Sample canonical correlation coefficients of high-dimensional random vectors: local law and Tracy-Widom limit

Fan Yang *[1]

[1]Department of Statistics, University of Pennsylvania

November 3, 2021

Abstract

Consider two random vectors $\mathbf{C}_1^{1/2}\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{C}_2^{1/2}\mathbf{y} \in \mathbb{R}^q$, where the entries of $\mathbf{x}$ and $\mathbf{y}$ are i.i.d. random variables with mean zero and variance one, and $\mathbf{C}_1$ and $\mathbf{C}_2$ are respectively $p \times p$ and $q \times q$ deterministic population covariance matrices. With n independent samples of $(\mathbf{C}_1^{1/2}\mathbf{x}, \mathbf{C}_2^{1/2}\mathbf{y})$, we study the sample correlation between these two vectors using canonical correlation analysis. Under the high-dimensional setting with $p/n \to c_1 \in (0,1)$ and $q/n \to c_2 \in (0, 1-c_1)$ as $n \to \infty$, we prove that the largest sample canonical correlation coefficient converges to the Tracy-Widom distribution as long as we have $\lim_{s\to\infty} s^4\mathbb{P}(|x_{ij}| \geqslant s) = 0$ and $\lim_{s\to\infty} s^4\mathbb{P}(|y_{ij}| \geqslant s) = 0$, which we believe to be a sharp moment condition. This extends the result in [19], which established the Tracy-Widom limit under the assumption that all moments exist for the entries of $\mathbf{x}$ and $\mathbf{y}$. Our proof is based on a new linearization method, which reduces the problem to the study of a $(p+q+2n) \times (p+q+2n)$ random matrix H. In particular, we shall prove an optimal local law on its inverse $G := H^{-1}$, called resolvent. This local law is the main tool for both the proof of the Tracy-Widom law in this paper, and the study in [26, 27] on the canonical correlation coefficients of high-dimensional random vectors with finite rank correlations.

1 Introduction

In multivariate statistics, the canonical correlation analysis (CCA) has been one of the most general and classical methods to study the correlations between two random vectors $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$ since the seminal work by Hotelling [21]. CCA seeks two sequence of orthonormal vectors, such that the projections of $\mathbf{x}$ and $\mathbf{y}$ onto these vectors have maximized correlations. The corresponding sequence of correlations are called the *canonical correlation coefficients* (CCC). More precisely, we first find the unit vectors $\mathbf{a}_1 \in \mathbb{R}^p$ and $\mathbf{b}_1 \in \mathbb{R}^p$ that maximize the correlation,

$$\rho(\mathbf{a}_1, \mathbf{b}_1) = \sup_{\|\mathbf{a}\|=1, \|\mathbf{b}\|=1} \rho(\mathbf{a}, \mathbf{b}), \quad \rho(\mathbf{a}, \mathbf{b}) := \mathrm{Corr}(\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{y}).$$

Then $\rho_1 := \rho(\mathbf{a}_1, \mathbf{b}_1)$ is the first CCC, and $(\mathbf{a}_1^T\mathbf{x}, \mathbf{b}_1^T\mathbf{y})$ is called the first pair of canonical variables. Suppose we have obtained the first k CCC, ρ_i, $1 \leqslant i \leqslant k$, and the corresponding pairs of canonical variables $(\mathbf{a}_i^T\mathbf{x}, \mathbf{b}_i^T\mathbf{y})$, $1 \leqslant i \leqslant k$. We then define inductively the $(k+1)$-th CCC by seeking the vectors $(\mathbf{a}_{k+1}, \mathbf{b}_{k+1})$

*E-mail: fyang75@wharton.upenn.edu

that maximize $\rho(\mathbf{a}_{k+1}, \mathbf{b}_{k+1})$ subject to the constraint that $(\mathbf{a}_{k+1}^T\mathbf{x}, \mathbf{b}_{k+1}^T\mathbf{y})$ is uncorrelated with the first k pairs of canonical variables. Then $\rho_{k+1} := \rho(\mathbf{a}_{k+1}, \mathbf{b}_{k+1})$ is the $(k+1)$-th CCC.

Define the population covariance and cross-covariance matrices

$$\Sigma_{xx} := \mathrm{Cov}(\mathbf{x}, \mathbf{x}), \quad \Sigma_{yy} := \mathrm{Cov}(\mathbf{y}, \mathbf{y}), \quad \Sigma_{xy} = \Sigma_{yx}^T := \mathrm{Cov}(\mathbf{x}, \mathbf{y}).$$

It is well-known that ρ_i^2 is the i-th largest eigenvalue of the population canonical correlation matrix $\mathbf{\Sigma} := \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$. Given n independent samples of $(\mathbf{x}, \mathbf{y})$, we study the CCC through their sample counterparts, which are defined as the eigenvalues of the *sample canonical correlation* (SCC) matrix

$$\mathcal{C}_{XY} := S_{xx}^{-1}S_{xy}S_{yy}^{-1}S_{yx},$$

where

$$S_{xx} := \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T, \quad S_{yy} := \frac{1}{n}\sum_{i=1}^n \mathbf{y}_i\mathbf{y}_i^T, \quad S_{xy} = S_{yx}^T := \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{y}_i^T.$$

We denote the eigenvalues of $\mathcal{C}_{XY}$, i.e. the sample CCC, as $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{p\wedge q}$.

In this paper, we consider the case where $\mathbf{x}$ and $\mathbf{y}$ are independent. We are interested in the behaviors of the eigenvalues of the SCC matrix $\mathcal{C}_{XY}$, including the convergence of (almost) all the eigenvalues and the limiting distribution of the largest few eigenvalues. If the entries of X and Y are i.i.d. Gaussian distributed, then the eigenvalues of $\mathcal{C}_{XY}$ reduce to those of the double Wishart matrices [22]. Moreover, the joint distribution of the eigenvalues of double Wishart matrices has been studied in the context of the so-called Jacobi ensemble and F-type matrices, and it has been shown that the largest eigenvalue converges to the type-1 Tracy-Widom distribution under a proper scaling [20, 22]. For general distribution of the entries of X and Y, the Tracy-Widom law of the largest eigenvalue of $\mathcal{C}_{XY}$ was established in [19] under the assumption that all the moments of the entries are finite. There have been many other works on high-dimensional CCA, and, without attempting to be comprehensive, we mention some of them that are most related to the topic of this paper. In [16], the author derived the asymptotic distributions of the canonical correlation coefficients when one of p and q is fixed as $n \to \infty$. When p and q are proportional to n, the asymptotic distributions of the spiked eigenvalues for CCA with finite rank correlations have been established in [4]. The CLT for linear spectral statistics of CCA was proved in [34]. Under certain sparsity assumptions, the theory of high-dimensional sparse CCA and it applications have been discussed in [17, 18]. In a recent paper [23], the authors studied the asymptotic behaviors of likelihood ratios of CCA under the null hypothesis of no spikes and the alternative hypothesis with a single spike.

One purpose of this paper is to extend the Tracy-Widom law in [19] to the case with weaker moment assumptions. In fact, we prove that the largest eigenvalue of $\mathcal{C}_{XY}$ converges to the Tracy-Widom distribution as long as the following tail condition holds (see Theorem 2.7):

$$\lim_{s\to\infty} s^4\left[\mathbb{P}(|x_i| \geqslant s) + \mathbb{P}(|y_i| \geqslant s)\right] = 0. \tag{1.1}$$

We believe it to be the sharp moment condition, because it has been shown to be necessary and sufficient for the Tracy-Widom limit of the largest eigenvalue of sample covariance matrices [9]. Besides the Tracy-Widom law for the largest eigenvalue, we will also prove a rigidity estimate for (almost) all the eigenvalues of $\mathcal{C}_{XY}$, including the ones in the bulk of the spectrum. This rigidity estimate was not presented [19], and we expect that it will be of independent interest. Different from the methods used in [19], we will develop a new linearization method, which reduces the problem to the study of a $(p+q+2n)\times(p+q+2n)$ random matrix H that is linear in X and Y; see (2.19) below. Moreover, we will prove an optimal local law on its inverse $G := H^{-1}$, i.e. the so-called resolvent, which is another main result of this paper. The linearization idea and the local law allow us to relax the moment assumptions in [19] and prove the Tracy-Widom law under the tail condition in (1.1).

Besides the Tracy-Widom distribution of the largest eigenvalues, the local law on G proved in this paper will also serve as the base of further studies of high-dimensional CCA with finite rank correlations. More precisely, we consider the sample CCC of two high-dimensional random vectors $\widehat{\mathbf{x}} \in \mathbb{R}^p$ and $\widehat{\mathbf{y}} \in \mathbb{R}^q$ with finite rank correlations as following:

$$\widehat{\mathbf{x}} = \mathbf{C}_1^{1/2}\mathbf{x} + A\mathbf{z}, \quad \widehat{\mathbf{y}} = \mathbf{C}_2^{1/2}\mathbf{y} + B\mathbf{z},$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are $p \times p$ and $q \times q$ deterministic non-negative definite symmetric matrices, which give the population covariances, and A and B are $p \times r$ and $q \times r$ deterministic matrices, which are the factor loading matrices. Moreover, suppose that the entries of $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$ and $\mathbf{z} \in \mathbb{R}^r$ are real independent random variables with zero mean and unit variance. For n independent samples $(\widehat{\mathbf{x}}_i, \widehat{\mathbf{y}}_i)$, $1 \leqslant i \leqslant n$, we can arrange them into the following data matrix with a conventional scaling $n^{-1/2}$:

$$\mathcal{X} := \mathbf{C}_1^{1/2}X + AZ, \quad \mathcal{Y} := \mathbf{C}_2^{1/2}Y + BZ.$$

Now X, Y and Z are respectively $p \times n$, $q \times n$ and $r \times n$ matrices with real independent entries with mean zero and variance n^{-1}. We consider the high-dimensional setting with low-rank perturbations, that is, $p/n \to c_1 \in (0,1)$ and $q/n \to c_2 \in (0, 1 - c_1)$ as $n \to \infty$, and $r = \mathrm{O}(1)$ is fixed. For this model, the canonical correlation matrix $\mathbf{\Sigma}$ is of rank $\leqslant r$, and has at most r nonzero eigenvalues $t_i := \rho_i^2$, $1 \leqslant i \leqslant r$. Bao et al. [4] consider this setting for Gaussian vectors, that is, X, Y and Z are all random matrices with i.i.d. Gaussian entries. They show that t_i will give rise to an outlier of the spectrum if it is above some threshold t_c. The outlier lies around a fixed location determined by t_i, and moreover, it is asymptotic Gaussian under the $\sqrt{n}$ scaling. The proof in [4] depends on the fact that multivariate Gaussian distribution is rotational invariant, which is not true for more general distributions. On the other hand, the linearization method developed in this paper allows us to circumvent this issue. Based on the main results of this paper, we will extend the results in [4] to more general distributions, assuming only certain moments conditions on the entries of X, Y and Z. Due to restraint of length of this paper, we will put those results in other papers [26, 27]. In [27] we will study the convergence of the spiked eigenvalues of the SCC matrices, and in [26] we will prove a central limit theorem for the spiked eigenvalues. For all these proofs, the local law for G and the eigenvalue rigidity proved in this paper play central roles.

This paper is organized as follows. In Section 2, we define our model and state the main results—Theorem 2.5 and Theorem 2.7, which give the eigenvalue rigidity and Tracy-Widom law, and Theorem 2.11 and Theorem 2.12, which give the local laws for the resolvent G. In Section 3, we introduce the notations and collect some basic tools that will be used in the proof. Section 4 is devoted to the proof of Theorem 2.5 and Theorem 2.12, and Section 5 contains the proof of Theorem 2.7. Finally, the proof of Theorem 2.11 is divided into two parts: in Section 6, we prove a weaker version of Theorem 2.11, which gives the entrywise local law for G; the proof of Theorem 2.11 is then completed in Section 7 based on the results in Section 6.

Conventions. The fundamental large parameter is n and we always assume that p, q are comparable to 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, whose value may change from one line to the next. Similarly, we use ε, τ, δ and c to denote generic small positive constants. If a constant depends 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 $|a_n| \leqslant C|b_n|$ 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 a matrix A, we use $\|A\| := \|A\|_{l^2 \to l^2}$ to denote the operator norm, $\|A\|_F$ to denote the Frobenius norm, and $\|A\|_{\max} := \max_{i,j} |A_{ij}|$ to denote the max norm. For a vector $\mathbf{v} = (v_i)_{i=1}^n$, $\|\mathbf{v}\| \equiv \|\mathbf{v}\|_2$ stands for the Euclidean norm. In this paper, we often write an identity matrix as I or 1 without causing any confusions. If two random variables X and Y have the same distribution, we write $X \stackrel{d}{=} Y$.

Acknowledgements. The author would like to thank Zongming Ma for bringing this problem to his attention and for helpful discussions.

2 Definitions and main results

2.1 The model

We consider two data matrices

$$\mathcal{X} := \mathbf{C}_1^{1/2} X, \quad \mathcal{Y} := \mathbf{C}_2^{1/2} Y,$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are $p \times p$ and $q \times q$ deterministic population covariance matrices, and $X = (x_{ij})$ and $Y = (y_{ij})$ are $p \times n$ and $q \times n$ random matrices, respectively. We assume that the entries x_{ij}, $1 \leqslant i \leqslant p$, $1 \leqslant j \leqslant n$ and y_{ij}, $1 \leqslant i \leqslant q$, $1 \leqslant j \leqslant n$ are independent (but not necessarily identically distributed) random variables satisfying

$$\mathbb{E} x_{ij} = \mathbb{E} y_{ij} = 0, \qquad \mathbb{E}|x_{ij}|^2 = \mathbb{E}|y_{ij}|^2 = n^{-1}. \tag{2.1}$$

For definiteness, in this paper we focus on the real case, that is, all the random variables are real. However, we remark that our proof can be applied to the complex case after minor modifications. In this paper, we consider the high dimensional setting, i.e.,

$$c_1(n) := \frac{p}{n} \to \hat{c}_1 \in (0,1), \quad c_2(n) := \frac{q}{n} \to \hat{c}_2 \in (0,1), \quad \text{with} \quad c_1(n) + c_2(n) \in (0,1). \tag{2.2}$$

For simplicity, we will always abbreviate $c_1(n) \equiv c_1$ and $c_2(n) \equiv c_2$ in the rest of the paper. Without loss of generality, we can assume that $c_1 \geqslant c_2$. In this paper, we are interested in the eigenvalues of the sample canonical correlation matrix

$$\mathcal{C}_{\mathcal{X}\mathcal{Y}} := \left(\mathcal{X}\mathcal{X}^T\right)^{-1/2} \left(\mathcal{X}\mathcal{Y}^T\right) \left(\mathcal{Y}\mathcal{Y}^T\right)^{-1} \left(\mathcal{Y}\mathcal{X}^T\right) \left(\mathcal{X}\mathcal{X}^T\right)^{-1/2}$$

Since the canonical correlations are invariant under block diagonal transformations $(X, Y) \to (\mathbf{C}_1^{1/2} X, \mathbf{C}_2^{1/2} Y)$, it is equivalent to study the eigenvalues of

$$\mathcal{C}_{XY} := S_{xx}^{-1/2} S_{xy} S_{yy}^{-1} S_{yx} S_{xx}^{-1/2},$$

where

$$S_{xx} := XX^T, \quad S_{yy} := YY^T, \quad S_{xy} = S_{yx}^T := XY^T. \tag{2.3}$$

We will also use the following matrix

$$\mathcal{C}_{YX} := S_{yy}^{-1/2} S_{yx} S_{xx}^{-1} S_{xy} S_{yy}^{-1/2},$$

and denote its eigenvalues by $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_q \geqslant 0$. Note that $\mathcal{C}_{XY}$ shares the same eigenvalues with $\mathcal{C}_{YX}$, except that it has $(p-q)$ more trivial zero eigenvalues $\lambda_{q+1} = \cdots = \lambda_p = 0$.

We now summarize the main assumptions for future reference. For our purpose, we shall relax the assumption (2.1) a little bit.

Assumption 2.1. *Fix a small constant $\tau > 0$. Let $X = (x_{ij})$ and $Y = (Y_{ij})$ be two real independent $p \times n$ and $q \times n$ matrices, whose entries are independent random variables that satisfy the following moment conditions:*

$$\max_{i,j} |\mathbb{E} x_{ij}| \leqslant n^{-2-\tau}, \quad \max_{i,j} |\mathbb{E} y_{ij}| \leqslant n^{-2-\tau}, \tag{2.4}$$

$$\max_{i,j} \left| \mathbb{E}|x_{ij}|^2 - n^{-1} \right| \leqslant n^{-2-\tau}, \quad \max_{i,j} \left| \mathbb{E}|y_{ij}|^2 - n^{-1} \right| \leqslant n^{-2-\tau}. \tag{2.5}$$

Note that (2.4) and (2.5) are slightly more general than (2.1). Moreover, we assume that

$$\tau \leqslant c_2 \leqslant c_1, \quad c_1 + c_2 \leqslant 1 - \tau. \tag{2.6}$$

2.2 The Tracy-Widom limit and eigenvalue rigidity

We denote the ESD of $\mathcal{C}_{YX}$ by

$$F_n(x) := \frac{1}{q}\sum_{i=1}^{q} \mathbf{1}_{\lambda_i \leqslant x}.$$

If X and Y are both i.i.d. Gaussian matrices, then it is known that, almost surely, F_n converges weakly to a deterministic probability distribution $F(x)$ with density [29]

$$f(x) = \frac{1}{2\pi c_2}\frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{x(1-x)}, \quad \lambda_- \leqslant x \leqslant \lambda_+, \tag{2.7}$$

where

$$\lambda_\pm := \left(\sqrt{c_1(1-c_2)} \pm \sqrt{c_2(1-c_1)}\right)^2. \tag{2.8}$$

The convergence of the ESD actually holds under a more general distribution assumption on the entries of X and Y as proved by [33]. We define the quantiles of the density (2.7), which correspond to the classical locations of the eigenvalues of $\mathcal{C}_{YX}$.

Definition 2.2 (Classical locations of eigenvalues)**.** *The classical location γ_j of the j-th eigenvalue is defined as*

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} f(x)\mathrm{d}x > \frac{j-1}{q} \right\}, \tag{2.9}$$

where f is defined in (2.7)*. Note that we have $\gamma_1 = \lambda_+$ and $\lambda_+ - \gamma_j \sim (j/n)^{2/3}$ for $j > 1$.*

Before stating the main results, we first define the following notion of stochastic domination, which was first introduced in [10] and subsequently used in many works on random matrix theory, such as [5, 6, 7, 11, 12, 24]. 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 2.3 (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 paper, 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 2.1. If for some complex family ξ we have $|\xi| \prec \zeta$, then we will also write $\xi \prec \zeta$ or $\xi = \mathrm{O}_\prec(\zeta)$.

(ii) We extend the definition of $\mathrm{O}_\prec(\cdot)$ 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 $A = \mathrm{O}_\prec(\zeta)$ means that $|\langle \mathbf{v}, A\mathbf{w}\rangle| \prec \zeta\|\mathbf{v}\|_2\|\mathbf{w}\|_2$ for any deterministic vectors $\mathbf{v}$ and $\mathbf{w}$.

(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.

For X and Y, we introduce the following bounded support condition.

Definition 2.4 (Bounded support condition)**.** *We say a random matrix X satisfies the bounded support condition with ϕ_n, if*

$$\max_{i,j} |x_{ij}| \leqslant \phi_n. \tag{2.10}$$

Usually ϕ_n is a deterministic parameter and satisfies $n^{-1/2} \leqslant \phi_n \leqslant n^{-c_\phi}$ for some (small) constant $c_\phi > 0$. Whenever (2.10) holds, we say that X has support ϕ_n.

Then we have the following eigenvalue rigidity and edge universality result for $\mathcal{C}_{YX}$, which extends the result in [19].

