi = \mathrm{O}_\prec(N^{2\delta}),\quad i\in\mathcal{I}_1,\ \mu\in\mathcal{I}_2\,. \tag{2.7.23}$$

Proof. The proof is the same as the one for [59, Lemma 7.14]. □

In the following proof, for simplicity of notations, we introduce $f_{(i\mu)}(\lambda) := F_{\mathbf{v}}^p(X_{(i\mu)}^{\theta,\lambda})$. We use $f_{(i\mu)}^{(r)}$ to denote the r-th derivative of $f_{(i\mu)}$. With Lemma 2.7.12 and (2.7.22), it is easy to prove the following result.

Lemma 2.7.13. *Suppose that y is a random variable satisfying $|y| \prec q$. Then for fixed $r \in \mathbb{N}$,*

$$\left|f_{(i\mu)}^{(r)}(y)\right| \prec N^{2\delta(r+p)}. \tag{2.7.24}$$

By this lemma, the Taylor expansion of $f_{(i\mu)}$ gives

$$f_{(i\mu)}(y) = \sum_{r=0}^{4p+4} \frac{y^r}{r!} f_{(i\mu)}^{(r)}(0) + \mathrm{O}_\prec\left(q^{p+4}\right), \tag{2.7.25}$$

provided C_a is chosen large enough in (2.7.8). Therefore we have for $u \in \{0,1\}$,

$$\begin{aligned}\mathbb{E}F_{\mathbf{v}}^p\left(X_{(i\mu)}^{\theta,X_{i\mu}^u}\right) - \mathbb{E}F_{\mathbf{v}}^p\left(X_{(i\mu)}^{\theta,0}\right) &= \mathbb{E}\left[f_{(i\mu)}\left(X_{i\mu}^u\right) - f_{(i\mu)}(0)\right] \\ &= \mathbb{E}\, f_{(i\mu)}(0) + \frac{1}{2N}\,\mathbb{E}\, f_{(i\mu)}^{(2)}(0) + \sum_{r=4}^{4p+4} \frac{1}{r!}\,\mathbb{E}\, f_{(i\mu)}^{(r)}(0)\,\mathbb{E}\left(X_{i\mu}^u\right)^r + \mathrm{O}_\prec(q^{p+4}),\end{aligned}$$

where we used that $X_{i\mu}^u$ has vanishing first and third moments and its variance is $1/N$. (Note that this is the only place where we need the condition (2.7.16).) By (2.4.15) and the bounded support condition, we have

$$\left|\mathbb{E}\left(X_{i\mu}^u\right)^r\right| \prec N^{-2} q^{r-4}, \quad r \geqslant 4. \tag{2.7.26}$$

Thus to show (2.7.20), we only need to prove for $r = 4, 5, ..., 4p+4$,

$$N^{-2} q^{r-4} \sum_{i\in\mathcal{I}_1} \sum_{\mu\in\mathcal{I}_2} \left|\mathbb{E}\, f_{(i\mu)}^{(r)}(0)\right| = \mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi)\right]^p + \mathbb{E}F_{\mathbf{v}}^p(X^\theta, z)\right). \tag{2.7.27}$$

In order to get a self-consistent estimate in terms of the matrix X^θ on the right-hand side of (2.7.27), we want to replace $X_{(i\mu)}^{\theta,0}$ in $f_{(i\mu)}(0) = F_{\mathbf{v}}^p(X_{(i\mu)}^{\theta,0})$ with $X^\theta = X_{(i\mu)}^{\theta,X_{i\mu}^\theta}$.

Lemma 2.7.14. *Suppose that*

$$N^{-2} q^{r-4} \sum_{i\in\mathcal{I}_1} \sum_{\mu\in\mathcal{I}_2} \left|\mathbb{E}\, f_{(i\mu)}^{(r)}(X_{i\mu}^\theta)\right| = \mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi)\right]^p + \mathbb{E}F_{\mathbf{v}}^p(X^\theta, z)\right) \tag{2.7.28}$$

holds for $r = 4, ..., 4p+4$. Then (2.7.27) holds for $r = 4, ..., 4p+4$.

Proof. We abbreviate $f_{(i\mu)} \equiv f$ and $X_{i\mu}^\theta \equiv \xi$. Then with (2.7.25) we can get

$$\mathbb{E}f^{(l)}(0) = \mathbb{E}f^{(l)}(\xi) - \sum_{k=1}^{4p+4-l} \mathbb{E}f^{(l+k)}(0)\frac{\mathbb{E}\xi^k}{k!} + \mathrm{O}_\prec(q^{p+4-l}). \tag{2.7.29}$$

The estimate (2.7.27) then follows from a repeated application of (2.7.29). Fix $r = 4, ..., 4p+4$. Using (2.7.29), we get

$$\begin{aligned}
&\mathbb{E}f^{(r)}(0) \\
&= \mathbb{E}f^{(r)}(\xi) - \sum_{k_1 \geqslant 1} \mathbf{1}(r + k_1 \leqslant 4p+4)\mathbb{E}f^{(r+k_1)}(0)\frac{\mathbb{E}\xi^{k_1}}{k_1!} + \mathrm{O}_{\prec}(q^{p+4-r}) \\
&= \mathbb{E}f^{(r)}(\xi) - \sum_{k_1 \geqslant 1} \mathbf{1}(r + k_1 \leqslant 4p+4)\mathbb{E}f^{(r+k_1)}(\xi)\frac{\mathbb{E}\xi^{k_1}}{k_1!} \\
&+ \sum_{k_1, k_2 \geqslant 1} \mathbf{1}(r + k_1 + k_2 \leqslant 4p+4)\mathbb{E}f^{(r+k_1+k_2)}(0)\frac{\mathbb{E}\xi^{k_1}}{k_1!}\frac{\mathbb{E}\xi^{k_2}}{k_2!} + \mathrm{O}_{\prec}(q^{p+4-r}) \\
&= \cdots = \sum_{t=0}^{4p+4-r} (-1)^t \sum_{k_1, \cdots, k_t \geqslant 1} \mathbf{1}\left(r + \sum_{j=1}^{t} k_j \leqslant 4p+4\right) \mathbb{E}f^{(r+\sum_{j=1}^{t} k_j)}(\xi) \prod_{j=1}^{t} \frac{\mathbb{E}\xi^{k_j}}{k_j!} + \mathrm{O}_{\prec}(q^{p+4-r}).
\end{aligned}$$

The lemma now follows easily by using (2.7.26). □

2.7.2 Proof with "Words"

What remains now is to prove (2.7.28). For simplicity, we abbreviate $X^{\theta} \equiv X$. In order to exploit the detailed structure of the derivatives on the left-hand side of (2.7.28), we introduce the following algebraic objects.

Definition 2.7.15 (Words). *Given $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_2$. Let $\mathcal{W}$ be the set of words of even length in two letters $\{\mathbf{i}, \boldsymbol{\mu}\}$. We denote the length of a word $w \in \mathcal{W}$ by $2m(w)$ with $m(w) \in \mathbb{N}$. We use bold symbols to denote the letters of words. For instance, $w = \mathbf{t}_1\mathbf{s}_2\mathbf{t}_2\mathbf{s}_3 \cdots \mathbf{t}_r\mathbf{s}_{r+1}$ denotes a word of length $2r$. Define $\mathcal{W}_r := \{w \in \mathcal{W} : m(w) = r\}$ to be the set of words of length $2r$, and such that each word $w \in \mathcal{W}_r$ satisfies that $\mathbf{t}_l\mathbf{s}_{l+1} \in \{\mathbf{i}\boldsymbol{\mu}, \boldsymbol{\mu}\mathbf{i}\}$ for all $1 \leqslant l \leqslant r$.*

Next we assign to each letter $*$ a value $[*]$ through $[\mathbf{i}] := \Sigma\, \mathbf{u}_i$, $[\boldsymbol{\mu}] := \widetilde{\Sigma}\mathbf{v}_{\mu}$, where $\mathbf{u}_i$ and $\mathbf{v}_{\mu}$ are defined in Lemma 2.7.1 and are regarded as summation indices. Note that it is important to distinguish the abstract letter from its value, which is a summation index. Finally, to each word w we assign a random variable $A_{\mathbf{v},i,\mu}(w)$ as follows. If $m(w) = 0$ we define

$$A_{\mathbf{v},i,\mu}(w) := G_{\mathbf{vv}} - \Pi_{\mathbf{vv}}.$$

If $m(w) \geqslant 1$, *say* $w = \mathbf{t}_1\mathbf{s}_2\mathbf{t}_2\mathbf{s}_3\cdots\mathbf{t}_r\mathbf{s}_{r+1}$, *we define*

$$A_{\mathbf{v},i,\mu}(w) := G_{\mathbf{v}[\mathbf{t}_1]}G_{[\mathbf{s}_2][\mathbf{t}_2]}\cdots G_{[\mathbf{s}_r][\mathbf{t}_r]}G_{[\mathbf{s}_{r+1}]\,\mathbf{v}}. \tag{2.7.30}$$

Notice the words are constructed such that, by (2.7.21) and (2.7.22) ,

$$\left(\frac{\partial}{\partial X_{i\mu}}\right)^r (G_{\mathbf{vv}} - \Pi_{\mathbf{vv}}) = (-1)^r r! \sum_{w\in\mathcal{W}_r} A_{\mathbf{v},i,\mu}(w), \quad r\in\mathbb{N},$$

with which we get that

$$\left(\frac{\partial}{\partial X_{i\mu}}\right)^r F^p_{\mathbf{v}}(X) = (-1)^r \sum_{m_1+\cdots+m_p=r}\prod_{t=1}^{p/2}\left(m_t!m_{t+p/2}!\right)\left(\sum_{w_t\in\mathcal{W}_{m_t}}\sum_{w_{t+p/2}\in\mathcal{W}_{m_{t+p/2}}} A_{\mathbf{v},i,\mu}(w_t)\overline{A_{\mathbf{v},i,\mu}(w_{t+p/2})}\right).$$

Then to prove (2.7.28), it suffices to show that

$$N^{-2}q^{r-4}\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left|\mathbb{E}\prod_{t=1}^{p/2}A_{\mathbf{v},i,\mu}(w_t)\overline{A_{\mathbf{v},i,\mu}(w_{t+p/2})}\right| = \mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi)\right]^p + \mathbb{E}F^p_{\mathbf{v}}(X,z)\right) \tag{2.7.31}$$

for $4 \leqslant r \leqslant 4p+4$ and all words $w_1, ..., w_p \in \mathcal{W}$ satisfying $m(w_1)+\cdots+m(w_p) = r$. To avoid the unimportant notational complications associated with the complex conjugates, we will actually prove that

$$N^{-2}q^{r-4}\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left|\mathbb{E}\prod_{t=1}^{p}A_{\mathbf{v},i,\mu}(w_t)\right| = \mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi)\right]^p + \mathbb{E}F^p_{\mathbf{v}}(X,z)\right). \tag{2.7.32}$$

The proof of (2.7.31) is essentially the same but with slightly heavier notations. Treating empty words separately, we find it suffices to prove

$$N^{-2}q^{r-4}\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\mathbb{E}\left|A^{p-l}_{\mathbf{v},i,\mu}(w_0)\prod_{t=1}^{l}A_{\mathbf{v},i,\mu}(w_t)\right| = \mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi)\right]^p + \mathbb{E}F^p_{\mathbf{v}}(X,z)\right) \tag{2.7.33}$$

for $4 \leqslant r \leqslant 4p+4$, $1 \leqslant l \leqslant p$, and words such that $m(w_0) = 0$, $\sum_t m(w_t) = r$ and $m(w_t) \geqslant 1$ for $t \geqslant 1$.

To estimate (2.7.33) we introduce the quantity

$$\mathcal{R}_a := |G_{\mathbf{v}\mathbf{w}_a}| + |G_{\mathbf{w}_a\mathbf{v}}| \tag{2.7.34}$$

for $a \in \mathcal{I}$, where $\mathbf{w}_i := \Sigma^{1/2}\,\mathbf{u}_i$ for $i \in \mathcal{I}_1$ and $\mathbf{w}_\mu := \widetilde{\Sigma}^{1/2}\,\mathbf{v}_\mu$ for $\mu \in \mathcal{I}_2$.

Lemma 2.7.16. *For $w \in \mathcal{W}$, we have the rough bound*

$$|A_{\mathbf{v},i,\mu}(w)| \prec N^{2\delta(m(w)+1)}. \tag{2.7.35}$$

Furthermore, for $m(w) \geqslant 1$ we have

$$|A_{\mathbf{v},i,\mu}(w)| \prec (\mathcal{R}_i^2 + \mathcal{R}_\mu^2) N^{2\delta(m(w)-1)}. \tag{2.7.36}$$

For $m(w) = 1$, we have the better bound

$$|A_{\mathbf{v},i,\mu}(w)| \prec \mathcal{R}_i \mathcal{R}_\mu. \tag{2.7.37}$$

Proof. The estimates (2.7.35) and (2.7.36) follow immediately from the rough bound (2.7.14) and definition (2.7.30). The estimate (2.7.37) follows from the constraint $\mathbf{t}_1 \neq \mathbf{s}_2$ in the definition (2.7.30). □

By pigeonhole principle, if $r \leqslant 2l-2$, then there exist at least two words w_t with $m(w_t) = 1$. Therefore by Lemma 2.7.16 we have

$$\left| A_{\mathbf{v},i,\mu}^{p-l}(w_0) \prod_{t=1}^{l} A_{\mathbf{v},i,\mu}(w_t) \right| \prec N^{2\delta(r+l)} F_{\mathbf{v}}^{p-l}(X) \left(\mathbf{1}(r \geqslant 2l-1)(\mathcal{R}_i^2 + \mathcal{R}_\mu^2) + \mathbf{1}(r \leqslant 2l-2)\mathcal{R}_i^2 \mathcal{R}_\mu^2 \right). \tag{2.7.38}$$

Let $\mathbf{v} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix}$ for $\mathbf{v}_1 \in \mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{v}_2 \in \mathbb{C}^{\mathcal{I}_2}$. Then using Lemma 2.7.1, we get

$$\begin{aligned} \frac{1}{N} \sum_{i \in \mathcal{I}_1} \mathcal{R}_i^2 + \frac{1}{N} \sum_{\mu \in \mathcal{I}_2} \mathcal{R}_\mu^2 &\prec \frac{\operatorname{Im}(z^{-1} G_{\mathbf{v}_1 \mathbf{v}_1}) + \operatorname{Im}(G_{\mathbf{v}_2 \mathbf{v}_2}) + \eta |G_{\mathbf{v}_1 \mathbf{v}_1}| + \eta |G_{\mathbf{v}_2 \mathbf{v}_2}|}{N\eta} \\ &\prec N^{2\delta} \frac{\operatorname{Im} m_{2c} + N^{C_a \delta}(q + \Psi(z))}{N\eta} \prec N^{(C_a+2)\delta} \left(\Psi^2(z) + \frac{q}{N\eta} \right), \end{aligned} \tag{2.7.39}$$

where in the second step we used the two bounds in Lemma 2.7.7 and $\eta = \mathrm{O}(\operatorname{Im} m_{2c})$ by (2.4.11), and in the last step the definition of Ψ in (2.4.18). Using the same method we can get

$$\frac{1}{N^2} \sum_{i \in \mathcal{I}_1} \sum_{\mu \in \mathcal{I}_2} \mathcal{R}_i^2 \mathcal{R}_\mu^2 \prec \left[N^{(C_a+2)\delta} \left(\Psi^2(z) + \frac{q}{N\eta} \right) \right]^2. \tag{2.7.40}$$

Plugging (2.7.39) and (2.7.40) into (2.7.38), we get that the left-hand side of (2.7.33) is bounded by

$$
\begin{aligned}
&q^{r-4}N^{2\delta(r+l+2)}\,\mathbb{E}\,F_{\mathbf{v}}^{p-l}(X)\Big[\mathbf{1}(r\geqslant 2l-1)\left(N^{C_a\delta/2}(q+\Psi)\right)^2+\mathbf{1}(r\leqslant 2l-2)\left(N^{C_a\delta/2}(q+\Psi)\right)^4\Big]\\
&\leqslant N^{2\delta(r+l+2)}\,\mathbb{E}\,F_{\mathbf{v}}^{p-l}(X)\Big[\mathbf{1}(r\geqslant 2l-1)\left(N^{C_a\delta/2}(q+\Psi)\right)^{r-2}+\mathbf{1}(r\leqslant 2l-2)\left(N^{C_a\delta/2}(q+\Psi)\right)^r\Big]\\
&\leqslant \mathbb{E}\,F_{\mathbf{v}}^{p-l}(X)\Big[\mathbf{1}(r\geqslant 2l-1)\left(N^{C_a\delta/2+12\delta}(q+\Psi)\right)^{r-2}+\mathbf{1}(r\leqslant 2l-2)\left(N^{C_a\delta/2+12\delta}(q+\Psi)\right)^r\Big],
\end{aligned}
$$

where we used that $l\leqslant r$ and $r\geqslant 4$ in the last step. If we choose $C_a\geqslant 25$, then by (2.7.8) we have $N^{C_a\delta/2+12\delta}\ll\min\{N^{\phi/2},N^{\varepsilon/2}\}$, and hence $N^{C_a\delta/2+12\delta}(q+\Psi)\ll 1$. Moreover, if $r\geqslant 4$ and $r\geqslant 2l-1$, then $r\geqslant l+2$. Therefore we conclude that the left-hand side of (2.7.33) is bounded by

$$
\mathbb{E}\,F_{\mathbf{v}}^{p-l}(X)\left[N^{C_a\delta}(q+\Psi)\right]^l. \tag{2.7.41}
$$

Now (2.7.33) follows from Hölder's inequality. This concludes the proof of (2.7.28), and hence of (2.7.20), and hence of Lemma 2.7.4. This proves (2.7.7), and hence (2.4.22) under the condition (2.7.16).

2.7.3 Non-vanishing third moment

In this subsection, we prove Lemma 2.7.5 under (2.4.20) for $z\in\widetilde{S}(c_0,C_0,\varepsilon)$. Following the arguments in Sections 2.7.1-2.7.2, we see that it suffices to prove the estimate (2.7.28) in the $r=3$ case. In other words, we need to prove the following lemma.

Lemma 2.7.17. *Fix $p\in 2\mathbb{N}$ and $m\leqslant\delta^{-1}$. Let $z\in\mathbf{S}_m$ and suppose $(\mathbf{A}_{m-1})$ holds. Then*

$$
b_N N^{-2}\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left|\mathbb{E}\,f_{(i\mu)}^{(3)}(X_{i\mu}^\theta)\right| = \mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi)\right]^p+\mathbb{E}F_{\mathbf{v}}^p(X^\theta,z)\right). \tag{2.7.42}
$$

Proof. The main new ingredient of the proof is a further iteration step at a fixed z. Suppose

$$
G-\widetilde{\Pi}=\mathrm{O}_\prec(\Phi) \tag{2.7.43}
$$

for some deterministic parameter $\Phi\equiv\Phi_N$. By the a priori bound (2.7.14), we can take $\Phi\leqslant N^{2\delta}$. Assuming (2.7.43), we shall prove a self-improving bound of the form

$$
b_N N^{-2}\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left|\mathbb{E}\,f_{(i\mu)}^{(3)}(X_{i\mu}^\theta)\right| = \mathrm{O}\left(\left[N^{C_a\delta}(q+\Psi)\right]^p+(N^{-\varepsilon/2}\Phi)^p+\mathbb{E}F_{\mathbf{v}}^p(X^\theta,w)\right). \tag{2.7.44}
$$

Once (2.7.44) is proved, we can use it iteratively to get an increasingly accurate bound for $|G_{\mathbf{vv}}(X,z) - \Pi_{\mathbf{vv}}(z)|$. After each step, we obtain a better bound (2.7.43) with Φ reduced by $N^{-\varepsilon/2}$. Hence after $\mathrm{O}(\varepsilon^{-1})$ many iterations we can get (2.7.42).

As in Section 2.7.2, to prove (2.7.44) it suffices to show

$$b_N N^{-2}\left|\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2} A^{p-l}_{\mathbf{v},i,\mu}(w_0)\prod_{t=1}^{l} A_{\mathbf{v},i,\mu}(w_t)\right| \prec F^{p-l}_{\mathbf{v}}(X)\left[N^{(C_0-1)\delta}(q+\Psi)+N^{-\varepsilon/2}\Phi\right]^l, \quad (2.7.45)$$

which follows from the bound

$$b_N N^{-2}\left|\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\prod_{t=1}^{l} A_{\mathbf{v},i,\mu}(w_t)\right| \prec \left[N^{(C_0-1)\delta}(q+\Psi)+N^{-\varepsilon/2}\Phi\right]^l. \quad (2.7.46)$$

We now list all the three cases with $l = 1, 2, 3$, and discuss each case separately.

When $l = 1$, the single factor $A_{\mathbf{v},i,\mu}(w_1)$ is of the form

$$G_{\mathbf{v}[\mathbf{t}_1]}G_{[\mathbf{s}_2][\mathbf{t}_2]}G_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}}.$$

Then we split it as

$$\begin{aligned} G_{\mathbf{v}[\mathbf{t}_1]}G_{[\mathbf{s}_2][\mathbf{t}_2]}G_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}} =& G_{\mathbf{v}[\mathbf{t}_1]}\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}} + G_{\mathbf{v}[\mathbf{t}_1]}\widetilde{G}_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}} \\ &+ G_{\mathbf{v}[\mathbf{t}_1]}\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}\widetilde{G}_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}} + G_{\mathbf{v}[\mathbf{t}_1]}\widetilde{G}_{[\mathbf{s}_2][\mathbf{t}_2]}\widetilde{G}_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}}, \end{aligned} \quad (2.7.47)$$

where we abbreviate $\widetilde{G} := G - \Pi$. For the second term, we have

$$b_N N^{-2}\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left|G_{\mathbf{v}[\mathbf{t}_1]}\widetilde{G}_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}}\right| \prec b_N\Phi\cdot N^{(C_a+2)\delta}\left(\Psi^2+\frac{q}{N\eta}\right) \prec N^{-\varepsilon/2}\Phi \quad (2.7.48)$$

provided δ is small enough, where we used (2.7.39), (2.7.43) and the definition (2.4.21). The third and fourth term of (2.7.47) can be dealt with in a similar way. For the first term, when $[\mathbf{t}_1] = \mathbf{w}_i$ and $[\mathbf{s}_4] = \mathbf{w}_\mu$, we have

$$\left|\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2} G_{\mathbf{v}\mathbf{w}_i}\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}G_{\mathbf{w}_\mu\mathbf{v}}\right| \prec N^{1+2\delta}\left(\sum_{\mu\in\mathcal{I}_2}|G_{\mathbf{w}_\mu\mathbf{v}}|^2\right)^{1/2} \prec N^{3/2+(C_a/2+3)\delta}(q+\Psi),$$

where we used (2.7.39) and the fact that Π is deterministic, such that the a priori bound (2.7.23) gives

$$\Big|\sum_{i\in\mathcal{I}_1} G_{\mathbf{v}\mathbf{w}_i}\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}\Big| \prec N^{1/2+2\delta}.$$

If $[\mathbf{t}_1]=\mathbf{w}_\mu$ and $[\mathbf{s}_4]=\mathbf{v}_i$, the proof is similar. If $[\mathbf{t}_1]=[\mathbf{s}_4]$, then at least one of the terms $\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}$ and $\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}$ must be of the form $\Pi_{\mathbf{w}_i\mathbf{w}_\mu}$ or $\Pi_{\mathbf{w}_\mu\mathbf{w}_i}$, and hence we have

$$\sum_i |\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}| = \mathrm{O}(N^{1/2}) \quad \text{ or } \quad \sum_\mu |\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}| = \mathrm{O}(N^{1/2}).$$

Therefore using (2.7.39) and (2.4.21), we get

$$\Big|\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2} G_{\mathbf{v}[\mathbf{t}_1]}\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}}\Big| \prec N^{3/2+(C_a+2)\delta}\left(q^2+\Psi^2\right) \leqslant N^{3/2}(q+\Psi).$$

provided δ is small enough. In sum, we obtain that

$$b_N N^{-2}\Big|\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2} G_{\mathbf{v}[\mathbf{t}_1]}\Pi_{[\mathbf{s}_2][\mathbf{t}_2]}\Pi_{[\mathbf{s}_3][\mathbf{t}_3]}G_{[\mathbf{s}_4]\mathbf{v}}\Big| \prec N^{(C_a-1)\delta}(q+\Psi)$$

provided that $C_a \geqslant 8$. Together with (2.7.48), this proves (2.7.46) for $l=1$.

When $l=2$, $\prod_{t=1}^2 A_{\mathbf{v},i,\mu}(w_t)$ is of the form

$$G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{v}}G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{v}}, \quad G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{v}}G_{\mathbf{v}\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{v}}, \tag{2.7.49}$$

$$G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{v}}G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{v}}, \quad G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{v}}G_{\mathbf{v}\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{v}}, \tag{2.7.50}$$

or an expression obtained from one of these four by exchanging $\mathbf{w}_i$ and $\mathbf{w}_\mu$. The first expression in (2.7.49) can be estimated using (2.7.23), (2.7.39) and (2.7.43):

$$\sum_\mu G_{\mathbf{w}_\mu\mathbf{v}}G_{\mathbf{w}_\mu\mathbf{w}_\mu} = \sum_\mu G_{\mathbf{w}_\mu\mathbf{v}}\widetilde{G}_{\mathbf{w}_\mu\mathbf{w}_\mu} + \sum_\mu G_{\mathbf{w}_\mu\mathbf{v}}\Pi_{\mathbf{w}_\mu\mathbf{w}_\mu} = \mathrm{O}_\prec\left[N^{1+(C_a/2+1)\delta}\Phi\left(\Psi^2+\frac{q}{N\eta}\right)^{1/2} + N^{1/2+2\delta}\right], \tag{2.7.51}$$

and

$$\Big|\sum_i G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_i\mathbf{v}}\Big| \prec N^{1+(C_a+4)\delta}\left(\Psi^2+\frac{q}{N\eta}\right). \tag{2.7.52}$$

Combining (2.4.21), (2.7.51) and (2.7.52), we get that

$$b_N N^{-2}\Big|\sum_i\sum_\mu G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{v}}G_{\mathbf{v}\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{v}}\Big| \prec \left(N^{(C_a-1)\delta}(q+\Psi) + N^{-\varepsilon/2}\Phi\right)^2,$$

provided δ is small enough. The second expression in (2.7.49) can be estimated similarly. The first expression of (2.7.50) can be estimated using (2.4.21), (2.7.23) and (2.7.39) by

$$b_N N^{-2} \left| \sum_i \sum_\mu G_{\mathbf{v}\mathbf{w}_i} G_{\mathbf{w}_\mu \mathbf{v}} G_{\mathbf{v}\mathbf{w}_i} G_{\mathbf{w}_\mu \mathbf{w}_i} G_{\mathbf{w}_\mu \mathbf{v}} \right| \prec b_N N^{-2+2\delta} \sum_i \sum_\mu |G_{\mathbf{v}\mathbf{w}_i}|^2 \left|G_{\mathbf{w}_\mu \mathbf{v}}\right|^2$$
$$\prec b_N N^{(2C_0+6)\delta} \left(\Psi^2 + \frac{q}{N\eta} \right)^2 \leqslant (q+\Psi)^2$$

for small enough δ. The second expression in (2.7.50) is estimated similarly. This proves (2.7.46) for $l = 2$.

When $l = 3$, $\prod_{t=1}^{3} A_{\mathbf{v},i,\mu}(w_t)$ is of the form $(G_{\mathbf{v}\mathbf{w}_i} G_{\mathbf{w}_\mu \mathbf{v}})^3$ or an expression obtained by exchanging $\mathbf{w}_i$ and $\mathbf{w}_\mu$ in some of the three factors. We use (2.7.39) and $\sum_i |\Pi_{\mathbf{v}\mathbf{w}_i}|^2 = \mathrm{O}(1)$ to get that

$$\left| \sum_i (G_{\mathbf{v}\mathbf{w}_i})^3 \right| \prec \sum_i |\widetilde{G}_{\mathbf{v}\mathbf{w}_i}|^3 + \sum_i |\Pi_{\mathbf{v}\mathbf{w}_i}|^3 \prec \Phi \sum_i \left(|G_{\mathbf{v}\mathbf{w}_i}|^2 + |\Pi_{\mathbf{v}\mathbf{w}_i}|^2 \right) + 1 \prec N^{1+(C_0+2)\delta} \left(\Psi^2 + \frac{q}{N\eta} \right) \Phi + \Phi + 1.$$

Now we conclude (2.7.46) for $l = 3$ using (2.4.21) and $N^{-1/2} = \mathrm{O}(q + \Psi)$. □

If A or B is diagonal, then we can still prove (2.4.22) for all $z \in S(c_0, C_0, \varepsilon)$ without using (2.7.16). This follows from an improved self-consistent comparison argument for sample covariance matrices (i.e. separable covariance matrices with $B = I$) in [59, Section 8]. The argument for separable covariance matrices with diagonal A or B is almost the same except for some notational differences, so we omit the details.

2.7.4 Weak averaged local law

In this section, we prove the weak averaged local laws in (2.4.23) and (2.4.24). The proof is similar to that for (2.4.22) in previous subsections, and we only explain the differences. Note that the bootstrapping argument is not necessary, since we already have a good a priori bound by (2.4.22). In analogy to (2.7.12), we define

$$\widetilde{F}(X, z) := |m(z) - m_c(z)| = \left| \frac{1}{nz} \sum_{i \in \mathcal{I}_1} \left(G_{ii}(X, z) - \Pi_{ii}(z) \right) \right|,$$

where we used (2.4.17). Moreover, by Proposition 2.6.1, (2.4.23) and (2.4.24) hold for Gaussian X (without the q^2 term). For now, we assume (2.7.16) and prove the following stronger estimates:

$$|m(z) - m_c(z)| \prec (N\eta)^{-1} \tag{2.7.53}$$

for $z \in S(c_0, C_0, \varepsilon)$, and

$$|m(z) - m_c(z)| \prec \frac{q}{N\eta} + \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}}, \tag{2.7.54}$$

for $z \in S(c_0, C_0, \varepsilon) \cap \{z = E + \mathrm{i}\eta : E \geqslant \lambda_r, N\eta\sqrt{\kappa+\eta} \geqslant N^{\varepsilon}\}$. At the end of this section, we will show how to relax (2.7.16) to (2.4.20) for $z \in \widetilde{S}(c_0, C_0, \varepsilon)$.

Note that

$$\Psi^2(z) \lesssim \frac{1}{N\eta}, \quad \text{and} \quad \Psi^2(z) \lesssim \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}} \text{ outside of the spectrum.} \tag{2.7.55}$$

Then following the argument in Section 2.7.1, analogous to (2.7.28), we only need to prove that

$$N^{-2}q^{r-4}\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left|\mathbb{E}\left(\frac{\partial}{\partial X_{i\mu}}\right)^r \widetilde{F}^p(X)\right| = \mathrm{O}\left(\left[N^{\delta}\left(\Psi^2 + \frac{q}{N\eta}\right)\right]^p + \mathbb{E}\widetilde{F}^p(X)\right) \tag{2.7.56}$$

for all $r = 4, ..., 4p+4$, where $\delta > 0$ is any positive constant. Analogous to (2.7.32), it suffices to prove that for $r = 4, ..., 4p+4$,

$$N^{-2}q^{r-4}\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\left|\mathbb{E}\prod_{t=1}^{p}\left(\frac{1}{n}\sum_{j\in\mathcal{I}_1}A_{\mathbf{e}_j,i,\mu}(w_t)\right)\right| = \mathrm{O}\left(\left[N^{\delta}\left(\Psi^2 + \frac{q}{N\eta}\right)\right]^p + \mathbb{E}\widetilde{F}^p(X)\right) \tag{2.7.57}$$

for $\sum_t m(w_t) = r$. Similar to (2.7.34) we define

$$\mathcal{R}_{j,a} := |G_{j\mathbf{w}_a}| + |G_{\mathbf{w}_a j}|.$$

Using (2.4.22) and Lemma 2.7.1, similarly to (2.7.39), we get that

$$\frac{1}{n}\sum_{j\in\mathcal{I}_1}\mathcal{R}_{j,a}^2 \prec \frac{\operatorname{Im}(z^{-1}G_{\mathbf{w}_i\mathbf{w}_i}) + \operatorname{Im}G_{\mathbf{w}_\mu\mathbf{w}_\mu} + \eta\left(|G_{\mathbf{w}_i\mathbf{w}_i}| + |G_{\mathbf{w}_\mu\mathbf{w}_\mu}|\right)}{N\eta} \prec \Psi^2 + \frac{q}{N\eta}. \tag{2.7.58}$$

Since $G = \mathrm{O}_\prec(1)$ by (2.4.22), we have

$$\left|\frac{1}{n}\sum_{j\in\mathcal{I}_1}A_{\mathbf{e}_j,i,\mu}(w)\right| \prec \frac{1}{n}\sum_{j\in\mathcal{I}_1}\left(\mathcal{R}_{j,i}^2 + \mathcal{R}_{j,\mu}^2\right) \prec \Psi^2 + \frac{q}{N\eta} \quad \text{for any } w \text{ such that } m(w) \geqslant 1. \tag{2.7.59}$$

With (2.7.59), for any $r \geqslant 4$, the left-hand side of (2.7.57) is bounded by

$$\mathbb{E}\,\widetilde{F}^{p-l}(X)\left(\Psi^2 + \frac{q}{N\eta}\right)^l.$$

Applying Holder's inequality, we get (2.7.56), which completes the proof of (2.7.53) and (2.7.54) under (2.7.16).

Then we prove the averaged local law for $z \in \widetilde{S}(c_0, C_0, \varepsilon)$ under (2.4.20). By (2.7.55), it suffices to prove

$$b_N N^{-2}\left|\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\mathbb{E}\left(\frac{\partial}{\partial X_{i\mu}}\right)^3 \widetilde{F}^p(X)\right| = \mathrm{O}\left(\left[N^\delta(q^2+\Psi^2)\right]^p + \left(\frac{N^{-\varepsilon/2}}{N\eta}\right)^p + \mathbb{E}\widetilde{F}^p(X)\right), \tag{2.7.60}$$

for any constant $\delta > 0$. Analogous to the arguments in Section 2.7.3, it reduces to showing that

$$b_N N^{-2}\left|\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\prod_{t=1}^{l}\left(\frac{1}{n}\sum_{j\in\mathcal{I}_1}A_{\mathbf{e}_j,i,\mu}(w_t)\right)\right| = \mathrm{O}_\prec\left(\left(q^2+\Psi^2\right)^l + \left(\frac{N^{-\varepsilon/2}}{N\eta}\right)^l\right), \tag{2.7.61}$$

where $l \in \{1,2,3\}$ is the number of words with nonzero length. Then we can discuss these three cases using a similar argument as in Section 2.7.3, with the only difference being that we now can use the anisotropic local law (2.4.22) instead of the a priori bounds (2.7.23) and (2.7.43).

In the $l = 1$ case, we first consider the expression $A_{\mathbf{e}_j,i,\mu}(w_1) = G_{j\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{w}_i}G_{\mathbf{w}_\mu j}$. We have

$$\left|\sum_i G_{j\mathbf{w}_i}G_{\mathbf{w}_i\mathbf{w}_i}\right| \leqslant \left|\sum_i G_{j\mathbf{w}_i}\Pi_{\mathbf{w}_i\mathbf{w}_i}\right| + \sum_i (q+\Psi)\left|G_{j\mathbf{w}_i}\right| \prec \sqrt{N} + N(q+\Psi)\left(\Psi^2 + \frac{q}{N\eta}\right)^{1/2},$$

where we used (2.4.22) and (2.7.39). Similarly, we also have

$$\left|\sum_\mu G_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_\mu j}\right| \prec \left|\sum_\mu \Pi_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_\mu j}\right| + \left|\sum_\mu \widetilde{G}_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_\mu j}\right| \prec \sqrt{N}(q+\Psi) + N(q+\Psi)\left(\Psi^2 + \frac{q}{N\eta}\right)^{1/2},$$

where we also used $\Pi_{\mathbf{w}_\mu j} = 0$ for any μ in the second step. Then with (2.4.21), we can see that the LHS of (2.7.61) is bounded by $\mathrm{O}_\prec(q^2+\Psi^2)$ in this case. For the case $A_{\mathbf{e}_j,i,\mu}(w_1) = G_{j\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{w}_\mu}G_{\mathbf{w}_i j}$, we can estimate that

$$\left|\sum_\mu G_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{w}_\mu}\right| \leqslant \left|\sum_\mu \Pi_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_i\mathbf{w}_\mu}\right| + \sum_\mu (q+\Psi)\left|G_{\mathbf{w}_i\mathbf{w}_\mu}\right| \prec \sqrt{N} + N(q+\Psi)\left(\Psi^2 + \frac{q}{N\eta}\right)^{1/2},$$

and

$$\sum_i |G_{j\mathbf{w}_i} G_{\mathbf{w}_i j}| \prec N\left(\Psi^2 + \frac{q}{N\eta}\right).$$

Thus in this case the LHS of (2.7.61) is also bounded by $\mathrm{O}_\prec(q^2+\Psi^2)$. The case $A_{\mathbf{e}_j,i,\mu}(w_1) = G_{j\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_i j}$ can be handled similarly. Finally in the case $A_{\mathbf{e}_j,i,\mu}(w_1) = G_{j\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu j}$, we can estimate that

$$\left|\sum_{i,\mu} G_{j\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu j}\right| \prec \sum_{i,\mu}\left(|G_{j\mathbf{w}_i}|^2 + |G_{\mathbf{w}_\mu j}|^2\right)|G_{\mathbf{w}_\mu\mathbf{w}_i}|^2 \prec N^2\left(\Psi^2 + \frac{q}{N\eta}\right)^2.$$

Again in this case the LHS of (2.7.61) is bounded by $\mathrm{O}_\prec(q^2+\Psi^2)$. All the other expressions are obtained from these four by exchanging $\mathbf{w}_i$ and $\mathbf{w}_\mu$.

In the $l=2$ case, $\prod_{t=1}^2\left(\frac{1}{n}\sum_{j\in\mathcal{I}_1} A_{\mathbf{e}_j,i,\mu}(w_t)\right)$ is of the forms

$$\frac{1}{N^2}\sum_{j_1,j_2} G_{j_1\mathbf{w}_i}G_{\mathbf{w}_\mu j_1}G_{j_2\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_\mu}G_{\mathbf{w}_i j_2} \quad \text{or} \quad \frac{1}{N^2}\sum_{j_1,j_2} G_{j_1\mathbf{w}_i}G_{\mathbf{w}_\mu j_1}G_{j_2\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu j_2},$$

or an expression obtained from one of these terms by exchanging $\mathbf{w}_i$ and $\mathbf{w}_\mu$. These two expressions can be written as

$$N^{-2}(G^{\times 2})_{\mathbf{w}_\mu\mathbf{w}_i}(G^{\times 2})_{\mathbf{w}_i\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_\mu}, \quad N^{-2}(G^{\times 2})^2_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}, \quad G^{\times 2} := G\begin{pmatrix} I_{\mathcal{I}_1\times\mathcal{I}_1} & 0 \\ 0 & 0\end{pmatrix}G. \tag{2.7.62}$$

For the second term, using (2.4.4), (2.4.5) and recalling that $Y = \Sigma^{1/2}U^*XV\widetilde{\Sigma}^{1/2}$, we can get that

$$\begin{aligned}
&\left|\frac{1}{N^2}\sum_{i,\mu}(G^{\times 2})^2_{\mathbf{w}_\mu\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_i}\right| \leqslant \frac{1}{N^2}\sum_{i,\mu}\left|(G^{\times 2})_{\mathbf{w}_\mu\mathbf{w}_i}\right|^2 = \frac{|z|^2}{N^2}\mathrm{Tr}\left[(\mathcal{G}_1^*)^2 YY^*(\mathcal{G}_1)^2\right] \\
&= \frac{|z|^2}{N^2}\mathrm{Tr}\left[\mathcal{G}_1^*(\mathcal{G}_1)^2\right] + \frac{\bar{z}|z|^2}{N^2}\mathrm{Tr}\left[(\mathcal{G}_1^*)^2(\mathcal{G}_1)^2\right] \qquad (2.7.63)\\
&\lesssim \frac{1}{N^2}\sum_k \frac{1}{[(\lambda_k-E)^2+\eta^2]^{3/2}} + \frac{1}{N^2}\sum_k\frac{1}{[(\lambda_k-E)^2+\eta^2]^2} \\
&\lesssim \frac{1}{N\eta^3}\left(\frac{1}{n}\sum_k\frac{\eta}{(\lambda_k-E)^2+\eta^2}\right) = \frac{\mathrm{Im}\, m}{N\eta^3} \prec \frac{\mathrm{Im}\, m_c + q + \Psi}{N\eta^3} \lesssim \eta^{-2}\left(\Psi^2 + \frac{q}{N\eta}\right). \qquad (2.7.64)
\end{aligned}$$

Using (2.4.22) and (2.7.39), it is easy to show that

$$\left|\sum_\mu (G^{\times 2})_{\mathbf{w}_\mu\mathbf{w}_i}\Pi_{\mathbf{w}_\mu\mathbf{w}_\mu}\right| \prec N^{3/2}\left(\Psi^2 + \frac{q}{N\eta}\right), \quad \text{and} \quad \left|(G^{\times 2})_{\mathbf{x}\mathbf{y}}\right| \prec N\left(\Psi^2+\frac{q}{N\eta}\right), \tag{2.7.65}$$

for any deterministic unit vectors $\mathbf{x}$, $\mathbf{y}$. Thus for the first term in (2.7.62), we have

$$\begin{aligned}
&\left|\frac{1}{N^2}\sum_{i,\mu}(G^{\times 2})_{\mathbf{w}_\mu\mathbf{w}_i}(G^{\times 2})_{\mathbf{w}_i\mathbf{w}_i}G_{\mathbf{w}_\mu\mathbf{w}_\mu}\right| \qquad (2.7.66)\\
&\leqslant \left|\frac{1}{N^2}\sum_{i,\mu}(G^{\times 2})_{\mathbf{w}_\mu\mathbf{w}_i}(G^{\times 2})_{\mathbf{w}_i\mathbf{w}_i}\widetilde{G}_{\mathbf{w}_\mu\mathbf{w}_\mu}\right| + \left|\frac{1}{N^2}\sum_{i,\mu}(G^{\times 2})_{\mathbf{w}_\mu\mathbf{w}_i}(G^{\times 2})_{\mathbf{w}_i\mathbf{w}_i}\Pi_{\mathbf{w}_\mu\mathbf{w}_\mu}\right|\\
&\prec N(q+\Psi)\left(\Psi^2+\frac{q}{N\eta}\right)\left(\frac{1}{N^2}\sum_{i,\mu}\left|(G^{\times 2})_{\mathbf{w}_\mu\mathbf{w}_i}\right|^2\right)^{1/2} + N^{3/2}\left(\Psi^2+\frac{q}{N\eta}\right)^2\\
&\prec N\eta^{-1}(q+\Psi)\left(\Psi^2+\frac{q}{N\eta}\right)^{3/2} + N^{3/2}\left(\Psi^2+\frac{q}{N\eta}\right)^2, \qquad (2.7.67)
\end{aligned}$$

where in the last step we used the bound in (2.7.64). Now using (2.7.64), (2.7.67) and (2.4.21), we get

$$b_N N^{-2}\left|\sum_{i\in\mathcal{I}_1}\sum_{\mu\in\mathcal{I}_2}\prod_{t=1}^{2}\left(\frac{1}{n}\sum_{j\in\mathcal{I}_1}A_{\mathbf{e}_j,i,\mu}(w_t)\right)\right| \prec \left(q^2+\Psi^2\right)^2 + \left(\frac{N^{-\varepsilon/2}}{N\eta}\right)^2.$$

Finally, in the $l=3$ case, $\prod_{t=1}^3\left(\frac{1}{N}\sum_{j\in\mathcal{I}_1}A_{\mathbf{e}_j,i,\mu}(w_t)\right)$ is of the form $N^{-3}(G^{\times 2})^3_{\mathbf{w}_i\mathbf{w}_\mu}$, or an expression obtained by exchanging $\mathbf{w}_i$ and $\mathbf{w}_\mu$ in some of the three factors. Using (2.7.65) and the bound in (2.7.64), we can estimate that

$$\frac{1}{N^3}\left|\sum_{i,\mu}(G^{\times 2})^3_{\mathbf{w}_i\mathbf{w}_\mu}\right| \prec \left(\Psi^2+\frac{q}{N\eta}\right)\frac{1}{N^2}\sum_{i,\mu}\left|(G^{\times 2})_{\mathbf{w}_\mu\mathbf{w}_i}\right|^2 \prec \eta^{-2}\left(\Psi^2+\frac{q}{N\eta}\right)^2,$$

Then the LHS of (2.7.61) is bounded by

$$O_\prec\left(\left(q^2+\Psi^2\right)\left(\frac{N^{-\varepsilon/2}}{N\eta}\right)^2\right).$$

Combining the above three cases, we conclude (2.7.60), which finishes the proof of (2.4.23) and (2.4.24).

If A or B is diagonal, then by the remark at the end of Section 2.7.3, the anisotropic local law (2.4.22) holds for all $z\in S(c_0,C_0,\varepsilon)$ even in the case with $b_N=N^{1/2}$ in (2.4.20). Then with (2.4.22) and the self-consistent comparison argument in [59, Section 9], we can prove (2.4.23) and (2.4.24) for $z\in S(c_0,C_0,\varepsilon)$. Again most of the arguments are the same as the ones in [59, Section 9], hence we omit the details.

2.8 Lindeberg replacement strategy

2.8.1 Proof of Theorem 2.4.8 and Lemma 2.4.7

With Lemma 2.4.12, given X satisfying the assumptions in Theorem 2.4.6, we can construct a matrix $\widetilde{X}$ with support $q = N^{-1/2}$ and have the same first four moments as X. By Theorem 2.4.6, the averaged local laws (2.4.26) and (2.4.27) hold for $G(\widetilde{X}, z)$. Thus it is easy to see that Theorem 2.4.8 is implied by the following lemma.

Lemma 2.8.1. *Let X, $\widetilde{X}$ be two matrices as in Lemma 2.4.12, and $G \equiv G(X,z)$, $\widetilde{G} \equiv G(\widetilde{X}, z)$ be the corresponding resolvents. We denote $m(z) \equiv m(X,z)$ and $\widetilde{m}(z) \equiv m(\widetilde{X}, z)$. Fix any constant $\varepsilon > 0$. For any $z \in \widetilde{S}(c_0, C_0, \varepsilon)$, if there exist deterministic quantities $J \equiv J(N)$ and $K \equiv K(N)$ such that*

$$\widetilde{G}(z) - \Pi = \mathrm{O}_{\prec}(J), \quad |\widetilde{m}(z) - m_c(z)| \prec K, \quad J + K \prec 1, \tag{2.8.1}$$

then for any fixed $p \in 2\mathbb{N}$, we have

$$\mathbb{E}|m(z) - m_c(z)|^p \prec \mathbb{E}|\widetilde{m}(z) - m_c(z)|^p + \left(\Psi^2(z) + J^2 + K\right)^p. \tag{2.8.2}$$

Proof of Theorem 2.4.8. By Theorem 2.4.6, one can choose $J = \Psi(z)$ and

$$K = \frac{1}{N\eta}, \quad \text{or} \quad \frac{1}{N(\kappa+\eta)} + \frac{1}{(N\eta)^2\sqrt{\kappa+\eta}} \text{ outside of the spectrum.}$$

Then using (2.8.2), (2.7.55) and Markov's inequality, we can prove (2.4.26) and (2.4.27). The eigenvalues rigidity results (2.4.28) and (2.4.30) follow from (2.4.26) and (2.4.27) through a standard argument using Helffer-Sjöstrand calculus, see e.g. the proofs for [34, Theorems 2.12-2.13], [38, Theorem 2.2] or [79, Theorem 3.3]. We omit the details. □

In order to prove Lemma 2.4.7 and Lemma 2.8.1, we will extend the resolvent comparison method developed in [62]. The basic idea is still to use the Lindeberg replacement strategy for $G(X, z)$. On the other hand, the main difference is that the resolvent estimates are only obtained from the entrywise local law in [62], while in our case we need to use the more general anisotropic local law (2.4.22). (We will use the anisotropic local law in (2.8.1) when proving Lemma 2.8.1. However, for simplicity of presentation, we will always mention (2.4.22) instead.)

Let $X = (x_{i\mu})$ and $\widetilde{X} = (\widetilde{x}_{i\mu})$ be two matrices as in Lemma 2.4.12. Define a bijective ordering map Φ on the index set of X as

$$\Phi : \{(i,\mu) : 1 \leqslant i \leqslant n,\ n+1 \leqslant \mu \leqslant n+N\} \to \{1, \ldots, \gamma_{\max} = nN\}.$$

For any $1 \leqslant \gamma \leqslant \gamma_{\max}$, we define the matrix $X^\gamma = (x^\gamma_{i\mu})$ such that $x^\gamma_{i\mu} = x_{i\mu}$ if $\Phi(i,\mu) \leqslant \gamma$, and $x^\gamma_{i\mu} = \widetilde{x}_{i\mu}$ otherwise. Note that we have $X^0 = \widetilde{X}$, $X^{\gamma_{\max}} = X$, and X^γ satisfies the bounded support condition with $q = N^{-\phi}$ for all $0 \leqslant \gamma \leqslant \gamma_{\max}$. Correspondingly, we define

$$H^\gamma := \begin{pmatrix} 0 & Y^\gamma \\ (Y^\gamma)^* & 0 \end{pmatrix}, \quad G^\gamma := \begin{pmatrix} -I_{n\times n} & Y^\gamma \\ (Y^\gamma)^* & -zI_{N\times N} \end{pmatrix}^{-1}, \tag{2.8.3}$$

where $Y^\gamma := \Sigma^{1/2} U^* X^\gamma V \widetilde{\Sigma}^{1/2}$. Then we define the $(n+N)\times(n+N)$ matrices V^γ and W^γ by (recall (2.7.21))

$$V^\gamma = \Delta^{x_{i\mu}}_{(i\mu)}, \quad W^\gamma := \Delta^{\widetilde{x}_{i\mu}}_{(i\mu)},$$

so that H^γ and $H^{\gamma-1}$ can be written as

$$H^\gamma = Q^\gamma + V^\gamma, \quad H^{\gamma-1} = Q^\gamma + W^\gamma, \tag{2.8.4}$$

for some matrix Q^γ that is independent of $x_{i\mu}$ and $\widetilde{x}_{i\mu}$. For simplicity of notations, for any γ we denote

$$S^\gamma := G^\gamma, \quad T^\gamma := G^{\gamma-1}, \quad R^\gamma := \left(Q^\gamma - \begin{pmatrix} I_{n\times n} & 0 \\ 0 & zI_{N\times N} \end{pmatrix} \right)^{-1}. \tag{2.8.5}$$

For convenience, we sometimes drop the superscript from R, S, T if γ is fixed. Under the above definitions, we can write

$$S = \left(Q^\gamma - \begin{pmatrix} I_{M\times M} & 0 \\ 0 & zI_{N\times N} \end{pmatrix} + V^\gamma \right)^{-1} = (I + RV^\gamma)^{-1} R. \tag{2.8.6}$$

Thus we can expand S using the resolvent expansion till order m:

$$S = R - RV^\gamma R + (RV^\gamma)^2 R + \cdots + (-1)^m (RV^\gamma)^m R + (-1)^{m+1} (RV^\gamma)^{m+1} S. \tag{2.8.7}$$

On the other hand, we can also expand R in terms of S,

$$R = (I - SV^\gamma)^{-1}S = S + SV^\gamma S + (SV^\gamma)^2 S + \ldots + (SV^\gamma)^m S + (SV^\gamma)^{m+1}R. \quad (2.8.8)$$

We have similar expansions for T and R by replacing (V^γ, S) with (W^γ, T) in (2.8.7) and (2.8.8). By the bounded support condition, we have

$$\max_\gamma \|V^\gamma\| = \mathrm{O}(|x_{i\mu}|) \prec N^{-\phi}, \quad \max_\gamma \|W^\gamma\| = \mathrm{O}(|\widetilde{x}_{i\mu}|) \prec N^{-1/2}. \quad (2.8.9)$$

Note that S, R and T satisfy the following deterministic bounds by (2.6.9):

$$\sup_{z\in\widetilde{S}(c_0,C_0,\varepsilon)} \max_\gamma \max\left\{\|S^\gamma\|, \|T^\gamma\|, \|R^\gamma\|\right\} \lesssim \sup_{z\in\widetilde{S}(c_0,C_0,\varepsilon)} \eta^{-1} \leqslant N. \quad (2.8.10)$$

Then using expansion (2.8.8) in terms of T, W^γ with $m = 3$, the isotropic local law (2.4.22) for T, and the bound (2.8.10) for R, we can get that for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$,

$$\sup_{z\in\widetilde{S}(c_0,C_0,\varepsilon)} \max_\gamma |R^\gamma_{\mathbf{uv}}| = \mathrm{O}(1) \quad \text{with high probability.} \quad (2.8.11)$$

From the definitions of V^γ and W^γ, one can see that it is helpful to introduce the following notations to simplify the expressions.

Definition 2.8.2 (Matrix operators $*_\gamma$)**.** *For any two $(n+N)\times(n+N)$ matrices A and B, we define*

$$A *_\gamma B := AI_\gamma B, \quad I_\gamma := \Delta^1_{(i\mu)}, \quad \Phi(i,\mu) = \gamma. \quad (2.8.12)$$

In other words, we have

$$A *_\gamma B = A\mathbf{w}_i\mathbf{w}_\mu^* B + A\mathbf{w}_\mu\mathbf{w}_i^* B, \quad \mathbf{w}_i := \Sigma^{1/2}\,\mathbf{u}_i, \quad \mathbf{w}_\mu := \widetilde{\Sigma}^{1/2}\,\mathbf{v}_\mu\,.$$

We denote the m-th power of A under the $*_\gamma$-product by $A^{*_\gamma m}$, i.e.

$$A^{*_\gamma m} := \underbrace{A *_\gamma A *_\gamma A *_\gamma \ldots *_\gamma A}_{m}. \quad (2.8.13)$$

Definition 2.8.3 ($\mathcal{P}_{\gamma,\mathbf{k}}$ and $\mathcal{P}_{\gamma,k}$)**.** *For $k \in \mathbb{N}$, $\mathbf{k} = (k_1, \cdots, k_s) \in \mathbb{N}^s$ and $1 \leqslant \gamma \leqslant \gamma_{\max}$, we define*

$$\mathcal{P}_{\gamma,k}G_{\mathbf{uv}} := G^{*_\gamma(k+1)}_{\mathbf{uv}}, \quad \mathcal{P}_{\gamma,\mathbf{k}}\left(\prod_{t=1}^p G_{\mathbf{u}_t\mathbf{v}_t}\right) := \prod_{t=1}^p \mathcal{P}_{\gamma,k_t}G_{\mathbf{u}_t\mathbf{v}_t}, \quad (2.8.14)$$

where we abbreviate $G^{*_\gamma(k+1)}_{\mathbf{uv}} \equiv (G^{*_\gamma(k+1)})_{\mathbf{uv}}$. *If* $\mathfrak{G}_1$ *and* $\mathfrak{G}_2$ *are products of resolvent entries as above, then we define*

$$\mathcal{P}_{\gamma,\mathbf{k}}(\mathfrak{G}_1 + \mathfrak{G}_2) := \mathcal{P}_{\gamma,\mathbf{k}}\mathfrak{G}_1 + \mathcal{P}_{\gamma,\mathbf{k}}\mathfrak{G}_2. \tag{2.8.15}$$

Note that $\mathcal{P}_{\gamma,k}$ *and* $\mathcal{P}_{\gamma,\mathbf{k}}$ *are not linear operators, but just notations we use for simplification. Similarly, for the product of the entries of* $G - \Pi$*, we define*

$$\mathcal{P}_{\gamma,\mathbf{k}}\left(\prod_{t=1}^{p}(G-\Pi)_{\mathbf{u}_t\mathbf{v}_t}\right) := \prod_{t=1}^{p}\mathcal{P}_{\gamma,k_t}(G-\Pi)_{\mathbf{u}_t\mathbf{v}_t}, \tag{2.8.16}$$

where

$$\mathcal{P}_{\gamma,k}(G-\Pi)_{\mathbf{uv}} := \begin{cases} (G-\Pi)_{\mathbf{uv}}, & \textit{if } k=0, \\ G^{*_\gamma(k+1)}_{\mathbf{uv}}, & \textit{otherwise.} \end{cases}$$

Remark 2.8.4. Using Definition 2.8.3, we may write, for example,

$$\mathcal{P}_{\gamma,\mathbf{k}}\left(\prod_{t=1}^{p} G^{\gamma}_{\mathbf{u}_t\mathbf{v}_t}\right) := \prod_{t=1}^{p} S^{*_\gamma(k_t+1)}_{\mathbf{u}_t\mathbf{v}_t}, \quad \mathcal{P}_{\gamma,\mathbf{k}}\left(\prod_{t=1}^{p} G^{\gamma-1}_{\mathbf{u}_t\mathbf{v}_t}\right) := \prod_{t=1}^{p} T^{*_\gamma(k_t+1)}_{\mathbf{u}_t\mathbf{v}_t}.$$

For $k, s \in \mathbb{N}$ and $\mathbf{k} \in \mathbb{N}^{s+1}$, it is easy to verify that

$$G^{*_\gamma s} *_\gamma G^{*_\gamma k} = G^{*_\gamma(s+k)}, \quad \mathcal{P}_{\gamma,\mathbf{k}}(\mathcal{P}_{\gamma,s}G_{\mathbf{uv}}) = \mathcal{P}_{\gamma,s+|\mathbf{k}|}G_{\mathbf{uv}}, \tag{2.8.17}$$

where $|\mathbf{k}| := \sum_{t=1}^{p} k_t$ is the l^1-norm of $\mathbf{k}$. For the second equality, note that $\mathcal{P}_{\gamma,s}G_{\mathbf{uv}}$ is a sum of the products of G entries, where each product contains $s+1$ terms.

Remark 2.8.5. It is easy to see that for any fixed $k \in \mathbb{N}$, $\mathcal{P}_{\gamma,k}G_{\mathbf{uv}}$ is a sum of finitely many products of $(k+1)$ resolvent entries of the form $\mathbb{G}_{\mathbf{xy}}$, $\mathbf{x},\mathbf{y} \in \{\mathbf{u},\mathbf{v},\mathbf{w}_i,\mathbf{w}_\mu\}$. Hence by (2.4.22) and (2.8.11), we can bound $\mathcal{P}_{\gamma,k}G_{\mathbf{uv}}$ by $\mathrm{O}_\prec(1)$. This is one of the main reasons why we need to prove the stronger anisotropic local law for G, rather than the entrywise local law only as in [62].

Now we begin to perform the resolvent comparison strategy. The basic idea is to expand S and T in terms of R using the resolvent expansions as in (2.8.7) and (2.8.8), and then compare the two expressions. We expect that the main terms will cancel since $x_{i\mu}$ and $\widetilde{x}_{i\mu}$ have the same first four moments, while the remaining error terms will be sufficiently small since $x_{i\mu}$ and $\widetilde{x}_{i\mu}$ have support bounded by $N^{-\phi}$. The key of the comparison argument is the following Lemma 2.8.6. Its proof is almost the same as the one for [62, Lemma 6.5]. In fact, we can copy their arguments almost verbatim, except for some notational differences. We leave the details to the reader.

Lemma 2.8.6. *Given $z \in \widetilde{S}(c_0, C_0, \varepsilon)$ and $\Phi(i,\mu) = \gamma$. For S, R in (2.8.5), we have*

$$\mathbb{E}\prod_{t=1}^{p} S_{\mathbf{u}_t\mathbf{v}_t} = \sum_{0\leqslant k\leqslant 4} A_k \mathbb{E}\left[(-x_{i\mu})^k\right] + \sum_{5\leqslant|\mathbf{k}|\leqslant r/\phi, \mathbf{k}\in\mathbb{N}^p} \mathcal{A}_{\mathbf{k}} \mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{p} S_{\mathbf{u}_t\mathbf{v}_t} + \mathrm{O}_{\prec}(N^{-r}), \quad (2.8.18)$$

where A_k, $0 \leqslant k \leqslant 4$, depend only on R, $\mathcal{A}_{\mathbf{k}}$'s do not depend on the deterministic unit vectors $(\mathbf{u}_t, \mathbf{v}_t)$, $1 \leqslant t \leqslant p$, and we have the bound

$$|\mathcal{A}_{\mathbf{k}}| \leqslant N^{-|\mathbf{k}|\phi/10-2}. \quad (2.8.19)$$

Similarly, we have

$$\mathbb{E}\prod_{t=1}^{p} (S-\Pi)_{\mathbf{u}_t\mathbf{v}_t} = \sum_{0\leqslant k\leqslant 4} \widetilde{A}_k \mathbb{E}\left[(-x_{i\mu})^k\right] + \sum_{5\leqslant|\mathbf{k}|\leqslant r/\phi, \mathbf{k}\in\mathbb{N}^p} \mathcal{A}_{\mathbf{k}} \mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{p} (S-\Pi)_{\mathbf{u}_t\mathbf{v}_t} + \mathrm{O}_{\prec}(N^{-r}), \quad (2.8.20)$$

where $\widetilde{A}_k$, $0 \leqslant k \leqslant 4$, again depend only on R. Finally, we have

$$\mathbb{E}\prod_{t=1}^{p} S_{\mathbf{u}_t\mathbf{v}_t} = \mathbb{E}\prod_{t=1}^{p} R_{\mathbf{u}_t\mathbf{v}_t} + \sum_{1\leqslant|\mathbf{k}|\leqslant r/\phi, \mathbf{k}\in\mathbb{N}^p} \widetilde{\mathcal{A}}_{\mathbf{k}} \mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{p} S_{\mathbf{u}_t\mathbf{v}_t} + \mathrm{O}_{\prec}(N^{-r}), \quad (2.8.21)$$

where $\widetilde{\mathcal{A}}_{\mathbf{k}}$'s do not depend on $(\mathbf{u}_t, \mathbf{v}_t)$, $1 \leqslant t \leqslant p$, and

$$|\widetilde{\mathcal{A}}_{\mathbf{k}}| \leqslant N^{-|\mathbf{k}|\phi/10}. \quad (2.8.22)$$

Note that the terms A, $\widetilde{A}$, $\mathcal{A}$ and $\widetilde{\mathcal{A}}$ do depend on γ and we have omitted this dependence in the above expressions.

We now use Lemma 2.8.6 to finish the proof of Lemma 2.4.7 and Lemma 2.8.1. It is obvious that a result similar to Lemma 2.8.6 also holds for the product of T entries. As in (2.8.18), we define the notation $\mathcal{A}^{\gamma,a}$, $a = 0, 1$ as follows:

$$\mathbb{E}\prod_{t=1}^{p} S_{\mathbf{u}_t\mathbf{v}_t} = \sum_{0\leqslant k\leqslant 4} A_k \mathbb{E}\left[(-x_{i\mu})^k\right] + \sum_{5\leqslant|\mathbf{k}|\leqslant r/\phi, \mathbf{k}\in\mathbb{N}^p} \mathcal{A}_{\mathbf{k}}^{\gamma,0} \mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{p} S_{\mathbf{u}_t\mathbf{v}_t} + \mathrm{O}_{\prec}(N^{-r}), \quad (2.8.23)$$

$$\mathbb{E}\prod_{t=1}^{p} T_{\mathbf{u}_t\mathbf{v}_t} = \sum_{0\leqslant k\leqslant 4} A_k \mathbb{E}\left[(-\widetilde{x}_{i\mu})^k\right] + \sum_{5\leqslant|\mathbf{k}|\leqslant r/\phi,\mathbf{k}\in\mathbb{N}^p} \mathcal{A}_{\mathbf{k}}^{\gamma,1}\mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} T_{\mathbf{u}_t\mathbf{v}_t} + \mathrm{O}_{\prec}(N^{-r}). \tag{2.8.24}$$

Since A_k, $0 \leqslant k \leqslant 4$, depend only on R and $x_{i\mu}$, $\widetilde{x}_{i\mu}$ have the same first four moments, we get from (2.8.23) and (2.8.24) that

$$\begin{aligned} &\mathbb{E}\prod_{t=1}^{p} G_{\mathbf{u}_t\mathbf{v}_t} - \mathbb{E}\prod_{t=1}^{p} \widetilde{G}_{\mathbf{u}_t\mathbf{v}_t} = \sum_{\gamma=1}^{\gamma_{\max}}\left(\mathbb{E}\prod_{t=1}^{p} G^{\gamma}_{\mathbf{u}_t\mathbf{v}_t} - \mathbb{E}\prod_{t=1}^{p} G^{\gamma-1}_{\mathbf{u}_t\mathbf{v}_t}\right) \\ &= \sum_{\gamma=1}^{\gamma_{\max}}\sum_{\mathbf{k}\in\mathbb{N}^p}^{5\leqslant|\mathbf{k}|\leqslant r/\phi}\left(\mathcal{A}_{\mathbf{k}}^{\gamma,0}\mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{\gamma}_{\mathbf{u}_t\mathbf{v}_t} - \mathcal{A}_{\mathbf{k}}^{\gamma,1}\mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{\gamma-1}_{\mathbf{u}_t\mathbf{v}_t}\right) + \mathrm{O}_{\prec}(N^{-r+2}). \end{aligned} \tag{2.8.25}$$

where we abbreviate $G := G(X, z)$ and $\widetilde{G} := G(\widetilde{X}, z)$. With a similar argument, we also have

$$\begin{aligned} &\mathbb{E}\prod_{t=1}^{p}(G-\Pi)_{\mathbf{u}_t\mathbf{v}_t} - \mathbb{E}\prod_{t=1}^{p}(\widetilde{G}-\Pi)_{\mathbf{u}_t\mathbf{v}_t} \\ &= \sum_{\gamma=1}^{\gamma_{\max}}\sum_{\mathbf{k}\in\mathbb{N}^p}^{5\leqslant|\mathbf{k}|\leqslant r/\phi}\left(\mathcal{A}_{\mathbf{k}}^{\gamma,0}\mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p}(G^{\gamma}-\Pi)_{\mathbf{u}_t\mathbf{v}_t} - \mathcal{A}_{\mathbf{k}}^{\gamma,1}\mathbb{E}\,\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p}(G^{\gamma-1}-\Pi)_{\mathbf{u}_t\mathbf{v}_t}\right) + \mathrm{O}_{\prec}(N^{-r+2}). \end{aligned} \tag{2.8.26}$$

Note that by (2.8.25), we have

$$\left|\mathbb{E}\prod_{t=1}^{p} G^{\gamma_{\max}}_{\mathbf{u}_t\mathbf{v}_t}\right| \leqslant \left|\mathbb{E}\prod_{t=1}^{p} G^{0}_{\mathbf{u}_t\mathbf{v}_t}\right| + \sum_{\gamma=1}^{\gamma_{\max}}\sum_{a=0,1}\sum_{5\leqslant|\mathbf{k}|\leqslant r/\phi,\mathbf{k}\in\mathbb{N}^p}|\mathcal{A}_{\mathbf{k}}^{\gamma,a}|\left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{\gamma-a}_{\mathbf{u}_t\mathbf{v}_t}\right| + \mathrm{O}_{\prec}(N^{-r+2}). \tag{2.8.27}$$

By (2.4.22) and (2.8.19), the second term in (2.8.27) is bounded by

$$\sum_{5\leqslant k\leqslant r/\phi}\sum_{\gamma=1}^{\gamma_{\max}}\sum_{a=0,1}\sum_{|\mathbf{k}|=k,\mathbf{k}\in\mathbb{N}^p}|\mathcal{A}_{\mathbf{k}}^{\gamma,a}|\left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{\gamma-a}_{\mathbf{u}_t\mathbf{v}_t}\right| \prec \sum_{5\leqslant k\leqslant r/\phi} N^{-k\phi/10} \lesssim N^{-\phi/2}. \tag{2.8.28}$$

However, the bound in (2.8.28) is not good enough. To improve it, we iterate the above arguments as following. Recall that $\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{\gamma-a}_{\mathbf{u}_t\mathbf{v}_t}$ is also a sum of the products of G entries. Applying (2.8.25) again to $\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{\gamma-a}_{\mathbf{u}_t\mathbf{v}_t}$ and replacing $\gamma_{\max}$ in (2.8.27) with $\gamma - a$, we obtain that

$$\begin{aligned} \left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{\gamma-a}_{\mathbf{u}_t\mathbf{v}_t}\right| &\leqslant \left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{0}_{\mathbf{u}_t\mathbf{v}_t}\right| \\ &+ \sum_{\gamma'=1}^{\gamma-a}\sum_{a'=0,1}\sum_{5\leqslant|\mathbf{k}'|\leqslant r/\phi,\mathbf{k}'\in\mathbb{N}^{p+|\mathbf{k}|}}\left|\mathcal{A}_{\mathbf{k}'}^{\gamma',a'}\right|\left|\mathbb{E}\mathcal{P}_{\gamma',\mathbf{k}'}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p} G^{\gamma'-a'}_{\mathbf{u}_t\mathbf{v}_t}\right| + \mathrm{O}_{\prec}(N^{-r+2}). \end{aligned}$$

Together with (2.8.27), we have

$$\left|\mathbb{E}\prod_{t=1}^{p}G^{\gamma_{\max}}_{\mathbf{u}_t\mathbf{v}_t}\right|\leqslant\left|\mathbb{E}\prod_{t=1}^{p}G^{0}_{\mathbf{u}_t\mathbf{v}_t}\right|+\sum_{\gamma=1}^{\gamma_{\max}}\sum_{a=0,1}\sum_{5\leqslant|\mathbf{k}|\leqslant r/\phi,\mathbf{k}\in\mathbb{N}^p}|\mathcal{A}^{\gamma,a}_{\mathbf{k}}|\left|\mathbb{E}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p}G^{0}_{a_t,b_t}\right|$$
$$+\sum_{\gamma,\gamma'}\sum_{a,a'}\sum_{\mathbf{k},\mathbf{k}'}\left|\mathcal{A}^{\gamma,a}_{\mathbf{k}}\mathcal{A}^{\gamma',a'}_{\mathbf{k}'}\right|\left|\mathbb{E}\mathcal{P}_{\gamma',\mathbf{k}'}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p}G^{\gamma'-a'}_{\mathbf{u}_t\mathbf{v}_t}\right|+\mathrm{O}_\prec(N^{-r+2}).$$

Again using (2.4.22) and (2.8.19), it is easy to see that

$$\sum_{\gamma,\gamma'}\sum_{a,a'}\sum_{\mathbf{k},\mathbf{k}'}\left|\mathcal{A}^{\gamma,a}_{\mathbf{k}}\mathcal{A}^{\gamma',a'}_{\mathbf{k}'}\right|\left|\mathbb{E}\mathcal{P}_{\gamma',\mathbf{k}'}\mathcal{P}_{\gamma,\mathbf{k}}\prod_{t=1}^{p}G^{\gamma'-a'}_{\mathbf{u}_t\mathbf{v}_t}\right|\prec N^{-\phi},$$

where we used that $k'+k\geqslant 10$. Repeating the above process for $m\leqslant 2r/\phi$ times, we obtain that

$$\left|\mathbb{E}\prod_{t=1}^{p}G^{\gamma_{\max}}_{\mathbf{u}_t\mathbf{v}_t}\right|\leqslant\sum_{m=0}^{2r/\phi}\sum_{\gamma_1,\cdots,\gamma_m}\sum_{a_1,\cdots,a_m}\sum_{\mathbf{k}_1,\cdots,\mathbf{k}_m}\left|\prod_{j}\mathcal{A}^{\gamma_j,a_j}_{\mathbf{k}_j}\right|\left|\mathbb{E}\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}G^{0}_{\mathbf{u}_t\mathbf{v}_t}\right|+\mathrm{O}_\prec(N^{-r+2}),$$

where

$$\mathbf{k}_1\in\mathbb{N}^p,\quad \mathbf{k}_2\in\mathbb{N}^{p+|\mathbf{k}_1|},\quad \mathbf{k}_3\in\mathbb{N}^{p+|\mathbf{k}_1|+|\mathbf{k}_2|},\ \ldots,\quad\text{and}\quad 5\leqslant|\mathbf{k}_i|\leqslant\frac{r}{\phi}. \tag{2.8.29}$$

Using (2.4.22) and (2.8.19), we obtain that

$$\left|\mathbb{E}\prod_{t=1}^{p}G^{\gamma_{\max}}_{\mathbf{u}_t\mathbf{v}_t}\right|\leqslant\left|\mathbb{E}\prod_{t=1}^{p}G^{0}_{\mathbf{u}_t\mathbf{v}_t}\right|$$
$$+\mathrm{O}_\prec\left(\max_{\mathbf{k},m}(N^{-2})^m(N^{-\phi/10})^{\sum_i|\mathbf{k}_i|}\sum_{\gamma_1,\cdots,\gamma_m}\left|\mathbb{E}\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}G^{0}_{\mathbf{u}_t\mathbf{v}_t}\right|\right)+\mathrm{O}_\prec(N^{-r+2}). \tag{2.8.30}$$

We remark that the above estimate still holds if we replace some of the G entries with $\overline{G}$ entries, since we have only used the absolute bounds for the resolvent entries. Of course, using (2.8.26) instead of (2.8.25), we can obtain a similar estimate

$$\left|\mathbb{E}\prod_{t=1}^{p}(G^{\gamma_{\max}}-\Pi)_{\mathbf{u}_t\mathbf{v}_t}\right|\leqslant\left|\mathbb{E}\prod_{t=1}^{p}\left(G^{0}-\Pi\right)_{\mathbf{u}_t\mathbf{v}_t}\right|$$
$$+\mathrm{O}_\prec\left(\max_{\mathbf{k},m}(N^{-2})^m(N^{-\phi/10})^{\sum_i|\mathbf{k}_i|}\sum_{\gamma_1,\cdots,\gamma_m}\left|\mathbb{E}\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}\left(G^{0}-\Pi\right)_{\mathbf{u}_t\mathbf{v}_t}\right|\right)+\mathrm{O}_\prec(N^{-r+2}). \tag{2.8.31}$$

Now we use Lemma 2.8.6, (2.8.30) and (2.8.31) to complete the proof of Lemma 2.4.7 and Lemma 2.8.1.

Proof of Lemma 2.4.7. We apply (2.8.31) to $(G-\Pi)_{\mathbf{uv}}\overline{(G-\Pi)_{\mathbf{uv}}}$ with $p=2$ and $r=3$. Recall that $\widetilde{X}$ is of bounded support $q=N^{-1/2}$. Then by (2.4.22) and Lemma 2.4.2, we have

$$\mathbb{E}|(\widetilde{G}-\Pi)_{\mathbf{uv}}|^2 \prec \Psi^2(z). \tag{2.8.32}$$

Moreover, by (2.4.19) the remainder term $\mathrm{O}_\prec(N^{-r+2})=\mathrm{O}_\prec(N^{-1})$ in (2.8.31) is negligible. Hence it remains to handle the second term on the right-hand side of (2.8.31), i.e.

$$(N^{-2})^m \sum_{\gamma_1,\cdots,\gamma_m} \left|\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\left|\left(G^0-\Pi\right)_{\mathbf{uv}}\right|^2\right|. \tag{2.8.33}$$

For each product in (2.8.33), $\mathbf{v}$ appears exactly twice in the indices of G. These two $\mathbf{v}$'s appear as $G_{\mathbf{v}\mathbf{w}_a}G_{\mathbf{w}_b\mathbf{v}}$ in the product, where $\mathbf{w}_a,\mathbf{w}_b$ come from some γ_k and γ_l $(1\leqslant k,l\leqslant m)$ via $\mathcal{P}$. Let $\mathbf{v}=\begin{pmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{pmatrix}$ for $\mathbf{v}_1\in\mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{v}_2\in\mathbb{C}^{\mathcal{I}_2}$. By Lemma 2.7.1, after taking the averages $N^{-2}\sum_{\gamma_k}$ and $N^{-2}\sum_{\gamma_l}$, the term $G_{\mathbf{v}\mathbf{w}_a}G_{\mathbf{w}_b\mathbf{v}}$ contributes a factor

$$\begin{aligned}&\mathrm{O}_\prec\left(\frac{\operatorname{Im}\left(z^{-1}G^0_{\mathbf{v}_1\mathbf{v}_1}\right)+\operatorname{Im}\left(G^0_{\mathbf{v}_2\mathbf{v}_2}\right)+\eta\left|G^0_{\mathbf{v}_1\mathbf{v}_1}\right|+\eta\left|G^0_{\mathbf{v}_2\mathbf{v}_2}\right|}{N\eta}\right)\\ &=\mathrm{O}_\prec\left(\frac{\operatorname{Im} m_{2c}+\Psi(z)}{N\eta}\right)=\mathrm{O}_\prec(\Psi^2(z)),\end{aligned} \tag{2.8.34}$$

where we used (2.4.22). For all the other G factors in the product, we control them by $\mathrm{O}_\prec(1)$ using (2.4.22). Thus for any $\mathbf{k}_1,\ldots,\mathbf{k}_m$, we have proved that (2.8.33) $\prec\Psi^2(z)$. Together with (2.8.31) and (2.8.32), this proves Lemma 2.4.7. □

Proof of Lemma 2.8.1. For simplicity of notations, instead of (2.8.2), we shall prove that

$$|\mathbb{E}\,(m(z)-m_c(z))^p| \prec |\mathbb{E}\,(\widetilde{m}(z)-m_c(z))^p| + \left(\Psi^2(z)+J^2+K\right)^p. \tag{2.8.35}$$

The proof for (2.8.2) is exactly the same but with slightly heavier notations.

Define a function $h(I,J)$ such that

$$\sum_{I,J} h(I,J)=1,\quad h(I,J)\geqslant 0,\quad I=(i_1,i_2,\cdots,i_p)\in\mathcal{I}_1^p,\quad J=(j_1,j_2,\cdots,j_p)\in\mathcal{I}_1^p. \tag{2.8.36}$$

Since $\mathcal{A}$'s do not depend on $\mathbf{u}_t$, $\mathbf{v}_t$, we may consider a linear combination of (2.8.31) with coefficients $f(I,J)$:

$$\left|\mathbb{E}\sum_{I,J} f(I,J)\prod_{t=1}^{p}(G-\Pi)_{i_t j_t}\right| = \left|\mathbb{E}\sum_{I,J} f(I,J)\prod_{t=1}^{p}(\widetilde{G}-\Pi)_{i_t j_t}\right|$$
$$+ \mathrm{O}_{\prec}\left(\max_{\mathbf{k},m,\gamma}(N^{-\phi/10})^{\sum_i |\mathbf{k}_i|}\left|\mathbb{E}\sum_{I,J} f(I,J)\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}(\widetilde{G}-\Pi)_{i_t j_t}\right|\right) + \mathrm{O}(N^{-r+2}). \tag{2.8.37}$$

If we take $r = p+2$ and $f(I,J) = n^{-p}\prod \delta_{i_t j_t}$, it is easy to check that

$$\mathbb{E}\sum_{I,J} f(I,J)\prod_{t=1}^{p}(G^{\alpha}-\Pi)_{i_t j_t} = \mathbb{E}(m^{\alpha}-m_c)^p, \quad \alpha = 0,\ \gamma_{\max}. \tag{2.8.38}$$

Now to conclude (2.8.35), it suffices to control the second term on the RHS of (2.8.37). We consider the terms

$$\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}(\widetilde{G}-\Pi)_{i_t i_t}, \tag{2.8.39}$$

for $\mathbf{k}_1,\ldots,\mathbf{k}_m$ satisfying (2.8.29). For each product in (2.8.39) and any $1 \leqslant t \leqslant p$, there are two i_t's in the indices of G. These two i_t's can only appear as (1) $(\widetilde{G}-\Pi)_{i_t i_t}$ in the product, or (2) $\widetilde{G}_{i_t \mathbf{w}_a}\widetilde{G}_{\mathbf{w}_b i_t}$, where $\mathbf{w}_a, \mathbf{w}_b$ come from some γ_k and γ_l via $\mathcal{P}$. Then after averaging over $n^{-p}\sum_{i_1,\cdots,i_p}$, this term becomes either (1) $\widetilde{m}-m_c$, which is bounded by K by (2.8.1), or (2) $n^{-1}\sum_{i_t} G_{i_t \mathbf{w}_a}G_{\mathbf{w}_b i_t}$, which is bounded as in (2.8.34) by

$$\mathrm{O}_{\prec}\left(\frac{\operatorname{Im} m_{2c}+J}{N\eta}\right) = \mathrm{O}_{\prec}\left(\Psi^2(z)+J^2\right).$$

For other G entries in the product with no i_t, we simply bound them by $O_{\prec}(1)$ using (2.8.1). Then for any fixed $\gamma_1,\ldots,\gamma_m$, $\mathbf{k}_1,\ldots,\mathbf{k}_m$, we have proved that

$$\left|\frac{1}{n^p}\sum_{i_1,\ldots,i_p}\mathbb{E}\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}\left(\widetilde{G}-\Pi\right)_{i_t i_t}\right| \prec \left(\Psi^2(z)+J^2+K\right)^p. \tag{2.8.40}$$

Together with (2.8.37), this concludes (2.8.35). □

2.8.2 Proof of Theorem 2.4.10

For the matrix $\widetilde{X}$ constructed in Lemma 2.4.12, it satisfies the edge universality by the following lemma.

Lemma 2.8.7. *Let $X^{(1)}$ and $X^{(2)}$ be two separable covariance matrices satisfying the assumptions in Theorem 2.4.6 and the bounded support condition* (2.4.1) *with $q = N^{-1/2}$. Suppose $b_N \leqslant N^{1/3-c}$ for some constant $c > 0$. Then there exist constants $\varepsilon, \delta > 0$ such that for any $s \in \mathbb{R}$, we have*

$$\begin{aligned} \mathbb{P}^{(1)}\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s - N^{-\varepsilon}\right) - N^{-\delta} &\leqslant \mathbb{P}^{(2)}\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s\right) \\ &\leqslant \mathbb{P}^{(1)}\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s + N^{-\varepsilon}\right) + N^{-\delta}, \end{aligned} \tag{2.8.41}$$

where $\mathbb{P}^{(1)}$ and $\mathbb{P}^{(2)}$ denote the laws of $X^{(1)}$ and $X^{(2)}$, respectively.

Proof. The proof of this lemma is similar to the ones in [31, Section 6], [38, Section 6], [79, Section 4] and [59, Section 10]. The main argument involves a routine application of the Green's function comparison method (as the one in Lemma 2.8.9) near the edge developed in [38, Section 6] and [79, Section 4]. The proofs there can be easily adapted to our case using the anisotropic local law (Theorem 2.4.6), the rigidity of eigenvalues (Theorem 2.4.8), and the resolvent identities in Lemma 2.6.3 and Lemma 2.7.1. □

Now it is easy to see that Theorem 2.4.10 follows from the following comparison lemma.

Lemma 2.8.8. *Let X and $\widetilde{X}$ be two matrices as in Lemma 2.4.12. Suppose $b_N \leqslant N^{1/3-c}$ for some constant $c > 0$. Then there exist constants $\varepsilon, \delta > 0$ such that, for any $s \in \mathbb{R}$ we have*

$$\begin{aligned} \mathbb{P}^{\widetilde{X}}\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s - N^{-\varepsilon}\right) - N^{-\delta} &\leqslant \mathbb{P}^{X}(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s) \\ &\leqslant \mathbb{P}^{\widetilde{X}}\left(N^{2/3}(\lambda_1 - \lambda_r) \leqslant s + N^{-\varepsilon}\right) + N^{-\delta}, \end{aligned} \tag{2.8.42}$$

where $\mathbb{P}^{X}$ and $\mathbb{P}^{\widetilde{X}}$ are the laws for X and $\widetilde{X}$, respectively.

To prove Lemma 2.8.8, it suffices to prove the following Green's function comparison result.

Lemma 2.8.9. *Let X and $\widetilde{X}$ be two matrices as in Lemma 2.4.12. Suppose $F : \mathbb{R} \to \mathbb{R}$ is a function whose derivatives satisfy*

$$\sup_x |F^{(k)}(x)|(1 + |x|)^{-C_1} \leqslant C_1, \quad k = 1, 2, 3, \tag{2.8.43}$$

for some constant $C_1 > 0$. Then for any sufficiently small constant $\delta > 0$ and for any

$$E, E_1, E_2 \in I_\delta := \left\{x : |x - \lambda_r| \leqslant N^{-2/3+\delta}\right\} \quad \text{and} \quad \eta := N^{-2/3-\delta},$$

we have

$$|\mathbb{E}F\left(N\eta \operatorname{Im} m(z)\right) - \mathbb{E}F\left(N\eta \operatorname{Im} \widetilde{m}(z)\right)| \leqslant N^{-\phi+C_2\delta}, \quad z = E + \mathrm{i}\eta, \tag{2.8.44}$$

and

$$\left|\mathbb{E}F\left(N\int_{E_1}^{E_2} \operatorname{Im} m(y+\mathrm{i}\eta)dy\right) - \mathbb{E}F\left(N\int_{E_1}^{E_2} \operatorname{Im} \widetilde{m}(y+\mathrm{i}\eta)dy\right)\right| \leqslant N^{-\phi+C_2\delta}, \tag{2.8.45}$$

where ϕ is as given in Theorem 2.4.6 and $C_2 > 0$ is some constant.

Proof of Lemma 2.8.8. Although not explicitly stated, it was shown in [38] that if Theorem 2.4.8 and Lemma 2.8.9 hold, then the edge universality (2.8.42) holds. More precisely, in Section 6 of [38], the edge universality problem was reduced to proving Theorem 6.3 of [38], which corresponds to our Lemma 2.8.9. In order for this conversion to work, only the the averaged local law and the rigidity of eigenvalues are used, which correspond to the statements in our Theorem 2.4.8. □

Proof of Lemma 2.8.9. For simplicity, we only prove (2.8.44). The proof for (2.8.45) is similar with only some notational differences. By (2.6.10), we have

$$N\eta \operatorname{Im} m(z) = \frac{N\eta^2}{n|z|^2} \sum_{i,j} |G_{ij}(z)|^2. \tag{2.8.46}$$

Since $N \sim n$ and $|z| \sim 1$, it is equivalent to prove that

$$\left|\mathbb{E}F\left(\eta^2 \sum_{i,j} G_{ij}\overline{G}_{ij}\right) - \mathbb{E}F\left(\eta^2 \sum_{i,j} \widetilde{G}_{ij}\overline{\widetilde{G}}_{ij}\right)\right| \leqslant N^{-\phi+C_2\varepsilon},$$

for $z = E + \mathrm{i}\eta$ with $E \in I_\delta$ and $\eta = N^{-2/3-\delta}$. Corresponding to the notations in (2.8.5), we denote

$$x^S := \eta^2 \sum_{i,j} S_{ij}\overline{S}_{ij}, \quad x^R := \eta^2 \sum_{i,j} R_{ij}\overline{R}_{ij}, \quad x^T := \eta^2 \sum_{i,j} T_{ij}\overline{T}_{ij}.$$

Applying (2.8.46) to S, T and using (2.4.26) and (2.4.11), we get that with high probability,

$$\max_\gamma \max\left\{\left|x^S\right|, \left|x^T\right|\right\} \prec 1. \tag{2.8.47}$$

Using (2.8.7), Lemma 2.7.1, (2.4.22) and (2.8.11), one can obtain that

$$|\operatorname{Tr} S - \operatorname{Tr} R| \prec \eta^{-1}. \tag{2.8.48}$$

Together with (2.8.47), we also get that

$$\max_{\gamma} \left| x^R \right| \prec 1. \tag{2.8.49}$$

By (2.4.22), (2.8.9) and the expansion (2.8.8), we also get that

$$S - \Pi = \mathrm{O}_{\prec}\left(N^{-\phi} + N^{-1/3+\delta}\right), \quad R - \Pi = \mathrm{O}_{\prec}\left(N^{-\phi} + N^{-1/3+\delta}\right). \tag{2.8.50}$$

Without loss of generality, we assume that $\phi \leqslant 1/3 - \delta$ in the following proof.

Applying the Lindeberg replacement strategy, we get that

$$\mathbb{E}F\left(\eta^2 \sum_{i,j} G_{ij}\overline{G}_{ij}\right) - \mathbb{E}F\left(\eta^2 \sum_{i,j} \widetilde{G}_{ij}\overline{\widetilde{G}}_{ij}\right) = \sum_{\gamma=1}^{\gamma_{\max}} \left[\mathbb{E}F\left(x^S\right) - \mathbb{E}F\left(x^T\right)\right]. \tag{2.8.51}$$

From the Taylor expansion, we have

$$F\left(x^S\right) - F\left(x^R\right) = \sum_{l=1}^{2} \frac{1}{l!} F^{(l)}\left(x^R\right)\left(x^S - X^R\right)^s + \frac{1}{3!} F^{(3)}\left(\zeta_S\right)\left(x^S - x^R\right)^3, \tag{2.8.52}$$

where ζ_S lies between x^S and x^R. We have a similar expansion for $F\left(x^T\right) - F\left(x^R\right)$ with ζ_S replaced by ζ_T.

Let $\Phi(i,\mu) = \gamma$ and fix $r \in \mathbb{N}$. We perform the expansion (2.8.7) to get that

$$S_{a_t b_t} = \sum_{0 \leqslant k \leqslant r} (-x_{i\mu})^k \mathcal{P}_{\gamma,k} R_{a_t b_t} + \mathrm{O}_{\prec}(N^{-r\phi}), \quad a_t, b_t \in \mathcal{I}. \tag{2.8.53}$$

Using this expansion and bound (2.8.11), we have that

$$\prod_{t=1}^{p} S_{a_t b_t} = \sum_{0 \leqslant k \leqslant rp} \sum_{\mathbf{k} \in I_{r,k}^p} \left(\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{p} R_{a_t b_t}\right)(-x_{i\mu})^k + \mathrm{O}_{\prec}\left(N^{-r\phi}\right), \tag{2.8.54}$$

where

$$\mathbf{k} := (k_1, \cdots, k_p), \quad I_{r,k}^p = \left\{\mathbf{k} \in \mathbb{N}^p : 0 \leqslant k_i \leqslant r, \sum k_i = k\right\}. \tag{2.8.55}$$

By (2.8.11), the $k > r$ terms in (2.8.54) can be bounded by

$$\left|\sum_{k>r} \sum_{\mathbf{k} \in I_{r,k}^p} \left(\mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{p} R_{a_t b_t}\right)(-x_{i\mu})^k\right| \prec \sum_{k>r} N^{-k\phi} = \mathrm{O}(N^{-r\phi}).$$

Hence (2.8.54) is reduced to

$$\prod_{t=1}^{p} S_{a_t b_t} = \prod_{t=1}^{p} R_{a_t b_t} + \sum_{1\leqslant k\leqslant r} (-x_{i\mu})^k \left(\sum_{\mathbf{k}\in I_{r,k}^p} \mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{p} R_{a_t b_t} \right) + \mathrm{O}_{\prec}\left(N^{-r\phi}\right). \tag{2.8.56}$$

Similarly, we also have

$$\prod_{t=1}^{p} T_{a_t b_t} = \prod_{t=1}^{p} R_{a_t b_t} + \sum_{1\leqslant k\leqslant r} (-\widetilde{x}_{i\mu})^k \left(\sum_{\mathbf{k}\in I_{r,k}^p} \mathcal{P}_{\gamma,\mathbf{k}} \prod_{t=1}^{p} R_{a_t b_t} \right) + \mathrm{O}_{\prec}\left(N^{-r\phi}\right). \tag{2.8.57}$$

Obviously we can replace some of the resolvent entries with their complex conjugates by modifying the notations slightly. Now we apply (2.8.56) and (2.8.57) with $p = 2$ and $r := 3/\phi$ to get that

$$x^S = x^R + \sum_{1\leqslant k\leqslant 3/\phi} \left(\sum_{\mathbf{k}\in I_{3/\phi,k}^2} \eta^2 \sum_{i,j} \mathcal{P}_{\gamma,\mathbf{k}}(R_{ij}\overline{R}_{ij}) \right) (-x_{i\mu})^k + \mathrm{O}_{\prec}(N^{-3}), \tag{2.8.58}$$

and

$$x^T = x^R + \sum_{1\leqslant k\leqslant 3/\phi} \left(\sum_{\mathbf{k}\in I_{3/\phi,k}^2} \eta^2 \sum_{i,j} \mathcal{P}_{\gamma,\mathbf{k}}(R_{ij}\overline{R}_{ij}) \right) (-\widetilde{x}_{i\mu})^k + \mathrm{O}_{\prec}(N^{-3}). \tag{2.8.59}$$

To control the second term in (2.8.58), we have the following lemma.

Lemma 2.8.10. *For any fixed* $\mathbf{k} \neq 0$ *and* $\mathbf{k} \in I_{3/\phi,k}^2$*, we have*

$$\left| \sum_{i,j} \mathcal{P}_{\gamma,\mathbf{k}} \left(R_{ij}\overline{R}_{ij}\right) \right| \prec N^{1+C\delta} \tag{2.8.60}$$

for some constant $C > 0$*.*

Before proving this lemma, we first use it to finish the proof of Lemma 2.8.9. Given (2.8.60) and $\eta = N^{-2/3-\delta}$, we see that there exists constant $C > 0$ such that

$$|\mathcal{P}_{\gamma,\mathbf{k}} x^R| := \left| \eta^2 \sum_{i,j} \mathcal{P}_{\gamma,\mathbf{k}} \left(R_{ij}\overline{R}_{ij}\right) \right| \prec N^{-1/3+C\delta}. \tag{2.8.61}$$

Combining (2.8.58), (2.8.61) and (2.4.15), we see that there exists a constant $C > 0$ such that

$$\mathbb{E}|x^S - x^R|^3 \leqslant N^{-5/2+C\delta}. \tag{2.8.62}$$

Since ζ_S is between x^S and x^R, we have $|\zeta_S| \prec 1$ by (2.8.47) and (2.8.49). Together with (2.8.62) and the assumption (2.8.43), we get

$$\left|\sum_{\gamma=1}^{\gamma_{\max}} \mathbb{E}\left[F^{(3)}(\zeta_S)\left(x^S - x^R\right)^3\right]\right| \leqslant N^{-1/2+C\delta}. \tag{2.8.63}$$

We have a similar estimate for $\mathbb{E}\left[F^{(3)}(\zeta_T)\left(x^T - x^R\right)^3\right]$. Now it only remains to deal with the first sum on the right-hand side of (2.8.52). Using (2.8.58), (2.8.59) and the fact that the first four moments of $x_{i\mu}$ and $\widetilde{x}_{i\mu}$ match, we obtain that for $l = 1, 2$,

$$\begin{aligned} &\left|\mathbb{E}\left[F^{(l)}(x^R)\left(x^S - x^R\right)^l\right] - \mathbb{E}\left[F^{(l)}(x^R)\left(x^T - x^R\right)^l\right]\right| \\ &\leqslant \sum_{k=5}^{6/\phi}\left|\sum_{\sum_{t=1}^{l}|\mathbf{k}_t|=k}\ \sum_{\mathbf{k}_t \in I^{2l}_{3/\phi,k}} \mathbb{E}\prod_{t=1}^{l}\left(\mathcal{P}_{\gamma,\mathbf{k}_t} x^R\right)\right|\left(\left|\mathbb{E}(-x_{i\mu})^k\right| + \left|\mathbb{E}(-\widetilde{x}_{i\mu})^k\right|\right) + \mathrm{O}_\prec(N^{-3+C\delta}). \end{aligned}$$

Recall that (2.4.15) holds for $x_{i\mu}$ and $\widetilde{x}_{i\mu}$, $x_{i\mu}$ has support $\mathrm{O}_\prec(N^{-\phi})$, and $\widetilde{x}_{i\mu}$ has support $\mathrm{O}_\prec(N^{-1/2})$. Then it is easy to check that $|\mathbb{E}(-\widetilde{x}_{i\mu})^k| \prec N^{-5/2}$ and $|\mathbb{E}(-x_{i\mu})^k| \prec N^{-2-\phi}$ for any fixed $k \geqslant 5$. Using (2.8.61), we obtain that for $l \in \{1, 2\}$,

$$\left|\mathbb{E}\left[F^{(l)}(x^R)(x^S - x^R)^l\right] - \mathbb{E}\left[F^{(l)}(x^R)(x^T - x^R)^l\right]\right| \leqslant N^{-2-\phi+C\delta},$$

Together with (2.8.51), (2.8.52) and (2.8.63), this concludes the proof of (2.8.44). □

Proof of Lemma 2.8.60. By Markov's inequality, it suffices to prove that for any fixed $p \in 2\mathbb{N}$,

$$\mathbb{E}\left|\sum_{i,j}\mathcal{P}_{\gamma,\mathbf{k}}\left(R_{ij}\overline{R}_{ij}\right)\right|^p \prec (N^{1+C\delta})^p. \tag{2.8.64}$$

For simplicity, we shall show the proof for

$$\mathbb{E}\left|\left[\sum_{i,j}\mathcal{P}_{\gamma,\mathbf{k}}\left(R_{ij}\overline{R}_{ij}\right)\right]^p\right| \prec (N^{1+C\delta})^p. \tag{2.8.65}$$

The proof for (2.8.64) is similar with slightly heavier notations.

Using (2.8.21) with $r = p$, we have

$$\begin{aligned} &\mathbb{E}\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(R_{i_t j_t}\overline{R}_{i_t j_t}\right) \\ &= \mathbb{E}\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(S_{i_t j_t}\overline{S}_{i_t j_t}\right) - \sum_{1\leqslant|\boldsymbol{\alpha}|\leqslant p/\phi}\widetilde{\mathcal{A}}_{\boldsymbol{\alpha}}\mathbb{E}\,\mathcal{P}_{\gamma,\boldsymbol{\alpha}}\left[\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(S_{i_t j_t}\overline{S}_{i_t j_t}\right)\right] + \mathrm{O}_\prec(N^{-p}). \end{aligned}$$

With (2.8.22), in order to show (2.8.65), it suffices to prove that

$$\left|\sum_{i_1,j_1,\cdots,i_p,j_p}\mathbb{E}\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(S_{i_tj_t}\overline{S}_{i_tj_t}\right)\right|\leqslant(N^{1+C\delta})^p, \tag{2.8.66}$$

and for $\boldsymbol{\alpha}\in\mathbb{N}^{(|\mathbf{k}|+2)p}$, $1\leqslant|\boldsymbol{\alpha}|\leqslant p/\phi$,

$$\left|\sum_{i_1,j_1,\cdots,i_p,j_p}\mathbb{E}\mathcal{P}_{\gamma,\boldsymbol{\alpha}}\left[\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(S_{i_tj_t}\overline{S}_{i_tj_t}\right)\right]\right|\leqslant(N^{1+C\delta})^p. \tag{2.8.67}$$

We only prove (2.8.66), and the proof for (2.8.67) is exactly the same except for the one more $\mathcal{P}_{\gamma,\boldsymbol{\alpha}}$ factor. Using a similar estimate as in (2.8.37) with

$$f(I,J)=n^{-2p},\quad I=(i_1,i_2,\cdots,i_p)\in\mathcal{I}_1^p,\quad J=(j_1,j_2,\cdots,j_p)\in\mathcal{I}_1^p,$$

and $\mathcal{P}_{\gamma,\mathbf{k}}\left(S_{i_tj_t}\overline{S}_{i_tj_t}\right)$ playing the role of $(G-\Pi)_{i_tj_t}$, we obtain that

$$\begin{aligned}&\left|\mathbb{E}\sum_{I,J}f(I,J)\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(S_{i_tj_t}\overline{S}_{i_tj_t}\right)\right|=\left|\mathbb{E}\sum_{I,J}f(I,J)\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(\widetilde{G}_{i_tj_t}\overline{\widetilde{G}}_{i_tj_t}\right)\right|\\&+\mathrm{O}_{\prec}\left(\max_{\mathbf{k},m,\gamma}(N^{-\phi/10})^{\sum_i|\mathbf{k}_i|}\left|\mathbb{E}\sum_{I,J}f(I,J)\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(\widetilde{G}_{i_tj_t}\overline{\widetilde{G}}_{i_tj_t}\right)\right|\right)+\mathrm{O}_{\prec}(N^{-r+2}),\end{aligned} \tag{2.8.68}$$

where

$$\mathbf{k}_1\in\mathbb{N}^{(|\mathbf{k}|+2)p},\quad\mathbf{k}_2\in\mathbb{N}^{(|\mathbf{k}|+2)p+|\mathbf{k}_1|},\quad\mathbf{k}_3\in\mathbb{N}^{(|\mathbf{k}|+2)p+|\mathbf{k}_1|+|\mathbf{k}_2|},\ \ldots,\ \text{ and }\ 5\leqslant|\mathbf{k}_i|\leqslant\frac{r}{\phi}.$$

Taking $r=p+2$, we see that to show (2.8.66), it suffices to prove that

$$\left|\mathbb{E}\sum_{I,J}f(I,J)\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(\widetilde{G}_{i_tj_t}\overline{\widetilde{G}}_{i_tj_t}\right)\right|\leqslant(N^{-1+C\delta})^p, \tag{2.8.69}$$

and

$$\left|\mathbb{E}\sum_{I,J}f(I,J)\mathcal{P}_{\gamma_m,\mathbf{k}_m}\cdots\mathcal{P}_{\gamma_1,\mathbf{k}_1}\prod_{t=1}^{p}\mathcal{P}_{\gamma,\mathbf{k}}\left(\widetilde{G}_{i_tj_t}\overline{\widetilde{G}}_{i_tj_t}\right)\right|\leqslant(N^{-1+C\delta})^p. \tag{2.8.70}$$

We only prove (2.8.70), and the proof for (2.8.69) is exactly the same (and actually easier).

For each product in (2.8.70) and any fixed $1\leqslant t\leqslant p$, each of the indices i_t and j_t only appears twice. Since $\mathbf{k}\neq0$, they cannot contain the term $\widetilde{G}_{i_tj_t}\overline{\widetilde{G}}_{i_tj_t}$ and we must have one of the following three forms

$$\widetilde{G}_{i_t\mathbf{w}_a}\widetilde{G}_{\mathbf{w}_bj_t}\overline{\widetilde{G}}_{i_tj_t},\quad\widetilde{G}_{i_tj_t}\overline{\widetilde{G}_{i_t\mathbf{w}_a}\widetilde{G}_{\mathbf{w}_bj_t}},\quad\widetilde{G}_{i_t\mathbf{w}_a}\widetilde{G}_{\mathbf{w}_bj_t}\overline{\widetilde{G}_{i_t\mathbf{w}_c}\widetilde{G}_{\mathbf{w}_dj_t}},$$

where $\mathbf{w}_{a,b,c,d}$ come from some (possibly different) γ's via $\mathcal{P}$'s. Following a similar argument as below (2.8.39), each of the above form contributes a factor

$$\mathrm{O}_{\prec}\left(\left[\frac{\operatorname{Im} m_{2c} + \Psi(z)}{N\eta}\right]^{3/2}\right) = \mathrm{O}_{\prec}\left(N^{-1+3\delta}\right)$$

after averaging over $n^{-2p}\sum_{i_1,\cdots,i_p,j_1,\cdots,j_p}$, where we used that $E \in I_\delta$, $\eta := N^{-2/3-\delta}$, (2.4.11) and (2.4.18). Applying Lemma 2.4.2 (iii), we conclude (2.8.70). □

CHAPTER 3

Convergence of eigenvector empirical spectral distribution

3.1 Introduction and main result

In this chapter, we continue to consider separable covariance matrices $\mathcal{Q}_1$ and $\mathcal{Q}_2$ satisfying the Assumption 2.2.1. We are interested in the eigenvector statistics of $\mathcal{Q}_1$ and $\mathcal{Q}_2$, which will be (partially) characterized by the so-called *eigenvector empirical spectral distribution* (VESD). We now give a brief introduction of VESD and its application in high-dimensional statistics.

3.1.1 Eigenvector empirical spectral distribution

In applications of spectral analysis of large dimensional random matrices, one important problem is the convergence rate of the empirical spectral distributions (ESD). For the simplest sample covariance matrix with $A = B = I$ (i.e. the null case), it is well-known that the ESD $F^{(n)}_{XX^*}$ of XX^* converges weakly to the Marčenko-Pastur (MP) law F_{MP} [66]. One way to measure the convergence rate of the ESD is to use the Kolmogorov distance

$$\|F^{(n)}_{XX^*} - F_{MP}\| := \sup_x |F^{(n)}_{XX^*}(x) - F_{MP}(x)|.$$

The convergence rate for sample covariance matrices was first established in [5], and later improved in [48] to $\mathrm{O}(n^{-1/2})$ in probability under the finite 8th moment condition. In [79], the authors proved an almost optimal bound that $\|F^{(n)}_{XX^*} - F_{MP}\| = \mathrm{O}_{\prec}(n^{-1})$ under the sub-exponential decay assumption.

The research on the asymptotic properties of eigenvectors of large dimensional random matrices is generally harder and much less developed. However, the eigenvectors play an important role in high dimensional statistics. In particular, the principal component analysis (PCA) is now favorably recognized as a powerful technique for dimensionality reduction, and the eigenvectors corresponding to the largest eigenvalues are the directions of the principal components. The earlier work on the properties of eigenvectors goes back to Anderson [3], where the author proved that the eigenvectors of the Wishart matrix are asymptotically normal and isotropic when n is fixed and $N \to \infty$. For the high dimensional case, Johnstone [54] proposed the spiked model to test the existence of principal components. Then Paul [76] studied the directions of eigenvectors corresponding to spiked eigenvalues. In [65], Ma proposed an iterative thresholding approach to estimate sparse principal subspaces in the setting of a high-dimensional spiked covariance model. Using a reduction scheme which reduces the sparse PCA problem to a high-dimensional multivariate regression problem, [21] established the optimal rates of convergence for estimating the principal subspace for a large class of spiked covariance matrices. One can see the references in [21, 65] for more literatures on sparse PCA and spiked covariance matrices.

For the test of the existence of spiked eigenvalues, we first need to study the properties of the eigenmatrices in the null case. If $A = B = I$, then the eigenmatrix is expected to be asymptotically Haar distributed (i.e. uniformly distributed over the unitary group). However, formulating the terminology "asymptotically Haar distributed" is far from trivial since the dimension n is increasing. Following the approach in [7, 86, 88, 106, 107], we will use the VESD to characterize the asymptotical Haar property. Suppose

$$A^{1/2}XB^{1/2} = \sum_{1\leqslant k\leqslant N\wedge n} \sqrt{\lambda_k}\xi_k\zeta_k^* \tag{3.1.1}$$

is a singular value decomposition, where

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_{N\wedge n} \geqslant 0 = \lambda_{N\wedge n+1} = \ldots = \lambda_{N\vee n},$$

$\{\xi_k\}_{k=1}^n$ are the left-singular vectors, and $\{\zeta_k\}_{k=1}^N$ are the right-singular vectors. Then for deterministic unit vectors $\mathbf{u} \in \mathbb{C}^n$ and $\mathbf{v} \in \mathbb{C}^N$, we define the VESD of $\mathcal{Q}_{1,2}$ as

$$F_{\mathcal{Q}_1,\mathbf{u}}^{(n)}(x) = \sum_{k=1}^n |\langle \xi_k, \mathbf{u}\rangle|^2 \mathbf{1}_{\{\lambda_k\leqslant x\}}, \quad F_{\mathcal{Q}_2,\mathbf{v}}^{(N)}(x) = \sum_{k=1}^N |\langle \zeta_k, \mathbf{v}\rangle|^2 \mathbf{1}_{\{\lambda_k\leqslant x\}}. \tag{3.1.2}$$

Now we apply the above formulations to the null case. Adopting the ideas of [86, 88], we define the stochastic process as

$$X_{n,\mathbf{u}}(t) := \sqrt{\frac{n}{2}} \sum_{k=1}^{\lfloor nt \rfloor} \left(|\langle \xi_k, \mathbf{u} \rangle|^2 - n^{-1} \right).$$

If the eigenmatrix of XX^* is Haar distributed, then the vector $\mathbf{y} := (\langle \xi_k, \mathbf{u} \rangle)_{k=1}^n$ is uniformly distributed over the unit sphere, and $X_{n,\mathbf{u}}(t)$ would converge to a Brownian bridge by Donsker's theorem. Thus the convergence of $X_{n,\mathbf{u}}$ to a Brownian bridge characterizes the asymptotical Haar property of the eigenmatrix. For convenience, we can consider the time transformation

$$X_{n,\mathbf{u}}(F_{XX^*}^{(n)}(x)) = \sqrt{\frac{n}{2}} \left(F_{XX^*,\mathbf{u}}^{(n)}(x) - F_{XX^*}^{(n)}(x) \right).$$

Thus the problem is reduced to the study of the difference between the VESD and the ESD. It was already proved in [7, 14] that $F_{XX^*,\mathbf{u}}^{(n)}$ also converges weakly to the MP law for any sequence of unit vectors $\mathbf{u} \in \mathbb{R}^M$. On the other hand, compared with ESD, much less has been known about the convergence rate of the VESD. The best result before was obtained in [107], where the authors proved that if $d_N < 1$ and the entries of X are *i.i.d.* centered random variables, then $\|\mathbb{E}F_{XX^*,\mathbf{u}}^{(n)} - F_{MP}\| = \mathrm{O}(n^{-1/2})$ under the finite 10th moment assumption, and $\|F_{XX^*,\mathbf{u}}^{(n)} - F_{MP}\| = \mathrm{O}(n^{-1/4+\varepsilon})$ almost surely under the finite 8th moment assumption. However, we find that both of these bounds are far away from being optimal, and can be improved with a different method. This is one of the purposes of this paper.

We will also extend the above formulation to include separable covariance matrices with general *diagonal* covariance matrices A and B. In the general case, the eigenmatrix of $\mathcal{Q}_1$ is not asymptotically Haar distributed anymore. For its distribution, we conjecture that the eigenvectors of $\mathcal{Q}_1$ are asymptotically independent, and each ξ_k is asymptotically normal with covariance matrix given by some $\mathbf{D}_k$. In fact, our results in this paper suggest that $\mathbf{D}_k$ takes the form $\mathbf{F}_{1c}(\gamma_k) - \mathbf{F}_{1c}(\gamma_{k+1})$, where γ_k is defined in (2.4.14) to denote the classical location for λ_k, and $\mathbf{F}_{1c}$ is a matrix-valued function defined in (3.1.8) with the property that $\langle \mathbf{u}, \mathbf{F}_{1c}\mathbf{u} \rangle$ is the asymptotic distribution of the VESD $F_{\mathcal{Q}_1,\mathbf{u}}$ for any $\mathbf{u} \in \mathbb{C}^n$. Again, since the dimension n increases to infinity, the above property is hard to formulate. One way is to consider the finite-dimensional restriction in the following sense: given $m \in \mathbb{N}$, for any fixed unit vector $\mathbf{u} \in \mathbb{C}^n$ and $\{i_1, \cdots, i_m\} \subseteq \{1, \cdots, N \wedge n\}$, we should have asymptotically

$$(\langle \xi_{i_1}, \mathbf{u} \rangle, \cdots, \langle \xi_{i_m}, \mathbf{u} \rangle) \sim \mathcal{N}_m \left(0, \langle \mathbf{u}, \mathbf{D}_{i_1} \mathbf{u} \rangle, \ldots, \langle \mathbf{u}, \mathbf{D}_{i_m} \mathbf{u} \rangle \right). \tag{3.1.3}$$

(In fact, for nice choices of A, B in the sense of Definition 3.1.1, $\langle \mathbf{u}, \mathbf{D}_k \mathbf{u} \rangle$ is typically of order N^{-1}.) We can also adopt the approach as above, that is to investigate the stochastic process

$$X^{A,B}_{M,\mathbf{u}}(t) := \sqrt{\frac{n}{2}} \sum_{k=1}^{\lfloor nt \rfloor} \left(|\langle \xi_k, \mathbf{u} \rangle|^2 - \langle \mathbf{u}, \mathbf{D}_k \mathbf{u} \rangle \right). \tag{3.1.4}$$

If $n < N$, we conjecture that $X^{A,B}_{M,\mathbf{u}}(t)$ converges to the following Gaussian process for $0 \leqslant t \leqslant 1$:

$$\mathbf{B}^{A,B}_{\mathbf{u}}(t) := \int_0^t \langle \mathbf{u}, \mathbf{F}_{1c} \mathbf{u} \rangle \circ F_{1c}^{-1} \mathrm{d}B_t \quad \text{conditioning on} \quad \mathbf{B}^{A,B}_{\mathbf{u}}(1) = 0, \tag{3.1.5}$$

where B_t is a standard Brownian motion, F_{1c} is the asymptotic ESD of $\mathcal{Q}_1$ defined as the cumulative distribution function of ρ_{1c} in (2.2.16), and F_{1c}^{-1} denotes the quantile function. As before, we can study the process (3.1.4) through the time transformaton $X^{A,B}_{n,\mathbf{u}}(F_{\mathcal{Q}_1}(x))$, where $F_{\mathcal{Q}_1}$ is the ESD of $\mathcal{Q}_1$. Due to the rigidity of eigenvalues (see Theorem 3.2.4), we have for all x,

$$\sqrt{\frac{2}{M}} X^{A,B}_{n,\mathbf{u}}(F_{\mathcal{Q}_1}(x)) = F_{\mathcal{Q}_1,\mathbf{u}}(x) - \langle \mathbf{u}, \mathbf{F}_{1c}(x) \mathbf{u} \rangle + \mathrm{O}_{\prec}(n^{-1}).$$

Thus we need to study the convergence rate of $F_{\mathcal{Q}_1,\mathbf{u}}$ to $\langle \mathbf{u}, \mathbf{F}_{1c} \mathbf{u} \rangle$, and this is our main goal. In fact, we will prove that the convergence rate of $\mathbb{E} F_{\mathcal{Q}_1,\mathbf{u}}$ is $\mathrm{O}_{\prec}(n^{-1})$, which shows that the limiting process is centered, and the convergence rate of $F_{\mathcal{Q}_1,\mathbf{u}}$ is $\mathrm{O}_{\prec}(n^{-1/2})$, which partially verify the $\sqrt{n}$ scaling.

We remark that great progress has been made in other directions of the research on eigenvector statistics. For example, one can refer to [14, 35] for the delocalization and isotropic delocalization of eigenvectors, [57, 94] for the universality of eigenvectors, [17] for the local quantum unique ergodicity of eigenvectors and [15] for the eigenvectors of principal components. Note that some of these results are proved for Wigner matrices, but their generalizations to separable (or sample) covariance matrices usually are straightforward.

3.1.2 Main result

We consider separable covariance matrices $\mathcal{Q}_1$ and $\mathcal{Q}_2$ satisfying the Assumption 2.2.1, where we made one more assumption, i.e. both A and B are diagonal. Following the notations in (2.2.3), we shall call them Σ and $\widetilde{\Sigma}$ instead. To establish our main result, we need to make some extra assumptions on π_Σ and $\pi_{\widetilde{\Sigma}}$ defined in (2.2.4), which takes the form of the following regularity conditions. Recall the notations in Lemma 2.2.5.

Definition 3.1.1 (Regularity). *(i) Fix a (small) constant $\tau > 0$. We say that the edge a_k, $k = 1, \dots, 2p$, is τ-regular if*

$$a_k \geqslant \tau, \quad \min_{l \neq k} |a_k - a_l| \geqslant \tau, \quad \min_i |1 + m_{1c}(a_k)\widetilde{\sigma}_i| \geqslant \tau, \quad \min_i |1 + m_{2c}(a_k)\sigma_i| \geqslant \tau. \quad (3.1.6)$$

(ii) We say that the bulk components $[a_{2k}, a_{2k-1}]$ is regular if for any fixed $\tau' > 0$ there exists a constant $c \equiv c_{\tau'} > 0$ such that the densities ρ_{1c} and $\rho_{2,c}$) in $[a_{2k} + \tau', a_{2k-1} - \tau']$ are bounded from below by c.

Remark 3.1.2. The edge regularity conditions in (3.1.6) is an extension of the ones in (2.2.19). They ensure a regular square-root behavior of $\rho_{1,2c}$ near a_k for any k (instead of the rightmost edge only as given by (2.2.19)). The bulk regularity condition (ii) was introduced in [59], and it imposes a lower bound on the density of eigenvalues away from the edges. These conditions are satisfied by quite general classes of A and B; see e.g. [59, Examples 2.8 and 2.9].

For any $\mathbf{u} \in \mathbb{C}^n$ and $z \in \mathbb{C}_+$, we define

$$m_{1c,\mathbf{u}}(z) := -\langle \mathbf{u}, z^{-1}(1 + m_{2c}(z)\Sigma)^{-1}\mathbf{u} \rangle. \quad (3.1.7)$$

Then $m_{1c,\mathbf{u}}$ is the Stieltjes transform of a distribution, which we shall denote by $F_{1c,\mathbf{u}}$. Moreover, we denote the density of $F_{1c,\mathbf{u}}$ as $\rho_{1c,\mathbf{u}}$. From (3.1.7), it is easy to see that there exists a matrix-valued function $\mathbf{F}_{1c}$ depending on Σ such that $F_{1c,\mathbf{u}} = \langle \mathbf{u}, \mathbf{F}_{1c}\mathbf{u} \rangle$, i.e., we have

$$m_{1c,\mathbf{u}}(z) = \int_{\mathbb{R}} \frac{\mathrm{d}F_{1c,\mathbf{u}}(x)}{x - z} = \langle \mathbf{u}, \int_{\mathbb{R}} \frac{\mathrm{d}\mathbf{F}_{1c}(x)}{x - z} \mathbf{u} \rangle. \quad (3.1.8)$$

Now we are ready to state our main results, i.e. Theorem 3.1.4. We first state the main assumptions.

Assumption 3.1.3. *Fix a (small) constant $\tau > 0$.*

(i) $X = (x_{ij})$ is an $n \times N$ real or complex matrix whose entries are independent random variables that satisfy the following moment conditions: there exist constants $C_0, c_0 > 0$ such that for all $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant N$,

$$|\mathbb{E}x_{ij}| \leqslant C_0 N^{-2-c_0}, \tag{3.1.9}$$

$$\left|\mathbb{E}|x_{ij}|^2 - N^{-1}\right| \leqslant C_0 N^{-2-c_0}, \tag{3.1.10}$$

$$\left|\mathbb{E}x_{ij}^2\right| \leqslant C_0 N^{-2-c_0}, \quad \text{if } x_{ij} \text{ is complex}, \tag{3.1.11}$$

$$\mathbb{E}|x_{ij}|^4 \leqslant C_0 N^{-2}. \tag{3.1.12}$$

Note that (3.1.9)-(3.1.11) are slightly more general than (2.2.1).

(ii) $\tau \leqslant d_N \leqslant \tau^{-1}$ and $|d_N - 1| \geqslant \tau$.

(iii) $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ and $\widetilde{\Sigma} = \operatorname{diag}(\widetilde{\sigma}_1, \widetilde{\sigma}_2, \ldots, \widetilde{\sigma}_N)$ are deterministic positive-definite matrices. We assume that (2.2.5) holds, all the edges of $\rho_{1,2c}$ are τ-regular, and all the bulk components of $\rho_{1,2c}$ are regular in the sense of Definition 3.1.1.

Theorem 3.1.4. *Suppose d_N, X and Σ satisfy the Assumption 3.1.3. Suppose there exist constants $C_1, \phi > 0$ such that*

$$\max_{1\leqslant i\leqslant n, 1\leqslant j\leqslant N} |x_{ij}| \leqslant C_1 N^{-\phi}. \tag{3.1.13}$$

Let $\mathbf{u} \equiv \mathbf{u}_n \in \mathbb{C}^n$ denote a sequence of deterministic unit vectors. Then for any fixed (small) $\varepsilon > 0$ and (large) $D > 0$, we have

$$\|\mathbb{E}F^{(n)}_{\mathcal{Q}_1,\mathbf{u}} - F^{(n)}_{1c,\mathbf{u}}\| \leqslant N^{-1+\varepsilon} \tag{3.1.14}$$

for sufficiently large N, and for $\mathfrak{a} := \min(2\phi, 1/2)$,

$$\mathbb{P}\left(\|F^{(n)}_{\mathcal{Q}_1,\mathbf{u}} - F^{(n)}_{1c,\mathbf{u}}\| \geqslant N^{-\mathfrak{a}+\varepsilon}\right) \leqslant N^{-D}. \tag{3.1.15}$$

As an immediate corollary of Theorem 3.1.4, we have the following result.

Corollary 3.1.5. *Suppose d_N and Σ satisfy the Assumption 3.1.3. Let $X = (x_{ij})$ be an $n \times N$ random matrix whose entries are independent and satisfy*

$$\mathbb{E}x_{ij} = 0, \quad \mathbb{E}|x_{ij}|^2 = N^{-1}, \quad 1 \leqslant i \leqslant n, \ 1 \leqslant j \leqslant N. \tag{3.1.16}$$

If the entries of X are complex, then we assume in addition that

$$\mathbb{E}x_{ij}^2 = 0, \quad 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant N. \tag{3.1.17}$$

Suppose there exist constants $a, A > 0$ such that

$$\limsup_{s\to\infty} s^a \max_{i,j} \mathbb{P}\left(|\sqrt{N}x_{ij}| \geqslant s\right) \leqslant A \tag{3.1.18}$$

for all N. Let $\mathbf{u} \equiv \mathbf{u}_n \in \mathbb{C}^n$ denote a sequence of deterministic unit vectors. Then for any fixed $\varepsilon > 0$, if $a \geqslant 6$, we have

$$\|\mathbb{E}F^{(n)}_{\mathcal{Q}_1,\mathbf{u}} - F^{(n)}_{1c,\mathbf{u}}\| \leqslant N^{-1+\varepsilon} \tag{3.1.19}$$

for sufficiently large N; if $a \geqslant 8$, we have

$$\mathbb{P}\left(\limsup_{N\to\infty} N^{1/2-\varepsilon}\|F^{(n)}_{\mathcal{Q}_1,\mathbf{u}} - F^{(n)}_{1c,\mathbf{u}}\| \leqslant 1\right) = 1. \tag{3.1.20}$$

Proof of Corollary 3.1.5. We use a standard cutoff argument. We fix $a > 4$ and choose a constant $\phi > 0$ small enough such that $\left(N^{1/2-\phi}\right)^a \geqslant N^{2+\omega}$ for some constant $\omega > 0$. Then we introduce the following truncation

$$\tilde{X} := 1_\Omega X, \quad \Omega := \left\{|x_{ij}| \leqslant N^{-\phi} \text{ for all } 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant N\right\}.$$

By the tail condition (3.1.18), we have

$$\mathbb{P}(\tilde{X} \neq X) = O(N^{2-a/2+a\phi}). \tag{3.1.21}$$

Moreover, we have

$$\begin{aligned}
\mathbb{P}(\tilde{X} \neq X \text{ i.o.}) &= \lim_{k\to\infty} \mathbb{P}\left(\cup_{N=k}^\infty \cup_{i=1}^n \cup_{j=1}^N \left\{|x_{ij}| \geqslant N^{-\phi}\right\}\right) \\
&= \lim_{k\to\infty} \mathbb{P}\left(\cup_{t=k}^\infty \cup_{N\in[2^t,2^{t+1})} \cup_{i=1}^n \cup_{j=1}^N \left\{|x_{ij}| \geqslant N^{-\phi}\right\}\right) \\
&\leqslant C \lim_{k\to\infty} \sum_{t=k}^\infty \left(2^{t+1}\right)^2 \left(2^{t(1/2-\phi)}\right)^{-a} \leqslant C \lim_{k\to\infty} \sum_{t=k}^\infty 2^{-\omega t} = 0,
\end{aligned} \tag{3.1.22}$$

i.e. $\tilde{X} = X$ almost surely as $N \to \infty$. Here in the above derivation, we regard $n = Nd_N$ as a function depending on N.

Using (3.1.18) and integration by parts, it is easy to verify that

$$\mathbb{E}\,|x_{ij}|\,1_{|x_{ij}|>N^{-\phi}} = \mathrm{O}(N^{-2-\omega/2}), \quad \mathbb{E}\,|x_{ij}|^2\,1_{|x_{ij}|>N^{-\phi}} = \mathrm{O}(N^{-2-\omega/2}),$$

which imply that

$$|\mathbb{E}\tilde{x}_{ij}| = \mathrm{O}(N^{-2-\omega/2}), \quad \mathbb{E}|\tilde{x}_{ij}|^2 = N^{-1} + \mathrm{O}(N^{-2-\omega/2}),$$

$$\left|\mathbb{E}\tilde{x}_{ij}^2\right| = \mathrm{O}(N^{-2-\omega/2}), \quad \text{if } x_{ij} \text{ is complex.}$$

Moreover, we trivially have

$$\mathbb{E}|\tilde{x}_{ij}|^4 \leqslant \mathbb{E}|x_{ij}|^4 = \mathrm{O}(N^{-2}).$$

Hence $\tilde{X}$ is a random matrix satisfying Assumption 3.1.3. Then using (3.1.14) and (3.1.21) with $a = 6$ and $\phi = \varepsilon/6$, we conclude (3.1.19); using (3.1.15) and (3.1.22) with $\phi = (1-\varepsilon)/4$ and $a = 8$, we conclude (3.1.20). □

Remark 3.1.6. By exchanging the roles of (n, Σ) and $(N, \widetilde{\Sigma})$, one can prove the same bounds (3.1.14), (3.1.15), (3.1.19) and (3.1.20) for $F_{\mathcal{Q}_2,\mathbf{v}}(x)$ for any deterministic vector $\mathbf{v} \in \mathbb{C}^N$.

Remark 3.1.7. The estimates (3.1.19) and (3.1.20) improve the bounds obtained in [107], and relax the assumptions on moments and Σ, $\widetilde{\Sigma}$ as well. The convergence rates in (3.1.19) and (3.1.20) are optimal up to an N^{ε} factor. In fact, it was proved in [7] that for an analytic function f,

$$\sqrt{N}\int f(x)\mathrm{d}\left(F_{\mathcal{Q}_1,\mathbf{u}}(x) - F_{1c,\mathbf{u}}(x)\right) \to \mathcal{N}(0, \sigma_{f,\mathbf{u}}), \tag{3.1.23}$$

where $\mathcal{N}(0, \sigma_{f,\mathbf{u}})$ denotes the Gaussian distribution with mean zero and variance $\sigma_{f,\mathbf{u}}$. This shows that the fluctuation of $F_{\mathcal{Q}_1,\mathbf{u}}(x)$ is of order $N^{-1/2}$ and suggests the bound in (3.1.20). Taking expectation of (3.1.23), one can see that the order of $|\mathbb{E}F_{\mathcal{Q}_1,\mathbf{u}}(x) - F_{1c,\mathbf{u}}(x)|$ should be even smaller. Moreover, the fluctuation of eigenvalues on the microscopic scale will lead to an error of order at least N^{-1} by the universality of eigenvalues [12, 61, 79]. This shows that the bound (3.1.19) should be close to being optimal. We check the bounds (3.1.19) and (3.1.20) below with some numerical simulations; see Fig. 3.1.

Remark 3.1.8. In [107], the authors only handle the $n < N$ (i.e. $d_N < 1$) case for $\mathcal{Q}_1$, while our proof works for both the $d_N > 1$ and $d_N < 1$ cases. However, in the case with $d_N \to 1$, we will encounter some difficulties near the leftmost edge a_{2L}, which converges to 0 as $N \to \infty$ and violates the regularity condition (3.1.6). We will try to relax this assumption in the future.

Remark 3.1.9. In Theorem 3.1.4, we have assumed that $A = \Sigma$ and $B = \widetilde{\Sigma}$ are diagonal. But our results can be extended immediately to the case with general non-diagonal covariance matrices A and B for multivariate normal data as discussed in (2.1.3). For generally distributed data, under sufficiently strong moment assumptions, it is possible to prove the same results for the case with non-diagonal covariance matrices A and B. In particular, if the entries of $\sqrt{N}X$ have arbitrarily high moments, it can be proved that (3.1.19) and (3.1.20) hold for the VESD of $\mathcal{Q}_1$. The main inputs for the proof will include: (a) the anisotropic local law in Theorem 2.4.6, (b) Theorem 3.1.4 proved for the diagonal A, B case, (c) a self-consistent comparison argument as in Section 2.7 which extends Theorem 3.1.4 to the non-diagonal case through comparison with the diagonal case, and (d) the Helffer-Sjöstrand arguments in Section **??**. However, under weaker moment assumptions as in Corollary 3.1.5, the proof will be much harder. The main issue will be that the error bounds in steps (a) and (c) are not sharp enough, which does not give the optimal convergence rates as in (3.1.19) and (3.1.20). We would like to deal with this problem in the future, and focus on proving a sharp bound for the convergence rate of VESD in the diagonal $\mathbf{C}$ case in this thesis.

Remark 3.1.10. As discussed above, the convergence of the stochastic process $X^{A,B}_{n,\mathbf{u}}$ defined in (3.1.4) to the Gaussian process $\mathbf{B}^{A,B}_{\mathbf{u}}$ in (3.1.5) is also a very important question, which is complementary to the results in Corollary 3.1.5. The convergence of $X^{I,I}_{n,\mathbf{u}}$ to the Brownian bridge was first proved in the null case $\Sigma = I$ and $\widetilde{\Sigma} = I$, for some special vectors of the form $\mathbf{u} = n^{-1/2}(\pm 1, \cdots, \pm 1)$ in [88]. The result was later extended to the case with a general fixed vector $\mathbf{u}$ in [7]. More precisely, it was proved in [7] that for any fixed vector $\mathbf{u}$ and analytic functions $g_1, \cdots, g_k$, the random vector

$$(\hat{X}_{n,\mathbf{u}}(g_1), \cdots, \hat{X}_{n,\mathbf{u}}(g_k)), \quad \hat{X}_{n,\mathbf{u}}(g_i) := \int g_i(x) \mathrm{d}X^{I}_{n,\mathbf{u}}(F_{\mathcal{Q}_1}(x)), \ 1 \leqslant i \leqslant k,$$

converges to a Gaussian vector with mean zero and certain covariance function. We expect that combining the method in [7] and the new tools in this paper, one can prove a similar convergence result for $X_{n,\mathbf{u}}^{A,B}$ in the case with general non-scalar A, B. This will be studied in a future paper.

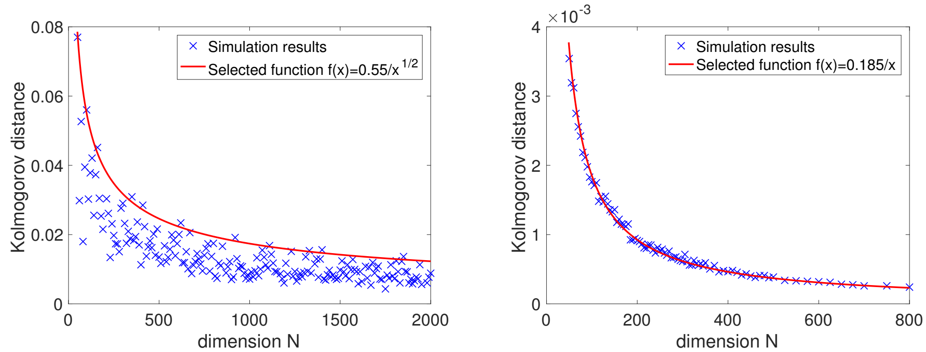
3.1.3 Simulations

In this subsection, we check the convergence rate of the (expected) VESD to the deformed MP law with some numerical simulations. The simulations are performed under the following setting: $\widetilde{\Sigma} = I$, i.e. we consider sample covariance matrices; $n = 2N$, i.e. $d_N = 2$; the entries $\sqrt{N}x_{ij}$ are drawn from a distribution ξ with mean zero, variance 1 and tail $\mathbb{P}(|\xi| \geqslant s) \sim s^{-6}$ for large s; the unit vector $\mathbf{v}$ is randomly chosen for each N. Note that for sample covariance matrices, $F_{\mathcal{Q}_2,\mathbf{v}}$ converges to the MP law F_{2c}, the cumulative distribution function of ρ_{2c} in (2.2.16), for any deterministic unit vector $\mathbf{v} \in \mathbb{C}^N$. In Fig. 3.1, we plot the Kolmogorov distances $\|F_{\mathcal{Q}_2,\mathbf{v}} - F_{2c}\|$ and $\|\mathbb{E}F_{\mathcal{Q}_2,\mathbf{v}} - F_{2c}\|$ for the following two choices of Σ: $\Sigma = I$ with ESD $\pi = \delta_1$, and

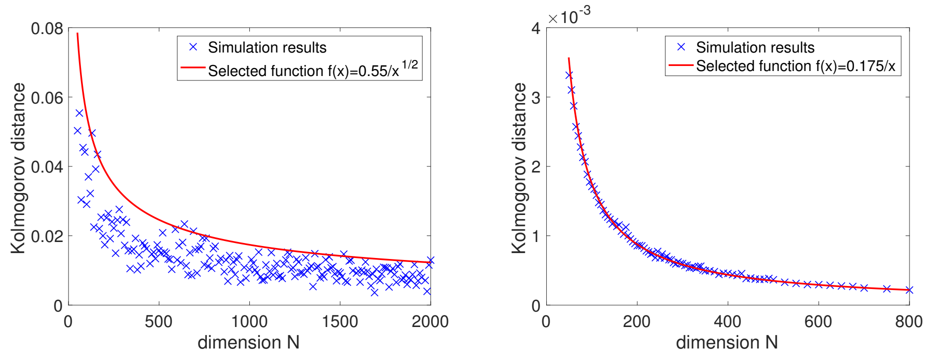
$$\Sigma = \operatorname{diag}(\underbrace{1, \cdots, 1}_{n/2}, \underbrace{4, \cdots, 4}_{n/2}), \quad \text{with ESD } \pi_\Sigma = 0.5\delta_1 + 0.5\delta_4. \tag{3.1.24}$$

For each N, we take an average over 10 repetitions to represent $F_{\mathcal{Q}_2,\mathbf{v}}^{(N)}$ and an average over $4N^2$ repetitions to approximate $\mathbb{E}F_{\mathcal{Q}_2,\mathbf{v}}^{(N)}$. Under each setting, we choose an appropriate function $f(x)$ to fit the simulation data. It is easy to observe that the convergence rate of the VESD is bounded by $\mathrm{O}(N^{-1/2})$, while the convergence rate of the expected VESD has order N^{-1}. This verifies the results in Corollary 3.1.5.

As discussed before, the convergence of $F_{\mathcal{Q}_2,\mathbf{v}}$ to F_{2c} for any sequence of deterministic unit vectors $\mathbf{v}$ can be used to characterize the asymptotical Haar property of the eigenmatrix of $\mathcal{Q}_2 = X^*\Sigma X$ (which also implies the asymptotical Haar property of the eigenmatrix of $\mathcal{Q}_1$ when $\Sigma = \sigma^2 I$). On the other hand, for a general Σ, the eigenmatrix of $\mathcal{Q}_1$ is not asymptotically Haar distributed anymore and the VESD of $\mathcal{Q}_1$ will depend on $\mathbf{v}$. Moreover, (3.1.7) gives an explicit dependence of $\mathbf{F}_{1c}$ on Σ, which should be of interest to statistical applications. In Fig. 3.2(a), we plot $F_{\mathcal{Q}_1,\mathbf{v}}$ for Σ in (3.1.24) and different choices of $\mathbf{v}_i$, $i = 1, 2, 3$. One can observe a transition of $F_{\mathcal{Q}_1,\mathbf{v}}$ when $\mathbf{v}$ changes from the direction corresponding



(a) $\pi = \delta_1$



(b) $\pi = 0.5\delta_1 + 0.5\delta_4$

Figure 3.1: The left figures of (a) and (b) plot $\|F^{(N)}_{\mathcal{Q}_2,\mathbf{v}} - F^{(N)}_{2c}\|$ as N increases from 50 to 2000, and we choose f to fit the upper envelope of the data. The right figures plot $\|\mathbb{E}F^{(N)}_{\mathcal{Q}_2,\mathbf{v}} - F^{(N)}_{2c}\|$ as N increases from 50 to 800.

to the smaller eigenvalues of Σ to the direction corresponding to the larger eigenvalues of Σ. In Fig. 3.2(b), we take $A = U\Sigma U^*$, where D is as in (3.1.24), U is a randomly chosen unitary matrix, and $\mathbf{w}_i = U\mathbf{v}_i$. One can see that even if A is non-diagonal, the convergence of the VESD of $\mathcal{Q}_1$ still holds (see Remark 3.1.9).

The rest of this chapter is organized as follows. We prove Theorem 3.1.4 in Section 3.2 using Stieltjes transforms. In the proof, we mainly use Theorems 3.2.1-3.2.3, which give the desired anisotropic local laws for the resolvents of $\mathcal{Q}_1$ and $\mathcal{Q}_2$. The proofs of Theorems 3.2.1-3.2.3 will be given in subsequent sections.

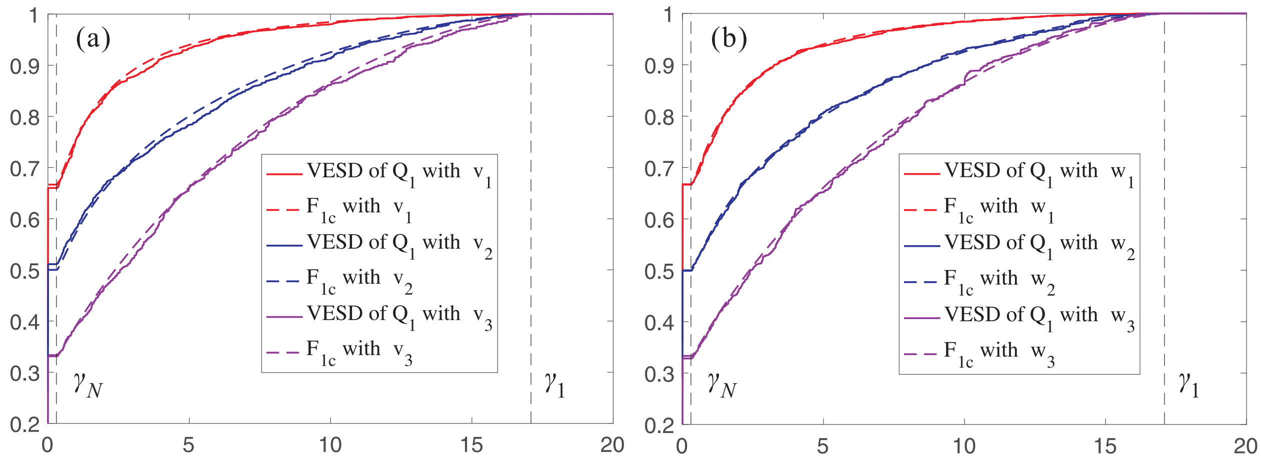


Figure 3.2: The plots for $F_{\mathcal{Q}_1,\mathbf{v}}(x)$ and $F_{1c,\mathbf{v}}(x)$ with $N = 2000$, $n = 2N$ and under the settings in Fig. 3.1. We take $\mathbf{v}_1 = \sqrt{2/n}(\underbrace{1,\cdots,1}_{n/2},0,\cdots,0)$, $\mathbf{v}_3 = \sqrt{2/n}(0,\cdots,0,\underbrace{1,\cdots,1}_{n/2})$, $\mathbf{v}_2 = (\mathbf{v}_1+\mathbf{v}_3)/\sqrt{2}$, and $\mathbf{w}_i = U\mathbf{v}_i$, $i = 1,2,3$. The dashed lines mark the places of the left edge γ_N and the right edge γ_1 of the spectrum (recall (2.4.14)).

3.2 Proof of the main result

For definiteness, we will focus on *real* sample covariance matrices during the proof. However, our proof also applies, after minor changes, to the *complex* case if we include the extra assumption (3.1.17) or (3.1.11).

Note that the Stieltjes transforms of $F^{(n)}_{\mathcal{Q}_1,\mathbf{u}}$ and $F^{(N)}_{\mathcal{Q}_2,\mathbf{v}}$ are equal to $\langle \mathbf{u}, \mathcal{G}_1(X,z)\mathbf{u}\rangle$ and $\langle \mathbf{v}, \mathcal{G}_2(X,z)\mathbf{v}\rangle$, respectively. We first state the local laws on $\langle \mathbf{u}, \mathcal{G}_1\mathbf{u}\rangle$ and $\langle \mathbf{v}, \mathcal{G}_2\mathbf{v}\rangle$, which will be used to prove Theorem 3.1.4. Recall the notations in Section 2.4. In the following proof, we will always assume that z lies in the spectral domain

$$\mathbf{D}(\omega, N) := \{z \in \mathbb{C}_+ : \omega \leqslant E \leqslant 2\lambda_r, N^{-1+\omega} \leqslant \eta \leqslant \omega^{-1}\}, \tag{3.2.1}$$

for some small constant $\omega > 0$, unless otherwise indicated. Recall the condition (3.1.6), we can take ω to be sufficiently small such that $\omega \leqslant \gamma_K/2$. Define the distance to the spectral edges as $\kappa := \min_{1\leqslant k\leqslant 2p} |E - a_k|$. Then we have the following estimates for $m_{1,2c}$:

$$|m_{1,2c}(z)| \sim 1, \quad \operatorname{Im} m_{1,2c}(z) \sim \begin{cases} \eta/\sqrt{\kappa+\eta}, & \text{if } E \notin \operatorname{supp}\rho_{1,2c} \\ \sqrt{\kappa+\eta}, & \text{if } E \in \operatorname{supp}\rho_{1,2c} \end{cases}, \tag{3.2.2}$$

and

$$\max_{i\in\mathcal{I}_1} |(1+m_{2c}(z)\sigma_i)^{-1}| + \max_{\mu\in\mathcal{I}_2} |(1+m_{1c}(z)\widetilde{\sigma}_\mu)^{-1}| = \mathrm{O}(1). \tag{3.2.3}$$

for $z \in \mathbf{D}$. Their proof is the same as in (2.2.6) using the regularity condition (3.1.6).

Theorem 3.2.1 (Local MP law)**.** *Suppose d_N, X, Σ and $\widetilde{\Sigma}$ satisfy the Assumption 3.1.3. Suppose X is real and satisfies (2.4.1) with $q \leqslant N^{-\phi}$ for some constant $\phi > 0$. Then the following estimates hold for $z \in \mathbf{D}$:*

(1) the averaged local law:

$$|m(X,z) - m_c(z)| + |m_{1,2}(X,z) - m_{1,2c}(z)| \prec (N\eta)^{-1}; \tag{3.2.4}$$

(2) the anisotropic local law: for deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$,

$$|\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec q + \Psi(z); \tag{3.2.5}$$

(3) for deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$ or $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}$,

$$|\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec q^2 + (N\eta)^{-1/2}. \tag{3.2.6}$$

All of the above estimates are uniform in the spectral parameter z and the deterministic vectors $\mathbf{u}, \mathbf{v}$.

The proof for Theorem 3.2.1 will be given in Section 3.4. Here we make some brief comments on it. If we assume (3.1.16) (instead of (3.1.9) and (3.1.10)), then (3.2.4) and (3.2.5) have already been proved in Theorem 2.4.6. Note that since Σ and $\widetilde{\Sigma}$ are diagonal, there is no need to use the domain in (2.4.21). Moreover, extending the domain $S(c_0, C_0, \varepsilon)$ in Theorem 2.4.6 to $\mathbf{D}(\omega, N)$ here does not change the proof, because only the estimates (3.2.2) and (3.2.3) are relevant for the proof. The main novelty of this theorem is the bound (3.2.6), which will be the main focus of our proof. Finally, if the variance assumption in

(3.1.16) is relaxed to the one in (3.1.10), we can still use the arguments for Theorem 2.4.6 to get the desired estimates (3.2.4)-(3.2.6). In fact, it is easy to check that the $\mathrm{O}(N^{-2-c_0})$ term leads to a negligible error at each step, and the whole proof remains unchanged. The relaxation of the mean zero assumption in (3.1.16) to the assumption (3.1.9) can be handled with the centralization Lemma 3.3.2.

After taking expectation, we have the following crucial improvement from (3.2.6) to (3.2.7), which is the main reason why we can improve the bound in [107] to the almost optimal one in (3.1.14). In fact, the leading order terms of $(\langle \mathbf{u}, \mathcal{G}_1 \mathbf{u}\rangle - m_{1c,\mathbf{u}})$ vanish after taking expectation, and hence leads to a bound that is one order smaller than the one in (3.2.6). The proof of Theorem 3.2.2 constitutes the main novelty of this chapter, and will be given in Section 3.3.

Theorem 3.2.2. *Suppose the assumptions in Theorem 3.2.1 hold. Then we have*

$$|\mathbb{E}\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec q^4 + (N\eta)^{-1} \tag{3.2.7}$$

uniformly in $z \in \mathbf{D}$ *and deterministic unit vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$ *or* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}$.

If $q = N^{-1/4}$, then (3.2.6) and (3.2.7) already give that

$$|\langle \mathbf{u}, \mathcal{G}_1 \mathbf{u}\rangle - m_{1c,\mathbf{u}}| \prec (N\eta)^{-1/2}, \quad |\mathbb{E}\langle \mathbf{u}, \mathcal{G}_1 \mathbf{u}\rangle - m_{1c,\mathbf{u}}| \prec (N\eta)^{-1},$$

which are sufficient to conclude Theorem 3.1.4. However, we find that the second bound on the expected VESD is still valid under a much weaker support assumption. More specifically, we have the following theorem, whose proof uses the Lindeberg replacement strategy, and is very similar to (actually, simpler than) the one for Theorem 2.4.8 and Lemma 2.4.7. Hence we shall omit the details and refer the reader to the supplementary material of [104].

Theorem 3.2.3. *Suppose the assumptions in Theorem 3.2.1 hold. Then we have*

$$|\mathbb{E}\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle| \prec (N\eta)^{-1}, \tag{3.2.8}$$

uniformly in $z \in \mathbf{D}$ *and deterministic unit vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$ *or* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}$.

As a corollary of (3.2.4), we have the following rigidity result for the eigenvalues. Its proof is the same as the ones for (2.4.30) and [59, Theorem 3.12]. So we omit the details. Recall the classical locations of eigenvalues γ_j defined in (2.4.14).

Theorem 3.2.4 (Rigidity of eigenvalues)**.** *Suppose Theorem 3.2.1 and the regularity condition (3.1.6) hold. Then we have*

$$|\lambda_1 - \gamma_1| + |\lambda_K - \gamma_K| \prec N^{-2/3}, \tag{3.2.9}$$

where $K := n \wedge N$.

In the rest of this subsection, we finish the proof of Theorem 3.1.4 using Theorems 3.2.1-3.2.4. The following arguments have been used previously to control the Kolmogorov distance between the ESD of a random matrix and the limiting law. For example, the reader can refer to [37, Lemma 6.1] and [79, Lemma 8.1]. By the remark below (3.2.1), we can choose the constant $\omega > 0$ such that $\gamma_K/2 > \omega$.

Proof of (3.1.14). The key inputs are the bounds (3.2.8) and (3.2.9). Suppose $\langle \mathbf{u}, \mathcal{G}_1(X,z)\mathbf{u}\rangle$ is the Stieltjes transform of $\hat{\rho}_{\mathbf{u}}$. Then we define

$$\hat{n}_{\mathbf{u}}(E) := \int \mathbf{1}_{[0,E]}(x)\hat{\rho}_{\mathbf{u}}dx, \quad n_c(E) := \int \mathbf{1}_{[0,E]}(x)\rho_{1c,\mathbf{u}}dx, \tag{3.2.10}$$

and $\rho_{\mathbf{u}} := \mathbb{E}\hat{\rho}_{\mathbf{u}}$, $n_{\mathbf{u}} := \mathbb{E}\hat{n}_{\mathbf{u}}$. Hence we would like to bound

$$\|\mathbb{E}F_{\mathcal{Q}_1,\mathbf{u}} - F_{1c,\mathbf{u}}\| = \sup_E |n_{\mathbf{u}}(E) - n_c(E)|.$$

For simplicity, we denote $\Delta\rho := \rho_{\mathbf{u}} - \rho_{2c}$ and its Stieltjes transform by

$$\Delta m(z) := \mathbb{E}\langle \mathbf{u}, \mathcal{G}_1(X,z)\mathbf{u}\rangle - m_{1c,\mathbf{u}}(z).$$

Let $\chi(y)$ be a smooth cutoff function with support in $[-1,1]$, with $\chi(y) = 1$ for $|y| \leqslant 1/2$ and with bounded derivatives. Fix $\eta_0 = N^{-1+\omega}$ and $3\gamma_K/4 \leqslant E_1 < E_2 \leqslant 3\gamma_1/2$. Let $f \equiv f_{E_1,E_2,\eta_0}$ be a smooth function supported in $[E_1 - \eta_0, E_2 + \eta_0]$ such that $f(x) = 1$ if $x \in [E_1 + \eta_0, E_2 - \eta_0]$, and $|f'| \leqslant C\eta_0^{-1}$, $|f''| \leqslant C\eta_0^{-2}$ if $|x - E_i| \leqslant \eta_0$. Using the Helffer-Sjöstrand calculus (see e.g. [24]), we have

$$f(E) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \frac{\mathrm{i}yf''(x)\chi(y) + \mathrm{i}(f(x) + \mathrm{i}yf'(x))\chi'(y)}{E - x - \mathrm{i}y}\mathrm{d}x\mathrm{d}y.$$

Then we obtain that

$$
\begin{aligned}
&\left|\int f(E)\Delta\rho(E)dE\right| \\
&\leqslant C\int_{\mathbb{R}^2}\left(|f(x)|+|y||f'(x)|\right)|\chi'(y)||\Delta m(x+\mathrm{i}y)|\mathrm{d}x\mathrm{d}y && (3.2.11)\\
&+C\sum_i\left|\int_{|y|\leqslant\eta_0}\int_{|x-E_i|\leqslant\eta_0} yf''(x)\chi(y)\mathrm{Im}\,\Delta m(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y\right| && (3.2.12)\\
&+C\sum_i\left|\int_{|y|\geqslant\eta_0}\int_{|x-E_i|\leqslant\eta_0} yf''(x)\chi(y)\mathrm{Im}\,\Delta m(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y\right|. && (3.2.13)
\end{aligned}
$$

By (3.2.8) with $\eta=\eta_0$, we have

$$\eta_0\mathrm{Im}\,\mathbb{E}\langle\mathbf{u},\mathcal{G}_1(X,E+\mathrm{i}\eta_0)\mathbf{u}\rangle\prec N^{-1+\omega}. \tag{3.2.14}$$

Since $\eta\mathrm{Im}\,\mathbb{E}\langle\mathbf{u},\mathcal{G}_1(X,E+\mathrm{i}\eta)\mathbf{u}\rangle$ and $\eta\mathrm{Im}\,m_{1c,\mathbf{u}}(E+\mathrm{i}\eta)$ are increasing with η, we obtain that

$$\eta\left|\mathrm{Im}\,\Delta m(E+\mathrm{i}\eta)\right|\prec N^{-1+\omega}\quad\text{for all } 0\leqslant\eta\leqslant\eta_0. \tag{3.2.15}$$

Moreover, since $G(X,z)^*=G(X,\bar{z})$, the estimates (3.2.8) and (3.2.15) also hold for $z\in\mathbb{C}_-$.