Theorem 2.5. *Suppose Assumption 2.1 holds. Suppose X and Y have bounded support ϕ_n such that $n^{-1/2} \leqslant \phi_n \leqslant n^{-c_\phi}$ for some constant $c_\phi > 0$. Assume that*

$$\begin{aligned} &\max_{i,j} \mathbb{E}|x_{ij}|^3 = \mathrm{O}(n^{-3/2}), \quad \max_{i,j} \mathbb{E}|x_{ij}|^4 \prec n^{-2}, \\ &\max_{i,j} \mathbb{E}|y_{ij}|^3 = \mathrm{O}(n^{-3/2}), \quad \max_{i,j} \mathbb{E}|y_{ij}|^4 \prec n^{-2}. \end{aligned} \tag{2.11}$$

Then the eigenvalues λ_i of the SCCA matrix $\mathcal{C}_{YX}$ satisfy the following eigenvalue rigidity estimate: if $\lambda_- \geqslant \varepsilon$ for some constant $\varepsilon > 0$, then

$$|\lambda_i - \gamma_i| \prec [i \wedge (q + 1 - i)]^{-1/3} n^{-2/3}, \quad 1 \leqslant i \leqslant q, \tag{2.12}$$

Otherwise, if $\lambda_- = \mathrm{o}(1)$, then (2.12) hold for all $1 \leqslant i \leqslant (1-\varepsilon)q$ for any constant $\varepsilon > 0$. Moreover, we have that for any fixed k,

$$\lim_{n\to\infty} \mathbb{P}\left(\left(n^{\frac{2}{3}} \frac{\lambda_i - \lambda_+}{c_{TW}} \leqslant s_i \right)_{1 \leqslant i \leqslant k} \right) = \lim_{n\to\infty} \mathbb{P}^{GOE} \left(\left(n^{\frac{2}{3}} (\lambda_i - 2) \leqslant s_i \right)_{1 \leqslant i \leqslant k} \right), \tag{2.13}$$

for all $s_1, s_2, \ldots, s_k \in \mathbb{R}$, where

$$c_{TW} := \left[\frac{\lambda_+^2 (1 - \lambda_+)^2}{\sqrt{c_1 c_2 (1 - c_1)(1 - c_2)}} \right]^{1/3},$$

and $\mathbb{P}^{GOE}$ stands for the law of the Gaussian orthogonal ensemble (GOE) of dimension $n \times n$.

Recall that the joint distribution of the k largest eigenvalues of GOE can be written in terms of the Airy kernel for any fixed k [15]. Moreover, taking $k = 1$ in (2.13), we obtain that

$$n^{\frac{2}{3}} \frac{\lambda_i - \lambda_+}{c_{TW}} \Rightarrow F_1,$$

where F_1 is the Type-1 Tracy-Widom distribution. The result (2.13) was proved in [19] under the assumption that all the moments of $\sqrt{n} x_{ij}$ and $\sqrt{n} y_{ij}$ exist. On the other hand, combining our result with a simple cutoff argument allows us to obtain the following corollary under the finite $(4+\varepsilon)$-th moment assumption. Since we do not assume the entries of X and Y are identically distributed, the means and variances of the truncated entries may be different. This is why we assume the slightly more general conditions (2.4) and (2.5).

Corollary 2.6. *Suppose* (2.6) *holds. Assume that* $X = (x_{ij})$ *and* $Y = (Y_{ij})$ *are two real independent* $p \times n$ *and* $q \times n$ *matrices, whose entries are independent random variables that satisfy* (2.1) *and*

$$\max_{i,j} \mathbb{E}|\sqrt{n} x_{ij}|^{4+\tau} \leqslant C, \quad \max_{i,j} \mathbb{E}|\sqrt{n} y_{ij}|^{4+\tau} \leqslant C, \tag{2.14}$$

for some constants $\tau, C > 0$. *Then both* (2.12) *and* (2.13) *hold with probability* $1 - \mathrm{o}(1)$.

Proof. We choose the constants $c_\phi > 0$ small enough such that $\left(n^{1/2-c_\phi}\right)^{4+\tau} \geqslant n^{2+\varepsilon}$ for some constant $\varepsilon > 0$. Then we introduce the following truncation

$$\widetilde{X} := \mathbf{1}_\Omega X, \quad \widetilde{Y} := \mathbf{1}_\Omega Y, \quad \Omega := \left\{ \max_{i,j} |x_{ij}| \leqslant n^{-c_\phi}, \max_{i,j} |y_{ij}| \leqslant n^{-c_\phi} \right\}.$$

By the moment conditions (2.14) and a simple union bound, we have

$$\mathbb{P}(\widetilde{X} \neq X, \widetilde{Y} \neq Y) = \mathrm{O}(n^{-\varepsilon}). \tag{2.15}$$

Using (2.14) and integration by parts, it is easy to verify that

$$\mathbb{E}\, |x_{ij}|\, 1_{|x_{ij}| > n^{-c_\phi}} = \mathrm{O}(n^{-2-\varepsilon}), \quad \mathbb{E}\, |x_{ij}|^2\, 1_{|x_{ij}| > n^{-c_\phi}} = \mathrm{O}(n^{-2-\varepsilon}),$$

which imply that

$$|\mathbb{E}\tilde{x}_{ij}| = \mathrm{O}(n^{-2-\varepsilon}), \quad \mathbb{E}|\tilde{x}_{ij}|^2 = n^{-1} + \mathrm{O}(n^{-2-\varepsilon}).$$

Moreover, we trivially have

$$\mathbb{E}|\tilde{x}_{ij}|^4 \leqslant \mathbb{E}|x_{ij}|^4 = \mathrm{O}(n^{-2}).$$

Similar estimates also hold for the entries of Y. Hence $\widetilde{X}$ and $\widetilde{Y}$ are random matrices satisfying Assumption 2.1 and condition (2.11). Now combing (2.15) and Theorem 2.5, we conclude the corollary. ☐

If we assume that the entries of X and Y are identically distributed, respectively, then the Tracy-Widom law actually holds under the weaker tail condition (2.16), which we believe to be sharp.

Theorem 2.7. *Suppose* (2.6) *holds. Assume that* $x_{ij} = n^{-1/2}\widehat{x}_{ij}$ *and* $y_{ij} = n^{-1/2}\widehat{y}_{ij}$, *where* $\{\widehat{x}_{ij}\}$ *and* $\{\widehat{y}_{ij}\}$ *are independent families of i.i.d. random variables with mean zero and variance one. Then for any fixed* k, (2.13) *holds under the following tail condition:*

$$\lim_{t\to\infty} t^4 \left[\mathbb{P}\left(|\widehat{x}_{11}| \geqslant t\right) + \mathbb{P}\left(|\widehat{y}_{11}| \geqslant t\right)\right] = 0. \tag{2.16}$$

2.3 The linearization method and local law

The self-adjoint linearization method has been proved to be useful in studying the local laws of random matrices of the Gram type [1, 2, 8, 19, 20, 24, 31, 32]. We now introduce a generalization of this method, which will be the starting point of this paper.

For now, we assume that XX^T and YY^T are both non-singular almost surely. This is trivially true if, say, the entries of X and Y have continuous densities. For any $\lambda > 0$, it is an eigenvalue of $\mathcal{C}_{XY}$ if and only if the following equation holds:

$$\det\left(\left(XY^T\right)\left(YY^T\right)^{-1}\left(YX^T\right) - \lambda XX^T\right) = 0. \tag{2.17}$$

By Schur complement, it is equivalent to

$$\det\begin{pmatrix}\lambda XX^T & \lambda^{1/2}XY^T\\ \lambda^{1/2}YX^T & \lambda YY^T\end{pmatrix}=0 \quad\Leftrightarrow\quad \det\begin{pmatrix}X & 0\\ 0 & Y\end{pmatrix}\begin{pmatrix}\lambda I_n & \lambda^{1/2}I_n\\ \lambda^{1/2}I_n & \lambda I_n\end{pmatrix}\begin{pmatrix}X^T & 0\\ 0 & Y^T\end{pmatrix}=0.$$

Using Schur complement again, if $\lambda \notin \{0,1\}$, then it is equivalent to

$$\det\begin{pmatrix} 0 & \begin{pmatrix}X & 0\\ 0 & Y\end{pmatrix}\\ \begin{pmatrix}X^T & 0\\ 0 & Y^T\end{pmatrix} & \begin{pmatrix}\lambda I_n & \lambda^{1/2}I_n\\ \lambda^{1/2}I_n & \lambda I_n\end{pmatrix}^{-1}\end{pmatrix}=0. \tag{2.18}$$

Inspired by the above discussion, we define the following $(p+q+2n)\times(p+q+2n)$ self-adjoint block matrix

$$H(\lambda):=\begin{pmatrix} 0 & \begin{pmatrix}X & 0\\ 0 & Y\end{pmatrix}\\ \begin{pmatrix}X^T & 0\\ 0 & Y^T\end{pmatrix} & \begin{pmatrix}\lambda I_n & \lambda^{1/2}I_n\\ \lambda^{1/2}I_n & \lambda I_n\end{pmatrix}^{-1}\end{pmatrix}. \tag{2.19}$$

We can also extend the argument λ to $z\in\mathbb{C}_+ := \{z\in\mathbb{C} : \operatorname{Im} z>0\}$ and define $H(z)$ in general, where we take $z^{1/2}$ to be the branch with positive imaginary part. We then define the resolvent (or Green's function) as

$$G(z):=[H(z)]^{-1},\quad z\in\mathbb{C}_+, \tag{2.20}$$

whenever the inverse exists.

Definition 2.8 (Index sets)**.** *For simplicity of notations, we define the index sets*

$$\mathcal{I}_1 := [\![1,p]\!],\qquad \mathcal{I}_2 := [\![p+1,p+q]\!],$$

and

$$\mathcal{I}_3 := [\![p+q+1,p+q+n]\!],\qquad \mathcal{I}_4 := [\![p+q+n+1,p+q+2n]\!].$$

We will consistently use the latin letters $i,j\in\mathcal{I}_{1,2}$ and greek letters $\mu,\nu\in\mathcal{I}_{3,4}$. Moreover, we shall use the notations $\mathfrak{a},\mathfrak{b}\in\mathcal{I}:=\cup_{i=1}^4\mathcal{I}_i$. We label the indices of the matrices according to

$$X=(x_{i\mu}: i\in\mathcal{I}_1,\mu\in\mathcal{I}_3),\quad Y=(y_{j\nu}: j\in\mathcal{I}_2,\nu\in\mathcal{I}_4).$$

Moreover, we denote $\bar{i}:=i+p$ for $i\in\mathcal{I}_1$, $\bar{j}:=j-p$ for $j\in\mathcal{I}_2$, $\bar{\mu}:=\mu+n$ for $\mu\in\mathcal{I}_3$, and $\bar{\nu}:=\nu-n$ for $\nu\in\mathcal{I}_4$.

Definition 2.9 (Resolvents)**.** *We denote the $\mathcal{I}_\alpha\times\mathcal{I}_\alpha$ block of $G(z)$ by $\mathcal{G}_\alpha(z)$ for $\alpha=1,2,3,4$. We denote the $(\mathcal{I}_1\cup\mathcal{I}_2)\times(\mathcal{I}_1\cup\mathcal{I}_2)$ block of $G(z)$ by $\mathcal{G}_L(z)$, the $(\mathcal{I}_1\cup\mathcal{I}_2)\times(\mathcal{I}_3\cup\mathcal{I}_4)$ block by $\mathcal{G}_{LR}(z)$, the $(\mathcal{I}_3\cup\mathcal{I}_4)\times(\mathcal{I}_1\cup\mathcal{I}_2)$ block by $\mathcal{G}_{RL}(z)$, and the $(\mathcal{I}_3\cup\mathcal{I}_4)\times(\mathcal{I}_3\cup\mathcal{I}_4)$ block by $\mathcal{G}_R(z)$. We introduce the following random quantities:*

$$m_\alpha(z):=\frac{1}{n}\operatorname{Tr}\mathcal{G}_\alpha(z)=\frac{1}{n}\sum_{\mathfrak{a}\in\mathcal{I}_\alpha}G_{\mathfrak{a}\mathfrak{a}}(z),\quad \alpha=1,2,3,4.$$

Recalling the notations in (2.3)*, we define $\mathcal{H}:=S_{xx}^{-1/2}S_{xy}S_{yy}^{-1/2}$ and*

$$\begin{aligned} R_1(z) &:= (\mathcal{C}_{XY}-z)^{-1}=(\mathcal{H}\mathcal{H}^T-z)^{-1},\\ R_2(z) &:= (\mathcal{C}_{YX}-z)^{-1}=(\mathcal{H}^T\mathcal{H}-z)^{-1},\quad m(z):=q^{-1}\operatorname{Tr}R_2(z).\end{aligned} \tag{2.21}$$

Note that we have $R_1\mathcal{H} = \mathcal{H}R_2$, $\mathcal{H}^T R_1 = R_2\mathcal{H}^T$, *and*

$$\operatorname{Tr} R_1 = \operatorname{Tr} R_2 - \frac{p-q}{z} = qm(z) - \frac{p-q}{z}, \tag{2.22}$$

since $\mathcal{C}_{XY}$ *has* $(p-q)$ *more zeros eigenvalues than* $\mathcal{C}_{YX}$.

By Schur complement formula, we immediately obtain that

$$\mathcal{G}_L = \begin{pmatrix} S_{xx}^{-1/2} R_1 S_{xx}^{-1/2} & -z^{-1/2} S_{xx}^{-1/2} R_1 \mathcal{H} S_{yy}^{-1/2} \\ -z^{-1/2} S_{yy}^{-1/2} \mathcal{H}^T R_1 S_{xx}^{-1/2} & S_{yy}^{-1/2} R_2 S_{yy}^{-1/2} \end{pmatrix}, \tag{2.23}$$

and

$$\mathcal{G}_1 = S_{xx}^{-1/2} R_1 S_{xx}^{-1/2} = \left(S_{xy} S_{yy}^{-1} S_{yx} - z S_{xx}\right)^{-1},$$
$$\mathcal{G}_2 = S_{yy}^{-1/2} R_2 S_{yy}^{-1/2} = \left(S_{yx} S_{xx}^{-1} S_{xy} - z S_{yy}\right)^{-1}.$$

The other blocks are

$$\mathcal{G}_R = \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix} + \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix} \begin{pmatrix} X^T & 0 \\ 0 & Y^T \end{pmatrix} \mathcal{G}_L \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}, \tag{2.24}$$

and

$$\begin{aligned} \mathcal{G}_{LR}(z) &= -\mathcal{G}_L(z) \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}, \\ \mathcal{G}_{RL}(z) &= -\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix} \begin{pmatrix} X^T & 0 \\ 0 & Y^T \end{pmatrix} \mathcal{G}_L(z). \end{aligned} \tag{2.25}$$

Expanding the product in (2.24) using (2.23) and calculating the partial traces, one can verify directly that

$$m_3(z) = z + \frac{1}{n}\left(-2zp - z^2 \operatorname{Tr} R_1 + z \operatorname{Tr} R_2\right) = c_2 z(1-z)m(z) + (1-c_1-c_2)z, \tag{2.26}$$

and

$$\begin{aligned} m_4(z) &= z + \frac{1}{n}\left(-2zq - z^2 \operatorname{Tr} R_2 + z \operatorname{Tr} R_1\right) \\ &= c_2 z(1-z)m(z) - (c_1-c_2) + (1-2c_2)z. \end{aligned} \tag{2.27}$$

where we also used (2.22). In particular, we have the identity

$$m_3(z) - m_4(z) = (1-z)(c_1-c_2). \tag{2.28}$$

We now give the deterministic limit of m_α, $\alpha = 1,2,3,4$, as $n \to \infty$:

$$m_{1c}(z) = \frac{-z + c_1 + c_2 + \sqrt{(z-\lambda_-)(z-\lambda_+)}}{2(1-c_1)z(1-z)} - \frac{c_1}{(1-c_1)z}, \tag{2.29}$$

$$m_{2c}(z) = \frac{-z + c_1 + c_2 + \sqrt{(z-\lambda_-)(z-\lambda_+)}}{2(1-c_2)z(1-z)} - \frac{c_2}{(1-c_2)z}, \tag{2.30}$$

$$m_{3c}(z) = \frac{1}{2}\left[(1-2c_1)z + c_1 - c_2 + \sqrt{(z-\lambda_-)(z-\lambda_+)}\right], \tag{2.31}$$

$$m_{4c}(z) = \frac{1}{2}\left[(1-2c_2)z + c_2 - c_1 + \sqrt{(z-\lambda_-)(z-\lambda_+)}\right], \tag{2.32}$$

where we recall (2.8). One can verify when $z \to 1$, $m_{1c}(z)$ and $m_{2c}(z)$ have finite limits, which we define as $m_{1c}(1)$ and $m_{2c}(1)$. Moreover, by (2.26) the deterministic limit of m is

$$m_c(z) = \frac{m_{3c}(z) + (c_1 + c_2 - 1)z}{c_2 z(1-z)} = \frac{1-c_2}{c_2} m_{2c}(z). \tag{2.33}$$

One can directly verify that the following equations hold:

$$m_{1c} = -\frac{c_1}{m_{3c}}, \quad m_2 = -\frac{c_2}{m_{4c}}, \quad m_{3c}(z) - m_{4c}(z) = (1-z)(c_1 - c_2), \tag{2.34}$$

$$m_{3c}(z) = \frac{1-(z-1)m_{2c}(z)}{z^{-1} - (m_{1c}(z) + m_{2c}(z)) + (z-1)m_{1c}(z)m_{2c}(z)}, \tag{2.35}$$

$$m_{3c}^2(z) + \left[(2c_1 - 1)z - c_1 + c_2\right] m_{3c}(z) + c_1(c_1 - 1)z(z-1) = 0. \tag{2.36}$$

We then define the matrix limit of $G(z)$ as

$$\Pi(z) := \begin{pmatrix} \begin{pmatrix} c_1^{-1} m_{1c}(z) I_p & 0 \\ 0 & c_2^{-1} m_{2c}(z) I_q \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} m_{3c}(z) I_n & h(z) I_n \\ h(z) I_n & m_{4c}(z) I_n \end{pmatrix} \end{pmatrix}, \tag{2.37}$$

where

$$\begin{aligned} h(z) : &= \frac{z^{-1/2} m_{3c}(z)}{1 + (1-z) m_{2c}(z)} = \frac{z^{-1/2} m_{4c}(z)}{1 + (1-z) m_{1c}(z)} \\ &= \frac{z^{1/2}}{2}\left[-z + (2 - c_1 - c_2) + \sqrt{(z-d_-)(z-d_+)}\right]. \end{aligned} \tag{2.38}$$

For simplicity of notations, we introduce the notion of generalized entries.

Definition 2.10 (Generalized entries)**.** *For $\mathbf{v}, \mathbf{w} \in \mathbb{C}^{\mathcal{I}}$, $\mathfrak{a} \in \mathcal{I}$ and an $\mathcal{I} \times \mathcal{I}$ matrix $\mathcal{A}$, we shall denote*

$$\mathcal{A}_{\mathbf{vw}} := \langle \mathbf{v}, \mathcal{A}\mathbf{w} \rangle, \quad \mathcal{A}_{\mathbf{v}\mathfrak{a}} := \langle \mathbf{v}, \mathcal{A}\mathbf{e}_{\mathfrak{a}} \rangle, \quad \mathcal{A}_{\mathfrak{a}\mathbf{w}} := \langle \mathbf{e}_{\mathfrak{a}}, \mathcal{A}\mathbf{w} \rangle, \tag{2.39}$$

where $\mathbf{e}_{\mathfrak{a}}$ is the standard unit vector along $\mathfrak{a}$-th coordinate axis, and the inner product is defined as $\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v}^* \mathbf{w}$ with $\mathbf{v}^*$ denoting the conjugate transpose. Given a vector $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_\alpha}$, $\alpha = 1,2,3,4$, we always identify it with its natural embedding in $\mathbb{C}^{\mathcal{I}}$. For example, we shall identify $\mathbf{v} \in \mathbb{C}^{\mathcal{I}_1}$ with $\begin{pmatrix} \mathbf{v} \\ \mathbf{0}_{q+2n} \end{pmatrix} \in \mathbb{C}^{\mathcal{I}}$.