Now we bound the terms (3.2.11), (3.2.12) and (3.2.13). Using (3.2.8) and that the support of χ' is in $1\geqslant|y|\geqslant 1/2$, the term (3.2.11) can be bounded by

$$\int_{\mathbb{R}^2}\left(|f(x)|+|y||f'(x)|\right)|\chi'(y)||\Delta m(x+\mathrm{i}y)|\mathrm{d}x\mathrm{d}y\prec N^{-1}. \tag{3.2.16}$$

Using $|f''|\leqslant C\eta_0^{-2}$ and (3.2.15), we can bound the terms in (3.2.12) by

$$\left|\int_{|y|\leqslant\eta_0}\int_{|x-E_i|\leqslant\eta_0} yf''(x)\chi(y)\mathrm{Im}\,\Delta m(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y\right|\prec N^{-1+\omega}. \tag{3.2.17}$$

Finally, we integrate the term (3.2.13) by parts first in x, and then in y (and use the Cauchy-Riemann equation $\partial\mathrm{Im}(\Delta m)/\partial x=-\partial\mathrm{Re}(\Delta m)/\partial y$) to get

$$
\begin{aligned}
&\int_{y\geqslant\eta_0}\int_{|x-E_i|\leqslant\eta_0} yf''(x)\chi(y)\mathrm{Im}\,\Delta m(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y \\
&=-\int_{|x-E_i|\leqslant\eta_0}\eta_0\chi(\eta_0)f'(x)\mathrm{Re}\,\Delta m(x+\mathrm{i}\eta_0)dx && (3.2.18)\\
&\quad-\int_{y\geqslant\eta_0}\int_{|x-E_i|\leqslant\eta_0}\left(y\chi'(y)+\chi(y)\right)f'(x)\mathrm{Re}\,\Delta m(x+\mathrm{i}y)\mathrm{d}x\mathrm{d}y. && (3.2.19)
\end{aligned}
$$

We bound the term in (3.2.18) by $\mathrm{O}_\prec(N^{-1})$ using (3.2.8) and $|f'| \leqslant C\eta_0^{-1}$. The first term in (3.2.19) can be estimated by $\mathrm{O}_\prec(N^{-1})$ as in (3.2.16). For the second term in (3.2.19), we again use (3.2.8) and $|f'| \leqslant C\eta_0^{-1}$ to get that

$$\left|\iint_{y \geqslant \eta_0} \int_{|x-E_i| \leqslant \eta_0} \chi(y) f'(x) \mathrm{Re}\, \Delta m(x+\mathrm{i}y) \mathrm{d}x\mathrm{d}y\right| \prec \int_{\eta_0}^1 \frac{1}{Ny} dy \prec N^{-1}.$$

Combining the above estimates, we obtain that

$$\left|\iint_{y \geqslant \eta_0} \int_{|x-E_i| \leqslant \eta_0} y f''(x) \chi(y) \mathrm{Im}\, \Delta m(x+\mathrm{i}y) \mathrm{d}x\mathrm{d}y\right| \prec N^{-1}.$$

Obviously, the same estimate also holds for the $y \leqslant -\eta_0$ part. Together with (3.2.16) and (3.2.17), we conclude that

$$\left|\int f(E) \Delta\rho(E) dE\right| \prec N^{-1+\omega}. \tag{3.2.20}$$

For any interval $I := [E-\eta_0, E+\eta_0]$ with $E \in [\gamma_K/2, 2\gamma_1]$, we have

$$\begin{aligned} \hat{n}_{\mathbf{u}}(E+\eta_0) - \hat{n}_{\mathbf{u}}(E-\eta_0) &= \sum_{\lambda_k \in (E-\eta_0, E+\eta_0]} |\langle \xi_k, \mathbf{u}\rangle|^2 \\ \leqslant 2\eta_0 \sum_{k=1}^N \frac{|\langle \xi_k, \mathbf{u}\rangle|^2 \eta_0}{(\lambda_k - E)^2 + \eta_0^2} &= 2\eta_0 \mathrm{Im}\, \langle \mathbf{u}, \mathcal{G}_1(X, E+\mathrm{i}\eta_0)\mathbf{u}\rangle, \end{aligned} \tag{3.2.21}$$

where in the last step we used the spectral decomposition in (2.4.5). Then by (3.2.14), we get that

$$n_{\mathbf{u}}(E+\eta_0) - n_{\mathbf{u}}(E-\eta_0) \prec N^{-1+\omega}. \tag{3.2.22}$$

On the other hand, since $\rho_{1c,\mathbf{u}}$ is bounded, we trivially have

$$n_c(E+\eta_0) - n_c(E-\eta_0) \leqslant C\eta_0 = CN^{-1+\omega}. \tag{3.2.23}$$

Now we set $E_2 = 3\gamma_1/2$. With (3.2.20), (3.2.22) and (3.2.23), we get that for any $E \in [3\gamma_K/4, E_2]$,

$$|(n_{\mathbf{u}}(E_2) - n_{\mathbf{u}}(E)) - (n_c(E_2) - n_c(E))| \prec N^{-1+\omega}. \tag{3.2.24}$$

Note that by (3.2.9), the eigenvalues of $\mathcal{Q}_2$ are inside $\{0\} \cup [3\gamma_K/4, E_2]$ with high probability. Hence we have that with high probability,

$$\hat{n}_{\mathbf{u}}(E_2) = n_c(E_2) = 1, \quad \hat{n}_{\mathbf{u}}(3\gamma_K/4) = \hat{n}_{\mathbf{u}}(0). \tag{3.2.25}$$

Together with (3.2.24), we get that

$$\sup_{E\geqslant 0}|n_{\mathbf{u}}(E)-n_c(E)|\prec N^{-1+\omega}. \tag{3.2.26}$$

This concludes (3.1.14) since ω can be arbitrarily small. □

Proof of (3.1.15). The proof for (3.1.15) is similar except that we shall use the estimate (3.2.6) instead of (3.2.8). By (3.2.6), we have for any $\mathbf{u}\in\mathbb{C}^{\mathcal{I}_1}$,

$$|\langle \mathbf{u},\mathcal{G}_1(X,z)\mathbf{u}\rangle - m_{1c,\mathbf{u}}(z)|\prec N^{-2\phi}+(N\eta)^{-1/2} \tag{3.2.27}$$

uniformly in $z\in\mathbf{D}$. Then we would like to bound (recall (3.2.10))

$$\|F^{(n)}_{\mathcal{Q}_1,\mathbf{u}}-F_{1c,\mathbf{u}}\|=\sup_E|\hat{n}_{\mathbf{u}}(E)-n_c(E)|\,,$$

where $\hat{n}_{\mathbf{u}}$ is defined in (3.2.10). We denote

$$\Delta\hat{\rho}:=\hat{\rho}_{\mathbf{u}}-\rho_{1c,\mathbf{u}},\quad \Delta\hat{m}:=\langle \mathbf{u},\mathcal{G}_1(X,z)\mathbf{u}\rangle - m_{1c,\mathbf{u}}(z).$$

Then for f_{E_1,E_2,η_0} defined above, we can repeat the Helffer-Sjöstrand argument with the estimate (3.2.27) to get that

$$\sup_{E_1,E_2}\left|\int f_{E_1,E_2,\eta_0}(E)\Delta\hat{\rho}(E)dE\right|\prec N^{-2\phi}+N^{-1/2}, \tag{3.2.28}$$

which, together with (3.2.21) and (3.2.25), implies that

$$\sup_{E\geqslant 0}|\hat{n}_{\mathbf{u}}(E)-n_c(E)|\prec N^{-2\phi}+N^{-1/2}.$$

This concludes (3.1.15) by the Definition 2.4.1. □

3.3 Proof of Theorem 3.2.2

First, we record the following simple lemma. In fact, it has been already used in the previous proof in Chapter 2, but we state it here for reader's convenience.

Lemma 3.3.1. *Suppose $\tilde{\Phi}(z)$ is a deterministic function on $\mathbf{D}$ satisfying $N^{-1/2}\leqslant\tilde{\Phi}(z)\leqslant N^{-c}$ for some constant $c>0$. Suppose $|G_{ab}(z)-\Pi_{ab}(z)|\prec\tilde{\Phi}(z)$ uniformly in $a,b\in\mathcal{I}$ and $z\in\mathbf{D}$. Then for any $\mathbb{T}\subseteq\mathcal{I}$ with $|\mathbb{T}|=\mathrm{O}(1)$, we have uniformly in $z\in\mathbf{D}$,*

$$\max_{a,b\in\mathcal{I}\setminus\mathbb{T}}\left|G_{ab}(z)-G^{(\mathbb{T})}_{ab}(z)\right|\prec\tilde{\Phi}^2(z). \tag{3.3.1}$$

Proof. The bound (3.3.1) can be proved by repeatedly applying the first resolvent expansion in (2.6.8) with respect to the indices in $\mathbb{T}$. □

For X satisfying the assumptions in Theorem 3.2.1, we write $X = X_1 + B$, where $X_1 := X - \mathbb{E}X$ is a real random matrix satisfying (3.1.10), (3.1.12) and

$$\mathbb{E}(X_1)_{i\mu} = 0, \quad i \in \mathcal{I}_1, \ \mu \in \mathcal{I}_2, \tag{3.3.2}$$

and $B := \mathbb{E}X$ is a deterministic matrix such that

$$\max_{i,\mu} |B_{i\mu}| \leqslant C_0 N^{-2-c_0}. \tag{3.3.3}$$

The next lemma shows that $G(X,z)$ is very close to $G(X_1,z)$ in the sense of anisotropic local law. Its proof will be given in the supplementary material.

Lemma 3.3.2. *If (3.2.5) holds for $G(X_1,z)$, then we have*

$$|\langle \mathbf{u}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{u}, G(X_1,z)\mathbf{v}\rangle| \prec (N\eta)^{-1} \tag{3.3.4}$$

uniformly in $z \in \mathbf{D}$ and deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$.

Proof. For $z \in \mathbf{D}$, we have

$$G(X,z) := \begin{pmatrix} -I_{M\times M} & X_1 + B \\ X_1^* + B^* & -zI_{N\times N} \end{pmatrix}^{-1} = \left(G_1^{-1} + V\right)^{-1}, \tag{3.3.5}$$

where we abbreviate $G_1(z) := G(X_1,z)$ and $V := \begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}$. Then we expand G using the resolvent expansion

$$G = G_1 - G_1VG_1 + (G_1V)^2G_1 - (G_1V)^3G. \tag{3.3.6}$$

We need to estimate the last three terms of the right-hand side. First, note that by (2.7.2)-(2.7.5) and (3.2.5), we have for $z \in \mathbf{D}$,

$$\max\left\{\sum_i |(G_1)_{\mathbf{v}i}|^2, \sum_i |(G_1)_{i\mathbf{v}}|^2, \sum_\mu |(G_1)_{\mathbf{v}\mu}|^2, \sum_\mu |(G_1)_{\mu\mathbf{v}}|^2\right\} \prec \eta^{-1}, \tag{3.3.7}$$

for any $\mathbf{v} \in \mathbb{C}^{\mathcal{I}}$ and $\mathbb{T} \subseteq \mathcal{I}$ with $|\mathbb{T}| = \mathrm{O}(1)$.

For any unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$, we have

$$
\begin{aligned}
|\langle \mathbf{u}, G_1 V G_1 \mathbf{v}\rangle| &\leqslant \sum_{b\in\mathcal{I}} \Big| \sum_{a\in\mathcal{I}} (G_1)_{\mathbf{u}a} V_{ab} \Big| \, |(G_1)_{b\mathbf{v}}| \\
&\prec \max_b \Big(\sum_{a\in\mathcal{I}} |V_{ab}|^2 \Big)^{1/2} \sum_{b\in\mathcal{I}} |(G_1)_{b\mathbf{v}}| \\
&\prec N^{-1-c_0} \Big(\sum_{b\in\mathcal{I}} |(G_1)_{b\mathbf{v}}|^2 \Big)^{1/2} \prec N^{-1-c_0}\eta^{-1/2},
\end{aligned}
\tag{3.3.8}
$$

where in the second step we used (3.2.5) for G_1, in the third step the Cauchy-Schwarz inequality and (3.3.3), and in the last step (3.3.7). With a similar argument, we obtain that

$$
\left|\langle \mathbf{u}, (G_1 V)^2 G_1 \mathbf{v}\rangle\right| \prec N^{-2-2c_0}\eta^{-1}. \tag{3.3.9}
$$

Combining (3.3.9) with the rough bound (4.3.12) for G, we get that

$$
\begin{aligned}
\left|\langle \mathbf{u}, (G_1 V)^3 G \mathbf{v}\rangle\right| &= \Big| \sum_{a,b} \big((G_1 V)^2 G_1\big)_{\mathbf{u}a} V_{ab} G_{b\mathbf{v}} \Big| \\
&\prec \left(N^{-2-2c_0}\eta^{-1}\right)\eta^{-1} \sum_a \Big(\sum_b |V_{ab}|^2 \Big)^{1/2} \leqslant C N^{-3/2-3c_0}\eta^{-1},
\end{aligned}
\tag{3.3.10}
$$

where we used $\eta \geqslant N^{-1}$ for $z \in \mathbf{D}$ in the last step. Plugging the estimates (3.3.8)-(3.3.10) into (3.3.6), we conclude that

$$
|\langle \mathbf{u}, G\mathbf{v}\rangle - \langle \mathbf{u}, G_1 \mathbf{v}\rangle| \prec N^{-1-c_0}\eta^{-1/2} \leqslant (N\eta)^{-1}. \tag{3.3.11}
$$

for all deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$. □

3.3.1 Sketch of the proof

In this subsection, we start proving the resolvent estimate (3.2.7). For simplicity, we denote $\Phi := q^2 + (N\eta)^{-1/2}$. By Lemma 3.3.2, we can assume that the entries of X are centered without loss of generality. We will only prove (3.2.7) for $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}$, while the proof in the case of $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$ is exactly the same. Also by polarization, it suffices to prove the following estimate

$$
|\mathbb{E}\langle \mathbf{v}, \mathcal{G}_2(X,z)\mathbf{v}\rangle - \Pi_{\mathbf{v}\mathbf{v}}(z)| \prec q^4 + (N\eta)^{-1}, \quad \mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}. \tag{3.3.12}
$$

We can obtain the more general bound (3.2.7) by applying (3.3.12) to the vectors $\mathbf{u}+\mathbf{v}$ and $\mathbf{u}+i\mathbf{v}$, respectively. Note that (3.2.6) gives the a priori bound

$$\Big|\sum_{\mu,\nu}\bar{v}_\mu v_\nu\left[\mathbb{E}\left(\mathcal{G}_2\right)_{\mu\nu}-\Pi_{\mu\mu}\delta_{\mu\nu}\right]\Big|\prec\Phi.$$

We will show that after taking expectation, the leading order term in $\left(\mathcal{G}_2\right)_{\mu\nu}-\Pi_{\mu\mu}\delta_{\mu\nu}$ vanishes and leads to the better estimate (3.3.12). We deal with the diagonal and off-diagonal parts separately:

$$\sum_{\mu}|v_\mu|^2\left[\mathbb{E}(\mathcal{G}_2)_{\mu\mu}-\Pi_{\mu\mu}\right],\quad \sum_{\mu\neq\nu}\bar{v}_\mu v_\nu\mathbb{E}\left(\mathcal{G}_2\right)_{\mu\nu}.$$

For any $\mathbb{T}\subseteq\mathcal{I}$, recall the Z variables defined in (2.6.22):

$$Z_\mu^{(\mathbb{T})}:=(1-\mathbb{E}_\mu)(G^{(\mathbb{T})})^{-1}_{\mu\mu}=\frac{\widetilde{\sigma}_\mu}{N}\sum_{i\in\mathcal{I}_1}\sigma_i G_{ii}^{(\mathbb{T}\mu)}-(Y^*G^{(\mathbb{T}\mu)}Y)_{\mu\mu},\quad \mu\notin\mathbb{T},$$

Note that by (3.2.6), (3.3.1) (with $\tilde{\Phi}=q+\Psi$ by (3.2.5)), and Lemma 2.4.2, we have

$$Z_\mu^{(\mathbb{T})}:=(1-\mathbb{E}_\mu)\left[(G^{(\mathbb{T})})^{-1}_{\mu\mu}-m_{2c}^{-1}\right]\prec\Phi, \tag{3.3.13}$$

for any $\mathbb{T}\subseteq\mathcal{I}$ with $|\mathbb{T}|=\mathrm{O}(1)$. Then using (2.6.5) we get that

$$\begin{aligned}\mathbb{E}G_{\mu\mu}-\Pi_{\mu\mu}&=\mathbb{E}\frac{1}{-z-z\widetilde{\sigma}_\mu m_{1c}(z)-z\widetilde{\sigma}_\mu(m_1^{(\mu)}-m_{1c}(z))+Z_\mu}-\Pi_{\mu\mu}\\&=-\Pi_{\mu\mu}^2\mathbb{E}Z_\mu+\mathrm{O}_\prec\left(\Phi^2+(N\eta)^{-1}\right)=\mathrm{O}_\prec\left(\Phi^2\right),\end{aligned}$$

where in the second step we used (3.2.4), (3.3.1), (3.3.13), and (2.4.16). So we can bound the diagonal part by

$$\sum_{\mu}|v_\mu|^2\left[\mathbb{E}(\mathcal{G}_2)_{\mu\mu}-\Pi_{\mu\mu}\right]=\sum_\mu|v_\mu|^2[\mathbb{E}G_{\mu\mu}-\Pi_{\mu\mu}]\prec q^4+\frac{1}{N\eta}. \tag{3.3.14}$$

For the off-diagonal part, we claim that for $\mu\neq\nu\in\mathcal{I}_2$,

$$\left|\mathbb{E}\left(\mathcal{G}_2\right)_{\mu\nu}\right|\prec N^{-1}\Phi^2. \tag{3.3.15}$$

Then using (3.3.15) and $\|\mathbf{v}\|_1\leqslant\sqrt{N}$, we obtain that

$$\Big|\sum_{\mu\neq\nu}\bar{v}_\mu v_\nu\mathbb{E}\left(\mathcal{G}_2\right)_{\mu\nu}\Big|\prec\|\mathbf{v}\|_1^2N^{-1}\Phi^2\leqslant C\left(q^4+\frac{1}{N\eta}\right).$$

This concludes (3.3.12) together with (3.3.14).

To prove (3.3.15), we extend the arguments in [14, Section 5]. We illustrate the basic idea with some simplified calculations. Using the resolvent identities (2.6.6) and (2.6.8), we get

$$\begin{aligned} \mathbb{E}G_{\mu\nu} &= \mathbb{E}G_{\mu\mu}G_{\nu\nu}^{(\mu)}\left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu} \\ &= \mathbb{E}G_{\mu\mu}^{(\nu)}G_{\nu\nu}^{(\mu)}\left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu} + \mathbb{E}\frac{G_{\mu\nu}G_{\nu\mu}}{G_{\nu\nu}}G_{\nu\nu}^{(\mu)}\left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu}. \end{aligned} \tag{3.3.16}$$

We now focus on the first term. Applying (2.6.5) gives that

$$\begin{aligned} \mathbb{E}G_{\mu\mu}^{(\nu)}G_{\nu\nu}^{(\mu)}\left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu} &= \mathbb{E}\frac{\left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu}}{[-z-(Y^*G^{(\mu\nu)}Y)_{\mu\mu}]\,[-z-(XG^{(\mu\nu)}X^*)_{\nu\nu}]} \\ &= \mathbb{E}\frac{\left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu}}{\left(\Pi_{\mu\mu}^{-1}+\varepsilon_\mu\right)\left(\Pi_{\nu\nu}^{-1}+\varepsilon_\nu\right)}. \end{aligned} \tag{3.3.17}$$

where we have

$$\varepsilon_\mu := \frac{\widetilde{\sigma}_\mu}{N}\sum_{i\in\mathcal{I}_1}\sigma_i\Pi_{ii} - (Y^*G^{(\mu\nu)}Y)_{\mu\mu} = \frac{\widetilde{\sigma}_\mu}{N}\sum_{i\in\mathcal{I}_1}\sigma_i(\Pi_{ii}-G_{ii}^{(\mu\nu)}) + Z_\mu^{(\nu)} \prec \Phi \tag{3.3.18}$$

by (2.4.16), (3.2.4), (3.3.1) (with $\tilde{\Phi} = q + \Psi$) and (3.3.13). We now expand the fractions in (3.3.17) in order to take the expectation. Note that the $G^{(\mu\nu)}$ entries are independent of the X entries in the μ,ν-th rows and columns. Thus to attain a nonzero expectation, each X entry must appear at least twice in the expression. Due to this reason, the leading and next-to-leading order terms in the expansion vanish. The "real" leading order term is proportional to

$$\begin{aligned} &\mathbb{E}\varepsilon_\mu\varepsilon_\nu\left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu} = \mathbb{E}(Y^*G^{(\mu\nu)}Y)_{\mu\mu}(Y^*G^{(\mu\nu)}Y)_{\nu\nu}(Y^*G^{(\mu\nu)}Y)_{\mu\nu} \\ &\propto \sum_{\mu,\nu}\frac{C_{i,j}}{N^3}\mathbb{E}G_{ii}^{(\mu\nu)}G_{jj}^{(\mu\nu)}G_{ij}^{(\mu\nu)} = \sum_{i\neq j}\frac{C_{i,j}}{N^3}\Pi_{ii}\Pi_{jj}\mathbb{E}G_{ij}^{(\mu\nu)} + \mathrm{O}_\prec(N^{-1}\Phi^2), \end{aligned} \tag{3.3.19}$$

where the constants $C_{i,j}$ depend on σ_i, σ_j and the 3rd moments of $X_{i\mu}$ and $X_{j\mu}$ (recall (3.1.12)). Here in the last step, we used $|G_{ii}^{(\mu\nu)} - \Pi_{ii}| \prec \Phi$ (by (3.2.6) and (3.3.1)) and $|\Pi_{ii}| = \mathrm{O}(1)$ (by (3.2.3)), and bounded the $i = j$ terms by $\mathrm{O}_\prec(N^{-2}) = \mathrm{O}_\prec(N^{-1}\Phi^2)$. Now applying (2.6.6) to $G_{ij}^{(\mu\nu)}$, we get that

$$\begin{aligned} \mathbb{E}G_{ij}^{(\mu\nu)} &= \mathbb{E}G_{ii}^{(\mu\nu)}G_{jj}^{(i\mu\nu)}\left(YG^{(ij\mu\nu)}Y^*\right)_{ij} \\ &= \Pi_{ii}\Pi_{jj}\mathbb{E}\left(YG^{(ij\mu\nu)}Y^*\right)_{ij} + \mathrm{O}_\prec(\Phi^2) = \mathrm{O}_\prec(\Phi^2), \end{aligned} \tag{3.3.20}$$

where in the second step we used $|G_{ii}^{(\mu\nu)} - \Pi_{ii}| + |G_{jj}^{(i\mu\nu)} - \Pi_{jj}| \prec \Phi$ and

$$\left(YG^{(ij\mu\nu)}Y^*\right)_{ij} = G_{ij}^{(\mu\nu)} \left(G_{ii}^{(\mu\nu)} G_{jj}^{(i\mu\nu)}\right)^{-1} \prec \Phi,$$

which follow easily from (3.2.6) and (3.3.1), and in the last step the leading order term vanishes since the two X entries are independent for $i \neq j$. Then with (3.3.20), the terms in (3.3.19) can be bounded by $\mathrm{O}_\prec(N^{-1}\Phi^2)$.

In general, after the expansion of the two fractions in (3.3.17), we get a summation of terms of the form

$$A_{m,n} := \mathbb{E}\varepsilon_\mu^m \varepsilon_\nu^n (Y^* G^{(\mu\nu)} Y)_{\mu\nu}, \quad \mu \neq \nu,$$

up to some deterministic coefficients of order $\mathrm{O}(1)$. Since $|\varepsilon_{\mu,\nu}| \prec \Phi \lesssim N^{-\omega/2}$ for $z \in \mathbf{D}$ (we can take ω small enough such that $N^{-\omega/2} \geqslant q^2$), we only need to include the terms with $m + n \leqslant 2 + 2/\omega$ and the tail terms will be smaller than $N^{-1}\Phi^2$. Note that in $A_{m,n}$, the $X_{*\mu}$ entries, $X_{*\nu}$ entries and $G^{(\mu\nu)}$ entries are mutually independent. Moreover, both the number of $X_{*\mu}$ entries and the number of $X_{*\nu}$ entries are odd. Thus to attain a nonzero expectation, we must pair the X entries such that there are products of the forms $X_{i\mu}^{n_1}$ and $X_{j\nu}^{n_2}$ for some $n_1, n_2 \geqslant 3$. As a result, we lose $(n_1 - 2)/2 + (n_2 - 2)/2 \geqslant 1$ free indices, and this contributes an N^{-1} factor. On the other hand, for the product of G entries, we have the following three cases: (1) if there are at least 2 off-diagonal G entries, then we bound them with $\mathrm{O}_\prec(\Phi^2)$; (2) if there is only 1 off-diagonal G entry, then we can use the trick in (3.3.19) and the bound (3.3.20); (3) if there is no off-diagonal G entry, then we lose one more free index and get an extra N^{-1} factor. This leads to the estimate (3.3.15) for the term in (3.3.17).

For the second term in (3.3.16), we again use Lemma 2.6.3 to expand the $G_{\mu\nu}$, $G_{\nu\mu}$ and $G_{\nu\nu}^{-1}$ entries. Our goal is to expand all the G entries into polynomials of the random variables

$$S_{\alpha\beta} := (Y^* G^{(\mu\nu)} Y)_{\alpha\beta}, \quad \alpha, \beta \in \{\mu, \nu\}, \tag{3.3.21}$$

so that the X entries and $G^{(\mu\nu)}$ entries are independent in the resulting expression. In particular, the *maximally expanded* terms (see (3.3.22) below) can be expanded into $S_{\alpha\beta}$ variables directly through (2.6.5) and (2.6.6). However, *non-maximally expanded* terms are also created along the expansions in (2.6.6) and (2.6.8). Then we need to further expand these newly appeared terms. In general, this process will not terminate. However, we will show in Lemma 3.3.6 that after sufficiently many expansions, the resulting expression either has enough off-diagonal terms, or is maximally expanded. In the former case, it suffices to bound each off-diagonal term by $\mathrm{O}_\prec(\Phi)$. In the latter case, the expression will only consist of $S_{\alpha\beta}$ variables. Following the argument in the previous paragraph, the expectation over the X entries produces an N^{-1} factor, while the expectation over the G entries produces a Φ^2 factor.

In the rest of section, we give a rigorous proof based on the above arguments.

3.3.2 Resolvent expansions

To perform the resolvent expansion in a systematic way, we introduce the following notions of *string* and *string operator*.

Definition 3.3.3 (Strings)**.** *Let* $\mathfrak{A}$ *be the alphabet containing all symbols that will appear during the expansion:*

$$\mathfrak{A} = \left\{G_{\alpha\beta}, G_{\alpha\alpha}^{-1}, S_{\alpha\beta} \text{ with } \alpha, \beta \in \{\mu, \nu\}\right\} \cup \left\{G_{\mu\mu}^{(\nu)}, G_{\nu\nu}^{(\mu)}, (G_{\mu\mu}^{(\nu)})^{-1}, (G_{\nu\nu}^{(\mu)})^{-1}\right\}.$$

We define a string $\mathbf{s}$ *to be a concatenation of the symbols from* $\mathfrak{A}$*, and we use* $[\![\mathbf{s}]\!]$ *to denote the random variable represented by* $\mathbf{s}$*. We denote an empty string by* $\varnothing$ *with value* $[\![\varnothing]\!] = 0$*.*

Remark 3.3.4. It is important to distinguish a string $\mathbf{s}$ from its value $[\![\mathbf{s}]\!]$. For example, "$G_{\mu\nu}$" and "$G_{\mu\mu}G_{\nu\nu}^{(\mu)}S_{\mu\nu}$" are different strings, but they represent the same random variable by (2.6.6).

We shall call the following symbols the *maximally expanded* symbols:

$$\mathfrak{A}_{\max} = \left\{G_{\mu\nu}, G_{\nu\mu}, G_{\mu\mu}^{(\nu)}, G_{\nu\nu}^{(\mu)}, (G_{\mu\mu}^{(\nu)})^{-1}, (G_{\nu\nu}^{(\mu)})^{-1}, S_{\mu\mu}, S_{\nu\nu}, S_{\mu\nu}, S_{\nu\mu}\right\}. \tag{3.3.22}$$

A string $\mathbf{s}$ is said to be maximally expanded if all of its symbols are in $\mathfrak{A}_{\max}$. We shall call $G_{\mu\nu}, G_{\nu\mu}, S_{\mu\nu}, S_{\nu\mu}$ the *off-diagonal* symbols and all the other symbols *diagonal*. By (3.2.6) and (3.3.1), we have $[\![\mathbf{a}_o]\!] \prec \Phi$ if $\mathbf{a}_o$ is off-diagonal (we have $S_{\mu\nu} \prec \Phi$ using (2.6.6)) and $[\![\mathbf{a}_d]\!] \prec 1$ if $\mathbf{a}_d$ is diagonal. We use $\mathcal{F}_{n\text{-}max}(\mathbf{s})$ and $\mathcal{F}_{\text{off}}(\mathbf{s})$ to denote the number of non-maximally expanded symbols and the number of off-diagonal symbols, respectively, in $\mathbf{s}$.

Definition 3.3.5 (String operators)**.** *Let $\alpha \neq \beta \in \{\mu, \nu\}$.*

(i) We define an operator τ_0 acting on a string $\mathbf{s}$ in the following sense. Find the first $G_{\alpha\alpha}$ or $G_{\alpha\alpha}^{-1}$ in $\mathbf{s}$. If $G_{\alpha\alpha}$ is found, replace it with $G_{\alpha\alpha}^{(\beta)}$; if $G_{\alpha\alpha}^{-1}$ is found, replace it with $(G_{\alpha\alpha}^{(\beta)})^{-1}$; if neither is found, set $\tau_0(\mathbf{s}) = \mathbf{s}$ and we say that τ_0 is trivial for $\mathbf{s}$.

(ii) We define an operator τ_1 acting on a string $\mathbf{s}$ in the following sense. Find the first $G_{\alpha\alpha}$ or $G_{\alpha\alpha}^{-1}$ in $\mathbf{s}$. If $G_{\alpha\alpha}$ is found, replace it with $\frac{G_{\alpha\beta}G_{\beta\alpha}}{G_{\beta\beta}}$; if $G_{\alpha\alpha}^{-1}$ is found, replace it with $-\frac{G_{\alpha\beta}G_{\beta\alpha}}{G_{\alpha\alpha}G_{\alpha\alpha}^{(\beta)}G_{\beta\beta}}$; if neither is found, set $\tau_1(\mathbf{s}) = \varnothing$ and we say that τ_1 is null for $\mathbf{s}$.

(iii) The operator ρ replaces each $G_{\alpha\beta}$ in the string $\mathbf{s}$ with $G_{\alpha\alpha}G_{\beta\beta}^{(\alpha)}S_{\alpha\beta}$.

By Lemma 2.6.3, it is clear that for any string $\mathbf{s}$,

$$[\![\tau_0(\mathbf{s})]\!] + [\![\tau_1(\mathbf{s})]\!] = [\![\mathbf{s}]\!], \quad [\![\rho(\mathbf{s})]\!] = [\![\mathbf{s}]\!]. \tag{3.3.23}$$

Moreover, a string $\mathbf{s}$ is trivial under τ_0 and null under τ_1 if and only if $\mathbf{s}$ is maximally expanded. Given a string $\mathbf{s}$, we abbreviate $\mathbf{s}_0 := \tau_0(\mathbf{s})$ and $\mathbf{s}_1 := \rho(\tau_1(\mathbf{s}))$. For any sequence $w = a_1a_2\ldots a_m$ with $a_i \in \{0, 1\}$, we denote

$$\mathbf{s}_w := \rho^{a_m}\tau_{a_m}\ldots\rho^{a_2}\tau_{a_2}\rho^{a_1}\tau_{a_1}(\mathbf{s}), \quad \text{where } \rho^0 := 1.$$

Then by (3.3.23) we have

$$\sum_{|w|=m} [\![\mathbf{s}_w]\!] = [\![\mathbf{s}]\!], \tag{3.3.24}$$

where the summation is over all binary sequences w with length $|w| = m$.

Lemma 3.3.6. *Consider the string $\mathbf{s} =$ "$G_{\mu\mu}G_{\nu\nu}^{(\mu)}S_{\mu\nu}$". Let w be any binary sequence with $|w| = 4l_0$ and such that $\mathbf{s}_w \neq \varnothing$. Then either $\mathcal{F}_{\mathrm{off}}(\mathbf{s}_w) \geqslant 2l_0$ or $\mathbf{s}_w$ is maximally expanded.*

Proof. It suffices to show that any nonempty string $\mathbf{s}_w$ with $\mathcal{F}_{\mathrm{off}}(\mathbf{s}_w) < 2l_0$ is maximally expanded. By Definition 3.3.5, a nontrivial τ_0 reduces the number of non-maximally expanded symbols by 1, and keeps the number of off-diagonal symbols the same; a $\rho\tau_1$ increases the number of non-maximally expanded symbols by 2 or 3, and increases the number of off-diagonal symbols by 2. Hence $\mathcal{F}_{\mathrm{off}}(\mathbf{s}_w) < 2l_0$ implies that there are at most $(l_0 - 1)$ 1's in w.

Those $\rho\tau_1$ operators increase $\mathcal{F}_{n\text{-}max}$ at most by $3(l_0-1)$ in total. On the other hand, there are at least $3l_0$ 0's in w, which is sufficient to eliminate all the non-maximally expanded symbols (whose number is at most $3(l_0-1)+1 = 3l_0-2$ in total since $\mathcal{F}_{n\text{-}max}(\mathbf{s}) = 1$ for the initial string). $\square$

Now we choose $l_0 = 1 + 1/\omega$. Then using $\Phi = \mathrm{O}(N^{-\omega/2})$, we have

$$\sum_{|w|=4l_0} [\![\mathbf{s}_w]\!] \cdot \mathbf{1}(\mathcal{F}_{\text{off}}(\mathbf{s}_w) \geqslant 2l_0) \prec 2^{4l_0}\Phi^{2l_0} \prec N^{-1}\Phi^2.$$

By Lemma 3.3.6, we see that to prove (3.3.15), it suffices to show that

$$|\mathbb{E}[\![\mathbf{s}_w]\!]| \prec N^{-1}\Phi^2 \tag{3.3.25}$$

for any maximally expanded string $\mathbf{s}_w$ with $|w| = 4l_0$. Note that the maximally expanded string $\mathbf{s}_w$ thus obtained consists only of the symbols

$$G^{(\beta)}_{\alpha\alpha},\ (G^{(\beta)}_{\alpha\alpha})^{-1},\ S_{\alpha\beta}, \quad \text{with } \alpha \neq \beta \in \{\mu, \nu\}.$$

By (2.6.5), we can replace $(G^{(\beta)}_{\alpha\alpha})^{-1}$ with

$$(G^{(\beta)}_{\alpha\alpha})^{-1} = -z - S_{\alpha\alpha}. \tag{3.3.26}$$

Note that $|S_{\alpha\alpha} - \widetilde{\sigma}_\mu N^{-1}\sum_i \sigma_i \Pi_{ii}| \prec \Phi$ by (3.3.18). Then we can expand $G^{(\beta)}_{\alpha\alpha}$ as

$$G^{(\beta)}_{\alpha\alpha} = \Pi_{\alpha\alpha}\sum_{k=0}^{2l_0} \Pi^k_{\alpha\alpha}\left(S_{\alpha\alpha} - \widetilde{\sigma}_\mu N^{-1}\sum_i \sigma_i\Pi_{ii}\right)^k + O_\prec(N^{-1}\Phi^2). \tag{3.3.27}$$

We apply the expansions (3.3.26) and (3.3.27) to the G symbols in $\mathbf{s}_w$, disregard the sufficiently small tails, and denote the resulting polynomial (in terms of the symbols $S_{\alpha\beta}$) by P_w. Then P_w can be written as a finite sum of maximally expanded strings (or monomials) consisting of the $S_{\alpha\beta}$ symbols. Moreover, the number of such monomials depends only on l_0. Hence we only need to prove that for any such monomial M_w,

$$|\mathbb{E}[\![M_w]\!]| \prec N^{-1}\Phi^2. \tag{3.3.28}$$

Let N_μ (N_ν) be the number of times that μ (ν) appears as a lower index of the S symbols in M_w. We have $N_\mu = N_\nu = 3$ for the initial string $\mathbf{s} =$ "$G_{\mu\mu}G^{(\mu)}_{\nu\nu}S_{\mu\nu}$". From Definition

3.3.5, it is easy to see that the operators τ_0, τ_1 and ρ do not change the parity of N_μ and N_ν. The expansions (3.3.26) and (3.3.27) also do not change the parity of N_μ and N_ν. This leads to the following key observation:

$$\textit{both } N_\mu \textit{ and } N_\nu \textit{ are odd in } M_w. \tag{3.3.29}$$

3.3.3 A graphical proof

In this subsection, we finish the proof of (3.3.28). Suppose $M_w = C(z)(S_{\mu\mu})^{m_1}(S_{\nu\nu})^{m_2}(S_{\mu\nu})^{m_3}(S_{\nu\mu})^{m_4}$, where $C(z)$ denotes a deterministic function of order 1 for all $z \in \mathbf{D}$. Then we write

$$\begin{aligned}\llbracket M_w \rrbracket \sim \sum_{i_*^{(*)}, j_*^{(*)} \in \mathcal{I}_1} & \prod_{a=1}^{m_1} X_{i_a^{(1)}\mu} G^{(\mu\nu)}_{i_a^{(1)} j_a^{(1)}} X_{j_a^{(1)}\mu} \prod_{b=1}^{m_2} X_{i_b^{(2)}\nu} G^{(\mu\nu)}_{i_b^{(2)} j_b^{(2)}} X_{j_b^{(2)}\nu} \\ & \prod_{c=1}^{m_3} X_{i_c^{(3)}\mu} G^{(\mu\nu)}_{i_c^{(3)} j_c^{(3)}} X_{j_c^{(3)}\nu} \prod_{d=1}^{m_4} X_{i_d^{(4)}\nu} G^{(\mu\nu)}_{i_d^{(4)} j_d^{(4)}} X_{j_d^{(4)}\mu},\end{aligned} \tag{3.3.30}$$

where we have ignored the coefficients containing σ_i's and $\widetilde{\sigma}_\mu$'s. To avoid heavy expressions, we introduce the following graphical notations. We use a connected graph (V, E) to represent the string M_w, where the vertex set V consists of the indices in (3.3.30) and the edge set E consists of the X and G variables. The indices μ, ν are represented by the black vertices in the graph, while the i, j indices are represented by the white vertices. The X edges are represented by the zig-zag lines and the G edges are represented by the straight lines. One can refer to Fig. 3.3 for an example of such a graph.

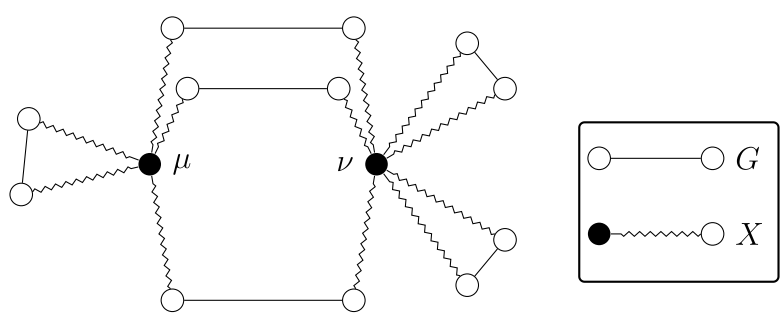


Figure 3.3: The graph representing $S_{\mu\mu}(S_{\mu\nu})^3(S_{\nu\nu})^2$.

We organize the summation in (3.3.30) in the following way. We first partition the white vertices into blocks by requiring that any pair of white vertices take the same value if they are in the same block, and take different values otherwise. Then we take the summation over the white blocks which take values in $\mathcal{I}_2$. Finally, we sum over all possible partitions. Note that the number of different partitions depends only on the total number of S variables in M_w, which in turn depends only on l_0.

Fix a partition Γ of the white vertices. We denote its blocks by $b_1, ..., b_k$, where k gives the number of distinct blocks in Γ. We denote by n_l^μ (n_l^ν) the number of white vertices in b_l that are connected to the vertex μ (ν). Let $G(\Gamma)$ be the product of all the G edges in the graph. Then we have

$$[\![M_w]\!] \sim \sum_\Gamma \sum_{b_1,\ldots,b_k}^{*} G(\Gamma) \prod_{l=1}^{k} (X_{b_l\mu})^{n_l^\mu} (X_{b_l\nu})^{n_l^\nu}, \tag{3.3.31}$$

where $\sum^*$ denotes the summation subject to the condition that $b_1, ..., b_k$ all take distinct values. Note that k, b_l, n_l^μ and n_l^ν all depend on Γ, and we have omitted the Γ dependence for simplicity of notations.

From (3.3.30), it is easy to observe that the X edges are independent of $G(\Gamma)$. Thus taking expectation of (3.3.31) gives that

$$\begin{aligned} |\mathbb{E}[\![M_w]\!]| \leqslant & C \sum_\Gamma \sum_{b_1,\ldots,b_k}^{*} |\mathbb{E}G(\Gamma)| \prod_{l=1}^{k} |\mathbb{E}(X_{b_l\mu})^{n_l^\mu}||\mathbb{E}(X_{b_l\nu})^{n_l^\nu}| \\ \leqslant & C \sum_\Gamma \sum_{b_1,\ldots,b_k}^{*} |\mathbb{E}G(\Gamma)| \prod_{l=1}^{k} \mathbb{E}|X_{b_l\mu}|^{n_l^\mu} \mathbb{E}|X_{b_l\nu}|^{n_l^\nu} \mathbf{1}(n_l^\mu \neq 1, n_l^\nu \neq 1). \end{aligned} \tag{3.3.32}$$

Note that we must have $n_l^\mu + n_l^\nu \geqslant 2$ for $1 \leqslant l \leqslant k$, because we only consider nonempty blocks. On the other hand, if all n_l^μ are even, then $N_\mu = \sum_{l=1}^k n_l^\mu$ must be even, which contradicts (3.3.29). Hence we can find some $1 \leqslant l_1 \leqslant k$ such that $n_{l_1}^\mu$ is odd and $n_{l_1}^\mu \geqslant 3$. Similarly, we can also find some $1 \leqslant l_2 \leqslant k$ such that $n_{l_2}^\nu$ is odd and $n_{l_2}^\nu \geqslant 3$. We abbreviate $\hat{n}_l^\mu := n_l^\mu \wedge 3$ and $\hat{n}_l^\nu := n_l^\nu \wedge 3$. From the above discussions, we see that

$$\frac{1}{2}\sum_{l=1}^{k} (\hat{n}_l^\mu + \hat{n}_l^\nu) \geqslant \frac{1}{2}\sum_{l \neq l_1, l_2}^{k} (\hat{n}_l^\mu + \hat{n}_l^\nu) + \frac{3}{2} + \frac{3}{2} \geqslant (k-2) + 3 = k + 1. \tag{3.3.33}$$

Now using the moment assumption (3.1.12), we can bound (3.3.32) by

$$|\mathbb{E}[\![M_w]\!]| \leqslant C \sum_\Gamma \sum_{b_1,\ldots,b_k}^{*} |\mathbb{E}G(\Gamma)| N^{-\sum_{l=1}^k (\hat{n}_l^\mu + \hat{n}_l^\nu)/2}. \tag{3.3.34}$$

Next we deal with $|\mathbb{E}G(\Gamma)|$. We consider the following 3 cases separately: (i) there are at least 2 off-diagonal G-edges in $G(\Gamma)$; (ii) there is only 1 off-diagonal G-edge in $G(\Gamma)$; (iii) there is no off-diagonal G-edge in $G(\Gamma)$.

In case (i), we trivially have $|\mathbb{E}G(\Gamma)| \prec \Phi^2$. In case (ii), we use the same trick as in (3.3.19). Let the off-diagonal G-edge be $G_{ij}^{(\mu\nu)}$. For each diagonal $G_{kk}^{(\mu\nu)}$, we replace it with $(G_{kk}^{(\mu\nu)} - \Pi_{kk}) + \Pi_{kk} = \Pi_{kk} + \mathrm{O}_\prec(\Phi)$. Plugging these expansions into $\mathbb{E}G(\Gamma)$, we obtain that $|\mathbb{E}G(\Gamma)| \prec \Phi^2 + |\mathbb{E}G_{ij}^{(\mu\nu)}| \prec \Phi^2$, where we used (3.3.20) in the second step. Finally, in case (iii), we have $|\mathbb{E}G(\Gamma)| \prec 1$. Moreover, $n_l^\mu + n_l^\nu$ is even for any $1 \leqslant l \leqslant k$. Take $1 \leqslant l_1, l_2 \leqslant k$ such that $n_{l_1}^\mu, n_{l_2}^\nu$ are odd and $n_{l_1}^\mu, n_{l_2}^\nu \geqslant 3$. If $l_1 \neq l_2$, then we must have $\hat{n}_{l_1}^\mu + \hat{n}_{l_1}^\nu \geqslant 4$, $\hat{n}_{l_2}^\mu + \hat{n}_{l_2}^\nu \geqslant 4$, and hence

$$\frac{1}{2}\sum_{l=1}^{k}(\hat{n}_l^\mu + \hat{n}_l^\nu) \geqslant \frac{1}{2}\sum_{l \neq l_1, l_2}^{k}(\hat{n}_l^\mu + \hat{n}_l^\nu) + 4 \geqslant k+2.$$

Otherwise, if $l_1 = l_2$, then

$$\frac{1}{2}\sum_{l=1}^{k}(\hat{n}_l^\mu + \hat{n}_l^\nu) \geqslant \frac{1}{2}\sum_{l \neq l_1}^{k}(\hat{n}_l^\mu + \hat{n}_l^\nu) + 3 \geqslant k+2.$$

Now applying the above estimates and (3.3.33) to (3.3.34), we obtain that

$$\begin{aligned}|\mathbb{E}[\![M_w]\!]| &\prec \sum_{\Gamma \text{ in Case (1), (2)}} \Phi^2 N^{k-\sum_{l=1}^k(\hat{n}_l^\mu+\hat{n}_l^\nu)/2} + \sum_{\Gamma \text{ in Case (3)}} N^{k-\sum_{l=1}^k(\hat{n}_l^\mu+\hat{n}_l^\nu)/2} \\ &\leqslant C(N^{-1}\Phi^2 + N^{-2}) \leqslant CN^{-1}\Phi^2.\end{aligned}$$

This concludes the proof of (3.3.28), and hence finishes the proof of (3.3.15).

3.4 Proof of Theorem 3.2.1

By Lemma 3.3.2, we can assume that the entries of X are centered without loss of generality. According to the comments below Theorem (3.2.1), we only need to prove the bound (3.2.6). By polarization, it suffices to prove that

$$|\langle \mathbf{v}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{v}, \Pi(z)\mathbf{v}\rangle| \prec q^2 + (N\eta)^{-1/2} \tag{3.4.1}$$

for $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$ or $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}$. It easy to see that (3.4.1) follows from the next two lemmas.

Lemma 3.4.1. *Let X be an $n \times N$ real random matrix whose entries are independent random variables satisfying (3.3.2), (3.1.12), and the bounded support condition* (2.4.1) *with* $q \leqslant N^{-\phi}$ *for some constant $\phi > 0$. If*

$$\max_{a,b\in\mathcal{I}} |G_{ab}(X,z) - \Pi_{ab}(z)| \prec q + \Psi(z), \tag{3.4.2}$$

and

$$|m_1(X,z) - m_{1c}(z)| + |m_2(X,z) - m_{2c}(z)| \prec (N\eta)^{-1} \tag{3.4.3}$$

hold uniformly in $z \in \mathbf{D}$, then the following local law also holds uniformly in $z \in \mathbf{D}$:

$$\max_{r=1,2}\max_{a,b\in\mathcal{I}_r} |G_{ab}(X,z) - \Pi_{ab}(z)| \prec q^2 + (N\eta)^{-1/2}. \tag{3.4.4}$$

Lemma 3.4.2. *Suppose the assumptions in Lemma 3.4.1 hold. Let $\Phi(z)$ be a deterministic function on $\mathbf{D}$ satisfying $c_1(N^{-1/2}+q^2) \leqslant \Phi(z) \leqslant N^{-c_1}$ for some constant $c_1 > 0$. If we have*

$$\max_{a,b\in\mathcal{I}} |G_{ab}(z) - \Pi_{ab}(z)|^2 \prec \Phi, \quad \max_{r=1,2}\max_{a,b\in\mathcal{I}_r} |G_{ab}(z) - \Pi_{ab}(z)| \prec \Phi, \tag{3.4.5}$$

uniformly in $z \in \mathbf{D}$, then

$$|\langle \mathbf{v}, G(X,z)\mathbf{v}\rangle - \langle \mathbf{v}, \Pi(z)\mathbf{v}\rangle| \prec \Phi(z) \tag{3.4.6}$$

uniformly in $z \in \mathbf{D}$ and deterministic unit vectors $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$ or $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}$.

Suppose (3.4.2) holds. Then using (2.7.2)-(2.7.5) and (3.3.1), it is easy to verify that for $z \in \mathbf{D}$,

$$\max\left\{ \sum_i \left|G_{ai}^{(\mathbb{T})}\right|^2, \sum_i \left|G_{ia}^{(\mathbb{T})}\right|^2, \sum_\mu \left|G_{a\mu}^{(\mathbb{T})}\right|^2, \sum_\mu \left|G_{\mu a}^{(\mathbb{T})}\right|^2 \right\} \prec \eta^{-1}, \tag{3.4.7}$$

for any $a \in \mathcal{I}$ and $\mathbb{T} \subseteq \mathcal{I}$ with $|\mathbb{T}| = \mathrm{O}(1)$.

3.4.1 Proof of Lemma 3.4.1

We only prove

$$\max_{i,j\in\mathcal{I}_1} |G_{ij}(X,z) - \Pi_{ij}(z)| \prec q^2 + (N\eta)^{-1/2}. \tag{3.4.8}$$

The proof for (3.4.4) with $a, b \in \mathcal{I}_2$ is exactly the same. First, we recall the large deviation lemma—Lemma 2.6.6. If we assume the fourth moment of x_i is bounded for all i as in (3.1.12), then we have a better bound for the LHS of (2.6.17).

Lemma 3.4.3. *Suppose the assumptions in Lemma 2.6.6 hold and x_i, $1 \leqslant i \leqslant K$, satisfy (3.1.12). Then we have*

$$\Big|\sum_i \left(|x_i|^2 - \mathbb{E}|x_i|^2\right) B_{ii}\Big| \prec \left(q^2 + N^{-1/2}\right) B_d. \tag{3.4.9}$$

Proof. We abbreviate $z_i := (|x_i|^2 - \mathbb{E}|x_i|^2)\, B_{ii}/B_d$. By Markov's inequality, it suffices to prove that for any fixed $p \in \mathbb{N}$,

$$\mathbb{E}\Big|\sum_i z_i\Big|^{2p} \prec \left(q^2 + N^{-1/2}\right)^{2p}. \tag{3.4.10}$$

Note that by the assumption, we have

$$\mathbb{E}z_i = 0, \quad \mathbb{E}|z_i|^n \prec q^{2n-4}N^{-2} \text{ for fixed } n \geqslant 2. \tag{3.4.11}$$

Now we expand the LHS of (3.4.10) as

$$\mathbb{E}\Big|\sum_i z_i\Big|^{2p} = \sum_{i_1,\ldots,i_{2p}} \mathbb{E}y_{i_1}\cdots y_{i_{2p}},$$

where we denote $y_{i_l} := z_{i_l}$ for $1 \leqslant l \leqslant p$ and $y_{i_l} := \bar{z}_{i_l}$ for $p+1 \leqslant l \leqslant 2p$. To organize the summation over the indices $i_1, \ldots, i_{2p}$, we look at the partitions Γ of the set of the labels $\{1, ..., 2p\}$ according to the equivalence relation that k, l are in the same class if and only if $i_k = i_l$. We use b_l, $1 \leqslant l \leqslant k$, to denote the equivalence classes of Γ and n_l to denote the size of b_l. Obviously, k, b_l and n_l all depend on Γ, but we will omit this dependence in the following expressions. Moreover, since the random variables are centered, we must have $n_l \geqslant 2$ for all l to attain a nonzero expectation. Hence we have

$$\mathbb{E}\Big|\sum_i z_i\Big|^{2p} \leqslant \sum_\Gamma \sum_{b_1,\ldots,b_k}^{*} \mathbb{E}|y_{i_{b_1}}|^{n_1} \ldots \mathbb{E}|y_{i_{b_k}}|^{n_k}, \tag{3.4.12}$$

where $\sum^*$ denotes the summation subject to the conditions that $b_1, \ldots, b_k$ are all distinct, $n_l \geqslant 2$ for all l, and $\sum_{l=1}^k n_l = 2p$. Note that under these conditions, we trivially have $k \leqslant p$.

Using (3.4.11), we obtain that

$$\begin{aligned}\sum_{b_1,\ldots,b_k}^{*} \mathbb{E}|y_{\mu_{b_1}}|^{n_1} \ldots \mathbb{E}|y_{\mu_{b_k}}|^{n_k} &\prec \sum_{b_1,\ldots,b_k}^{*} (q^{2n_1-4}N^{-2}) \ldots (q^{2n_k-4}N^{-2}) \\ &= \sum_{b_1,\ldots,b_k}^{*} N^{-2k}q^{4p-4k} \leqslant CN^{-k}q^{4p-4k}.\end{aligned}$$

Since the number of partitions of $\{1, ..., 2p\}$ is finite and depends only on p, (3.4.12) can then be bounded by

$$\mathbb{E}\Big|\sum_i z_i\Big|^{2p} \prec \max_{1\leqslant k\leqslant p} N^{-k} q^{4p-4k} \leqslant q^{4p} + N^{-p},$$

where in the last step, q^{4p} and N^{-p} can be obtained from the extreme cases $k = 0$ and $k = p$, respectively. This concludes (3.4.10). □

Now using (2.6.6) and (2.6.16), we get that for $i \neq j \in \mathcal{I}_1$,

$$\begin{aligned}|G_{ij}| \prec \left|\sum_{\mu,\nu} Y_{i\mu} G^{(ij)}_{\mu\nu} Y^*_{\nu j}\right| &\prec q^2 \max_{\mu} |G^{(ij)}_{\mu\mu}| + q \max_{\mu\neq\nu} |G^{(ij)}_{\mu\nu}| + \Big(\frac{1}{N^2}\sum_{\mu\neq\nu} |G^{(ij)}_{\mu\nu}|^2\Big)^{1/2} \\ &\prec q^2 + q(q+\Psi) + \left(\frac{1}{N\eta}\right)^{1/2} \prec q^2 + (N\eta)^{-1/2}, \end{aligned} \tag{3.4.13}$$

where we used (3.4.2), (3.3.1) and the bound (3.4.7). For the diagonal estimate, we need to control the Z variables. Using (2.6.18) and (3.4.9), we get that

$$\begin{aligned}|Z_i| &= \sigma_i \Big|\sum_{\mu} \widetilde{\sigma}_\mu G^{(i)}_{\mu\mu} \left(|X_{i\mu}|^2 - \mathbb{E}|X_{i\mu}|^2\right) + \sum_{\mu\neq\nu} \sqrt{\widetilde{\sigma}_\mu\widetilde{\sigma}_\nu} X_{i\mu} G^{(i)}_{\mu\nu} X^*_{\nu i}\Big| \\ &\prec \left(q^2 + N^{-1/2}\right) + q \max_{\mu\neq\nu} |G^{(i)}_{\mu\nu}| + \frac{1}{N}\Big(\sum_{\mu\neq\nu} |G^{(i)}_{\mu\nu}|^2\Big)^{1/2} \\ &\prec q^2 + (N\eta)^{-1/2}, \end{aligned} \tag{3.4.14}$$

where we used (3.4.2), (3.3.1) and (3.4.7) again. Then with (2.6.5), we get that

$$\begin{aligned} G_{ii} - \Pi_{ii} &= \frac{1}{-1 - \sigma_i m_{2c} - \sigma_i (m_2^{(i)} - m_{2c}) + Z_i} - \frac{1}{-1 - \sigma_i m_{2c}} \\ &= \mathrm{O}_\prec\left(q^2 + (N\eta)^{-1/2}\right) \end{aligned}$$

where in the second step we used (3.4.14), (3.4.3) and (3.3.1) (with $\tilde{\Phi} = q + \Psi$). Together with (3.4.13), we conclude (3.4.8).

3.4.2 Proof of Lemma 3.4.2

We only prove (3.4.6) for $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$. The proof for the case with $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_2}$ is exactly the same. Note that by (3.4.5), we immediately get $\sum_i |v_i|^2 (G_{ii} - \Pi_{ii}) \prec \Phi$. Hence it remains to prove

that

$$\sum_{i\neq j}\bar{v}_i v_j G_{ij} \prec \Phi.$$

By Markov's inequality, it suffices to show that

$$\mathbb{E}\Big|\sum_{i\neq j}\bar{v}_i v_j G_{ij}\Big|^{2p} \prec \Phi^{2p} \tag{3.4.15}$$

for any fixed $p\in\mathbb{N}$. The proof of (3.4.15) is similar to the ones in [14, Section 5]. The main difference is that in [14], the matrix entries are assumed to have arbitrarily high moments, while here we assume that the X entries have finite third moment and support bounded by q. In particular, for any fixed $k \geqslant 3$, we have

$$\mathbb{E}|X_{i\mu}|^k \prec q^{k-3}N^{-3/2}, \quad i\in\mathcal{I}_1,\ \mu\in\mathcal{I}_2. \tag{3.4.16}$$

(Note that we have a stronger moment assumption in (3.1.12). However, the finite fourth moment condition will not be used in the proof below. We only need the weaker bound (3.4.16).) We will use a graphical tool as in Section 3.3.

We first rewrite the product in (3.4.15) as

$$\begin{aligned}\Big|\sum_{i\neq j}\bar{v}_i G_{ij} v_j\Big|^{2p} &= \sum_{i_k\neq j_k\in\mathcal{I}_1}\prod_{k=1}^{p}\bar{v}_{i_k}G_{i_kj_k}v_{j_k}\cdot\prod_{k=p+1}^{2p}\overline{\bar{v}_{i_k}G_{i_kj_k}v_{j_k}}\\ &=\sum_{\Gamma}\sum_{b_1,\ldots,b_r}^{*}\prod_{k=1}^{p}\bar{v}_{\Gamma(i_k)}G_{\Gamma(i_k)\Gamma(j_k)}v_{\Gamma(j_k)}\cdot\prod_{k=p+1}^{2p}\overline{\bar{v}_{\Gamma(i_k)}G_{\Gamma(i_k)\Gamma(j_k)}v_{\Gamma(j_k)}},\end{aligned}$$

where Γ ranges over all partitions of the set of labels $\{i_1,...,i_{2p},j_1,...,j_{2p}\}$ with the restriction that i_k, j_k cannot be in the same equivalence class for all k, $\{b_1,...,b_r\}$ is the set of equivalence classes for a fixed Γ, $\Gamma(\cdot)$ is regarded as a mapping from the set of labels to the set of equivalence classes, and $\sum^*$ denotes the summation subject to the condition that $b_1,\ldots,b_r$ all take distinct values and $\Gamma(i_k)\neq\Gamma(j_k)$ for all k. Since the number of such partitions Γ is finite and depends only on p, it suffices to prove that for any fixed Γ,

$$\mathbb{E}\sum_{b_1,\ldots,b_r}^{*}\prod_{k=1}^{p}\bar{v}_{\Gamma(i_k)}G_{\Gamma(i_k)\Gamma(j_k)}v_{\Gamma(j_k)}\cdot\prod_{k=p+1}^{2p}\overline{\bar{v}_{\Gamma(i_k)}G_{\Gamma(i_k)\Gamma(j_k)}v_{\Gamma(j_k)}} \prec \Phi^{2p}. \tag{3.4.17}$$

We abbreviate

$$P(b_1,...,b_r) := \prod_{k=1}^{p}G_{\Gamma(i_k)\Gamma(j_k)}\cdot\prod_{k=p+1}^{2p}\overline{G_{\Gamma(i_k)\Gamma(j_k)}}.$$

For simplicity, we shall omit the overline for complex conjugate in the following proof. In this way, we can avoid a lot of immaterial notational complexities that do not affect the proof.

For $k = 1, ..., r$, we denote by $\deg(b_k, P)$ the number of times that b_k appears as an index of the G entries in P, i.e. $\deg(b_k, P) := |\Gamma^{-1}(b_k)|$. We define $h := \#\{1 \leqslant k \leqslant r : \deg(b_k, P) = 1\}$, i.e. h is the number of b_k's that only appear once in the indices of P. Without loss of generality, we assume these b_k's are $b_1, ..., b_h$. Then we have the following properties:

$$\sum_{k=1}^{r} \deg(b_k, P) = 4p, \quad \text{and} \quad \deg(b_k, P) = 1 \;\text{ for } k = 1, ..., h. \tag{3.4.18}$$

Now we claim that

$$|\mathbb{E}P| \prec N^{-h/2}\Phi^{2p}. \tag{3.4.19}$$

Note that by $\|\mathbf{v}\|_2 = 1$ and Cauchy-Schwarz inequality, we have $\sum_i |v_i| \leqslant \sqrt{M}$ and $\sum_i |v_i|^k \leqslant 1$ for $k \geqslant 2$. Then if (3.4.19) holds, we can bound the left hand side of (3.4.17) by

$$N^{-h/2}\Phi^{2p} \prod_{k=1}^{r} \left(\sum_{b_k} |v_{b_k}|^{\deg(b_k, P)} \right) \leqslant N^{-h/2}\Phi^{2p}(\sqrt{M})^h \leqslant C\Phi^{2p}.$$

Hence it suffices to prove (3.4.19).

We define the S variables as

$$S_{ij} := (YG^{(L)}Y^*)_{ij}, \tag{3.4.20}$$

for $i, j \in \mathcal{I}_1$ and $L := \{b_1, ..., b_r\}$. As in (3.4.13) and (3.4.14), we can verify that $|S_{ij} - \sigma_i m_{2c}\delta_{ij}| \prec \Phi$ for $i, j \in \mathcal{I}_1$ using (3.4.5), (3.3.1) and Lemmas 2.6.6 and Lemma 3.4.3. Then as in Section 3.3.2, we keep expanding the G entries in P using the resolvent expansions in Lemma 2.6.3, until each monomial in the expression either consists of S variables only or has sufficiently many off-diagonal terms. The following lemma has been proved in [14, Lemma 5.9] and [105, Lemma 5.9].

Lemma 3.4.4. *After finitely many expansions, we can write P as*

$$P = \sum_{\alpha=1}^{A} c_\alpha Q_\alpha + O_\prec(N^{-h/2}\Phi^{2p}), \tag{3.4.21}$$

where $A \in \mathbb{N}$ depends only on p and c_1 (recall that $\Phi(z) \leqslant N^{-c_1}$ by our assumption), c_α's are constants of order $\mathrm{O}(1)$, and Q_α are monomials of S variables only, where the number of S variables in each Q_α depends only on p and c_1. Moreover, we have that

$$\deg_o(b_k, Q_\alpha) \geqslant \deg_o(b_k, P), \quad \deg_o(b_k, Q_\alpha) = \deg_o(b_k, P) \mod 2, \tag{3.4.22}$$

for $k = 1, ..., r$ and $\alpha = 1, ..., A$, and the number of off-diagonal S variables in Q_α is at least $2p$. Here $\deg_o(b_k, Q_\alpha)$ denotes the number of times that b_k appears as an index of the off-diagonal S variables in Q_α, and $\deg_o(b_k, P) := \deg(b_k, P)$ (which is consistent with the previous definition since P only contains off-diagonal entries).

Now given the expansion in (3.4.21), we see that to conclude (3.4.19), it suffices to show that for any Q_α,

$$|\mathbb{E}Q_\alpha| \prec N^{-h/2}\Phi^{2p}. \tag{3.4.23}$$

In the following proof, we fix one such $Q \equiv Q_\alpha$ and write

$$\begin{aligned} Q = \prod_{j=1}^{J} S_{b_{k_j} b_{l_j}} &= \sum_{\mu_j, \nu_j \in \mathcal{I}_2} \prod_{j=1}^{J} Y_{b_{k_j}\mu_j} G^{(L)}_{\mu_j\nu_j} Y^*_{\nu_j b_{l_j}} \\ &= \sum_{W} \sum_{w_1,...,w_m}^{*} \prod_{j=1}^{J} Y_{b_{k_j} W(\mu_j)} G^{(L)}_{W(\mu_j)W(\nu_j)} Y_{b_{l_j} W(\nu_j)} \end{aligned}$$

where J is the number of S-variables in Q, W ranges over all partitions of the set of indices $\{\mu_1, ..., \mu_J, \nu_1, ..., \nu_J\}$, $\{w_1, ..., w_m\}$ denotes the set of equivalence classes for a particular W, $W(\cdot)$ is regarded as a symbolic mapping from the set of indices to the set of equivalence classes, and $\sum^*$ denotes the summation subject to the condition that $w_1, \ldots, w_m$ all take distinct values. Note that the number of partitions W depends only on J. For a fixed partition W, we denote

$$R(w_1, ..., w_m; W) := \prod_{j=1}^{J} X_{b_{k_j} W(\mu_j)} G^{(L)}_{W(\mu_j)W(\nu_j)} X_{b_{l_j} W(\nu_j)}.$$

Then to prove (3.4.23), it suffices to show that

$$|\mathbb{E}R(w_1, ..., w_m; W)| \prec N^{-m-h/2}\Phi^{2p}. \tag{3.4.24}$$

for any partition W.

To facilitate the proof, we introduce the graphical notations as in Section 3.3.3. We use a connected graph (V, E) to represent R, where the vertex set V consists of black vertices $b_1, \ldots, b_r$ and white vertices $w_1, \ldots, w_m$, and the edge set E consists of (k, α) edges representing $X_{b_k w_\alpha}$ and (α, β) edges representing $G^{(L)}_{w_\alpha w_\beta}$. We denote

$$e_{k\alpha} := \text{number of } (k, \alpha) \text{ edges in } R, \quad d_\alpha := \text{number of } (\alpha, \alpha) \text{ edges in } R.$$

Note that to attain a nonzero expectation, we must have

$$e_{k\alpha} = 0 \text{ or } e_{k\alpha} \geqslant 2 \quad \text{for all } k, \alpha. \tag{3.4.25}$$

We also define

$$e^{(o)}_{k\alpha} := \text{number of } (k, \alpha) \text{ edges that are from off-diagonal } S \text{ in } Q.$$

Then we have

$$\sum_\alpha e^{(o)}_{k\alpha} = \deg_o(b_k, Q) \tag{3.4.26}$$

By (3.4.18), (3.4.25) and the parity conservation due to (3.4.22), there exist edges $(1, \alpha_1), ..., (h, \alpha_h)$ such that $e_{k\alpha_k}$ is odd and $e_{k\alpha_k} \geqslant 3$, $1 \leqslant k \leqslant h$. Let $H := \{(1, \alpha_1), ..., (h, \alpha_h)\}$ be the set of these edges. Denote by F the set of (k, α) edge such that $e_{k\alpha} \geqslant 2$ and $(k, \alpha) \notin H$. Denote

$$s_\alpha := \sum_{k=1}^r e_{k\alpha}, \quad h_{k\alpha} := \mathbf{1}_{(k,\alpha)\in H}, \quad h_\alpha := \sum_{k=1}^r h_{k\alpha}, \quad f_\alpha := \sum_{k=1}^r \mathbf{1}_{(k,\alpha)\in F}$$

for all $k = 1, ..., r$ and $\alpha = 1, ..., m$. By the above definitions, we have $s_\alpha \geqslant 2$ and $h_\alpha + f_\alpha > 0$ (since the classes w_α are nonempty), $s_\alpha \geqslant 2d_\alpha$, and

$$\sum_\alpha h_{k\alpha} = \mathbf{1}(1 \leqslant k \leqslant h), \quad \sum_\alpha h_\alpha = h. \tag{3.4.27}$$

Note that there are $\frac{1}{2}\sum_{k,\alpha} e_{k\alpha} - \sum_\alpha d_\alpha$ off-diagonal G edges in R. Hence by (3.4.5) and (3.4.16), we have

$$\begin{aligned} |\mathbb{E}R| &\prec \prod_{\alpha=1}^m \left(\Phi^{-d_\alpha} \prod_{k=1}^n \Phi^{\frac{1}{2}e_{k\alpha}} \mathbb{E}|X_{b_k w_\alpha}|^{e_{k\alpha}} \right) \\ &\prec \prod_{\alpha=1}^m \Phi^{s_\alpha/2 - d_\alpha} \Big(\prod_{(k,\alpha)\in H} q^{e_{k\alpha}-3} N^{-3/2} \Big) \Big(\prod_{(k,\alpha)\in F} q^{e_{k\alpha}-2} N^{-1} \Big) =: \prod_{\alpha=1}^m R_\alpha. \end{aligned}$$

Now we consider the following four cases for R_α.

(i) $d_\alpha = 0$. In this case we have

$$R_\alpha \prec \Phi^{s_\alpha/2} \prod_{(k,\alpha)\in H} N^{-3/2} \prod_{(k,\alpha)\in F} N^{-1} = \Phi^{s_\alpha/2}\big(N^{-1}\big)^{h_\alpha+f_\alpha} N^{-h_\alpha/2}$$
$$\prec \Phi^{s_\alpha/2} N^{-1} N^{-h_\alpha/2} \prec \Phi^{\sum_{k=1}^h h_{k\alpha}/2+\sum_{k=h+1}^n e_{k\alpha}^{(o)}/2} N^{-1} N^{-h_\alpha/2},$$

where in the third step we used $h_l + f_l > 0$, and in the fourth step we used

$$s_\alpha \geqslant \sum_k e_{k\alpha}^{(o)} \geqslant \sum_{k=1}^h h_{k\alpha} + \sum_{k=h+1}^r e_{k\alpha}^{(o)},$$

where we used that $e_{k\alpha}^{(o)} \geqslant h_{k\alpha}$ for $1 \leqslant k \leqslant h$ (recall that if $(k,\alpha) \in H$, then $e_{k\alpha_k}$ is odd and hence at least one of the edges must come from the off-diagonal S).

(ii) $d_\alpha \neq 0$, $h_\alpha = 1$ and $f_\alpha = 0$. Then there is only one k such that $e_{k\alpha} > 0$ and $s_\alpha = e_{k\alpha}$ is odd. Hence we have $s_\alpha/2 \geqslant d_\alpha + 1/2$ and we can bound R_α as

$$R_\alpha \prec \Phi^{\frac{1}{2}s_\alpha - d_\alpha}\big(N^{-1}\big)^{h_\alpha+f_\alpha} N^{-h_\alpha/2} \prec \Phi^{1/2} N^{-1} N^{-h_\alpha/2}$$
$$= \Phi^{\sum_{k=1}^h h_{k\alpha}/2+\sum_{k=h+1}^n e_{k\alpha}^{(o)}/2} N^{-1} N^{-h_\alpha/2},$$

where in the last step we used

$$1 = \sum_{k=1}^h h_{k\alpha} + \sum_{k=h+1}^r e_{k\alpha}^{(o)},$$

since all the summands except one $h_{k\alpha}$ are 0.

(iii) $d_\alpha \neq 0$, $h_\alpha = 0$ and $f_\alpha = 1$. Then there is only one k such that $e_{k\alpha} > 0$ and $s_\alpha = e_{k\alpha}$. Thus the (α, α) edges are expanded from the diagonal S variables (otherwise α must connect to at least two different k's), which implies $\frac{1}{2}s_\alpha - d_\alpha = \frac{1}{2}e_{k\alpha}^{(o)}$. Then we can bound R_α by

$$R_\alpha \prec \Phi^{\frac{1}{2}s_\alpha - d_\alpha}\big(N^{-1}\big)^{h_\alpha+f_\alpha} N^{-h_\alpha/2} = \Phi^{\sum_k e_{k\alpha}^{(o)}/2} N^{-1} N^{-h_\alpha/2}$$
$$\prec \Phi^{\sum_{k=1}^h h_{k\alpha}/2+\sum_{k=h+1}^n e_{k\alpha}^{(o)}/2} N^{-1} N^{-h_\alpha/2}$$

where, as in Case (i), we used $e_{k\alpha}^{(o)} \geqslant h_{k\alpha}$ for $1 \leqslant k \leqslant h$.

(iv) $d_\alpha \neq 0$ and $h_\alpha + f_\alpha \geqslant 2$. Then using $s_\alpha \geqslant 2d_\alpha$, $q \prec \Phi^{1/2}$ and $N^{-1/2} \prec \Phi$, we get that

$$\begin{aligned} R_\alpha &\prec \prod_{(k,\alpha)\in H} \Phi^{e_{k\alpha}/2-3/2} N^{-3/2} \prod_{(k,\alpha)\in F} \Phi^{e_{k\alpha}/2-1} N^{-1} \\ &\prec \prod_{(k,\alpha)\in H} \Phi^{e_{k\alpha}/2-1/2} N^{-1} \prod_{(k,\alpha)\in F} \Phi^{e_{k\alpha}/2} N^{-1/2} \\ &= \Phi^{(s_\alpha - h_\alpha)/2} N^{-(h_\alpha+f_\alpha)/2} N^{-h_\alpha/2} \leqslant \Phi^{(s_\alpha-h_\alpha)/2} N^{-1} N^{-h_\alpha/2} \\ &\leqslant \Phi^{\sum_{k=1}^h h_{k\alpha}/2 + \sum_{k=h+1}^r e_{k\alpha}^{(o)}/2} N^{-1} N^{-h_\alpha/2} \end{aligned}$$

where in the last step we used the definitions of s_α and h_α, $e_{k\alpha} \geqslant 2h_{k\alpha}$ for $1 \leqslant k \leqslant h$ (since $e_{k\alpha} \geqslant 3$ whenever $h_{k\alpha} = 1$), and $h_{k\alpha} = 0$ for $k \geqslant h+1$.