Now we are ready to state the local laws for $G(z)$. For any constant $\varepsilon > 0$, we define a domain of the spectral parameter z as

$$S(\varepsilon) := \left\{ z = E + \mathrm{i}\eta : \varepsilon \leqslant E \leqslant 1, n^{-1+\varepsilon} \leqslant \eta \leqslant \varepsilon^{-1} \right\}. \tag{2.40}$$

We define the distance to the two edges as

$$\kappa \equiv \kappa_E := \min\left\{ |E - \lambda_-|, |E - \lambda_+| \right\}, \quad \text{for } z = E + \mathrm{i}\eta. \tag{2.41}$$

Theorem 2.11 (Local laws)**.** *Suppose the assumptions of Theorem 2.5 hold. Then for any fixed $\varepsilon > 0$, the following estimates hold.*

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

$$|G_{\mathbf{uv}}(z) - \Pi_{\mathbf{uv}}(z)| \prec \phi_n + \Psi(z), \tag{2.42}$$

where $\Psi(z)$ is a deterministic control parameter defined as

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m_c(z)}{n\eta}} + \frac{1}{n\eta}, \quad z = E + \mathrm{i}\eta. \tag{2.43}$$

(2) **Weak averaged local law***: For any $z \in S(\varepsilon)$, we have*

$$|m_\alpha(z) - m_{\alpha c}(z)| \prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{1}{n\eta}, \quad \alpha = 1,2,3,4. \tag{2.44}$$

Moreover, outside of the spectrum we have the following stronger estimate

$$|m_\alpha(z) - m_{\alpha c}(z)| \prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}, \quad \alpha = 1,2,3,4, \tag{2.45}$$

uniformly in $z \in S_{out}(\varepsilon) := S(\varepsilon) \cap \{z = E + \mathrm{i}\eta : E \notin [\lambda_-, \lambda_+], n\eta\sqrt{\kappa+\eta} \geqslant n^{\varepsilon}\}$.

The above estimates are uniform in the spectral parameter z and any set of deterministic vectors of cardinality $n^{\mathrm{O}(1)}$.

With Theorem 2.11 as input, we can prove an even stronger estimate on $m(z)$ that is independent of ϕ_n. This averaged local law will give the rigidity of eigenvalues for $\mathcal{Q}_1$ in (2.12). For fixed $\widetilde{\varepsilon} > 0$, we define the following domains

$$\widetilde{S}(\varepsilon, \widetilde{\varepsilon}) := \left\{z = E + \mathrm{i}\eta : \varepsilon \leqslant E \leqslant 1 - \widetilde{\varepsilon}, n^{-1+\varepsilon} \leqslant \eta \leqslant \varepsilon^{-1}\right\}, \quad \widetilde{S}_{out}(\varepsilon, \widetilde{\varepsilon}) := \widetilde{S}(\varepsilon, \widetilde{\varepsilon}) \cap S_{out}(\varepsilon).$$

Note that these two domains are away from $z = 1$.

Theorem 2.12 (Strong averaged local law)**.** *Suppose the assumptions of Theorem 2.5 hold. Then for any fixed $\varepsilon, \widetilde{\varepsilon} > 0$, we have*

$$|m(z) - m_c(z)| \prec (n\eta)^{-1}, \tag{2.46}$$

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

$$|m(z) - m_c(z)| \prec \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}, \tag{2.47}$$

uniformly in $z \in \widetilde{S}_{out}(\varepsilon, \widetilde{\varepsilon})$. These estimates also hold for $(m_\alpha(z) - m_{\alpha c}(z))$, $\alpha = 1,2,3,4$. Finally, given any small constant $0 < \varepsilon_0 < 1 - \lambda_+$, we have

$$\max_{E \geqslant \varepsilon_0} |n(E) - n_c(E)| \prec n^{-1}, \tag{2.48}$$

where

$$n(E) := \frac{1}{q}\#\{\lambda_j \geqslant E\}, \qquad n_c(E) := \int_E^{1-\varepsilon_0} f(x)dx, \tag{2.49}$$

for $f(x)$ defined in (2.7).

The rest of the paper is devoted to proving these main results—Theorems 2.5, 2.7, 2.11 and 2.12.

3 Basic notations and tools

In this preliminary section, we introduce some basic notations and tools that will be used in the proof. First, the following lemma collects basic properties of stochastic domination $\prec$, which will be used tacitly in the proof.

Lemma 3.1 (Lemma 3.2 in [5]). *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.

We have the following lemma, which can be verified through direct calculation using (2.29)-(2.32).

Lemma 3.2. *Fix any constants $c, C > 0$. If (2.6) holds, then for $z \in \mathbb{C}_+ \cap \{z : c \leqslant |z| \leqslant C\}$ we have*

$$\left|z^{-1} - (m_{1c}(z) + m_{2c}(z)) + (z-1)m_{1c}(z)m_{2c}(z)\right| \sim 1, \tag{3.1}$$

and

$$|m_{3c}(z)| \sim |h(z)| \sim 1, \quad 0 \leqslant \operatorname{Im} m_{3c}(z) \sim \begin{cases} \eta/\sqrt{\kappa+\eta}, & \text{if } E \notin [\lambda_-, \lambda_+] \\ \sqrt{\kappa+\eta}, & \text{if } E \in [\lambda_-, \lambda_+] \end{cases}. \tag{3.2}$$

The estimate (3.2) also holds for m_{1c}, $m_{2c}(z)$, $m_{4c}(z)$ and $m_c(z)$.

By (3.1) and (3.2), we have for $z \in S(\varepsilon)$ (recall (2.43)),

$$\begin{aligned} &\|\Pi\| = \mathrm{O}(1), \quad \Psi \gtrsim n^{-1/2}, \quad \Psi^2 \lesssim (n\eta)^{-1}, \\ &\Psi(z) \sim \sqrt{\frac{\operatorname{Im} m_{\alpha c}(z)}{n\eta}} + \frac{1}{n\eta} \ \text{ with } \ \alpha = 1,2,3,4. \end{aligned} \tag{3.3}$$

Note that S_{xx} (resp. S_{yy}) is a standard sample covariance matrix, and it is well-known that its eigenvalues are all inside the support of the Marchenko-Pastur law $[(1-\sqrt{c_1})^2, (1+\sqrt{c_1})^2]$ (resp. $[(1-\sqrt{c_2})^2, (1+\sqrt{c_2})^2]$) with probability $1 - \mathrm{o}(1)$ [3]. Hence both S_{xx}^{-1} and S_{yy}^{-1} behaves well under the assumption (2.6). In our proof, we shall need a slightly stronger probability bound, which is given by the following lemma. Denote the eigenvalues of S_{xx} and S_{yy} by $\lambda_1(S_{xx}) \geqslant \cdots \geqslant \lambda_p(S_{xx})$ and $\lambda_1(S_{yy}) \geqslant \cdots \geqslant \lambda_q(S_{yy})$.

Lemma 3.3. *Suppose Assumption 2.1 holds. Suppose X and Y have bounded support ϕ_n such that $n^{-1/2} \leqslant \phi \leqslant n^{-c_\phi}$ for some constant $c_\phi > 0$. Then for any constant $\varepsilon > 0$, we have with high probability,*

$$(1-\sqrt{c_1})^2 - \varepsilon \leqslant \lambda_p(S_{xx}) \leqslant \lambda_1(S_{xx}) \leqslant (1+\sqrt{c_1})^2 + \varepsilon, \tag{3.4}$$

and

$$(1-\sqrt{c_2})^2 - \varepsilon \leqslant \lambda_q(S_{yy}) \leqslant \lambda_1(S_{yy}) \leqslant (1+\sqrt{c_2})^2 + \varepsilon. \tag{3.5}$$

Proof. Note that X can be written as

$$X = \mathcal{M}_1 \odot \widetilde{X} + \mathcal{M}_2,$$

where $\odot$ denotes the Hadamard product, $\widetilde{X}$ is a $p\times n$ random matrices whose entries are independent random variables that satisfy (2.1), and $\mathcal{M}_1$ and $\mathcal{M}_2$ are $p \times n$ deterministic matrices with $(\mathcal{M}_1)_{ij} = 1 + \mathrm{O}(n^{-2-\tau})$ and $(\mathcal{M}_2)_{ij} = \mathrm{O}(n^{-2-\tau})$. In particular, we have that

$$\|\mathcal{M}_2\| \leqslant \|\mathcal{M}_2\|_F = \mathrm{O}(n^{-1-\tau}). \tag{3.6}$$

Moreover, $\widetilde{X}$ has bounded support $\mathrm{O}(\phi_n)$. Then we claim that for any constant $\varepsilon > 0$,

$$(1-\sqrt{c_1})^2 - \varepsilon \leqslant \lambda_p(\widetilde{X}\widetilde{X}^T) \leqslant \lambda_1(\widetilde{X}\widetilde{X}^T) \leqslant (1+\sqrt{c_1})^2 + \varepsilon \tag{3.7}$$

with high probability. This result essentially follows from [5, Theorem 2.10], although the authors considered the case with $\phi_n \prec n^{-1/2}$ only. The results for more general ϕ_n follows from [9, Lemma 3.12], but only the bounds for the largest eigenvalues are given there in order to avoid the issue with the smallest eigenvalue when c_1 is close to 1. However, under the assumption (2.6), the lower bound for the smallest eigenvalue follows from the exactly the same arguments as in [9]. Hence we omit the details. Now using (3.6), (3.7) and the estimates on the entries of $\mathcal{M}_1$, we conclude (3.4). The estimate (3.5) can be proved in the same way. $\square$

Next we provide a rough bound on the operator norms of the resolvents.

Lemma 3.4. *For $z = E + \mathrm{i}\eta \in \mathbb{C}_+$ such that $c \leqslant |z| \leqslant c^{-1}$ for some constant $c > 0$, we have*

$$\|R(z)\| \leqslant \frac{C}{\eta}, \quad R(z) := \begin{pmatrix} R_1 & -z^{-1/2}R_1\mathcal{H} \\ -z^{-1/2}\mathcal{H}^T R_1 & R_2 \end{pmatrix}, \tag{3.8}$$

and

$$\|G(z)\| \leqslant \frac{C(1+\|S_{xx}^{-1}\| + \|S_{yy}^{-1}\|)}{\eta}, \tag{3.9}$$

for some constant $C > 0$.

Proof. Let $\mathcal{H} = \sum_{k=1}^q \sqrt{\lambda_k}\xi_k\zeta_k^T$ be a singular value decomposition of $\mathcal{H}$, where

$$\lambda_1 \geqslant \ldots \geqslant \lambda_q \geqslant 0 = \lambda_{q+1} = \ldots = \lambda_p,$$

$\{\xi_k\}_{k=1}^p$ are the left-singular vectors, and $\{\zeta_k\}_{k=1}^q$ are the right-singular vectors. Then we have

$$R(z) = \sum_{k=1}^q \frac{1}{\lambda_k - z}\begin{pmatrix} \xi_k\xi_k^T & -z^{-1/2}\sqrt{\lambda_k}\xi_k\zeta_k^T \\ -z^{-1/2}\sqrt{\lambda_k}\zeta_k\xi_k^T & \zeta_k\zeta_k^T \end{pmatrix} - \frac{1}{z}\begin{pmatrix} \sum_{k=q+1}^p \xi_k\xi_k^T & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.10}$$

The estimate (3.8) follows immediately from this representation and the fact that $|\lambda_k - z| \geqslant \eta$. The bound (3.9) holds for $\mathcal{G}_L$ by noticing that

$$\mathcal{G}_L = \begin{pmatrix} S_{xx}^{-1/2} & 0 \\ 0 & S_{yy}^{-1/2} \end{pmatrix} R(z) \begin{pmatrix} S_{xx}^{-1/2} & 0 \\ 0 & S_{yy}^{-1/2} \end{pmatrix}. \tag{3.11}$$

For $\mathcal{G}_R$, $\mathcal{G}_{LR}$ and $\mathcal{G}_{RL}$, stronger bounds hold by (3.11) and (2.24)-(2.25):

$$\|\mathcal{G}_R(z)\| \leqslant \frac{C}{\eta}, \quad \|\mathcal{G}_{LR}(z)\| + \|\mathcal{G}_{RL}(z)\| \leqslant \frac{C(1+\|S_{xx}^{-1/2}\| + \|S_{yy}^{-1/2}\|)}{\eta},$$

where we used $\|S_{xx}^{-1/2}X\| \leqslant 1$ and $\|S_{yy}^{-1/2}Y\| \leqslant 1$. $\square$

One subtle point is that in order to apply Lemma 3.1 (iii) in our proof, we need a bound on the high moments of $\|S_{xx}^{-1}\|$ and $\|S_{yy}^{-1}\|$ (since we will take expectation over the products of many resolvent entries). However instead of using such a bound, we shall regularize the resolvents a little bit in the following way.

Definition 3.5 (Regularized resolvents). *For $z = E + \mathrm{i}\eta \in \mathbb{C}_+$, we define the regularized resolvent $\widehat{G}(z)$ as*

$$\widehat{G}(z) := \left[H(z) - zn^{-10} \begin{pmatrix} I_{p+q} & 0 \\ 0 & 0 \end{pmatrix} \right]^{-1}.$$

Then $\widehat{\mathcal{G}}_L(z)$, $\widehat{\mathcal{G}}_R(z)$, $\widehat{\mathcal{G}}_\alpha(z)$ and $\widehat{m}_\alpha(z)$, $\alpha = 1,2,3,4$, are defined in the obvious way. We define

$$\widehat{\mathcal{H}} := \widehat{S}_{xx}^{-1/2} S_{xy} \widehat{S}_{yy}^{-1/2}, \quad \widehat{S}_{xx} := S_{xx} + n^{-10}, \quad \widehat{S}_{yy} := S_{yy} + n^{-10}.$$

Then $\widehat{R}_1$, $\widehat{R}_2$ and $\widehat{m}(z)$ are defined in the obvious way. We also define $\widehat{R}(z)$ and the spectral decomposition

$$\begin{aligned} \widehat{R}(z) &:= \begin{pmatrix} \widehat{R}_1 & -z^{-1/2}\widehat{R}_1\widehat{\mathcal{H}} \\ -z^{-1/2}\widehat{\mathcal{H}}^T\widehat{R}_1 & \widehat{R}_2 \end{pmatrix} \\ &= \sum_{k=1}^q \frac{1}{\widehat{\lambda}_k - z} \begin{pmatrix} \widehat{\xi}_k\widehat{\xi}_k^T & -z^{-1/2}\sqrt{\widehat{\lambda}_k}\widehat{\xi}_k\widehat{\zeta}_k^T \\ -z^{-1/2}\sqrt{\widehat{\lambda}_k}\widehat{\zeta}_k\widehat{\xi}_k^T & \widehat{\zeta}_k\widehat{\zeta}_k^T \end{pmatrix} - \frac{1}{z}\begin{pmatrix} \sum_{k=q+1}^p \widehat{\xi}_k\widehat{\xi}_k^T & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned} \tag{3.12}$$

By Schur complement formula, we have

$$\widehat{\mathcal{G}}_L = \begin{pmatrix} \widehat{S}_{xx}^{-1/2}\widehat{R}_1\widehat{S}_{xx}^{-1/2} & -z^{-1/2}\widehat{S}_{xx}^{-1/2}\widehat{R}_1\widehat{\mathcal{H}}\widehat{S}_{yy}^{-1/2} \\ -z^{-1/2}\widehat{S}_{yy}^{-1/2}\widehat{\mathcal{H}}^T\widehat{R}_1\widehat{S}_{xx}^{-1/2} & \widehat{S}_{yy}^{-1/2}\widehat{R}_2\widehat{S}_{yy}^{-1/2} \end{pmatrix}, \tag{3.13}$$

$$\widehat{\mathcal{G}}_R = \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix} + \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}\begin{pmatrix} X^T & 0 \\ 0 & Y^T \end{pmatrix}\widehat{\mathcal{G}}_L\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}, \tag{3.14}$$

and

$$\begin{aligned} \widehat{\mathcal{G}}_{LR}(z) &= -\widehat{\mathcal{G}}_L(z)\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}, \\ \widehat{\mathcal{G}}_{RL}(z) &= -\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}\begin{pmatrix} X^T & 0 \\ 0 & Y^T \end{pmatrix}\widehat{\mathcal{G}}_L(z). \end{aligned} \tag{3.15}$$

With a straightforward calculation and using (2.28), we obtain that

$$\widehat{m}_3(z) - \widehat{m}_4(z) = (1-z)(c_1 - c_2) + z(1-z)n^{-11}\left(\operatorname{Tr}\widehat{\mathcal{G}}_1(z) - \operatorname{Tr}\widehat{\mathcal{G}}_2(z)\right). \tag{3.16}$$

For the regularized resolvents, it is easy to prove the following result using the same argument for the proof of Lemma 3.4.

Lemma 3.6. *For $z = E + \mathrm{i}\eta \in \mathbb{C}_+$ such that $c \leqslant |z| \leqslant c^{-1}$ for some constant $c > 0$, (3.8) and (3.9) hold for $\widehat{G}(z)$ and $\widehat{R}(z)$. Moreover, we have*

$$\left\|\widehat{G}(z)\right\| \leqslant \frac{Cn^{10}}{\eta}, \tag{3.17}$$

for some constant $C > 0$.

In the proof, we will take $\eta \gg n^{-1}$, and the deterministic bound (3.17) then justifies the application of Lemma 3.1 (iii) when we calculate the expectation of polynomials of the entries of $\widehat{G}(z)$. For simplicity of presentation, we will not repeat this again in the proof.

Remark 3.7. The results for $\widehat{G}(z)$ can be extended to $G(z)$ with a standard perturbative argument. We will show that there exists a high probability event Ξ on which $\|\widehat{G}(z)\|_{\max} = \mathrm{O}(1)$ for z in some bounded regime. Then we define

$$G_t(z) := \left[H(z) - tzn^{-10}\begin{pmatrix} I_{p+q} & 0 \\ 0 & 0 \end{pmatrix}\right]^{-1}, \quad G_0(z) \equiv G(z), \quad G_1(z) \equiv \widehat{G}(z).$$

Taking the derivative, we get

$$\partial_t G_t(z) = zn^{-10} G_t(z) \begin{pmatrix} I_{p+q} & 0 \\ 0 & 0 \end{pmatrix} G_t(z). \tag{3.18}$$

Thus applying the Gronwall's inequality to

$$\|G_t(z)\|_{\max} \leqslant \|\widehat{G}(z)\|_{\max} + Cn^{-9}\int_t^1 \|G_s(z)\|_{\max}^2 \mathrm{d}s,$$

we can obtain that $\|G_t(z)\|_{\max} \leqslant C$ for all $0 \leqslant t \leqslant 1$ on Ξ. Then using (3.18) again, we obtain that $\|G(z) - \widehat{G}(z)\|_{\max} \leqslant n^{-8}$ on Ξ. Such a small error will not affect any of our results. We emphasize that the above argument is purely deterministic on Ξ, so we do not lose any probability.

We record the following resolvent estimates, which will be used in the proof of Theorem 2.11.