Combining the above four cases, we obtain that

$$|\mathbb{E}R| \prec \prod_{\alpha=1}^m R_\alpha \prec N^{-m} N^{-\frac{1}{2}\sum_\alpha h_\alpha} \Phi^{\sum_\alpha \left(\sum_{k=1}^h h_{k\alpha}/2 + \sum_{k=h+1}^r e_{k\alpha}^{(o)}/2\right)}.$$

Recall that $\sum_\alpha h_\alpha = h$. Then to prove (3.4.24), it remains to show that

$$\sum_\alpha \left(\sum_{k=1}^h h_{k\alpha} + \sum_{k=h+1}^r e_{k\alpha}^{(o)} \right) \geqslant 4p. \tag{3.4.28}$$

For $k = 1, ..., h$, using (3.4.27) and (3.4.18) we get that

$$\sum_{\alpha=1}^m h_{k\alpha} = 1 = \deg(b_k, P).$$

For $k = h+1, ..., n$, using (3.4.26) and (3.4.22) we get that

$$\sum_{\alpha=1}^m e_{k\alpha}^{(o)} = \deg_o(b_k, Q) \geqslant \deg(b_k, P).$$

With (3.4.18), we then conclude (3.4.28), which finishes our proof.

CHAPTER 4

Local circular law for deformed non-Hermitian random matrices

4.1 Introduction

Circular law for non-Hermitian random matrices.

The study of the eigenvalue spectral of non-Hermitian random matrices goes back to the celebrated paper [46] by Ginibre, where he calculated the joint probability density for the eigenvalues of non-Hermitian random matrix with independent complex Gaussian entries. The joint density distribution is integrable with an explicit kernel (see [46, 67]), which allowed him to derive the circular law for the eigenvalues. For the Gaussian random matrix with real entries, the joint distribution of the eigenvalues is more complicated but still integrable, which leads to a proof of the circular law as well [16, 27, 43, 89].

For the random matrix with non-Gaussian entries, there is no explicit formula for the joint distribution of the eigenvalues. However, in many cases the eigenvalue spectrum of the non-Gaussian random matrices behaves similarly to the Gaussian case as $N \to \infty$, known as the universality phenomena. A key step in this direction is made by Girko in [47], where he partially proved the circular law for non-Hermitian matrices with independent entries. The crucial insight of this chapter is the *Hermitization technique*, which allowed Girko to translate the convergence of complex empirical measures of a non-Hermitian matrix into the convergence of logarithmic transforms for a family of Hermitian matrices, or, to be more precise,

$$\operatorname{Tr}\log[(X-z)^*(X-z)] = \log\left[\det((X-z)^*(X-z))\right], \tag{4.1.1}$$

with X being the non-Hermitian random matrix and $z \in \mathbb{C}$. Due to the singularity of the log function at 0, the small eigenvalues of $(X-z)^*(X-z)$ play a special role. The estimate on the smallest singular value of $X-z$ was not obtained in [47], but the gap was remedied later in a series of paper. Bai [6, 8] analyzed the ESD of $(X-z)^*(X-z)$ through its Stieltjes transform and handled the logarithmic singularity by assuming bounded density and bounded high moments for the entries of X. Lower bounds on the smallest singular values were given by Rudelson and Vershynin [80, 81], and subsequently by Tao and Vu [91], Pan and Zhou [74] and Gőtze and Tikhomirov [49] under weakened moments and smoothness assumptions. The final result was presented in [96], where the circular law is proved under the optimal L^2 assumption. These papers studied the circular law in the global regime, i.e. the convergence of ESD on subsets containing ηN eigenvalues for some small constant $\eta > 0$. Later in a series of papers [18, 19, 111], Bourgade, Yau and Yin proved the $l$ocal version of the circular law up to the optimal scale $N^{-1/2+\varepsilon}$ under the assumption that the distributions of the matrix entries satisfy a uniform sub-exponential decay condition. In [95], the local universality was proved by Tao and Vu under the assumption of first four moments matching the moments of a Gaussian random variable.

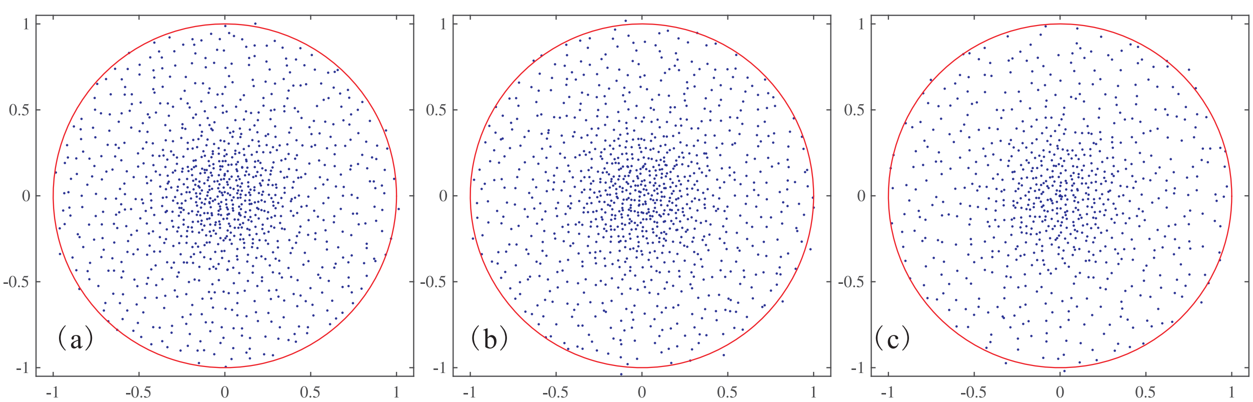


Figure 4.1: The eigenvalue distribution of the product TX of a deterministic $N \times M$ matrix T with a Gaussian random $M \times N$ matrix X. The entries of X have zero mean and variance $(N \wedge M)^{-1}$, and TT^* has $0.5(N \wedge M)$ eigenvalues as $2/17$ and $0.5(N \wedge M)$ eigenvalues as $32/17$. (a) $N = M = 1000$. (b) $N = 1000$, $M = 2000$. (c) $N = 1500$, $M = 750$.

In this chapter, we study the ESD of the product of a deterministic $N \times M$ matrix T

with a random $M \times N$ matrix X, where we assume $N \sim M$. In Figure 4.1, we plot the eigenvalue distribution of TX when T has two distinct singular values (except the trivial zero singular values). The goal of this chapter is to prove a local circular law for the ESD of TX at any point z away from the boundary circle. Following the idea in [18], the key ingredients for the proof are (a) the upper bound for the largest singular value of $TX - z$, (b) the lower bound for the least singular value of $TX - z$, and (c) rigidity of the singular values of $TX - z$. The upper bound for the largest singular value can be obtained by controlling the norm of X through a standard large deviation estimate (see e.g. [23, 64, 82]) or by studying the eigenvalue rigidity of X^*X (see e.g. [14] and (4.2.63)). The lower bound for the least singular value of $TX - z$ will be proved in the Appendix B. The bulk of this chapter is devoted to establishing (c).

Basic ideas

To obtain the rigidity of the singular values of $TX - z$, we study the ESD of $Q := (TX - z)^*(TX - z)$ using Stieltjes transform as in [18]. We normalize X so that its entries have variance $(N \wedge M)^{-1}$. Then Q is an $N \times N$ Hermitian matrix with eigenvalues being typically of order 1. We denote its resolvent by $R(w) := (Q - w)^{-1}$, where $w = E + \mathrm{i}\eta$ is a spectral parameter with positive imaginary part η. The key to the proof is again an averaged local law for $R(w)$ of the form

$$N^{-1}\mathrm{Tr}\, R(w) \approx m_c(w), \qquad (4.1.2)$$

where m_c is the Stieltjes transform of the asymptotic eigenvalue density. In [18], such a local law for the resolvent of $(X - z)^*(X - z)$ was established to prove the local circular law.

In generalizing the proof in [18] to our setting, a main difficulty is that the entries of TX are not independent. We will again use the self-consistent comparison method as in Section 2.7. For definiteness, we assume $N = M$ for now, and let T be a square matrix with singular decomposition $T = UDV$. For a Gaussian $X \equiv X^{Gauss}$, we have $VX^{Gauss}U \overset{d}{=} \tilde{X}^{Gauss}$, where $\tilde{X}$ is another Gaussian random matrix. Then for the determinant in (4.1.1), we have

$$\det(TX^{Gauss} - z) = \det(DVX^{Gauss}U - z) \overset{d}{=} \det(D\tilde{X}^{Gauss} - z). \qquad (4.1.3)$$

The problem is now reduced to the study of the singular values of $D\tilde{X}^{Gauss} - z$, which has independent entries. Notice that since D is not scalar, the entries of $D\tilde{X}^{Gauss}$ are not identically distributed, which will make our proof much more complicated than the one in [18].

Now we briefly outline the three steps to establish the averaged local law for $Q = (TX - z)^*(TX - z)$: (A) the entrywise local law and averaged local law when T is diagonal (Theorem 4.2.22); (B) the anisotropic local law when T is diagonal (Theorem 4.2.22); (C) the anisotropic local law and averaged local law when T is a general (rectangular) matrix (Theorem 4.2.23). In performing Step (A), our proof is basically based on the methods in [18]. However, our multi-variable self-consistent equations and their solutions are much more complicated here. Thus a key part of the proof is to establish some basic properties of the asymptotic eigenvalue density and prove the stability of the self-consistent equations under small perturbations. These work need some new ideas and analytic techniques (see Appendix A). In performing Step (B), we applied and extended the polynomialization method developed in [14, section 5]. Finally, as remarked above, (B) implies the anisotropic local law and averaged local law for Gaussian X and general T. Based on this fact we perform Step (C) using a self-consistent comparison argument of [59] as in Section 2.7. With the averaged local law proved in Step (C), we can obtain a generalized (inhomogeneous) local circular law for TX. In general, the averaged local law we get is up to the non-optimal scale $\eta \gg (N \wedge M)^{-1/2}$. As a result, we can only prove the local circular law for TX up to the scale $(N \wedge M)^{-1/4+\varepsilon}$. One observation is that the non-optimal averaged local law can lead to the optimal local circular law for TX outside the unit circle (i.e. $|z| > 1$) (see Section 4.2.3). To prove the optimal local circular law inside the unit circle (i.e. $|z| < 1$), we need the optimal averaged local law up to the scale $\eta \gg (N \wedge M)^{-1}$, which can be obtained under the extra assumption that the entries of X have vanishing third moments.

4.2 Main result

In this section, we state and prove the main result of this chapter. In Section 4.2.1, we first define the asymptotic eigenvalue density ρ_{2c} of $Q = (TX - z)^*(TX - z)$, and then state the main theorem—Theorem 4.2.6—of this chapter. Its proof depends crucially on local laws of the resolvent of Q, which are presented in Section 4.2.2. In Section 4.2.3, we prove Theorem 4.2.6 based on the local laws.

We want to understand the local statistics of the eigenvalues of $TX - zI$, where T is a deterministic $N \times M$ matrix, X is a random $M \times N$ matrix, $z \in \mathbb{C}$ and I is the $N \times N$ identity matrix. We assume $M \sim N$, i.e.

$$\tau \leqslant \frac{M}{N} \leqslant \tau^{-1} \tag{4.2.1}$$

for some small constant $\tau > 0$. We assume the entries $X_{i\mu}$ of X are independent (not necessarily identically distributed) random variables satisfying

$$\mathbb{E}\, X_{i\mu} = 0, \quad \mathbb{E}\, |X_{i\mu}|^2 = \frac{1}{N \wedge M}, \tag{4.2.2}$$

for all $1 \leqslant i \leqslant M,\ 1 \leqslant \mu \leqslant N$. For definiteness, in this chapter we only focus on the case where all the X entries are real. However, our results and proofs also hold, after minor changes, in the complex case if we assume in addition $\mathbb{E}X_{i\mu}^2 = 0$ for $X_{i\mu} \in \mathbb{C}$. We assume that for all $p \in \mathbb{N}$, there is an N-independent constant C_p such that

$$\mathbb{E}\, |\sqrt{N \wedge M} X_{i\mu}|^p \leqslant C_p \tag{4.2.3}$$

for all $1 \leqslant i \leqslant M,\ 1 \leqslant \mu \leqslant N$. We define $\Sigma := TT^*$, and assume the eigenvalues of Σ satisfy that

$$\tau^{-1} \geqslant \sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_{N \wedge M} \geqslant \tau \tag{4.2.4}$$

and all other eigenvalues are 0. Furthermore, we can normalize T by multiplying a scalar such that

$$\frac{1}{N \wedge M} \sum_{i=1}^{N \wedge M} \sigma_i = 1. \tag{4.2.5}$$

We summarize our basic assumptions here for future reference.

Assumption 4.2.1. *We suppose that (4.2.1), (4.2.2), (4.2.3), (4.2.4) and (4.2.5) hold.*

4.2.1 Main theorem

To state the main theorem, we need to define the asymptotic eigenvalue density function for Q. We first introduce the self-consistent equations, then the asymptotic eigenvalue density will be closely related to their solutions. Let

$$\rho_\Sigma := \frac{1}{N \wedge M} \sum_{i=1}^{N\wedge M} \delta_{\sigma_i} \tag{4.2.6}$$

denote the empirical spectral density of Σ. Let $n := |\operatorname{supp} \rho_\Sigma|$ be the number of distinct nonzero eigenvalues of Σ, which are denoted as

$$\tau^{-1} \geqslant s_1 > s_2 > \cdots > s_n \geqslant \tau. \tag{4.2.7}$$

Let l_i be the multiplicity of s_i. By (4.2.5), l_i and s_i satisfy the normalization conditions

$$\frac{1}{N \wedge M}\sum_{i=1}^{n} l_i = 1, \quad \frac{1}{N \wedge M}\sum_{i=1}^{n} l_i s_i = 1. \tag{4.2.8}$$

For each $w \in \mathbb{C}_+$, we define the self-consistent equations of (m_1, m_2) as

$$\frac{1}{m_2} = -w(1+m_1) + \frac{|z|^2}{1+m_1}, \tag{4.2.9}$$

$$m_1 = \frac{1}{N}\sum_{i=1}^{n} l_i s_i \left[-w\left(1 + s_i m_2\right) + \frac{|z|^2}{1+m_1} \right]^{-1}. \tag{4.2.10}$$

If we plug (4.2.9) into (4.2.10), we get the self-consistent equation for m_1 only:

$$m_1 = \frac{1}{N}\sum_{i=1}^{n} l_i s_i \left[-w\left(1 + \frac{s_i}{-w(1+m_1) + \frac{|z|^2}{1+m_1}}\right) + \frac{|z|^2}{1+m_1} \right]^{-1}. \tag{4.2.11}$$

The next lemma states that the solution to the functional equation (4.2.11) in $\mathbb{C}_+$ is unique if z is away from the unit circle. It will be proved in Appendix A.3.

Lemma 4.2.2. *Fix $z \in \mathbb{C}$ such that $|z| \neq 1$. For $w \in \mathbb{C}_+$, there exists at most one analytic function $m_{1c,z,\Sigma}(w) : \mathbb{C}_+ \to \mathbb{C}_+$ such that (4.2.11) holds and $w m_{1c,z,\Sigma}(w) \in \mathbb{C}_+$. Moreover, $m_{1c,z,\Sigma,N}(w)$ is the Stieltjes transform of a positive integrable function ρ_{1c} with compact support in $[0, \infty)$.*

We shall abbreviate $m_{1c}(w) := m_{1c,z,\Sigma}(w)$. We also define $m_{2c}(w) := m_{2c,z,\Sigma}(w)$ by taking $m_1 = m_{1c}(w)$ in (4.2.9). Obviously, m_{2c} is also an analytic function of w. Moreover, for any $w \in \mathbb{C}_+$ we can verify that $m_{2c}(w), wm_{2c}(w) \in \mathbb{C}_+$ by using (4.2.9) and that $m_{1c}, wm_{1c} \in \mathbb{C}_+$. We define two functions on $\mathbb{R}$ as

$$\rho_{1,2c}(x) = \frac{1}{\pi} \lim_{\eta \searrow 0} \operatorname{Im} m_{1,2c}(x + \mathrm{i}\eta), \quad x \in \mathbb{R}. \tag{4.2.12}$$

It is easy to see that $\rho_{1,2c} \geqslant 0$ and $\operatorname{supp}(\rho_{1,2c}) \subseteq [0, \infty)$. Moreover, $\operatorname{supp} \rho_{2c} = \operatorname{supp} \rho_{1c}$ by (4.2.9). We shall call ρ_{2c} the asymptotic eigenvalue density of $Q = (TX - z)^*(TX - z)$. Since $\operatorname{Im}(wm_{2c}) \geqslant 0$, we have

$$\operatorname{Im}\left[-w\left(1 + s_i m_{2c}\right) + \frac{|z|^2}{1 + m_{1c}}\right] \leqslant -\operatorname{Im} w,$$

and (4.2.10) gives $|m_{1c}| \leqslant 1/\operatorname{Im} w \to 0$ as $\operatorname{Im} w \to \infty$. Similarly, $|m_{2c}| \leqslant 1/\operatorname{Im} w \to 0$ as $\operatorname{Im} w \to \infty$. Thus $m_{1,2c}(w)$ is indeed the Stieltjes transform of $\rho_{1,2c}$:

$$m_{1,2c}(w) = \int_{\mathbb{R}} \frac{\rho_{1,2c}(x)}{x - w} dx. \tag{4.2.13}$$

We now state the basic properties of ρ_{1c} and ρ_{2c}, which can be obtained by studying the solutions $m_{1,2c}(w)$ to the self-consistent equations (4.2.9) and (4.2.11) when $w \in (0, \infty)$. Here we extend the definition of $m_{1,2c}$ continuously down to the real axis by setting

$$m_{1,2c}(x) = \lim_{\eta \searrow 0} m_{1,2c}(x + \mathrm{i}\eta), \quad x \in \mathbb{R}.$$

As a convention, for $w \in \overline{\mathbb{C}_+}$, we take $\sqrt{w}$ to be the branch with positive imaginary part. Denote $m := \sqrt{w}(1 + m_1)$ and $m_c := \sqrt{w}(1 + m_{1c})$. Equation (4.2.11) then becomes

$$f(\sqrt{w}, m) = 0, \tag{4.2.14}$$

where

$$f(\sqrt{w}, m) = -\sqrt{w} + m + \frac{1}{N} \sum_{i=1}^{n} l_i s_i \frac{m(m^2 - |z|^2)}{\sqrt{w} m^3 - (s_i + |z|^2) m^2 - \sqrt{w}|z|^2 m + |z|^4}. \tag{4.2.15}$$

The following lemma gives the basic structure of $\operatorname{supp} \rho_{1,2c}$. Its proof will be given in Appendix A.1.

Lemma 4.2.3. *Fix $\tau \leqslant ||z|^2 - 1| \leqslant \tau^{-1}$. The support of $\rho_{1,2c}$ is a union of connected components:*

$$\operatorname{supp} \rho_{1,2c} \cap (0, +\infty) = \left(\bigcup_{1 \leqslant k \leqslant L} [e_{2k}, e_{2k-1}] \right) \cap (0, \infty), \tag{4.2.16}$$

where $L \equiv L(n) \in \mathbb{N}$ and $C_1 \tau^{-1} \geqslant e_1 > e_2 > \ldots > e_{2L} \geqslant 0$ for some constant $C_1 > 0$ that does not depend on τ. If $|z|^2 \leqslant 1 - \tau$, we have $e_{2L} = 0$; if $1 + \tau \leqslant |z|^2 \leqslant 1 + \tau^{-1}$, we have $e_{2L} \geqslant \varepsilon(\tau)$ for some constant $\varepsilon(\tau) > 0$. Moreover, for every $e_i > 0$, there exists a unique $m_c(e_i)$ such that

$$\partial_m f(\sqrt{e_i}, m_c(e_i)) = 0. \tag{4.2.17}$$

We shall call e_i the edges of $\rho_{1,2c}$. For any $w \in (0, \infty)$ and $1 \leqslant i \leqslant n$, the cubic polynomial $\sqrt{w} m^3 - (s_i + |z|^2) m^2 - \sqrt{w} |z|^2 m + |z|^4$ in (4.2.15) has three distinct roots $a_i(w) > 0$, $b_i(w) > 0$ and $-c_i(w) < 0$ (see Lemma A.1.1). Our next assumption on ρ_Σ and $|z|$ takes the form of the following regularity conditions. They take the similar form as in Definition 3.1.1.

Definition 4.2.4 (Regularity)**.** *Fix $\tau \leqslant ||z|^2 - 1| \leqslant \tau^{-1}$.*

(i) We say that the edge $e_k \neq 0$, $k = 1, \ldots, 2L$, is regular if

$$\min_{1 \leqslant i \leqslant n} \{ |m_c(e_k) - a_i(e_k)|, |m_c(e_k) - b_i(e_k)|, |m_c(e_k) + c_i(e_k)| \} \geqslant \varepsilon, \tag{4.2.18}$$

and

$$\left| \partial_m^2 f(\sqrt{e_k}, m_c(e_k)) \right| \geqslant \varepsilon, \tag{4.2.19}$$

for some small constant $\varepsilon > 0$. In the case $|z|^2 \leqslant 1 - \tau$, we always call $e_{2L} = 0$ a regular edge.

(ii) We say that the bulk components $[e_{2k}, e_{2k-1}]$ is regular if for any fixed $\tau' > 0$ there exists a constant $c(\tau, \tau') > 0$ such that the density of ρ_{1c} in $[e_{2k} + \tau', e_{2k-1} - \tau']$ is bounded from below by c.

Remark 4.2.5. The regularity conditions in Definition 4.2.4 are stable under perturbations of $|z|$ and ρ_Σ. In particular, fix ρ_Σ, suppose the regularity conditions are satisfied at $z = z_0$ with $\tau \leqslant ||z_0|^2 - 1| \leqslant \tau^{-1}$. Then for sufficiently small $c > 0$, the regularity conditions hold uniformly in $z \in \{z : ||z| - |z_0|| \leqslant c\}$. For a detailed discussion, see Remark A.3.3.

In the following, we denote the eigenvalues of TX by μ_j, $1 \leqslant j \leqslant N$. We are now ready to state our main theorem, i.e. the generalized local circular law for TX.

Theorem 4.2.6 (Local circular law for TX)**.** *Suppose Assumption 4.2.1 holds, and $\tau \leqslant ||z_0|^2 - 1| \leqslant \tau^{-1}$ for any N (z_0 can depend on N). Suppose ρ_Σ (defined in (4.2.6)) and $|z_0|$ are such that all the edges and bulk components of ρ_{1c} are regular in the sense of Definition 4.2.4. We assume in addition that each entry of X has a density bounded by N^{C_2} for some $C_2 > 0$. Let F be a smooth non-negative function which may depend on N, such that $\|F\|_\infty \leqslant C_1$, $\|F'\|_\infty \leqslant N^{C_1}$ and $F(z) = 0$ for $|z| \geqslant C_1$, for some constant $C_1 > 0$ independent of N. Let $F_{z_0,a}(z) = K^{2a}F(K^a(z - z_0))$, where $K := N \wedge M$. Then TX has $(N - K)$ trivial zero eigenvalues, and for the other eigenvalues μ_j, $1 \leqslant j \leqslant K$, we have*

$$\frac{1}{K}\sum_{j=1}^{K} F_{z_0,a}(\mu_j) - \frac{1}{\pi}\int F_{z_0,a}(z)\tilde{\chi}_{\mathbb{D}}(z)\mathrm{d}A(z) \prec K^{-1/2+2a}\|\Delta F\|_{L^1}, \tag{4.2.20}$$

for any $a \in (0, 1/4]$. Here

$$\tilde{\chi}_{\mathbb{D}}(z) := \frac{1}{4}\int_0^\infty (\log x)\Delta_z \rho_{2c}(x, z)\mathrm{d}x, \tag{4.2.21}$$

where $\rho_{2c} \equiv \rho_{2c,z,\Sigma}$ is defined in (4.2.12). If $1 + \tau \leqslant |z_0|^2 \leqslant 1 + \tau^{-1}$ or the entries of X have vanishing third moments,

$$\mathbb{E}X_{i\mu}^3 = 0, \quad 1 \leqslant i \leqslant M, 1 \leqslant \mu \leqslant N, \tag{4.2.22}$$

then we have the improved result

$$\frac{1}{K}\sum_{j=1}^{K} F_{z_0,a}(\mu_j) - \frac{1}{\pi}\int F_{z_0,a}(z)\tilde{\chi}_{\mathbb{D}}(z)\mathrm{d}A(z) \prec K^{-1+2a}\|\Delta F\|_{L^1}, \tag{4.2.23}$$

for any $a \in (0, 1/2]$. If the entries of X are identically distributed, then the bounded density condition is not necessary.

Remark 4.2.7. Note that $F_{z_0,a}(z) = K^{2a}F(K^a(z - z_0))$ is an approximate delta function obtained from rescaling F to the size of order K^{-a} around z_0. Thus (4.2.20) gives a generalized circular law up to scale $K^{-1/4+\varepsilon}$, while (4.2.23) gives a generalized circular law up to scale $K^{-1/2+\varepsilon}$. The $\tilde{\chi}_{\mathbb{D}}$ in (4.2.21) gives the distribution of the eigenvalues of TX. It is rotationally symmetric, because $\rho_{2c}(x, z)$ depends only on $|z|$ (see (4.2.9) and (4.2.10)). If $TT^* = 1$ or $T^*T = 1$ (i.e. all the nontrivial singular values of T are equal to 1), then $\tilde{\chi}_{\mathbb{D}}$ becomes the indicator function $\chi_{\mathbb{D}}$ on the unit disk $\mathbb{D}$, and we get the well-known local circular law for X (see [18] for the $T = I$ case). For a general T, we do not have much understanding of $\tilde{\chi}_{\mathbb{D}}$ so far. This will be one of the topics of the future study. Also, we have assumed that z is strictly away from the unit circle. Our proof may be extended to the $|z - 1| = o(1)$ case if we have a better understanding of the solutions $m_{1,2c}$.

Remark 4.2.8. As explained in the Introduction, the basic strategy of this chapter is first to prove the anisotropic local law for the resolvent of Q when X is Gaussian, and then to get the anisotropic local law for a general X through a comparison with the Gaussian case. Without (4.2.22), our comparison arguments cannot give the anisotropic local law up to the optimal scale, so we can only prove the weaker bound (4.2.20). We will try to remove this assumption in the future.

Remark 4.2.9. If the entries of X are identically distributed, then it was proved in Appendix B that the smallest singular value of $TX - z$ is larger than $N^{-1-\varepsilon}$ with high probability for any $\varepsilon > 0$. Otherwise, we need the extra bounded density condition, which is only used in Lemma 4.2.26 to get a lower bound for the smallest singular value of $TX - z$.

We conclude this section with two examples verifying the regularity conditions of Definition 4.2.4.

Example 4.2.10 (Bounded number of distinct eigenvalues). We suppose that n is fixed, and that $s_1, \ldots, s_n$ and $\rho_\Sigma(\{s_1\}), \ldots, \rho_\Sigma(\{s_n\})$ all converge as $N \to \infty$. We suppose that $\lim_N e_k > \lim_N e_{k+1}$ for all k, and furthermore for all e_k we have $\partial_m^2 f(\sqrt{e_k}, m_c(e_k)) \neq 0$. Then it is easy to check that all the edges and bulk components are regular in the sense of Definition 4.2.4 for small enough ε.

Example 4.2.11 (Continuous limit). We suppose ρ_Σ is supported in some interval $[a, b] \subset (0, \infty)$, and that ρ_Σ converges in distribution to some measure ρ_∞ that is absolutely continuous and whose density satisfies $\tau \leqslant \mathrm{d}\rho_\infty(E)/\mathrm{d}E \leqslant \tau^{-1}$ for $E \in [a, b]$. Then there are only a small number (which is independent of n) of connected components for $\operatorname{supp} \rho_{1c}$, and all the edges and bulk components are regular; see Remark A.1.6.

4.2.2 Hermitization and local laws

In the following, we use the notation

$$Y \equiv Y_z := TX - zI, \tag{4.2.24}$$

where I is the identity matrix. Following Girko's Hermitization technique [47], the first step in proving the local circular law is to understand the local statistics of singular values of Y. In this subsection, we present the main local estimates concerning the resolvents $(YY^* - w)^{-1}$ and $(Y^*Y - w)^{-1}$. These results will be used later to prove Theorem 4.2.6. First as in Section 2.4, we shall use a linearization trick for YY^* and Y^*Y.

Definition 4.2.12 (Index sets)**.** *We define the index sets*

$$\mathcal{I}_1 := \{1, ..., N\}, \quad \mathcal{I}_1^M := \{1, \ldots, M\}, \quad \mathcal{I}_2 := \{N+1, ..., 2N\},$$

and

$$\mathcal{I} := \mathcal{I}_1 \cup \mathcal{I}_2, \quad \mathcal{I}^M := \mathcal{I}_1^M \cup \mathcal{I}_2 .$$

We will consistently use the latin letters $i, j \in \mathcal{I}_1$ or $\mathcal{I}_1^M$, greek letters $\mu, \nu \in \mathcal{I}_2$, and $s, t \in \mathcal{I}$. We label the indices of the matrices according to

$$X = (X_{i\mu} : i \in \mathcal{I}_1^M, \mu \in \mathcal{I}_2), \quad T = (T_{ij} : i \in \mathcal{I}_1, j \in \mathcal{I}_1^M).$$

When $M = N$, we always identify $\mathcal{I}_1^M$ with $\mathcal{I}_1$. For $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_2$, we introduce the notations $\bar{i} := i + N \in \mathcal{I}_2$ and $\bar{\mu} := \mu - N \in \mathcal{I}_1$.

Definition 4.2.13 (Groups)**.** *For an $\mathcal{I} \times \mathcal{I}$ matrix A, we define the 2×2 matrices $A_{[ij]}$ as*

$$A_{[ij]} = \begin{pmatrix} A_{ij} & A_{i\bar{j}} \\ A_{\bar{i}j} & A_{\bar{i}\bar{j}} \end{pmatrix}. \tag{4.2.25}$$

We shall call $A_{[ij]}$ a diagonal group if $i = j$, and an off-diagonal group otherwise .

Definition 4.2.14 (Linearizing block matrix)**.** *For $w := E + \mathrm{i}\eta \in \mathbb{C}_+$, we define the $\mathcal{I} \times \mathcal{I}$ matrix*

$$H(w) \equiv H(T, X, z, w) := \begin{pmatrix} -wI & w^{1/2}Y \\ w^{1/2}Y^* & -wI \end{pmatrix}, \tag{4.2.26}$$

where we take the branch of $\sqrt{w}$ with positive imaginary part. Define the $\mathcal{I} \times \mathcal{I}$ matrix

$$G(w) \equiv G(T, X, z, w) := H(w)^{-1}, \tag{4.2.27}$$

as well as the $\mathcal{I}_1 \times \mathcal{I}_1$ and $\mathcal{I}_2 \times \mathcal{I}_2$ matrices

$$G_L(w) = (YY^* - w)^{-1}, \quad G_R(w) = (Y^*Y - w)^{-1}. \tag{4.2.28}$$

Throughout the rest of this chapter, we frequently omit the argument w from our notations.

By Schur's complement formula, it is easy to see that

$$G(w) = \begin{pmatrix} G_L & w^{-1/2}G_L Y \\ w^{-1/2}Y^* G_L & w^{-1}Y^* G_L Y - w^{-1}I \end{pmatrix} = \begin{pmatrix} w^{-1}YG_R Y^* - w^{-1}I & w^{-1/2}YG_R \\ w^{-1/2}G_R Y^* & G_R \end{pmatrix}. \tag{4.2.29}$$

Therefore a control of G immediately yields a control of the resolvents G_L and G_R.

In the following, we only consider the $N \leqslant M$ case. The $N > M$ case, as we will see, will be built easily upon $N \leqslant M$ case. We introduce a deterministic matrix Π, which is the asymptotic limit of G.

Definition 4.2.15 (Deterministic limit of G)**.** *Suppose $N \leqslant M$ and T has a singular decomposition*

$$T = U\bar{D}V, \quad \bar{D} = (D, 0), \tag{4.2.30}$$

where $D = \mathrm{diag}(d_1, d_2, \ldots, d_N)$ is a diagonal matrix. Define $\pi_{[i]c}$ to be the 2×2 matrix such that

$$\left(\pi_{[i]c}\right)^{-1} = \begin{pmatrix} -w(1 + |d_i|^2 m_{2c}) & -w^{1/2}z \\ -w^{1/2}\bar{z} & -w(1 + m_{1c}) \end{pmatrix}. \tag{4.2.31}$$

Let Π_d be the $2N \times 2N$ matrix with $(\Pi_d)_{[ii]} = \pi_{[i]c}$ and all other entries being zero. Define

$$\Pi \equiv \Pi(\Sigma, z, w) := \begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix} \Pi_d \begin{pmatrix} U^* & 0 \\ 0 & U^* \end{pmatrix} = \begin{pmatrix} -(1 + m_{1c})A(\Sigma) & w^{-1/2}zA(\Sigma) \\ w^{-1/2}\bar{z}A(\Sigma) & -(1 + m_{2c}\Sigma)A(\Sigma) \end{pmatrix}, \tag{4.2.32}$$

where $\Sigma = TT^*$ *and* $A(\Sigma) = \left[w(1+m_{2c}\Sigma)(1+m_{1c}) - |z|^2\right]^{-1}$.

Definition 4.2.16 (Averaged variables)**.** *Suppose* $N \leqslant M$*. Define the averaged random variables*

$$m_1 := \frac{1}{N}\sum_{i\in\mathcal{I}_1}\left(\bar{\Sigma}G\right)_{ii}, \quad m_2 := \frac{1}{N}\sum_{\mu\in\mathcal{I}_2}\left(\bar{\Sigma}G\right)_{\mu\mu}, \tag{4.2.33}$$

where

$$\bar{\Sigma} := \begin{pmatrix} \Sigma & 0 \\ 0 & I \end{pmatrix}. \tag{4.2.34}$$

Define $\pi_{[i]}$ *to be the* 2×2 *matrix such that*

$$\left(\pi_{[i]}\right)^{-1} = \begin{pmatrix} -w(1+|d_i|^2 m_2) & -w^{1/2}z \\ -w^{1/2}\bar{z} & -w(1+m_1) \end{pmatrix}. \tag{4.2.35}$$

Remark 4.2.17*.* Note that under the above definition we have

$$m_2 = \frac{1}{N}\operatorname{Tr} G_R = \frac{1}{N}\operatorname{Tr} G_L,$$

which is the Stieltjes transform of the empirical eigenvalue density of YY^* and Y^*Y. Moreover, we will see from the proof that $m_{1,2c}$ are the almost sure limits of $m_{1,2}$ as $N\to\infty$ with

$$m_{1c} = \frac{1}{N}\sum_{i\in\mathcal{I}_1}\left(\bar{\Sigma}\Pi\right)_{ii}, \quad m_{2c} = \frac{1}{N}\sum_{\mu\in\mathcal{I}_2}\left(\bar{\Sigma}\Pi\right)_{\mu\mu}. \tag{4.2.36}$$

The following two propositions summarize the properties of $\rho_{1,2c}$ and $m_{1,2c}$ that are needed to understand the main results in this section. They will be proved in Appendix A. In Fig. 4.2, we plot ρ_{2c} for the example from Fig. 4.1 for different values of z.

Proposition 4.2.18 (Basic properties of $\rho_{1,2c}$)**.** *The density* ρ_{1c} *is compactly supported in* $[0,\infty)$ *and the following properties regarding* ρ_{1c} *hold.*

(i) The support of ρ_{1c} *is* $\bigcup_{1\leqslant k\leqslant L(n)}[e_{2k}, e_{2k-1}]$ *where* $e_1 > e_2 > \ldots > e_{2L} \geqslant 0$*. If* $1+\tau \leqslant |z|^2 \leqslant 1+\tau^{-1}$*, then* $e_{2L}\geqslant\varepsilon$ *for some constant* $\varepsilon > 0$*; if* $|z|^2\leqslant 1-\tau$*, then* $e_{2L} = 0$*.*

(ii) Suppose $[e_{2k}, e_{2k-1}]$ *is a regular bulk component. For any* $\tau' > 0$*, if* $x\in[e_{2k}+\tau', e_{2k-1}-\tau']$*, then* $\rho_{1c}(x)\sim 1$*.*

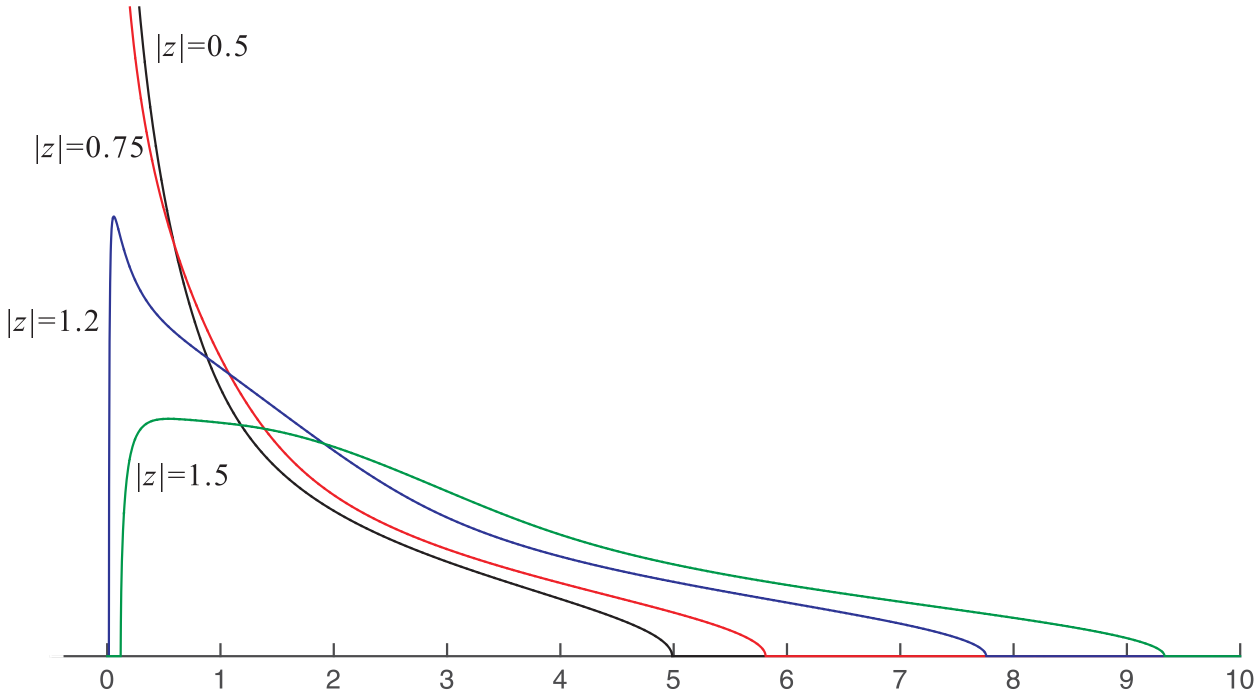


Figure 4.2: The densities $\rho_{2c}(x,z)$ when $|z| = 0.5,\ 0.75,\ 1.2,\ 1.5$. Here $\rho_\Sigma = 0.5\delta_{\sqrt{2/17}} + 0.5\delta_{4\sqrt{2/17}}$.

(iii) Suppose e_j is a nonzero regular edge. If j is even, then $\rho_{1c}(x) \sim \sqrt{x - e_j}$ as $x \to e_j$ from above. Otherwise if j is odd, then $\rho_{1c}(x) \sim \sqrt{e_j - x}$ as $x \to e_j$ from below.

(iv) If $|z|^2 \leqslant 1 - \tau$, then $\rho_{1c}(x) \sim x^{-1/2}$ as $x \searrow e_{2L} = 0$.

The same results also hold for ρ_{2c}. In addition, ρ_{2c} is a probability density.

Proposition 4.2.19. *The preceding proposition implies that, uniformly in w in any compact set of $\mathbb{C}_+$,*

$$|m_{1,2c}(w)| = O(|w|^{-1/2}). \tag{4.2.37}$$

Moreover, if $1 + \tau \leqslant |z|^2 \leqslant 1 + \tau^{-1}$, then $|m_{1,2c}(w)| \sim 1$ for w in any compact set of $\mathbb{C}_+$; if $|z|^2 \leqslant 1 - \tau$, then $|m_{1,2c}(w)| \sim |w|^{-1/2}$ for w in any compact set of $\mathbb{C}_+$.

We will consistently use the notation $E + \mathrm{i}\eta$ for the spectral parameter w. In this chapter, we regard the quantities $E(w)$ and $\eta(w)$ as functions of w and usually omit the argument w. In the following we would like to define several spectral domains of w that will be used in the proof.

Definition 4.2.20 (Spectral domains)**.** *Fix a small constant $\zeta > 0$ which may depend on τ. The spectral parameter w is always assumed to be in the fundamental domain*

$$\mathbf{D} \equiv \mathbf{D}(\zeta, N) := \{w \in \mathbb{C}_+ : \zeta e_{2L} \leqslant E \leqslant \zeta^{-1}, N^{-1+\zeta}|m_{2c}|^{-1} \leqslant \eta \leqslant \zeta^{-1}\}, \tag{4.2.38}$$

unless otherwise indicated. Given a regular edge e_k, we define the subdomain

$$\mathbf{D}_k^e \equiv \mathbf{D}_k^e(\zeta, \tau', N) := \{w \in \mathbf{D}(\zeta, N) : |E - e_k| \leqslant \tau', E \geqslant 0\}. \tag{4.2.39}$$

Corresponding to a regular bulk component $[e_{2k}, e_{2k-1}]$, we define the subdomain

$$\mathbf{D}_k^b \equiv \mathbf{D}_k^b(\zeta, \tau', N) := \{w \in \mathbf{D}(\zeta, N) : E \in [e_{2k} + \tau', e_{2k-1} - \tau']\}. \tag{4.2.40}$$

For the component outside supp ρ_{1c}, *we define the subdomain*

$$\mathbf{D}^o \equiv \mathbf{D}^o(\zeta, \tau', N) := \{w \in \mathbf{D}(\zeta, N) : \operatorname{dist}(E, \operatorname{supp} \rho_{1c}) \geqslant \tau'\}. \tag{4.2.41}$$

We also need the following domain with large η,

$$\mathbf{D}_L \equiv \mathbf{D}_L(\zeta) := \{w \in \mathbb{C}_+ : 0 \leqslant E \leqslant \zeta^{-1}, \eta \geqslant \zeta^{-1}\}, \tag{4.2.42}$$

and the subdomain of $\mathbf{D} \cup \mathbf{D}_L$,

$$\widehat{\mathbf{D}} \equiv \widehat{\mathbf{D}}(\zeta, N) := \{w \in \mathbf{D}(\zeta, N) : \eta \geqslant N^{-1/2+\zeta}|m_{2c}|^{-1}\} \cup \mathbf{D}_L(\zeta). \tag{4.2.43}$$

We call $\mathbf{S}$ a regular domain if it is a regular $\mathbf{D}_k^e$ domain, a regular $\mathbf{D}_k^b$ domain, a $\mathbf{D}^o$ domain or a $\mathbf{D}_L$ domain.

Remark: In the definition of $\mathbf{D}$, we have suppressed the explicit w-dependence. Notice that when $|z|^2 < 1 - \tau$, since $|m_{2c}| \sim |w|^{-1/2}$ as $w \to 0$, we allow $\eta \sim |w| \sim N^{-2+2\zeta}$ in $\mathbf{D}$. In the definition of $\mathbf{D}_k^e$, the condition $E \geqslant 0$ is only useful for the edge at 0 when $|z|^2 \leqslant 1 - \tau$.

Now we are prepared to state the local laws satisfied by G defined in (4.2.27). Let

$$\Psi \equiv \Psi(w) := \sqrt{\frac{\operatorname{Im}(m_{1c} + m_{2c})}{N\eta}} + \frac{1}{N\eta} \tag{4.2.44}$$

be the deterministic control parameter.

Definition 4.2.21 (Local laws)**.** *Suppose $N \leqslant M$. Recall $G \equiv G(T, X, z, w)$ defined in (4.2.27) and $\Pi \equiv \Pi(\Sigma, z, w)$ defined in (4.2.32). Let $\mathbf{S}$ be a regular domain.*

(i) We say that the entrywise local law holds with parameters $(T, X, z, \mathbf{S})$ if

$$[G(T, X, z, w) - \Pi(\Sigma, z, w)]_{st} \prec \Psi(w) \tag{4.2.45}$$

uniformly in $w \in \mathbf{S}$ and $s, t \in \mathcal{I}$.

(ii) We say that the anisotropic local law holds with parameters $(T, X, z, \mathbf{S})$ if

$$G(T, X, z, w) - \Pi(\Sigma, z, w) = \mathrm{O}_{\prec}(\Psi(w)) \tag{4.2.46}$$

uniformly in $w \in \mathbf{S}$ (recall Definition 2.4.1 (ii)).

(iii) We say that the averaged local law holds with parameters $(T, X, z, \mathbf{S})$ if

$$|m_2(T, X, z, w) - m_{2c}(\Sigma, z, w)| \prec \frac{1}{N\eta} \tag{4.2.47}$$

uniformly in $w \in \mathbf{S}$.

The local laws for G with a general T will be built upon the following result with a diagonal T.

Theorem 4.2.22 (Local laws when T is diagonal)**.** *Fix $\tau \leqslant \left||z|^2 - 1\right| \leqslant \tau^{-1}$. Suppose Assumption 4.2.1 holds, $N = M$, and $T \equiv D := \operatorname{diag}(d_1, ..., d_N)$ is a diagonal matrix. Let $\mathbf{S}$ be a regular domain. Then the entrywise local law, anisotropic local law and averaged local law hold with parameters $(D, X, z, \mathbf{S})$.*

Now suppose that $N \leqslant M$ and T is an $N \times M$ matrix such that the eigenvalues of Σ satisfy (4.2.4) and (4.2.5). Consider the singular decomposition $T = U\bar{D}V$, where U is an $N \times N$ unitary matrix, V is an $M \times M$ unitary matrix and $\bar{D} = (D, 0)$ is an $N \times M$ matrix such that $D = \operatorname{diag}(d_1, d_2, \ldots, d_N)$. Then we have

$$TX - z = UDV_1X - z, \tag{4.2.48}$$

where V_1 is an $N\times M$ matrix and V_2 is an $(M-N)\times M$ matrix defined through $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$.

If $X = X^{Gauss}$ is Gaussian, then $V_1 X^{Gauss} \overset{d}{=} \tilde{X}^{Gauss} U^*$, where $\tilde{X}^{Gauss}$ is another $N \times N$ Gaussian random matrix. Then by the definition of G in (4.2.27),

$$G(T, X^{Gauss}, z, w) \overset{d}{=} \begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix} G(D, \tilde{X}^{Gauss}, z, w) \begin{pmatrix} U^* & 0 \\ 0 & U^* \end{pmatrix}. \tag{4.2.49}$$

Since the anisotropic local law holds for $G(D, \tilde{X}^{Gauss}, z, w)$ by Theorem 4.2.22, we get immediately the anisotropic local law for $G(T, X^{Gauss}, z, w)$. The next theorem states that the anisotropic local law holds for general TX provided that the anisotropic local law holds for TX^{Gauss}.

Theorem 4.2.23 (Local law when $N \leqslant M$)**.** *Fix $\tau \leqslant ||z|^2 - 1| \leqslant \tau^{-1}$. Suppose Assumption 4.2.1 holds and $N \leqslant M$. Let $T = U\bar{D}V$ be a singular decomposition of T, where $\bar{D} = (D, 0)$ with $D = \operatorname{diag}(d_1, d_2, \dots, d_N)$. Let $\mathbf{S}$ be a regular domain. Then the anisotropic local law and averaged local law hold with parameters $(T, X, z, \mathbf{S} \cap \widehat{\mathbf{D}})$. If in addition (4.2.22) holds, then the anisotropic local law and averaged local law hold with parameters $(T, X, z, \mathbf{S})$.*

Proof. Using Theorem 4.2.22, we can prove Theorem 4.2.23 with a self-consistent comparison method in [59]. Since the proof is almost the same as the one in Section 2.7, we omit the details and refer the reader to Section 6 of the author's work [105]. □

Finally we turn to the $N > M$ case. Suppose $T = U\bar{D}V$ is a singular decomposition of T, where U is an $N \times N$ unitary matrix, V is an $M \times M$ unitary matrix and $\bar{D} = \begin{pmatrix} D \\ 0 \end{pmatrix}$ is an $N \times M$ matrix such that $D = \operatorname{diag}(d_1, d_2, \dots, d_M)$. Let $U = (U_1, U_2)$, where U_1 has size $N \times M$ and U_2 has size $N \times (N - M)$. Following Girko's idea of Hermitization [47], to prove the local circular law in Theorem 4.2.6 when $N > M$, it suffices to study $\det(TX - z)$ (see (4.2.51) below), for which we have

$$\det(TX - z) = \det \begin{pmatrix} DVXU_1 - z & DVXU_2 \\ 0 & -z \end{pmatrix} = \det(V^T D^T U_1^T X^T - z)(-z)^{N-M}. \tag{4.2.50}$$

Comparing with (4.2.48), we see that this case is reduced to the $N \leqslant M$ case. The only difference is that the extra $(-z)^{N-M}$ term now corresponds to the $N-M$ zero eigenvalues of TX. Thus we make the following claim.

Claim 4.2.24. *The $N < M$ case of Theorem 4.2.6 implies the $N > M$ case of Theorem 4.2.6.*

4.2.3 Proof of the main theorem

We now prove Theorem 4.2.6. By Claim 4.2.24, it suffices to assume $N \leqslant M$. Our main tool will be Theorem 4.2.23. A major part of the proof follows from [18, Section 5].

The Girko's Hermitization technique [47] can be reformulated as the following (see e.g. [50]): for any smooth function g,

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N} g(\mu_j) &= \frac{1}{4\pi N}\int \Delta g(z)\sum_{j=1}^{N}\log(\mu_j - z)(\bar{\mu}_j - \bar{z})dA(z) \\ &= \frac{1}{4\pi N}\int \Delta g(z)\log\left|\det(Y(z)Y^*(z))\right| dA(z) \\ &= \frac{1}{4\pi N}\int \Delta g(z)\sum_{j=1}^{N}\log\lambda_j(z)dA(z), \end{aligned} \tag{4.2.51}$$

where $0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$ are the ordered eigenvalues of $Y(z)Y^*(z)$. For $g = F_{z_0,a}$, we use the new variable $\xi = N^a(z - z_0)$ to write the above equation as

$$\frac{1}{N}\sum_{i=1}^{N} F_{z_0,a}(\mu_j) = \frac{N^{-1+2a}}{4\pi}\int (\Delta F)(\xi)\sum_{j=1}^{N}\log\lambda_j(z)dA(\xi). \tag{4.2.52}$$

Define the classical location $\gamma_j(z)$ of the j-th eigenvalue of $Y(z)Y^*(z)$ by

$$\gamma_j(z) := \begin{cases} \sup_x \left\{\int_0^x \rho_{2c}(x)dx \leqslant \frac{j}{N}\right\}, & \text{if } \ 1 \leqslant j \leqslant N-1 \\ e_1, & \text{if } \ j = N \end{cases}. \tag{4.2.53}$$

In fact, if γ_j lies in the bulk, then by the positivity of ρ_{2c} we can simply define γ_j through

$$\int_0^{\gamma_j}\rho_{2c}(x)dx = \frac{j}{N}.$$

By Proposition 4.2.18, we have that for any $\delta > 0$,

$$\left|\sum_{j=1}^{N} \log \gamma_j(z) - N \int_0^\infty (\log x)\rho_{2c}(x,z)dx\right| \leqslant \sum_{j=1}^{N} N \int_{\gamma_{j-1}(z)}^{\gamma_j(z)} \left|\log \gamma_j(z) - \log x\right| \rho_{2c}(x,z)dx \leqslant N^\delta \tag{4.2.54}$$

for large enough N. Suppose we have the bound

$$\left|\sum_j \log \lambda_j - \sum_j \log \gamma_j\right| \prec N^b. \tag{4.2.55}$$

Plugging (4.2.54) and (4.2.55) into (4.2.52), we get

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} F_{z_0}(\mu_j) &= \frac{N^{2a}}{4\pi}\int (\Delta F)(\xi)\int_0^\infty (\log x)\rho_{2c}(x,z)dxdA(\xi) + \mathrm{O}_\prec(N^{-1+b+2a}\|\Delta F\|_{L_1})\\ &= \frac{1}{4\pi}\int F(\xi)\int_0^\infty (\log x)\Delta_z\rho_{2c}(x,z)dxdA(\xi) + \mathrm{O}_\prec(N^{-1+b+2a}\|\Delta F\|_{L_1}).\end{aligned}$$

Thus we obtain (4.2.20) if we can prove (4.2.55) for $b = 1/2$, and we obtain (4.2.23) if we can prove (4.2.55) for $b = 0$ when $1 + \tau \leqslant |z_0|^2 \leqslant 1 + \tau^{-1}$ or when the assumption (4.2.22) holds.

We need the following lemma which is a consequence of Theorem 4.2.23. Recall (4.2.16) and (4.2.18), the number L of the connected components is of order 1 and the number of γ_j's in each component $[e_{2k}, e_{2k-1}]$ is of order N. We define the classical number of eigenvalues to the left of e_k, $1 \leqslant k \leqslant 2L$, as

$$N_k := \left\lceil N \int_0^{e_k} \rho_{2c}(x) \right\rceil. \tag{4.2.56}$$

Note that $N_{2L} = 0$, $N_1 = N$ and $N_{2k+1} = N_{2k}$, $1 \leqslant k \leqslant L-1$.

Lemma 4.2.25 (Singular value rigidity)**.** *Fix a small $\varepsilon > 0$.*

(i) If the averaged local law holds with parameters $(T, X, z, \mathbf{D}(\zeta, N) \cap \widehat{\mathbf{D}}(\zeta, N))$ for arbitrarily small ζ, then the following estimates hold. For any $e_{2k} > 0$ and $N_{2k} + N^{1/2+\varepsilon} \leqslant j \leqslant N_{2k-1} - N^{1/2+\varepsilon}$,

$$\frac{|\lambda_j - \gamma_j|}{\gamma_j} \prec \left(\min\left\{\frac{j - N_{2k}}{N}, \frac{N_{2k-1} - j}{N}\right\}\right)^{-1/3} N^{-1/2}. \tag{4.2.57}$$

In the case $|z|^2 \leqslant 1 - \tau$ with $e_{2L} = 0$, we have for any $N^{1/2+\varepsilon} \leqslant j \leqslant N_{2L-1} - N^{1/2+\varepsilon}$,

$$\frac{|\lambda_j - \gamma_j|}{\gamma_j} \prec j^{-1}\left(\frac{N_{2L-1} - j}{N}\right)^{-1/3} N^{1/2}. \tag{4.2.58}$$

Moreover, if $1+\tau \leqslant |z|^2 \leqslant 1+\tau^{-1}$*, then for any fixed* $0 < c < e_{2L}$*,*

$$\#\{j : 0 < \lambda_j < c\} \prec 1. \tag{4.2.59}$$

(ii) If the averaged local law holds with parameters $(T, X, z, \mathbf{D}(\zeta, N))$ *for arbitrarily small* ζ*, then the following estimates hold. For any* $e_{2k} > 0$ *and* $N_{2k} + N^{\varepsilon} \leqslant j \leqslant N_{2k-1} - N^{\varepsilon}$*,*

$$\frac{|\lambda_j - \gamma_j|}{\gamma_j} \prec \left(\min\left\{\frac{j - N_{2k}}{N}, \frac{N_{2k-1} - j}{N}\right\}\right)^{-1/3} N^{-1}. \tag{4.2.60}$$

In the case $|z|^2 \leqslant 1-\tau$ *with* $e_{2L} = 0$*, we have for any* $N^{\varepsilon} \leqslant j \leqslant N_{2L-1} - N^{\varepsilon}$*,*

$$\frac{|\lambda_j - \gamma_j|}{\gamma_j} \prec j^{-1}\left(\frac{N_{2L-1} - j}{N}\right)^{-1/3}. \tag{4.2.61}$$

Proof. The proof is similar to the proof of [18, Lemma 5.1]. See also [14, Theorem 2.10] or [33, Theorem 7.6] □

Using (4.2.57) and (4.2.58), we get that

$$\sum_{N_{2k}+N^{1/2+\varepsilon} \leqslant j \leqslant N_{2k-1}-N^{1/2+\varepsilon}} |\log\lambda_j - \log\gamma_j| \prec \sum_{N_{2k}+N^{1/2+\varepsilon} \leqslant j \leqslant N_{2k-1}-N^{1/2+\varepsilon}} \frac{|\lambda_j - \gamma_j|}{\gamma_j} \prec N^{1/2}. \tag{4.2.62}$$

By Theorem 2.10 of [14], there exists a constant $C > 0$ such that

$$\|X^*X\| \leqslant C \quad \text{with high probability.} \tag{4.2.63}$$

Thus we have

$$\lambda_j \leqslant \|Y\|^2 \leqslant (\|T\|\|X\| + |z|)^2 \prec 1, \quad 1 \leqslant j \leqslant N. \tag{4.2.64}$$

Together with Lemma 4.2.26 concerning the smallest singular value of $TX - z$, we get

$$\sum_{k=1}^{2L} \sum_{|j-e_k| < N^{1/2+\varepsilon}} |\log\lambda_j| \prec N^{1/2+\varepsilon}. \tag{4.2.65}$$

Since $|\log\gamma_j| \prec 1$ by Proposition 4.2.18, we conclude

$$\sum_{k=1}^{2L} \sum_{|j-e_k| < N^{1/2+\varepsilon}} |\log\lambda_j - \log\gamma_j| \prec N^{1/2+\varepsilon}. \tag{4.2.66}$$

Combining (4.2.62) and (4.2.66), we get for any $\varepsilon > 0$,

$$\sum_{1 \leqslant j \leqslant N} |\log \lambda_j - \log \gamma_j| \prec N^{1/2+\varepsilon} \tag{4.2.67}$$

for large enough N. This implies (4.2.55) for $b = 1/2$. If in addition the assumption (4.2.22) holds, the averaged local law holds with parameters $(T, X, z, \mathbf{D}(\zeta, N))$ for arbitrarily small ζ by Theorem 4.2.23. Then we can prove (4.2.55) for $b = 0$ using the better bounds (4.2.60) and (4.2.61).

Finally we show that when $|z_0|^2 \geqslant 1 + \tau$, with the bounds (4.2.57) we can still prove the estimate (4.2.55) for $b = 0$. By the averaged local law and the definition of γ_j in (4.2.53), we have

$$\left| \sum_{j=1}^{N} \frac{1}{\lambda_j - \mathrm{i}\eta} - \sum_{j=1}^{N} \frac{1}{\gamma_j - \mathrm{i}\eta} \right| \prec \frac{1}{\eta}, \tag{4.2.68}$$

uniformly in $N^{-1/2+\varepsilon} \leqslant \eta \leqslant N^{1/2}$. Taking integral of (4.2.68) over η from $N^{-1/2+\varepsilon}$ to $N^{1/2}$, we get

$$\left| \sum_{j=1}^{N} \log \left(\frac{\lambda_j - iN^{-1/2+\varepsilon}}{\gamma_j - iN^{-1/2+\varepsilon}} \right) - \sum_{j=1}^{N} \log \left(\frac{\lambda_j - iN^{1/2}}{\gamma_j - iN^{1/2}} \right) \right| \prec 1. \tag{4.2.69}$$

Then we use (4.2.57) and the bound (4.2.64) to estimate that

$$\left| \sum_{j=1}^{N} \log \left(\frac{\lambda_j - iN^{1/2}}{\gamma_j - iN^{1/2}} \right) \right| \prec \sum_{j=1}^{N} \left| (\lambda_j - \gamma_j)\, N^{-1/2} \right| \prec N^{\varepsilon}.$$

Thus we conclude

$$\left| \sum_{j=1}^{N} \log \left(\frac{\lambda_j - iN^{-1/2+\varepsilon}}{\gamma_j - iN^{-1/2+\varepsilon}} \right) \right| \prec N^{\varepsilon}. \tag{4.2.70}$$

Using $\gamma_j \sim 1$, (4.2.59) and (4.2.72), we get

$$\begin{aligned} \left| \sum_{j=1}^{N} \log \left(\frac{\lambda_j - iN^{-1/2+\varepsilon}}{\gamma_j - iN^{-1/2+\varepsilon}} \right) - \sum_{j=1}^{N} \log \frac{\lambda_j}{\gamma_j} \right| &\prec 1 + \left| \sum_{\lambda_j \geqslant c} \log \left(\frac{\lambda_j - iN^{-1/2+\varepsilon}}{\gamma_j - iN^{-1/2+\varepsilon}} \right) - \sum_{\lambda_j \geqslant c} \log \frac{\lambda_j}{\gamma_j} \right| \\ &\prec 1 + \sum_{\lambda_j \geqslant c} \left| (\lambda_j - \gamma_j)\, N^{-1/2+\varepsilon} \right| \prec N^{2\varepsilon}. \end{aligned} \tag{4.2.71}$$

Combing (4.2.70) and (4.2.71), we conclude (4.2.55) for $b = 0$.

If the entries of X are identically distributed, then instead of Lemma 4.2.26 below, we shall use Theorem B.1.1 in Appendix B to get a lower bound for the smallest singular value of $TX - z$. In particular, the bounded density condition for the entries of X is not needed anymore. This concludes the last statement of Theorem 4.2.6.

Lemma 4.2.26 (Lower bound on the smallest singular value)**.** *If $N \leqslant M$ and the entries of X have a density bounded by N^{C_3} for some $C_3 > 0$, then*

$$|\log \lambda_1(z)| \prec 1 \tag{4.2.72}$$

holds uniformly for z in any fixed compact set.

Proof. We already have an upper bound for λ_1; see (4.2.64). Hence to get (4.2.72), we still need to prove that

$$\mathbb{P}\left(\lambda_1(z) \leqslant e^{-N^{\varepsilon}}\right) \leqslant N^{-C} \tag{4.2.73}$$

for any $\varepsilon, C > 0$. By (4.2.48), we have that

$$TX - z = UD(V_1 X - D^{-1}U^{-1}z) =: UD\tilde{Y}(z).$$

Hence it suffices to control the smallest singular value of $\tilde{Y}(z)$, call it $\tilde{\lambda}_1(z)$. Notice the columns $\tilde{Y}_1, \ldots, \tilde{Y}_N$ of $\tilde{Y}(z)$ are independent vectors. From the variational characterization

$$\tilde{\lambda}_1(z) = \min_{|u|=1} \|\tilde{Y}(z)u\|^2,$$

we can easily get

$$\tilde{\lambda}_1(z)^{1/2} \geqslant N^{-1/2} \min_{1\leqslant k\leqslant N} \operatorname{dist}\left(\tilde{Y}_k, \operatorname{span}\{\tilde{Y}_l, l \neq k\}\right) = N^{-1/2} \min_{1\leqslant k\leqslant N} \left|\langle \tilde{Y}_k, u_k\rangle\right|, \tag{4.2.74}$$

where u_k is the unit normal vector of $\operatorname{span}\{\tilde{Y}_l, l \neq k\}$ and hence is independent of $\tilde{Y}_k$. By conditioning on u_k, we get immediately that

$$\mathbb{P}(\tilde{\lambda}_1(z) \leqslant N^{-C_0}) \leqslant CN^{-C_0/2+C_3+3/2}, \tag{4.2.75}$$

which is a much stronger result than (4.2.73). Here we have used Theorem 1.2 of [84] to conclude that $\langle \bar{Y}_k, u_k\rangle$ for fixed u_k has density bounded by CN^{C_3}. □

The rest of this chapter is devoted to the proof of Theorem 4.2.22. In Section 4.3, we collect the basic tools that we shall use in the proof. In Section 4.4, we prove the entrywise local law and averaged local law in Theorem 4.2.22 under the assumption that T is diagonal. In Section 4.5, we prove the anisotropic local law in Theorem 4.2.22 using the entrywise local law proved in Section 4.4.

The Appendix A.1 establishes the basic properties of $\rho_{1,2c}$ stated in Lemma 4.2.3 and Proposition 4.2.18. Our proof of the entrywise local depends crucially on some key estimates about $m_{1,2c}$ and the stability of the self-consistent equation (4.2.11) on regular domains, which will be proved in Sections A.2 and A.3. Finally, in Appendix B, we prove a lower tail estimate for the smallest singular value of $TX - z$.

4.3 Basic tools

In this preliminary section, we collect various identities and estimates that we shall use throughout the following. We first state the resolvent identities in current case. They (and their proof) are very similar to the ones in Lemma 2.6.3, and we state them for reader's convenience.

Lemma 4.3.1. *(Resolvent identities).*

(i) For $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_2$, we have

$$\frac{1}{G_{ii}} = -w - w\left(YG^{(i)}Y^*\right)_{ii}, \quad \frac{1}{G_{\mu\mu}} = -w - w\left(Y^*G^{(\mu)}Y\right)_{\mu\mu}. \tag{4.3.1}$$

For $i \neq j \in \mathcal{I}_1$ and $\mu \neq \nu \in \mathcal{I}_2$, we have

$$G_{ij} = wG_{ii}G_{jj}^{(i)}\left(YG^{(ij)}Y^*\right)_{ij}, \quad G_{\mu\nu} = wG_{\mu\mu}G_{\nu\nu}^{(\mu)}\left(Y^*G^{(\mu\nu)}Y\right)_{\mu\nu}. \tag{4.3.2}$$

(ii) For $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_2$, we have

$$G_{i\mu} = G_{ii}G_{\mu\mu}^{(i)}\left(-w^{1/2}Y_{i\mu} + w\big(YG^{(i\mu)}Y\big)_{i\mu}\right), \tag{4.3.3}$$

$$G_{\mu i} = G_{\mu\mu}G_{ii}^{(\mu)}\left(-w^{1/2}Y_{\mu i}^* + w\left(Y^*G^{(\mu i)}Y^*\right)_{\mu i}\right). \tag{4.3.4}$$

(iii) For $r \in \mathcal{I}$ and $s, t \in \mathcal{I}\backslash\{r\}$,

$$G_{st}^{(r)} = G_{st} - \frac{G_{sr}G_{rt}}{G_{rr}}, \quad \frac{1}{G_{ss}} = \frac{1}{G_{ss}^{(r)}} - \frac{G_{sr}G_{rs}}{G_{ss}G_{ss}^{(r)}G_{rr}}. \tag{4.3.5}$$

(iv) All of the above identities hold for $G^{(J)}$ instead of G for $J \subset \mathcal{I}$.

Lemma 4.3.2. *(Resolvent identities for $G_{[ij]}$ groups).*

(i) For $i \in \mathcal{I}_1$, we have

$$G_{[ii]}^{-1} = H_{[ii]} - \sum_{k,l \neq i} H_{[ik]} G_{[kl]}^{[i]} H_{[li]}. \tag{4.3.6}$$

For $i \neq j \in \mathcal{I}_1$, we have

$$G_{[ij]} = -G_{[ii]} \sum_{k \neq i} H_{[ik]} G_{[kj]}^{[i]} = -\sum_{k \neq j} G_{[ik]}^{[j]} H_{[kj]} G_{[jj]} \tag{4.3.7}$$

$$= -G_{[ii]} H_{[ij]} G_{[jj]}^{[i]} + G_{[ii]} \sum_{k,l \notin \{i,j\}} H_{[ik]} G_{[kl]}^{[ij]} H_{[lj]} G_{[jj]}^{[i]}. \tag{4.3.8}$$

(ii) For $k \in \mathcal{I}_1$ and $i, j \in \mathcal{I}_1 \backslash \{k\}$,

$$G_{[ij]}^{[k]} = G_{[ij]} - G_{[ik]} G_{[kk]}^{-1} G_{[kj]}, \tag{4.3.9}$$

and

$$G_{[ii]}^{-1} = \left(G_{[ii]}^{[k]}\right)^{-1} - G_{[ii]}^{-1} G_{[ik]} G_{[kk]}^{-1} G_{[ki]} \left(G_{[ii]}^{[k]}\right)^{-1}. \tag{4.3.10}$$

(iii) All of the above identities hold for $G^{[J]}$ instead of G for $J \subset \mathcal{I}$.

Proof. These identities can be proved using Schur's complement formula. □

Next we introduce the spectral decomposition of G. Let

$$Y = \sum_{k=1}^{N} \sqrt{\lambda_k} \xi_k \zeta_{\bar{k}}^*$$

be the singular decomposition of Y, where $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_N \geqslant 0$ and $\{\xi_k\}_{k=1}^N$ and $\{\zeta_{\bar{k}}\}_{k=1}^N$ are orthonormal bases of $\mathbb{C}^{\mathcal{I}_1}$ and $\mathbb{C}^{\mathcal{I}_2}$ respectively. Then by (4.2.29), we have

$$G(w) = \sum_{k=1}^{N} \frac{1}{\lambda_k - w} \begin{pmatrix} \xi_k \xi_k^* & w^{-1/2} \sqrt{\lambda_k} \xi_k \zeta_{\bar{k}}^* \\ w^{-1/2} \sqrt{\lambda_k} \zeta_{\bar{k}} \xi_k^* & \zeta_{\bar{k}} \zeta_{\bar{k}}^* \end{pmatrix}. \tag{4.3.11}$$

The following lemma corresponds to Lemma 2.7.1, and we leave its proof to the reader.

Lemma 4.3.3. *Fix $\tau > 0$. The following estimates hold uniformly for any $w \in \mathbf{D}(\zeta, N) \cup \mathbf{D}_L(\zeta)$. We have*

$$\|G\| \leqslant C\eta^{-1}, \ \|\partial_w G\| \leqslant C\eta^{-2}. \tag{4.3.12}$$

Let $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$ and $\mathbf{w} \in \mathbb{C}^{\mathcal{I}_2}$, we have the bounds

$$\sum_{\mu\in\mathcal{I}_2} |G_{\mathbf{w}\mu}|^2 = \sum_{\mu\in\mathcal{I}_2} |G_{\mu\mathbf{w}}|^2 = \frac{\operatorname{Im} G_{\mathbf{ww}}}{\eta}, \tag{4.3.13}$$

$$\sum_{i\in\mathcal{I}_1} |G_{\mathbf{v}i}|^2 = \sum_{i\in\mathcal{I}_1} |G_{i\mathbf{v}}|^2 = \frac{\operatorname{Im} G_{\mathbf{vv}}}{\eta}, \tag{4.3.14}$$

$$\sum_{i\in\mathcal{I}_1} |G_{\mathbf{w}i}|^2 = \sum_{i\in\mathcal{I}_1} |G_{i\mathbf{w}}|^2 = |w|^{-1} G_{\mathbf{ww}} + \bar{w}\,|w|^{-1} \frac{\operatorname{Im} G_{\mathbf{ww}}}{\eta}, \tag{4.3.15}$$

$$\sum_{\mu\in\mathcal{I}_2} |G_{\mathbf{v}\mu}|^2 = \sum_{\mu\in\mathcal{I}_2} |G_{\mu\mathbf{v}}|^2 = |w|^{-1} G_{\mathbf{vv}} + \bar{w}\,|w|^{-1} \frac{\operatorname{Im} G_{\mathbf{vv}}}{\eta}. \tag{4.3.16}$$

All of the above estimates remain true for $G^{(J)}$ instead of G for $J \subset \mathcal{I}$.

We have stated some basic properties of $\rho_{1,2c}$ and $m_{1,2c}$ in Lemma 4.2.3 and Proposition 4.2.18. Now we collect more estimates for $m_{1,2c}$ that will be used in the proof. The next lemma is proved in Appendix A.2. For $w = E + \mathrm{i}\eta$, we define the distance to the spectral edge through

$$\kappa \equiv \kappa(E) := \min_{1\leqslant k\leqslant 2L, e_k>0} |E - e_k|. \tag{4.3.17}$$

Notice in the $|z| < 1$ case, we do not take into consideration the edge at $e_{2L} = 0$.

Lemma 4.3.4. *Fix $\tau > 0$ and suppose $\tau \leqslant ||z|^2 - 1| \leqslant \tau^{-1}$. We denote $w = E + \mathrm{i}\eta$.*

Case 1 Fix $\tau' > 0$. Suppose the bulk component $[e_{2k}, e_{2k-1}]$ is regular in the sense of Definition 4.2.4. Then for $w \in \mathbf{D}_k^b(\zeta, \tau', N)$, we have

$$|1 + m_{1c}| \sim \operatorname{Im} m_{1c} \sim 1, \ |m_{2c}| \sim \operatorname{Im} m_{2c} \sim 1. \tag{4.3.18}$$

Case 2 Fix $\tau' > 0$. Then for $w \in \mathbf{D}^o(\zeta, \tau', N)$, we have

$$\operatorname{Im} m_{1,2c} \sim \eta, \ |1 + m_{1c}| \sim 1, \ |m_{2c}| \sim 1. \tag{4.3.19}$$

Case 3 *Suppose $e_k \neq 0$ is a regular edge. Then for $w \in \mathbf{D}^e_k(\zeta, \tau', N)$, if $\tau' > 0$ is small enough, we have*

$$\operatorname{Im} m_{1,2c} \sim \begin{cases} \sqrt{\kappa + \eta} & \text{if } E \in \operatorname{supp} \rho_{1,2c} \\ \eta/\sqrt{\kappa+\eta} & \text{if } E \notin \operatorname{supp} \rho_{1,2c} \end{cases}, \quad |1 + m_{1c}| \sim 1, \quad |m_{2c}| \sim 1. \tag{4.3.20}$$

Case 4 *Suppose $|z|^2 \leqslant 1 - \tau$ so that $e_{2L} = 0$. We take $\tau' > 0$ to be small enough. Then for $w \in \mathbf{D}^e_{2L}(\zeta, \tau', N)$, if $\operatorname{Im} w \geqslant \tau'$, we have*

$$|1 + m_{1c}| \sim \operatorname{Im} m_{1c} \sim 1, \; |m_{2c}| \sim \operatorname{Im} m_{2c} \sim 1; \tag{4.3.21}$$

if $|w| \leqslant 2\tau'$, we have

$$m_{1c} = i\frac{\sqrt{t}}{\sqrt{w}} + O(1), \; m_{2c} = \frac{i\sqrt{t}}{\sqrt{w}(t + |z|^2)} + O(1), \tag{4.3.22}$$

for some constant $t > 0$, and

$$\operatorname{Im} m_{1,2c} \sim |w|^{-1/2}. \tag{4.3.23}$$

Case 5 *For $w \in \mathbf{D}_L(\zeta)$, we have*

$$|m_{1c}| \sim \operatorname{Im} m_{1c} \sim \frac{1}{\eta}, \; |m_{2c}| \sim \operatorname{Im} m_{2c} \sim \frac{1}{\eta}. \tag{4.3.24}$$

In Cases 1-4, we have

$$\left| w\left(1 + s_i m_{2c}\right)\left(1 + m_{1c}\right) - |z|^2 \right| \geqslant c, \tag{4.3.25}$$

where $c > 0$ is some constant that may depend on τ, τ' and ζ. In Case 5, we have

$$\left| w\left(1 + s_i m_{2c}\right)\left(1 + m_{1c}\right) - |z|^2 \right| \geqslant \eta, \tag{4.3.26}$$

Note that the uniform bounds (4.3.25) and (4.3.26) guarantee that the matrix entries of $\Pi(w)$ remain bounded. We have the following Lemma, which will be proved in Appendix A.2.

Lemma 4.3.5. *In Cases 1-4 of Lemma 4.3.4, we have*

$$\|\pi_{[i]c}\| \leqslant C|w|^{-1/2}, \quad \left\| \left(\pi_{[i]c}\right)^{-1} \right\| \leqslant C|w|^{1/2}, \tag{4.3.27}$$

and in Case 5 of Lemma 4.3.4, we have

$$\|\pi_{[i]c}\| \leqslant C\eta^{-1}, \quad \left\|\left(\pi_{[i]c}\right)^{-1}\right\| \leqslant C\eta. \tag{4.3.28}$$

For all the cases in Lemma 4.3.4,

$$\operatorname{Im} \Pi_{\mathbf{v}\mathbf{v}} \leqslant C \operatorname{Im}(m_{1c} + m_{2c}), \tag{4.3.29}$$

uniformly in w and any deterministic unit vector $\mathbf{v} \in \mathbb{C}^{\mathcal{I}}$.

The self-consistent equation (4.2.11) can be written as

$$\Upsilon(w, m_1) = 0, \tag{4.3.30}$$

where

$$\Upsilon(w, m_1) = m_1 + \frac{1}{N}\sum_{i=1}^{n} l_i s_i (1+m_1)\left[w\left(1 + s_i \frac{1+m_1}{-w(1+m_1)^2 + |z|^2}\right)(1+m_1) - |z|^2\right]^{-1}. \tag{4.3.31}$$

The stability of (4.3.30) roughly says that if $\Upsilon(w, m_1)$ is small and $m_1(w') - m_{1c}(w')$ is small for $w' := w + iN^{-10}$, then $m_1(w) - m_{1c}(w)$ is small. For an arbitrary $w \in \mathbf{D}$, we define the discrete set

$$L(w) := \{w\} \cup \{w' \in \mathbf{D} : \operatorname{Re} w' = \operatorname{Re} w, \operatorname{Im} w' \in [\operatorname{Im} w, 1] \cap (N^{-10}\mathbb{N})\}, \tag{4.3.32}$$

Thus, if $\operatorname{Im} w \geqslant 1$ then $L(w) = \{w\}$, and if $\operatorname{Im} w < 1$ then $L(w)$ is a 1-dimensional lattice with spacing N^{-10} plus the point w. Obviously, we have $|L(w)| \leqslant N^{10}$.

Definition 4.3.6 (Stability)**.** *We say that (4.3.30) is stable on $\mathbf{D}$ if the following holds. Suppose that $N^{-2}|m_{1c}| \leqslant \delta(w) \leqslant (\log N)^{-1}|m_{1c}|$ for $w \in \mathbf{D}$ and that δ is Lipschitz continuous with Lipschitz constant $\leqslant N^4$. Suppose moreover that for each fixed E, the function $\eta \mapsto \delta(E + \mathrm{i}\eta)$ is non-increasing for $\eta > 0$. Suppose that $u_1 : \mathbf{D} \to \mathbb{C}$ is the Stieltjes transform of a positive integrable function. Let $w \in \mathbf{D}$ and suppose that for all $w' \in L(w)$ we have*

$$|\Upsilon(w, u_1)| \leqslant \delta(w). \tag{4.3.33}$$

Then

$$|u_1(w) - m_{1c}(w)| \leqslant \frac{C\delta}{\sqrt{\kappa + \eta + \delta}}, \tag{4.3.34}$$

for some constant $C > 0$ independent of w and N.

We say that (4.3.30) is stable on $\mathbf{D}_L$ if for $0 \leqslant \delta(w) \leqslant (\log N)^{-1}|m_{1c}|$, (4.3.33) implies

$$|u_1(w) - m_{1c}(w)| \leqslant C\delta, \tag{4.3.35}$$

for some constant $C > 0$ independent of w and N.

In the following lemma, we establish the stability on each regular domain. The proof is given in Appendix A.3. This lemma leaves the case $|w|^{1/2} + |z|^2 = \mathrm{o}(1)$ alone. We will handle this case in a different way in Section 4.4.5.

Lemma 4.3.7. *Fix $\tau > 0$ and let $\tau' > 0$ be sufficiently small depending on τ. Let $\tau \leqslant ||z|^2 - 1| \leqslant \tau^{-1}$.*

Case 1 Suppose the bulk component $[e_{2k}, e_{2k-1}]$ is regular in the sense of Definition 4.2.4. Then (4.3.30) is stable on $\mathbf{D}_k^b(\zeta, \tau', N)$ in the sense of Definition 4.3.6.

Case 2 (4.3.30) is stable on $\mathbf{D}^o(\zeta, \tau', N)$ in the sense of Definition 4.3.6.

Case 3 Suppose $e_k \neq 0$ is a regular edge in the sense of Definition 4.2.4. Then (4.3.30) is stable on $\mathbf{D}_k^e(\zeta, \tau', N)$ in the sense of Definition 4.3.6.

Case 4 Suppose $|z|^2 \leqslant 1 - \tau$ so that $e_{2L} = 0$. If $|w|^{1/2} + |z|^2 \geqslant \varepsilon$ for some constant $\varepsilon > 0$, then (4.3.30) is stable on $\mathbf{D}_{2L}^e(\zeta, \tau', N)$ in the sense of Definition 4.3.6.

Case 5 (4.3.30) is stable on $\mathbf{D}_L(\zeta)$ in the sense of Definition 4.3.6.

4.4 Entrywise local law when T is diagonal

In this section we prove the entrywise local law and averaged local law in Theorem 4.2.22 when T is diagonal. The proof is similar to the previous proofs of the entrywise local law in e.g. [14, 15, 18, 59]. We basically follow the idea in [18], and we will provide necessary details for the parts that are different from the previous proofs.

Before we start the proof, we make the following remark. In this section we mainly focus on the domain $\mathbf{D}$. On the domain $\mathbf{D}_L$, the proofs are much simpler and we only describe them briefly. The parameter z can be either inside or outside of the unit circle. Recall Lemma 4.3.4 and Lemma 4.3.7, the domain $\mathbf{D}$ of w can be divided roughly into four regions: w near a *nonzero* regular edge, $w \to 0$, w in the bulk, or w outside the spectrum. In this section we will only consider the case $|z|^2 \leqslant 1-\tau$ since it covers all four different behaviors of $m_{1,2c}$. Note that in this case $|m_{1,2c}(w)| \sim |w|^{-1/2}$ for w in any compact set of $\mathbb{C}_+$ by Proposition 4.2.19. Also due to the remark above Lemma 4.3.7, in Sections 4.4.1-4.4.4, we assume $|w|^{1/2} + |z|^2 \geqslant c$ for some $c > 0$. We will handle the $|w|^{1/2} + |z|^2 = o(1)$ case in Section 4.4.5.

4.4.1 The self-consistent equations

To begin with, we prove the following weak version of the entrywise local law.

Proposition 4.4.1 (Weak entrywise law)**.** *Fix $|z|^2 \leqslant 1-\tau$ and a small constant $c>0$. Suppose Assumption 4.2.1 holds, $N = M$ and $T \equiv D := \operatorname{diag}(d_1, ..., d_N)$. Then for any regular domain $\mathbf{S} \subset \mathbf{D}$,*

$$\max_{i,j \in \mathcal{I}_1} \left\| (G(w) - \Pi(w))_{[ij]} \right\| \prec \frac{1}{|w|^{1/2}} \left(\frac{|w|^{1/2}}{N\eta} \right)^{1/4} \tag{4.4.1}$$

for all $w \in \mathbf{S}$ such that $|w|^{1/2} + |z|^2 \geqslant c$. For $w \in \mathbf{D}_L$, we have

$$\max_{i,j \in \mathcal{I}_1} \left\| (G(w) - \Pi(w))_{[ij]} \right\| \prec \frac{1}{\eta}\sqrt{\frac{1}{N}}. \tag{4.4.2}$$

For the purpose of proof, we define the following random control parameters.

Definition 4.4.2 (Control parameters)**.** *Suppose $N = M$ and $T \equiv D := \operatorname{diag}(d_1, ..., d_N)$. We define*

$$\Lambda := \max_{i,j \in \mathcal{I}_1} \left\| (G - \Pi)_{[ij]} \right\|, \quad \Lambda_o := \max_{i \neq j \in \mathcal{I}_1} \left\| (G - \Pi)_{[ij]} \right\|. \tag{4.4.3}$$

For $J \subseteq \mathcal{I}$, define the averaged variables $m_{1,2}^{(J)}$ ($m_{1,2}^{[J]}$) by replacing G in (4.2.33) with $G^{(J)}$ ($G^{[J]}$), i.e.

$$m_1^{(J)} := \frac{1}{N} \sum_{i \notin J} |d_i|^2 G_{ii}^{(J)}, \quad m_2^{(J)} := \frac{1}{N} \sum_{\mu \notin J} G_{\mu\mu}^{(J)}. \tag{4.4.4}$$

The averaged error and the random control parameter are defined as

$$\theta := |m_1 - m_{1c}| + |m_2 - m_{2c}| \quad \text{and} \quad \Psi_\theta := \sqrt{\frac{\operatorname{Im}(m_{1c} + m_{2c}) + \theta}{N\eta}} + \frac{1}{N\eta}, \tag{4.4.5}$$

respectively.

Remark 4.4.3. By (4.2.4), we immediately get that

$$\tau \operatorname{Im} m_1^{(J)} \leqslant \operatorname{Im} m_2^{(J)} \leqslant \tau^{-1} \operatorname{Im} m_2^{(J)}, \tag{4.4.6}$$

and $\theta = O(\Lambda)$, since $|m_1 - m_{1c}| \leqslant \tau^{-1}\Lambda$ and $|m_2 - m_{2c}| \leqslant \Lambda$.

As in (2.6.22), we introduce the Z variables:

$$Z_{[i]}^{[J]} := (1 - \mathbb{E}_{[i]}) \left(G_{[ii]}^{[J]} \right)^{-1}.$$

By the identity (4.3.6), we have

$$G_{[ii]}^{-1} = \mathbb{E}_{[i]} G_{[ii]}^{-1} + Z_{[i]} = \begin{pmatrix} -w - w\,|d_i|^2\, m_2^{[i]} & -w^{1/2} z \\ -w^{1/2}\bar{z} & -w - w m_1^{[i]} \end{pmatrix} + Z_{[i]}, \tag{4.4.7}$$

where

$$Z_{[i]} = w \begin{pmatrix} |d_i|^2 m_2^{[i]} - |d_i|^2 \left(X G^{[i]} X^* \right)_{ii} & w^{-1/2} d_i X_{i\bar{i}} - \left(D X G^{[i]} D X \right)_{i\bar{i}} \\ w^{-1/2} \bar{d}_i X^*_{\bar{i}i} - \left(X^* D^* G^{[i]} X^* D^* \right)_{\bar{i}i} & m_1^{[i]} - \left(X^* D^* G^{[i]} D X \right)_{\bar{i}\bar{i}} \end{pmatrix}. \tag{4.4.8}$$

We have the following lemma, whose proof is the same as the one for (2.6.14).

Lemma 4.4.4. *For $J \subseteq \mathcal{I}_1$, the following crude bound on the difference between m_a and $m_a^{[J]}$ ($a = 1, 2$) holds:*

$$\left| m_a - m_a^{[J]} \right| \leqslant \frac{C\,|J|}{N\eta}, \quad a = 1, 2, \tag{4.4.9}$$

where $C = C(\tau)$ is a constant depending only on τ.

With (4.4.9), we can prove the following estimates on the Z variables.

Lemma 4.4.5. *Suppose $|z|^2 \leqslant 1 - \tau$. For $i \in \mathcal{I}_1$, we have*

$$\left| \left(Z_{[i]} \right)_{11} \right| \prec |w| \sqrt{\frac{\operatorname{Im} m_2^{[i]}}{N\eta}}, \quad \left| \left(Z_{[i]} \right)_{22} \right| \prec |w| \sqrt{\frac{\operatorname{Im} m_1^{[i]}}{N\eta}}, \tag{4.4.10}$$

$$\left| \left(Z_{[i]} \right)_{st} \right| \prec |w| \left(\frac{|w|^{-1/2}}{\sqrt{N}} + \sqrt{\frac{|m_1^{[i]}|}{N\,|w|}} + \sqrt{\frac{\operatorname{Im} m_1^{[i]}}{N\eta}} \right) \quad \text{for } s \neq t \in \{1, 2\}, \tag{4.4.11}$$

uniformly in $w \in \mathbf{D} \cup \mathbf{D}_L$. In particular, these imply that

$$Z_{[i]} \prec |w|\Psi_\theta, \tag{4.4.12}$$

uniformly in $w \in \mathbf{D}$, and

$$Z_{[i]} \prec |w|(N\eta)^{-1/2}, \tag{4.4.13}$$

uniformly in $w \in \mathbf{D}_L$.

Proof. Applying the large deviation Lemma 2.6.6 to $Z_{[i]}$ in (4.4.8), we get that

$$\begin{aligned}\left|\frac{\left(Z_{[i]}\right)_{11}}{w}\right| \prec & \frac{1}{N}\left[\left(\sum_\mu |G^{[i]}_{\mu\mu}|^2\right)^{1/2} + \left(\sum_{\mu\neq\nu} |G^{[i]}_{\mu\nu}|^2\right)^{1/2}\right] \leqslant \frac{C}{N}\left(\sum_{\mu,\nu}|G^{[i]}_{\mu\nu}|^2\right)^{1/2} \\ =& \frac{C}{N}\left(\sum_\mu \frac{\operatorname{Im} G^{[i]}_{\mu\mu}}{\eta}\right)^{1/2} = C\sqrt{\frac{\operatorname{Im} m_2^{[i]}}{N\eta}}.\end{aligned}$$

where in the third step we used the equality (4.3.13). Similarly we can prove the bound for $\left(Z_{[i]}\right)_{22}$ using Lemma 2.6.6 and (4.3.14). Now we consider $\left(Z_{[i]}\right)_{12}$. First, we have $X_{i\bar{i}} \prec N^{-1/2}$ by (4.2.3). For the other part, we use Lemma 2.6.6 and (4.3.16) to get that

$$\begin{aligned}\left|\left(DXG^{[i]}DX\right)_{i\bar{i}}\right| \prec & \frac{1}{N}\left(\sum_{j,\mu}|d_j|^2\left|G^{[i]}_{\mu j}\right|^2\right)^{1/2} = \frac{1}{N}\left[\sum_j |d_j|^2\left(|w|^{-1}G^{[i]}_{jj} + \frac{\bar{w}}{|w|}\frac{\operatorname{Im} G^{[i]}_{jj}}{\eta}\right)\right]^{1/2} \\ \leqslant & \left[\frac{|m_1^{[i]}|}{N|w|} + \frac{\operatorname{Im} m_1^{[i]}}{N\eta}\right]^{1/2} \leqslant C\left(\sqrt{\frac{|m_1^{[i]}|}{N|w|}} + \sqrt{\frac{\operatorname{Im} m_1^{[i]}}{N\eta}}\right).\end{aligned} \tag{4.4.14}$$

Similarly we can prove the estimate for $\left(Z_{[i]}\right)_{21}$.

Now we prove (4.4.12). By the definitions (4.4.5) and using (4.4.9), we get that

$$\left|\left(Z_{[i]}\right)_{11}\right| \prec |w|\sqrt{\frac{\operatorname{Im} m_2^{[i]}}{N\eta}} = |w|\sqrt{\frac{\operatorname{Im} m_{2c} + \operatorname{Im}\left(m_2^{[i]} - m_2\right) + \operatorname{Im}\left(m_2 - m_{2c}\right)}{N\eta}} \leqslant C|w|\Psi_\theta. \tag{4.4.15}$$

We can estimate $\left(Z_{[i]}\right)_{22}$ and the third term in (4.4.11) in a similar way. For the Cases 1-4 in Lemma 4.3.4, we have $|m_{1c}| \sim 1$ for $|w| \sim 1$, $\operatorname{Im} m_{1c} \sim |w|^{-1/2} \sim |m_{1c}|$ for $|w| \to 0$, and $\eta \leqslant C\operatorname{Im} m_{1c}$. Thus

$$\sqrt{\frac{|m_{1c}|}{N|w|}} \leqslant \frac{C}{\sqrt{N}} \leqslant C\Psi_\theta \text{ for } |w| \sim 1, \quad \text{and} \quad \sqrt{\frac{|m_{1c}|}{N|w|}} \leqslant C\sqrt{\frac{\operatorname{Im} m_{1c}}{N\eta}} \leqslant C\Psi_\theta \text{ for } |w| \to 0.$$

Then for the second term in (4.4.11), we have that

$$\sqrt{\frac{|m_1^{[i]}|}{N\,|w|}} \leqslant C\left(\frac{1}{N\eta}+\sqrt{\frac{\theta}{N\eta}}+\sqrt{\frac{|m_{1c}|}{N\,|w|}}\right) \leqslant C\Psi_\theta.$$

This concludes (4.4.12). Finally, the estimate (4.4.13) follows directly from (4.4.10), (4.4.11) and (4.3.12). □

Lemma 4.4.6. *Suppose* $|z|^2 \leqslant 1-\tau$. *Define the* w*-dependent event* $\Xi(w) := \{\theta \leqslant |w|^{-1/2}(\log N)^{-1}\}$. *Then we have that for* $w \in \mathbf{D}$,

$$\mathbf{1}(\Xi)m_2 = \mathbf{1}(\Xi)\left[\frac{1+m_1}{-w\,(1+m_1)^2+|z|^2}+\mathrm{O}_\prec(\Psi_\theta)\right], \quad \mathbf{1}(\Xi)\Upsilon(w,m_1) \prec \mathbf{1}(\Xi)\Psi_\theta, \qquad (4.4.16)$$

where Υ *is defined in (4.3.31). For* $w \in \mathbf{D}_L$, *we have*

$$m_2 = \frac{1+m_1}{-w\,(1+m_1)^2+|z|^2}+\mathrm{O}_\prec\left(\eta^{-1}(N\eta)^{-1/2}\right), \quad \Upsilon(w,m_1) \prec \eta^{-1}\,(N\eta)^{-1/2}. \qquad (4.4.17)$$

Proof. First, suppose that $w \in \mathbf{D}$. Using (4.4.7), we get

$$G_{[ii]}^{-1} = \pi_{[i]}^{-1} + \varepsilon_{[i]}, \qquad (4.4.18)$$

where $\pi_{[i]}$ is defined in (4.2.35) and

$$\varepsilon_{[i]} = w\begin{pmatrix} |d_i|^2\left(m_2-m_2^{[i]}\right) & 0 \\ 0 & m_1-m_1^{[i]} \end{pmatrix} + Z_{[i]}.$$

By (4.4.9) and (4.4.12), we have that $\varepsilon_{[i]} \prec |w|\Psi_\theta$. Let $B_i = \pi_{[i]}^{-1} - \pi_{[i]c}^{-1}$, where $\pi_{[i]c}$ is defined in (4.2.31). By (4.3.27) and the definition of Ξ, we have $\mathbf{1}(\Xi)\|B_i\pi_{[i]c}\| \leqslant C(\log N)^{-1}$. Thus we have the expansion

$$\mathbf{1}(\Xi)\pi_{[i]} = \mathbf{1}(\Xi)(\pi_{[i]c}^{-1}+B_i)^{-1} = \mathbf{1}(\Xi)\pi_{[i]c}\left(1-B_i\pi_{[i]c}+(B_i\pi_{[i]c})^2+\ldots\right) = \mathbf{1}(\Xi)(\pi_{[i]c}+\varepsilon_a), \qquad (4.4.19)$$

where ε_a can be estimated as $\mathbf{1}(\Xi)\|\varepsilon_a\| \leqslant \mathbf{1}(\Xi)C|w|^{-1/2}(\log N)^{-1}$. This shows that $\mathbf{1}(\Xi)\|\pi_{[i]}\| = \mathbf{1}(\Xi)O(|w|^{-1/2})$, and so $\mathbf{1}(\Xi)\left\|\varepsilon_{[i]}\pi_{[i]}\right\| \prec \mathbf{1}(\Xi)|w|^{1/2}\Psi_\theta \leqslant \mathbf{1}(\Xi)\,CN^{-\zeta/2}$ by the definition of $\mathbf{D}$ in (4.2.38). Again we do the expansion for (4.4.18):

$$\mathbf{1}(\Xi)G_{[ii]} = \mathbf{1}(\Xi)\left(\pi_{[i]}^{-1}+\varepsilon_{[i]}\right)^{-1} = \mathbf{1}(\Xi)\pi_{[i]}\left(1+\sum_{l=1}^{\infty}\left(-\varepsilon_{[i]}\pi_{[i]}\right)^l\right) = \mathbf{1}(\Xi)\left(\pi_{[i]}+\varepsilon_b\right), \qquad (4.4.20)$$

where $\mathbf{1}(\Xi)\|\varepsilon_b\| \prec \mathbf{1}(\Xi)\Psi_\theta$. Now the 11 entry of (4.4.20) gives that

$$\mathbf{1}(\Xi)G_{ii} = \mathbf{1}(\Xi)\frac{-1-m_1}{w\left(1+|d_i|^2 m_2\right)\left(1+m_1\right)-|z|^2} + \mathbf{1}(\Xi)\,\mathrm{O}_\prec\left(\Psi_\theta\right), \tag{4.4.21}$$

from which we get that

$$\mathbf{1}(\Xi)G_{ii}\left[-w\left(1+|d_i|^2 m_2\right)+\frac{|z|^2}{1+m_1}\right] = \mathbf{1}(\Xi)\left[1+\mathrm{O}_\prec\left(|w|^{1/2}\Psi_\theta\right)\right]. \tag{4.4.22}$$

Here we used that

$$\mathbf{1}(\Xi)\left[-w\left(1+|d_i|^2 m_2\right)+\frac{|z|^2}{1+m_1}\right] = \mathrm{O}(|w|^{1/2}),$$

which follows from Lemma 4.3.4 and the definition of Ξ. Summing (4.4.22) over i, we get

$$\mathbf{1}(\Xi)\left[-w\left(m_2+m_1 m_2\right)+\frac{|z|^2 m_2}{1+m_1}\right] = \mathbf{1}(\Xi)\left[1+\mathrm{O}_\prec\left(|w|^{1/2}\Psi_\theta\right)\right],$$

which gives

$$\mathbf{1}(\Xi)m_2 = \mathbf{1}(\Xi)\frac{1+m_1}{-w\left(1+m_1\right)^2+|z|^2} + \mathbf{1}(\Xi)\,\mathrm{O}_\prec\left(\Psi_\theta\right). \tag{4.4.23}$$

Now plugging (4.4.23) into (4.4.21), multiplying with $|d_i|^2$ and summing over i, we obtain that

$$\mathbf{1}(\Xi)m_1 = \mathbf{1}(\Xi)\left[\frac{1}{N}\sum_{i=1}^{n} l_i s_i \frac{-1-m_1}{w\left(1+s_i\frac{1+m_1}{-w(1+m_1)^2+|z|^2}\right)\left(1+m_1\right)-|z|^2} + \mathrm{O}_\prec\left(\Psi_\theta\right)\right], \tag{4.4.24}$$

where we used (4.3.25) and $\mathbf{1}(\Xi)(1+m_1) = \mathbf{1}(\Xi)O(|w|^{-1/2})$. This concludes the proof.

Similarly, when $w \in \mathbf{D}_L$, it is easy to prove (4.4.17) using the estimates (4.4.13) and (4.3.12). Note that $|m_{1,2}| = \mathrm{O}(\eta^{-1})$ by (4.3.12), which implies immediately the bounds $\|\pi_{[i]}\| = \mathrm{O}(\eta^{-1})$ and $\|\left(\pi_{[i]}\right)^{-1}\| = \mathrm{O}(\eta)$. Hence without introducing the event Ξ, we can obtain directly

$$G_{[ii]} = \pi_{[i]} + \mathrm{O}_\prec(\eta^{-1}(N\eta)^{-1/2}). \tag{4.4.25}$$

The rest of the proof is essentially the same. □

Notice that applying Lemma 4.3.7 to (4.4.17), we obtain that $|m_{1,2}-m_{1,2c}| \prec \eta^{-1}(N\eta)^{-1/2}$. Plugging it into (4.4.25), we immediately get (4.4.2) for $w \in \mathbf{D}_L$. This proves the entrywise law on $\mathbf{D}_L$, since $\eta^{-1}N^{-1/2} \leqslant C\Psi$ by the definition (4.2.44) and the estimate (4.3.24).

4.4.2 The large η case

It remains to prove Proposition 4.4.1 on domain $\mathbf{D}$. We would like to fix E and then apply a continuity argument in η by first showing that the rough bound $\Lambda \leqslant |w|^{-1/2}(\log N)^{-1}$ in Lemma 4.4.6 holds for large η. To start the argument, we first need to establish the estimates on G when $\eta \sim 1$. The next lemma is a trivial consequence of (4.3.12).

Lemma 4.4.7. *For any $w \in \mathbf{D}$ and $\eta \geqslant c$ for fixed $c > 0$, we have the bound*

$$\max_{s,t} |G_{st}(w)| \leqslant C \tag{4.4.26}$$

for some $C > 0$. This estimate also holds if we replace G with $G^{(J)}$ for $J \subset \mathcal{I}$.

Lemma 4.4.8. *Fix $c > 0$ and $|z|^2 \leqslant 1 - \tau$. We have the following estimate*

$$\max_{w \in \mathbf{D}, \eta \geqslant c} \Lambda(w) \prec N^{-1/2}. \tag{4.4.27}$$

Proof. By the previous lemma, we have $|m_{1,2}^{[i]}| = \mathrm{O}(1)$. So by Lemma 4.4.5, $\|Z_{[i]}\| \prec N^{-1/2}$ uniformly in $\eta \geqslant c$. Then as in (4.4.18), we have

$$G_{[ii]} = \left(\pi_{[i]}^{-1} + \varepsilon_{[i]}\right)^{-1}, \tag{4.4.28}$$

where $\|\pi_{[i]}^{-1}\| = \mathrm{O}(1)$ and $\|\varepsilon_{[i]}\| \prec N^{-1/2}$. Notice since $G_{[ii]} = \mathrm{O}(1)$, we have the estimate

$$\pi_i = \left(G_{[ii]}^{-1} - \varepsilon_{[i]}\right)^{-1} = G_{[ii]}\left(1 - \varepsilon_{[i]} G_{[ii]}\right)^{-1} = \mathrm{O}_\prec(1).$$

Then we can expand (4.4.28) to get that

$$G_{[ii]} = \pi_i + \mathrm{O}_\prec\left(N^{-1/2}\right). \tag{4.4.29}$$

The 11 and 22 entries of (4.4.29) lead to the equations

$$m_1 = \frac{1}{N}\sum_{i=1}^N |d_i|^2 \left[-w\left(1 + |d_i|^2 m_2\right) + \frac{|z|^2}{1 + m_1}\right]^{-1} + \mathrm{O}_\prec\left(N^{-1/2}\right), \tag{4.4.30}$$

$$m_2 = \frac{1}{N}\sum_{i=1}^N \left[-w\left(1 + m_1\right) + \frac{|z|^2}{1 + |d_i|^2 m_2}\right]^{-1} + \mathrm{O}_\prec\left(N^{-1/2}\right). \tag{4.4.31}$$

We claim that $\operatorname{Im} m_{1,2} \geqslant C(\log N)^{-1}$ with high probability for some $C > 0$.

Using the spectral decomposition (4.3.11), we note that for $l > 1$,

$$\frac{1}{N} \sum_{|\lambda_k - E| \geqslant l\eta} \frac{|E - \lambda_k|}{(\lambda_k - E)^2 + \eta^2} \leqslant \frac{1}{l\eta},$$
$$\frac{1}{N} \sum_{|\lambda_k - E| \leqslant l\eta} \frac{|E - \lambda_k|}{(\lambda_k - E)^2 + \eta^2} \leqslant \frac{1}{N} \sum_{|\lambda_k - E| \leqslant l\eta} \frac{l\eta}{(\lambda_k - E)^2 + \eta^2} \leqslant l \operatorname{Im} m_2.$$

Summing up these two inequalities and optimizing l, we get

$$|\operatorname{Re} m_2| \leqslant 2\sqrt{\frac{\operatorname{Im} m_2}{\eta}}. \tag{4.4.32}$$

Assume that $\operatorname{Im} m_2 \leqslant C(\log N)^{-1}$, then by (4.4.6) we also have $\operatorname{Im} m_1 \leqslant C\tau^{-1}(\log N)^{-1}$. From (4.4.32), we get $|m_2| \leqslant C(\log N)^{-1/2}$. Together with the estimate $m_1 = \mathrm{O}(1)$, we get

$$\left| -w\left(1 + m_1\right) + \frac{|z|^2}{1 + |d_i|^2 m_2} \right| \leqslant C \text{ with high probability.} \tag{4.4.33}$$

On the other hand

$$\operatorname{Im}\left[-w\left(1 + m_1\right) + \frac{|z|^2}{1 + |d_i|^2 m_2} \right] \leqslant -\operatorname{Im} w = -\eta, \tag{4.4.34}$$

where we used $\operatorname{Im}[|z|^2/(1 + |d_i|^2 m_2)] < 0$ and

$$\operatorname{Im}(w m_1) = \operatorname{Im}\left[\frac{1}{N} \sum_{k=1}^{N} |d_i|^2 |\xi_k(i)|^2 \left(-1 + \frac{\lambda_k}{\lambda_k - w} \right) \right] \geqslant 0.$$

With (4.4.33) and (4.4.34), we get from (4.4.31) that $\operatorname{Im} m_2 \geqslant c'$ with high probability for some $c' > 0$. This contradicts $\operatorname{Im} m_2 \leqslant C(\log N)^{-1}$. Thus we must have $\operatorname{Im} m_2 \geqslant C(\log N)^{-1}$ with high probability, which also implies $\operatorname{Im} m_1 \geqslant C(\log N)^{-1}$ by (4.4.6).

Now we can proceed as in the proof of Lemma 4.4.6 and get that

$$m_2 = \frac{1 + m_1}{-w\left(1 + m_1\right)^2 + |z|^2} + \mathrm{O}_\prec\left(N^{-1/2}\right), \quad \Upsilon(w, m_1) \prec N^{1/2}. \tag{4.4.35}$$

We omit the details. Applying Lemma 4.3.7 to (4.4.35), we conclude $|m_{1,2} - m_{1,2c}| \prec N^{-1/2}$ uniformly in $\eta \geqslant c$. By (4.4.29), we get $\|(G - \Pi)_{[ii]}\| \prec N^{-1/2}$ uniformly in $\eta \geqslant c$ and $i \in \mathcal{I}_1$. Finally using (4.3.8), Lemma 4.3.3 and Lemma 2.6.6, we can prove the off-diagonal estimate; see (4.4.48) below. $\square$

4.4.3 Weak entrywise local law

In this subsection, we finish the proof of Proposition 4.4.1 on domain **D**. We shall fix the real part E of $w = E + \mathrm{i}\eta$ and decrease the imaginary part η. Recall that Lemma 4.4.6 is based on the condition $\theta \leqslant |w|^{-1/2}(\log N)^{-1}$. So far this is established only for large η in (4.4.27). We want to show that this condition also holds for small η by using a continuity argument.

It is convenient to introduce the random function

$$v(w) = \max_{w' \in L(w)} \theta(w')|w'|^{1/2} \left(\frac{N \mathrm{Im}\, w'}{|w'|^{1/2}} \right)^{1/4},$$

where $L(w)$ is defined in (4.3.32). Fix a regular domain **S**, $\varepsilon < \zeta/4$ and a large constant $D > 0$. Our goal is to prove that with high probability there is a gap in the range of v, i.e.

$$\mathbb{P}\left(v(w) \leqslant N^{\varepsilon}, v(w) > N^{3\varepsilon/4} \right) \leqslant N^{-D+21} \tag{4.4.36}$$

for all $w \in \mathbf{S}$ and large enough $N \geqslant N(\varepsilon, D)$.

Suppose $v(w) \leqslant N^{\varepsilon}$, then it is easy to verify that

$$\theta(w') \leqslant C|w'|^{-1/2}(\log N)^{-1} \tag{4.4.37}$$

for all $w' \in L(w)$. Hence $\{v(w) \leqslant N^{\varepsilon}\} \subset \Xi(w')$ for all $w' \in \mathbf{S} \cap L(w)$. Then by (4.4.16), for all $w' \in \mathbf{S} \cap L(w)$, there exists an $N_0 \equiv N_0(\varepsilon, D)$ such that

$$P\left(v(w) \leqslant N^{\varepsilon}, \Upsilon(w') > \frac{N^{\varepsilon}}{|w'|^{1/2}} \sqrt{\frac{|w'|^{1/2}}{N \mathrm{Im}\, w'}} \right) \leqslant N^{-D}, \tag{4.4.38}$$

for all $N > N_0$. Taking the union bound we get

$$P\left(v(w) \leqslant N^{\varepsilon}, \max_{w' \in L(w)} \Upsilon(w') \sqrt{\frac{N \mathrm{Im}\, w'}{|w'|^{-1/2}}} > N^{\varepsilon} \right) \leqslant N^{-D+10}. \tag{4.4.39}$$

Now consider the event

$$\Xi_1 := \left\{ v(w) \leqslant N^{\varepsilon}, \max_{w' \in L(w)} \Upsilon(w') \sqrt{\frac{N \mathrm{Im}\, w'}{|w'|^{-1/2}}} \leqslant N^{\varepsilon} \right\}. \tag{4.4.40}$$

We have $1(\Xi_1)\Upsilon(w') \leqslant \delta(w')$ for all $w' \in L(w)$ with $\delta(w') := \frac{N^\varepsilon}{|w'|^{1/2}}\sqrt{\frac{|w'|^{1/2}}{N\operatorname{Im} w'}}$. We now apply Lemma 4.3.7. If $\kappa \ll 1$ (recall (4.3.17)), then $|w| \sim 1$ and we have

$$1(\Xi_1)|m_1(w') - m_{1c}(w')| \leqslant C\sqrt{\delta(w')} \leqslant CN^{\varepsilon/2}\left(\frac{1}{N\operatorname{Im} w'}\right)^{1/4}$$

for all $w' \in L(w)$; if $\kappa \geqslant c > 0$ for some constant $c > 0$, then

$$1(\Xi_1)|m_1(w') - m_{1c}(w')| \leqslant C\delta(w') \leqslant C\frac{N^\varepsilon}{|w'|^{1/2}}\left(\frac{|w'|^{1/2}}{N\operatorname{Im} w'}\right)^{1/2}$$

for all $w' \in L(w)$. Combining these two cases we get

$$1(\Xi_1)|m_1(w') - m_{1c}(w')| \leqslant C\frac{N^{\varepsilon/2}}{|w'|^{1/2}}\left(\frac{|w'|^{1/2}}{N\operatorname{Im} w'}\right)^{1/4} \tag{4.4.41}$$

for all $w' \in L(w)$. By (4.4.16), we have

$$1(\Xi_1)|m_2(w') - m_{2c}(w')| \prec 1(\Xi_1)|m_1(w') - m_{1c}(w')| + 1(\Xi_1)\Psi_\theta \prec \frac{N^{\varepsilon/2}}{|w'|^{1/2}}\left(\frac{|w'|^{1/2}}{N\operatorname{Im} w'}\right)^{1/4},$$

for all $w' \in \mathbf{S} \cap L(w)$. Together with (4.4.41), this shows that there exists an $N_1 \equiv N_1(\varepsilon, D)$ such that

$$\mathbb{P}\left(v(w) \leqslant N^\varepsilon, \max_{w' \in L(w)} \Upsilon(w')\sqrt{\frac{N\operatorname{Im} w'}{|w'|^{-1/2}}} \leqslant N^\varepsilon, \max_{w' \in L(w)} \theta(w')|w'|^{1/2}\left(\frac{N\operatorname{Im} w'}{|w'|^{1/2}}\right)^{1/4} > N^{3\varepsilon/4}\right) \leqslant N^{-D} \tag{4.4.42}$$

for $N \geqslant \max\{N_0, N_1\}$. Adding (4.4.39) and (4.4.42), we get

$$\mathbb{P}\left(v(w) \leqslant N^\varepsilon, \max_{w' \in L(w)} \theta(w')|w'|^{1/2}\left(\frac{N\operatorname{Im} w'}{|w'|^{1/2}}\right)^{1/4} > N^{3\varepsilon/4}\right) \leqslant N^{-D+11}.$$

Taking the union bound over $L(w)$ we get (4.4.36) for all $N \geqslant \max\{N_0, N_1\}$.

Now we conclude the proof of Proposition 4.4.1 by combining (4.4.36) with the large η estimate (4.4.27). We choose a lattice $\Delta \subset \mathbf{S}$ such that $|\Delta| \leqslant N^{20}$ and for any $w \in \mathbf{S}$ there is a $w' \in \Delta$ with $|w' - w| \leqslant N^{-9}$. Taking the union bound we get

$$\mathbb{P}\left(\exists w \in \Delta : v(w) \in (N^{3\varepsilon/4}, N^\varepsilon]\right) \leqslant N^{-D+41}. \tag{4.4.43}$$

Since v has Lipshcitz constant bounded by, say, N^6, then we have

$$\mathbb{P}\left(\exists w \in \mathbf{S} : v(w) \in (2N^{3\varepsilon/4}, N^{\varepsilon}/2]\right) \leqslant N^{-D+41}. \tag{4.4.44}$$

Combining with (4.4.27), we see that there exists $N_2 \equiv N_2(\varepsilon, D)$ such that for all $N > N_2$,

$$\mathbb{P}\left(\forall w \in \mathbf{S} : v(w) \leqslant 2N^{3\varepsilon/4}\right) \geqslant 1 - 2N^{-D+41}.$$

Since ε and D are arbitrary, the above inequality shows that $v(w) \prec 1$ uniformly in $w \in \mathbf{S}$, or

$$\theta(w) \prec \frac{1}{|w|^{1/2}} \left(\frac{|w|^{1/2}}{N\eta}\right)^{1/4}. \tag{4.4.45}$$

In particular this shows that for all $w \in \mathbf{S}$, the event Ξ holds with high-probability.

Now using (4.4.20) and (4.4.45), we get

$$\left\|G_{[ii]} - \pi_{[i]^c}\right\| \leqslant \left\|G_{[ii]} - \pi_{[i]}\right\| + \left\|\pi_{[i]} - \pi_{[i]^c}\right\| \prec \Psi_\theta + \theta \prec \frac{1}{|w|^{1/2}} \left(\frac{|w|^{1/2}}{N\eta}\right)^{1/4}. \tag{4.4.46}$$

To conclude Proposition 4.4.1, it remains to prove the estimate for the off-diagonal $G_{[ij]}$ groups. Using (4.4.9), it is not hard to get that

$$\left\|G_{[ii]}^{[J]} - \pi_{[i]^c}\right\| \prec \frac{1}{|w|^{1/2}} \left(\frac{|w|^{1/2}}{N\eta}\right)^{1/4} \tag{4.4.47}$$

for any $|J| \leqslant l$ with $l \in \mathbb{N}$ fixed. Thus we have $\left\|G_{[ii]}^{[J]}\right\| = \mathrm{O}\left(|w|^{-1/2}\right)$ and $\left\|\left(G_{[ii]}^{[J]}\right)^{-1}\right\| = \mathrm{O}\left(|w|^{1/2}\right)$ with high probability. Let $i \neq j \in \mathcal{I}_1$, using (4.3.8) and the above diagonal estimates, we get that

$$\left\|G_{[ij]}\right\| \prec |w|^{-1} \frac{|w|^{1/2}}{\sqrt{N}} + |w|^{-1} \left\|\sum_{k,l \notin \{i,j\}} H_{[ik]} G_{[kl]}^{[ij]} H_{[lj]}\right\| \prec \Psi_\theta \prec \frac{1}{|w|^{1/2}} \left(\frac{|w|^{1/2}}{N\eta}\right)^{1/4}, \tag{4.4.48}$$

where we used Lemma 4.3.3 and Lemma 2.6.6 to obtain that

$$|w|^{-1} \left\|\sum_{k,l \notin \{i,j\}} H_{[ik]} G_{[kl]}^{[ij]} H_{[lj]}\right\| = \left\|\begin{pmatrix} \sum_{k,l \notin \{i,j\}} X_{i\bar{k}} G_{\bar{k}\bar{l}}^{[ij]} X_{\bar{l}j}^* & \sum_{k,l \notin \{i,j\}} X_{i\bar{k}} G_{\bar{k}l}^{[ij]} X_{l\bar{j}} \\ \sum_{k,l \notin \{i,j\}} X_{\bar{i}k}^* G_{k\bar{l}}^{[ij]} X_{\bar{l}j}^* & \sum_{k,l \notin \{i,j\}} X_{\bar{i}k}^* G_{kl}^{[ij]} X_{l\bar{j}} \end{pmatrix}\right\| \prec \Psi_\theta. \tag{4.4.49}$$

Its proof is very similar to the proof of Lemma 4.4.5, so we omit the details.

4.4.4 Strong enterywise local law

In this section, we finish the proof of the (strong) entrywise local law and averaged local law in Theorem 4.2.22 on domain $\mathbf{D}$ and under the condition $|w|^{1/2}+|z|^2 \geqslant c$. In Lemma 4.4.6, we have proved an error estimate of the self-consistent equations of $m_{1,2}$ linearly in Ψ_θ. The core part of the proof is to improve this estimate to quadratic in Ψ_θ. For the sequence of random variables $Z_{[i]}$, we define the averaged quantities

$$[Z] = \frac{1}{N}\sum_{i=1}^{N} \pi_{[i]} Z_{[i]} \pi_{[i]}, \quad \langle Z \rangle = \frac{1}{N}\sum_{i=1}^{N} |d_i|^2 \pi_{[i]} Z_{[i]} \pi_{[i]}.$$

The following Lemma gives an improvement of Lemma 4.4.6.

Lemma 4.4.9. *Fix* $|z|^2 \leqslant 1-\tau$. *Then for* $w \in \mathbf{D}$*,*

$$m_2 = \frac{1+m_1}{-w\left(1+m_1\right)^2+|z|^2} + \mathrm{O}_\prec\left(|w|^{1/2}\Psi_\theta^2 + \|[Z]\| + \|\langle Z\rangle\|\right), \tag{4.4.50}$$

and

$$\Upsilon(w, m_1) \prec |w|^{1/2}\Psi_\theta^2 + \|[Z]\| + \|\langle Z\rangle\|. \tag{4.4.51}$$

For $w \in \mathbf{D}_L$*,*

$$m_2 = \frac{1+m_1}{-w\left(1+m_1\right)^2+|z|^2} + \mathrm{O}_\prec\left((N\eta)^{-1} + \|[Z]\| + \|\langle Z\rangle\|\right), \tag{4.4.52}$$

and

$$\Upsilon(w, m_1) \prec (N\eta)^{-1} + \|[Z]\| + \|\langle Z\rangle\|. \tag{4.4.53}$$

Proof. The proof is almost the same as the one for Lemma 4.4.6, we only lay out the difference. We first consider the case $w \in \mathbf{D}$. By Proposition 4.4.1, the event Ξ holds with high probability. Hence without loss of generality, we may assume Ξ holds throughout the following proof. Using (4.3.9), we get

$$\frac{1}{N}\sum_{k\in\mathcal{I}_1}\begin{pmatrix} |d_k|^2 & 0 \\ 0 & 1\end{pmatrix}\left(G_{[kk]} - G_{[kk]}^{[i]}\right) = \begin{pmatrix} |d_i|^2 & 0 \\ 0 & 1\end{pmatrix}\frac{G_{[ii]}}{N} + \frac{1}{N}\sum_{k\neq i}\begin{pmatrix} |d_k|^2 & 0 \\ 0 & 1\end{pmatrix} G_{[ki]} G_{[ii]}^{-1} G_{[ik]}. \tag{4.4.54}$$

By Proposition 4.4.1, (4.3.27) and (4.4.48), we have

$$\left\| G_{[ki]} G_{[ii]}^{-1} G_{[ik]} \right\| \prec |w|^{1/2} \Psi_\theta^2.$$

By Lemma 4.3.4, it is easy to verify that $\|G_{[ii]}/N\| \leqslant C|w|^{1/2}\Psi_\theta^2$. Plugging it into (4.4.54), we get

$$\left| m_{1,2}^{[i]} - m_{1,2} \right| \prec |w|^{1/2} \Psi_\theta^2. \tag{4.4.55}$$

By (4.4.12) and (4.4.55), the error ε_b in (4.4.20) is

$$\varepsilon_b = \mathrm{O}_\prec(|w|^{1/2}\Psi_\theta^2) - \pi_{[i]} Z_{[i]} \pi_{[i]} \left[1 + \mathrm{O}_\prec(|w|^{1/2}\Psi_\theta) \right] = \mathrm{O}_\prec(|w|^{1/2}\Psi_\theta^2) - \pi_{[i]} Z_{[i]} \pi_{[i]}.$$

Then following the arguments in Lemma 4.4.6, we can prove the desired result. For $w \in \mathbf{D}_L$, the proof is similar by using (4.4.2). □

In the following lemma, we shall give stronger bounds on $[Z]$ and $\langle Z \rangle$ by keeping track of the cancellation effects due to the average over the index i.

Lemma 4.4.10. *(Fluctuation averaging) Fix $|z|^2 \leqslant 1 - \tau$. Suppose Φ and Φ_o are positive, N-dependent deterministic functions satisfying $N^{-1/2} \leqslant \Phi, \Phi_o \leqslant N^{-c}$ for some constant $c > 0$. Suppose moreover that $\Lambda \prec |w|^{-1/2}\Phi$ and $\Lambda_o \prec |w|^{-1/2}\Phi_o$. Then for $w \in \mathbf{D}$,*

$$\|[Z]\| + \|\langle Z \rangle\| \prec |w|^{-1/2} \, \Phi_o^2. \tag{4.4.56}$$

Proof. Our proof of (4.4.56) is an extension of [14, Lemma 4.9], [18, Lemma 7.3] and [33, Theorem 4.7]. Here we only prove the bound for $\|[Z]\|$. The proof for $\|\langle Z \rangle\|$ is exactly the same. For $i \in \mathcal{I}_1$, we define $P_i := \mathbb{E}_{[i]}$ and $Q_i := 1 - P_i$. Recall that $Z_{[i]} = Q_i G_{[ii]}^{-1}$. Hence we need to prove

$$[Z] = \frac{1}{N} \sum_{i=1}^N \pi_{[i]} \left(Q_i G_{[ii]}^{-1} \right) \pi_{[i]} \prec |w|^{-1/2} \, \Phi_o^2,$$

for $w \in \mathbf{D}$. For $J \subset \mathcal{I}$, we define $\pi_{[i]}^{[J]}$ by replacing $m_{1,2}$ in (4.2.35) with $m_{1,2}^{[J]}$ defined in (4.4.4). As in (4.4.55), we can prove that $|m_{1,2}^{[i]} - m_{1,2}| \prec |w|^{-1/2} \, \Phi_o^2$, which further gives that

$$[Z] = \frac{1}{N} \sum_{i=1}^N \pi_{[i]}^{[i]} \left(Q_i G_{[ii]}^{-1} \right) \pi_{[i]}^{[i]} + \mathrm{O}_\prec \left(|w|^{-1/2} \, \Phi_o^2 \right) = \frac{1}{N} \sum_{i=1}^N Q_i \left(\pi_{[i]}^{[i]} G_{[ii]}^{-1} \pi_{[i]}^{[i]} \right) + \mathrm{O}_\prec \left(|w|^{-1/2} \, \Phi_o^2 \right).$$

Thus if we abbreviate $B_i := |w|^{1/2} Q_i \left(\pi_{[i]}^{[i]} G_{[ii]}^{-1} \pi_{[i]}^{[i]}\right)$, it suffices to prove that $B := N^{-1} \sum_i B_i \prec \Phi_o^2$. We will estimate B by bounding the p-th moment of its norm by Φ_o^{2p} for $p = 2n \in 2\mathbb{N}$, i.e. $\mathbb{E}\|B\|^p \prec \Phi_o^{2p}$. The lemma then follows from the Markov's inequality. Using $\|KK^*\| = \|K\|^2$, we have that

$$\operatorname{Tr}(BB^*)^n \geqslant \|BB^*\|^n = \|B\|^{2n}.$$

Thus it suffices to prove that

$$\mathbb{E}\operatorname{Tr}(BB^*)^{p/2} \prec \Phi_o^{2p}, \quad \text{for } p = 2n. \tag{4.4.57}$$

This estimate can be proved with the same method as in [33, Appendix B], with the only complication being that $\pi_{[i]}$ is random and depends on i. In principle, this can be handle by using (4.3.9) and (4.3.10) to put any indices $j, k, ... \in \mathcal{I}_1$ (that we wish to include) into the superscripts of $\pi_{[i]}$. This leads to a minor modification of the proof in [33, Appendix B]. Here we describe the basic ideas of the proof, without writing down all the details.

The proof is based on a decomposition of the space of random variables using P_s and Q_s. It is evident that P_s and Q_s are projections, $P_s + Q_s = 1$ and all of these projections commute with each other. For a set $J \subset \mathcal{I}$, we denote $P_J := \prod_{s \in J} P_s$ and $Q_J := \prod_{s \in J} Q_s$. Let $p = 2n$ and introduce the shorthand notation $\tilde{B}_{k_s} := B_{k_s}$ for odd $s \leqslant p$ and $\tilde{B}_{k_s} := B_{k_s}^*$ for even $s \leqslant p$. Then we get

$$\mathbb{E}\operatorname{Tr}(BB^*)^{p/2} = \frac{1}{N^p} \sum_{k_1, k_2, \ldots, k_p} \mathbb{E}\operatorname{Tr} \prod_{s=1}^{p} \tilde{B}_{k_s} = \frac{1}{N^p} \sum_{k_1, k_2, \ldots, k_p} \mathbb{E}\operatorname{Tr} \prod_{s=1}^{p} \left(\prod_{r=1}^{p} (P_{k_r} + Q_{k_r}) \tilde{B}_{k_s} \right). \tag{4.4.58}$$

Introducing the notations $\mathbf{k} = (k_1, k_2, \ldots, k_p)$ and $\{\mathbf{k}\} = \{k_1, k_2, \ldots, k_p\}$, we can write

$$\mathbb{E}\operatorname{Tr}(BB^*)^{p/2} = \frac{1}{N^p} \sum_{\mathbf{k}} \sum_{I_1, \ldots, I_p \subset \{\mathbf{k}\}} \mathbb{E}\operatorname{Tr} \prod_{s=1}^{p} \left(P_{I_s^c} Q_{I_s} \tilde{B}_{k_s} \right). \tag{4.4.59}$$

Following the arguments in [33, Appendix B], one can see that to conclude (4.4.57) it suffices to prove that for $k \in I$,

$$\|Q_I B_k\| \prec \Phi_o^{|I|}. \tag{4.4.60}$$

As in [33, Appendix B], it is not hard to prove that for $k \in I$,

$$|w|^{-1/2} \left\| Q_I G_{[kk]}^{-1} \right\| \prec \Phi_o^{|I|}, \quad \text{and} \quad |w|^{-1/2} \left\| Q_{I \setminus \{k\}} G_{[kk]}^{-1} \right\| \prec \Phi_o^{|I|} \text{ if } |I| \geqslant 2. \tag{4.4.61}$$

Now we extend the proof to obtain the estimate (4.4.60). For the case $|I| = 1$ (i.e. $I = \{k\}$),

$$\|B_k\| = |w|^{1/2} \|\pi_{[i]}^{[i]} Z_{[k]} \pi_{[i]}^{[i]}\| \leqslant |w|^{-1/2} \|Z_{[k]}\| \prec \Phi_o,$$

where we used $\|Z_{[k]}\| \prec |w|^{1/2}\Phi_o$, which can be proved with the same arguments as in Lemma 4.4.5. For the case $|I| \geqslant 2$, WLOG, we may assume $k = 1$ and $I = \{1, \ldots, t\}$ with $t \geqslant 2$. It is enough to prove that

$$|w|^{1/2} \left\| Q_t \ldots Q_2 Q_1 \pi_{[1]}^{[1]} G_{[11]}^{-1} \pi_{[1]}^{[1]} \right\| \prec \Phi_o^t. \tag{4.4.62}$$

We take $t = 3$ as an example to describe the ideas for the proof of (4.4.62). Using (4.3.9), we get

$$\pi_{[1]}^{[1]} = \pi_{[1]}^{[12]} + |w|^{1/2} \varepsilon_{11}^{[1]} \pi_{[1]}^{[12]} A_1 \pi_{[1]}^{[12]} + |w|^{1/2} \varepsilon_{\bar{1}\bar{1}}^{[1]} \pi_{[1]}^{[12]} A_2 \pi_{[1]}^{[12]} + \mathrm{error}_{1,2}, \tag{4.4.63}$$

where $\varepsilon_{11}^{[1]}$ and $\varepsilon_{\bar{1}\bar{1}}^{[1]}$ are the upper left and lower right entries of

$$\varepsilon_{[1]}^{[1]} := |w|^{1/2} \left(\frac{G_{[22]}^{[1]}}{N} + \frac{1}{N} \sum_{k \notin \{1,2\}} G_{[k2]}^{[1]} \left(G_{[22]}^{[1]} \right)^{-1} G_{[2k]}^{[1]} \right) \prec \Phi_o^2,$$

$A_{1,2}$ are deterministic matrices with operator norm $O(1)$, and $\|\mathrm{error}_{1,2}\| \prec |w|^{-1/2}\Phi_o^4$. Then we get

$$\begin{aligned} \pi_{[1]}^{[1]} G_{[11]}^{-1} \pi_{[1]}^{[1]} &= \pi_{[1]}^{[12]} G_{[11]}^{-1} \pi_{[1]}^{[12]} + |w|^{1/2} \varepsilon_{11}^{[1]} \pi_{[1]}^{[12]} A_1 \pi_{[1]}^{[12]} G_{[11]}^{-1} \pi_{[1]}^{[12]} + |w|^{1/2} \varepsilon_{\bar{1}\bar{1}}^{[1]} \pi_{[1]}^{[12]} A_2 \pi_{[1]}^{[12]} G_{[11]}^{-1} \pi_{[1]}^{[12]} \\ &+ |w|^{1/2} \pi_{[1]}^{[12]} G_{[11]}^{-1} \varepsilon_{11}^{[1]} \pi_{[1]}^{[12]} A_1 \pi_{[1]}^{[12]} + |w|^{1/2} \pi_{[1]}^{[12]} G_{[11]}^{-1} \varepsilon_{\bar{1}\bar{1}}^{[1]} \pi_{[1]}^{[12]} A_2 \pi_{[1]}^{[12]} + \mathrm{O}_\prec(|w|^{-1/2}\Phi_o^4). \end{aligned} \tag{4.4.64}$$

We first handle the $\pi_{[1]}^{[12]} G_{[11]}^{-1} \pi_{[1]}^{[12]}$ term. By (4.4.61), we have

$$Q_2 \pi_{[1]}^{[12]} G_{[11]}^{-1} \pi_{[1]}^{[12]} = \pi_{[1]}^{[12]} \left(Q_2 G_{[11]}^{-1} \right) \pi_{[1]}^{[12]} \prec |w|^{-1/2} \Phi_o^2.$$

For the remaining term, we first expand $\pi_{[1]}^{[12]} = \pi_{[1]}^{[123]} + \mathrm{O}_\prec(|w|^{-1/2}\Phi_o^2)$ and use (4.4.61) to get

$$Q_3 Q_2 \pi_{[1]}^{[12]} G_{[11]}^{-1} \pi_{[1]}^{[12]} = \pi_{[1]}^{[123]} \left(Q_3 Q_2 G_{[11]}^{-1} \right) \pi_{[1]}^{[123]} + \mathrm{O}_\prec \left(|w|^{-1/2} \Phi_o^4 \right) \prec |w|^{-1/2} \Phi_o^3.$$

Then we deal with the second terms in (4.4.64). We first expand $\varepsilon_{[1]}^{[1]} = e_{[1]}^{[3]} + \mathrm{O}_\prec(\Phi_o^3)$, where

$$e_{[1]}^{[3]} := |w|^{1/2} \left(\frac{G_{[22]}^{[13]}}{N} + \frac{1}{N} \sum_{k \notin \{1,2,3\}} G_{[k2]}^{[13]} \left(G_{[22]}^{[13]} \right)^{-1} G_{[2k]}^{[13]} \right).$$

Using the similar arguments as above, we get

$$
\begin{aligned}
Q_3|w|^{1/2}e_{11}^{[3]}\pi_{[1]}^{[12]}A_1\pi_{[1]}^{[12]}G_{[11]}^{-1}\pi_{[1]}^{[12]} &= |w|^{1/2}e_{11}^{[3]}\pi_{[1]}^{[123]}A_1\pi_{[1]}^{[123]}\left(Q_3G_{[11]}^{-1}\right)\pi_{[1]}^{[123]} + \mathrm{O}_{\prec}(|w|^{-1/2}\Phi_o^4) \\
&\prec |w|^{-1/2}\Phi_o^4.
\end{aligned}
$$

Thus we have

$$
Q_2Q_3|w|^{1/2}\varepsilon_{11}^{[1]}\pi_{[1]}^{[12]}A_1\pi_{[1]}^{[12]}G_{[11]}^{-1}\pi_{[1]}^{[12]} \prec |w|^{-1/2}\Phi_o^3.
$$

Obviously this kind of estimate works for the rest of the terms in (4.4.64). This proves (4.4.62) when $t = 3$.

We can continue in this manner for a general t. At the l-th step, we expand the leading order terms using (4.3.9) and (4.3.10), and after applying $Q_l \dots Q_3Q_2$ on them, the number of Φ_o factors increases by one at each step by (4.4.61). Trough induction we can prove (4.4.62). In fact the expansions can be performed in a systematic way using the method in [33, Appendix B], and we leave the details to the reader. Also we remark that similar techniques are used in the proof of anisotropic local law in Section 4.5, and we choose to present the details there (in fact the proof here is much easier than the one in Section 4.5). □

Now we finish the proof of the entrywise local law and averaged local law on the domain **D**. By Proposition 4.4.1, we can take

$$
\Phi_o = |w|^{1/2}\sqrt{\frac{\mathrm{Im}(m_{1c}+m_{2c}) + |w|^{-3/8}(N\eta)^{-1/4}}{N\eta}}, \quad \Phi = \left(\frac{|w|^{1/2}}{N\eta}\right)^{1/4},
$$

in Lemma 4.4.10, with $\Lambda_o \prec \Psi_\theta \prec |w|^{-1/2}\Phi_o$ and $\Lambda \prec \Psi_\theta + \theta \prec |w|^{-1/2}\Phi$. Then (4.4.51) gives

$$
\Upsilon(w, m_1) \prec \frac{|w|^{1/2}\mathrm{Im}(m_{1c}+m_{2c}) + |w|^{1/4}(N\eta)^{-1/4}}{N\eta}.
$$

Using the stability Lemma 4.3.7, we get

$$
|m_1 - m_{1c}| \prec \frac{|w|^{1/2}\mathrm{Im}(m_{1c}+m_{2c})}{N\eta\sqrt{\kappa+\eta}} + \frac{|w|^{1/8}}{(N\eta)^{5/8}} \prec \frac{1}{N\eta} + \frac{|w|^{1/8}}{(N\eta)^{5/8}} \prec |w|^{-1/2}\left(\frac{|w|^{1/2}}{N\eta}\right)^{1/2+1/8}.
$$

Here if $\sqrt{\kappa+\eta} \geqslant (\log N)^{-1}$, we use

$$
\frac{|w|^{1/2}\mathrm{Im}(m_{1c}+m_{2c})}{N\eta\sqrt{\kappa+\eta}} \leqslant \frac{C\log N}{N\eta} \prec \frac{1}{N\eta};
$$

if $\sqrt{\kappa+\eta} \leqslant (\log N)^{-1}$, we have $\mathrm{Im}(m_{1c}+m_{2c}) = O(\sqrt{\kappa+\eta})$, which also gives that

$$\frac{|w|^{1/2}\mathrm{Im}(m_{1c}+m_{2c})}{N\eta\sqrt{\kappa+\eta}} \prec \frac{1}{N\eta}.$$

We then use (4.4.50) to get that

$$\theta \prec |m_1 - m_{1c}| + \frac{|w|^{1/2}\mathrm{Im}(m_{1c}+m_{2c}) + |w|^{1/4}(N\eta)^{-1/4}}{N\eta} \prec |w|^{-1/2}\left(\frac{|w|^{1/2}}{N\eta}\right)^{1/2+1/8}. \quad (4.4.65)$$

Repeating the previous steps with the new estimate (4.4.65), we get the bound

$$\theta \prec |w|^{-1/2}\left(\frac{|w|^{1/2}}{N\eta}\right)^{\sum_{k=1}^{l} 1/2^k + 1/2^{l+2}}$$

after l iterations. This implies the averaged local law $\theta \prec (N\eta)^{-1}$ since l can be arbitrarily large. Finally as in (4.4.46) and (4.4.48), we have for $i \neq j \in \mathcal{I}_1$,

$$\left\|G_{[ii]} - \pi_{[i]c}\right\| + \left\|G_{[ij]}\right\| \prec \Psi_\theta + \theta \prec \sqrt{\frac{\mathrm{Im}(m_{1c}+m_{2c})}{N\eta}} + \frac{1}{N\eta}.$$

This concludes the proof of the entrywise local law and averaged local law on domain $\mathbf{D}$ when $|w|^{1/2} + |z|^2 \geqslant c$.

When $w \in \mathbf{D}_L$, we have proved the entrywise law (see the remark after (4.4.25)). Also we can prove a similar estimate as in Lemma 4.4.10, which implies

$$m_2 = \frac{1+m_1}{-w\left(1+m_1\right)^2 + |z|^2} + \mathrm{O}_\prec\left((N\eta)^{-1}\right), \quad \Upsilon(w, m_1) \prec (N\eta)^{-1}. \quad (4.4.66)$$

The averaged local law then follows from Lemma 4.3.7. We leave the details to the reader.

4.4.5 The small $|z|$ and $|w|$ case

In the previous proof, we did not include the case where $|w|^{1/2} + |z|^2 \leqslant \varepsilon$ for some sufficiently small constant $\varepsilon > 0$. The only reason is that Lemma 4.3.7 does not apply in this case. We deal with this problem in this subsection.

The main idea of this subsection is to use a different set of self-consistent equations, which has the desired stability when $|w|$ and $|z|$ are small. Multiplying (4.4.21) with $|d_i|^2$ and summing over i, we get

$$1(\Xi)m_1 = 1(\Xi)\left[\frac{1}{N}\sum_{i=1}^{n} l_i s_i \frac{-1-m_1}{w\left(1+s_i m_2\right)\left(1+m_1\right) - |z|^2} + \mathrm{O}_\prec\left(\Psi_\theta\right)\right]. \quad (4.4.67)$$

Recall that $\Sigma := DD^* = D^*D$. We introduce a new matrix

$$\tilde{H}(w) := \begin{pmatrix} -w\Sigma^{-1} & w^{1/2}(X - D^{-1}z) \\ w^{1/2}(X - D^{-1}z)^* & -wI \end{pmatrix}, \tag{4.4.68}$$

and define $\tilde{G} := \tilde{H}^{-1}$. By Schur's complement formula, the upper left block of $\tilde{G}$ is

$$\tilde{G}_L = \left[(X - D^{-1}z)(X - D^{-1}z)^* - w\Sigma^{-1}\right]^{-1},$$

and the lower right block is

$$\tilde{G}_R = \left[(X - D^{-1}z)^*\Sigma(X - D^{-1}z) - w\right]^{-1} = \left[(DX - z)^*(DX - z) - w\right]^{-1} = G_R.$$

Now we write $m_{1,2}$ in another way as

$$\begin{aligned} m_1 &= \frac{1}{N}\mathrm{Tr}\left[D^*\left(YY^* - w\right)^{-1}D\right] = \frac{1}{N}\mathrm{Tr}\,\tilde{G}_L, && (4.4.69) \\ m_2 &= \frac{1}{N}\mathrm{Tr}\,\tilde{G}_R = \frac{1}{N}\mathrm{Tr}\left[(X - D^{-1}z)^*\Sigma(X - D^{-1}z) - w\right]^{-1} \\ &= \frac{1}{N}\mathrm{Tr}\left[(X - D^{-1}z)(X - D^{-1}z)^*\Sigma - w\right]^{-1} = \frac{1}{N}\mathrm{Tr}\left(\Sigma^{-1}\tilde{G}_L\right). && (4.4.70) \end{aligned}$$

We apply the arguments in the proof of Lemma 4.4.6 to $\tilde{H}$, and obtain that

$$\tilde{G}_{[ii]}^{-1} = \begin{pmatrix} -w|d_i|^{-2} - wm_2 & -w^{1/2}zd_i^{-1} \\ -w^{1/2}\bar{z}\bar{d}_i^{-1} & -w - wm_1 \end{pmatrix} + O_\prec(|w|\Psi_\theta), \tag{4.4.71}$$

from which we get that

$$1(\Xi)\tilde{G}_{ii} = 1(\Xi)\left[\frac{-1 - m_1}{w(|d_i|^{-2} + m_2)(1 + m_1) - |z|^2|d_i|^{-2}} + \mathrm{O}_\prec(\Psi_\theta)\right].$$

Plugging this into (4.4.70), we get

$$1(\Xi)m_2 = 1(\Xi)\left[\frac{1}{N}\sum_{i=1}^{n}\frac{l_i}{s_i}\frac{-1 - m_1}{w(s_i^{-1} + m_2)(1 + m_1) - |z|^2 s_i^{-1}} + \mathrm{O}_\prec(\Psi_\theta)\right]. \tag{4.4.72}$$

We take the equations in (4.4.67) and (4.4.72) as our new self-consistent equations, namely,

$$1(\Xi)f_1(m_1, m_2) = 1(\Xi)\,\mathrm{O}_\prec(\Psi_\theta), \quad 1(\Xi)f_2(m_1, m_2) = 1(\Xi)\,\mathrm{O}_\prec(\Psi_\theta), \tag{4.4.73}$$

where

$$f_1(m_1, m_2) := m_1 + \frac{1}{N} \sum_i l_i s_i \frac{1 + m_1}{w\left(1 + s_i m_2\right)\left(1 + m_1\right) - |z|^2}, \tag{4.4.74}$$

$$f_2(m_1, m_2) := m_2 + \frac{1}{N} \sum_i l_i \frac{1 + m_1}{w(1 + s_i m_2)(1 + m_1) - |z|^2}. \tag{4.4.75}$$

According to the following lemma, this system of self-consistent equations are stable when $|w|$ and $|z|^2$ are small enough .

Lemma 4.4.11. *Suppose that* $N^{-2}|w|^{-1/2} \leqslant \delta(w) \leqslant (\log N)^{-1}|w|^{-1/2}$ *for* $w \in \mathbf{D}$. *Suppose* $u_{1,2} : \mathbf{D} \to \mathbb{C}$ *are Stieltjes transforms of positive integrable functions such that*

$$\max\left\{\left|f_1(u_1, u_2)(w)\right|, \left|f_2(u_1, u_2)(w)\right|\right\} \leqslant \delta(w).$$

Then there exists an $\varepsilon > 0$ *such that if* $|w|^{1/2} + |z|^2 \leqslant \varepsilon$, *we have*

$$|u_1(w) - m_{1c}(w)| + |u_2(w) - m_{2c}(w)| \leqslant C\delta, \tag{4.4.76}$$

for some constant $C > 0$ *independent of* w, z *and* N.

Proof. The proof depends on the estimate of the Jacobian at (m_{1c}, m_{2c}). By (4.3.22) and (A.1.35), we have

$$m_{1c} = \frac{i\sqrt{t_0} + O(|w|^{1/2} + |z|^2)}{\sqrt{w}}, \quad m_{2c} = \frac{it_0^{-1/2} + O(|w|^{1/2} + |z|^2)}{\sqrt{w}},$$

where $t_0 = (N^{-1}\sum_{i=1}^n l_i/s_i)^{-1}$. Then we can calculate that

$$\det \begin{pmatrix} \partial_1 f_1 & \partial_2 f_1 \\ \partial_1 f_2 & \partial_2 f_2 \end{pmatrix}_{u_{1,2}=m_{1,2c}} = \det \begin{pmatrix} 1 + O(|z|^2) & t_0 + O(|w|^{1/2} + |z|^2) \\ O(|z|^2) & 2 + O(|w|^{1/2} + |z|^2) \end{pmatrix} = 2 + O(|w|^{1/2} + |z|^2).$$

We can conclude the stability by expanding $f_{1,2}(u_1, u_2)$ around (m_{1c}, m_{2c}) and using a fixed point argument as in the proof of Lemma 4.3.7 in Section A.3. □

With this stability lemma, we can repeat all the arguments in the previous subsections to conclude the entrywise local law and averaged local law when $|w|^{1/2} + |z|^2 \leqslant \varepsilon$.

4.5 Anisotropic local law when T is diagonal

In this section we prove the anisotropic local law in Theorem 4.2.22 when T is diagonal. The basic idea of the proof follows from [14, section 5], and the core part of our proof is a novel way to perform the combinatorics. By the Definition 4.2.21 (ii) and Definition 2.4.1 (ii), it suffices to prove the following proposition for generalized entries of G.

Proposition 4.5.1. *Fix $|z|^2 \leqslant 1-\tau$ and suppose that the assumptions of Theorem 4.2.22 hold. Then for any regular domain $\mathbf{S} \subseteq \mathbf{D}$,*

$$|\langle \mathbf{u}, (G(w) - \Pi(w))\, \mathbf{v}\rangle| \prec \Psi \tag{4.5.1}$$

uniformly in $w \in \mathbf{S}$ and any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$.

It is equivalent to prove that

$$\sum_{i,j\in\mathcal{I}_1} u_{[i]}^* \left(G_{[ij]} - \Pi_{[ij]}\right) v_{[j]} \prec \Psi, \quad u_{[i]} := \begin{pmatrix} u_i \\ u_{\bar{i}} \end{pmatrix}, \quad v_{[j]} := \begin{pmatrix} v_j \\ v_{\bar{j}} \end{pmatrix}. \tag{4.5.2}$$

By the entrywise local law,

$$\left|\sum_{i,j} u_{[i]}^* \left(G_{[ij]} - \Pi_{[ij]}\right) v_{[j]}\right| \leqslant \sum_i \left\|G_{[ii]} - \Pi_{[ii]}\right\| \left|u_{[i]}\right| \left|v_{[i]}\right| + \left|\sum_{i\neq j} u_{[i]}^* G_{[ij]} v_{[j]}\right| \prec \Psi + \left|\sum_{i\neq j} u_{[i]}^* G_{[ij]} v_{[j]}\right|.$$

Thus to show (4.5.2), it suffices to prove

$$\left|\sum_{i\neq j} u_{[i]}^* G_{[ij]} v_{[j]}\right| \prec \Psi. \tag{4.5.3}$$

Note that with the entrywise local law, one can only get that

$$\left|\sum_{i\neq j} u_{[i]}^* G_{[ij]} v_{[j]}\right| \prec \Psi \|\mathbf{u}\|_1 \|\mathbf{v}\|_1 \leqslant N\Psi,$$

using $\|\mathbf{u}\|_1 \leqslant N^{1/2}\|\mathbf{u}\|_2$ and $\|\mathbf{v}\|_1 \leqslant N^{1/2}\|\mathbf{v}\|_2$. In particular, this estimate of the ℓ^1 norm is sharp when $\mathbf{u}, \mathbf{v}$ are delocalized, i.e. their entries have size of order $N^{-1/2}$.

The estimate (4.5.3) follows from the Markov's inequality if we can prove the following lemma.

Lemma 4.5.2. *Suppose the assumptions in Proposition 4.5.1 hold. For any* $p \in 2\mathbb{N}$*, we have*

$$\mathbb{E}\left|\sum_{i\neq j} u_{[i]}^* G_{[ij]} v_{[j]}\right|^p \prec \Psi^p.$$

The proof of Lemma 4.5.2 is based on the polynomialization method developed in [14, Section 5], which is also used in Section 3.3 and Section 3.4. But our proof here is much more complicated than the ones before. For simplicity, we only consider the case with $w \in \mathbf{D}$ and $|z|^2 \leqslant 1-\tau$ in this section. If $w \in \mathbf{D}_L$ or $1+\tau \leqslant |z|^2 \leqslant 1+\tau^{-1}$, the proof is almost the same.