Lemma 3.8. *For any deterministic unit $\mathbf{v}_\alpha \in \mathbb{C}^{\mathcal{I}_\alpha}$, $\alpha = 1, 2$, we have for $\beta = 1, 2, 3, 4$,*

$$\sum_{\mathfrak{a} \in \mathcal{I}_\beta} |G_{\mathfrak{a}\mathbf{v}_\alpha}(z)|^2 = \sum_{\mathfrak{a} \in \mathcal{I}_\beta} |G_{\mathbf{v}_\alpha\mathfrak{a}}(z)|^2 \prec |G_{\mathbf{v}_\alpha\mathbf{v}_\alpha}(z)| + \frac{\operatorname{Im} G_{\mathbf{v}_\alpha\mathbf{v}_\alpha}(z)}{\eta}, \quad z = E + \mathrm{i}\eta. \tag{3.19}$$

For any deterministic unit $\mathbf{v}_\beta \in \mathbb{C}^{\mathcal{I}_\beta}$, $\beta = 3, 4$, we have for $\alpha = 1, 2, 3, 4$,

$$\sum_{\mathfrak{a} \in \mathcal{I}_\alpha} \left|G_{\mathfrak{a}\mathbf{v}_\beta}\right|^2 \prec 1 + \frac{\operatorname{Im}(\mathcal{U}\mathcal{G}_R)_{\mathbf{v}_\beta\mathbf{v}_\beta}}{\eta}, \quad \sum_{\mathfrak{a} \in \mathcal{I}_\alpha} \left|G_{\mathbf{v}_\beta\mathfrak{a}}\right|^2 \prec 1 + \frac{\operatorname{Im}(\mathcal{G}_R\mathcal{U}^T)_{\mathbf{v}_\beta\mathbf{v}_\beta}}{\eta}, \tag{3.20}$$

where

$$\mathcal{U} := z^{1/2}\begin{pmatrix} \bar{z} I_n & \bar{z}^{1/2} I_n \\ \bar{z}^{1/2} I_n & \bar{z} I_n \end{pmatrix}\begin{pmatrix} z I_n & z^{1/2} I_n \\ z^{1/2} I_n & z I_n \end{pmatrix}^{-1}.$$

Similar estimates hold for $\widehat{G}$.

Proof. First, we prove some simple resolvent estimates on $R(z)R^*(z)$ and $R^*(z)R(z)$ for $R(z)$ in (3.8). Using spectral decomposition (3.10), for any vector $\mathbf{w} \in \mathbb{C}^p$ and $z = E + \mathrm{i}\eta$, we have

$$\mathbf{w}^* R_1^*(z) R_1(z)\, \mathbf{w} = \mathbf{w}^* R_1(z) R_1^*(z)\, \mathbf{w} = \sum_{k=1}^p \frac{|\langle \mathbf{w}, \xi_k\rangle|^2}{|\lambda_k - E|^2 + \eta^2} = \frac{\operatorname{Im}(R_1(z))_{\mathbf{w}\mathbf{w}}}{\eta}. \tag{3.21}$$

For R_2, we have a similar estimate. Notice that with Schur complement, we can write

$$R(z) = z^{-1/2}\left(\mathscr{H} - z^{1/2}\right)^{-1}, \quad \mathscr{H} = \begin{pmatrix} 0 & -\mathcal{H} \\ -\mathcal{H}^T & 0 \end{pmatrix}.$$

Then using a same argument as in (3.21), we obtain that for any vector $\mathbf{w} \in \mathbb{C}^{p+q}$,

$$\mathbf{w}^* R^*(z)R(z)\,\mathbf{w} = |z|^{-1}\frac{\operatorname{Im}\left(z^{1/2}\,\mathbf{w}^*\,R(z)\,\mathbf{w}\right)}{\operatorname{Im} z^{1/2}}. \tag{3.22}$$

We first pick $\alpha = 1$ and $\beta = 1, 2$. Using (2.23) and Lemma 3.3, we get

$$\begin{aligned}\sum_{\mathfrak{a}\in\mathcal{I}_\beta} |G_{\mathfrak{a}\mathbf{v}_\alpha}(z)|^2 &\leqslant \sum_{\mathfrak{a}\in\mathcal{I}_1\cup\mathcal{I}_2} |G_{\mathfrak{a}\mathbf{v}_\alpha}(z)|^2 = (\mathcal{G}_L^*\mathcal{G}_L)_{\mathbf{v}_\alpha\mathbf{v}_\alpha} \prec \left(\mathbf{v}_\alpha^* S_{xx}^{-1/2}, 0\right) R^*(z)R(z)\begin{pmatrix} S_{xx}^{-1/2}\mathbf{v}_\alpha \\ 0\end{pmatrix} \\ &= \mathbf{v}_\alpha^* S_{xx}^{-1/2}\left[R_1^*(z)R_1(z) + |z|^{-1}R_1^*(z)\left(\mathcal{H}\mathcal{H}^T\right)R_1(z)\right]S_{xx}^{-1/2}\mathbf{v}_\alpha \\ &= \left(1+\frac{\overline{z}}{|z|}\right)\mathbf{v}_\alpha^* S_{xx}^{-1/2}R_1^*(z)R_1(z)S_{xx}^{-1/2}\mathbf{v}_\alpha + |z|^{-1}\mathbf{v}_\alpha^* S_{xx}^{-1/2}R_1(z)S_{xx}^{-1/2}\mathbf{v}_\alpha \\ &= \left(1+\frac{\overline{z}}{|z|}\right)\frac{\operatorname{Im} G_{\mathbf{v}_\alpha\mathbf{v}_\alpha}}{\eta} + |z|^{-1}G_{\mathbf{v}_\alpha\mathbf{v}_\alpha},\end{aligned}$$

where in the third step we used (3.11), in the fourth step $\mathcal{H}\mathcal{H}^T = (R_1^*(z))^{-1} + \overline{z}$, and in the last step (3.11) and (3.21). For $\alpha = 1$ and $\beta = 3, 4$, the proof is similar except that we need to use (2.25) and

$$\left\|\begin{pmatrix} X & 0 \\ 0 & Y\end{pmatrix}\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n\end{pmatrix}\begin{pmatrix} \overline{z}I_n & \overline{z}^{1/2}I_n \\ \overline{z}^{1/2}I_n & \overline{z}I_n\end{pmatrix}\begin{pmatrix} X^T & 0 \\ 0 & Y^T\end{pmatrix}\right\| \prec 1$$

by Lemma 3.3. For $\alpha = 2$, the proof is the same. This concludes (3.19).

Then we consider the case $\alpha = 3$ and $\beta = 3, 4$. Using (2.24), (3.11), Lemma 3.3 and (3.22), we get that

$$\sum_{\mathfrak{a}\in\mathcal{I}_\beta} |G_{\mathfrak{a}\mathbf{v}_\alpha}(z)|^2 \leqslant \sum_{\mathfrak{a}\in\mathcal{I}_3\cup\mathcal{I}_4} |G_{\mathfrak{a}\mathbf{v}_\alpha}(z)|^2 \prec 1 + \mathbf{w}_\alpha^*\,R^*(z)R(z)\,\mathbf{w}_\alpha \lesssim \frac{\operatorname{Im}\left(z^{1/2}\,\mathbf{w}^*\,R(z)\,\mathbf{w}\right)}{\eta},$$

where

$$\mathbf{w}_\alpha := \begin{pmatrix} S_{xx}^{-1/2}X & 0 \\ 0 & S_{yy}^{-1/2}Y\end{pmatrix}\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n\end{pmatrix}\mathbf{v}_\alpha.$$

By (2.24), we have

$$z^{1/2}\,\mathbf{w}^*\,R(z)\,\mathbf{w} = \mathbf{v}_\alpha^*\mathcal{U}\left[\mathcal{G}_R - \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n\end{pmatrix}\right]\mathbf{v}_\alpha.$$

Then using

$$\operatorname{Im}\mathbf{v}_\alpha^*\mathcal{U}\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n\end{pmatrix}\mathbf{v}_\alpha = \mathrm{O}(\eta),$$

we conclude (3.20). □

The anisotropic local law (2.42) together with the rigidity estimate (2.12) implies the following delocalization properties of eigenvectors.

Lemma 3.9 (Isotropic delocalization of eigenvectors)**.** *Suppose* (2.12) *hold, and* (2.42) *holds for G. Then for any fixed $\delta > 0$ and any deterministic unit vectors $\mathbf{u}_\alpha \in \mathbb{C}^{\mathcal{I}_\alpha}$, $\alpha = 1, 2, 3, 4$, the following estimates hold:*

$$\left|\left\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_k\right\rangle\right|^2 + \left|\left\langle \mathbf{u}_2, S_{yy}^{-1/2}\zeta_k\right\rangle\right|^2 \prec n^{-1}, \quad 1 \leqslant k \leqslant q, \tag{3.23}$$

and

$$\left|\langle \mathbf{u}_3, X^T S_{xx}^{-1/2}\xi_k\rangle\right|^2 + \left|\langle \mathbf{u}_4, Y^T S_{yy}^{-1/2}\zeta_k\rangle\right|^2 \prec n^{-1}, \quad 1 \leqslant k \leqslant q. \tag{3.24}$$

If (2.12) *only holds for* $i \leqslant (1-\varepsilon)q$, *then* (3.23) *and* (3.24) *hold for* $1 \leqslant k \leqslant (1-\varepsilon)q$.

Proof. Choose $z_0 = E + \mathrm{i}\eta_0 \in S(\varepsilon)$ with $\eta_0 = n^{-1+\varepsilon}$. By (2.42) for G, we have $\mathrm{Im}\langle \mathbf{u}_1, G(z_0)\mathbf{u}_1\rangle = \mathrm{O}(1)$ with high probability. Then using (2.23) and the spectral decomposition (3.10), we get

$$\sum_{k=1}^{q} \frac{\eta_0 |\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_k\rangle|^2}{(\lambda_k - E)^2 + \eta_0^2} = \mathrm{Im}\,\langle \mathbf{u}_1, G(z_0)\mathbf{u}_1\rangle = \mathrm{O}(1) \quad \text{with high probability.} \tag{3.25}$$

By (2.12), we have that $\lambda_k + \mathrm{i}\eta_0 \in S(\varepsilon)$ with high probability. Then choosing $E = \lambda_k$ in (3.25) yields that

$$|\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_k\rangle|^2 \lesssim \eta_0 \quad \text{with high probability.}$$

Since ε is arbitrary, we get $|\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_k\rangle|^2 \prec n^{-1}$. In a similar way, we can prove $\left|\langle \mathbf{u}_2, S_{yy}^{-1/2}\zeta_k\rangle\right|^2 \prec \eta_0$.

Now for $z_0 = \lambda_k + \mathrm{i}\eta_0 \in S(\varepsilon)$, we denote

$$\widetilde{\mathbf{u}}_3 := \begin{pmatrix} z_0 I_n & z_0^{1/2} I_n \\ z_0^{1/2} I_n & z_0 I_n \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{u}_3 \\ 0 \end{pmatrix}.$$

Note that z_0 is well-separated from 0 and 1 by a distance of order 1 by (2.12), so we have $\|\widetilde{\mathbf{u}}_3\|_2 = \mathrm{O}(1)$. By (2.42), we have $\mathrm{Im}\,\widetilde{\mathbf{u}}_3^T G(z_0)\widetilde{\mathbf{u}}_3 = \mathrm{O}(1)$ with high probability. Using (2.24) and the spectral decomposition (3.10), we get that with high probability,

$$\eta_0^{-1}|\langle \mathbf{u}_3, X^T S_{xx}^{-1/2}\xi_k\rangle|^2 \leqslant \sum_{l=1}^{q} \frac{\eta_0|\langle \mathbf{u}_3, X^T S_{xx}^{-1/2}\xi_l\rangle|^2}{(\lambda_l - E)^2 + \eta_0^2} = \mathrm{Im}\,\widetilde{\mathbf{u}}_3^T G(z_0)\widetilde{\mathbf{u}}_3 + \mathrm{O}(1).$$

This gives $\left|\langle \mathbf{u}_3, X^T S_{xx}^{-1/2}\xi_k\rangle\right|^2 \prec n^{-1}$. Similarly, we get $\left|\langle \mathbf{u}_4, Y^T S_{yy}^{-1/2}\zeta_k\rangle\right|^2 \prec n^{-1}$. ☐

The second moment of the error $\langle \mathbf{u}, (G(z) - \Pi(z))\mathbf{v}\rangle$ in fact satisfies a stronger bound. It will be used in the proof of Theorem 2.7.

Lemma 3.10. *Suppose the assumptions of Theorem 2.5 hold. Then for any fixed* $\varepsilon > 0$, *we have*

$$\mathbb{E}|\langle \mathbf{u}, G(z)\mathbf{v}\rangle - \langle \mathbf{u}, \Pi(z)\mathbf{v}\rangle|^2 \prec \Psi^2(z), \tag{3.26}$$

for any $z \in S(\varepsilon)$ *(recall* (2.40)*) and deterministic unit vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$.

4 Proof of Theorem 2.5, Theorem 2.12 and Lemma 3.10

In this section, we prove Theorem 2.5, Theorem 2.12 and Lemma 3.10 using Theorem 2.11, whose proof is postponed to Sections 6-7. The following proofs will use a comparison argument developed in [25] for Wigner matrices, which was later extended to sample covariance matrices [9] and separable covariance matrices [32]. This argument can be extended to our setting without difficulties, where the only inputs are the linearization in (2.19) and the anisotropic local law, Theorem 2.11. Hence we will not give all the details, and only focus on the part that is significantly different from the pervious works.

Given any random matrices X and Y satisfying the assumptions in Theorem 2.5, we can construct matrices $\widetilde{X}$ and $\widetilde{Y}$ that match the first four moments as X and Y but with smaller support $\phi_n \prec n^{-1/2}$, which is the content of the next lemma.

Lemma 4.1 (Lemma 5.1 in [25])**.** *Suppose X and Y satisfy the assumptions in Theorem 2.5. Then there exists another matrix $\widetilde X = (\widetilde x_{ij})$ and $\widetilde Y = (\widetilde y_{ij})$, such that $\widetilde X$ and $\widetilde Y$ satisfy the bounded support condition (2.10) with $\phi_n \prec n^{-1/2}$, and the following moments matching holds:*

$$\mathbb{E}x_{ij}^k = \mathbb{E}\widetilde x_{ij}^k, \quad \mathbb{E}y_{ij}^k = \mathbb{E}\widetilde y_{ij}^k, \quad k = 1,2,3,4. \tag{4.1}$$

We can define $\widetilde H(z)$ and $\widetilde G(z)$ by replacing (X,Y) with $(\widetilde X,\widetilde Y)$. Of course $\widetilde{\mathcal G}_L(z)$, $\widetilde{\mathcal G}_R(z)$, $\widetilde m_\alpha(z)$, $\widetilde{\mathcal H}$, $\widetilde R_{1,2}$, etc. can be defined in the obvious way.

Proof of Lemma 3.10. By Theorem 2.11, we see that (3.26) hold for $\widetilde G(z)$ using (3.3). Thus Lemma 3.10 follows immediately from the following comparison lemma.

Lemma 4.2. *Let (X,Y) and $(\widetilde X,\widetilde Y)$ be pairs of random matrices defined as above. Suppose Theorem 2.11 holds for both $G(z)$ and $\widetilde G(z)$. For any small constant $\varepsilon > 0$, we have that*

$$\mathbb{E}\left|\langle \mathbf{u}, G(z)\mathbf{v}\rangle - \langle \mathbf{u}, \widetilde G(z)\mathbf{v}\rangle\right|^2 \prec \Psi^2(z), \tag{4.2}$$

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

The proof of this lemma is the same as the one for Lemma 3.7 in [32, Section 7] and the one for Lemma 3.8 in [25, Section 6]. The only inputs are Theorem 2.11 and the moment matching conditions in (4.1). Hence we omit the details. □

We have a similar comparison lemma for the estimates (2.44) and (2.45). Notice that (2.46) and (2.47) holds for $\widetilde G(z)$ since

$$\Psi^2(z) \lesssim (n\eta)^{-1}, \quad \text{and} \quad \Psi^2(z) \lesssim \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}} \quad \text{for} \quad z \in S_{out}(\varepsilon),$$

by (3.2).

Lemma 4.3. *Let (X,Y) and $(\widetilde X,\widetilde Y)$ be pairs of random matrices defined as above. Fix any small constant $\varepsilon > 0$. For $z \in S(\varepsilon)$ or $z \in S_{out}(\varepsilon)$, if there exist deterministic quantities $J \equiv J(N)$ and $K \equiv K(N)$ such that $J \leqslant n^{-c}$ and $K \leqslant n^{-c}$ for some constant $c > 0$, and*

$$\widetilde G(z) - \Pi = \mathrm{O}_\prec(J), \quad |\widetilde m_\alpha(z) - m_{\alpha c}(z)| \prec K, \quad \alpha = 1,2,3,4. \tag{4.3}$$

Then we have

$$|m_\alpha(z) - m_{\alpha c}(z)| \prec \Psi^2(z) + J^2 + K, \quad \alpha = 1,2,3,4. \tag{4.4}$$

Proof. The proof is similar to the one for [25, Lemma 5.4] or [32, Lemma 7.1] (the latter is closer to our current setting and we can copy its proof almost verbatim). Hence we omit the details. □

Now we are ready to give the proof of Theorem 2.12 using this lemma.

Proof of Theorem 2.12. By Theorem 2.11 for $\widetilde G$ with $\phi_n = n^{-1/2}$, one can choose $J = \Psi(z)$ and

$$K = \frac{1}{n\eta}, \quad \text{or} \quad \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}} \quad \text{for} \quad z \in S_{out}(\varepsilon).$$

Then using (4.4) and $|m(z) - m_c(z)| \lesssim |1-z|^{-1}|m_3(z) - m_{3c}(z)|$ by (2.26), we get (2.46) and (2.47). Note that due to the $|1-z|^{-1}$ factor, we need to stay away from $z = 1$, which is the main reason why we need to restrict ourself to the domain $\widetilde S(\varepsilon,\widetilde\varepsilon)$ or $\widetilde S_{out}(\varepsilon,\widetilde\varepsilon)$. The estimate (2.48) follows from (2.46) through a standard argument, see e.g. the proofs for [13, Theorems 2.12-2.13], [14, Theorem 2.2] or [28, Theorem 3.3]. □

Next we give the proof of Theorem 2.5. We first prove the rigidity result (2.12).

Proof of (2.12). Without loss of generality, we only consider the case $\lambda_- \gtrsim 1$ in the proof. For the case with $\lambda_- = \mathrm{o}(1)$, since we only need to prove a weaker result with $1 \leqslant i \leqslant (1-\varepsilon)q$, the proof is the same except that we do not need to provide the bound in (4.6) below.

Using (2.48) and the method in [13, 14], we can prove the following rigidity estimate: for any fixed $\delta > 0$ and all $n^\delta \leqslant i \leqslant q - n^\delta$, (2.12) holds. To obtain this estimate for the largest and smallest n^δ eigenvalues, we still need to provide the following upper and lower bounds: for any constant $\varepsilon > 0$,

$$\lambda_1 \leqslant \lambda_+ + n^{-2/3+\varepsilon} \quad \text{with high probability,} \tag{4.5}$$

and

$$\lambda_q \geqslant \lambda_- - n^{-2/3+\varepsilon} \quad \text{with high probability.} \tag{4.6}$$

Given these bounds, the estimate (2.48) and the method in [13, 14] allow us to conclude (2.12) for all i.

First, we claim that for any small constants $c, \varepsilon > 0$, with high probability,

$$\#\{i : \lambda_i \in [\lambda_+ + n^{-2/3+\varepsilon}, 1-c]\} = 0, \quad \text{and} \quad \#\{i : \lambda_i \in [c, \lambda_- - n^{-2/3+\varepsilon}]\} = 0. \tag{4.7}$$

We choose $\eta = n^{-2/3}$ and $E = \lambda_+ + \kappa \leqslant 1 - c$ outside of the spectrum with some $\kappa \geqslant n^{-2/3+2\varepsilon} \gg n^\varepsilon \eta$. Then using (2.47), we get that

$$|\operatorname{Im} m(z) - \operatorname{Im} m_c(z)| \prec \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}} \lesssim \frac{n^{-\varepsilon}}{n\eta}. \tag{4.8}$$

On the other hand, if there is an eigenvalue λ_j satisfying $|\lambda_j - E| \leqslant \eta$ for some $1 \leqslant j \leqslant n$, then

$$\operatorname{Im} m(z) = \frac{1}{q}\sum_{i=1}^{q} \frac{\eta}{|\lambda_i - E|^2 + \eta^2} \gtrsim \frac{1}{n\eta}. \tag{4.9}$$

On the other hand, by (3.2) we have

$$\operatorname{Im} m_c(z) = \mathrm{O}\left(\frac{\eta}{\sqrt{\kappa+\eta}}\right) = \mathrm{O}\left(\frac{n^{-\varepsilon}}{n\eta}\right).$$

Together with (4.9), this contradicts (4.8). Hence we conclude the first estimate in (4.7) since ε can be arbitrarily chosen. The second estimate in (4.7) can be proved in the same way by choosing $E = \lambda_- - \kappa$.