4.5.1 Rescaling and partition of indices

For our purpose, it is convenient to define the rescaled matrix

$$R^{(J)} := w^{1/2} G^{(J)}, \tag{4.5.4}$$

for any $J \subset \mathcal{I}$ with $|J| \leqslant l$ for some fixed l. Consequently we define the control parameter Φ

$$\Phi = |w|^{1/2}\, \Psi. \tag{4.5.5}$$

By the entrywise law, for $w \in \mathbf{D}$,

$$R_{[ii]}^{(J)} = O_\prec(1), \quad \left(R_{[ii]}^{(J)}\right)^{-1} = O_\prec(1), \quad R_{[ij]}^{(J)} = \mathrm{O}_\prec(\Phi) \text{ for } i \neq j, \tag{4.5.6}$$

under the above scaling. Now to prove Lemma 4.5.2, it is equivalent to prove

$$\mathbb{E}\left|\sum_{i\neq j} u_{[i]}^* R_{[ij]} v_{[j]}\right|^p \prec \Phi^p. \tag{4.5.7}$$

We expand the product in (4.5.7) as

$$\left|\sum_{i\neq j} u_{[i]}^* R_{[ij]} v_{[j]}\right|^p = \sum_{i_k\neq j_k\in\mathcal{I}_1} \prod_{k=1}^{p/2} u_{[i_k]}^* R_{[i_k j_k]} v_{[j_k]} \cdot \prod_{k=p/2+1}^{p} \overline{u_{[i_k]}^* R_{[i_k j_k]} v_{[j_k]}}.$$

Formally, we regard $\{i_1,...,i_p,j_1,...,j_p\}$ as the set of $2p$ (index) variables that take values in $\mathcal{I}_1$. Let $\mathcal{B}_p$ be the collection of all partitions of $\{i_1,...,i_p,j_1,...,j_p\}$ such that i_k, j_k are not in the same block for all $k = 1, ..., p$. For $\Gamma \in \mathcal{B}_p$, let $n(\Gamma)$ be the number of its blocks and define a set of $\mathcal{I}_1$-valued variables as

$$L(\Gamma) := \{b_1, ..., b_{n(\Gamma)}\}. \tag{4.5.8}$$

Now it is convenient to regard Γ as a symbol-to-symbol function,

$$\Gamma : \{i_1, ..., i_p, j_1, ..., j_p\} \to L(\Gamma), \tag{4.5.9}$$

such that each $\Gamma^{-1}(b_k)$ is a block of the partition. Then we can rewrite the sum as

$$\left|\sum_{i \neq j} u^*_{[i]} R_{[ij]} v_{[j]}\right|^p = \sum_{\Gamma \in \mathcal{B}_p} \sum^*_{\substack{b_l \in \mathcal{I}_1, \\ l=1,...,n(\Gamma)}} \prod_{k=1}^{p/2} u^*_{[\Gamma(i_k)]} R_{[\Gamma(i_k)\Gamma(j_k)]} v_{[\Gamma(j_k)]} \cdot \prod_{k=p/2+1}^{p} \overline{u^*_{[\Gamma(i_k)]} R_{[\Gamma(i_k)\Gamma(j_k)]} v_{[\Gamma(j_k)]}}, \tag{4.5.10}$$

where $\sum^*$ denotes the summation subject to the condition that the values of $b_1, ...b_n$ are ordered as $b_1 < b_2 < \ldots < b_n$. We pick one term from the above summation and denote

$$\Delta(\Gamma) := \prod_{k=1}^{p/2} u^*_{[\Gamma(i_k)]} R_{[\Gamma(i_k)\Gamma(j_k)]} v_{[\Gamma(j_k)]} \cdot \prod_{k=p/2+1}^{p} \overline{u^*_{[\Gamma(i_k)]} R_{[\Gamma(i_k)\Gamma(j_k)]} v_{[\Gamma(j_k)]}}. \tag{4.5.11}$$

For any $b_k \in L$, we can also define a corresponding $\mathcal{I}_2$-valued variable $\bar{b}_k$ in the obvious way, and we denote

$$[L] := \{b_1, ..., b_n, \overline{b_1}, ..., \overline{b_n}\}. \tag{4.5.12}$$

For notational convenience, we will also use letters i, j, k, l to denote the symbols in L.

4.5.2 String and string operators

During the proof we will frequently use the following resolvent identities for rescaled matrix R. They follow immediately from Lemma 4.3.2.

Lemma 4.5.3 (Resolvent identities for $R_{[ij]}$ groups)**.** *For $k \notin J$ and $i, j \in \mathcal{I}_1 \backslash J \cup \{k\}$, we have*

$$R^{[J]}_{[ij]} = R^{[Jk]}_{[ij]} + R^{[J]}_{[ik]} \left(R^{[J]}_{[kk]}\right)^{-1} R^{[J]}_{[kj]}, \tag{4.5.13}$$

$$\left(R^{[J]}_{[ii]}\right)^{-1} = \left(R^{[Jk]}_{[ii]}\right)^{-1} - \left(R^{[J]}_{[ii]}\right)^{-1} R^{[J]}_{[ik]} \left(R^{[J]}_{[kk]}\right)^{-1} R^{[J]}_{[ki]} \left(R^{[Jk]}_{[ii]}\right)^{-1}, \tag{4.5.14}$$

$$\left(R^{[J]}_{[ii]}\right)^{-1} = w^{-1/2} H^{[J]}_{[ii]} - w^{-1} \sum_{l,l' \notin J \cup \{i\}} H^{[J]}_{[il]} R^{[Ji]}_{[ll']} H^{[J]}_{[l'i]}. \tag{4.5.15}$$

Furthermore, for $i \neq j$ and L defined in (4.5.8), we have

$$R_{[ij]}^{[L\setminus\{ij\}]} = R_{[ii]}^{[L\setminus\{ij\}]} S_{[ij]} R_{[jj]}^{[L\setminus\{j\}]}, \quad \textit{with} \;\; S_{[ij]} = -w^{-1/2} H_{[ij]} + w^{-1} \sum_{k,l \notin L} H_{[ik]} R_{[kl]}^{[L]} H_{[lj]}. \tag{4.5.16}$$

In this section, we expand the R variables in $\Delta(\Gamma)$ using the identities in Lemma 4.5.3. During the expansion, we need to distinguish carefully between an algebraic expression and its value as a random variable. Our notations below are extensions of the ones defined in Section 3.3.2.

Definition 4.5.4 (Strings)**.** *Let $\mathfrak{A}$ be an alphabet containing all symbols that may appear during the expansion, such as $R_{[ij]}^{[J]}$, $\left(R_{[ij]}^{[J]}\right)^{-1}$, $S_{[ij]}$, $u_{[i]}^*$ and $v_{[j]}$ for $J \subset L(\Gamma)$. We define a string $\mathbf{s}$ to be a formal expression consisting of the symbols from $\mathfrak{A}$, and denote by $[\![\mathbf{s}]\!]$ the random variable represented by it. Let $\mathfrak{M}$ be the collection of all possible strings. We denote an empty string by $\varnothing$.*

Given a string $\mathbf{s}$, after an expansion of R's in it, we will get a different string $\mathbf{s}'$. However, they represent the same random variable $[\![\mathbf{s}]\!] = [\![\mathbf{s}']\!]$. During the proof, we will identify more elements of $\mathfrak{A}$ (see the symbols in (4.5.32)).

To perform the expansions in a systematical way, we define the following operators acting on strings. We call the symbols $R_{[ij]}^{[J]}$, $\left(R_{[ij]}^{[J]}\right)^{-1}$ to be *m*aximally expanded if $J \cup \{i,j\} = L$. We call a string $\mathbf{s}$ to be *m*aximally expanded if all the R symbols in $\mathbf{s}$ is maximally expanded.

Definition 4.5.5 (String operators)**.** *(i) Define an operator $\tau_0^{(k)}$ for $\Omega \in \mathfrak{M}$, in the following sense. Find the first $R_{[ij]}^{[J]}$ in Ω such that $k \notin J \cup \{i,j\}$, or the first $\left(R_{[ii]}^{[J]}\right)^{-1}$ such that $k \notin J \cup \{i\}$. If $R_{[ij]}^{[J]}$ is found, replace it with $R_{[ij]}^{[Jk]}$; if $\left(R_{[ii]}^{[J]}\right)^{-1}$ is found, replace it with $\left(R_{[ii]}^{[Jk]}\right)^{-1}$; if neither is found, $\tau_0^{(k)}(\Omega) = \Omega$ and we say that $\tau_0^{(k)}$ is trivial for Ω.*

(ii) Define an operator $\tau_1^{(k)}$ for $\Omega \in \mathfrak{M}$, in the following sense. Find the first $R_{[ij]}^{[J]}$ in Ω such that $k \notin J \cup \{i,j\}$, or the first $\left(R_{[ii]}^{[J]}\right)^{-1}$ such that $k \notin T \cup \{i\}$. If $R_{[ij]}^{[J]}$ is found, replace it with $R_{[ik]}^{[J]} \left(R_{[kk]}^{[J]}\right)^{-1} R_{[kj]}^{[J]}$; if $\left(R_{[ii]}^{[J]}\right)^{-1}$ is found, replace it with

$$-\left(R_{[ii]}^{[J]}\right)^{-1} R_{[ik]}^{[J]} \left(R_{[kk]}^{[J]}\right)^{-1} R_{[ki]}^{[J]} \left(R_{[ii]}^{[Jk]}\right)^{-1};$$

if neither is found, $\tau_1^{(k)}(\Omega) = \varnothing$ *and we say that* $\tau_1^{(k)}$ *is null for* Ω*.*

(iii) Define an operator ρ *for* $\Omega \in \mathfrak{M}$*, in the following sense. Find each maximally expanded off-diagonal* $R_{[ij]}^{[L\backslash\{ij\}]}$ *in* Ω *and replace it with* $R_{[ii]}^{[L\backslash\{ij\}]} S_{[ij]} R_{[jj]}^{[L\backslash\{j\}]}$*. If nothing is found,* $\rho(\Omega) = \Omega$*.*

According to Lemma 4.5.3, for any $\Omega \in \mathfrak{M}$ we have

$$\left[\!\!\left[\left(\tau_0^{(k)} + \tau_1^{(k)} \right)(\Omega) \right]\!\!\right] = [\![\Omega]\!], \quad [\![\rho(\Omega)]\!] = [\![\Omega]\!]. \tag{4.5.17}$$

Definition 4.5.6. *Define the function* $\mathcal{F}_{\text{d-max}} : \mathfrak{M} \to \mathbb{N}$ *(where the subscript "d-max" stands for "distance to being maximally expanded") through*

$$\mathcal{F}_{\textit{d-max}}\left(R_{[ij]}^{[J]\,\star} \right) = |L \backslash (J \cup \{i,j\})|,$$

where $\star$ *could be* 1 *or* -1*, and*

$$\mathcal{F}_{\textit{d-max}}(\Omega) = \sum_{R \text{ variables in } \Omega} \mathcal{F}_{\textit{d-max}}(R).$$

Define another function $\mathcal{F}_{\textit{off}} : \mathfrak{M} \to \mathbb{N}$ *with* $\mathcal{F}_{\textit{off}}(\Omega)$ *being the number of off-diagonal symbols in* Ω*.*

By off-diagonal symbols, we mean the terms of the form A_{st} with $s \notin \{t, \bar{t}\}$ or $A_{[ij]}$ with $i \neq j$, e.g. $R_{[ij]}^{[J]}$ and $S_{[ij]}$ with $i \neq j$. Later we will define other types of off-diagonal symbols (see (4.5.32)). Note that a R symbol is maximally expanded if and only if $\mathcal{F}_{\text{d-max}}(R) = 0$ and a string Ω is maximally expanded if and only if $\mathcal{F}_{\text{d-max}}(\Omega) = 0$. The next two lemmas are almost trivial by Definition 4.5.5.

Lemma 4.5.7. *Fix* $k \in L$*. If* $\tau_0^{(k)}(\Omega) = \Omega$ *and* $\tau_1^{(k)}(\Omega) = \varnothing$*,*

$$\mathcal{F}_{\textit{d-max}}\left(\tau_0^{(k)}(\Omega) \right) = \mathcal{F}_{\textit{d-max}}(\Omega), \quad \mathcal{F}_{\textit{d-max}}\left(\tau_1^{(k)}(\Omega) \right) = 0; \tag{4.5.18}$$

otherwise,

$$\mathcal{F}_{\textit{d-max}}\left(\tau_0^{(k)}(\Omega) \right) = \mathcal{F}_{\textit{d-max}}(\Omega) - 1, \quad \mathcal{F}_{\textit{d-max}}\left(\tau_1^{(k)}(\Omega) \right) \leqslant \mathcal{F}_{\textit{d-max}}(\Omega) + 4n(\Gamma). \tag{4.5.19}$$

For ρ*, we have*

$$\mathcal{F}_{\textit{d-max}}\left(\rho(\Omega) \right) = \mathcal{F}_{\textit{d-max}}(\Omega) + a, \tag{4.5.20}$$

where a *is the number of maximally expanded off-diagonal* R*'s in* Ω*.*

Lemma 4.5.8. *Fix $k \in L$. For any $\Omega \in \mathfrak{M}$, we have*

$$\mathcal{F}_{\mathit{off}}\left(\tau_0^{(k)}(\Omega)\right) = \mathcal{F}_{\mathit{off}}(\Omega), \quad \mathcal{F}_{\mathit{off}}\left(\rho(\Omega)\right) = \mathcal{F}_{\mathit{off}}(\Omega), \tag{4.5.21}$$

and

$$\mathcal{F}_{\mathit{off}}(\Omega) + 1 \leqslant \mathcal{F}_{\mathit{off}}\left(\tau_1^{(k)}(\Omega)\right) \leqslant \mathcal{F}_{\mathit{off}}(\Omega) + 2 \quad \text{if} \quad \tau_1^{(k)}(\Omega) \neq \varnothing. \tag{4.5.22}$$

4.5.3 Expansion of the strings

For simplicity of notations, throughout the rest of this section we omit the complex conjugates on the right hand side of (4.5.11) (if we keep the complex conjugates, the proof is the same but with slightly heavier notations). Suppose the right hand side of (4.5.11) is represented by a string Ω_Δ. Given a binary word $\mathbf{w} = a_1 a_2 ... a_m$ with $a_i \in \{0, 1\}$, we define the operation

$$(\Omega_\Delta)_{\mathbf{w}} := \rho\tau_{a_m}^{(b_m)} \cdots \rho\tau_{a_2}^{(b_2)} \rho\tau_{a_1}^{(b_1)} \left(\Omega_\Delta\right) \tag{4.5.23}$$

where $b_{qn+r} := b_r$ (recall (4.5.8)) for any $1 \leqslant r \leqslant n$ and $q \in \mathbb{N}$. So a binary word $\mathbf{w}$ uniquely determines an operator composition. By (4.5.17), $[\![(\Omega_\Delta)_{\mathbf{w}0}]\!] + [\![(\Omega_\Delta)_{\mathbf{w}1}]\!] = [\![(\Omega_\Delta)_{\mathbf{w}}]\!]$ and so we get

$$\sum_{|\mathbf{w}|=m} [\![(\Omega_\Delta)_{\mathbf{w}}]\!] = [\![\Omega_\Delta]\!]$$

for any $m \geqslant 1$, where $|\mathbf{w}|$ denotes the length of $\mathbf{w}$.

Lemma 4.5.9. *Given any $\mathbf{w}$ such that $|\mathbf{w}| = (n^2 + 1)(p + 6l_0)$ and $(\Omega_\Delta)_{\mathbf{w}} \neq \varnothing$, then either $\mathcal{F}_{\mathit{off}}((\Omega_\Delta)_{\mathbf{w}}) \geqslant l_0 := (8/\zeta + 2)\, p$, or $(\Omega_\Delta)_{\mathbf{w}}$ is maximally expanded.*

Proof. We use m_0 to denote the number of 0's in $\mathbf{w}$, and m_1 to denote the number of 1's. Furthermore, we use $m_0^{(0)}$ to denote the number of 0's corresponding to the trivial τ_0's, and $m_0^{(1)}$ to denote the number of 0's corresponding to the non-trivial τ_0's. Assume $\mathcal{F}_{\text{off}}((\Omega_\Delta)_{\mathbf{w}}) < l_0$ and $(\Omega_\Delta)_{\mathbf{w}}$ is not maximally expanded. By (4.5.21) and (4.5.22), we have $m_1 \leqslant l_0 - p \leqslant l_0$. By (4.5.18)-(4.5.20), we have

$$\mathcal{F}_{\text{d-max}}((\Omega_\Delta)_{\mathbf{w}}) \leqslant \mathcal{F}_{\text{d-max}}(\Omega_\Delta) + l_0 + 4nm_1 - m_0^{(1)}.$$

Then with $\mathcal{F}_{\text{d-max}}(\Omega_\Delta) = np$, we get a rough bound $m_0^{(1)} + m_1 < n(p + 6l_0)$. By pigeonhole principle, there are at least n 0's in a row in $\mathbf{w}$ that correspond to trivial τ_0's. This indicates that $(\Omega_\Delta)_\mathbf{w}$ is maximally expanded, which gives a contradiction. □

Lemma 4.5.10. *There exists constants $C_{p,l_0}, C_{p,\zeta} > 0$ such that*

$$\sum_{\Gamma \in \mathcal{B}_p} \sum^{*}_{\substack{b_l \in \mathcal{I}_1, \\ l=1,\dots,n(\Gamma)}} \left| \mathbb{E} \sum_{\substack{|\mathbf{w}|=(n^2+1)(p+6l_0), \\ \mathcal{F}_{\mathit{off}}((\Omega_{\Delta(\Gamma)})_\mathbf{w}) \geqslant l_0}} \left[\!\left[(\Omega_{\Delta(\Gamma)})_\mathbf{w} \right]\!\right] \right| \prec C_{p,l_0} N^{2p} \Phi^{l_0} \leqslant C_{p,\zeta} \Phi^p. \tag{4.5.24}$$

Proof. The first bound is due to the fact that each summand is of the order $\mathrm{O}_\prec(\Phi^{l_0})$ and there are at most N^{2p} of them. For the second bound, we used $\Phi \leqslant C N^{-\zeta/2}$. □

This lemma shows that all the strings with sufficiently many off-diagonal symbols contribute at most Φ^p. It remains to handle the maximally expanded strings. Define a diagonal symbol as

$$S_{[ii]} := -\begin{pmatrix} 0 & d_i X_{i\bar{i}} \\ \bar{d}_i X^\dagger_{\bar{i}i} & 0 \end{pmatrix} + w^{-1} \sum_{k,l \notin L} H_{[ik]} R^{[L]}_{[kl]} H_{[li]}, \tag{4.5.25}$$

such that

$$\left(R^{[L\setminus\{i\}]}_{[ii]} \right)^{-1} = \begin{pmatrix} -w^{1/2} & -z \\ -\bar{z} & -w^{1/2} \end{pmatrix} - S_{[ii]}. \tag{4.5.26}$$

Notice all the R symbols in a maximally expanded string are diagonal. We taylor expand $R^{[L\setminus\{i\}]}_{[ii]}$ as

$$R^{[L\setminus\{i\}]}_{[ii]} = \left[w^{-1/2} \pi^{-1}_{[i]c} + \left(S_{[ii]} - B_i \right) \right]^{-1} = \sum_{k=0}^{l_0 - 1} \tilde{\pi}_{ic} \left[\left(S_{[ii]} - B_i \right) \tilde{\pi}_{ic} \right]^k + O_\prec \left(\Phi^{l_0} \right), \tag{4.5.27}$$

where $\tilde{\pi}_{[i]c} := w^{1/2} \pi_{[i]c}$, $B_i := \begin{pmatrix} w^{1/2} |d_i|^2 m_{2c} & 0 \\ 0 & w^{1/2} m_{1c} \end{pmatrix}$, and for the error term,

$$S_{[ii]} - B_i = w^{-1/2} Z^{[L\setminus\{i\}]}_{[i]} + w^{1/2} \begin{pmatrix} |d_i|^2 (m_{2c} - m_2^{[L]}) & 0 \\ 0 & m_{1c} - m_1^{[L]} \end{pmatrix} \prec \Phi$$

by (4.4.12) and the averaged local law. Now for all maximally expanded $(\Omega_\Delta)_\mathbf{w}$ with $|\mathbf{w}| = (n^2 + 1)(p + 6l_0)$, denote by $\sigma \left[\!\left[(\Omega_\Delta)_\mathbf{w} \right]\!\right]$ the expression after plugging in (4.5.26) and (4.5.27) without the tail terms. Similar to Lemma 4.5.10, we have

$$\sum_{\Gamma\in\mathcal{B}_p}\sum^{*}_{\substack{b_l\in\mathcal{I}_1,\\ l=1,\dots,n(\Gamma)}}\left|\mathbb{E}\sum_{\substack{|\mathbf{w}|=(n^2+1)(p+6l_0),\\ (\Omega_\Delta)_\mathbf{w}\text{ maximally expanded}}}\left(\left[\!\left[(\Omega_{\Delta(\Gamma)})_\mathbf{w}\right]\!\right]-\sigma\left[\!\left[(\Omega_{\Delta(\Gamma)})_\mathbf{w}\right]\!\right]\right)\right|\prec C_{p,\zeta}\Phi^p.$$

From the above bound and Lemmas 4.5.9-4.5.10, we see that to prove (4.5.7), it suffices to show

$$\sum_{\Gamma\in\mathcal{B}_p}\sum^{*}_{\substack{b_l\in\mathcal{I}_1,\\ l=1,\dots,n(\Gamma)}}\left|\mathbb{E}\sum_{\substack{|\mathbf{w}|=(n^2+1)(p+6l_0),\\ (\Omega_\Delta)_\mathbf{w}\text{ maximally expanded}}}\sigma\left[\!\left[(\Omega_{\Delta(\Gamma)})_\mathbf{w}\right]\!\right]\right|\prec C_{p,\zeta}\Phi^p. \tag{4.5.28}$$

We write $\sigma\left[\!\left[(\Omega_\Delta)_\mathbf{w}\right]\!\right]$ as a sum of monomials in terms of $S_{[ij]}$:

$$\sigma\left[\!\left[(\Omega_\Delta)_\mathbf{w}\right]\!\right]=\sum_i M(\mathbf{w},\Delta(\Gamma),i), \tag{4.5.29}$$

where i is an index to label these monomials. Note that after plugging (4.5.29) into (4.5.28), the number of summands $M(\mathbf{w},\Delta(\Gamma),i)$ inside the expectation depends only on p and ζ. Thus to show (4.5.28), it suffices to prove the following lemma.

Lemma 4.5.11. *Fix any $\Gamma\in\mathcal{B}_p$ and binary word $\mathbf{w}$ with $|\mathbf{w}|=(n^2+1)(p+6l_0)$. Suppose $(\Omega_\Delta)_\mathbf{w}$ is maximally expanded. Let $M(\mathbf{w},\Delta(\Gamma))$ be a monomial in $\sigma\left[\!\left[(\Omega_{\Delta(\Gamma)})_\mathbf{w}\right]\!\right]$. Then we have*

$$\sum^{*}_{b_l\in\mathcal{I}_1,l=1,\dots,n(\Gamma)}\left|\mathbb{E}M(\mathbf{w},\Delta(\Gamma))\right|\prec C_{p,\zeta}\Phi^p \tag{4.5.30}$$

for some constant $C_{p,\zeta}$ that only depends on p and ζ.

For the rest of this section, we fix a $\Gamma\in\mathcal{B}_p$ and a maximally expanded $(\Omega_{\Delta(\Gamma)})_\mathbf{w}$ with $|\mathbf{w}|=(n^2+1)(p+6l_0)$. Then we fix a monomial $M(\mathbf{w},\Delta(\Gamma))$ in $\sigma\left[\!\left[(\Omega_{\Delta(\Gamma)})_\mathbf{w}\right]\!\right]$. Let Ω_M be the string form of $M(\mathbf{w},\Delta(\Gamma))$ in terms of $S_{[ij]}$. It is not hard to see that

$$\mathcal{F}_{\text{off}}\left(\Omega_M\right)=\mathcal{F}_{\text{off}}\left((\Omega_\Delta)_\mathbf{w}\right). \tag{4.5.31}$$

Now we decompose $S_{[ij]}$ as

$$S_{[ij]}=S^X_{\bar{i}\bar{j}}+S^X_{\bar{i}j}+S^R_{\bar{i}\bar{j}}+S^R_{ij}+S^R_{\bar{i}\bar{j}}+S^R_{\bar{i}j}, \tag{4.5.32}$$

where we define the following symbols in $\mathfrak{A}$:

$$S^X_{i\bar{j}} := d_i X_{i\bar{j}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad S^X_{\bar{i}j} := \bar{d}_i X^\dagger_{\bar{i}j} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{4.5.33}$$

$$S^R_{i\bar{j}} := \sum_{k,l\notin L} d_i d_l X_{i\bar{k}} X_{l\bar{j}} \begin{pmatrix} 0 & R^{[L]}_{\bar{k}l} \\ 0 & 0 \end{pmatrix}, \quad S^R_{ij} := \sum_{k,l\notin L} d_i \bar{d}_l X_{i\bar{k}} X^\dagger_{\bar{l}j} \begin{pmatrix} R^{[L]}_{\bar{k}\bar{l}} & 0 \\ 0 & 0 \end{pmatrix}, \tag{4.5.34}$$

$$S^R_{\bar{i}\bar{j}} := \sum_{k,l\notin L} \bar{d}_i d_l X^\dagger_{\bar{i}k} X_{l\bar{j}} \begin{pmatrix} 0 & 0 \\ 0 & R^{[L]}_{kl} \end{pmatrix}, \quad S^R_{\bar{i}j} := \sum_{k,l\notin L} \bar{d}_i \bar{d}_l X^\dagger_{\bar{i}k} X^\dagger_{\bar{l}j} \begin{pmatrix} 0 & 0 \\ R^{[L]}_{k\bar{l}} & 0 \end{pmatrix}. \tag{4.5.35}$$

We expand the $S_{[ij]}$'s in $M(\mathbf{w}, \Delta(\Gamma))$ using (4.5.32), and write $M(\mathbf{w}, \Delta(\Gamma))$ as a sum of monomials in terms of S^X_{st} and S^R_{st}:

$$M(\mathbf{w}, \Delta(\Gamma)) = \sum_i Q(\mathbf{w}, \Delta(\Gamma), i), \tag{4.5.36}$$

where i is an index to label these monomials. Again it is not hard to see that

$$\mathcal{F}_{\text{off}}(\Omega_Q) = \mathcal{F}_{\text{off}}(\Omega_M) = \mathcal{F}_{\text{off}}((\Omega_\Delta)_{\mathbf{w}}). \tag{4.5.37}$$

Since the number of summands in (4.5.36) is independent of N, to prove (4.5.30) it suffices to show

$$\sum^*_{b_l\in\mathcal{I}_1, l=1,\dots,n(\Gamma)} |\mathbb{E}Q(\mathbf{w}, \Delta(\Gamma))| \prec C_{p,\zeta}\Phi^p \tag{4.5.38}$$

for any monomial $Q(\mathbf{w}, \Delta(\Gamma))$ in (4.5.36). Throughout the following, we fix a $Q(\mathbf{w}, \Delta(\Gamma))$ with nonzero expectation, and denote by Ω_Q the string form of $Q(\mathbf{w}, \Delta(\Gamma))$ in terms of S^X_{st} and S^R_{st}. Notice the R variables in S^R_{st} are maximally expanded. As a result, the S^X_{st} variables are independent of S^R_{st} variables in $Q(\mathbf{w}, \Delta(\Gamma))$. Therefore we make the following observation: if S^X_{st} appears as a symbol in Ω_Q, then Ω_Q contains at least two of them.

Definition 4.5.12. *Recall Γ defined in (4.5.9). Let h be the number of blocks of Γ whose size is* 1*, i.e.*

$$h := \sum_{l=1}^{n(\Gamma)} \mathbf{1}\left(\left|\Gamma^{-1}(b_l)\right| = 1\right). \tag{4.5.39}$$

For $l = 1, ..., n$, define

$$I_l := \left|\{i_1, \dots, i_p\} \cap \Gamma^{-1}(b_l)\right|, \quad J_l := \left|\{j_1, \dots, j_p\} \cap \Gamma^{-1}(b_l)\right|.$$

Lemma 4.5.13. *Suppose for any $b_1, ..., b_n$ taking distinct values in $\mathcal{I}_1$,*

$$|\,\mathbb{E}\, Q(\mathbf{w}, \Delta(\Gamma))| \prec CN^{-h/2}\Phi^p \prod_{l=1}^{n} \left|u_{[b_l]}\right|^{I_l} \left|v_{[b_l]}\right|^{J_l} \tag{4.5.40}$$

holds for some constant C independent of N. Then the estimate (4.5.38) holds.

Proof. By Cauchy-Schwarz inequality,

$$\sum_{k=1}^{N} \left|u_{[k]}\right|^{a} \left|v_{[k]}\right|^{b} \leqslant \begin{cases} N^{1/2} & \text{if } a+b=1 \\ 1 & \text{if } a+b \geqslant 2 \end{cases}.$$

Then using $h = \sum_{l=1}^{n} \mathbf{1}\left(I_l + J_l = 1\right)$, we get

$$\sum_{b_l \in \mathcal{I}_1, l=1,...,n(\Gamma)}^{*} |\mathbb{E}Q(\mathbf{w}, \Delta(\Gamma))| \prec C\Phi^p N^{-h/2} \prod_{l=1}^{n} \sum_{b_l \in \mathcal{I}_1} \left|u_{[b_l]}\right|^{I_l} \left|v_{[b_l]}\right|^{J_l} \leqslant C\Phi^p.$$

□

Hence it suffices to prove (4.5.40). The key is to extract the $N^{-h/2}$ factor from $\mathbb{E}\, Q(\mathbf{w}, \Delta(\Gamma))$. For this purpose, we need to keep track of the indices in L during the expansion.

Definition 4.5.14. *Define a function $\mathcal{F}_{in} : L \times \mathfrak{M} \to \mathbb{N}$ with $\mathcal{F}_{in}(l, \Omega)$ giving the number of times l or $\bar{l}$ appears as an index of an off-diagonal R or S symbol in Ω.*

The following lemma follows immediately from Definition 4.5.5 and the expansions we have done to obtain Ω_Q from $(\Omega_\Delta)_\mathbf{w}$.

Lemma 4.5.15. *(1) For any string Ω, if $\tau_0^{(k)}$ is not trivial for Ω, then*

$$\mathcal{F}_{in}\left(l, \tau_0^{(k)}(\Omega)\right) = \mathcal{F}_{in}(l, \Omega), \quad \mathcal{F}_{in}\left(l, \tau_1^{(k)}(\Omega)\right) = \mathcal{F}_{in}(l, \Omega) + a, \;\; a \in \{0, 2\}. \tag{4.5.41}$$

(2) For any string Ω,

$$\mathcal{F}_{in}\left(l, \rho(\Omega)\right) = \mathcal{F}_{in}(l, \Omega). \tag{4.5.42}$$

(3) For any maximally expanded $(\Omega_\Delta)_\mathbf{w}$,

$$\mathcal{F}_{in}(l, \Omega_Q) = \mathcal{F}_{in}(l, (\Omega_\Delta)_\mathbf{w}). \tag{4.5.43}$$

Let Ω_Q^X be the substring of Ω_Q containing only S^X symbols, and Ω_Q^R be the substring of Ω_Q containing only S^R symbols. Define

$$\mathcal{V} := \{l \in L | \mathcal{F}_{\text{in}}(l, \Omega_\Delta) = 1\}, \tag{4.5.44}$$

and

$$\mathcal{V}_0 := \{l \in L | \mathcal{F}_{\text{in}}(l, \Omega_\Delta) = 1 \text{ and } \mathcal{F}_{\text{in}}(l, \Omega_Q^X) = 0\}, \tag{4.5.45}$$

$$\mathcal{V}_1 := \{l \in L | \mathcal{F}_{\text{in}}(l, \Omega_\Delta) = 1 \text{ and } \mathcal{F}_{\text{in}}(l, \Omega_Q^X) \geqslant 2\}. \tag{4.5.46}$$

Recall the observation above Definition 4.5.12, we have $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_1$ and

$$h = |\mathcal{V}| = |\mathcal{V}_0| + |\mathcal{V}_1|.$$

Let n_X be the number of off-diagonal S^X symbols in Ω_Q^X and n_R be the number of off-diagonal S^R symbols in Ω_Q^R. Note that $n_o := n_X + n_R$ is the total number of off-diagonal symbols in Ω_Q.

4.5.4 A graphical proof

We introduce graphs to conclude the proof of (4.5.40). We use a connected graph to represent the string Ω_Q, call it by $\mathfrak{G}_{Q0}$. The indices in $[L]$ are represented by black nodes in $\mathfrak{G}_{Q0}$. The S_{st}^X or S_{st}^R symbols in Ω_Q are represented by edges connecting the nodes s and t. We also define colors for the nodes and edges, where the color set for nodes is $\{black, white\}$ and the color set for edges is $\{S^X, S^R, X, R\}$. In $\mathfrak{G}_{Q0}$, all the nodes are black, all S^X edges are assigned S^X color and all S^R edges are assigned S^R color. We show a possible graph in Fig. 4.3. In this subsection, we identify an index with its node representation, and a symbol with its edge representation.

Definition 4.5.16. *Define function* deg *on the nodes set* $[L]$ *such that* $\deg(l)$ *gives the number of* S^R *edges connecting to the node* l.

By Lemma 4.5.15, we see that for any $l \in \mathcal{V}_0$,

$$\mathcal{F}_{\text{in}}(l, \Omega_Q) \equiv \deg(l) + \deg(\bar{l}) \equiv 1 \pmod 2. \tag{4.5.47}$$

Hence

$$|\mathcal{V}_0| = \sum_{l\in\mathcal{V}_0} \left[\mathcal{F}_{\rm in}\left(l,\Omega_Q\right) \mod 2\right] \leqslant \sum_{l\in\mathcal{V}_0} \Big[(\deg(l) \mod 2) + \big(\deg(\bar{l}) \mod 2\big)\Big]. \qquad (4.5.48)$$

Now we expand the S^R edges. Take the $S^R_{i\bar{j}}$ edge as an example (recall (4.5.34)). We replace the $S^R_{i\bar{j}}$ edge with an R-group, defined as following. We add two white colored nodes to represent the summation indices $\bar{k}, l \notin [L]$, two X-colored edges to represent $X_{i\bar{k}}$ and $X_{l\bar{j}}$, and an R-colored edge connecting $\bar{k}$ and l to represent $\begin{pmatrix} 0 & R^{[L]}_{\bar{k}l} \\ 0 & 0 \end{pmatrix}$. We call the subgraph consisting of the three new edges and their nodes an R-group. If $i = j$, we call it a diagonal R-group; otherwise, call it an off-diagonal R-group. We expand all the S^R edges in $\mathfrak{G}_{Q0}$ into R-groups and call the resulting graph $\mathfrak{G}_{Q1}$. For example, after expanding the S^R edges in Fig. 4.3, we get the graph in Fig. 4.4. In the graph $\mathfrak{G}_{Q1}$, the R edges, X edges and S^X edges are mutually independent, since the R symbols are maximally expanded, and the white nodes are different from the black nodes.

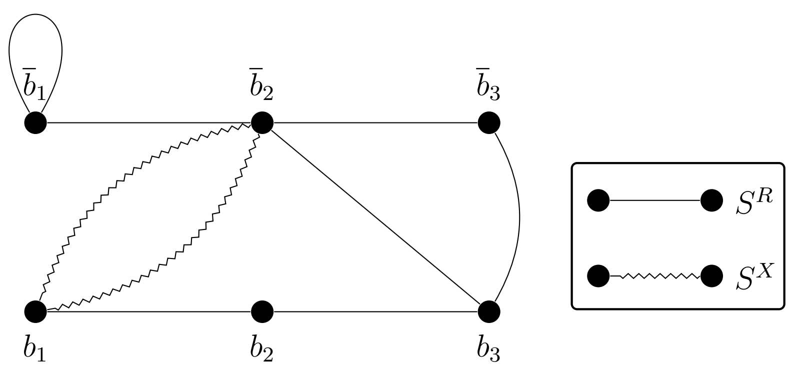


Figure 4.3: An example of the graph $\mathfrak{G}_{Q0}$.

Notice that each white node represents a summation index. As we have done for the black nodes, we first partition the white nodes into blocks and then assign values to the blocks when doing the summation. Let W be the set of all white nodes in $\mathfrak{G}_{Q1}$, and let $\mathcal{W}$ be the collection of all partitions of W. Fix a partition $\gamma \in \mathcal{W}$ and denote its blocks by $W_1, ..., W_{m(\gamma)}$. If two white nodes of some off-diagonal R-group happen to lie in the same block, then we merge the two nodes into one diamond white node (Fig. 4.5). All the other

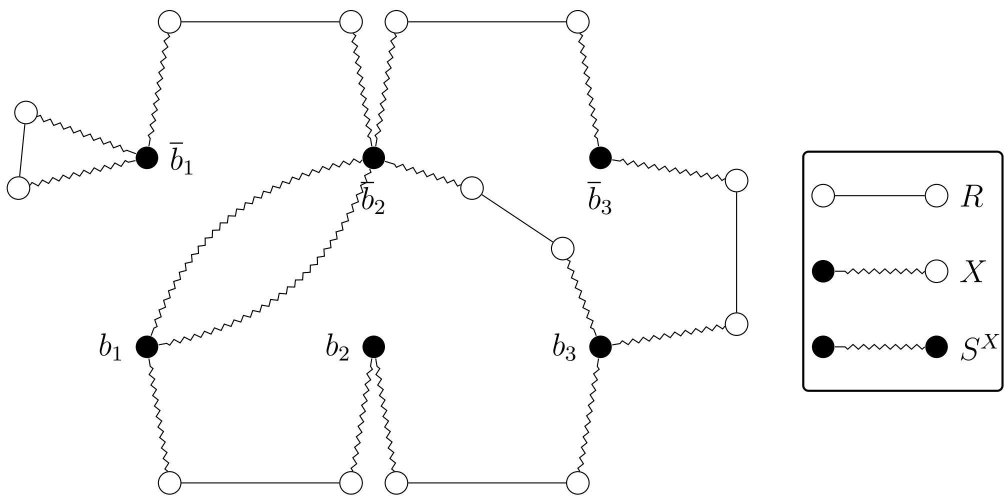


Figure 4.4: The resulting graph $\mathfrak{G}_{Q1}$ after expanding each S^R in Fig. 4.3 into R-groups.

white nodes are called normal (Fig. 4.6). Let $n_R^{(d)}$ be the number of diamond nodes (which is $\leqslant$ the number of diagonal R-edges in $\mathfrak{G}_{Q1}$). Then we trivially have (recall Definition 4.5.16)

$$\#\text{ of white nodes} = -n_R^{(d)} + \sum_{k=1}^{n} \left[\deg(b_k) + \deg(\bar{b}_k)\right]. \tag{4.5.49}$$

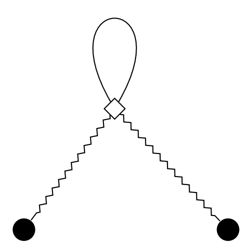


Figure 4.5: Diamond white node.

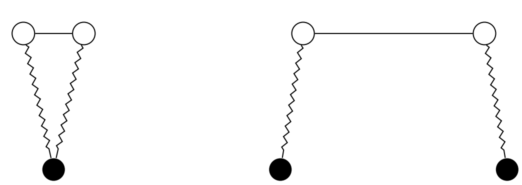


Figure 4.6: Normal white nodes.

By (4.5.48), there are at least $|\mathcal{V}_0|$ black nodes with odd deg in $[\mathcal{V}_0]$ (where $[\mathcal{V}_0]$ is defined in the obvious way). WLOG, we may assume these nodes are $b_1, ..., b_{|\mathcal{V}_0|}$. To have nonzero expectation, each white block must contain at least two white nodes. Therefore for each $k = 1, ..., |\mathcal{V}_0|$, there exists a block connecting to b_k which contains at least 3 white nodes. Call such a block $W(b_k)$, and denote by $A(b_k)$ the set of the adjacent white nodes to b_k in $W(b_k)$. Be careful that the $W(b_k)$'s or $A(b_k)$'s are not necessarily distinct. WLOG, let $W_1, ..., W_d$ be the distinct blocks among all $W(b_k)$'s. Define

$$\mathcal{V}_{00} := \{b_k |\ A(b_k) \text{ has no normal white nodes}, 1 \leqslant k \leqslant |\mathcal{V}_0|\},$$

and

$$\mathcal{V}_{01} := \{b_k |\ A(b_k) \text{ has at least one normal white node}, 1 \leqslant k \leqslant |\mathcal{V}_0|\}.$$

The following lemma gives the key estimates we need.

Lemma 4.5.17. *For any partition* $\gamma \in \mathcal{W}$*,*

$$m(\gamma) \leqslant \frac{-|\mathcal{V}_{00}| - |\mathcal{V}_{01}|/2 - n_R^{(d)} + \sum_{k=1}^n \left[\deg(b_k) + \deg(\bar{b}_k)\right]}{2}, \tag{4.5.50}$$

and

$$n_X + n_R \geqslant p + |\mathcal{V}_1| + |\mathcal{V}_{00}|, \quad n_X \geqslant |\mathcal{V}_1|, \quad n_R^{(d)} \geqslant |\mathcal{V}_{00}|. \tag{4.5.51}$$

Proof. The second inequality of (4.5.51) can be proved easily through

$$|\mathcal{V}_1| \leqslant \left|\{k \in L | \mathcal{F}_{\text{in}}(k, \Omega_Q^X) \geqslant 2\}\right| \leqslant n_X.$$

Notice for $b_k \in \mathcal{V}_0$, $A(b_k)$ contains at least three diamond white nodes, while each of the white node is shared by another b_l. Thus we trivially have $|\mathcal{V}_{00}| \leqslant n_R^{(d)}$.

Now we prove (4.5.50). A diamond white node is connected to two black nodes and a normal white node is connected to one black node. Hence a diamond white node belongs to two sets $A(b_{k_1}), A(b_{k_2})$, and a normal white node belongs to exactly one set $A(b_k)$. Therefore for each $i = 1, ..., d$, if W_i contains exactly one $A(b_k)$, then

$$|W_i| \geqslant 3 \geqslant 2 + \mathbf{1}_{\mathcal{V}_{01}}(b_k) + \frac{\mathbf{1}_{\mathcal{V}_{00}}(b_k)}{2}.$$

Otherwise if W_i contains more than one $A(b_k)$, then

$$|W_i| \geqslant \sum_{b_k : A(b_k) \subseteq W_i} \left(2 \cdot \mathbf{1}_{\mathcal{V}_{01}}(b_k) + \frac{3}{2} \cdot \mathbf{1}_{\mathcal{V}_{00}}(b_k)\right) \geqslant 2 + \sum_{b_k : A(b_k) \subseteq W_i} \left(\mathbf{1}_{\mathcal{V}_{01}}(b_k) + \frac{\mathbf{1}_{\mathcal{V}_{00}}(b_k)}{2}\right).$$

Here the first inequality can be understood as following. For each black node b_k with $A(b_k) \subseteq W_i$, we count the number of white nodes in $A(b_k)$ and add them together. During the counting, we assign weight-1 to a normal white node and weight-1/2 to a diamond white node (since it is shared by two different black nodes). If $b_k \in \mathcal{V}_{00}$, there are at least three diamond white nodes in $A(b_k)$ with total weight $\geqslant 3/2$; if $b_k \in \mathcal{V}_{01}$, there are at least one normal white node and two other white nodes in $A(b_k)$ with total weight $\geqslant 2$. Thus $\sum_{b_k:A(b_k)\subseteq W_i} (2 \cdot \mathbf{1}_{\mathcal{V}_{01}}(b_k) + 3/2 \cdot \mathbf{1}_{\mathcal{V}_{00}}(b_k))$ is smaller than the number of white nodes in W_i. Then summing $|W_i|$ over i, we get

$$\sum_{i=1}^{d} |W_i| \geqslant 2d + |\mathcal{V}_{01}| + \frac{|\mathcal{V}_{00}|}{2}.$$

For the other $m - d$ blocks, each of them contains at least two white nodes. Therefore

$$2m + |\mathcal{V}_{01}| + \frac{|\mathcal{V}_{00}|}{2} \leqslant \sum_{i=1}^{d} |W_i| + 2(m-d) \leqslant -n_R^{(d)} + \sum_{k=1}^{n} \left[\deg(b_k) + \deg(\bar{b}_k)\right],$$

where we used (4.5.49) in the last step. This proves (4.5.50).

For $b_k \in \mathcal{V}_{00}$, $A(b_k)$ contains at least three white nodes from off-diagonal R-groups,

$$\mathcal{V}_{00} \subseteq \{b_k \in L \mid \mathcal{F}_{\text{in}}(b_k, \Omega_\Delta) = 1 \text{ and } \mathcal{F}_{\text{in}}(b_k, \Omega_Q^R) \geqslant 3\} =: \mathcal{V}_2.$$

Recall Lemma 4.5.15, only $\tau_1^{(k)}$ can increase $\mathcal{F}_{\text{in}}$. Thus $\mathbf{w}$ contains $\tau_1^{(b_k)}$ for each $b_k \in \mathcal{V}_1 \cup \mathcal{V}_2$ (recall the definition of $\mathcal{V}_1$ in (4.5.46)). Therefore by (4.5.22), (4.5.37) and the fact that $\mathcal{V}_{00}$ and $\mathcal{V}_1$ are disjoint, we have

$$n_X + n_R = \mathcal{F}_{\text{off}}((\Omega_\Delta)_{\mathbf{w}}) \geqslant \mathcal{F}_{\text{off}}(\Omega_\Delta) + |\mathcal{V}_1 \cup \mathcal{V}_2| \geqslant p + |\mathcal{V}_1| + |\mathcal{V}_{00}|.$$

This proves the first inequality of (4.5.51). □

Now we prove (4.5.40). By (4.2.3) and (4.5.6), a diagonal R edge contributes 1, an off-diagonal R edge contributes Φ, and an S^X or X edge contributes $N^{-1/2}$. Denote

$$\mathcal{U} = \prod_{l=1}^{n} \left|u_{[b_l]}\right|^{I_l} \left|v_{[b_l]}\right|^{J_l}.$$

Then using Lemma 2.4.2, we get

$$
\begin{aligned}
|\mathbb{E}Q(\mathbf{w},\Delta(\Gamma))| &\prec C\mathcal{U}\big(N^{-1/2}\big)^{n_X}\sum_{\gamma\in\mathcal{W}}\sum^{*}_{\gamma(W_1),\ldots,\gamma(W_m)\in\mathcal{I}\setminus L}\Phi^{n_R-n_R^{(d)}}\prod_{k=1}^{n}\big(N^{-1/2}\big)^{\deg(b_k)+\deg\left(\bar{b}_k\right)} \\
&\leqslant C\mathcal{U}N^{-n_X/2}\sum_{\gamma\in\mathcal{W}}N^{m-\frac{\sum_{k=1}^{n}\deg\left(b_k\right)+\deg\left(\bar{b}_k\right)}{2}}\Phi^{n_R-n_R^{(d)}} \\
&\leqslant C\mathcal{U}N^{-n_X/2}\sum_{\gamma\in\mathcal{W}}N^{\frac{-|\mathcal{V}_{01}|-|\mathcal{V}_{00}|/2-n_R^{(d)}}{2}}\Phi^{n_R-n_R^{(d)}} \\
&\leqslant C\mathcal{U}N^{-h/2}\sum_{\gamma\in\mathcal{W}}N^{-(n_X-|\mathcal{V}_1|)/2}N^{-(n_R^{(d)}-|\mathcal{V}_{00}|)/2}\Phi^{n_R-n_R^{(d)}} \\
&\leqslant C\mathcal{U}N^{-h/2}\sum_{\gamma\in\mathcal{W}}\Phi^{n_X+n_R-|\mathcal{V}_1|-|\mathcal{V}_{00}|}\leqslant C\mathcal{U}N^{-h/2}\Phi^{p},
\end{aligned}
$$

where in the third step we used (4.5.50), in the fourth step $h=|\mathcal{V}|=|\mathcal{V}_1|+|\mathcal{V}_{00}|+|\mathcal{V}_{01}|$, in the fifth step $N^{-1/2}\leqslant\Phi$ and (4.5.51), and in the last step (4.5.51). Thus we have proved (4.5.40), which concludes the proof of Proposition 4.5.1.

APPENDIX A

Stability of self-consistent equations

A.1 Proof of Lemma 4.2.3 and Proposition 4.2.18

We now prove Lemma 4.2.3. First is a technical lemma for f defined in (4.2.15).

Lemma A.1.1. *For $w > 0$ and $|z| > 0$, f can be written as*

$$f(\sqrt{w}, m) = -\sqrt{w} + m + w^{-1/2} + \frac{1}{N}\sum_{i=1}^{n} l_i s_i \left(\frac{A_i}{m - a_i} + \frac{B_i}{m - b_i} + \frac{C_i}{m + c_i}\right), \tag{A.1.1}$$

where we have the following estimates for the poles and the coefficients,

$$\max\left(|z|, \frac{s_i + |z|^2}{\sqrt{w}}\right) < a_i < \frac{s_i + |z|^2}{\sqrt{w}} + |z|, \quad a_n < a_{n-1} < \ldots < a_1, \tag{A.1.2}$$

$$0 < b_1 < b_2 < \ldots < b_n < \min\left(|z|, \frac{|z|^2}{\sqrt{w}}\right), \tag{A.1.3}$$

$$\frac{-(s_i + |z|^2) + \sqrt{(s_i + |z|^2)^2 + 4w|z|^2}}{2\sqrt{w}} < c_i < |z|, \quad c_1 < c_2 < \ldots < c_n, \tag{A.1.4}$$

and

$$0 < A_i \leqslant 2\frac{s_i + |z|^2 + \sqrt{w}|z|}{w}, \; 0 < B_i \leqslant 2\frac{s_i + |z|^2 + \sqrt{w}|z|}{w}, \; 0 < C_i \leqslant \frac{s_i + |z|^2 + \sqrt{w}|z|}{w}. \tag{A.1.5}$$

Proof. The proof is based on basic algebraic arguments. Let

$$p_i = \sqrt{w}m^3 - (s_i + |z|^2)m^2 - \sqrt{w}|z|^2 m + |z|^4.$$

It is easy to verify that

$$\Delta = 18(s_i + |z|^2)w|z|^6 + 4(s_i + |z|^2)^3|z|^4 + (s_i + |z|^2)^2 w|z|^4 + 4w^2|z|^6 - 27w|z|^8 > 0.$$

Thus p_i has three distinct real roots. By the form of p_i, we see that there are two positive roots and one negative root, call them $a_i > b_i > 0 > -c_i$. Now we perform the partial fraction expansion for the rational functions in (4.2.15):

$$\frac{m^2 - |z|^2}{\sqrt{w}m^3 - (s_i + |z|^2)m^2 - \sqrt{w}|z|^2 m + |z|^4} = \frac{A_i'}{m - a_i} + \frac{B_i'}{m - b_i} - \frac{C_i'}{m + c_i}, \tag{A.1.6}$$

where

$$A_i' = \frac{a_i^2 - |z|^2}{\sqrt{w}(a_i - b_i)(a_i + c_i)},\ B_i' = \frac{b_i^2 - |z|^2}{\sqrt{w}(b_i - a_i)(b_i + c_i)},\ C_i' = \frac{-c_i^2 + |z|^2}{\sqrt{w}(c_i + a_i)(c_i + b_i)}. \tag{A.1.7}$$

We take $s_i = 0$ in p_i and call the resulting polynomial as

$$p_0 = \sqrt{w}m^3 - |z|^2 m^2 - \sqrt{w}|z|^2 m + |z|^4 = \sqrt{w}\left(m - \frac{|z|^2}{\sqrt{w}}\right)\left(m^2 - |z|^2\right),$$

which has roots $m = \pm|z|, |z|^2/\sqrt{w}$. By (4.2.7), we have $p_1 < p_2 < \ldots < p_n < p_0$ for all $m \neq 0$. Comparing the graphs of p_i's (as cubic functions of m) for $0 \leqslant i \leqslant n$, we get that

$$\max\left(|z|, \frac{|z|^2}{\sqrt{w}}\right) < a_n < a_{n-1} < \ldots < a_1,\ 0 < b_1 < b_2 < \ldots < b_n < \min\left(|z|, \frac{|z|^2}{\sqrt{w}}\right), \tag{A.1.8}$$

and

$$0 < c_1 < c_2 < \ldots < c_n < |z|. \tag{A.1.9}$$

Thus we get (A.1.3). By these bounds, we see that $a_i^2 - |z|^2 > 0$, $b_i^2 - |z|^2 < 0$ and $-c_i^2 + |z|^2 > 0$, which, by (A.1.7), give that $A_i' > 0$, $B_i' > 0$ and $C_i' > 0$. Plugging (A.1.6) into f, we get immediately (A.1.1) with $A_i = A_i' a_i$, $B_i = B_i' b_i$ and $C_i = C_i' c_i$. The $w^{-1/2}$ term can be obtained by comparing the coefficients of the m^3 terms in (4.2.15) and using the normalization condition (4.2.8).

Now we compare p_i with $p_i' := \sqrt{w}m^3 - (s_i + |z|^2)m^2 - \sqrt{w}|z|^2 m$, which has roots

$$m = 0,\ \frac{(s_i + |z|^2) \pm \sqrt{(s_i + |z|^2)^2 + 4w|z|^2}}{2\sqrt{w}}.$$

Since $p_i' < p_i$ for all m, we get

$$a_i < \frac{(s_i + |z|^2) + \sqrt{(s_i + |z|^2)^2 + 4w|z|^2}}{2\sqrt{w}} < \frac{s_i + |z|^2}{\sqrt{w}} + |z|, \tag{A.1.10}$$

and

$$c_i > \frac{-(s_i + |z|^2) + \sqrt{(s_i + |z|^2)^2 + 4w|z|^2}}{2\sqrt{w}}. \tag{A.1.11}$$

Combining (A.1.9) and (A.1.11), we get (A.1.4). Then we compare p_i with $p_i'' := \sqrt{w}m^3 - (s_i + |z|^2)m^2$, which has roots $w = 0$, $(s_i + |z|^2)/\sqrt{w}$. Note that $p_i'' > p_i$ for $m > |z|^2/\sqrt{w}$, which gives $a_i > (s_i + |z|^2)/\sqrt{w}$ since $a_i > |z|^2/\sqrt{w}$. Combining this bound with (A.1.8) and (A.1.10), we get (A.1.2).

Finally we estimate the coefficients A_i, B_i and C_i. Using (A.1.7) and (A.1.2)-(A.1.4), we first can estimate that

$$\begin{aligned}
A_i' &= \frac{(a_i - |z|)(a_i + |z|)}{\sqrt{w}(a_i - b_i)(a_i + c_i)} \leqslant \frac{a_i + |z|}{\sqrt{w}(a_i + c_i)} \leqslant \frac{2}{\sqrt{w}}, \\
B_i' &= \frac{(|z| + b_i)(|z| - b_i)}{\sqrt{w}(a_i - b_i)(b_i + c_i)} \leqslant \frac{|z| + b_i}{\sqrt{w}(b_i + c_i)} \leqslant 2\frac{s_i + |z|^2 + \sqrt{w}|z|}{w|z|}, \\
C_i' &= \frac{(|z| - c_i)(c_i + |z|)}{\sqrt{w}(c_i + a_i)(c_i + b_i)} \leqslant \frac{|z| - c_i}{\sqrt{w}(c_i + b_i)} \leqslant \frac{s_i + |z|^2 + \sqrt{w}|z|}{w|z|},
\end{aligned}$$

with which we can get that

$$A_i = A_i' a_i \leqslant \frac{2}{\sqrt{w}} \left(\frac{s_i + |z|^2}{\sqrt{w}} + |z| \right) = 2\frac{s_i + |z|^2 + \sqrt{w}|z|}{w}, \tag{A.1.12}$$

$$B_i = B_i' b_i \leqslant 2\frac{s_i + |z|^2 + \sqrt{w}|z|}{w|z|}|z| = 2\frac{s_i + |z|^2 + \sqrt{w}|z|}{w}, \tag{A.1.13}$$

$$C_i = C_i' c_i \leqslant \frac{s_i + |z|^2 + \sqrt{w}|z|}{w|z|}|z| = \frac{s_i + |z|^2 + \sqrt{w}|z|}{w}. \tag{A.1.14}$$

This completes the proof. □

In (A.1.1), it is sometimes convenient to reorder the terms and rename the constants to write f as

$$f(m) = -\sqrt{w} + m + w^{-1/2} + \frac{1}{N}\sum_{k=1}^{2n} \frac{C_k^+}{m - x_k} + \frac{1}{N}\sum_{l=1}^{n} \frac{C_l^-}{m + y_l}. \tag{A.1.15}$$

where all the constants C_k^+ and C_l^- are positive and chosen such that

$$0 < x_1 < x_2 < \ldots < x_{2n},\ 0 < y_1 < y_2 < \ldots < y_n. \tag{A.1.16}$$

Clearly, f is smooth on the $3n + 1$ open intervals of $\mathbb{R}$ defined by

$$I_{-n} := (-\infty, -y_n),\ I_{-k} := (-y_{k+1}, -y_k)\ (k = 1, \ldots, n-1),\ I_0 := (-y_1, x_1),$$

$$I_k := (x_k, x_{k+1}) \ (k = 1, \dots, 2n-1), \ I_{2n} := (x_{2n}, +\infty).$$

Next, we introduce the multiset $\mathcal{C}$ of critical points of f (as a function of m), using the conventions that a nondegenerate critical point is counted once and a degenerated critical point twice. First we will prove the following elementary lemma about the structure of $\mathcal{C}$ (see Fig. A.1 and Fig. A.2).

Lemma A.1.2. *(Critical points) We have* $|\mathcal{C} \cap I_{-n}| = |\mathcal{C} \cap I_{2n}| = 1$ *and* $|\mathcal{C} \cap I_k| \in \{0, 2\}$ *for* $k = -n+1, \dots, 2n-1$.

Proof. We omit the dependence of f on w for now. By (A.1.15) we have

$$f'(m) = 1 - \frac{1}{N} \sum_{k=1}^{2n} \frac{C_k^+}{(m - x_k)^2} - \frac{1}{N} \sum_{l=1}^{n} \frac{C_l^-}{(m + y_l)^2},$$

and

$$f''(m) = \frac{1}{N} \sum_{k=1}^{2n} \frac{2C_k^+}{(m - x_k)^3} + \frac{1}{N} \sum_{l=1}^{n} \frac{2C_l^-}{(m + y_l)^3}.$$

We see that f'' is decreasing on all the intervals I_k for $k = -n+1, \dots, 2n-1$. Thus there is at most one point $m \in I_k$ such that $f''(m) = 0$. We conclude that f has at most two critical points on I_k. By the boundary conditions of f' on ∂I_k, we get $|\mathcal{C} \cap I_k| \in \{0, 2\}$ for $k = -n+1, \dots, 2n-1$. For $m < -y_n$, we have $f''(m) < 0$, while for $m > x_{2n}$, we have $f''(m) > 0$. By the boundary conditions of f' on ∂I_{-n} and ∂I_{2n}, we see that f' decreases from 1 to $-\infty$ when m increases from $-\infty$ to $-y_n$, while f' increases from $-\infty$ to 1 when m increases from x_{2n} to $+\infty$. Hence we conclude that each of the intervals $(-\infty, -y_n)$ and $(x_{2n}, +\infty)$ contains a unique critical point in it, i.e. $|\mathcal{C} \cap I_{-n}| = |\mathcal{C} \cap I_{2n}| = 1$. □

From this lemma, we deduce that $|\mathcal{C}| = 2p$ is even. We denote by z_{2p} the critical point in I_{-n}, z_1 the critical point in I_{2n}, and $z_2 \geqslant \dots \geqslant z_{2p-1}$ the $2p-2$ critical points in $I_{-n+1} \cup \dots \cup I_{2n-1}$. For $k = 1, \dots, 2p$, we define the critical values $h_k := f(z_k)$. The next lemma is crucial in establishing the basic properties of ρ_{1c} (see e.g. Fig. A.1).

Lemma A.1.3. *(Orderings of the critical values) The critical values are ordered as* $h_1 \geqslant h_2 \geqslant \dots \geqslant h_{2p}$. *Furthermore, there is an absolute constant* $C_0 > 0$ *independent of* τ *such that* $h_k \in [-C_0(\tau^{-1}|w|^{-1/2} + |z|) - \sqrt{w}, C_0(\tau^{-1}|w|^{-1/2} + |z|) - \sqrt{w}]$ *for* $k = 1, \dots, 2p$.

Proof. Notice for the equation (4.2.14), if we multiply both sides with the product of all denominators in f, we get a polynomial equation $P_w(m) = 0$ with P_w being a polynomial of degree $3n + 1$. An immediate consequence is that for any fixed $w > 0$ and $E \in \mathbb{R}$, $f(\sqrt{w}, m) = E$ can have at most $3n + 1$ roots in m. This fact will be useful in the proof of this lemma and Lemma 4.2.3.

For $i = -n, \ldots, 2n$, define the subset $J_i(w) := \{m \in I_i : \partial_m f(\sqrt{w}, m) > 0\}$. From Lemma A.1.2, we deduce that if $i = -n + 1, \ldots, 2n - 1$, then $J_i \neq \varnothing$ if and only if I_i contains two distinct critical points of f, in which case J_i is an interval. Moreover, we have $J_{-n} = (-\infty, z_{2p})$ and $J_{2n} = (z_1, +\infty)$. Next, we observe that for any $-n \leqslant i < j \leqslant 2n$, we have $f(J_i) \cap f(J_j) = \varnothing$. Otherwise if there were $E \in f(J_i) \cap f(J_j)$, we would have $|\{x : f(x) = E\}| > 3n + 1$. We hence conclude that the sets $f(J_i)$, $-n \leqslant i \leqslant 2n$ can be strictly ordered. The claim $h_1 \geqslant h_2 \geqslant \ldots \geqslant h_{2p}$ is now reformulated as

$$f(J_i) < f(J_j) \text{ whenever } i < j \text{ and } J_i, J_j \neq \varnothing. \tag{A.1.17}$$

To prove (A.1.17), we use a continuity argument. Let $t \in (0, 1]$ and introduce

$$f^t(m) = -\sqrt{w} + m + w^{-1/2} + \frac{t}{N}\sum_{k=1}^{2n} \frac{C_k^+}{m - x_k} + \frac{t}{N}\sum_{l=1}^{n} \frac{C_l^-}{m + y_l}.$$

It is easy to check (A.1.17) holds for small enough $t > 0$. We claim that

$$J_i \neq \varnothing \Rightarrow J_i^t \neq \varnothing \text{ for all } t \in (0, 1]. \tag{A.1.18}$$

This is trivial for $i = -n, 2n$. Recall that for $-n+1 \leqslant i \leqslant 2n-1$, $J_i^t \neq \varnothing$ is equivalent to I_i containing two distinct critical points. Moreover, $\partial_t \partial_m f^t(m) < 0$ in $I_{-n+1} \cup \ldots \cup I_{2n-1}$, from which we deduce that the number of distinct critical points in each I_i, $i = -n+1, \ldots, 2n-1$, does not decreases as t decreases. This proves (A.1.18).

Next, suppose that there exist $i < j$ such that $J_i, J_j \neq \varnothing$ and $f(J_i) > f(J_j)$. From (A.1.18), we deduce that $J_i^t, J_j^t \neq \varnothing$ for all $t \in (0, 1]$. By a simple continuity argument, we get that $f^t(J_i^t) > f^t(J_j^t)$ for all $t \in (0, 1]$. However, this is impossible for small enough t as explained before (A.1.18). This concludes the proof of (A.1.17).

To prove the second statement of Lemma A.1.3, we only need to show that $h_1 \leqslant C_0(\tau^{-1}|w|^{-1/2} + |z|) - \sqrt{w}$ and $h_{2p} \geqslant -C_0(\tau^{-1}|w|^{-1/2} + |z|) - \sqrt{w}$ for some absolute constant C_0. We only give the proof for h_1; the proof for h_{2p} is similar. At z_1, we have

$$f(z_1)+\sqrt{w} \leqslant (z_1+y_n)\left[1 + \frac{1}{N}\sum_{k=1}^{2n} \frac{C_k^+}{(z_1 - x_k)^2} + \frac{1}{N}\sum_{l=1}^{n} \frac{C_l^-}{(z_1 + y_l)^2}\right] + w^{-1/2} = 2(z_1+y_n)+w^{-1/2},$$

where we used

$$0 = f'(z_1) = 1 - \frac{1}{N}\sum_{k=1}^{2n} \frac{C_k^+}{(z_1 - x_k)^2} - \frac{1}{N}\sum_{l=1}^{n} \frac{C_l^-}{(z_1 + y_l)^2}. \tag{A.1.19}$$

Now we would like to estimate $z_1 + y_n$. Again using (A.1.19), we have that

$$\frac{1}{N}\sum_{k=1}^{2n} \frac{C_k^+}{(z_1 - x_{2n})^2} + \frac{1}{N}\sum_{l=1}^{n} \frac{C_l^-}{(z_1 - x_{2n})^2} \geqslant 1.$$

Then by (A.1.5) we get

$$z_1 - x_{2n} \leqslant \sqrt{\frac{1}{N}\sum_{k=1}^{2n} C_k^+ + \frac{1}{N}\sum_{l=1}^{n} C_l^-} \leqslant \sqrt{5\frac{\tau^{-1} + |z|^2 + \sqrt{w}|z|}{w}}.$$

Using the above estimates and (A.1.2)-(A.1.4), we obtain that

$$\begin{aligned} f(z_1) &\leqslant 2\left(\sqrt{5\frac{\tau^{-1} + |z|^2 + \sqrt{w}|z|}{w}} + \frac{s_1 + |z|^2}{\sqrt{w}} + 2|z|\right) + w^{-1/2} - \sqrt{w} \\ &\leqslant C_0(\tau^{-1}|w|^{-1/2} + |z|) - \sqrt{w}. \end{aligned}$$

for some constant $C_0 > 0$ that does not depend on τ. □

Proof of Lemma 4.2.3. Let $J(w) := \bigcup_{i=-n}^{2n} J_i(w)$. Given $w > 0$ such that $0 \in f(J(w))$, then the set $\{m \in \mathbb{R} : f(\sqrt{w}, m) = 0\}$ has $3n + 1$ points. Since $f(\sqrt{w}, m) = 0$ has at most $3n + 1$ solutions in m, we deduce that $m_c(w)$ is real and hence $m_{1c}(w)$ is also real. Since m_{1c} is the Stieltjes transform of ρ_{1c}, we conclude that $w \notin \operatorname{supp} \rho_{1c}$. On the other hand, suppose $w > 0$ and $0 \notin f(J(w))$. Then the set of preimages $\{m \in \mathbb{R} : f(\sqrt{w}, m) = 0\} = \{m \in \mathbb{R} : P_w(m) = 0\}$ has $3n - 1$ points. Since $P_w(m)$ is a degree $3n + 1$ polynomial with real coefficients, we conclude that P_w has a unique root with positive imaginary part. By the uniqueness of the solution of $P_{w+i\eta}$ in $\mathbb{C}_+$ (Lemma 4.2.2) and the continuity of the roots of $P_{w+i\eta}$ in η, we

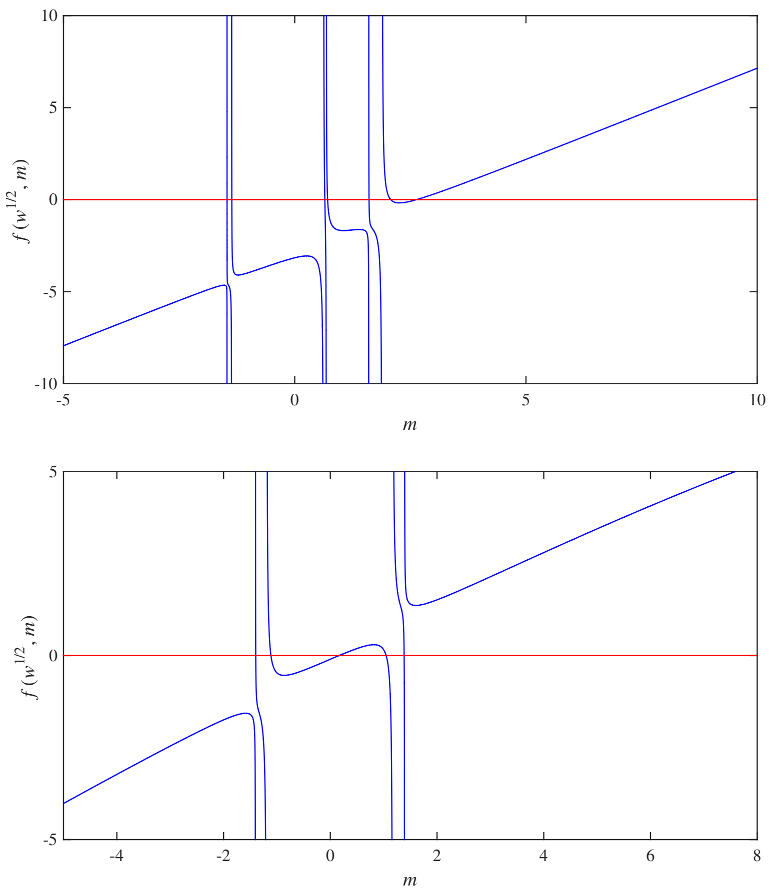


Figure A.1: The graphs of $f(\sqrt{w}, m)$ for the example from Figure 4.1, i.e. $\rho_\Sigma = 0.5\delta_{\sqrt{2/17}} + 0.5\delta_{4\sqrt{2/17}}$. We take $|z| = 1.5$, and $w = 10$ and 0.01 in the upper and lower graphs, respectively. In the lower graph, we only plot the five branches near $m = 0$. The remaining two branches are far away.

conclude that $\operatorname{Im} m_c(w) > 0$ and hence $\operatorname{Im} m_{1c}(w) > 0$ by taking $\eta \searrow 0$, i.e. $w \in \operatorname{supp} \rho_{1c}$. In sum, we get

$$\operatorname{supp} \rho_{1c} = \overline{\{w > 0 : 0 \notin f\,(J(w))\}}. \tag{A.1.20}$$

From Lemma A.1.3, we see that there exists an absolute constant $C_1 > 0$ such that if $w \geqslant C_1\tau^{-1}$, then $h_1(\omega) \leqslant C_0(\tau^{-1}|w|^{-1/2} + |z|) - \sqrt{w} < 0$. Hence fix $w \geqslant C_1\tau^{-1}$, we have $0 \in f(J_{2n}(w))$ and $w \notin \operatorname{supp} \rho_{1c}$ (see the upper graphs in Fig. A.1 and Fig. A.2). This shows that ρ_{1c} is compactly supported in $[0, C_1\tau^{-1}]$. Now we decrease w so that $w < s_1 + |z|^2 + 1$.

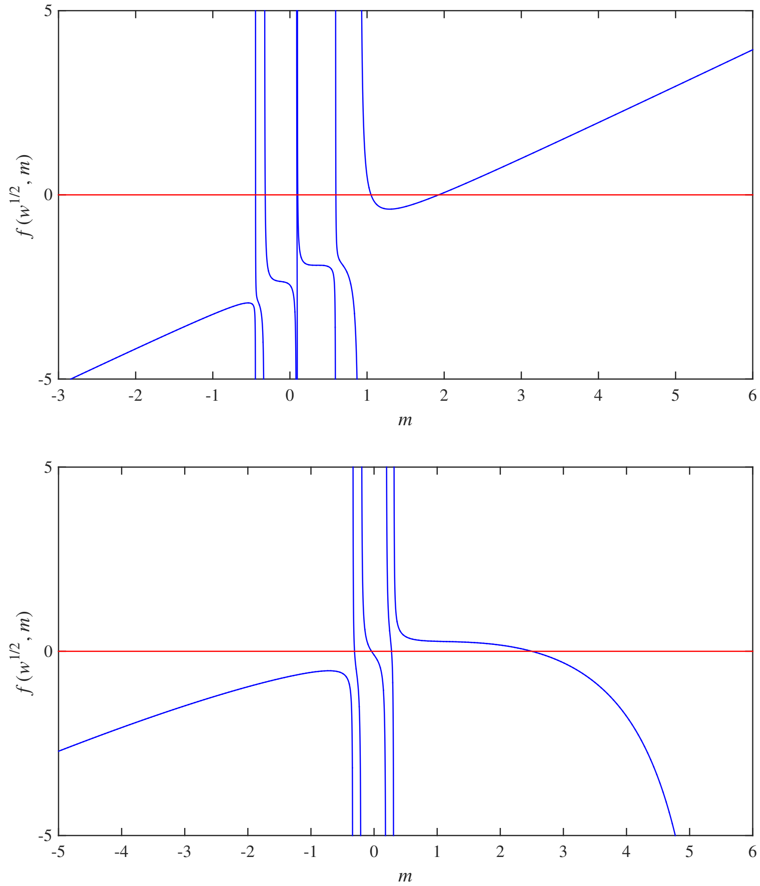


Figure A.2: The graphs of $f(\sqrt{w}, m)$ for the example from Figure 4.1, i.e. $\rho_\Sigma = 0.5\delta_{\sqrt{2/17}} + 0.5\delta_{4\sqrt{2/17}}$. We take $|z| = 0.5$, and $w = 6$ and 0.01 in the upper and lower graphs, respectively. In the lower graph, we only plot the five branches near $m = 0$. The remaining two branches are far away.

Then using (A.1.2), we have

$$h_1(w) > z_1 + w^{-1/2} - \sqrt{w} > \frac{s_1 + |z|^2 + 1 - w}{\sqrt{w}} > 0.$$

By continuity, there must be some $0 < w < C\tau^{-1}$ such that $0 \notin f\left(J(w)\right)$. Thus $\operatorname{supp} \rho_{1c} \neq \varnothing$. By (A.1.20), it is not hard to see that $\operatorname{supp} \rho_{1c}$ is a disjoint union of (countably many) closed intervals,

$$\operatorname{supp} \rho_{1c} = \bigcup_k [e_{2k}, e_{2k-1}], \tag{A.1.21}$$

where $C_1\tau^{-1} \geqslant e_1 \geqslant e_2 \geqslant \cdots$. Furthermore, for e_i to be a boundary point, we must have that 0 is a critical value of $f(\sqrt{e_i}, m)$, i.e. there is a unique critical point $m = m_c(e_i)$ such

that

$$f(\sqrt{e_i}, m_c(e_i)) = 0, \quad \partial_m f(\sqrt{e_i}, m_c(e_i)) = 0. \tag{A.1.22}$$

Notice the two equations in (A.1.22) are equivalent to two polynomial equations in $(\sqrt{w}, m)$ with order $3n+1$ and $6n$, respectively. By Bézout's theorem, there are at most finitely many solutions to the equations (A.1.22). Hence there are finitely many e_i's, call them $e_1 \geqslant \ldots \geqslant e_{2L}$, where $L \equiv L(n) \in \mathbb{N}$. The statement about e_{2L} follows from Lemma A.1.4 below. This concludes Lemma 4.2.3. □

Lemma A.1.4. *If $1+\tau \leqslant |z|^2 \leqslant 1+\tau^{-1}$, there is a constant $\varepsilon(\tau) > 0$ so that $e_{2L} \geqslant \varepsilon(\tau)$. If $|z|^2 \leqslant 1-\tau$, $e_{2L} = 0$ and $\rho_{1c}(x) \sim x^{-1/2}$ when $x \searrow 0$.*

By this lemma, the behavior of the leftmost edge e_{2L} changes essentially when z crosses the unit circle. From the following proof, we will see that the singularity happens at $|z|^2 = N^{-1}\sum_{i=1}^n l_i s_i$. Thus the fact that the singular circle has radius 1 is due to our normalization (4.2.5) for T.

Proof of Lemma A.1.4. We first study the equation (4.2.14) when $w \searrow 0$ in the case $1+\tau \leqslant |z|^2 \leqslant 1+\tau^{-1}$. We calculate the derivative of f as

$$\begin{aligned}\partial_m f(\sqrt{w}, m) = 1 &+ \frac{1}{N}\sum_{i=1}^n l_i s_i \frac{m^2 - |z|^2}{\sqrt{w}m^3 - (s_i + |z|^2)m^2 - \sqrt{w}|z|^2 m + |z|^4} \\ &- \frac{m}{N}\sum_{i=1}^n l_i s_i \frac{\sqrt{w}\left(m^2 - |z|^2\right)^2 + 2s_i|z|^2 m}{\left[\sqrt{w}m^3 - (s_i + |z|^2)m^2 - \sqrt{w}|z|^2 m + |z|^4\right]^2}.\end{aligned} \tag{A.1.23}$$

Recall the definition of J_i in the proof of Lemma A.1.3. It is easy to see that $J_0 \neq \varnothing$ for all $w > 0$, since $\partial_m f(\sqrt{w}, 0) = 1 - |z|^{-2} > 0$ (see the lower graph in Fig. A.1). Call the end points of J_0 as $z_k(w) > 0$ and $z_{k+1}(w) < 0$. By the definition of I_0, we have $z_k < b_1 < |z|$. Suppose $z_k = o(|z|)$ as $w \to 0$, then (A.1.23) gives that $0 = 1 - |z|^{-2} + o(1)$, which gives a contradiction. Thus $z_k \sim |z|$ as $w \to 0$. Now using $\partial_m f(\sqrt{w}, z_k) = 0$, we can estimate that

$$\begin{aligned}f(\sqrt{w}, z_k) &= -\sqrt{w} + \frac{z_k^2}{N}\sum_{i=1}^n l_i s_i \frac{\sqrt{w}\left(z_k^2 - |z|^2\right)^2 + 2s_i|z|^2 z_k}{\left[\sqrt{w}z_k^3 - (s_i + |z|^2)z_k^2 - \sqrt{w}|z|^2 z_k + |z|^4\right]^2} \\ &\geqslant -\sqrt{w} + \frac{1}{N}\sum_{i=1}^n l_i s_i \frac{2s_i|z|^2 z_k^3}{|z|^8} \geqslant c - \sqrt{w}\end{aligned} \tag{A.1.24}$$

for some constant $c > 0$ independent of w, where in the second step we used that

$$\sqrt{w}z_k^3 - (s_i + |z|^2)z_k^2 - \sqrt{w}|z|^2 z_k + |z|^4 > 0, \text{ and } \sqrt{w}z_k^3 - (s_i + |z|^2)z_k^2 - \sqrt{w}|z|^2 z_k < 0$$

which come from the fact that $0 < z_k < b_i < |z|$ for all $1 \leqslant i \leqslant n$. By (A.1.24), we can find ε small enough such that $f(\sqrt{w}, z_k) > 0$ for all $0 < w \leqslant \varepsilon$. In this case, $0 \in f(J_0(w))$ and hence $w \notin \operatorname{supp} \rho_{1c}$. In fact, it is not hard to see that there is a solution $m_0 = \sqrt{w}|z|^2/(|z|^2 - 1) + o(\sqrt{w}) \in I_0$ such that $f(\sqrt{w}, m_0) = 0$ and $\partial_m f(\sqrt{w}, m_0) > 0$. This proves the first statement of Lemma A.1.4.

Now we study equation (4.2.14) when $|z|^2 \leqslant 1 - \tau$ and $w \to 0$. For later purpose, we allow w to be complex and prove a more general result than what we need for this lemma. Let $w = 0$ in the equation (4.2.14), we get $m = 0$ or

$$0 = 1 + \frac{1}{N}\sum_{i=1}^{n} l_i s_i \frac{m^2 - |z|^2}{-(s_i + |z|^2)m^2 + |z|^4}. \tag{A.1.25}$$

We define

$$g(x) := 1 + \frac{1}{N}\sum_{i=1}^{n} l_i s_i \frac{x - |z|^2}{-(s_i + |z|^2)x + |z|^4} = \frac{|z|^2}{N}\sum_{i=1}^{n} l_i \frac{-x + |z|^2 - s_i}{-(s_i + |z|^2)x + |z|^4}. \tag{A.1.26}$$

It is easy to see that g is smooth and decreasing on the intervals defined through

$$K_1 := \left(-\infty, \frac{|z|^4}{s_1 + |z|^2}\right), \quad K_{n+1} := \left(\frac{|z|^4}{s_n + |z|^2}, \infty\right),$$

and

$$K_i := \left(\frac{|z|^4}{s_{i-1} + |z|^2}, \frac{|z|^4}{s_i + |z|^2}\right), \quad i = 2, \ldots, n.$$

By the boundary values of g on these intervals, we see that $g(x)$ has exactly one zero on intervals K_i for $i = 1, \ldots, n$, and has no zero on K_{n+1}. Since $g(x) = 0$ is equivalent to a polynomial equation of order n, it has at most n solutions. We conclude that all of its solutions are real. Obviously, the zeros on the intervals K_i are positive for $i = 2, \ldots, n$. Now we study the zero on K_1. Observe that $g(0) = 1 - |z|^{-2} < 0$ (as $|z|^2 \leqslant 1 - \tau$), hence the zero on K_1 is negative, call it $-t$. Moreover, it is easy to verify that $g(-\tau^{-1}) > 0$ using (A.1.26), so $t < \tau^{-1}$. If $|z|^2 \geqslant \tau/2$, then by the concavity of g on the K_1, we get

$$t \geqslant \frac{g(0)}{g'(0)} \geqslant \frac{|z|^4(1 - |z|^2)}{s_1} \geqslant \frac{\tau^4}{4}. \tag{A.1.27}$$

In the case $|z|^2 \leqslant \tau/2$, we have $|z|^2 - s_n \leqslant -\tau/2$ and $g(|z|^2 - s_n) \leqslant 0$ by (A.1.26). Hence we have

$$-t \leqslant |z|^2 - s_n \leqslant -\tau/2. \tag{A.1.28}$$

Combining (A.1.27) and (A.1.28), we get that $c\tau^4 \leqslant t \leqslant \tau^{-1}$ for some constant $c > 0$.

Now we return to the self-consistent equation (4.2.14). The previous discussion shows that

$$f(0, i\sqrt{t}) = 0, \quad \text{with } t \geqslant c\tau^4.$$

It is easy to see that there exist constants $c_1, \tau' > 0$ such that

$$\left|-(s_i + |z|^2)m^2 + |z|^4 + \sqrt{w}\left(m^3 - |z|^2 m\right)\right| \geqslant c_1 \text{ for } |m - i\sqrt{t}| \leqslant \tau'. \tag{A.1.29}$$

First we consider the case $|z| \geqslant \varepsilon > 0$. Expanding $f(\sqrt{w}, m)$ around $(0, i\sqrt{t})$ and using (A.1.29), we get

$$0 = \partial_{\sqrt{w}} f(0, i\sqrt{t})\sqrt{w} + \partial_m f(0, i\sqrt{t})(m - i\sqrt{t}) + o(\sqrt{w}) + o(m - i\sqrt{t}). \tag{A.1.30}$$

By (A.1.23), the partial derivative

$$\partial_{\sqrt{w}} f(\sqrt{w}, m) = -1 - \frac{m^2}{N}\sum_{i=1}^n l_i s_i \frac{(m^2 - |z|^2)^2}{\left[-(s_i + |z|^2)m^2 + |z|^4 + \sqrt{w}\,(m^3 - |z|^2 m)\right]^2}, \tag{A.1.31}$$

and (A.1.29), we obtain that $\left|\partial_{\sqrt{w}} f(0, i\sqrt{t})\right| \leqslant C$ and

$$\partial_m f(0, i\sqrt{t}) = \frac{t}{N}\sum_{i=1}^n l_i s_i \frac{2 s_i |z|^2}{\left[(s_i + |z|^2)t + |z|^4\right]^2} \geqslant c_2 \tag{A.1.32}$$

for some constant $c_2 > 0$. Using (A.1.32), we get from (A.1.30) that

$$m - i\sqrt{t} = O(\sqrt{w}), \quad \text{if } |z| \geqslant \varepsilon. \tag{A.1.33}$$

Then assume that $|z|^2 < \varepsilon$ for sufficiently small ε. From $g(-t) = 0$ and (A.1.26), we get that

$$\frac{1}{N}\sum_{i=1}^n l_i \frac{t + |z|^2 - s_i}{(s_i + |z|^2)t + |z|^4} = 0. \tag{A.1.34}$$

From the leading order term, we get $t^{-1} = t_0^{-1} + O(|z|^2)$, where $t_0 := \left(N^{-1}\sum_i l_i/s_i\right)^{-1}$. Expanding (A.1.34) up to the first order of $|z|^2$, we get

$$t = t_0 + \left(\frac{t_0^2}{N}\sum_i \frac{l_i}{s_i^2} - 2\right)|z|^2 + O(|z|^4). \tag{A.1.35}$$

Now we write equation (4.2.14) as

$$F(\sqrt{w}, m) = 0, \tag{A.1.36}$$

where $F(\sqrt{w}, m) := f(\sqrt{w}, m)/m$. Expanding F around $(0, i\sqrt{t})$ and using (A.1.29), we get

$$\begin{aligned} 0 =& \partial_{\sqrt{w}} F(0, i\sqrt{t})\sqrt{w} + \partial_m F(0, i\sqrt{t})(m - i\sqrt{t}) + \partial_m \partial_{\sqrt{w}} F(0, i\sqrt{t})(m - i\sqrt{t})\sqrt{w} \\ &+ \frac{1}{2}\partial^2_{\sqrt{w}} F(0, i\sqrt{t})w + \frac{1}{2}\partial_m^2 F(0, i\sqrt{t})(m - i\sqrt{t})^2 + o(w, |m - i\sqrt{t}|^2, |m - i\sqrt{t}|\sqrt{w}). \end{aligned} \tag{A.1.37}$$

We can calculate that (the partial derivatives of F can be obtained using (A.1.23) and (A.1.31))

$$\partial_m F(\sqrt{w}, i\sqrt{t}) = -\frac{2i|z|^2 + 2\sqrt{wt_0}}{t_0^{3/2}} + o(|z|^2, \sqrt{w}), \tag{A.1.38}$$

$$\partial_{\sqrt{w}} F(\sqrt{w}, i\sqrt{t}) = \left(i|z|^2 + 2\sqrt{wt_0}\right)\frac{\sqrt{t_0}}{N}\sum_{j=1}^n \frac{l_j}{s_j^2} + o(|z|^2, \sqrt{w}). \tag{A.1.39}$$

From (A.1.38) and (A.1.39), we get that

$$\begin{aligned} &\partial_m F(0, i\sqrt{t}) = -\frac{2i|z|^2}{t_0^{3/2}} + o(|z|^2),\ \partial_{\sqrt{w}} F(0, i\sqrt{t}) = \frac{i|z|^2\sqrt{t_0}}{N}\sum_{j=1}^n \frac{l_j}{s_j^2} + o(|z|^2), \\ &\partial_m\partial_{\sqrt{w}} F(0, i\sqrt{t}) = -\frac{2}{t_0} + O(|z|^2),\ \partial^2_{\sqrt{w}} F(0, i\sqrt{t}) = \frac{2t_0}{N}\sum_{j=1}^n \frac{l_j}{s_j^2} + O(|z|^2), \\ &\partial_m^2 F(0, i\sqrt{t}) = O(|z|^2). \end{aligned}$$

Plugging the above results into (A.1.37), we get that

$$\begin{aligned} 0 =& \left[\frac{i|z|^2\sqrt{t_0} + \sqrt{w}t_0}{N}\sum_{j=1}^n \frac{l_j}{s_j^2} + o(|z|^2)\right]\sqrt{w} + \left[-2\frac{i|z|^2 + \sqrt{wt_0}}{t_0^{3/2}} + o(|z|^2)\right](m - i\sqrt{t}) \\ &+ o(w, |m - i\sqrt{t}|^2, |m - i\sqrt{t}|\sqrt{w}). \end{aligned} \tag{A.1.40}$$

Observing that $|i|z|^2\sqrt{t_0} + \sqrt{w}t_0| \sim |z|^2 + \sqrt{|w|}$, we get

$$m - i\sqrt{t} = \left[\frac{t_0^2}{2N}\sum_{j=1}^{n}\frac{l_j}{s_j^2} + O(|w|^{1/2} + |z|^2)\right]\sqrt{w}, \quad \text{if } |z| < \varepsilon. \tag{A.1.41}$$

Combing (A.1.33) and (A.1.41), we get that if $|z|^2 < 1 - \tau$, $m = i\sqrt{t} + O(\sqrt{w})$ when $w \to 0$. In particular, this shows that $|m| \approx \operatorname{Im} m \sim 1$ when $w \to 0$. Finally, we conclude the proof of Lemma A.1.4 by using that $m_{1c}(w) = m_c(w)w^{-1/2} - 1$. □

To prove Proposition 4.2.18, we need the following lemma, which is a consequence of the edge regularity conditions (4.2.18) and (4.2.19).

Lemma A.1.5. *Suppose $e_k \neq 0$ is a regular edge. Then $|m_{1c}(w) - m_{1c}(e_k)| \sim |w - e_k|^{1/2}$ as $w \to e_k$ and $\min_{l\neq k}|e_l - e_k| \geqslant \delta$ for some constant $\delta > 0$.*

Proof. Denote $m_k := m_c(e_k)$ and let $w \to e_k$. Note that by Lemma 4.2.3 and Lemma A.1.4, if $e_k \neq 0$, we have

$$\varepsilon' \leqslant e_k \leqslant C\tau^{-1}, \tag{A.1.42}$$

for some constant $\varepsilon' > 0$. Then we expand f around $(\sqrt{e_k}, m_k)$ to get that

$$\begin{aligned} 0 =& \partial_{\sqrt{w}} f(\sqrt{e_k}, m_k)(\sqrt{w} - \sqrt{e_k}) + \frac{1}{2}\partial_m^2 f(\sqrt{e_k}, m_k)(m_c(w) - m_k)^2 \\ &+ O\left[|\sqrt{w} - \sqrt{e_k}|^2 + |m_c(w) - m_k|^3 + |\sqrt{w} - \sqrt{e_k}||m_c(w) - m_k|\right], \end{aligned} \tag{A.1.43}$$

where by (A.1.31),

$$\partial_{\sqrt{w}} f(\sqrt{e_k}, m_k) = -1 - \frac{m_k^2}{N}\sum_{i=1}^{n} l_i s_i \frac{(m_k^2 - |z|^2)^2}{e_k(m_k - a_i)^2(m_k - b_i)^2(m_k + c_i)^2}, \tag{A.1.44}$$

and by (A.1.1),

$$\partial_m^2 f(\sqrt{e_k}, m_k) = \frac{2}{N}\sum_{i=1}^{n} l_i s_i \left[\frac{A_i}{(m_k - a_i)^3} + \frac{B_i}{(m_k - b_i)^3} + \frac{C_i}{(m_k + c_i)^3}\right]. \tag{A.1.45}$$

Applying (A.1.2)-(A.1.5), (A.1.42) and the conditions (4.2.18)-(4.2.19) to (A.1.44) and (A.1.45), we get that

$$1 \leqslant \left|\partial_{\sqrt{w}} f(\sqrt{e_k}, m_k)\right| \leqslant C_1, \quad \varepsilon \leqslant \left|\partial_m^2 f(\sqrt{e_k}, m_k)\right| \leqslant C_2 \tag{A.1.46}$$

for some $C_1, C_2 > 0$. Similarly, if $|w - e_k| \leqslant \tau'$ and $|m_c(w) - m_k| \leqslant \tau'$ for some sufficiently small τ', using the condition (4.2.18) we can get that

$$\max\left\{\left|\partial_m^3 f(\sqrt{w}, m_c(w))\right|, \left|\partial_{\sqrt{w}}^2 f(\sqrt{w}, m_c(w))\right|, \left|\partial_m \partial_{\sqrt{w}} f(\sqrt{w}, m_c(w))\right|\right\} \leqslant C_3. \qquad (A.1.47)$$

Plugging them into equation (A.1.43), for $|w - e_k| \leqslant \tau'$ and $|m_c(w) - m_k| \leqslant \tau'$, we get $|m_c(w) - m_k| \sim |\sqrt{w} - \sqrt{e_k}|^{1/2}$ and

$$-\partial_{\sqrt{w}} f(\sqrt{e_k}, m_k)(\sqrt{w} - \sqrt{e_k}) + O(|\sqrt{w} - \sqrt{e_k}|^{3/2}) = \frac{1}{2}\partial_m^2 f(\sqrt{e_k}, m_k)(m_c(w) - m_k)^2. \quad (A.1.48)$$

By (A.1.42), we immediately get that $|\sqrt{w} - \sqrt{e_k}| \sim |w - e_k|$ and $|m_c(w) - m_k| \sim |m_{1c}(w) - m_{1c}(e_k)|$, which proves the first part of the lemma. By (A.1.48), if w is real and $|w - e_k| \leqslant \tau'$, we have that

$$m_c(w) - m_k = \left[\frac{-2\partial_{\sqrt{w}} f(\sqrt{e_k}, m_k)}{\partial_m^2 f(\sqrt{e_k}, m_k)} + O(|\sqrt{w} - \sqrt{e_k}|^{1/2})\right]^{1/2} \left(\sqrt{w} - \sqrt{e_k}\right)^{1/2}. \qquad (A.1.49)$$

Thus in a sufficiently small interval $U = [e_k - \delta, e_k + \delta]$, $m_c(w)$ has positive imaginary part for w on one side of e_k, while $m_c(w)$ is real for w on the other side. Hence U does not contain another edge. This shows that $\min_{l \neq k} |e_l - e_k| \geqslant \delta$. □

Proof of Proposition 4.2.18. The properties of ρ_{1c} have been proved in Lemmas 4.2.3, A.1.4 and A.1.5, and included in Definition 4.2.4. Since $\operatorname{supp} \rho_{2c} = \operatorname{supp} \rho_{1c}$ by the discussion after Lemma 4.2.2, we immediately get property (i) for ρ_{2c}. The conclusion ρ_{2c} being a probability measure is due to the definition of m_2 in (4.2.33) and the fact that m_{2c} is the almost sure limit of m_2.

The properties (ii) and (iv) for ρ_{2c} can be easily obtained by plugging m_{1c} into (4.2.9). To prove the property (iii) for ρ_{2c}, we need to know the behavior of $\operatorname{Im} m_{2c}(w)$ when $w \to e_j$ along the real line. By (4.2.9), it suffices to prove that if $|x - e_j| \leqslant \tau'$ for some small enough $\tau' > 0$, then

$$\left|-w(1 + m_{1c})^2 + |z|^2\right| = \left|m_c^2 - |z|^2\right| \geqslant \varepsilon$$

for some constant $\varepsilon > 0$. Suppose that $|m_c^2(w) - |z|^2| = o(1)$. Then plugging m_c into $\partial_m f(\sqrt{w}, m_c)$ in (A.1.23), and using condition (4.2.18) and Lemma A.1.5, we get that

$$\partial_m f(\sqrt{w}, m_c(w)) = -1 + O(|m_c^2 - |z|^2|). \qquad (A.1.50)$$

Again using condition (4.2.18) and Lemma A.1.5, we can bound $\partial_{\sqrt{w}}\partial_m f(\sqrt{w}, m_c(w))$ and $\partial_m^2 f(\sqrt{w}, m_c(w))$ for w near e_j. Thus we shall have that

$$0 = \partial_m f(\sqrt{e_j}, m_c(e_j)) = \partial_m f(\sqrt{w}, m_c(w)) + O(|w - e_j|^{1/2}) = -1 + O(|m_c^2 - |z|^2| + |w - e_j|^{1/2}). \tag{A.1.51}$$

This gives a contradiction. Thus we must have a lower bound for $|m_c^2 - |z|^2|$. $\square$

Remark A.1.6. Here we add a small remark on Example 4.2.11. Given the assumptions in Example 4.2.11, it is easy to see that f can only take critical values on intervals I_{-n}, I_0, I_n and I_{2n}, since $\max\{|a_i - a_{i-1}|, |b_i - b_{i-1}|, |c_i - c_{i-1}|\} \to 0$ in this case. Thus the number of connected components of $\operatorname{supp}\rho_{1c}$ is independent of n, and all the edges and the bulk components are regular as in Example 4.2.10.

A.2 Proof of Lemmas 4.3.4 and 4.3.5

We first prove Lemma 4.3.4. We consider the five cases separately.

Case 1: For $w = E + \mathrm{i}\eta \in \mathbf{D}_k^b(\zeta, \tau', N)$, we have

$$m_{1c}(w) = \int_{\mathbb{R}} \frac{\rho_{1c}(x)}{x - (E + \mathrm{i}\eta)} dx, \quad \operatorname{Im} m_{1c}(w) = \int_{\mathbb{R}} \frac{\rho_{1c}(x, z)\eta}{(x - E)^2 + \eta^2} dx. \tag{A.2.1}$$

By the regularity condition of Definition 4.2.4 (ii), we get immediately $\operatorname{Im} m_{1c} \sim 1$. Since $\operatorname{Im} m_{1c} \leqslant |1 + m_{1c}| \leqslant C$ by Proposition 4.2.19, we get $|1 + m_{1c}| \sim 1$. Notice wm_{1c} can be expressed as

$$wm_{1c}(w) = \int_{\mathbb{R}} \frac{w\rho_{1c}(x, z)}{x - w} dx = -\int_{\mathbb{R}} \rho_{1c}(x, z) dx + \int_{\mathbb{R}} \frac{x\rho_c(x, z)}{x - w} dx.$$

By the same argument as above and using the fact that $x \geqslant \tau'$ for $x \in [e_{2k} + \tau', e_{2k-1} - \tau']$, we get

$$\operatorname{Im}(wm_{1c}) = \operatorname{Im} \int_{\mathbb{R}} \frac{x\rho_{1c}(x, z)}{x - w} dx \sim 1.$$

Since the imaginary parts of $-w$ and $|z|^2/(1 + m_{1c})$ are both negative, we get

$$\operatorname{Im}\left[-w(1 + m_{1c}) + \frac{|z|^2}{1 + m_{1c}}\right] \leqslant -\operatorname{Im}(wm_{1c}). \tag{A.2.2}$$

Using the bounds for m_{1c} and $\operatorname{Im} m_{1c}$ proved above, it is easy to see that

$$\left|-w(1+m_{1c})+\frac{|z|^2}{1+m_{1c}}\right| = O(1). \tag{A.2.3}$$

Equations (A.2.2) and (A.2.3) together give that $\operatorname{Im} m_{2c} \sim 1$ and $|m_{2c}| \sim 1$ by (4.2.9). Similarly, we can also prove that

$$wm_{2c} = \left[-(1+m_{1c})+\frac{|z|^2}{w(1+m_{1c})}\right]^{-1} \in \mathbb{C}_+$$

and $\operatorname{Im}(wm_{2c}) \sim 1$. Then (4.3.25) follows from the bound

$$\operatorname{Im}\left(w + s_i w m_{2c} - \frac{|z|^2}{1+m_{1c}}\right) \geqslant s_i \operatorname{Im}(wm_{2c}).$$

Case 2: For $w = E + \mathrm{i}\eta \in \mathbf{D}^o(\zeta, \tau', N)$, using (A.2.1) and $\operatorname{dist}(E, \operatorname{supp} \rho_{1,2c}) \geqslant \tau'$, we immediately get $\operatorname{Im} m_{1,2c} \sim \eta$. Now we prove the other estimates.

We first prove (4.3.25). If $\eta \sim 1$, the proof is the same as in Case 1. Hence we assume $\eta \leqslant c'$, where $c' \equiv c'(\tau, \tau') > 0$ is sufficiently small. Recall the definitions of $\mathbf{D}$ and $\mathbf{D}^o$ in (4.2.38) and (4.2.41), we always have $E \sim 1$ in this case.

We shall prove that

$$\min_i \{|m_c(w) - a_i(w)|, |m_c(w) - b_i(w)|, |m_c(w) + c_i(w)|\} \geqslant \varepsilon', \tag{A.2.4}$$

for some constant ε'. This leads immediately to (4.3.25) since

$$\left|w\left(1 + s_i \frac{1+m_{1c}}{-w(1+m_{1c})^2+|z|^2}\right)(1+m_{1c}) - |z|^2\right| = \left|\frac{\sqrt{w}(m_c-a_i)(m_c-b_i)(m_c+c_i)}{-m_c^2+|z|^2}\right|. \tag{A.2.5}$$

For $p_i = \sqrt{E} m^3 - (s_i + |z|^2)m^2 - \sqrt{E}|z|^2 m + |z|^4$, it is not hard to prove that the roots $a_i(E)$, $b_i(E)$ and $-c_i(E)$ decrease as E increase. Since $E \notin \operatorname{supp} \rho_{1c}$, we have $m_{1c}(E) \in \mathbb{R}$ and

$$\frac{dm_{1c}(E)}{dE} = \int_{\mathbb{R}} \frac{\rho_{1c}(x,z)}{(x-E)^2} dx \geqslant 0.$$

So $m_{1c}(E)$ (and hence $m_c(E)$) increases as E increases. Suppose e_k is the smallest edge that is bigger than E, then for $a_i(E)$ bigger than $m_c(E)$, we have that

$$a_i(E) - m_c(E) \geqslant a_i(e_k) - m_c(e_k) + \varepsilon'(\tau') \geqslant \varepsilon'(\tau'), \tag{A.2.6}$$

by using $|E - e_k| \geqslant \tau'$ (see (4.2.41)). On the other hand, If e_{k-1} is the largest edge value that is smaller than E, then for $a_i(E)$ smaller than $m_c(E)$, we have that

$$m_c(E) - a_i(E) \geqslant m_c(e_{k-1}) - a_i(e_{k-1}) + \varepsilon'(\tau') \geqslant \varepsilon'(\tau'). \tag{A.2.7}$$

Applying the same arguments to $b_i(E)$ and $-c_i(E)$, we get

$$\min_i\{|m_c(E) - a_i(E)|, |m_c(E) - b_i(E)|, |m_c(E) + c_i(E)|\} \geqslant \varepsilon' \tag{A.2.8}$$

for $E \in (e_{2k+1}, e_{2k})$ for some k. Now we are only left with the case $E < e_{2L}$, the rightmost edge, when $|z|^2 \geqslant 1 + \tau$. In this case, we have seen that $0 < m_c(E) < b_i(E)$ for all i in the proof of Lemma A.1.4. Thus we can use (A.2.6) to get lower bounds for $|m_c(E) - a_i(E)|$ and $|m_c(E) - b_i(E)|$. Since $c_i(E) \sim 1$ in this case (by (A.1.4) and using $E, |z| \sim 1$), $|m_c(E) + c_i(E)| \geqslant \varepsilon$ is trivial. Again we get the estimate (A.2.8).

Then we consider $w = E + \mathrm{i}\eta$ with $\eta \leqslant c'$. First, it is easy to check that $a_i(E + \mathrm{i}\eta)$, $b_i(E + \mathrm{i}\eta)$ and $c_i(E + \mathrm{i}\eta)$ are continuous in η. On the other hand for $m_c(E + \mathrm{i}\eta)$, we have

$$|\partial_w m_{1c}(w)| = \left|\int_{\mathbb{R}} \frac{\rho_{1c}(x, z)}{(x - w)^2} dx\right| \leqslant C \tag{A.2.9}$$

by the condition $\operatorname{dist}(E, \operatorname{supp} \rho_{1c}) \geqslant \tau'$. Thus we immediately get $|m_c(E + \mathrm{i}\eta) - m_c(E)| = O(\eta)$. Hence as long as c' is small enough, (A.2.4) still holds true, which further gives (4.3.25).

Now we show that $|1 + m_{1c}| \sim 1$ for $w \in \mathbf{D}^o$ and $\eta \leqslant c'$. In fact, if $|m_c|$ can be arbitrarily small, then by (4.3.25) we get that

$$f(\sqrt{w}, m_c) = -\sqrt{w} + O(m_c) \neq 0,$$

which gives a contradiction. Finally we have $|m_{2c}| \sim 1$ for $w \in \mathbf{D}^o$ and $\eta \leqslant c'$ by Proposition 4.2.19.

Case 3: For a regular edge $e_k \neq 0$, we always have $e_k \geqslant \varepsilon$ for some $\varepsilon > 0$ by Lemma A.1.4. Thus we always have $|w| \sim 1$ for $w = E + \mathrm{i}\eta \in \mathbf{D}_k^e(\zeta, \tau', N)$ as long as τ' is sufficiently small. If $\eta \sim 1$, then $\sqrt{\kappa + \eta} \sim \eta/\sqrt{\kappa + \eta} \sim 1$ and the proof is the same as in Case 1. Now we pick

τ' small and consider the case $\eta \leqslant \tau'$. By the regularity assumption (4.2.18) and Lemma A.1.5, we have

$$\min_{1\leqslant i\leqslant n}\{|m_c(w)-a_i(w)|,|m_c(w)-b_i(w)|,|m_c(w)+c_i(w)|\}\geqslant \varepsilon/2 \tag{A.2.10}$$

uniformly in $w \in \{w \in \mathbf{D}_k^e(\zeta,\tau',N) : \kappa(w)+\eta(w)\leqslant 2\tau'\}$, provided τ' is sufficiently small. The above bound implies (4.3.25). If $m_c(w)\to 0$, then using (4.3.25) we get from $f(\sqrt{w},m_c)=0$ that $-\sqrt{w}+O(m_c)=0$, which gives a contradiction. Thus we must have $|1+m_{1c}|\sim|m_c|\sim 1$. To show $|m_{2c}|\sim 1$, we can use Proposition 4.2.19.

We still need to prove the estimates for $\operatorname{Im} m_{1,2c}$ when $\eta\leqslant\tau'$. Recall the expansion (A.1.48) around e_k and equation (A.1.49), where both $\partial_{\sqrt{w}}f(\sqrt{e_k},m_k)$ and $\partial_m^2 f(\sqrt{e_k},m_k)$ are real (as e_k and m_k are real). Suppose k is odd, then $\operatorname{Im} m_c(E)=0$ for $E\searrow e_k$ (i.e. $E\notin \operatorname{supp}\rho_c$) and $\operatorname{Im} m_c(E)>0$ for $E\nearrow e_k$ (i.e. $E\in\operatorname{supp}\rho_c$). Thus (A.1.49) gives

$$m_c(w)-m_k=C_k(w)(w-e_k)^{1/2}+D_k(w),$$

with $C_k>0$, $C_k\sim 1$, $|D_k|=O(|w-e_k|)$ and $\operatorname{Im} D_k=O(\eta)$. Then for $E\geqslant e_k$, we have

$$\operatorname{Im} m_c(E+\mathrm{i}\eta)\sim \operatorname{Im}(\kappa+\mathrm{i}\eta)^{1/2}+O(\eta)\sim\frac{\eta}{\sqrt{\kappa+\eta}},$$

and for $E\leqslant e_k$, we have

$$\operatorname{Im} m_c(E+\mathrm{i}\eta)\sim \operatorname{Im}(-\kappa+\mathrm{i}\eta)^{1/2}+O(\eta)\sim\sqrt{\kappa+\eta}.$$

If k is even, the proof is the same except that in this case, we have

$$m_c(w)-m_k=C_k(w)(e_k-w)^{1/2}+D_k(w).$$

For $m_{1c}(w)$ and $m_{2c}(w)$, we get the conclusion by noticing $w\approx e_k$ and

$$\operatorname{Im} m_{1c}=\operatorname{Im}\left(w^{-1/2}m_c\right)\sim\operatorname{Im} m_c(w),\quad \operatorname{Im} m_{2c}=\operatorname{Im}\left[\frac{m_c}{\sqrt{w}(-m_c^2+|z|^2)}\right]\sim\operatorname{Im} m_c(w),$$

where we used that $|m_c^2-|z|^2|\sim 1$ as observed in the proof of Proposition 4.2.18 in Section A.1.

Case 4: Again if $\eta\sim 1$, the proof is the same as in Case 1. If $|w|\leqslant 2\tau'$ for small enough τ', in the proof of Lemma A.1.4, we have seen that $m_c=i\sqrt{t}+O(\sqrt{w})$, which gives the first equation in (4.3.22). Plugging it into (4.2.9), we get the second equation in (4.3.22). Taking the imaginary part, we obtain (4.3.23). Finally using (4.3.22), we can verify (4.3.25) easily.

Case 5: For $w = E + \mathrm{i}\eta \in \mathbf{D}_L(\zeta, N)$, the bounds for $m_{1,2}$ and $\operatorname{Im} m_{1,2}$ in (4.3.24) follows from (A.2.1) directly.

Now we have finished the proof of Lemma 4.3.4 by combining the above five cases. Next we prove Lemma 4.3.5.

Proof of Lemma 4.3.5. The estimates (4.3.27) and (4.3.28) follow immediately from (4.2.31), (4.3.25) and (4.3.26). For (4.3.29), we can write

$$\Pi_{\mathbf{vv}} = \left\langle \mathbf{v}, \begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix} \Pi_d \begin{pmatrix} U^* & 0 \\ 0 & U^* \end{pmatrix} \mathbf{v} \right\rangle = (\Pi_d)_{\mathbf{uu}} = \sum_{i=1}^{N} \left\langle u_{[i]}, \pi_{[i]c} u_{[i]} \right\rangle,$$

where

$$\mathbf{u} := \begin{pmatrix} U^* & 0 \\ 0 & U^* \end{pmatrix} \mathbf{v}, \quad u_{[i]} := \begin{pmatrix} u_i \\ u_{\bar{i}} \end{pmatrix}.$$

To control $\operatorname{Im} \Pi_{\mathbf{vv}}$, it is enough to bound $\left\langle u_{[i]}, \pi_{[i]c} u_{[i]} \right\rangle$ for each i.

We first consider Cases 1-4 of Lemma 4.3.4. By the definition of $\pi_{[i]c}$ in (4.2.31), we get

$$\begin{aligned} \operatorname{Im} \pi_{ii,c} &= \operatorname{Im} \left[-w(1 + |d_i|^2 m_{2c}) + \frac{|z|^2}{1 + m_{1c}} \right]^{-1} \leqslant \frac{C}{|w|} \operatorname{Im} \left[w(1 + |d_i|^2 m_{2c}) - \frac{|z|^2}{1 + m_{1c}} \right] \\ &= \frac{C}{|w|} \left[(1 + |d_i|^2 \operatorname{Re} m_{2c}) \operatorname{Im} w + |d_i|^2 (\operatorname{Re} w) \operatorname{Im} m_{2c} + \frac{|z|^2}{|1 + m_{1c}|^2} \operatorname{Im} m_{1c} \right], \end{aligned}$$

where in the second step we used (4.3.25) and $|1 + m_{1c}| \sim |w|^{-1/2}$. In the first three cases of Lemma 4.3.4, we have $|w| \sim 1$ and $\operatorname{Im} w = O(\operatorname{Im} m_{1c})$, which give that $\operatorname{Im} \pi_{ii,c} \leqslant C\operatorname{Im}(m_{1c} + m_{2c})$. In case 4 of Lemma 4.3.4, we use $|\operatorname{Im} w| + |\operatorname{Re} w| + |1 + m_{1c}|^{-2} = O(|w|)$ and $\operatorname{Im} m_{1,2c} \sim |w|^{-1/2}$ to get that $\operatorname{Im} \pi_{ii,c} \leqslant C\operatorname{Im}(m_{1c} + m_{2c})$. Similarly, we can get the bound $\operatorname{Im} \pi_{\bar{i}\bar{i},c} \leqslant C\operatorname{Im}(m_{1c} + m_{2c})$. Finally we can estimate the following term using similar methods,

$$\begin{aligned} \operatorname{Im} \left(\bar{u}_{\bar{i}} u_i \pi_{\bar{i}i,c} + \bar{u}_i u_{\bar{i}} \pi_{i\bar{i},c} \right) &= 2\operatorname{Re} \left(\bar{u}_i u_{\bar{i}} z \right) \operatorname{Im} \left\{ w^{-1/2} \left[w(1 + |d_i|^2 m_{2c})(1 + m_{1c}) - |z|^2 \right]^{-1} \right\} \\ &\leqslant C\operatorname{Re} \left(\bar{u}_i u_{\bar{i}} z \right) \operatorname{Im}(m_{1c} + m_{2c}) \leqslant C \left(|u_i|^2 + |u_{\bar{i}}|^2 \right) \operatorname{Im}(m_{1c} + m_{2c}). \end{aligned}$$

Combining the above estimates we get $\operatorname{Im}\left\langle u_{[i]}, \pi_{[i]^c} u_{[i]}\right\rangle \leqslant C|u_{[i]}|^2 \operatorname{Im}(m_{1c}+m_{2c})$, which implies (4.3.29). For the Case 5 of Lemma 4.3.4, we use (4.3.24) and (4.3.28) to get

$$\operatorname{Im}\left\langle u_{[i]}, \pi_{[i]^c} u_{[i]}\right\rangle \leqslant |u_{[i]}|^2 \|\pi_{[i]^c}\| \leqslant C|u_{[i]}|^2 \operatorname{Im}(m_{1c}+m_{2c}).$$

This completes the proof. □

A.3 Proof of Lemma 4.3.7 and Lemma 4.2.2

We first prove Lemma 4.3.7. During the proof, we also use the following equivalent definition of the stability expressed in terms of $m = \sqrt{w}(1+m_1)$, $u = \sqrt{w}(1+u_1)$ and $f(\sqrt{w}, m)$. Suppose the assumptions in Definition 4.3.6 holds. Let $w \in \mathbf{D}$ and suppose that for all $w' \in L(w)$ we have $|f(\sqrt{w}, u)| \leqslant |w|^{1/2}\delta(w)$. Then

$$|u(w) - m_c(w)| \leqslant \frac{C|w|^{1/2}\delta}{\sqrt{\kappa+\eta+\delta}}. \tag{A.3.1}$$

Case 1: We take over the notations in Definition 4.3.6 and abbreviate $R := f(\sqrt{w}, u)$, so that $|R| \leqslant |w|^{1/2}\delta$. Then we write the equation $f(\sqrt{w}, u) - f(\sqrt{w}, m_c) = R$ as

$$\alpha(u-m_c)^2 + \beta(u-m_c) = R, \tag{A.3.2}$$

where using (A.1.1), α and β can be expressed as

$$\alpha := \frac{1}{N}\sum_{i=1}^n l_i s_i \left[\frac{A_i}{(u-a_i)(m_c-a_i)^2} + \frac{B_i}{(u-b_i)(m_c-b_i)^2} + \frac{C_i}{(u+c_i)(m_c+c_i)^2}\right], \tag{A.3.3}$$

and

$$\beta := 1 - \frac{1}{N}\sum_{i=1}^n l_i s_i \left[\frac{A_i}{(m_c-a_i)^2} + \frac{B_i}{(m_c-b_i)^2} + \frac{C_i}{(m_c+c_i)^2}\right] = \partial_m f(\sqrt{w}, m_c). \tag{A.3.4}$$

We shall prove that

$$|\alpha| + |\partial_u \alpha| \leqslant C, \ \ |\beta| \sim 1, \tag{A.3.5}$$

for $w \in \mathbf{D}_k^b$ and u satisfying $|u - m_c| \leqslant (\log N)^{-1/3}$. If $|u - m_c| \leqslant (\log N)^{-1/3}$, we also have $\operatorname{Im} u \sim 1$. By (4.3.25), we have

$$\min_i\{|m_c - a_i|, |m_c - b_i|, |m_c + c_i|\} \geqslant \varepsilon \tag{A.3.6}$$

for some $\varepsilon > 0$. Replacing the m_c in (4.3.25) with u, we also get that

$$\min_i\{|u-a_i|, |u-b_i|, |u+c_i|\} \geqslant \varepsilon' \tag{A.3.7}$$

for some $\varepsilon' > 0$. Using (A.3.6) and (A.3.7), we get immediately that $|\alpha| + |\partial_u \alpha| + |\beta| \leqslant C$. What remains is the proof of the lower bound $|\beta| \geqslant c$. If $\operatorname{Im} w \geqslant \varepsilon$ for some constant $\varepsilon > 0$, the lower bound follows from Lemma A.3.1 below. If $\operatorname{Im} w \leqslant \varepsilon$ for a sufficiently small $\varepsilon > 0$, the lower bound follows from Lemma A.3.2 below. Now given the estimate (A.3.5), it is easy to prove (A.3.1) with a fixed point argument. This proves the stability of (4.3.30).

Lemma A.3.1. *Suppose that* $\operatorname{Im} w \sim 1$ *and* $|m_c| \sim \operatorname{Im} m_c \sim 1$. *Then* $|\partial_m f(\sqrt{w}, m_c)| \geqslant c$ *for some constant* $c > 0$.

Proof. Using (4.2.13), $m_c = \sqrt{w}(1+m_{1c})$ and the conditions $\operatorname{Im} w \sim 1$, $\operatorname{Im} m_c \sim 1$, we can get that

$$\left|\frac{\partial_{\sqrt{w}} f(\sqrt{w}, m_c)}{\partial_m f(\sqrt{w}, m_c)}\right| = \left|\frac{\partial m_c}{\partial \sqrt{w}}\right| \leqslant C \Rightarrow \left|\partial_{\sqrt{w}} f(\sqrt{w}, m_c)\right| \leqslant C\left|\partial_m f(\sqrt{w}, m_c)\right|, \tag{A.3.8}$$

for some constant $C > 0$. Now we assume that $|\partial_m f(\sqrt{w}, m_c)|$ can be arbitrarily small. Then $\left|\partial_{\sqrt{w}} f(\sqrt{w}, m_c)\right|$ can also be arbitrarily small. Denote $a := \partial_m f(\sqrt{w}, m_c)$ and $b := \partial_{\sqrt{w}} f(\sqrt{w}, m_c)$. Using (A.1.23) and (A.1.31), we get that

$$a = \frac{\sqrt{w}}{m_c} - \frac{m_c}{N}\sum_{i=1}^n l_i s_i \frac{\sqrt{w}\,(m_c^2 - |z|^2)^2 + 2s_i|z|^2 m_c}{\left[-(s_i+|z|^2)m_c^2 + |z|^4 + \sqrt{w}\,(m_c^3 - |z|^2 m_c)\right]^2} \tag{A.3.9}$$

and

$$b = -1 - \frac{m_c^2}{N}\sum_{i=1}^n l_i s_i \frac{(m_c^2 - |z|^2)^2}{\left[-(s_i+|z|^2)m_c^2 + |z|^4 + \sqrt{w}\,(m_c^3 - |z|^2 m)\right]^2}. \tag{A.3.10}$$

Using (A.3.9) and (A.3.10), we can get that

$$\frac{(\sqrt{w}m_c - |z|^2)|z|^2}{m_c} b - \frac{1}{2}(m_c^2 - |z|^2)(m_c a - \sqrt{w} b) = \frac{(|z|^2 - \sqrt{w}m_c)(m_c^2 + |z|^2)}{m_c}, \tag{A.3.11}$$

where we used the equation $f(\sqrt{w}, m_c) = 0$ in the derivation. By our assumption, the left-hand side of (A.3.11) can be arbitrarily small. For the right-hand side of (A.3.11), we have $|m_c| \sim 1$ and $|\sqrt{w}m_c - |z|^2| \sim 1$ (since $\operatorname{Im}(\sqrt{w}m_c) = \operatorname{Im}(w + wm_{1c}) \sim 1$). Then if $|m_c - i|z|| \geqslant c'$ for some constant $c' > 0$, we have $|m^2 + |z|^2| \sim 1$, and hence

$$\left|\frac{(\sqrt{w}m_c - |z|^2)|z|^2}{m_c} b - \frac{1}{2}(m_c^2 - |z|^2)(m_c a - \sqrt{w} b)\right| \sim 1,$$

which gives a contradiction. Thus we must have a lower bound $|\partial_m f(\sqrt{w}, m_c)| \geqslant c$ if $|m - i|z|| \geqslant c'$.

We still need to deal with the case with $|m_c - i|z|| \leqslant c'$ for some sufficiently small c'. Notice $|z| \sim 1$ in this case. It is easy to calculate that

$$\frac{\partial f}{\partial \sqrt{w}}(\sqrt{w}, i|z|) = -1 + \frac{|z|^2}{N} \sum_{k=1}^{n} l_k s_k \frac{4|z|^4}{\left[(s_k + |z|^2)|z|^2 + |z|^4 - 2i\sqrt{w}|z|^3\right]^2}. \tag{A.3.12}$$

Denote $L_k := (s_k + |z|^2)|z|^2 + |z|^4 - 2i\sqrt{w}|z|^3$. Since $i\sqrt{w} = i(x + iy) = ix - y$ for some $x, y > 0$ and $x, y \sim 1$, we have $\operatorname{Re} L_k > 0$, $\operatorname{Im} L_k < 0$ and $|\operatorname{Re} L_k|, |\operatorname{Im} L_k| \sim 1$. In particular, this gives that $\operatorname{Im} L_k^2 < 0$ and $|\operatorname{Im} L_k^2| \sim 1$. Thus each fraction $4|z|^4/L_k^2$ in (A.3.12) has positive imaginary part of order 1. Therefore

$$\left|\frac{\partial f}{\partial \sqrt{w}}(\sqrt{w}, i|z|)\right| \geqslant \operatorname{Im}\left[\frac{\partial f}{\partial \sqrt{w}}(\sqrt{w}, i|z|)\right] \sim 1.$$

Then by (A.3.8), we get that $|\partial_m f(\sqrt{w}, i|z|)| \geqslant c$ for some $c > 0$. Using (4.3.25), it is easy to see that

$$\partial_m f(\sqrt{w}, m_c) = \partial_m f(\sqrt{w}, i|z|) + O(|m_c - i|z||).$$

Thus in the case $|m_c - i|z|| \leqslant c'$, we still have $|\partial_m f(\sqrt{w}, m_c)| \geqslant c/2$, provided that c' is sufficiently small. □

Lemma A.3.2. *Suppose that $w \in \mathbf{D}_k^b$ and $\operatorname{Im} w \leqslant \varepsilon$. Then for sufficiently small $\varepsilon > 0$, we have $|\partial_m f(\sqrt{w}, m_c)| \sim 1$.*

Proof. By (4.3.18) and (4.3.25), we have $\partial_{\sqrt{w}} \partial_m f(w, m_c) = O(1)$ and $\partial_m^2 f(w, m_c) = O(1)$. Denote $w = E + \mathrm{i}\eta$. Taking the imaginary part of the following equation

$$0 = f(\sqrt{E}, m_c(E)) = -\sqrt{E} + m_c + E^{-1/2} + \frac{1}{N} \sum_{i=1}^{n} l_i s_i \left(\frac{A_i}{m_c - a_i} + \frac{B_i}{m_c - b_i} + \frac{C_i}{m_c + c_i}\right), \tag{A.3.13}$$

and noticing that A_i, B_i, C_i and a_i, b_i, c_i are all positive real numbers for real E, we get

$$\frac{1}{N} \sum_{i=1}^{n} l_i s_i \left(\frac{A_i}{|m_c - a_i|^2} + \frac{B_i}{|m_c - b_i|^2} + \frac{C_i}{|m_c + c_i|^2}\right) = 1. \tag{A.3.14}$$

Using the above equation, we get

$$\partial_m f(\sqrt{E}, m_c(E)) = 1 - \frac{1}{N}\sum_{i=1}^n l_i s_i \left[\frac{A_i}{(m_c - a_i)^2} + \frac{B_i}{(m_c - b_i)^2} + \frac{C_i}{(m_c + c_i)^2}\right]$$
$$= \frac{1}{N}\sum_{i=1}^n l_i s_i \left[\frac{A_i}{|m_c - a_i|^2} - \frac{A_i}{(m_c - a_i)^2} + \frac{B_i}{|m_c - b_i|^2} - \frac{B_i}{(m_c - b_i)^2} + \frac{C_i}{|m_c + c_i|^2} - \frac{C_i}{(m_c + c_i)^2}\right]. \tag{A.3.15}$$

We look at, for example, the term

$$\frac{A_i}{|m_c - a_i|^2} - \frac{A_i}{(m_c - a_i)^2} = \frac{A_i}{|m_c - a_i|^2}(1 - e^{-2i\theta_i}),$$

where $m_c - a_i := |m_c - a_i|e^{i\theta_i}$. Using $\operatorname{Im} m_c \sim 1$, it is easy to see that $\operatorname{Re}(1 - e^{-2i\theta_i}) \geqslant c'$ for some constant $c' > 0$. Applying the same estimates to the B, C terms in (A.3.15), we get

$$\left|\partial_m f(\sqrt{E}, m_c(E))\right| \geqslant \operatorname{Re}\left[\partial_m f(\sqrt{E}, m_c(E))\right] \geqslant c \tag{A.3.16}$$

for some constant $c > 0$.

Now for $w = E + \mathrm{i}\eta$ with $\eta \leqslant \varepsilon$, we can expand $\partial_m f(\sqrt{w}, m_c(w))$ around $\partial_m f(\sqrt{E}, m_c(E))$:

$$\partial_m f(\sqrt{w}, m_c(w)) = \partial_m f(E, m_c(E)) + O(\eta),$$

where we used (4.3.25). Combing with (A.3.16), we get $|\partial_m f(w, m_c(w))| \sim 1$ for small enough ε. □

Case 2: We mimic the argument in the proof of Case 1. We see that it suffices to prove $|\alpha| + |\partial_u \alpha| \leqslant C$ and $|\beta| \sim 1$ for α, β defined in (A.3.3) and (A.3.4) and $|u - m_c| \leqslant (\log N)^{-1/3}$. Using (4.3.25), it is not hard to prove that $|\alpha| + |\partial_u \alpha| + |\beta| \leqslant C$. What remains is the proof of the lower bound $|\beta| \geqslant c$. For the $\operatorname{Im} w \sim 1$ case, the bound follows from Lemma A.3.1. We are left with the case where $E = \operatorname{Re} w \sim 1$ and $\eta = \operatorname{Im} w \to 0$. Using (4.2.13), $m_c = \sqrt{w}(1 + m_{1c})$, $|w| \sim 1$ and $\operatorname{dist}(E, \operatorname{supp} \rho_{1c}) \geqslant \tau'$, we can get that

$$\left|\frac{\partial_{\sqrt{w}} f(\sqrt{w}, m_c)}{\partial_m f(\sqrt{w}, m_c)}\right| = \left|\frac{\partial m_c}{\partial \sqrt{w}}\right| \leqslant C$$

for some constant $C > 0$. Thus it suffices to prove that $\left|\partial_{\sqrt{w}} f(\sqrt{w}, m_c)\right|$ has a lower bound. Using (A.1.31) and noticing that $m_c(E) \in \mathbb{R}$, we get

$$\partial_{\sqrt{w}} f(\sqrt{E}, m_c(E)) = -1 - \frac{m_c^2}{N}\sum_{i=1}^n l_i s_i \frac{(m_c^2 - |z|^2)^2}{\left[-(s_i + |z|^2)m_c^2 + |z|^4 + \sqrt{E}\,(m_c^3 - |z|^2 m_c)\right]^2} \leqslant -1.$$

Expanding $\partial_{\sqrt{w}} f(\sqrt{w}, m_c(w))$ around $\partial_{\sqrt{w}} f(\sqrt{E}, m_c(E))$, using (4.3.25) and $|m_c(E+\mathrm{i}\eta) - m_c(E)| \sim \eta$, we get for η small

$$\left|\partial_{\sqrt{w}} f(\sqrt{w}, m_c)\right| \geqslant 1 + O(\eta) \geqslant c.$$

This concludes the proof for *Case 2*.

Case 3: The case $\operatorname{Im} w \geqslant \tau'$ can be proved with the same method as in the proof of case 1. Hence we only consider the case $|w - e_k| \leqslant 2\tau'$ in the following. Note that $|w| \sim 1$ in this case. Suppose

$$|w - e_k| \leqslant 2\tau', \quad |u - m_c| \leqslant (\log N)^{-1/3}. \tag{A.3.17}$$

Then we claim that

$$|\alpha| \sim 1, \quad |\beta| \sim \sqrt{\kappa + \eta} \tag{A.3.18}$$

for small enough τ'. Using (A.3.17), (4.3.25), (4.2.19) and Lemma A.1.5, we can get that

$$\alpha = \frac{1}{2}\partial_m^2 f(\sqrt{e_k}, m_c(e_k)) + O(|w - e_k|^{1/2} + (\log N)^{-1/3}) \sim 1.$$

To prove the estimate for β, we use (4.2.17), (4.3.25) and Lemma A.1.5 to get that

$$\begin{aligned}
\beta &= \int_{e_k}^{w} \frac{d}{dw'} \partial_m f(\sqrt{w'}, m_c(w')) dw' \\
&= \int_{e_k}^{w} \frac{\partial_{\sqrt{w'}}\partial_m f(\sqrt{w'}, m_c(w'))}{2\sqrt{w'}} dw' + \int_{e_k}^{w} \partial_m^2 f(\sqrt{w'}, m_c(w')) \frac{dm_c(w')}{dw'} dw' \\
&= \int_{e_k}^{w} \frac{\partial_{\sqrt{w}}\partial_m f(\sqrt{e_k}, m_c(e_k)) + O(|w-e_k|^{1/2})}{2\sqrt{w'}} dw' + \int_{m_c(e_k)}^{m_c(w)} \left[\partial_m^2 f(\sqrt{e_k}, m_c(e_k)) + O(|w-e_k|^{1/2})\right] dm \\
&= \partial_m^2 f(\sqrt{e_k}, m_k)(m_c(w) - m_c(e_k)) + O(|w - e_k|).
\end{aligned} \tag{A.3.19}$$

Thus we conclude for small enough τ' that

$$|\beta| \sim |w - e_k|^{1/2} \sim \sqrt{\kappa + \eta}.$$

With the estimate (A.3.18), we now proceed as in the proof of [14, Lemma 4.5], by solving the quadratic equation (A.3.2) for $u - m_c$ explicitly. We select the correct solution by a continuity argument using that (A.3.1) holds by assumption at $z + iN^{-10}$. The second assumption of (A.3.17) is obtained by continuity from the estimate on $|u - m_c|$ at the neighboring point $z+iN^{-10}$. We refer to [14, Lemma 4.5] for the full details. This concludes the proof for *Case 3*.

Case 4: The case when $\operatorname{Im} w \geqslant \tau'$ can be proved using the same method as in the proof of Case 1. Now we are left with the case $|w| \leqslant 2\tau'$ for some sufficiently small τ'. First we assume $|z| \geqslant c > 0$ for some small $c > 0$. Then mimicking the argument in the proof of Case 1, we see that it suffices to prove $|\alpha| + |\partial_u \alpha| \leqslant C$ and $|\beta| \sim 1$ when $|u - m_c| \leqslant (\log N)^{-1/3}$. Using (4.3.25), it is not hard to prove that $|\alpha| + |\partial_u \alpha| + |\beta| \leqslant C$. The lower bound $|\beta| \geqslant c$ can be obtained easily from (A.1.32).

Then suppose $|z|^2 < c$, but $|w|^{1/2} + |z|^2 \geqslant \varepsilon$. According to (A.1.38) and using that $|i|z|^2 + \sqrt{wt_0}| \sim |w|^{1/2} + |z|^2$, we can verify that

$$\beta = \partial_m f(\sqrt{w}, m_c(w)) \sim |w|^{1/2} + |z|^2 \sim 1.$$

With (4.3.25), it is easy to check that

$$\partial_m^2 f(\sqrt{w}, \xi) = O(1), \quad \partial_m^3 f(\sqrt{w}, \xi) = O(1),$$

for $|\xi - m_c| \leqslant (\log N)^{-1/3}$, from which we get that $|\alpha| + |\partial_u \alpha| = O(1)$. With a fixed point argument, we conclude (A.3.1).

Case 5: Again we following the arguments in the proof of Case 1. However, instead of $f(\sqrt{w}, m)$, we shall study $\Upsilon(w, m_1)$ in (4.3.31) directly. We take over the notations in Definition 4.3.6 and abbreviate $R := \Upsilon(w, u_1)$, so that $|R| \leqslant \delta$. Then we write the equation $\Upsilon(w, u_1) - \Upsilon(w, m_{1c}) = R$ as

$$\alpha(u_1)(u_1 - m_{1c})^2 + \beta(u_1 - m_{1c}) = R, \tag{A.3.20}$$

where we used the same symbols as in (A.3.2) for notational convenience. As in *Case 1*, we have $\beta = \partial_{m_1} \Upsilon(w, m_{1c})$, and we can estimate that $|\alpha| + |\partial_{u_1} \alpha| \leqslant C$ for $w \in \mathbf{D}_L$ and u_1 satisfying $|u_1 - m_{1c}| \ll |m_{1c}|$. Now to conclude (4.3.35), it suffices to prove $|\beta| \sim 1$ for $w \in \mathbf{D}_L$. In fact with (4.3.31), we can obtain that

$$\beta = 1 + O\left(\eta^{-1}\right) \sim 1,$$

for $\eta \geqslant \zeta^{-1}$. This concludes the proof of Lemma 4.3.7.

Proof of Lemma 4.2.2. The fact that ρ_{1c} has compact support follows from Lemma 4.2.3; ρ_{1c} being integrable follows from Lemma A.1.4. Note that in proving Lemmas 4.2.3 and A.1.4, we do not make use of the regularity assumptions in Definition 4.2.4. It remains to show that for fixed $w \in \mathbb{C}_+$ and $|z| \neq 1$, there exists a unique $m_{1c}(w) \in \mathbb{C}_+$ satisfying equation (4.2.11). This follows from the proof of *Case 1* in this section under the extra condition $\eta \sim 1$. Again, we do not need the regularity assumptions for the proof, because η^{-1} provides a nice bound for the Stieltjes transforms in the global region with $\eta \sim 1$. □

Remark A.3.3. The estimate (4.3.25) has been used repeatedly during the proof of Lemma 4.3.7. Here we remark that it also gives the stability of the regularity conditions in Definition 4.2.4 under perturbations of $|z|$ and ρ_Σ. For example, we define the shifted empirical spectral density

$$\rho_{\Sigma,t} := \frac{1}{N \wedge M} \sum_{i=1}^{N \wedge M} \delta_{\sigma_i + t}, \tag{A.3.21}$$

and the associated $m_c(w,t)$ and function $f(\sqrt{w}, m, t)$. Given a regular edge e_k, we have that

$$f(\sqrt{e_k}, m_k, t=0) = 0, \quad \partial_m f(\sqrt{e_k}, m_k, t=0) = 0,$$

where we denote $m_k := m_c(e_k)$. We have the Jacobian

$$J := \det \begin{pmatrix} \partial_{\sqrt{w}} f & \partial_m f \\ \partial_{\sqrt{w}} \partial_m f & \partial_m^2 f \end{pmatrix}_{(\sqrt{w},m,t)=(\sqrt{e_k},m_k,0)} = \partial_{\sqrt{w}} f(\sqrt{e_k}, m_k, 0) \partial_m^2 f(\sqrt{e_k}, m_k, 0).$$

By (A.1.31), we have $\left|\partial_{\sqrt{w}} f(\sqrt{e_k}, m_k, 0)\right| \gtrsim 1$. Combining with (4.2.19), we get $|J| \gtrsim \varepsilon$. Using (4.3.25), we can verify that $\partial_t f(\sqrt{e_k}, m_k, 0) = O(1)$ and $\partial_t \partial_m f(\sqrt{e_k}, m_k, 0) = O(1)$. Thus if we regard e_k and m_k as functions of t, then $\partial_t m_k(t=0) = O(1)$ and $\partial_t e_k(t=0) = O(1)$ by the implicit function theorem. Then it is easy to verify

$$\partial_m^2 f\left(\sqrt{e_k(t)}, m_c(e_k,t)\right) = \partial_m^2 f\left(\sqrt{e_k}, m_c(e_k)\right) + O(t),$$
$$|m_c(e_k,t) - a_i(e_k,t)| = |m_c(e_k) - a_i(e_k)| + O(t),$$

and similar estimates for $|m_c - b_i|$ and $|m_c + c_i|$. Thus if Definition 4.2.4 (i) holds for some ρ_Σ, then it holds for all $\rho_{\Sigma,t}$ provided that t is small enough.

Now given a regular bulk component $[e_{2k}, e_{2k-1}]$ and $E \in [e_{2k} + \tau', e_{2k-1} - \tau']$. Differentiating the equation $f(\sqrt{E}, m_c(E,t), t) = 0$ in t yields

$$\partial_t m_c(E,t) = -\frac{\partial_t f(\sqrt{E}, m_c(E,t), t)}{\partial_m f(\sqrt{E}, m_c(E,t), t)}.$$

By (4.3.25), we find that $\partial_t f(\sqrt{E}, m_c(E), 0) = O(1)$, while by (A.3.5), $|\partial_m f(\sqrt{E}, m_c(E), 0)| = \beta \sim 1$. Thus $\partial_t m_c(E, 0) = O(1)$. A simple extension of this argument shows that $m_c(E,t) = m_c(E) + O(t)$ and hence $\operatorname{Im} m_c(E,t)$ is bounded from below by some $c' = c'(\tau, \tau')$. Thus we conclude that if Definition 4.2.4 (ii) holds for some ρ_Σ, then it holds for all $\rho_{\Sigma,t}$ with t in some fixed small interval around zero. Obviously, the above arguments also work for $|z|$ perturbation.

APPENDIX B

The smallest singular value

One of the main purposes of this appendix is to prove a lower tail estimate for the smallest singular value of $TX - z$, which will be used in the proof of Theorem 4.2.6. However, we shall prove an estimate on the smallest singular value of a more general type of deformed random matrices (see Theorem B.1.1), because of the importance of the problem and its possible applications in many other problems in mathematics and statistics. This appendix is relatively independent of the other parts of this thesis, and can be read separately.

B.1 Introduction

Smallest singular values of random matrices

Consider an $N \times n$ real or complex matrix A. The singular values $s_i(A)$ of A are the eigenvalues of $(A^*A)^{1/2}$ arranged in the non-increasing order:

$$s_1(A) \geqslant s_2(A) \geqslant \ldots \geqslant s_n(A).$$

Of particular importance are the largest singular value $s_1(A)$, which gives the spectral norm $\|A\|$, and the smallest singular value $s_n(A)$, which measures the invertibility of A^*A in the $N \geqslant n$ case.

A natural random matrix model is given by a rectangular matrix X whose entries are independent random variables with mean zero, unit variance and certain moment assumptions. In this appendix, we focus on random variables with *arbitrarily high moments* (see (B.1.6)), which include all the *subgaussian* and *subexponential* random variables. The asymptotic behavior of the extreme singular values of X has been well-studied. Suppose X has dimensions

$N \times n$. If $n/N \to \lambda \in (0,1)$ as $N \to \infty$, then the ESD μ_N of $N^{-1}X^*X$ converges weakly to the famous Marchěnko-Pastur (MP) law [66]. Moreover, the MP distribution has a density with positive support on $[(1-\sqrt{\lambda})^2, (1+\sqrt{\lambda})^2]$, which suggests that asymptotically,

$$s_1(X) \to \sqrt{N}(1+\sqrt{\lambda}) = \sqrt{N} + \sqrt{n}, \quad \text{and} \quad s_n(X) \to \sqrt{N}(1-\sqrt{\lambda}) = \sqrt{N} - \sqrt{n}. \quad \text{(B.1.1)}$$

The almost sure convergence of the largest singular value was proved in [45] for random matrices whose entries have arbitrarily high moments. The almost sure convergence of the smallest singular value was proved in [85] for Gaussian random matrices (i.e. the Wishart matrix). These results were later generalized to random matrices with *i.i.d.* entries with finite fourth moment in [112] and [9].

A considerably harder problem is to establish non-asymptotic versions of (B.1.1), which would hold for any fixed dimensions N and n. Most often needed are upper bounds for the largest singular value $s_1(X)$ and lower bounds for the smallest singular value $s_n(X)$. With a standard ε-net argument, it is not hard to prove that $\|X\|$ is at most of the optimal order $\sqrt{N}$ for all dimensions, see e.g. [23, 64, 82]. On the other hand, the smallest singular value is much harder to bound below. There has been much progress in this direction during the last decade.

Tall matrices. It was proved in [64] that for arbitrary aspect ratios $\lambda < 1 - c/\log N$ and for random matrices with independent subgaussian entries, one has

$$\mathbb{P}\left(s_n(X) \leqslant c_\lambda \sqrt{N}\right) \leqslant e^{-cN}, \quad \text{(B.1.2)}$$

where $c_\lambda > 0$ depends only on λ and the maximal subgaussian moment of the entries.

Square matrices. For square random matrices with $N = n$, a lower bound for the smallest singular value was first obtained in [80], where it was proved that for subgaussian random matrix X, $s_N(X) \geqslant \varepsilon N^{-3/2}$ with high probability. This result was later improved in [81] to

$$\mathbb{P}\left(s_N(X) \leqslant \varepsilon N^{-1/2}\right) \leqslant C\varepsilon + e^{-cN}, \quad \text{(B.1.3)}$$

an essentially optimal estimate for subgaussian matrices. Subsequently, different lower bounds for $s_N(X)$ were proved under weakened moments assumptions [49, 74, 91].

Almost square matrices. The gap $1-c/\log N \leqslant \lambda < 1$ was filled in [82]. It was shown that for subgaussian random rectangular matrices,

$$\mathbb{P}\left(s_n(X) \leqslant \varepsilon(\sqrt{N}-\sqrt{n-1})\right) \leqslant (C\varepsilon)^{N-n+1} + e^{-cN}, \tag{B.1.4}$$

for all fixed dimensions $N \geqslant n$. This bound is essentially optimal for subgaussian matrices with all aspect ratios. It is easy to see that (B.1.2) and (B.1.3) are the special cases of the estimate (B.1.4).

In this appendix, we are interested in the extreme singular values of a multiplicatively and additively deformed random rectangular matrix. Given an $M \times n$ random matrix X with independent entries, we consider the matrix $TX - A$, where T and A are $N \times M$ and $N \times n$ deterministic matrices, respectively. It is easy to bound above the largest singular value using $\|TX - A\| \leqslant \|T\|\|X\| + \|A\|$. On the other hand, we expect that if $n \leqslant N \leqslant M$ and the singular values of T satisfy $c \leqslant s_N(T) \leqslant s_1(T) \leqslant c^{-1}$, then a similar estimate as in (B.1.4) would still hold for $TX - A$. In fact, if $M = N$ and X is subgaussian, one can prove that the estimate (B.1.4) holds for the matrix $X - T^{-1}A$ with a direct generalization of the method in [82]. Together with $s_n(TX - A) \geqslant s_N(T)s_n(X - T^{-1}A)$, this already gives the desired lower bound for $s_n(TX - A)$. In this appendix, we will consider more general case where $N \leqslant M$ and X is not necessarily subgaussian, see Theorem B.1.1.

Main result

Let $\xi_1, \ldots, \xi_n$ be independent random variables such that for $1 \leqslant i \leqslant n$,

$$\mathbb{E}\xi_i = 0, \quad \mathbb{E}|\xi_i|^2 = 1, \tag{B.1.5}$$

and for any $p \in \mathbb{N}$, there is an N-independent constant σ_p such that

$$\mathbb{E}|\xi_i|^p \leqslant \sigma_p. \tag{B.1.6}$$

We assume that X is an $M \times n$ random matrix, whose rows are independent copies of the random vector $(\xi_1, \ldots, \xi_n)$. In this appendix, we consider the deformed random rectangular matrix $TX - B$, where T and B are $N \times M$ and $N \times n$ deterministic matrices, respectively.

We assume that

$$n \leqslant N \leqslant M \leqslant \Lambda N, \quad \|B\| \leqslant K_0\sqrt{N} \tag{B.1.7}$$

for some constants $K_0, \Lambda \geqslant 1$. Moreover, we assume the eigenvalues of TT^* satisfy that

$$K_0^{-1} \leqslant \sigma_N \leqslant \ldots \leqslant \sigma_2 \leqslant \sigma_1 \leqslant K_0. \tag{B.1.8}$$

For definiteness, in this appendix we focus on the case with *real* matrices. However, our results and proof also hold, after minor changes, in the *complex* case if we assume in addition that X_{ij} have independent real and imaginary parts, such that

$$\mathbb{E}\left(\operatorname{Re} X_{ij}\right) = 0, \quad \mathbb{E}\left(\operatorname{Re} X_{ij}\right)^2 = \frac{1}{2},$$

and similarly for $\operatorname{Im} X_{ij}$. The main result of this paper is the following theorem.

Theorem B.1.1. *Suppose the assumptions (B.1.5), (B.1.6), (B.1.7) and (B.1.8) hold. Fix any constants $\tau > 0$ and $\Gamma > 0$. Then for every $\varepsilon \geqslant 0$, we have*

$$\mathbb{P}\left(s_n(TX - B) \leqslant \varepsilon N^{-\tau}\left(\sqrt{N} - \sqrt{n-1}\right)\right) \leqslant (C\varepsilon)^{N-n+1} + N^{-\Gamma} \tag{B.1.9}$$

for large enough $N \geqslant N_0$, where the constant $C > 0$ depends only on σ_p, Λ and K_0, and N_0 depends only on σ_p, Λ, Γ and τ.

To prove this theorem, we first truncate the entries of X at level N^ω for some small $\omega > 0$. Combining condition (B.1.6) with Markov's inequality, we get that for any (small) $\omega > 0$ and (large) $\Gamma > 0$, there exists $N(\omega, \Gamma)$ such that

$$\mathbb{P}\left(|\xi_i| > N^\omega/2\right) \leqslant N^{-\Gamma-2}$$

for all $N \geqslant N(\omega, \Gamma)$. Hence with a loss of probability $O(N^{-\Gamma})$, it suffices to control the smallest singular values of the random matrix $T\tilde{X} - B$, where

$$\tilde{X} := 1_\Omega X, \quad \Omega := \left\{|X_{ij}| \leqslant N^\omega/2 \text{ for all } 1 \leqslant i \leqslant M, 1 \leqslant j \leqslant n\right\}.$$

By (B.1.6) and integration by parts, we can check that for $1 \leqslant i \leqslant n$,

$$\mathbb{E}\left(\xi_i \mathbf{1}_{\{|\xi_i| \leqslant N^\omega/2\}}\right) = O\left(N^{-\Gamma-2+\omega}\right), \quad \operatorname{Var}\left(\xi_i \mathbf{1}_{\{|\xi_i| \leqslant N^\omega/2\}}\right) = 1 + O\left(N^{-\Gamma-2+2\omega}\right). \tag{B.1.10}$$

We define D_1 to be an $n \times n$ diagonal matrix with $(D_1)_{ii} = \text{Var}\left(\xi_i \mathbf{1}_{\{|\xi_i| \leqslant N^\omega/2\}}\right)^{1/2}$.

Let $T = U\tilde{D}V$ be a singular value decomposition of T, where U is an $N \times N$ unitary matrix, V is an $M \times M$ unitary matrix and $\tilde{D} = (D, 0)$ is an $N \times M$ rectangular diagonal matrix such that $D = \text{diag}(d_1, d_2, \ldots, d_N)$ with $d_i^2 = \sigma_i$. We denote $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$, where V_1 has size $N \times M$ and V_2 has size $(M - N) \times M$. Then we have

$$\begin{aligned} T\tilde{X} - B &= UDV_1(\tilde{X} - \mathbb{E}\tilde{X}) - (B - T\mathbb{E}\tilde{X}) \\ &= UD\left[V_1(\tilde{X} - \mathbb{E}\tilde{X})D_1^{-1} - \left(D^{-1}U^{-1}B - V_1\mathbb{E}\tilde{X}\right)D_1^{-1}\right]D_1. \end{aligned}$$

Due to (B.1.8) and (B.1.10), we only need to bound $s_n(V_1 Y - A)$, where

$$Y := (\tilde{X} - \mathbb{E}\tilde{X})D_1^{-1}, \quad \text{and} \quad A := (D^{-1}U^{-1}B - V_1\mathbb{E}\tilde{X})D_1^{-1}.$$

Using (B.1.7), (B.1.8), (B.1.10) and the definition of Ω, it is easy to check that A is a deterministic matrix with

$$\|A\| \leqslant C\left(\|B\| + \|\mathbb{E}\tilde{X}\|\right) \leqslant C\left(\sqrt{N} + N^{-\Gamma-1+\omega}\right) \leqslant C\sqrt{N}, \tag{B.1.11}$$

and Y is a random matrix with independent entries satisfying

$$\mathbb{E}(Y_{ij}) = 0, \quad \text{Var}(Y_{ij}) = 1, \quad |Y_{ij}| \leqslant N^\omega. \tag{B.1.12}$$

Recall that a random variable ξ is called subgaussian if there exists $K > 0$ such that

$$\mathbb{P}\left(|\xi| > t\right) \leqslant 2\exp(-t^2/K^2) \quad \text{for all } t > 0. \tag{B.1.13}$$

The infimum of such K is called the subgaussian moment of ξ or the ψ_2-norm $\|\xi\|_{\psi_2}$. By (B.1.12), it is obvious that Y_{ij} are subgaussian random variables with $\|Y_{ij}\|_{\psi_2} \leqslant N^\omega$. Moreover, by Theorem 2.10 of [14], there exists a constant $C > 0$ such that

$$\mathbb{P}\left(\|X\| \leqslant C\sqrt{N}\right) \geqslant 1 - N^{-\Gamma}$$

for large enough N. Then using $\|\tilde{X}\| \leqslant \|X\|$, we get that

$$\mathbb{P}(\|Y\| \leqslant C\sqrt{N}) \geqslant 1 - N^{-\Gamma}. \tag{B.1.14}$$

From the above discussion, we see that Theorem B.1.1 follows from the following theorem.

Theorem B.1.2. *Let $\xi_1, \dots, \xi_n$ be independent centered random variables with unit variance, finite fourth moments and subgaussian moments bounded by K for some $K \equiv K(N) \leqslant N^\omega$. Let Y be an $M \times n$ random matrix, whose rows are independent copies of the random vector $(\xi_1, \dots, \xi_n)$. Let P be an $N \times M$ deterministic matrix with $PP^T = 1$, and let A be an $N \times n$ deterministic matrix. Suppose that $\|Y\| + \|A\| \leqslant C_1\sqrt{N}$ for some constant $C_1 > 0$. Then for every $0 < \omega < \omega_0$ and every $\varepsilon \geqslant 0$, we have*

$$\mathbb{P}\left(s_n(PY - A) \leqslant \varepsilon\left(\sqrt{N} - \sqrt{n-1}\right)\right) \leqslant \left(CK^L\varepsilon\right)^{N-n+1} + e^{-cN/K^4}, \tag{B.1.15}$$

where the constants $\omega_0, c, C, L > 0$ depend only on Λ, C_1 and the maximal fourth moment.

Remark B.1.3. Suppose X_{ij} are subgaussian random variables with $\max_{i,j} \|X_{ij}\| \leqslant K$ for some constant $K > 0$. Then we have

$$\mathbb{P}(\|X\| \geqslant t\sqrt{N}) \leqslant e^{-c_0 t^2 N} \quad \text{for } t \geqslant C_0,$$

where $c_0, C_0 > 0$ depend only on K (see [82, Proposition 2.4]). Combining with Theorem B.1.2, we obtain the optimal estimate for the smallest singular value of $TX - B$:

$$\mathbb{P}\left(s_n(TX - B) \leqslant \varepsilon\left(\sqrt{N} - \sqrt{n-1}\right)\right) \leqslant (C\varepsilon)^{N-n+1} + e^{-cN}. \tag{B.1.16}$$

The rest of this appendix is devoted to the proof of Theorem B.1.2. In the preliminary Section B.2, we introduce some notations and tools that will be used in the proof. In Section B.3, we first reduce the problem into bounding below $\|(PY - A)x\|_2$ for *compressible unit vectors* $x \in S^{n-1}$, whose l^2-norm is concentrated in a small number of coordinates, and for *incompressible unit vectors* comprising the rest of the sphere S^{n-1}. Then we prove a lower bound for compressible unit vectors using a small ball probability result (Lemma B.2.7) and a standard ε-net argument. The incompressible unit vectors are dealt with in Sections B.4 and B.5. In Section B.4, we consider the case $1 \leqslant n \leqslant \lambda_0 N$ for some constant $\lambda_0 \in (0, 1)$, i.e. when $PY - A$ is a tall matrix. The proof can be finished with another small ball probability result (Lemma B.2.6) and the ε-net argument. The almost square case with $\lambda_0 N < n \leqslant N$ is considered in Section B.5. We first reduce the problem into bounding the distance between a random vector and a random subspace, and then complete the proof with a random distance lemma—Lemma B.5.3, whose proof will be given in Section B.6.