Then it remains to prove that for a sufficiently small constant $c > 0$, with high probability,

$$\#\{i : \lambda_i \in [1-c, 1]\} = 0 \quad \text{and} \quad \#\{i : \lambda_i \in [0, c]\} = 0. \tag{4.10}$$

We pick i.i.d. Gaussian X^G and Y^G, which are independent of the matrices X and Y we are considering. We denote the eigenvalues of $\mathcal{C}_{X^G Y^G}$ by $\lambda_1^G \geqslant \lambda_2^G \geqslant \cdots \geqslant \lambda_p^G$. Then with Lemma 1 in Section 8.2 of [19], we know that $|\lambda_1^G - \lambda_+| \prec n^{-2/3}$, which implies

$$\#\{i : \lambda_i^G \in [\lambda_+ + n^{-2/3+\varepsilon}, 1]\} = 0 \quad \text{with high probability.} \tag{4.11}$$

Now we define a continuous path of random matrices as

$$X_t := \sqrt{1-t}X^G + \sqrt{t}X, \quad Y_t := \sqrt{1-t}Y^G + \sqrt{t}Y, \quad t \in [0,1]. \tag{4.12}$$

Correspondingly, we define $H_t(z)$ and $G_t(z)$ by replacing (X, Y) with (X_t, Y_t) in the definitions (2.19) and (2.20). We denote the eigenvalues of $\mathcal{C}_{X_tY_t}$ by λ_i^t. We claim that with high probability,

$$\text{for any } 1 \leqslant i \leqslant q, \quad \lambda_i^t \text{ is continuous in } t \text{ for all } t \in [0,1]. \tag{4.13}$$

and

$$\|G_t(1-c)\|_{\max} \text{ is finite for all } t \in [0,1]. \tag{4.14}$$

Recall that with high probability, the eigenvalues of $\mathcal{C}_{X_0Y_0}$ are all inside $[0, \lambda_+ + n^{-2/3+\varepsilon}]$. Moreover, if (4.14) holds, then

$$m_t(1-c) = \frac{1}{q}\sum_{i=1}^{q} \frac{1}{\lambda_i^t - (1-c)}$$

is finite for all $t \in [0,1]$, which means that there is no eigenvalue λ_i^t crossing the point $E = 1-c$ for all $t \in [0,1]$. Hence using the continuity of eigenvalues in (4.13), we conclude the first estimate in (4.10), which, together with (4.7), concludes (4.5).

By the definition of $\mathcal{C}_{X_tY_t}$, to prove (4.13), it suffices to prove that with high probability, $(X_tX_t^T)^{-1}$ and $(Y_tY_t^T)^{-1}$ are continuous in t for all $t \in [0,1]$. For this purpose, we only need to show that with high probability,

$$X_tX_t^T \text{ and } Y_tY_t^T \text{ are non-singular for all } t \in [0,1].$$

We consider discrete times $t_k = kn^{-10}$. Note that X_t satisfies the assumptions of Lemma 3.3. Hence with a simple union bound we get that there exists a high probability event Ξ_1 such that

$$\mathbf{1}(\Xi_1) \min_{0\leqslant k\leqslant n^{10}} \lambda_p(X_{t_k}X_{t_k}^T) \geqslant \mathbf{1}(\Xi_1)\frac{1}{2}(1-\sqrt{c_1})^2.$$

Moreover, using the bounded support condition for X, we get that there exists a high probability event Ξ_2 such that

$$\mathbf{1}(\Xi_2)\max_{i,\mu} |(X_t)_{i\mu}| \leqslant 1 \ \Rightarrow\ \mathbf{1}(\Xi_2)\sup_{0\leqslant t\leqslant 1} \|X_t\| = \mathrm{O}(n). \tag{4.15}$$

This implies

$$\sup_{t_{k-1}\leqslant t\leqslant t_k} \|X_tX_t^T - X_{t_k}X_{t_k}^T\| \lesssim n^{-5}\cdot n^2 = n^{-3}$$

Therefore, on the event $\Xi_1 \cap \Xi_2$ we have

$$\inf_{0\leqslant t\leqslant 1} \lambda_p(X_tX_t^T) = \min_{1\leqslant k\leqslant n^{10}} \inf_{t_{k-1}\leqslant t\leqslant t_k} \lambda_p(X_tX_t^T) \geqslant \frac{1}{2}(1-\sqrt{c_1})^2 - \mathrm{O}\left(n^{-3}\right) \gtrsim 1.$$

We have a similar estimate for $Y_tY_t^T$. This concludes (4.13).

To prove (4.14), we consider discrete times $t_k = kn^{-100}$. Note that X_t and Y_t satisfy the assumptions of Theorem 2.11, hence the local law (2.42) holds for any $t \in [0,1]$. We claim that there exists a high probability event Ξ, on which

$$\max_{0\leqslant k\leqslant n^{100}} \|\widehat{G}_{t_k}(1-c+\mathrm{i}n^{-10})\|_{\max} = \mathrm{O}(1), \tag{4.16}$$

where we recall that $\widehat{G}$ is defined in Definition 3.5. Now suppose (4.16) holds. With the deterministic bound (3.17) and (4.15), we have that for any $t_{k-1} \leqslant t \leqslant t_k$,

$$\begin{aligned}&\left|\widehat{G}_t(1-c+\mathrm{i}n^{-10}) - \widehat{G}_{t_k}(1-c+\mathrm{i}n^{-10})\right| \\ &\leqslant Cn^{-50}\|\widehat{G}_t(1-c+\mathrm{i}n^{-10})\|\left(\|X\| + \|X^G\|\right)\|\widehat{G}_{t_k}(1-c+\mathrm{i}n^{-10})\|\end{aligned}$$

$$\leqslant n^{-50} \cdot \left(Cn^{20}\right)^2 \cdot n = \mathrm{O}(n^{-9}), \quad \text{on } \Xi_2.$$

Thus we conclude that on $\Xi \cap \Xi_2$,

$$\max_{0 \leqslant t \leqslant 1} \|\widehat{G}_t(1 - c + \mathrm{i}n^{-10})\|_{\max} \leqslant C.$$

Finally, the perturbation argument in Remark 3.7 allows us to remove the $\mathrm{i}n^{-10}$ and the regularization in $\widehat{G}$, which gives (4.14) on Ξ.

It remains to prove (4.16). Since X_t and Y_t also satisfy the assumptions of Theorem 2.5, by (4.7) we know that for any fixed t, the eigenvalues λ_i^t are either inside $[0, \lambda_+ + n^{-2/3+\varepsilon}]$ or $[1 - c/2, 1]$ with high probability. With a simple union bound, we obtain that

$$\min_{0 \leqslant k \leqslant n^{100}} \min_{1 \leqslant i \leqslant p} |(1 - c) - \lambda_i^{t_k}| \gtrsim 1 \quad \text{with high probability.}$$

Together with (3.10), this immediately gives that

$$\max_{0 \leqslant k \leqslant n^{100}} \|R_{t_k}(z)\| \leqslant C, \quad z = 1 - c + \mathrm{i}n^{-10}.$$

Combining this bound with (2.23)-(2.25) and Lemma 3.3, we get

$$\max_{0 \leqslant k \leqslant n^{100}} \|G_{t_k}(z)\| \leqslant C, \quad z = 1 - c + \mathrm{i}n^{-10}.$$

Finally, applying the arguments in Remark 3.7 gives (4.16) for $\widehat{G}$. This concludes (4.14), which further gives the first estimate in (4.10).

Finally, the second estimate in (4.10) can be proved in the same way using the continuous interpolation in (4.12), except that we still need to provide a similar estimate as in (4.11) for the smallest eigenvalues in the Gaussian case: there exists a constant $c > 0$ such that

$$\#\{i : \lambda_i^G \in [0, c]\} = 0 \quad \text{with high probability.} \tag{4.17}$$

In fact, it is known that the eigenvalues of $\mathcal{C}_{X^G Y^G}$ reduce to those of the double Wishart matrices [22], that is, the eigenvalues of $(\mathcal{W}_1 + \mathcal{W}_2)^{-1}\mathcal{W}_1$, where $\mathcal{W}_1 \sim W_q(p, I)$ (i.e. $\mathcal{W}_1$ is a $q \times q$ Wishart matrix with p samples) and $\mathcal{W}_1 \sim W_q(n - p, I)$ (i.e. $\mathcal{W}_2$ is a $q \times q$ Wishart matrix with $(n - p)$ samples). Note that we have $1 - q/p \gtrsim 1$ and $1 - q/(n - p) \gtrsim 1$ under (2.6) and the assumption that $\lambda_- \gtrsim 1$. Hence by Lemma 3.3, we have

$$\lambda_q^G \gtrsim \frac{\lambda_q(\mathcal{W}_1)}{\lambda_1(\mathcal{W}_1) + \lambda_1(\mathcal{W}_2)} \gtrsim 1 \quad \text{with high probability.}$$

This gives (4.17), which further concludes the second estimate in (4.10). □

Finally we prove (2.13), which will conclude Theorem 2.5.

Proof of (2.13). The proof is similar to the one for [9, Theorem 3.16], so we only outline the argument. For the matrices $\widetilde{X}$ and $\widetilde{Y}$ constructed in Lemma 4.1, the Tracy-Widom limit of $\mathcal{C}_{\widetilde{X}\widetilde{Y}}$ has been proved in [19].

Lemma 4.4 (Theorem 2.1 of [19])**.** *Let X and Y be random matrices satisfying the assumptions in Theorem 2.5 and the bounded support condition with $\phi_n \prec n^{-1/2}$. Then* (2.13) *holds.*

Now it is easy to see that (2.13) in the general case follows from the following comparison lemma.

Lemma 4.5. *Let (X, Y) and $(\widetilde{X}, \widetilde{Y})$ be two pairs of random matrices as in Lemma 4.1. Then for any fixed k, there exist constants $\varepsilon, \delta > 0$ such that, for all $s_1, s_2, \ldots, s_k \in \mathbb{R}$, we have*

$$\begin{aligned}\widetilde{\mathbb{P}}\left(\left(n^{\frac{2}{3}}(\lambda_i - \lambda_+) \leqslant s_i - n^{-\varepsilon}\right)_{1 \leqslant i \leqslant k}\right) - n^{-\delta} \leqslant \mathbb{P}\left(\left(n^{\frac{2}{3}}(\lambda_i - \lambda_+) \leqslant s_i\right)_{1 \leqslant i \leqslant k}\right) \\ \leqslant \widetilde{\mathbb{P}}\left(\left(n^{\frac{2}{3}}(\lambda_i - \lambda_+) \leqslant s_i + n^{-\varepsilon}\right)_{1 \leqslant i \leqslant k}\right) + n^{-\delta},\end{aligned} \tag{4.18}$$

where $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ denote the laws for (X, Y) and $(\widetilde{X}, \widetilde{Y})$, respectively.

To prove Lemma 4.5, it suffices to prove the following Green's function comparison result. Its proof is the same as the ones for [25, Lemma 5.5] and [9, Lemma 5.5], so we omit the details.

Lemma 4.6. *Let (X, Y) and $(\widetilde{X}, \widetilde{Y})$ be two pairs of random matrices as in Lemma 4.1. Suppose $F : \mathbb{R} \to \mathbb{R}$ is a function whose derivatives satisfy*

$$\sup_x |F^{(k)}(x)|(1 + |x|)^{-C_1} \leqslant C_1, \quad k = 1, 2, 3, \tag{4.19}$$

for some constant $C_1 > 0$. Then for any sufficiently small constant $\delta > 0$ and for any

$$E, E_1, E_2 \in I_\delta := \left\{x : |x - \lambda_+| \leqslant n^{-2/3+\delta}\right\} \quad \text{and} \quad \eta := n^{-2/3-\delta}, \tag{4.20}$$

we have

$$|\mathbb{E}F(n\eta \operatorname{Im} m(z)) - \mathbb{E}F(n\eta \operatorname{Im} \widetilde{m}(z))| \leqslant n^{-c_\phi + C_2\delta}, \quad z = E + \mathrm{i}\eta, \tag{4.21}$$

and

$$\left|\mathbb{E}F\left(n \int_{E_2}^{E_1} \operatorname{Im} m(y + \mathrm{i}\eta) dy\right) - \mathbb{E}F\left(n \int_{E_2}^{E_1} \operatorname{Im} \widetilde{m}(y + \mathrm{i}\eta) dy\right)\right| \leqslant n^{-c_\phi + C_2\delta}, \tag{4.22}$$

where c_ϕ is a constant as given in Theorem 2.5 and $C_2 > 0$ is some constant independent of c_ϕ and δ. Moreover, a general multivariate comparison estimate as in [14, Theorem 6.4] holds: fix any $k \in \mathbb{N}$ and let $F : \mathbb{R}^k \to \mathbb{R}$ be a bounded smooth function with bounded derivatives, then for any sequence of real numbers $E_k < \cdots < E_1 < E_0$ satisfying (4.20), we have

$$\begin{aligned}\left|\mathbb{E}F\left(\left(n \int_{E_k}^{E_0} \operatorname{Im} m(y + \mathrm{i}\eta) dy\right)_{1 \leqslant i \leqslant k}\right) - \mathbb{E}F\left(\left(n \int_{E_k}^{E_0} \operatorname{Im} \widetilde{m}(y + \mathrm{i}\eta) dy\right)_{1 \leqslant i \leqslant k}\right)\right| \\ \leqslant n^{-c_\phi + C_2\delta}.\end{aligned} \tag{4.23}$$

Proof of Lemma 4.5. Although not explicitly stated, it was shown in [14] that if (2.46), (2.12) and Lemma 4.6 hold, then the edge universality (4.18) holds. More precisely, in Section 6 of [14], the edge universality problem was reduced to proving Theorem 6.3 of [14], which corresponds to our Lemma 4.6. In order for this conversion to work, only the the averaged local law and the rigidity of eigenvalues are used, which correspond to (2.46) and (2.12), respectively. ☐

Finally, (2.13) follows immediately from Lemma 4.1, Lemma 4.4 and Lemma 4.5. ☐

5 Proof of Theorem 2.7

In this section, we prove Theorem 2.7. The proof is an extension of the one for Theorem 2.7 in [32]. Given the matrices X and Y satisfying Assumption 2.1 and the tail condition (2.16), we introduce a cutoff on their matrix entries at the level $n^{-\varepsilon}$. For any fixed $\varepsilon > 0$, define

$$\alpha_n^{(1)} := \mathbb{P}\left(|\widehat{x}_{11}| > n^{1/2-\varepsilon}\right), \quad \beta_n^{(1)} := \mathbb{E}\left[\mathbf{1}\left(|\widehat{x}_{11}| > n^{1/2-\varepsilon}\right)\widehat{x}_{11}\right].$$

By (2.16) and integration by parts, we get that for any fixed $\delta > 0$ and large enough n,

$$\alpha_n^{(1)} \leqslant \delta n^{-2+4\varepsilon}, \quad |\beta_n^{(1)}| \leqslant \delta n^{-3/2+3\varepsilon}. \tag{5.1}$$

Let $\rho^{(1)}(\mathrm{d}x)$ be the law of $\widehat{x}_{11}$. Then we define independent random variables $\widehat{x}_{ij}^s$, $\widehat{x}_{ij}^l$, $c_{ij}^{(1)}$, $1 \leqslant i \leqslant p, 1 \leqslant j \leqslant n$, in the following ways.

- $\widehat{x}_{ij}^s$ has law ρ_s, which is defined such that

$$\rho_s^{(1)}(\Omega) = \frac{1}{1-\alpha_n^{(1)}}\int \mathbf{1}\left(x + \frac{\beta_n^{(1)}}{1-\alpha_n^{(1)}} \in \Omega\right)\mathbf{1}\left(|x| \leqslant n^{1/2-\varepsilon}\right)\rho^{(1)}(\mathrm{d}x)$$

  for any event Ω. Note that if $\widehat{x}_{11}$ has density $\rho(x)$, then the density for $\widehat{x}_{11}^s$ is

$$\rho_s^{(1)}(x) = \mathbf{1}\left(\left|x - \frac{\beta_n^{(1)}}{1-\alpha_n^{(1)}}\right| \leqslant n^{1/2-\varepsilon}\right)\frac{\rho\left(x - \frac{\beta_n^{(1)}}{1-\alpha_n^{(1)}}\right)}{1-\alpha_n^{(1)}}.$$

- $\widehat{x}_{ij}^l$ has law ρ_l, such that

$$\rho_l^{(1)}(\Omega) = \frac{1}{\alpha_n^{(1)}}\int \mathbf{1}\left(x + \frac{\beta_n^{(1)}}{1-\alpha_n^{(1)}} \in \Omega\right)\mathbf{1}\left(|x| > n^{1/2-\varepsilon}\right)\rho^{(1)}(\mathrm{d}x)$$

  for any event Ω.

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

Let X^s, X^l and X^c be random matrices such that $x_{ij}^s = n^{-1/2}\widehat{x}_{ij}^s$, $x_{ij}^l = n^{-1/2}\widehat{x}_{ij}^l$ and $x_{ij}^c = c_{ij}^{(1)}$. It is easy to check that for independent X^s, X^l and X^c,

$$x_{ij} \overset{d}{=} x_{ij}^s\left(1 - x_{ij}^c\right) + x_{ij}^l x_{ij}^c - \frac{1}{\sqrt{n}}\frac{\beta_n^{(1)}}{1-\alpha_n^{(1)}}. \tag{5.2}$$

We have a similar decompostion for Y:

$$y_{ij} \overset{d}{=} y_{ij}^s\left(1 - y_{ij}^c\right) + y_{ij}^l y_{ij}^c - \frac{1}{\sqrt{n}}\frac{\beta_n^{(2)}}{1-\alpha_n^{(2)}}, \tag{5.3}$$

where the relevant terms are defined in the obvious way using

$$\alpha_n^{(2)} := \mathbb{P}\left(|\widehat{y}_{11}| > n^{1/2-\varepsilon}\right), \quad \beta_n^{(2)} := \mathbb{E}\left[\mathbf{1}\left(|\widehat{y}_{11}| > n^{1/2-\varepsilon}\right)\widehat{y}_{11}\right].$$

Notice that the deterministic matrices consist of the constant terms in (5.2) or (5.3) have operator norms $\mathrm{O}(n^{-1+3\varepsilon})$, which perturb the eigenvalues at most by $\mathrm{O}(n^{-1+3\varepsilon})$. Such a small error is negligible for our result, and hence we will omit the constant terms in (5.2) or (5.3) throughout the proof.

By (2.16) and integration by parts, it is easy to check that

$$\mathbb{E}\widehat{x}^s_{11} = 0, \quad \mathbb{E}|\widehat{x}^s_{11}|^2 = 1 - \mathrm{O}(n^{-1+2\varepsilon}), \quad \mathbb{E}|\widehat{x}^s_{11}|^4 = \mathrm{O}(\log n). \tag{5.4}$$

We have similar estimates for the y^s_{11} variable. Thus $X_1 := (\mathbb{E}|\widehat{x}^s_{11}|^2)^{-1/2}X^s$ and $Y_1 := (\mathbb{E}|\widehat{y}^s_{11}|^2)^{-1/2}Y^s$ are random matrices that satisfy the assumptions for X and Y in Theorem 2.5 with $\phi_n = \mathrm{O}(n^{-\varepsilon})$. Again, the $\mathrm{O}(n^{-1+2\varepsilon})$ in the denominator can be neglected.