B.2 Basic notations and tools

In the proof, the unit sphere centered at the origin in $\mathbb{R}^n$ is denoted S^{n-1}. The orthogonal projection in $\mathbb{R}^n$ onto a subspace E is denoted P_E. For a subset of coordinates $J \subseteq \{1, \ldots, n\}$, we often write P_J for $P_{\mathbb{R}^J}$. The unit sphere of E is denoted $S(E) := S^{n-1} \cap E$.

The following tensorization lemma is Lemma 2.2 of [81]

Lemma B.2.1 (Tensorization)**.** *Let $\zeta_1, \ldots, \zeta_n$ be independent non-negative random variables, and let $B, \varepsilon_0 \geqslant 0$.*

(1) Assume that for each k,

$$\mathbb{P}(\zeta_k < \varepsilon) \leqslant B\varepsilon \quad \textit{for all } \varepsilon \geqslant \varepsilon_0.$$

Then

$$\mathbb{P}\left(\sum_{k=1}^{n} \zeta_k^2 < \varepsilon^2 n\right) \leqslant (CB\varepsilon)^n \quad \textit{for all } \varepsilon \geqslant \varepsilon_0,$$

where C is an absolute constant.

(2) Assume that there exist $\lambda > 0$ and $\mu \in (0,1)$ such that for each k,

$$\mathbb{P}(\zeta_k < \lambda) \leqslant \mu.$$

Then there exists $\lambda_1 > 0$ and $\mu_1 \in (0,1)$ that depend on λ and μ only and such that

$$\mathbb{P}\left(\sum_{k=1}^{n} \zeta_k^2 < \lambda_1 n\right) \leqslant \mu_1^n.$$

Consider a subset $\Omega \subset \mathbb{R}^n$, and let $\varepsilon > 0$. An ε-net of Ω is a subset $\mathcal{N} \subseteq \Omega$ such that for every $x \in \Omega$ one has $\operatorname{dist}(x, \mathcal{N}) \leqslant \varepsilon$. The following lemma is proved as Propositions 2.1 and 2.2 in [82].

Lemma B.2.2 (Nets)**.** *Fix any $\varepsilon > 0$.*

(1) There exists an ε-net of S^{n-1} of cardinality at most

$$\min\left\{\left(1 + 2\varepsilon^{-1}\right)^n, 2n\left(1 + 2\varepsilon^{-1}\right)^{n-1}\right\}.$$

(2) Let S be a subset of S^{n-1}. There exists an ε-net of S of cardinality at most

$$\min\left\{\left(1+4\varepsilon^{-1}\right)^n, 2n\left(1+4\varepsilon^{-1}\right)^{n-1}\right\}.$$

Next we define the small ball probability for a random vector.

Definition B.2.3. *The Lévy concentration function of a random vector $S \in \mathbb{R}^m$ is defined for $\varepsilon > 0$ as*

$$\mathcal{L}(S,\varepsilon) = \sup_{v\in\mathbb{R}^m} \mathbb{P}(\|S-v\|_2 \leqslant \varepsilon),$$

which measures the small ball probabilities.

With Definition B.2.3, it is easy to prove the following lemma. It will allow us to select a nice subset of the coefficients a_k when computing the small ball probability.

Lemma B.2.4. *Let $\xi_1,\ldots,\xi_n$ be independent random variables. For any $\sigma \subseteq \{1,\ldots,n\}$, any $a \in \mathbb{R}^n$ and any $\varepsilon \geqslant 0$, we have*

$$\mathcal{L}\left(\sum_{k=1}^n a_k\xi_k, \varepsilon\right) \leqslant \mathcal{L}\left(\sum_{k\in\sigma} a_k\xi_k, \varepsilon\right).$$

The following three lemmas give some useful small ball probability bounds. They correspond to [82, Lemma 3.2], [81, Corollary 2.9] and [83, Corollary 2.4] respectively.

Lemma B.2.5. *Let ξ be a random variable with mean zero, unit variance, and finite fourth moment. Then for every $\varepsilon \in (0,1)$, there exists a $p \in (0,1)$ which depends only on ε and on the fourth moment, and such that*

$$\mathcal{L}(\xi,\varepsilon) \leqslant p.$$

Lemma B.2.6. *Let $\xi_1,\ldots,\xi_n$ be independent centered random variables with variances at least 1 and third moments bounded by B. Then for every $a \in \mathbb{R}^n$ and every $\varepsilon \geqslant 0$, one has*

$$\mathcal{L}\left(\sum_{k=1}^n a_k\xi_k, \varepsilon\right) \leqslant \sqrt{\frac{2}{\pi}}\frac{\varepsilon}{\|a\|_2} + \tilde{C}B\left(\frac{\|a\|_3}{\|a\|_2}\right)^3,$$

where $\tilde{C}$ is an absolute constant.

Lemma B.2.7. *Let A be a fixed $N \times M$ matrix. Consider a random vector $\xi = (\xi_1, \ldots, \xi_M)$ where ξ_i are independent random variables satisfying $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 = 1$ and $\|\xi_i\|_{\psi_2} \leqslant K$. Then for every $y \in \mathbb{R}^N$, we have*

$$\mathbb{P}\left\{\|A\xi - y\|_2 \leqslant \frac{1}{2}\|A\|_{HS}\right\} \leqslant 2\exp\left(-\frac{c\|A\|_{HS}^2}{K^4\|A\|^2}\right).$$

B.3 Decomposition of the sphere

Now we begin the proof of Theorem B.1.2. We will make use of a partition of the unit sphere into two sets of compressible and incompressible vectors. They are first defined in [81].

Definition B.3.1. *Let $\delta, \rho \in (0, 1]$. A vector $x \in \mathbb{R}^n$ is called sparse if $|supp(x)| \leqslant \delta n$. A vector $x \in S^{n-1}$ is called compressible if x is within Euclidean distance ρ from the set of all sparse vectors. A vector $x \in S^{n-1}$ is called incompressible if it is not compressible. The sets of sparse, compressible and incompressible vectors will be denoted by $Sparse_n(\delta)$, $Comp_n(\delta, \rho)$ and $Incomp_n(\delta, \rho)$. We sometimes omit the subindex n when the dimension is clear.*

Using the decomposition $S^{n-1} = Comp \cup Incomp$, we break the invertibility problem into two subproblems, for compressible and incompressible vectors:

$$\begin{aligned}\mathbb{P}\left(s_n(PY - A) \leqslant \varepsilon(\sqrt{N} - \sqrt{n-1})\right) &\leqslant \mathbb{P}\left(\inf_{x \in Comp_n(\delta,\rho)} \|(PY - A)x\|_2 \leqslant \varepsilon(\sqrt{N} - \sqrt{n-1})\right) \\ &+ \mathbb{P}\left(\inf_{x \in Incomp_n(\delta,\rho)} \|(PY - A)x\|_2 \leqslant \varepsilon(\sqrt{N} - \sqrt{n-1})\right). \end{aligned} \tag{B.3.1}$$

The bound for compressible vectors follows from the following lemma, which is a variant of Lemma 3.3 from [81].

Lemma B.3.2. *Suppose the assumptions in Theorem B.1.2 hold. Then there exist $\rho, c_0, c_1 > 0$ that depend only on C_1, and such that for $\delta \leqslant \min\{c_1 N/(nK^4 \log K), 1\}$, we have*

$$\mathbb{P}\left(\inf_{x \in Comp_n(\delta,\rho)} \|(PY - A)x\|_2 \leqslant c_0\sqrt{N}\right) \leqslant e^{-c_0 N/K^4}.$$

Proof. We first prove a similar estimate for sparse vectors. For any $x \in S^{n-1}$, we define the random vector $\zeta := Yx \in \mathbb{R}^N$. It is easy to verify that $\mathbb{E}\zeta_i = 0$, $\mathbb{E}\zeta_i^2 = 1$ and $\|\zeta_i\|_{\psi_2} \leqslant CK$.

Then with $\|P\| = 1$ and $\|P\|_{HS}^2 = N$, we conclude from Lemma B.2.7 that

$$\mathbb{P}\left\{\|(PY - A)x\|_2 \leqslant \frac{1}{2}\sqrt{N}\right\} \leqslant 2\exp\left(-\frac{cN}{K^4}\right). \tag{B.3.2}$$

Let $S_1 := \{x \in S^{n-1} : x_k = 0, k > \lceil \delta n \rceil\}$. By Lemma B.2.2, there exists an ε-net $\mathcal{N}$ of S_1 with $|\mathcal{N}| \leqslant (5/\varepsilon)^{\lceil \delta n \rceil}$. Then using (B.3.2) and taking the union bound, we get

$$\mathbb{P}\left(\inf_{x \in \mathcal{N}} \|(PY - A)x\|_2 \leqslant \frac{1}{2}\sqrt{N}\right) \leqslant 2e^{-cN/K^4}\left(5\varepsilon^{-1}\right)^{\lceil \delta n \rceil}. \tag{B.3.3}$$

Let V be the event that $\|(PY - A)y\| \leqslant \sqrt{N}/4$ for some $y \in S_1$. By the assumptions of Theorem B.1.2, we have

$$\|PY - A\| \leqslant \|Y\| + \|A\| \leqslant C_1\sqrt{N}.$$

Assume that V occurs and choose a point $x \in \mathcal{N}$ such that $\|y - x\| \leqslant \varepsilon$. Then

$$\|(PY - A)x\|_2 \leqslant \|(PY - A)y\|_2 + \|PY - A\|\|x - y\|_2 \leqslant \frac{1}{4}\sqrt{N} + C_1\varepsilon\sqrt{N} \leqslant \frac{1}{2}\sqrt{N},$$

if we choose $\varepsilon \leqslant 1/(4C_1)$. Fix one such ε, using (B.3.3) we obtain that

$$\mathbb{P}\left(\inf_{x \in S_1} \|(PY - A)x\|_2 \leqslant \frac{1}{4}\sqrt{N}\right) = \mathbb{P}(V) \leqslant 2e^{-cN/K^4}\left(5\varepsilon^{-1}\right)^{\lceil \delta n \rceil} \leqslant e^{-c_2N/K^4},$$

if we choose c_1 (and hence δ) to be sufficiently small. We use this result and take the union bound over all $\lceil \delta n \rceil$-element subsets σ of $\{1, \ldots, n\}$:

$$\begin{aligned}
&\mathbb{P}\left(\inf_{x \in Sparse(\delta) \cap S^{n-1}} \|(PY - A)x\|_2 \leqslant \frac{1}{4}\sqrt{N}\right) \\
=&\mathbb{P}\left(\exists \sigma, |\sigma| = \lceil \delta n \rceil : \inf_{x \in \mathbb{R}^\sigma \cap S^{n-1}} \|(PY - A)x\|_2 \leqslant \frac{1}{4}\sqrt{N}\right) \\
\leqslant& \binom{n}{\lceil \delta n \rceil} e^{-c_2N/K^4} \leqslant \exp\left(4e\delta \log\left(\frac{e}{\delta}\right) n - \frac{c_2N}{K^4}\right) \leqslant \exp\left(-\frac{c_2N}{2K^4}\right),
\end{aligned} \tag{B.3.4}$$

with an appropriate choice of c_1.

Now we deduce the estimate for compressible vectors. Let $c_3 > 0$ and $\rho \in (0, 1/2)$ to be chosen later. We need to control the event W that $\|(PY - A)x\|_2 \leqslant c_3\sqrt{N}$ for some vector $x \in Comp(\delta, \rho)$. Assume W occurs, then every such vector x can be written as a sum $x = y + z$ with $y \in Sparse(\delta)$ and $\|z\|_2 \leqslant \rho$. Thus $\|y\|_2 \geqslant 1 - \rho \geqslant 1/2$, and

$$\|(PY - A)y\|_2 \leqslant \|(PY - A)x\|_2 + \|(PY - A)\|\|z\|_2 \leqslant c_3\sqrt{N} + \rho C_1\sqrt{N}.$$

We choose $c_3 = 1/16$ and $\rho = 1/(16C_1)$, so that $\|(PY - A)y\|_2 \leqslant \sqrt{N}/8$. Since $\|y\|_2 \geqslant 1/2$, we can find a unit vector $u = y/\|y\|_2 \in Sparse(\delta)$ such that $\|(PX - A)u\|_2 \leqslant \sqrt{N}/4$. This shows that event W implies the event in (B.3.4), so we have $\mathbb{P}(W) \leqslant e^{-c_2 N/(2K^4)}$. This concludes the proof. □

Remark B.3.3. If $n < c_1 N/(K^4 \log K)$, then all the vectors in S^{n-1} are in $Comp(\delta, \rho)$ and Lemma B.3.2 already concludes the proof of Theorem B.1.2. Hence throughout the following sections, it suffices to assume

$$n \geqslant c_1 N/(K^4 \log K). \tag{B.3.5}$$

It remains to prove the bound for incompressible vectors in (B.3.1). Define the aspect ratio $\lambda := n/N$. We will divide the proof into two cases: the case where $c_1/(K^4 \log K) \leqslant \lambda \leqslant \lambda_0$ for some constant $0 < \lambda_0 < 1$, and the case where $\lambda_0 < \lambda \leqslant 1$. We record here an important property of the incompressible vectors, which is proved in Lemma 3.4 of [81].

Lemma B.3.4 (Incompressible vectors are spread)**.** *Let $x \in Incomp_n(\delta, \rho)$. Then there exists a set $\sigma \equiv \sigma(x) \subseteq \{1, \ldots, n\}$ of cardinality $|\sigma| \geqslant \frac{1}{2}\rho^2 \delta n$ and such that*

$$\frac{\rho}{\sqrt{2n}} \leqslant |x_k| \leqslant \frac{1}{\sqrt{\delta n}} \quad \textit{for all } k \in \sigma. \tag{B.3.6}$$

B.4 Tall matrices

In this section, we deal with the probability in (B.3.1) when $c_1/(K^4 \log K) \leqslant \lambda \leqslant \lambda_0$ for some constant $\lambda_0 \in (0, 1)$. The value of λ_0 will be chosen later in Section B.5 (see (B.5.8)), and it only depends on Λ, C_1 and the maximal fourth moment of the entries of Y. Then it is equivalent to control the probability

$$\mathbb{P}\left(\inf_{x \in Incomp_n(\delta,\rho)} \|(PY - A)x\|_2 \leqslant t\sqrt{N} \right)$$

for any $t \geqslant 0$.

Let x be a vector in $Incomp_n(\delta, \rho)$, where we fix $\delta = c_1 N/(nK^4 \log K)$ and $0 < \rho \leqslant 1/(16C_1)$ (see Lemma B.3.2). Take the set σ given by Lemma B.3.4. Note that the entries of Yx are of the form $(Yx)_i = \sum_{k=1}^n Y_{ik}x_k$, $1 \leqslant i \leqslant M$, where Y_{ik} are independent centered random variables with unit variance and bounded fourth moment. Hence we can use Lemma B.2.4 and Lemma B.2.6 to get that

$$\mathcal{L}\left((Yx)_i, t\right) \leqslant \sqrt{\frac{2}{\pi}}\frac{t}{\|P_\sigma x\|_2} + C\left(\frac{\|P_\sigma x\|_3}{\|P_\sigma x\|_2}\right)^3 \leqslant C_2(\rho)\left(\frac{t}{\sqrt{\delta}} + \frac{1}{\delta\sqrt{n}}\right), \tag{B.4.1}$$

for some constant $C_2(\rho) > 0$ depending only on ρ and the maximal fourth moment. Here we used the bound

$$\|P_\sigma x\|_2 \geqslant \frac{1}{2}\rho^2\sqrt{\delta}, \quad \left(\frac{\|P_\sigma x\|_3}{\|P_\sigma x\|_2}\right)^3 \leqslant \frac{2}{\rho^2\delta\sqrt{n}},$$

deduced from Lemma B.3.4. With (B.4.1) as the input, the next lemma provides a small ball probability bound for the random vector PYx.

Lemma B.4.1 (Corollary 1.4 of [84]). *Consider a random vector $X = (\xi_1, \ldots, \xi_M)$ where ξ_i are real-valued independent random variables. Let $t, p \geqslant 0$ be such that*

$$\mathcal{L}(\xi_i, t) \leqslant p \quad \text{for all } i = 1, \ldots, M.$$

Let P be an orthogonal projection in $\mathbb{R}^M$ onto an N-dimensional subspace. Then

$$\mathcal{L}\left(PX, t\sqrt{N}\right) \leqslant (Cp)^N,$$

where C is an absolute constant.

Applying the above lemma to random vector Yx, we obtain that

$$\mathbb{P}\left(\|(PY - A)x\|_2 \leqslant t\sqrt{N}\right) \leqslant \mathcal{L}\left(PYx, t\sqrt{N}\right) \leqslant \left[C_3\left(\frac{t}{\sqrt{\delta}} + \frac{1}{\delta\sqrt{n}}\right)\right]^N \tag{B.4.2}$$

for some constant $C_3 > 0$. Now we can take a union bound over all x in an ε-net of $Incomp_n(\delta, \rho)$ and complete the proof by approximation.

We first assume that $t \geqslant 1/\sqrt{\delta n}$. Then the $t/\sqrt{\delta}$ term in (B.4.2) dominates and we obtain that

$$\mathbb{P}\left(\|(PY - A)x\|_2 \leqslant t\sqrt{N}\right) \leqslant \left(2C_3 t/\sqrt{\delta}\right)^N.$$

By Lemma B.2.2, there exists an ε-net $\mathcal{N}$ in $Incomp_n(\delta, \rho)$ of cardinality $|\mathcal{N}| \leqslant 2n(5/\varepsilon)^{n-1}$. Taking the union bound, we get

$$\mathbb{P}\left(\inf_{x\in\mathcal{N}} \|(PY - A)x\|_2 \leqslant t\sqrt{N}\right) \leqslant 2n\left(\frac{2C_3 t}{\sqrt{\delta}}\right)^N \left(\frac{5}{\varepsilon}\right)^{n-1}. \tag{B.4.3}$$

Let V be the event that $\|(PY-A)y\|_2 \leqslant t\sqrt{N}/2$ for some $y \in Incomp_n(\delta,\rho)$. Assume that V occurs and choose a point $x \in \mathcal{N}$ such that $\|x-y\|_2 \leqslant \varepsilon$. Then if $\varepsilon \leqslant t/(2C_1)$, we have

$$\|(PY-A)x\|_2 \leqslant \|(PY-A)y\|_2 + \|PY-A\|\|x-y\|_2 \leqslant \frac{1}{2}t\sqrt{N} + C_1\varepsilon\sqrt{N} \leqslant t\sqrt{N},$$

where we used that $\|PY-A\| \leqslant C_1\sqrt{N}$. Fix one such ε, using (B.4.3) we obtain that

$$\begin{aligned} &\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY-A)x\|_2 \leqslant \frac{t}{2}\sqrt{N}\right) = \mathbb{P}(V) \\ &\leqslant 2n\left(\frac{2C_3 t}{\sqrt{\delta}}\right)^N \left(\frac{10C_1}{t}\right)^{n-1} \leqslant \left[\left(C_4\delta^{-1/2}\right)^{1/(1-\lambda_0)} t\right]^{N-n+1}, \end{aligned} \tag{B.4.4}$$

where in the last step we used $n/N \leqslant \lambda_0$. If $t \leqslant 1/\sqrt{\delta n}$, we use (B.4.4) to get

$$\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY-A)x\|_2 \leqslant \frac{t}{2}\sqrt{N}\right) \leqslant \left[\left(C_4\delta^{-1/2}\right)^{1/(1-\lambda_0)} (\delta n)^{-1/2}\right]^{N-n+1} \leqslant e^{-c_0 N/K^4},$$

if $K \leqslant N^{\omega}$ for some sufficiently small ω. Together with (B.4.4) and Lemma B.3.2, this concludes the proof of Theorem B.1.2 for the $\lambda \leqslant \lambda_0$ case.

B.5 Almost square matrices

In this section, we deal with the probability in (B.3.1) for the $\lambda_0 < \lambda \leqslant 1$ case. In particular, when $\lambda \to 1$, $PY - A$ becomes an almost square matrix and (B.4.4) cannot provide a satisfactory probability bound. For instance, for the square case with $N = n$, it is easy to see that the $(C\delta^{-1/2})^N$ term dominates over the t term. To handle this difficulty, we will use the method in [82], which reduces the problem of bounding $\|(PY-A)x\|_2$ for $x \in Incomp_n(\delta,\rho)$ to a random distance problem. We denote $N = n-1+d$ for some $d \geqslant 1$. Note that $\sqrt{N} - \sqrt{n-1} \leqslant d/\sqrt{n}$. Hence to bound (B.3.1), it suffices to bound

$$\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY-A)x\|_2 \leqslant \varepsilon\frac{d}{\sqrt{n}}\right), \quad \text{for } \delta = \frac{c_1 N}{nK^4\log K},\ \rho \leqslant 1/(16C_1). \tag{B.5.1}$$

We denote

$$m := \min\left\{d, \left\lfloor \frac{1}{2}\rho^2\delta n \right\rfloor\right\}. \tag{B.5.2}$$

Let $Z_1 := PY_1 - A_1, \ldots, Z_n := PY_n - A_n$ be the columns of the matrix $Z := PY - A$. Given a subset $J \subset \{1,\ldots,n\}$ of cardinality m, we define the subspace

$$H_{J^c} := \operatorname{span}(Z_k)_{k\in J^c} \subseteq \mathbb{R}^N. \tag{B.5.3}$$

For levels $K_1 := \rho\sqrt{\delta/2}$ and $K_2 := K_1^{-1}$, we define the set of totally spread vectors

$$S^J := \left\{ y \in S^{n-1} \cap \mathbb{R}^J : \frac{K_1}{\sqrt{m}} \leqslant |y_k| \leqslant \frac{K_2}{\sqrt{m}} \text{ for all } k \in J \right\}. \tag{B.5.4}$$

In the following lemma, we let J be a random subset uniformly chosen over all subsets of $\{1, \dots, n\}$ of cardinality m. We shall write P_J for $P_{\mathbb{R}^J}$, the orthogonal projection onto the subspace $\mathbb{R}^J$. We denote the probability and expectation over the random subset J by $\mathbb{P}_J$ and $\mathbb{E}_J$.

Lemma B.5.1. *There exists constant $c_2 > 0$ depending only on ρ such that for every $x \in Incomp_n(\delta, \rho)$, the event*

$$\mathcal{E}(x) := \left\{ \frac{P_J x}{\|P_J x\|_2} \in S^J \text{ and } \frac{\rho\sqrt{m}}{\sqrt{2n}} \leqslant \|P_J x\|_2 \leqslant \frac{\sqrt{m}}{\sqrt{\delta n}} \right\}$$

satisfies $\mathbb{P}_J(\mathcal{E}(x)) \geqslant (c_2\delta)^m$.

Proof. Let $\sigma \subset \{1, \dots, n\}$ be the subset from Lemma B.3.4. Then we have

$$\mathbb{P}_J\left(J \subset \sigma\right) = \binom{|\sigma|}{m} \Big/ \binom{n}{m}.$$

Using Stirling's approximation, for $d \leqslant \frac{1}{4}\rho^2\delta n$, we have

$$\mathbb{P}_J\left(J \subset \sigma\right) \geqslant \left(\frac{c|\sigma|}{n}\right)^m \geqslant \left(c_2\delta\right)^m,$$

and for $d > \frac{1}{4}\rho^2\delta n$, we have

$$\mathbb{P}_J\left(J \subset \sigma\right) \geqslant \binom{n}{m}^{-1} \geqslant \frac{m!}{n^m} \geqslant \left(\frac{cm}{n}\right)^m \geqslant \left(c_2\delta\right)^m.$$

If $J \subset \sigma$, then summing (B.3.6) over $k \in J$, we obtain the required two-sided bound for $\|P_J x\|_2$. This and (B.3.6) yield $P_J x/\|P_J x\|_2 \in S^J$. Hence $\mathcal{E}(x)$ holds. □

Lemma B.5.1 implies the following lemma, whose proof is similar to the one for [82, Lemma 6.2].

Lemma B.5.2. *Let J denote the m-element subsets of $\{1, \ldots, n\}$. Then for every $\varepsilon > 0$,*

$$\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|Zx\|_2 < \varepsilon\rho\sqrt{\frac{m}{2n}}\right) \leqslant (c_2\delta)^{-m} \max_J \mathbb{P}\left(\inf_{x\in S^J} \mathrm{dist}(Zx, H_{J^c}) < \varepsilon\right). \tag{B.5.5}$$

It remains to bound $\mathbb{P}(\inf_{x\in S^J} \mathrm{dist}(Zx, H_{J^c}) < \varepsilon)$ for any m-element subset J. We shall need the following lemma to bound below the distance between a random vector in $\mathbb{R}^N$ and an independent random subspace of codimension l. It will be proved in Section B.6.

Lemma B.5.3 (Distance to a random subspace)**.** *Let J be any m-element subset of $\{1, \ldots, n\}$ and let H_{J^c} be the random subspace of $\mathbb{R}^N$ defined in (B.5.3). Let X be a random vector in $\mathbb{R}^M$ whose coordinates are i.i.d. centered random variables with unit variance and finite fourth moments, independent of H_{J^c}. Assume that $l := m + d - 1 \leqslant \beta N$. Then for every $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\sup_{v\in\mathbb{R}^N} \mathbb{P}\left(\mathrm{dist}(PX - v, H_{J^c}) < \varepsilon\sqrt{l}\,\middle|\, H_{J^c}\right) > (\tilde{C}\varepsilon)^l + e^{-\tilde{c}N}\right) \leqslant e^{-\tilde{c}N}, \tag{B.5.6}$$

where $\beta, \tilde{c}, \tilde{C} > 0$ depend only on Λ, C_1 and the maximal fourth moment.

It is easy to see that (B.5.6) implies the weaker result:

$$\sup_{v\in\mathbb{R}^N} \mathbb{P}\left(\mathrm{dist}(PX - v, H_{J^c}) < \varepsilon\sqrt{l}\right) \leqslant (\tilde{C}\varepsilon)^l + 2e^{-\tilde{c}N}. \tag{B.5.7}$$

In the following proof, we choose λ_0 such that

$$d \leqslant \beta N/2 \Rightarrow l \leqslant 2d \leqslant \beta N. \tag{B.5.8}$$

Note that for any fixed $x \in S^J$, we have $Zx = PYx - Ax$, where Yx is a random vector satisfying the assumptions for X in Lemma B.5.3. So (B.5.7) gives a useful probability bound for a single $x \in S^J$. Then we will try to take a union bound over all x in an ε-net of S^J and obtain a uniform distance bound. This is stated in the following theorem.

Theorem B.5.4 (Uniform distance bound)**.** *Let Y be a random matrix satisfying the assumptions in Theorem B.1.2. Then for every m-element subset J and $t > 0$,*

$$\mathbb{P}\left(\inf_{x\in S^J} \mathrm{dist}\,(Zx, H_{J^c}) < t\sqrt{d}\right) \leqslant (\bar{C}tK^5 \log K)^d + e^{-\bar{c}N}, \tag{B.5.9}$$

where $\bar{C}, \bar{c} > 0$ depend only on $\tilde{C}$ and $\tilde{c}$.

By the definition of m in (B.5.2), we have

$$(c_2\delta)^{-m} \leqslant \left[(c_2\delta)^{-\rho^2\delta/2}\right]^n \leqslant e^{\bar{c}N/2},$$

with an appropriate choice of ρ. Then we conclude from Lemma B.5.2 and Theorem B.5.4 that

$$\begin{aligned}\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|Zx\|_2 < \varepsilon\rho\sqrt{\frac{md}{2n}}\right) &\leqslant (c_2\delta)^{-m}(\bar{C}\varepsilon K^5 \log K)^d + e^{-\bar{c}N/2} \\ &\leqslant \left(CK^9(\log K)^2\varepsilon\right)^d + e^{-\bar{c}N/2}, \end{aligned} \quad \text{(B.5.10)}$$

where we used $m \leqslant d$ and δ in (B.5.1). Changing ε to $\varepsilon\rho^{-1}\sqrt{2d/m}$ in (B.5.10) and using $d/m \leqslant CK^4 \log K$, we get

$$\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY - A)x\|_2 < \varepsilon\frac{d}{\sqrt{n}}\right) \leqslant \left(CK^{11}(\log K)^{5/2}\varepsilon\right)^d + e^{-\bar{c}N/2},$$

which, together with Lemma B.3.2, concludes the proof of Theorem B.1.2.

Now we begin the proof of Theorem B.5.4. Without loss of generality, we can assume that the entries of Y have absolute continuous distributions. In fact we can add to each entry an independent Gaussian random variable with small variance σ, and later let $\sigma \to 0$ (all the estimates below do not depend on σ). Under this assumption, we have the following convenient fact:

$$\dim(H_{J^c}) = n - m \quad \text{a.s.} \quad \text{(B.5.11)}$$

Let $P_{H^\perp}$ be the orthogonal projection in $\mathbb{R}^N$ onto $H_{J^c}^\perp$, and define

$$W := P_{H^\perp}PY|_{\mathbb{R}^J}. \quad \text{(B.5.12)}$$

Then for every $x \in \mathbb{R}^n$, we have

$$\operatorname{dist}(PYx - v, H_{J^c}) = \|Wx - w\|_2, \quad \text{where } w = P_{H^\perp}v. \quad \text{(B.5.13)}$$

By (B.5.11), $\dim(H_{J^c}^\perp) = N - n + m = l$ almost surely. Thus W acts as an operator from an m-dimensional subspace into an l-dimensional subspace. If we have a proper operator bound for W, we can run the approximation argument on S^J and prove a uniform distance bound over all $x \in S^J$.

Proposition B.5.5. *Let W be a random matrix as in (B.5.12). Then*

$$\mathbb{P}\left(\|W\| > sK\sqrt{d} \,\middle|\, H_{J^c}\right) \leqslant e^{-c_0 s^2 d}, \quad \textit{for } s \geqslant C_0,$$

where $C_0, c_0 > 0$ are absolute constants.

Proof. For simplicity of notations, we fix a realization of H_{J^c} and omit the conditioning on it from the expressions below. Let $\mathcal{N}$ be an (1/2)-net of $S^{n-1} \cap \mathbb{R}^J$ and $\mathcal{M}$ be an (1/2)-net of $S^{n-1} \cap H_{J^c}^{\perp}$. By Lemma B.2.2, we can choose $\mathcal{N}$ and $\mathcal{M}$ such that

$$|\mathcal{N}| \leqslant 5^m, \quad |\mathcal{M}| \leqslant 5^l.$$

It is easy to prove that

$$\|W\| \leqslant 4 \sup_{x\in\mathcal{N}, y\in\mathcal{M}} |\langle Wx, y\rangle| \,. \tag{B.5.14}$$

For every $x \in \mathcal{N}$ and $y \in \mathcal{M}$, $\langle Wx, y\rangle = \langle PYx, y\rangle = \langle Yx, P^T y\rangle$ is a random variable with subgaussian moment bounded by CK for some absolute constant $C > 0$. Hence by (B.1.13) we have

$$\mathbb{P}\left(|\langle Wx, y\rangle| > \frac{1}{4} sK\sqrt{d}\right) \leqslant 2e^{-cs^2 d}.$$

Using (B.5.14) and taking the union bound, we get that for large enough C_0,

$$\mathbb{P}\left(\|W\| > sK\sqrt{d}\right) \leqslant 5^m \cdot 5^l \cdot 2e^{-cs^2 d} \leqslant e^{-c_0 s^2 d}, \quad \text{for } s \geqslant C_0,$$

where we used that $m \leqslant l \leqslant 2d$. □

Lemma B.5.6. *Let W be a random matrix as in (B.5.12) and let w be a random vector as in (B.5.13). Then for every $t \geqslant 0$, we have*

$$\mathbb{P}\left(\inf_{x\in S^J} \|Wx - w\|_2 < t\sqrt{d}, \|W\| \leqslant C_0 K\sqrt{d}\right) \leqslant K^{m-1}(C_2 t)^d + 2e^{-\tilde{c}N/4}, \tag{B.5.15}$$

where C_2 depends only on $\tilde{C}$.

Proof. Fix any $x \in S^J$. It is easy to verify that Yx is a random vector that satisfies the assumptions for X in Lemma B.5.3. Hence by (B.5.13) and (B.5.7), we have

$$\mathbb{P}\left(\|Wx - w\|_2 < t\sqrt{d}\right) \leqslant \mathbb{P}\left(\text{dist}(PYx - v, H_{J^c}) < t\sqrt{l}\right) \leqslant (\tilde{C}t)^l + 2e^{-\tilde{c}N}. \tag{B.5.16}$$

Let $\varepsilon = t/(C_0 K)$. By Lemma B.2.2, there exists an ε-net $\mathcal{N}$ of S^J with $|\mathcal{N}| \leqslant 2m(5C_0K/t)^{m-1}$. Consider the event

$$\mathcal{E}_t := \left\{ \inf_{x \in \mathcal{N}} \|Wx - w\|_2 < 2t\sqrt{d} \right\}.$$

Taking the union bound, we get that

$$\mathbb{P}(\mathcal{E}_t) \leqslant 2m \left(\frac{5C_0K}{t} \right)^{m-1} \left[(2\tilde{C}t)^{m+d-1} + 2e^{-\tilde{c}N} \right] \leqslant K^{m-1}(C_2 t)^d + 4m \left(\frac{5C_0K}{t} \right)^{m-1} e^{-\tilde{c}N}.$$

For $t \geqslant t_0 := e^{-\tilde{c}N/(4d)}/(C_2 K)$, we have

$$4m \left(\frac{5C_0K}{t} \right)^{m-1} \leqslant \left(C_0' K^2 \right)^{\rho^2 \delta n/2} e^{\tilde{c}N/4} \leqslant e^{\tilde{c}N/2}$$

with an appropriate choice of ρ. Thus we get

$$\mathbb{P}(\mathcal{E}_t) \leqslant K^{m-1}(C_2 t)^d + e^{-\tilde{c}N/2}, \quad \text{for } t \geqslant t_0.$$

For $t < t_0$, we have

$$\mathbb{P}(\mathcal{E}_t) \leqslant \mathbb{P}(\mathcal{E}_{t_0}) \leqslant K^{m-1}(C_2 t_0)^d + e^{-\tilde{c}N/2} \leqslant 2e^{-\tilde{c}N/4}.$$

Then applying the standard approximation argument, we can check that the probability in (B.5.15) is bounded by $\mathbb{P}(\mathcal{E}_t)$, which concludes the proof. □

With Proposition B.5.5 and Lemma B.5.6, we obtain that

$$\mathbb{P} \left(\inf_{x \in S^J} \|Wx - w\|_2 < t\sqrt{d} \right) \leqslant K^{m-1}(C_2 t)^d + 2e^{-\tilde{c}N/4} + e^{-c_0 C_0^2 d}.$$

Unfortunately, the bound $e^{-c_0 C_0^2 d}$ is too weak for small d. Following the idea in [82], we refine the probability bound by decoupling the information about $\|Wx - w\|_2$ from the information about $\|W\|$. The proof of next lemma is essentially the same as the one for Proposition 7.5 of [82]. We omit the details.

Lemma B.5.7 (Decoupling)**.** *Let X be an $N \times m$ matrix whose columns are independent random vectors, and let A be an $N \times N$ deterministic matrix. Let $z \in S^{m-1}$ be a vector satisfying $|z_k| \geqslant K_1/\sqrt{m}$ for all $k \in \{1, \ldots, m\}$. Then for every $v \in \mathbb{R}^N$ and every $0 < a < b$, we have*

$$\mathbb{P}\left(\|AXz - Av\|_2 < a, \|AX\| > b \right) \leqslant 2 \sup_{y \in S^{m-1}, u \in \mathbb{R}^N} \mathbb{P}\left(\|AXy - Au\|_2 < \frac{\sqrt{2}a}{K_1} \right) \mathbb{P}\left(\|AX\| > \frac{b}{\sqrt{2}} \right).$$

Remark B.5.8. By (B.5.4), all the vectors in S^J satisfy the assumption for z in Lemma B.5.7.

With this decoupling lemma, we can prove the following refinement of Lemma B.5.6.

Lemma B.5.9. *Let W be a random matrix as in (B.5.12) and let w be a random vector as in (B.5.13). For every $s \geqslant 1$ and every $t \geqslant 0$, we have*

$$\mathbb{P}\left(\inf_{x\in S^J}\|Wx-w\|_2 < t\sqrt{d} \text{ and } sC_0K\sqrt{d} < \|W\| \leqslant 2sC_0K\sqrt{d}\right) \leqslant \left[\frac{K^{m-1}\left(C_3t\right)^d}{K_1^{m+d-1}} + 2e^{-\tilde{c}N/4}\right]e^{-c_1s^2d},$$

where c_1 is an absolute constant and C_3 depends only on $\tilde{C}$.

Proof. Let $\varepsilon = t/(2sC_0K)$. By Lemma B.2.2, there exists an ε-net $\mathcal{N}$ of S^J with $|\mathcal{N}| \leqslant 2m\left(9sC_0K/t\right)^{m-1}$. Consider the event

$$\mathcal{E}_t := \left\{\inf_{x\in\mathcal{N}}\|Wx-w\|_2 < 2t\sqrt{d} \text{ and } \|W\| > sC_0K\sqrt{d}\right\}.$$

Conditioning on H_{J^c}, we can apply Lemma B.5.7 to get that

$$\mathbb{P}\left(\mathcal{E}_t\middle|\, H_{J^c}\right) \leqslant |\mathcal{N}|\cdot 2\sup_{x\in S^{m-1},v\in\mathbb{R}^N}\mathbb{P}\left(\|Wx-P_{H^\perp}v\|_2 < \frac{\sqrt{2}}{K_1}\cdot 2t\sqrt{d}\,\middle|\, H_{J^c}\right)\mathbb{P}\left(\|W\| \geqslant \frac{sC_0K\sqrt{d}}{\sqrt{2}}\,\middle|\, H_{J^c}\right)$$

Taking expectation over H_{J^c} and using Proposition B.5.5, we obtain that

$$\begin{aligned}\mathbb{P}(\mathcal{E}_t) &\leqslant 4m\left(\frac{9sC_0K}{t}\right)^{m-1}e^{-c_0C_0^2s^2d/2}\mathbb{E}\left[\sup_{x\in S^{m-1},v\in\mathbb{R}^N}\mathbb{P}\left(\|Wx-P_{H^\perp}v\|_2 < \frac{\sqrt{2}}{K_1}\cdot 2t\sqrt{d}\,\middle|\, H_{J^c}\right)\right]\\ &\leqslant 4m\left(\frac{9C_0K}{t}\right)^{m-1}\left(s^{m-1}e^{-c_0C_0^2s^2d/2}\right)\left[\left(\frac{2\sqrt{2}\tilde{C}t}{K_1}\right)^{m+d-1} + 2e^{-\tilde{c}N}\right],\end{aligned}$$

where in the second step we used the representation in (B.5.13) and the estimate (B.5.6). Since $s \geqslant 1$ and $1 \leqslant m \leqslant d$, we can bound this as

$$\mathbb{P}(\mathcal{E}_t) \leqslant \left[\frac{K^{m-1}\left(C_3t\right)^d}{K_1^{m+d-1}} + C_4m\left(\frac{9C_0K}{t}\right)^{m-1}e^{-\tilde{c}N}\right]e^{-c_1s^2d},$$

where $C_4 > 0$ is an absolute constant. For $t \geqslant t_1 := e^{-\tilde{c}N/(4d)}K_1^2/(C_3K)$, we have

$$C_4m\left(\frac{9C_0K}{t}\right)^{m-1} \leqslant \left(C_0'K^2K_1^{-2}\right)^{\rho^2\delta n/2}e^{\tilde{c}N/4} \leqslant e^{\tilde{c}N/2}$$

with an appropriate choice of ρ. Thus we get

$$\mathbb{P}(\mathcal{E}_t) \leqslant \left[\frac{K^{m-1}\,(C_3 t)^d}{K_1^{m+d-1}} + e^{-\tilde{c}N/2}\right] e^{-c_1 s^2 d}, \quad \text{for } t \geqslant t_1.$$

For $t < t_1$, we have

$$\mathbb{P}(\mathcal{E}_t) \leqslant \mathbb{P}(\mathcal{E}_{t_1}) \leqslant \left[\frac{K^{m-1}\,(C_3 t_1)^d}{K_1^{m+d-1}} + e^{-\tilde{c}N/2}\right] e^{-c_1 s^2 d} \leqslant 2e^{-\tilde{c}N/4} e^{-c_1 s^2 d}.$$

Suppose there exists $y \in S^J$ such that

$$\|Wy - w\|_2 < t\sqrt{d} \;\text{ and }\; sC_0 K\sqrt{d} < \|W\| \leqslant 2sC_0 K\sqrt{d}.$$

Then we choose $x \in \mathcal{N}$ such that $\|x - y\|_2 \leqslant \varepsilon$, and by triangle inequality we obtain that

$$\|Wx - w\|_2 \leqslant \|Wy - w\|_2 + \|W\|\|x - y\|_2 < t\sqrt{d} + 2sC_0 K\sqrt{d}\varepsilon \leqslant 2t\sqrt{d},$$

i.e. the event $\mathcal{E}_t$ holds. Then the bound for $P(\mathcal{E}_t)$ concludes the proof. □

Proof of Theorem B.5.4. Summing the probability bounds in Lemma B.5.6 and Lemma B.5.9 for $s = 2^k$, $k \in \mathbb{Z}_+$, we conclude that

$$\mathbb{P}\left(\inf_{x \in S^J} \|Wx - w\|_2 < t\sqrt{d}\right) \leqslant K^{m-1}(C_2 t)^d + 2e^{-\tilde{c}N/4} + \left[\frac{K^{m-1}\,(C_3 t)^d}{K_1^{m+d-1}} + 2e^{-\tilde{c}N/4}\right] \sum_{s=2^k, k \in \mathbb{Z}_+} e^{-c_1 s^2 d}$$
$$\leqslant \left(C_5 K K_1^{-2} t\right)^d + C_6 e^{-\tilde{c}N/4}.$$

Using that $K_1 = \rho\sqrt{\delta/2}$ (see (B.5.4)) and $\delta = c_1 N/(nK^4 \log K)$ (see (B.5.1)), we get

$$\mathbb{P}\left(\inf_{x \in S^J} \|Wx - w\|_2 < t\sqrt{d}\right) \leqslant \left(CtK^5 \log K\right)^d + Ce^{-\tilde{c}N/4}.$$

In view of the representation (B.5.13), this concludes the proof. □

B.6 Proof of Lemma B.5.3

We will first prove a general inequality that holds for any fixed subspace H in $\mathbb{R}^N$ of codimension $l = m + d - 1$. This probability bound will depend on the arithmetic structure of H, which can be expressed using the *least common denominator* (LCD). Following the notations in [82], for $\alpha > 0$ and $\gamma \in (0,1)$, we define the least common denominator of a vector $a \in \mathbb{R}^M$ as

$$\mathrm{LCD}_{\alpha,\gamma}(a) := \inf\left\{\theta > 0 : \mathrm{dist}(\theta a, \mathbb{Z}^M) < \min(\gamma\|\theta a\|_2, \alpha)\right\}.$$

More generally, let $a = (a_1, \ldots, a_M)$ be a sequence of vectors $a_k \in \mathbb{R}^l$. We define the product of such multi-vector a and a vector $\theta \in \mathbb{R}^l$ as

$$\theta \cdot a := (\langle\theta, a_1\rangle, \ldots, \langle\theta, a_M\rangle) \in \mathbb{R}^M.$$

Then we define, for $\alpha > 0$ and $\gamma \in (0,1)$,

$$\mathrm{LCD}_{\alpha,\gamma}(a) := \inf\left\{\|\theta\|_2 : \theta \in \mathbb{R}^l, \mathrm{dist}(\theta \cdot a, \mathbb{Z}^M) < \min(\gamma\|\theta \cdot a\|_2, \alpha)\right\}.$$

Finally, the least common denominator of a subspace $E \subseteq \mathbb{R}^M$ is defined as

$$\mathrm{LCD}_{\alpha,\gamma}(E) := \inf\{\mathrm{LCD}_{\alpha,\gamma}(a) : a \in S(E)\} = \inf\left\{\|\theta\|_2 : \theta \in E, \mathrm{dist}(\theta, \mathbb{Z}^M) < \min(\gamma\|\theta\|_2, \alpha)\right\}. \tag{B.6.1}$$

A key to the proof is the next small ball probability theorem.

Theorem B.6.1 (Theorem 3.3 of [82]). *Consider a sequence $a = (a_1, \ldots, a_M)$ of vectors $a_k \in \mathbb{R}^l$, which satisfies*

$$\sum_{k=1}^{M}\langle x, a_k\rangle^2 \geqslant \|x\|_2^2 \quad \text{for every } x \in \mathbb{R}^l. \tag{B.6.2}$$

Let $\xi_1, \ldots, \xi_M$ be i.i.d. centered random variables, such that $\mathcal{L}(\xi_k, 1) \leqslant 1 - b$ for some $b > 0$. Consider the random sum $S := \sum_{k=1}^{M} a_k\xi_k \in \mathbb{R}^l$. Then, for every $\alpha > 0$ and $\gamma \in (0,1)$, and for

$$\varepsilon \geqslant \frac{\sqrt{l}}{LCD_{\alpha,\gamma}(a)},$$

we have

$$\mathcal{L}\left(S, \varepsilon\sqrt{l}\right) \leqslant \left(\frac{C\varepsilon}{\gamma\sqrt{b}}\right)^l + C^l e^{-2b\alpha^2}.$$

Let H be a fixed subspace in $\mathbb{R}^N$ of codimension l. We denote an orthonormal basis of $H^\perp$ by $\{n_1, \ldots, n_l\} \subseteq \mathbb{R}^N$, and write X in coordinates as $X = (\xi_1, \ldots, \xi_M)$. Then using

$PP^T = 1$, we get

$$\operatorname{dist}(PX - v, H) = \|P_{H^\perp}(PX - v)\|_2 = \left\| \sum_{r=1}^{l} \langle PX, n_r \rangle n_r - P_{H^\perp} v \right\|_2 = \left\| \sum_{r=1}^{l} \langle X, P^T n_r \rangle n_r - P_{H^\perp} v \right\|_2$$

$$= \left\| \sum_{r=1}^{l} \langle X, P^T n_r \rangle P^T n_r - P^T P_{H^\perp} v \right\|_2 = \|P_E X - w\|_2 = \left\| \sum_{k=1}^{M} a_k \xi_k - w \right\|_2,$$

where

$$E \equiv E(H) := P^T H^\perp, \quad a_k := P_E e_k, \quad w := P^T P_{H^\perp} v,$$

and where $e_1, \ldots, e_M$ denote the canonical basis of $\mathbb{R}^M$. Notice that

$$\sum_{k=1}^{M} \langle x, a_k \rangle^2 = \|x\|_2^2, \quad \text{for any } x \in E.$$

Hence we can use Theorem B.6.1 in the space E (identified with $\mathbb{R}^l$ by a suitable isometry). For every $\theta = (\theta_1, \ldots, \theta_M) \in E$ and every k, we have $\langle \theta, a_k \rangle = \langle \theta, e_k \rangle = \theta_k$, so $\theta \cdot a = \theta$, where the right hand side is considered as a vector in $\mathbb{R}^M$. Therefore, we have

$$\operatorname{LCD}_{\alpha,\gamma}(E) = \operatorname{LCD}_{\alpha,\gamma}(a).$$

By Lemma B.2.5, $\mathcal{L}(\xi_k, 1/2) \leqslant 1 - b$ for some $b > 0$ that depends only on the fourth moment of ξ_k. Hence we can apply Theorem B.6.1 to $S = \sum_{k=1}^{M} a_k \xi_k$ and conclude that for every $\varepsilon > 0$,

$$\mathbb{P}\left(\operatorname{dist}(PX - v, H) < \varepsilon\sqrt{l}\right) \leqslant \mathcal{L}(S, \varepsilon\sqrt{l}) \leqslant \left(\frac{C\varepsilon}{\gamma}\right)^l + \left(\frac{C\sqrt{l}}{\gamma \operatorname{LCD}_{\alpha,\gamma}(E)}\right)^l + C^l e^{-c\alpha^2}. \quad \text{(B.6.3)}$$

Now it suffices to bound below the least common denominator of the random subspace E. Heuristically, the randomness should remove any arithmetic structure from the subspace E and make the LCD exponentially large. The next theorem shows that this is indeed true.

Theorem B.6.2. *Suppose $\xi_1, \ldots, \xi_{N-l}$ are independent centered random variables with unit variance and uniformly bounded fourth moment. Let $\tilde{Y}$ be an $M \times (N - l)$ random matrix whose rows are independent copies of the random vector $(\xi_1, \ldots, \xi_{N-l})$, and $\tilde{A}$ be an $N \times (N - l)$ deterministic matrix. Suppose that $\|\tilde{Y}\| + \|\tilde{A}\| \leqslant C_1\sqrt{N}$ for some constant $C_1 > 0$.*

Let H be the random subspace of $\mathbb{R}^N$ spanned by the column vectors of $P\tilde{Y} - \tilde{A}$, and define the subspace $E \equiv E(H) := P^T H^{\perp} \subseteq \mathbb{R}^M$. Then for $\alpha = c\sqrt{N}$, we have

$$\mathbb{P}\left(LCD_{\alpha,c}(E) < c\sqrt{N}e^{cN/l}\right) \leqslant e^{-cN},$$

where c depends only on Λ, C_1 and the maximal fourth moment.

Proof of Lemma B.5.3. Consider the event $\mathcal{E} := \left\{\mathrm{LCD}_{\alpha,c}\left(E(H_{J^c})\right) \geqslant c\sqrt{N}e^{cN/l}\right\}$. The above theorem shows that $\mathbb{P}(\mathcal{E}) \geqslant 1 - e^{-cN}$. Conditioning on a realization of H_{J^c} in $\mathcal{E}$, we obtain from (B.6.3) that

$$\sup_{v \in \mathbb{R}^N} \mathbb{P}\left(\mathrm{dist}(PX - v, H_{J^c}) < \varepsilon\sqrt{l} \,\middle|\, H_{J^c}\right) \leqslant (C'\varepsilon)^l + (C')^l e^{-c'N}, \quad \text{for } H_{J^c} \in \mathcal{E}. \tag{B.6.4}$$

Since $l \leqslant \beta N$, with an appropriate choice of β we get

$$(C')^l \leqslant e^{c'N/2}.$$

Then the proof is completed by the estimate on the probability of $\mathcal{E}^c$. □

The rest of this section is devoted to proving Theorem B.6.2. Note that if $a \in E(H)$, then $a = P^T b$ for some $b \in H^{\perp}$. Then with $b = Pa$, we have that

$$b \in H^{\perp} \Leftrightarrow \tilde{Y}^T P^T b - \tilde{A}^T b = 0 \Leftrightarrow \tilde{Y}^T a - \tilde{A}^T P a = 0.$$

We denote $\tilde{B} := \tilde{A}^T P$. For every set S in E, we have

$$\inf_{x \in S} \left\|\tilde{Y}^T x - \tilde{B}x\right\|_2 > 0 \text{ implies } S \cap E = \varnothing. \tag{B.6.5}$$

This helps us to "navigate" the random subspace E away from undesired sets S on the unit sphere.

As in Definition B.3.1, we can define the compressible and incompressible vectors on S^{M-1}, which are denoted by $Comp_M(\delta,\rho)$ and $Incomp_M(\delta,\rho)$, respectively. First, we have the following result for compressible vectors.

Lemma B.6.3 (Random subspaces are incompressible)**.** *There exist $\delta, \rho \in (0,1)$ such that*

$$\mathbb{P}\left(E \cap Comp_M(\delta,\rho) = \varnothing\right) \geqslant 1 - e^{-c_0 N}, \tag{B.6.6}$$

where the constants $\delta, \rho, c_0 > 0$ depend only on Λ, C_1 and the maximal fourth moment.

Proof. Due to (B.6.5), it suffices to prove that

$$\mathbb{P}\left(\inf_{x\in Comp_M(\delta,\rho)}\left\|\left(\tilde{Y}^T-\tilde{B}\right)x\right\|_2 \leqslant c_0\sqrt{N}\right)\leqslant e^{-c_0N}. \tag{B.6.7}$$

In fact, the proof is similar to the one for Lemma B.3.2. However, instead of Lemma B.2.7, we will use the fact that $\tilde{Y}^T$ has independent row vectors $\tilde{Y}_1,\ldots,\tilde{Y}_{N-l}$. For any $x\in S^{M-1}$, it is easy to verify that $\langle\tilde{Y}_k,x\rangle$ has variance 1 and uniformly bounded fourth moment. Then by Lemma B.2.5, there exists a $p\in(0,1)$ such that for any fixed $v=(v_1,\ldots,v_{N-l})\in\mathbb{R}^{N-l}$,

$$\mathbb{P}\left(|\langle\tilde{Y}_k,x\rangle-v_k|\leqslant 1/2\right)\leqslant p.$$

By Lemma B.2.1, we can find constants $\eta,\nu\in(0,1)$ depending on p only and such that

$$\mathbb{P}\left\{\|\tilde{Y}^Tx-v\|_2\leqslant\eta\sqrt{N-l}\right\}\leqslant\nu^{N-l}. \tag{B.6.8}$$

Recall that $l\leqslant\beta N$ and $M\leqslant\Lambda N$ by our assumptions. Then using (B.6.8) instead of (B.3.2), we can complete the proof of (B.6.7) as in Lemma B.3.2. □

Fix the constants δ and ρ given by Lemma B.6.3 for the rest of this section. Note that in contrast to the case in Lemma B.3.2, δ is now an N-independent constant. We will further decompose $Incomp_M(\delta,\rho)$ into level sets S_D according to the value D of the LCD. We shall prove a nontrivial lower bound on $\inf_{x\in S_D}\|(\tilde{Y}^T-\tilde{B})x\|_2$ for each level set up to D of the exponential order. By (B.6.5), this means that E is disjoint from every such level set. Therefore, E must have exponentially large LCD. First, as a consequences of Lemma B.3.4, we have the following lemma, which gives a weak lower bound for the LCD.

Lemma B.6.4 (Lemma 3.6 of [82])**.** *For every $\delta,\rho\in(0,1)$, there exist $c_1(\delta,\rho)>0$ and $c_2(\delta)>0$ such that the following holds. Let $a\in Incomp_M(\delta,\rho)$. Then for every $0<c<c_1(\delta,\rho)$ and every $\alpha>0$, one has*

$$LCD_{\alpha,c}(a)>c_2(\delta)\sqrt{M}.$$

Definition B.6.5 (Level sets)**.** *Let $D\geqslant c_2(\delta)\sqrt{M}$. Define $S_D\subseteq S^{M-1}$ as*

$$S_D:=\{x\in Incomp_M(\delta,\rho):D\leqslant LCD_{\alpha,c}(x)<2D\}\cap\left(P^T\mathbb{R}^N\right).$$

To obtain a lower bound for $\|(\tilde{Y}^T - \tilde{B})x\|_2$ on S_D, we use the ε-net argument again. We first need such a bound for a single vector x. The proof of next lemma is very similar to the one for Lemma 4.6 in [82]. We omit the details.

Lemma B.6.6. *Let $x \in S_D$. Then for every $t > 0$ we have*

$$\mathbb{P}\left(\|(\tilde{Y}^T - \tilde{B})x\|_2 < t\sqrt{N}\right) \leqslant \left(Ct + \frac{C}{D} + Ce^{-c\alpha^2}\right)^{N-l}. \tag{B.6.9}$$

Now we construct a small ε-net of S_D. Our argument here is a little harder than the one in [82], because the ε-net lies in a subspace $P^T\mathbb{R}^N \subseteq \mathbb{R}^M$, whose direction is quite arbitrary. We shall need the following classical result in geometric functional analysis [**?**].

Lemma B.6.7. *If $S \subseteq \mathbb{R}^M$ is a subspace of codimension k, then*

$$|S \cap Q_M| \leqslant (\sqrt{2})^k,$$

where $Q_M = [-1/2, 1/2]^M$ is the unit cube centered at the origin.

Lemma B.6.8. *There exists a $(4\alpha/D)$-net of S_D of cardinality at most $(CD/\sqrt{N})^N$.*

Proof. We can assume that $4\alpha/D \leqslant 1$, otherwise the conclusion is trivial. For $x \in S_D$, we denote $D(x) := \mathrm{LCD}_{\alpha,c}(x)$. By the definition of S_D, we have $D \leqslant D(x) < 2D$. By the definition of LCD, there exists $p \in \mathbb{Z}^M$ such that

$$\|D(x)x - p\|_2 < \alpha. \tag{B.6.10}$$

Therefore,

$$\left\|x - \frac{p}{D(x)}\right\| < \frac{\alpha}{D(x)} \leqslant \frac{1}{4}.$$

Since $\|x\|_2 = 1$, it follows that

$$\left\|x - \frac{p}{\|p\|_2}\right\|_2 \leqslant \frac{2\alpha}{D}.$$

We can chose p such that it is the closest integer point to $D(x)x$. Since $\|D(x)x\|_2 < 2D$, p must lie in the "cube covering" $\tilde{F}$ of $F := B(0, 2D) \cap P^T\mathbb{R}^N$, defined as

$$\tilde{F} := \bigcup_{b \in F}\left(\prod_{i=1}^{M}[b_i - 1/2, b_i + 1/2]\right).$$

On the other hand, by (B.6.10) and using that $\|D(x)x\|_2 < 2D$ and $4\alpha/D \leqslant 1$, we obtain

$$\|p\|_2 < D(x) + \alpha \leqslant 3D.$$

In sum, we get a $(2\alpha/D)$-net of S_D as:

$$\mathcal{N} := \left\{ \frac{p}{\|p\|_2} : p \in \mathbb{Z}^M \cap B(0, 3D) \cap \tilde{F} \right\}.$$

The cardinality of $\mathcal{N}$ can be bounded by the volume of $B(0, 3D) \cap \tilde{F}$. By Fubini's theorem, we have

$$\left|B(0, 3D) \cap \tilde{F}\right| \leqslant |B(0, 3D) \cap S| \cdot \left|S^{\perp} \cap Q_M\right|, \quad S := P^T \mathbb{R}^N.$$

Then using the volume formula for an N-dimension ball and Lemma B.6.7, we obtain that

$$|\mathcal{N}| \leqslant (CD/\sqrt{N})^N.$$

Finally, we can find a $4\alpha/D$-net of the same cardinality, which lies in S_D (see Lemma 5.7 of [81]). This completes the proof. □

Lemma B.6.9. *There exist $c_3, c_4, \mu \in (0, 1)$ such that the following holds. Let $\alpha = \mu\sqrt{N} \geqslant 1$ and $D \leqslant c_3\sqrt{N}e^{c_3 N/l}$. Then*

$$\mathbb{P}\left(\inf_{x \in S_D} \left\| \left(\tilde{Y}^T - \tilde{B}\right) x \right\|_2 < c_4 N/D \right) \leqslant e^{-N}.$$

Proof. To conclude the proof, it is enough to find $\nu > 0$ such that the event

$$\mathcal{E} := \left\{ \inf_{x \in S_D} \left\| \left(\tilde{Y}^T - \tilde{B}\right) x \right\|_2 < \frac{\nu N}{2D} \right\}$$

has probability $\leqslant e^{-N}$. Let $\nu > 0$ be a small constant to be chosen later. We apply Lemma B.6.6 with $t = \nu\sqrt{N}/D$. By the assumptions on α and D, the term Ct dominates in the right hand side of (B.6.9). This gives for arbitrary $x \in S_D$,

$$\mathbb{P}\left(\left\| \left(\tilde{Y}^T - \tilde{B}\right) x \right\|_2 < \frac{\nu N}{D} \right) \leqslant \left(\frac{C\nu\sqrt{N}}{D} \right)^{N-l}.$$

We take the $(4\alpha/D)$-net $\mathcal{N}$ of S_D given by Lemma B.6.8, and take the union bound to get

$$p := \mathbb{P}\left(\inf_{x \in \mathcal{N}} \left\| \left(Y^T - B\right) x \right\|_2 < \frac{\nu N}{D} \right) \leqslant \left(\frac{CD}{\sqrt{N}} \right)^N \left(\frac{C\nu\sqrt{N}}{D} \right)^{N-l} \leqslant \left(\frac{CD}{\sqrt{N}} \right)^l (C'\nu)^{N-l}.$$

Using the assumption on D, we can choose ν small enough such that

$$p \leqslant (C'')^l e^{c_3 N} (C'\nu)^{N-l} \leqslant e^{-N},$$

where we used $l \leqslant \beta N$ in the last step.

Now assume $\mathcal{E}$ holds. By the assumption of Theorem B.6.2, we have

$$\|\tilde{Y}^T - \tilde{B}\| \leqslant \|\tilde{Y}\| + \|\tilde{A}\| \leqslant C_1\sqrt{N}.$$

Fix $x \in S_D$ such that $\|(\tilde{Y}^T - \tilde{B})x\| < \nu N/(2D)$. Then we can find $y \in \mathcal{N}$ such that $\|x - y\| \leqslant 4\alpha/D$. Then, by the triangle inequality we have

$$\left\|\left(\tilde{Y}^T - \tilde{B}\right) y\right\|_2 \leqslant \left\|\left(\tilde{Y}^T - \tilde{B}\right) x\right\|_2 + \left\|\tilde{Y}^T - \tilde{B}\right\| \cdot \|x - y\|_2 \leqslant \frac{\nu N}{2D} + C_1\sqrt{N}\frac{4\mu\sqrt{N}}{D} < \frac{\nu N}{D},$$

if we choose $\mu < \nu/(8C_1)$. Thus we get

$$\mathbb{P}(\mathcal{E}) \leqslant \mathbb{P}\left(\inf_{x\in\mathcal{N}} \left\|\left(Y^T - B\right) x\right\|_2 < \frac{\nu N}{D}\right) \leqslant e^{-N},$$

which concludes the proof. □

Proof of Theorem B.6.2. Consider $x \in S^{M-1} \cap E$ such that

$$\mathrm{LCD}_{\alpha,c}(x) < c_3\sqrt{N}e^{c_3 N/l}.$$

Then, by Lemma B.6.4 and Definition B.6.5, either x is compressible or $x \in S_D$ for some $D \in \mathcal{D}$, where

$$\mathcal{D} := \left\{D : c_2\sqrt{N} \leqslant D < c_3\sqrt{N}e^{c_3 N/l}, D = 2^k, k \in \mathbb{N}\right\},$$

where we used that $M \geqslant N$. Therefore, we can decompose the desired probability as follows:

$$p := \mathbb{P}\left(\mathrm{LCD}_{\alpha,c}(E) < c_3\sqrt{N}e^{c_3 N/l}\right) \leqslant \mathbb{P}\left(E \cap Comp_M(\delta,\rho) \neq \varnothing\right) + \sum_{D\in\mathcal{D}} \mathbb{P}(E \cap S_D \neq \varnothing).$$

The first term can be bounded by $e^{-c_0 N}$ by Lemma B.6.3. The other terms can be bounded with (B.6.5) and Lemma B.6.9:

$$\mathbb{P}(E \cap S_D \neq \varnothing) \leqslant \mathbb{P}\left(\inf_{x\in S_D} \left\|\left(\tilde{Y}^T - \tilde{B}\right) x\right\|_2 = 0\right) \leqslant e^{-N}.$$

Since there are $|\mathcal{D}| \leqslant CN$ terms in the sum, we conclude that

$$p \leqslant e^{-c_0 N} + CNe^{-N} \leqslant e^{-c' N}.$$

This concludes the proof. □

REFERENCES

[1] J. Alt. Singularities of the density of states of random Gram matrices. *Electron. Commun. Probab.*, 22:13 pp., 2017.

[2] J. Alt, L. Erdős, and T. Krüger. Local law for random Gram matrices. *Electron. J. Probab.*, 22:41 pp., 2017.

[3] T. W. Anderson. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34(1):122–148, 1963.

[4] Z. Bai, J. W. Silverstein, and Y. Yin. A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 26:166 – 168, 1988.

[5] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. part II. sample covariance matrices. *Ann. Probab.*, 21(2):649–672, 1993.

[6] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.

[7] Z. D. Bai, B. Q. Miao, and G. M. Pan. On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.*, 35(4):1532–1572, 2007.

[8] Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, volume 2 of *Mathematics Monograph Series*. Science Press, Beijing, 2006.

[9] Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294, 1993.

[10] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[11] J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382 – 1408, 2006.

[12] Z. Bao, G. Pan, and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.*, 43(1):382–421, 2015.

[13] P. Bianchi, M. Debbah, M. Maida, and J. Najim. Performance of statistical tests for single-source detection using random matrix theory. *IEEE Transactions on Information Theory*, 57(4):2400–2419, 2011.

[14] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19(33):1–53, 2014.

[15] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Prob. Theor. Rel. Fields*, 164(1):459–552, 2016.

[16] A. Borodin and C. D. Sinclair. The Ginibre ensemble of real random matrices and its scaling limits. *Commun. Math. Phys.*, 291(1):177–224, 2009.

[17] P. Bourgade and H.-T. Yau. The eigenvector moment flow and local quantum unique ergodicity. *Communications in Mathematical Physics*, 350(1):231–278, 2017.

[18] P. Bourgade, H.-T. Yau, and J. Yin. Local circular law for random matrices. *Probab. Theory Relat. Fields*, 159:545–595, 2014.

[19] P. Bourgade, H.-T. Yau, and J. Yin. The local circular law II: the edge case. *Probab. Theory Relat. Fields*, 159(3):619–660, 2014.

[20] Z. Burda, J. Jurkiewicz, and B. Wacław. Spectral moments of correlated Wishart matrices. *Phys. Rev. E*, 71:026111, 2005.

[21] T. T. Cai, Z. Ma, and Y. Wu. Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.*, 41(6):3074–3110, 2013.

[22] R. Couillet and W. Hachem. Analysis of the limiting spectral measure of large random matrices of the separable covariance type. *Random Matrices: Theory and Applications*, 03(04):1450016, 2014.

[23] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and banach spaces. volume 1 of *Handbook of the Geometry of Banach Spaces*, pages 317 – 366. North-Holland, Amsterdam, 2001.

[24] E. B. Davies. The functional calculus. *J. London Math. Soc. (2)*, 52:166–176, 1995.

[25] M. Dieng and C. A. Tracy. *Application of Random Matrix Theory to Multivariate Statistics*, pages 443–507. Springer New York, New York, NY, 2011.

[26] X. Ding and F. Yang. A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *Ann. Appl. Probab.*, 28(3):1679–1738, 2018.

[27] A. Edelman. The probability that a random real gaussian matrix has k real eigenvalues, related distributions, and the circular law. *J. Multivar. Anal.*, 60(2):203 – 232, 1997.

[28] N. El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 2007.

[29] N. El Karoui. Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.*, 19(6):2362–2405, 2009.

[30] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14:1837–1926, 2013.

[31] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.*, 314:587–640, 2012.

[32] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Delocalization and diffusion profile for random band matrices. *Commun. Math. Phys.*, 323:367–416, 2013.

[33] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electron. J. Probab.*, 18:1–58, 2013.

[34] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.

[35] L. Erdős, B. Schlein, and H.-T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Commun. Math. Phys.*, 287(2):641–655, 2008.

[36] L. Erdős and H.-T. Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics.* American Mathematical Society, Providence, RI, 2017.

[37] L. Erdős, H.-T. Yau, and J. Yin. Universality for generalized Wigner matrices with Bernoulli distribution. *J. of Combinatorics*, 2(1):15–81, 2011.

[38] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics*, 229:1435 – 1515, 2012.

[39] E. F. Fama and K. R. French. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465.

[40] J. Fan, Y. Fan, and J. Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186 – 197, 2008.

[41] J. Fan, Y. Liao, and H. Liu. An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, 19(1):C1–C32.

[42] P. Forrester. The spectrum edge of random matrix ensembles. *Nucl. Phys. B*, 402(3):709 – 728, 1993.

[43] P. J. Forrester and T. Nagao. Eigenvalue statistics of the real Ginibre ensemble. *Phys. Rev. Lett.*, 99:050603, 2007.

[44] M. Fuentes. Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference*, 136(2):447 – 466, 2006.

[45] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.

[46] J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Math. Phys.*, 6(3):440–449, 1965.

[47] V. Girko. The circular law. *Russ. Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984.

[48] F. Götze and A. Tikhomirov. Rate of convergence in probability to the Marchenko-Pastur law. *Bernoulli*, 10(3):503–548, 2004.

[49] F. Gőtze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010.

[50] A. Guionnet, M. Krishnapur, and O. Zeitouni. The single ring theorem. *Ann. Math.*, 174(2):1189–1217, 2011.

[51] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.*, 17(3):875–930, 2007.

[52] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, 3 edition, 2007.

[53] I. M. Johnstone. High dimensional statistical inference and random matrices. *arXiv:0611589*.

[54] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29:295–327, 2001.

[55] I. M. Johnstone. Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.*, 36(6):2638–2716, 2008.

[56] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory.* Prentice-Hall, 1998.

[57] A. Knowles and J. Yin. Eigenvector distribution of Wigner matrices. *Probability Theory and Related Fields*, 155(3):543–582, 2013.

[58] A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66:1663–1749, 2013.

[59] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, pages 1–96, 2016.

[60] P. C. Kyriakidis and A. G. Journel. Geostatistical space–time models: A review. *Mathematical Geology*, 31(6):651–684, 1999.

[61] J. O. Lee and K. Schnelli. Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.*, 26:3786–3839, 2016.

[62] J. O. Lee and J. Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.*, 163:117–173, 2014.

[63] B. Li, M. G. Genton, and M. Sherman. Testing the covariance structure of multivariate random fields. *Biometrika*, 95(4):813–829, 2008.

[64] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.*, 195(2):491 – 523, 2005.

[65] Z. Ma. Sparse principal component analysis and iterative thresholding. *Ann. Statist.*, 41(2):772–801, 2013.

[66] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457, 1967.

[67] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics*. Elsevier, Amsterdam, 3 edition, 2004.

[68] M. W. Mitchell and M. L. Gumpertz. Spatio-temporal prediction inside a free-air CO_2 enrichment system. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(3):310, 2003.

[69] R. R. Nadakuditi and A. Edelman. Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples. *IEEE Transactions on Signal Processing*, 56(7):2625–2638, 2008.

[70] R. R. Nadakuditi and J. W. Silverstein. Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE Journal of Selected Topics in Signal Processing*, 4(3):468–480, 2010.

[71] A. Onatski. The Tracy-Widom limit for the largest eigenvalues of singular complex Wishart matrices. *Ann. Appl. Probab.*, 18:470–490, 04 2008.

[72] A. Onatski. Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5):1447–1479, 2009.

[73] A. Onatski, M. J. Moreira, and M. Hallin. Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics*, 41(3):1204–1231, 2013.

[74] G. Pan and W. Zhou. Circular law, extreme singular values and potential theory. *J. Multivar. Anal.*, 101(3):645–656, 2010.

[75] N. Patterson, A. L. Price, and D. Reich. Population structure and eigenanalysis. *PLOS Genetics*, 2(12):1–20, 2006.

[76] D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.

[77] D. Paul and A. Aue. Random matrix theory in statistics: A review. *J. Stat. Plan. Inference*, 150:1 – 29, 2014.

[78] D. Paul and J. W. Silverstein. No eigenvalues outside the support of the limiting empirical spectral distribution of a separable covariance matrix. *Journal of Multivariate Analysis*, 100(1):37 – 57, 2009.

[79] N. S. Pillai and J. Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24:935–1001, 2014.

[80] M. Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. Math.*, 168(2):575–600, 2008.

[81] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218:600–633, 2008.

[82] M. Rudelson and R. Vershynin. The smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62:1707–1739, 2009.

[83] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, 18:9 pp., 2013.

[84] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Notices*, 2015(19):9594–9617, 2015.

[85] J. W. Silverstein. The smallest eigenvalue of a large dimensional Wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.

[86] J. W. Silverstein. On the eigenvectors of large dimensional sample covariance matrices. *Journal of Multivariate Analysis*, 30(1):1 – 16, 1989.

[87] J. W. Silverstein. On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 30:307 – 311, 1989.

[88] J. W. Silverstein. Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. *Ann. Probab.*, 18(3):1174–1194, 1990.

[89] C. D. Sinclair. Averages over Ginibre's ensemble of random real matrices. *Int. Math. Res. Not.*, 2007:1–15.

[90] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[91] T. Tao and V. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.

[92] T. Tao and V. Vu. Random matrices: Universality of local eigenvalue statistics up to the edge. *Communications in Mathematical Physics*, 298(2):549–572, 2010.

[93] T. Tao and V. Vu. Random matrices: Universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011.

[94] T. Tao and V. Vu. Random matrices: Universal properties of eigenvectors. *Random Matrices: Theory and Applications*, 01(01):1150001, 2012.

[95] T. Tao and V. Vu. Random matrices: Universality of local spectral statistics of non-Hermitian matrices. *Ann. Probab.*, 43(2):782–874, 2015.

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

[97] C. A. Tracy and H. Widom. Level-spacing distributions and the airy kernel. *Comm. Math. Phys.*, 159:151–174, 1994.

[98] C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177:727–754, 1996.

[99] A. M. Tulino and S. Verdú. Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1, 2004.

[100] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Transactions on Information Theory*, 48(6):1319–1343, 2002.

[101] L. Wang and D. Paul. Limiting spectral distribution of renormalized separable sample covariance matrices when $p/n \to 0$. *Journal of Multivariate Analysis*, 126:25 – 52, 2014.

[102] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955.

[103] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A(1/2):32–52, 1928.

[104] H. Xi, F. Yang, and J. Yin. Convergence of eigenvector empirical spectral distribution of sample covariance matrices. *arXiv:1705.03954*.

[105] H. Xi, F. Yang, and J. Yin. Local circular law for the product of a deterministic matrix with a random matrix. *Electron. J. Probab.*, 22:77 pp., 2017.

[106] N. Xia and Z. Bai. Convergence rate of eigenvector empirical spectral distribution of large Wigner matrices. *Statistical Papers*, pages 1–33, 2016.

[107] N. Xia, Y. Qin, and Z. Bai. Convergence rates of eigenvector empirical spectral distribution of large dimensional sample covariance matrix. *Ann. Statist.*, 41(5):2572–2607, 2013.

[108] F. Yang. The smallest singular value of deformed random rectangular matrices. *arXiv:1702.04050*, 2017.

[109] F. Yang. Edge universality of separable covariance matrices. *arXiv:1809.04572*, 2018.

[110] J. Yao, Z. Bai, and S. Zheng. *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press, 2015.

[111] J. Yin. The local circular law III: general case. *Probab. Theory Relat. Fields*, 160(3):679–732, 2014.

[112] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, 78:509–521, 1988.

[113] L. Zhang. *Spectral Analysis of Large Dimensional Random Matrices*. PhD thesis, National University of Singapore.