We define the SCCA matrix $\mathcal{C}^s_{XY}$ by replacing (X, Y, Z) with (X^s, Y^s, Z^s) in the definition, and let λ^s_i be its eigenvalues. Then by Theorem 2.5,

$$\lim_{n\to\infty} \mathbb{P}\left(n^{2/3}\frac{\lambda^s_1 - \lambda_+}{c_{TW}} \leqslant s_1\right) = \lim_{n\to\infty} \mathbb{P}^{GOE}\left(n^{2/3}(\lambda_1 - 2) \leqslant s_1\right). \tag{5.5}$$

Here and throughout the following proof, we only consider the largest eigenvalue. It is easy to extend to the case with multiple largest eigenvalues. Now we write the first two terms on the right-hand side of (5.2) as

$$x^s_{ij}\left(1 - x^c_{ij}\right) + x^l_{ij}x^c_{ij} = x^s_{ij} + \Delta^{(1)}_{ij}x^c_{ij}, \quad \Delta^{(1)}_{ij} := x^l_{ij} - x^s_{ij}.$$

Similarly, we have

$$y^s_{ij}\left(1 - y^c_{ij}\right) + y^l_{ij}y^c_{ij} = y^s_{ij} + \Delta^{(1)}_{ij}y^c_{ij}, \quad \Delta^{(2)}_{ij} := y^l_{ij} - y^s_{ij}.$$

We define the matrices $\mathcal{E}^{(1)} := (\Delta^{(1)}_{ij}x^c_{ij})$ and $\mathcal{E}^{(2)} := (\Delta^{(2)}_{ij}y^c_{ij})$. It remains to show that the effect of $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ on the eigenvalue λ_1 is negligible.

We introduce the following event

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

Using Bernstein inequality, we have that

$$\mathbb{P}\left(\left\{\#\{(i,j) : x^c_{ij} = 1\} \leqslant n^{5\varepsilon}\right\}\right) \geqslant 1 - \exp(-n^{\varepsilon}), \tag{5.6}$$

for sufficiently large n. Suppose the number n_0 of the nonzero elements in X^c is given with $n_0 \leqslant n^{5\varepsilon}$. Then it is easy to check that

$$\begin{aligned}\mathbb{P}\left(\exists\, i = k, j \neq l \text{ or } i \neq k, j = l \text{ such that } x^c_{ij} = x^c_{kl} = 1 \,\middle|\, \#\{(i,j) : x^c_{ij} = 1\} = n_0\right)& \\ = \mathrm{O}(n_0^2 n^{-1}).&\end{aligned} \tag{5.7}$$

Combining the estimates (5.6) and (5.7), we get that

$$\mathbb{P}(\mathscr{A}) \geqslant 1 - \mathrm{O}(n^{-1+10\varepsilon}). \tag{5.8}$$

On the other hand, by condition (2.16), we have

$$\mathbb{P}\left(|\mathcal{E}^{(1)}_{ij}| \geqslant \omega\right) \leqslant \mathbb{P}\left(|\widehat{x}_{ij}| \geqslant \frac{\omega}{2}n^{1/2}\right) = \mathrm{o}(n^{-2}),$$

for any fixed constant $\omega > 0$. With a simple union bound, we get

$$\mathbb{P}\left(\max_{i,j}|\mathcal{E}^{(1)}_{ij}| \geqslant \omega\right) = \mathrm{o}(1). \tag{5.9}$$

Similarly, we can define the event

$$\mathscr{B} := \left\{\#\{(i,j) : y^c_{ij} = 1\} \leqslant n^{5\varepsilon}\right\} \cap \left\{y^c_{ij} = y^c_{kl} = 1 \Rightarrow (i,j) = (k,l) \text{ or } \{i,j\} \cap \{k,l\} = \varnothing\right\}.$$

By (5.8), (5.9) and similar estimates for matrix Y, we get

$$\mathbb{P}(\mathscr{A} \cap \mathscr{B}) = 1 - \mathrm{o}(1), \quad \mathbb{P}(\mathscr{C}_1) = 1 - \mathrm{o}(1), \tag{5.10}$$

where

$$\mathscr{C}_1 := \left\{\max_{i,j} |\mathcal{E}^{(1)}_{ij}| \leqslant \omega\right\} \cap \left\{\max_{i,j} |\mathcal{E}^{(2)}_{ij}| \leqslant \omega\right\}.$$

Then recall (2.18), we only need to study the determinant of $H_1(\lambda)$ on event $\mathscr{A} \cap \mathscr{B} \cap \mathscr{C}_1$, where we define $H_t(\lambda)$, $t \in [0,1]$, as

$$H_t(\lambda) := H^s(\lambda) + t \begin{pmatrix} 0 & \begin{pmatrix} \mathcal{E}^{(1)} & 0 \\ 0 & \mathcal{E}^{(2)} \end{pmatrix} \\ \begin{pmatrix} (\mathcal{E}^{(1)})^T & 0 \\ 0 & (\mathcal{E}^{(3)})^T \end{pmatrix} & 0 \end{pmatrix},$$

where

$$H^s(\lambda) := \begin{pmatrix} 0 & \begin{pmatrix} X^s & 0 \\ 0 & Y^s \end{pmatrix} \\ \begin{pmatrix} (X^s)^T & 0 \\ 0 & (Y^s)^T \end{pmatrix} & \begin{pmatrix} \lambda I_n & \lambda^{1/2} I_n \\ \lambda^{1/2} I_n & \lambda I_n \end{pmatrix}^{-1} \end{pmatrix}$$

We would like to use a continuity argument to extend (5.5) in the $t = 0$ case all the way to the $t = 1$ case. It is easy to observe that with probability $1 - \mathrm{o}(1)$, the eigenvalues $\lambda^t_1 \equiv \lambda_1(t)$ is continuous in t for all $t \in [0,1]$. In fact, on even $\mathscr{A} \cap \mathscr{B} \cap \mathscr{C}_1$, we have

$$\|\mathcal{E}^{(1)}\| \leqslant \omega, \quad \|\mathcal{E}^{(2)}\| \leqslant \omega. \tag{5.11}$$

Hence with Lemma 3.3, as long as ω is chosen sufficiently small, $[(X^s + t\mathcal{E}^{(1)})(X^s + t\mathcal{E}^{(1)})^T]^{-1}$ and $[(Y^s + t\mathcal{E}^{(2)})(Y^s + t\mathcal{E}^{(2)})^T]^{-1}$ will be continuous in t on a high probability event, which implies the continuity of eigenvalues. Now we claim that for $\mu := \lambda_1(0) \pm n^{-3/4} \equiv \lambda^s_1 \pm n^{-3/4}$,

$$\mathbb{P}\left(\det H_t(\mu) \neq 0 \text{ for all } 0 \leqslant t \leqslant 1\right) = 1 - \mathrm{o}(1). \tag{5.12}$$

Suppose (5.12) holds true, then by continuity $\lambda_1 \equiv \lambda_1(t = 1) \in [\lambda^s_1 - n^{-3/4}, \lambda^s_1 + n^{-3/4}]$ with probability $1 - \mathrm{o}(1)$, which concludes the proof together with (5.5).

The rest of the proof is devoted to proving (5.12). In the following proof, we condition on the event $\mathscr{A} \cap \mathscr{B}$ and the event $\mathscr{C}_{n_x n_y}$ that X^c and Y^c have n_x and n_y nonzero entries for some $\max\{n_x, n_y\} \leqslant n^{5\varepsilon}$. Without loss of generality, we can assume the positions of the n_x nonzero entries of X^c are $(1,1), \cdots, (n_x, n_x)$, and the positions of the n_y nonzero entries of Y^c are $(1,1), \cdots, (n_y, n_y)$, that is, we also condition on these two event. For other choices of the positions of nonzero entries, the proof is the same. Then we rewrite

$$\widetilde{H}_t(\mu) = H^s(\mu) + tO \begin{pmatrix} 0 & \mathcal{D}_e \\ \mathcal{D}_e & 0 \end{pmatrix} O^T, \quad O := \begin{pmatrix} \mathbf{F}_1 & 0 \\ 0 & \mathbf{F}_2 \end{pmatrix},$$

where

$$\mathcal{D}_e := \begin{pmatrix} \Sigma^{(1)}_e & 0 \\ 0 & \Sigma^{(2)}_e \end{pmatrix}, \quad \Sigma^{(1)}_e := \mathrm{diag}\left(\mathcal{E}^{(1)}_{11}, \cdots, \mathcal{E}^{(1)}_{n_x n_x}\right), \quad \Sigma^{(2)}_e := \mathrm{diag}\left(\mathcal{E}^{(2)}_{11}, \cdots, \mathcal{E}^{(2)}_{n_y n_y}\right),$$

and

$$\mathbf{F}_1 := \begin{pmatrix} \left(\mathbf{e}_1^{(p)}, \cdots, \mathbf{e}_{n_x}^{(p)}\right) & 0 \\ 0 & \left(\mathbf{e}_1^{(q)}, \cdots, \mathbf{e}_{n_y}^{(q)}\right) \end{pmatrix}, \quad \mathbf{F}_2 := \begin{pmatrix} \left(\mathbf{e}_1^{(n)}, \cdots, \mathbf{e}_{n_x}^{(n)}\right) & 0 \\ 0 & \left(\mathbf{e}_1^{(n)}, \cdots, \mathbf{e}_{n_y}^{(n)}\right) \end{pmatrix}.$$

Here $\mathbf{e}_i^{(l)}$ means the standard unit vector along i-th coordinate direction in $\mathbb{R}^l$.

Applying the identity $\det(1 + \mathcal{AB}) = \det(1 + \mathcal{BA})$, we obtain that if μ is such that $\det G^s(\mu) \neq 0$, then

$$\det H_t(\mu) = \det G^s(\mu) \cdot \det\left(1 + tF(\mu)\right), \quad F(\mu) := t \begin{pmatrix} 0 & \mathcal{D}_e \\ \mathcal{D}_e & 0 \end{pmatrix} O^T G^s(\mu) O. \tag{5.13}$$

In the following proof, we use $z = \lambda_+ + \mathrm{i} n^{-2/3}$. Then we can write

$$O^T G^s(\mu) O = O^T \left[G^s(\mu) - G^s(z)\right] O + O^T [G^s(z) - \Pi(z)] O + O^T \Pi(z) O. \tag{5.14}$$

By Lemma 3.10, we have that

$$\mathbb{E}\left|\left[O^T \left(G^s(z) - \Pi(z)\right) O\right]_{ij}^2\right| \prec \Psi^2(z) = \mathrm{O}(n^{-2/3}), \quad 1 \leqslant i, j \leqslant n_x + n_y,$$

where we used (3.2) and (2.43) in the second step. Then with Markov's inequality and a union bound, we can get that

$$\max_{1 \leqslant i,j \leqslant n_x + n_y} \left|\left[O^T \left(G^s(z) - \Pi(z)\right) O\right]_{ab}\right| \leqslant n^{-1/6}$$

holds with probability $1 - \mathrm{O}(n^{-1/3+5\varepsilon})$. In particular, this gives that with probability $1 - \mathrm{O}(n^{-1/3+5\varepsilon})$,

$$\|O^T [G^s(z) - \Pi(z)] O\| \lesssim n^{-1/6+5\varepsilon}. \tag{5.15}$$

On the other hand, we claim that

$$\|O_t^T \left[G^s(\mu) - G^s(z)\right] O_t\| \leqslant n^{-1/6} \quad \text{with probability } 1 - \mathrm{o}(1). \tag{5.16}$$

If (5.16) holds, together with (5.14) and (5.15), we get that with probability $1 - \mathrm{o}(1)$,

$$\|O^T G^s(\mu) O\| \leqslant \|\Pi(z)\| + \mathrm{O}(n^{-1/6+5\varepsilon}) \leqslant 2\|\Pi(z)\| \;\Rightarrow\; \max_{0 \leqslant t \leqslant 1} t\|F(\mu)\| \leqslant 2\omega\|\Pi(z)\| \leqslant \frac{1}{2},$$

as long as ω is chosen small enough. Hence we have with probability $1 - \mathrm{o}(1)$, $1 + tF(\mu)$ is non-singular for all $t \in [0, 1]$, which concludes (5.12).

Finally it remains to prove (5.16). Since the largest eigenvalues for GOE are separated in the scale $n^{-2/3}$, by (2.13) we have that

$$\mathbb{P}\left(\min_i |\lambda_i^s - \mu| \geqslant n^{-3/4}\right) = 1 - \mathrm{o}(1). \tag{5.17}$$

On the other hand, the rigidity result (2.12) gives that

$$|\mu - \lambda_+| \prec n^{-2/3}. \tag{5.18}$$

Then using Lemma 3.3, Lemma 3.9, (5.17), (5.18) and (2.12), we can get that for any set Ω of deterministic unit vectors of cardinality $n^{\mathrm{O}(1)}$,

$$\sup_{\mathbf{u},\mathbf{v} \in \Omega} \left|\mathbf{u}^* \left(G^s(z) - G^s(\mu)\right) \mathbf{v}\right| \leqslant n^{-1/4+3\varepsilon} \tag{5.19}$$

with probability $1-\mathrm{o}(1)$. We only give the derivation of (5.19) for $\mathbf{u},\mathbf{v}\in\mathbb{C}^{\mathcal{I}_1}$. For the rest of the cases $\mathbf{u}\in\mathbb{C}^{\mathcal{I}_\alpha}$ and $\mathbf{v}\in\mathbb{C}^{\mathcal{I}_\beta}$, $\alpha,\beta=1,2,3,4$, the proof is similar. For deterministic unit vectors $\mathbf{u},\mathbf{v}\in\mathbb{C}^{\mathcal{I}_1}$, we have with probability $1-\mathrm{o}(1)$ that

$$
\begin{aligned}
&|\langle \mathbf{u}, (G^s(z)-G^s(\mu))\,\mathbf{v}\rangle| \\
&\leqslant \sum_k \frac{|z-\mu|\left|\langle \mathbf{u}, S_{xx}^{-1/2}\xi_k\rangle\langle \xi_k S_{xx}^{-1/2},\mathbf{v}\rangle\right|}{|\lambda_k^s-z||\lambda_k^s-\mu|} + \frac{|z-\mu|}{|z\mu|}\sum_{k=q+1}^{p}\left|\langle \mathbf{u}, S_{xx}^{-1/2}\xi_k\rangle\right|\left|\langle \xi_k S_{xx}^{-1/2},\mathbf{v}\rangle\right| \\
&\prec \frac{1}{n^{2/3}}\sum_{k\geqslant q/2}\left|\langle \mathbf{u}, S_{xx}^{-1/2}\xi_k\rangle\right|\left|\langle \xi_k S_{xx}^{-1/2},\mathbf{v}\rangle\right| + \frac{1}{n^{5/3}}\sum_{k<q/2}\frac{1}{|\lambda_k^s-z||\lambda_k^s-\mu|} \\
&\leqslant \frac{\|S_{xx}^{-1/2}\,\mathbf{u}\,\|^2+\|S_{xx}^{-1/2}\,\mathbf{v}\,\|^2}{n^{2/3}} + \frac{1}{n^{5/3}}\sum_{1\leqslant k\leqslant n^{\varepsilon}}\frac{1}{|\lambda_k^s-z||\lambda_k^s-\mu|} + \frac{1}{n^{5/3}}\sum_{n^{\varepsilon}<k<q/2}\frac{1}{|\lambda_k^s-z||\lambda_k^s-\mu|} \\
&\prec \frac{1}{n^{2/3}} + \frac{n^{\varepsilon}}{n^{1/4}} + \frac{1}{n^{2/3}}\left(\frac{1}{n}\sum_{n^{\varepsilon}<k<q/2}\frac{1}{|\lambda_k^s-z||\lambda_k^s-\mu|}\right) \prec n^{-1/4+\varepsilon},
\end{aligned}
$$

where in the first step we used (3.10) and (3.11); in the second step we used (3.23) and $|\lambda_k-z||\lambda_k-\mu|\gtrsim 1$ for $k\geqslant q/2$ due to (2.12); in the third step we used Cauchy-Schwarz inequality; in the fourth step we used Lemma 3.3 and (5.17); in the last step we used $|\lambda_k^s-z||\lambda_k^s-\mu|\sim (k/n)^{-4/3}$ for $k>n^{\varepsilon}$ by the rigidity estimate (2.12).

Thus we have proved (5.19), which implies (5.16), which further concludes (5.12). This completes the proof of Theorem 2.7.

6 Proof of Theorem 2.11: the entrywise local law

The proof of Theorem 2.11 is divided into two steps. In this section, we prove a weaker local law as in Proposition 6.1 below. Based on this estimate, we shall complete the proof of Theorem 2.11 in next section.

Note that (2.42) justifies the arguments in Remark 3.7, so we only need to prove this theorem for $\widehat{G}(z)$. However, for simplicity of notations, we will still use the notations $G(z)$, while keeping in mind that there is a small regularization term in $G(z)$ such that the deterministic bounds in (3.17) holds. In particular, we will tacitly use the following fact: for $z\in S(\varepsilon)$ and a (complex) polynomial of the entries of $G(z)$, say $\mathcal{P}(G)$, if $|\mathcal{P}(G)|\prec\Phi(z)$ for some deterministic parameter $\Phi(z)\geqslant n^{-C}$, then

$$|\mathbb{E}\mathcal{P}(G)|\prec\Phi(z)$$

by Lemma 3.1 (iii).

The goal of this section is to prove the averaged local laws (2.44) and (2.45), and the entrywise local law in Proposition 6.1 below. For simplicity of notations, we first assume that the entries of X and Y satisfy

$$\mathbb{E}x_{i\mu}=\mathbb{E}y_{j\nu}=0, \tag{6.1}$$

and

$$\mathbb{E}|x_{i\mu}|^2=\mathbb{E}|y_{j\nu}|^2=n^{-1}, \tag{6.2}$$

for $i\in\mathcal{I}_1$, $j\in\mathcal{I}_2$, $\mu\in\mathcal{I}_3$ and $\nu\in\mathcal{I}_4$. Later in Section 6.3, we will discuss how to relax (6.1) and (6.2) to (2.4) and (2.5).

Proposition 6.1. *Suppose that* (6.1), (6.2) *and the assumptions of Theorem 2.5 hold. Then for any fixed* $\varepsilon > 0$, *we have that*

$$\left|\left[\Pi^{-1}(z)\left(G(z)-\Pi(z)\right)\Pi^{-1}(z)\right]_{\mathfrak{a}\mathfrak{b}}\right| \prec \phi_n + \Psi(z), \tag{6.3}$$

uniformly in $\mathfrak{a},\mathfrak{b} \in \mathcal{I}$ *and* $z \in S(\varepsilon)$.

With (6.3), we shall use a polynomialization method as in [5, Section 5] and [31, Section 5] to get the anisotropic local law (2.42). This will be presented in Section 7.

6.1 Basic tools

In this subsection, we introduce more notations and collect some basic tools that will be used in the proof.

Definition 6.2 (Minors)**.** *For any* $\mathcal{J}\times\mathcal{J}$ *matrix* $\mathcal{A}$ *and* $\mathbb{T}\subseteq\mathcal{J}$, *where* $\mathcal{J}$ *and* $\mathbb{T}$ *are some index stes, we define the minor* $\mathcal{A}^{(\mathbb{T})} := (\mathcal{A}_{ab} : a, b \in \mathcal{J}\backslash\mathbb{T})$ *as the* $(\mathcal{J}\backslash\mathbb{T})\times(\mathcal{J}\backslash\mathbb{T})$ *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*

$$G^{(\mathbb{T})}(z) : = (H^{(\mathbb{T})}(z))^{-1}.$$

As in Definition 2.9, its blocks are denoted as $\mathcal{G}_\alpha^{(\mathbb{T})}(z)$, $\alpha = 1,2,3,4$, *and* $\mathcal{G}_L^{(\mathbb{T})}(z)$, $\mathcal{G}_{LR}^{(\mathbb{T})}(z)$, $\mathcal{G}_{RL}^{(\mathbb{T})}(z)$, $\mathcal{G}_R^{(\mathbb{T})}(z)$; *its partial traces are*

$$m_\alpha^{(\mathbb{T})}(z) := \frac{1}{n}\operatorname{Tr}\mathcal{G}_\alpha^{(\mathbb{T})}(z) = \frac{1}{n}\sum_{\mathfrak{a}\in\mathcal{I}_\alpha} G_{\mathfrak{a}\mathfrak{a}}^{(\mathbb{T})}(z), \quad \alpha = 1,2,3,4.$$

Moreover, we define $S_{xx}^{(\mathbb{T})}$, $S_{xy}^{(\mathbb{T})}$, $S_{yy}^{(\mathbb{T})}$, $\mathcal{H}^{(\mathbb{T})}$, $R^{(\mathbb{T})}(z)$ *and* $m^{(\mathbb{T})}(z)$ *by replacing* (X,Y) *with* $(X^{(\mathbb{T})}, Y^{(\mathbb{T})})$.

For $\mathbb{T}\subset\mathcal{I}_3\cup\mathcal{I}_4$, *we denote* $[\mathbb{T}] := \{\mu\in\mathcal{I}_3\cup\mathcal{I}_4 : \mu\in\mathbb{T} \text{ or } \overline{\mu}\in\mathbb{T}\}$ *(recall Definition 2.8). Then we define the minor* $H^{[\mathbb{T}]} := H^{([\mathbb{T}])}$, *and correspondingly* $G^{[\mathbb{T}]} := (H^{[\mathbb{T}]})^{-1}$.

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

Using Schur complement formula, one can obtain the following resolvent identities.

Lemma 6.3. *(Resolvent identities).*

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

$$\frac{1}{G_{ii}} = -zn^{-10} - \left(WG^{(i)}W^T\right)_{ii}. \tag{6.4}$$

where we abbreviate

$$W := \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}.$$

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

$$G_{ij} = -G_{ii}\left(WG^{(i)}\right)_{ij} = -\left(G^{(j)}W^T\right)_{ij}G_{jj} = G_{ii}G_{jj}^{(i)}\left(WG^{(ij)}W^T\right)_{ij}. \tag{6.5}$$

For $i\in\mathcal{I}_1\cup\mathcal{I}_2$ *and* $\mu\in\mathcal{I}_3\cup\mathcal{I}_4$, *we have*

$$\begin{aligned} G_{i\mu} &= -G_{ii}(WG^{(i)})_{i\mu} = -(G^{(\mu)}W)_{i\mu}G_{\mu\mu}, \\ G_{\mu i} &= -G_{\mu\mu}(W^TG^{(\mu)})_{\mu i} = -(G^{(i)}W^T)_{\mu i}G_{ii}. \end{aligned} \tag{6.6}$$

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

$$G_{\mathfrak{b}\mathfrak{c}} = G_{\mathfrak{b}\mathfrak{c}}^{(\mathfrak{a})} + \frac{G_{\mathfrak{b}\mathfrak{a}} G_{\mathfrak{a}\mathfrak{c}}}{G_{\mathfrak{a}\mathfrak{a}}}, \quad \frac{1}{G_{\mathfrak{b}\mathfrak{b}}} = \frac{1}{G_{\mathfrak{b}\mathfrak{b}}^{(\mathfrak{a})}} - \frac{G_{\mathfrak{b}\mathfrak{a}} G_{\mathfrak{a}\mathfrak{b}}}{G_{\mathfrak{b}\mathfrak{b}} G_{\mathfrak{b}\mathfrak{b}}^{(\mathfrak{a})} G_{\mathfrak{a}\mathfrak{a}}}. \tag{6.7}$$

(iv) All of the above identities hold for $G^{(\mathbb{T})}$ instead of G for $\mathbb{T} \subset \mathcal{I}$.

For $\mathcal{I} \times \mathcal{I}$ matrix $\mathcal{A}$ and $\mu, \nu \in \mathcal{I}_3$, we define the 2×2 minors as

$$\mathcal{A}_{[\mu\nu]} = \begin{pmatrix} \mathcal{A}_{\mu\nu} & \mathcal{A}_{\mu\overline{\nu}} \\ \mathcal{A}_{\overline{\mu}\nu} & \mathcal{A}_{\overline{\mu}\overline{\nu}} \end{pmatrix}, \tag{6.8}$$

where we recall the notations in Definition 2.8. We shall call $A_{[\mu\nu]}$ a diagonal group if $\mu = \nu$, and an off-diagonal group otherwise. Similarly, for $i \in \mathcal{I}_1$, $j \in \mathcal{I}_2$ and $\mu \in \mathcal{I}_3$, we define the minors

$$\begin{gathered} \mathcal{A}_{ij,[\mu]} = \begin{pmatrix} \mathcal{A}_{i\mu} & \mathcal{A}_{i\overline{\mu}} \\ \mathcal{A}_{j\mu} & \mathcal{A}_{j\overline{\mu}} \end{pmatrix}, \quad \mathcal{A}_{[\mu],ij} = \begin{pmatrix} \mathcal{A}_{\mu i} & \mathcal{A}_{\mu j} \\ \mathcal{A}_{\overline{\mu} i} & \mathcal{A}_{\overline{\mu} j} \end{pmatrix}, \\ \mathcal{A}_{i,[\mu]} = (\mathcal{A}_{i\mu}, \mathcal{A}_{i\overline{\mu}}), \quad \mathcal{A}_{[\mu],i} = \begin{pmatrix} \mathcal{A}_{\mu i} \\ \mathcal{A}_{\overline{\mu} i} \end{pmatrix}. \end{gathered} \tag{6.9}$$

For G, sometimes it is convenient to deal with 2×2 blocks directly, and we record the following resolvent identities obtained from Schur complement formula.

Lemma 6.4. *(Resolvent identities for $G_{[\mu\nu]}$ groups).*

(i) For $\mu \in \mathcal{I}_3$, we have

$$G_{[\mu\mu]}^{-1} = \frac{1}{z-1} \begin{pmatrix} 1 & -z^{-1/2} \\ -z^{-1/2} & 1 \end{pmatrix} - \begin{pmatrix} (X^T G^{[\mu]} X)_{\mu\mu} & (X^T G^{[\mu]} Y)_{\mu\overline{\mu}} \\ (Y^T G^{[\mu]} X)_{\overline{\mu}\mu} & (Y^T G^{[\mu]} Y)_{\overline{\mu}\overline{\mu}} \end{pmatrix}. \tag{6.10}$$

(ii) For $i \in \mathcal{I}_1$ and $\mu \in \mathcal{I}_3$, we have

$$\begin{aligned} G_{i,[\mu]} &= -\left((G^{[\mu]} X)_{i\mu}, (G^{[\mu]} Y)_{i\overline{\mu}}\right) G_{[\mu\mu]} \\ &= G_{ii}^{[\mu]} \left((X G^{(i\mu\overline{\mu})} X)_{i\mu}\ ,\ (X G^{(i\mu\overline{\mu})} Y)_{i\overline{\mu}}\right) G_{[\mu\mu]}, \end{aligned} \tag{6.11}$$

and

$$G_{[\mu],i} = -G_{[\mu\mu]} \begin{pmatrix} (X^T G^{[\mu]})_{\mu i} \\ (Y^T G^{[\mu]})_{\overline{\mu} i} \end{pmatrix} = G_{[\mu\mu]} \begin{pmatrix} (X^T G^{(i\mu\overline{\mu})} X^T)_{\mu i} \\ (Y^T G^{(i\mu\overline{\mu})} X^T)_{\overline{\mu} i} \end{pmatrix} G_{ii}^{[\mu]}. \tag{6.12}$$

We have similar expansions for $G_{j,[\mu]}$ and $G_{[\mu],j}$ for $j \in \mathcal{I}_2$ by interchanging X and Y.

(iii) For $\mu \neq \nu \in \mathcal{I}_3$, we have

$$\begin{aligned} G_{[\mu\nu]} &= -G_{[\mu\mu]} \begin{pmatrix} (X^T G^{[\mu]})_{\mu\nu} & (X^T G^{[\mu]})_{\mu\overline{\nu}} \\ (Y^T G^{[\mu]})_{\overline{\mu}\nu} & (Y^T G^{[\mu]})_{\overline{\mu}\overline{\nu}} \end{pmatrix} = -\begin{pmatrix} (G^{[\nu]} X)_{\mu\nu} & (G^{[\nu]} Y)_{\mu\overline{\nu}} \\ (G^{[\nu]} X)_{\overline{\mu}\nu} & (G^{[\nu]} Y)_{\overline{\mu}\overline{\nu}} \end{pmatrix} G_{[\nu\nu]} \\ &= G_{[\mu\mu]} G_{[\nu\nu]}^{[\mu]} \begin{pmatrix} (X^T G^{[\mu\nu]} X)_{\mu\nu} & (X^T G^{[\mu\nu]} Y)_{\mu\overline{\nu}} \\ (Y^T G^{[\mu\nu]} X)_{\overline{\mu}\nu} & (Y^T G^{[\mu\nu]} Y)_{\overline{\mu}\overline{\nu}} \end{pmatrix}. \end{aligned} \tag{6.13}$$

(iv) For $\mu \in \mathcal{I}_3$ and $\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{b}_1, \mathfrak{b}_2 \in \mathcal{I} \backslash \{\mu, \nu\}$, we have

$$\begin{pmatrix} G_{\mathfrak{a}_1\mathfrak{b}_1} & G_{\mathfrak{a}_1\mathfrak{b}_2} \\ G_{\mathfrak{a}_2\mathfrak{b}_1} & G_{\mathfrak{a}_2\mathfrak{b}_2} \end{pmatrix} = \begin{pmatrix} G^{[\mu]}_{\mathfrak{a}_1\mathfrak{b}_1} & G^{[\mu]}_{\mathfrak{a}_1\mathfrak{b}_2} \\ G^{[\mu]}_{\mathfrak{a}_2\mathfrak{b}_1} & G^{[\mu]}_{\mathfrak{a}_2\mathfrak{b}_2} \end{pmatrix} + \begin{pmatrix} G_{\mathfrak{a}_1\mu} & G_{\mathfrak{a}_1\overline{\mu}} \\ G_{\mathfrak{a}_2\mu} & G_{\mathfrak{a}_2\overline{\mu}} \end{pmatrix} G^{-1}_{[\mu\mu]} \begin{pmatrix} G_{\mu\mathfrak{b}_1} & G_{\mu\mathfrak{b}_2} \\ G_{\overline{\mu}\mathfrak{b}_1} & G_{\overline{\mu}\mathfrak{b}_2} \end{pmatrix}, \tag{6.14}$$

and

$$\begin{aligned} \begin{pmatrix} G_{\mathfrak{a}_1\mathfrak{a}_1} & G_{\mathfrak{a}_1\mathfrak{a}_2} \\ G_{\mathfrak{a}_2\mathfrak{a}_1} & G_{\mathfrak{a}_2\mathfrak{a}_2} \end{pmatrix}^{-1} &= \begin{pmatrix} G^{[\mu]}_{\mathfrak{a}_1\mathfrak{a}_1} & G^{[\mu]}_{\mathfrak{a}_1\mathfrak{a}_2} \\ G^{[\mu]}_{\mathfrak{a}_2\mathfrak{a}_1} & G^{[\mu]}_{\mathfrak{a}_2\mathfrak{a}_2} \end{pmatrix}^{-1} \\ - \begin{pmatrix} G_{\mathfrak{a}_1\mathfrak{a}_1} & G_{\mathfrak{a}_1\mathfrak{a}_2} \\ G_{\mathfrak{a}_2\mathfrak{a}_1} & G_{\mathfrak{a}_2\mathfrak{a}_2} \end{pmatrix}^{-1} & \begin{pmatrix} G_{\mathfrak{a}_1\mu} & G_{\mathfrak{a}_1\overline{\mu}} \\ G_{\mathfrak{a}_2\mu} & G_{\mathfrak{a}_2\overline{\mu}} \end{pmatrix} G^{-1}_{[\mu\mu]} \begin{pmatrix} G_{\mu\mathfrak{b}_1} & G_{\mu\mathfrak{b}_2} \\ G_{\overline{\mu}\mathfrak{b}_1} & G_{\overline{\mu}\mathfrak{b}_2} \end{pmatrix} \begin{pmatrix} G^{[\mu]}_{\mathfrak{a}_1\mathfrak{a}_1} & G^{[\mu]}_{\mathfrak{a}_1\mathfrak{a}_2} \\ G^{[\mu]}_{\mathfrak{a}_2\mathfrak{a}_1} & G^{[\mu]}_{\mathfrak{a}_2\mathfrak{a}_2} \end{pmatrix}^{-1}. \end{aligned} \tag{6.15}$$

(iii) All of the above identities hold for $G^{(\mathbb{T})}$ instead of G for $\mathbb{T} \subset \mathcal{I}$.

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

Lemma 6.5 (Lemma 3.8 of [13])**.** *Let (x_i), (y_j) be independent families of centered and independent random variables, and $(\mathcal{A}_i)$, $(\mathcal{B}_{ij})$ 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 (2.10) with $\phi_n \leqslant n^{-c_\phi}$ for some constant $c_\phi > 0$. Then we have the following bound:*

$$\begin{aligned} \Big|\sum_i \mathcal{A}_i x_i\Big| &\prec \phi_n \max_i |\mathcal{A}_i| + \frac{1}{\sqrt{n}}\Big(\sum_i |\mathcal{A}_i|^2\Big)^{1/2}, \\ \Big|\sum_{i,j} x_i \mathcal{B}_{ij} y_j\Big| &\prec \phi_n^2 \mathcal{B}_d + \phi_n \mathcal{B}_o + \frac{1}{n}\Big(\sum_{i\neq j} |\mathcal{B}_{ij}|^2\Big)^{1/2}, \\ \Big|\sum_i \overline{x}_i \mathcal{B}_{ii} x_i - \sum_i (\mathbb{E}|x_i|^2)\mathcal{B}_{ii}\Big| &\prec \phi_n \mathcal{B}_d, \\ \Big|\sum_{i\neq j} \overline{x}_i \mathcal{B}_{ij} x_j\Big| &\prec \phi_n \mathcal{B}_o + \frac{1}{n}\Big(\sum_{i\neq j} |\mathcal{B}_{ij}|^2\Big)^{1/2}, \end{aligned}$$

where $\mathcal{B}_d := \max_i |\mathcal{B}_{ii}|$ and $\mathcal{B}_o := \max_{i\neq j} |\mathcal{B}_{ij}|$.

Corresponding to the lower right block of $\Pi(z)$ in (2.37), we define the 2×2 matrix

$$\begin{aligned} \pi(z) &:= \begin{pmatrix} m_{3c}(z) & h(z) \\ h(z) & m_{4c}(z) \end{pmatrix} \\ &= \frac{1 - c_1 - c_2 + \sqrt{(z-\lambda_-)(z-\lambda_+)}}{2} \begin{pmatrix} 1 & z^{1/2} \\ z^{1/2} & 1 \end{pmatrix} + \frac{z-1}{2} \begin{pmatrix} 1-2c_1 & -z^{1/2} \\ -z^{1/2} & 1-2c_2 \end{pmatrix}. \end{aligned} \tag{6.16}$$

Using (2.34)-(2.36), one can check that

$$\pi(z)^{-1} = \frac{1}{z-1} \begin{pmatrix} 1-(z-1)m_{1c} & -z^{-1/2} \\ -z^{-1/2} & 1-(z-1)m_{2c} \end{pmatrix}. \tag{6.17}$$

For the proof of Proposition 6.1, it is convenient to introduce the following random control parameters.

Definition 6.6 (Control parameters). *We define the random errors*

$$\begin{aligned}\Lambda_o : &= \max_{i\neq j\in\mathcal{I}_1\cup\mathcal{I}_2}|G_{ij}| + \max_{\mu\neq\nu\in\mathcal{I}_3}\|\pi^{-1}G_{[\mu\nu]}\pi^{-1}\| \\ &+ \max_{i\in\mathcal{I}_1\cup\mathcal{I}_2,\mu\in\mathcal{I}_3}\left(\|G_{i,[\mu]}\pi^{-1}\| + \|\pi^{-1}G_{[\mu],i}\|\right), \\ \Lambda : &= \Lambda_o + \max_{i\in\mathcal{I}_1\cup\mathcal{I}_2}|G_{ii}-\Pi_{ii}| + \max_{\mu\in\mathcal{I}_3}\left\|\pi^{-1}\left(G_{[\mu\mu]}-\pi\right)\pi^{-1}\right\|,\end{aligned} \tag{6.18}$$

and

$$\theta := |m_1 - m_{1c}| + |m_2 - m_{2c}| + \left\|\frac{1}{n}\sum_\mu\left(G_{[\mu\mu]}-\pi\right)\right\|. \tag{6.19}$$

Here Λ controls the entrywise error, Λ_o controls the size of the off-diagonal entries, and θ gives the averaged error. Note that these parameters all depend on z, and we did not write down this dependence explicitly in the definitions. Moreover, replacing G with $G^{(\mathbb{T})}$ for any $\mathbb{T}\subset\mathcal{I}$, we can define parameters $\Lambda_o^{(\mathbb{T})}$, $\Lambda^{(\mathbb{T})}$ and $\theta^{(\mathbb{T})}$. We then define the random control parameter (recall Ψ defined in (2.43))

$$\Psi_\Lambda(z) := \sqrt{\frac{\operatorname{Im} m_c(z) + \Lambda(z)}{n\eta}} + \frac{1}{n\eta}. \tag{6.20}$$

Define a z-dependent event

$$\Xi(z) := \{\Lambda(z) \leqslant (\log n)^{-1}\}.$$

Then on Ξ, using (3.2) and (6.18), we have that for $i\in\mathcal{I}_1\cup\mathcal{I}_2$ and $\mu\in\mathcal{I}_3$,

$$G_{ii}\sim 1, \quad G_{[\mu\mu]} = \pi(z) + \pi(z)\mathcal{E}(z)\pi(z) \ \text{ with } \ \|\mathcal{E}(z)\| = \mathrm{O}\left((\log n)^{-1}\right). \tag{6.21}$$

Then using (6.21) and (6.7), we obtain that for $k\in\mathcal{I}_1\cup\mathcal{I}_2$, $i,j\in\mathcal{I}_1\cup\mathcal{I}_2\backslash\{k\}$ and $\mu,\nu\in\mathcal{I}_3$,

$$\mathbf{1}(\Xi)|G_{ij}-G_{ij}^{(k)}| \lesssim \Lambda_o^2, \quad \mathbf{1}(\Xi)\left\|\pi^{-1}\left(G_{[\mu\nu]}-G_{[\mu\nu]}^{(k)}\right)\pi^{-1}\right\| \lesssim \|\pi^{-1}G_{[\mu],k}\|\,\|G_{k,[\nu]}\pi^{-1}\| \lesssim \Lambda_o^2,$$

and

$$\mathbf{1}(\Xi)\left\|\pi^{-1}\left(G_{[\mu],i}-G_{[\mu],i}^{(k)}\right)\right\| \lesssim \|\pi^{-1}G_{[\mu],k}\|\,|G_{ik}| \lesssim \Lambda_o^2.$$

Similarly, for $i,j\in\mathcal{I}_1\cup\mathcal{I}_2$, $\mu\in\mathcal{I}_3$ and $\alpha,\beta\in\mathcal{I}_3\backslash\{\mu\}$, using (6.21) and (6.14) we obtain that

$$\mathbf{1}(\Xi)\left|G_{ij}-G_{ij}^{[\mu]}\right| \lesssim \|G_{i,[\mu]}\|\,\|\pi^{-1}G_{[\mu],j}\| \lesssim \Lambda_o^2,$$

$$\mathbf{1}(\Xi)\left\|\pi^{-1}\left(G_{[\alpha\beta]}-G_{[\alpha\beta]}^{[\mu]}\right)\pi^{-1}\right\| \lesssim \|\pi^{-1}G_{[\alpha\mu]}\pi^{-1}\|\,\|G_{[\mu\beta]}\pi^{-1}\| \lesssim \Lambda_o^2,$$

and

$$\mathbf{1}(\Xi)\left\|\pi^{-1}\left(G_{[\alpha],i}-G_{[\alpha],i}^{[\mu]}\right)\right\| \lesssim \|\pi^{-1}G_{[\alpha\mu]}\pi^{-1}\|\,\|G_{[\mu],i}\| \lesssim \Lambda_o^2.$$

Thus with an induction on the indices, we obtain that for any $\mathbb{T}\subset\mathcal{I}$ such that $\mathbb{T}=\{i_1,\cdots,i_k,\mu_1,\overline{\mu}_1,\cdots,\mu_l,\overline{\mu}_l\}$ for some fixed integers $k,l\in\mathbb{N}$, we have

$$\begin{aligned}&\mathbf{1}(\Xi)\max_{i,j\in\mathcal{I}_1\cup\mathcal{I}_2}\left|G_{ij}-G_{ij}^{(\mathbb{T})}\right| + \mathbf{1}(\Xi)\max_{\mu,\nu\in\mathcal{I}_3}\left\|\pi^{-1}\left(G_{[\mu\nu]}-G_{[\mu\nu]}^{(\mathbb{T})}\right)\pi^{-1}\right\| \\ &+\mathbf{1}(\Xi)\max_{i\in\mathcal{I}_1\cup\mathcal{I}_2,\mu\in\mathcal{I}_3}\left(\left\|\left(G_{i,[\mu]}-G_{i,[\mu]}^{(\mathbb{T})}\right)\pi^{-1}\right\| + \left\|\pi^{-1}\left(G_{[\mu],i}-G_{[\mu],i}^{(\mathbb{T})}\right)\right\|\right) \lesssim \Lambda_o^2.\end{aligned} \tag{6.22}$$

In particular, we have

$$\mathbf{1}(\Xi)\Lambda_o^{(\mathbb{T})} = \mathrm{O}(\Lambda_o), \quad \mathbf{1}(\Xi)\Lambda^{(\mathbb{T})} = \mathrm{O}(\Lambda), \quad \mathbf{1}(\Xi)\theta^{(\mathbb{T})} = \mathbf{1}(\Xi)\theta + \mathrm{O}(\Lambda_o^2), \tag{6.23}$$

which we shall use tacitly in the proof.

6.2 Entrywise local law

In analogy to [13, Section 3] and [24, Section 5], we introduce the Z variables

$$Z_{\mathfrak{a}}^{(\mathbb{T})} := (1-\mathbb{E}_{\mathfrak{a}})\big(G_{\mathfrak{a}\mathfrak{a}}^{(\mathbb{T})}\big)^{-1}, \quad \mathfrak{a} \notin \mathbb{T},$$

where $\mathbb{E}_{\mathfrak{a}}[\cdot] := \mathbb{E}[\cdot \mid H^{(\mathfrak{a})}]$, i.e. it is the partial expectation over the randomness of the $\mathfrak{a}$-th row and column of H. By (6.4), we have that for $i \in \mathcal{I}_\alpha$, $\alpha = 1,2$,

$$Z_i = (\mathbb{E}_i - 1)\left(WG^{(i)}W^T\right)_{ii} = \sum_{\mu,\nu\in\mathcal{I}_{\alpha+2}} G_{\mu\nu}^{(i)}\left(\frac{1}{n}\delta_{\mu\nu} - W_{i\mu}W_{i\nu}\right). \tag{6.24}$$

We also introduce the matrix value Z variables:

$$Z_{[\mu]}^{(\mathbb{T})} := (1-\mathbb{E}_{[\mu]})\left(G_{[\mu\mu]}^{[\mathbb{T}]}\right)^{-1}, \quad \mu,\overline{\mu} \notin \mathbb{T},$$

where $\mathbb{E}_{[\mu]}[\cdot] := \mathbb{E}[\cdot \mid H^{[\mu]}]$, i.e. it is the partial expectation over the randomness of the μ and $\overline{\mu}$-th rows and columns of H. By (6.10), we have

$$G_{[\mu\mu]}^{-1} = \frac{1}{z-1}\begin{pmatrix} 1 & -z^{-1/2} \\ -z^{-1/2} & 1 \end{pmatrix} - \begin{pmatrix} m_1^{[\mu]} & 0 \\ 0 & m_2^{[\mu]} \end{pmatrix} + Z_{[\mu]}, \tag{6.25}$$

where

$$Z_{[\mu]} = \begin{pmatrix} \sum_{i,j\in\mathcal{I}_1} G_{ij}^{[\mu]}(n^{-1}\delta_{ij} - X_{i\mu}X_{j\mu}) & \sum_{i\in\mathcal{I}_1,j\in\mathcal{I}_2} G_{ij}^{[\mu]} X_{i\mu}Y_{j\overline{\mu}} \\ \sum_{i\in\mathcal{I}_1,j\in\mathcal{I}_2} G_{ji}^{[\mu]} X_{i\mu}Y_{j\overline{\mu}} & \sum_{i,j\in\mathcal{I}_2} G_{ij}^{[\mu]}(n^{-1}\delta_{ij} - Y_{i\overline{\mu}}Y_{j\overline{\mu}}) \end{pmatrix}. \tag{6.26}$$

Then using Lemma 6.5, we can prove the following large deviation estimates for the Z variables and off-diagonal entries.

Lemma 6.7. *Suppose the assumptions in Proposition 6.1 hold. Let $c_0 > 0$ be a sufficiently small constant and fix $C_0, \varepsilon > 0$. Then the following estimates hold uniformly for all $i \in \mathcal{I}_1 \cup \mathcal{I}_2$, $\mu \in \mathcal{I}_3$, and $z = E + \mathrm{i}\eta \in S(\varepsilon)$:*

$$\mathbf{1}(\Xi(z))\left(|Z_i| + \|Z_{[\mu]}\|\right) \prec \phi_n + \Psi_\Lambda; \tag{6.27}$$

$$\mathbf{1}(\Xi(z))\Lambda_o \prec \phi_n + \Psi_\Lambda; \tag{6.28}$$

$$\mathbf{1}(\eta \geqslant 1)\left(|Z_i| + \|Z_{[\mu]}\| + \Lambda_o\right) \prec \phi_n. \tag{6.29}$$

Proof. For $i \in \mathcal{I}_1$, applying Lemma 6.5 to Z_i in (6.24), we get that on Ξ,

$$\begin{aligned}|Z_i| &= \left|\sum_{\mu,\nu\in\mathcal{I}_3} G_{\mu\nu}^{(i)}\left(\frac{1}{n}\delta_{\mu\nu} - X_{i\mu}X_{i\nu}\right)\right| \prec \phi_n + \frac{1}{n}\left(\sum_{\mu,\nu\in\mathcal{I}_3}\left|G_{\mu\nu}^{(i)}\right|^2\right)^{1/2} \\ &\prec \phi_n + \frac{1}{n}\left[\sum_{\nu\in\mathcal{I}_3}\left(1 + \frac{\operatorname{Im}\left(U(z)G_{[\nu\nu]}^{(i)}\right)_{11}}{\eta}\right)\right]^{1/2},\end{aligned}$$

where we used (3.20), $U(z)$ is the 2×2 matrix

$$U(z) := z^{1/2}\begin{pmatrix} \overline{z} & \overline{z}^{1/2} \\ \overline{z}^{1/2} & \overline{z} \end{pmatrix}\begin{pmatrix} z & z^{1/2} \\ z^{1/2} & z \end{pmatrix}^{-1},$$

and the subscript "11" means the $(1,1)$-th entry of the 2×2 matrix. Then using (6.18), (6.23) and $n^{-1/2} \leqslant \phi_n$, we obtain that

$$\mathbf{1}(\Xi(z))\,|Z_i| \prec \phi_n + \left(\frac{\operatorname{Im}(U(z)\pi(z))_{11} + \Lambda}{n\eta}\right)^{1/2} \prec \phi_n + \Psi_\Lambda. \tag{6.30}$$

Here we used that for π in (6.16),

$$\begin{aligned} U(z)\pi(z) &= \frac{1}{2}\left(1 - c_1 - c_2 + \sqrt{(z-\lambda_-)(z-\lambda_+)}\right)\begin{pmatrix} \overline{z}^{1/2} & \overline{z} \\ \overline{z} & \overline{z}^{1/2}\end{pmatrix} \\ &+ \frac{1}{2}\begin{pmatrix} \overline{z} & \overline{z}^{1/2} \\ \overline{z}^{1/2} & \overline{z}\end{pmatrix}\begin{pmatrix} z^{1/2} & -1 \\ -1 & z^{1/2}\end{pmatrix}\begin{pmatrix} 1-2c_1 & -z^{1/2} \\ -z^{1/2} & 1-2c_2,\end{pmatrix}\end{aligned}$$

which, together with (3.2), implies that

$$\|\operatorname{Im}(U(z)\pi(z))\| = \mathrm{O}(\operatorname{Im} m_c(z)). \tag{6.31}$$

Similarly, for $i \in \mathcal{I}_2$, we can prove the same estimate (6.30) for $\mathbf{1}(\Xi)|Z_i|$. Next we pick $\mu \in \mathcal{I}_3$. Applying Lemma 6.5 again to $Z_{[\mu]}$ in (6.26) and using (3.19), we obtain that on Ξ,

$$\begin{aligned} \|Z_{[\mu]}\| &\prec \phi_n + \frac{1}{n}\left(\sum_{i,j\in\mathcal{I}_1\cup\mathcal{I}_2}\left|G^{[\mu]}_{ij}\right|^2\right)^{1/2} \\ &\prec \phi_n + \left(\frac{\operatorname{Im} m_1^{[\mu]}(z) + \operatorname{Im} m_2^{[\mu]}(z)}{n\eta}\right)^{1/2} \prec \phi_n + \Psi_\Lambda. \end{aligned} \tag{6.32}$$

This completes the proof of (6.27).

Then we prove (6.28). For $i \neq j \in \mathcal{I}_1 \cup \mathcal{I}_2$, using (6.5), Lemma 6.5 and Lemma 3.8, we obtain that on Ξ,

$$|G_{ij}| \prec \phi_n + \frac{1}{n}\left(\sum_{\mu,\nu\in\mathcal{I}_3}\left|G^{(i)}_{\mu\nu}\right|^2\right)^{1/2} \prec \phi_n + \Psi_\Lambda. \tag{6.33}$$

For $\mu \neq \nu \in \mathcal{I}_3$, using (6.13), Lemma 6.5 and Lemma 3.8, we obtain that on Ξ,

$$\left\|\pi^{-1}G_{[\mu\nu]}\pi^{-1}\right\| \prec \phi_n + \frac{1}{n}\left(\sum_{i,j\in\mathcal{I}_1\cup\mathcal{I}_2}\left|G^{[\mu]}_{ij}\right|^2\right)^{1/2} \prec \phi_n + \Psi_\Lambda.$$

For $i \in \mathcal{I}_1 \cup \mathcal{I}_2$ and $\mu \in \mathcal{I}_3$, using (6.12), Lemma 6.5 and Lemma 3.8, we obtain that on Ξ,

$$\begin{aligned} \left\|\pi^{-1}G_{[\mu],i}\right\| &\prec \phi_n + \frac{1}{n}\left(\sum_{j\in\mathcal{I}_1\cup\mathcal{I}_2,\nu\in\mathcal{I}_3\cup I_4}\left|G^{(i\mu\overline{\mu})}_{j\nu}\right|^2\right)^{1/2} \\ &\prec \phi_n + \left(\frac{\operatorname{Im} m_1^{(i\mu\overline{\mu})} + \operatorname{Im} m_2^{(i\mu\overline{\mu})}}{n\eta}\right)^{1/2} \prec \Psi_\Lambda. \end{aligned}$$

Thus we conclude (6.28).

The proof of (6.29) is similar, except that when $\eta \geqslant 1$ we use $\mathbf{1}(\eta \geqslant 1)\|G^{(\mathbb{T})}(z)\| = \mathrm{O}(1)$ with high probability by (3.9) and $\|\pi^{-1}(z)\| = \mathrm{O}(1)$. For example, for the estimate (6.32), we have that for $\eta \geqslant 1$,

$$\|Z_{[\mu]}\| \prec \phi_n + \left(\frac{\operatorname{Im} m_1^{[\mu]}(z) + \operatorname{Im} m_2^{[\mu]}(z)}{n\eta}\right)^{1/2} \prec \phi_n + n^{-1/2} = \mathrm{O}(\phi_n).$$

We omit the rest of the details. □

A key component of the proof for Proposition 6.1 is an analysis of the self-consistent equation. Recall the equations in (2.34)-(2.36).

Lemma 6.8. *Fix any constant $\varepsilon > 0$. The following estimates hold uniformly in $z \in S(\varepsilon)$:*

$$\mathbf{1}(\Xi)\left(\left|m_1 + c_1 m_3^{-1}\right| + \left|m_2 + c_2 m_4^{-1}\right|\right) \prec \phi_n + \Psi_\Lambda, \tag{6.34}$$

$$\mathbf{1}(\Xi)\left|m_3^2 + \left[(2c_1 - 1)z - c_1 + c_2\right] m_3 + c_1(c_1 - 1)z(z-1)\right| \prec \phi_n + \Psi_\Lambda. \tag{6.35}$$

Moreover, we have the finer estimates

$$\mathbf{1}(\Xi)\left(\left|m_1 + c_1 m_3^{-1}\right| + \left|m_2 + c_2 m_4^{-1}\right|\right) \prec |\langle Z\rangle_1| + |\langle Z\rangle_2| + \phi_n^2 + \Psi_\Lambda^2, \tag{6.36}$$

$$\begin{aligned}\mathbf{1}(\Xi)&\left|m_3^2 + \left[(2c_1 - 1)z - c_1 + c_2\right] m_3 + c_1(c_1 - 1)z(z-1)\right| \\ &\prec |\langle Z\rangle_1| + |\langle Z\rangle_2| + \|[Z]\| + \phi_n^2 + \Psi_\Lambda^2,\end{aligned} \tag{6.37}$$

where

$$\langle Z\rangle_1 := \frac{1}{n}\sum_{i\in\mathcal{I}_1} Z_i, \quad \langle Z\rangle_2 := \frac{1}{n}\sum_{j\in\mathcal{I}_2} Z_j, \quad [Z] := \frac{1}{n}\sum_{\mu\in\mathcal{I}_3} G_{[\mu\mu]}. \tag{6.38}$$

Finally, there exists a constant $C_0 > 0$ such that

$$\mathbf{1}(C_0 \leqslant \eta \leqslant 2C_0)\left(\left|m_1 + c_1 m_3^{-1}\right| + \left|m_2 + c_2 m_4^{-1}\right|\right) \prec \phi_n, \tag{6.39}$$

$$\mathbf{1}(C_0 \leqslant \eta \leqslant 2C_0)\left|m_3^2 + \left[(2c_1 - 1)z - c_1 + c_2\right] m_3 + c_1(c_1 - 1)z(z-1)\right| \prec \phi_n. \tag{6.40}$$

Proof. We first prove (6.36) and (6.37), from which (6.34) and (6.35) follow due to (6.27). By (6.4), (6.24) and (6.25), we have that for $i \in \mathcal{I}_1$, $j \in \mathcal{I}_2$ and $\mu \in \mathcal{I}_3$,

$$\frac{1}{G_{ii}} = -m_3 + \varepsilon_i, \quad \frac{1}{G_{jj}} = -m_4 + \varepsilon_j, \tag{6.41}$$

and

$$G_{[\mu\mu]}^{-1} = \frac{1}{z-1}\begin{pmatrix} 1 & -z^{-1/2} \\ -z^{-1/2} & 1 \end{pmatrix} - \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} + \varepsilon_\mu, \tag{6.42}$$

where

$$\varepsilon_i := Z_i + \left(m_3 - m_3^{(i)}\right) + \mathrm{O}(n^{-10}), \quad \varepsilon_j := Z_j + \left(m_4 - m_4^{(j)}\right) + \mathrm{O}(n^{-10}),$$

and

$$\varepsilon_\mu := Z_\mu + \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} - \begin{pmatrix} m_1^{[\mu]} & 0 \\ 0 & m_2^{[\mu]} \end{pmatrix}.$$

By (6.22), (6.27) and (6.28), we have

$$\mathbf{1}(\Xi)\left(|\varepsilon_i| + |\varepsilon_j| + \|\varepsilon_\mu\|\right) \prec \phi_n + \Psi_\Lambda, \tag{6.43}$$

and

$$\mathbf{1}(\Xi)\left(|m_1 - m_1^{[\mu]}| + |m_2 - m_2^{[\mu]}| + |m_3 - m_3^{(i)}| + |m_4 - m_4^{(j)}|\right) \prec \phi_n^2 + \Psi_\Lambda^2. \tag{6.44}$$

Now using (6.41), (6.43), (6.44), (3.1) and the definition of Ξ, we can obtain that for $i \in \mathcal{I}_1$ and $j \in \mathcal{I}_2$,

$$\begin{aligned} \mathbf{1}(\Xi)G_{ii} &= \mathbf{1}(\Xi)\left(-\frac{1}{m_3} - \frac{Z_i}{m_3^2} + \mathrm{O}_\prec\left(\phi_n^2 + \Psi_\Lambda^2\right)\right), \\ \mathbf{1}(\Xi)G_{jj} &= \mathbf{1}(\Xi)\left(-\frac{1}{m_4} - \frac{Z_j}{m_4^2} + \mathrm{O}_\prec\left(\phi_n^2 + \Psi_\Lambda^2\right)\right). \end{aligned} \tag{6.45}$$

Taking average $\frac{1}{n}\sum_{i\in\mathcal{I}_1}$ and $\frac{1}{n}\sum_{j\in\mathcal{I}_2}$, we get

$$\begin{aligned} \mathbf{1}(\Xi)m_1 &= \mathbf{1}(\Xi)\left(-\frac{c_1}{m_3} - \frac{\langle Z\rangle_1}{m_3^2} + \mathrm{O}_\prec\left(\phi_n^2 + \Psi_\Lambda^2\right)\right), \\ \mathbf{1}(\Xi)m_2 &= \mathbf{1}(\Xi)\left(-\frac{c_2}{m_4} - \frac{\langle Z\rangle_2}{m_4^2} + \mathrm{O}_\prec\left(\phi_n^2 + \Psi_\Lambda^2\right)\right), \end{aligned} \tag{6.46}$$

which proves (6.36). On the other hand, using (6.42), (6.43), (6.44) and the definition of Ξ, we obtain that for $\mu \in \mathcal{I}_3$,

$$\mathbf{1}(\Xi)G_{[\mu\mu]} = \mathbf{1}(\Xi)\left(\widetilde{\pi}^{-1} + \varepsilon_\mu\right)^{-1} = \mathbf{1}(\Xi)\left(\widetilde{\pi} - \widetilde{\pi}\varepsilon_\mu\widetilde{\pi} + \mathrm{O}_\prec(\phi_n^2 + \Psi_\Lambda^2)\right). \tag{6.47}$$

where we define $\widetilde{\pi}(z)$ as

$$\widetilde{\pi}(z)^{-1} = \frac{1}{z-1}\begin{pmatrix} 1-(z-1)m_1 & -z^{-1/2} \\ -z^{-1/2} & 1-(z-1)m_2 \end{pmatrix}. \tag{6.48}$$

(Note $\widetilde{\pi}$ is actually a random version of π in (6.17).) On Ξ, we have the estimate $\pi^{-1}\widetilde{\pi} = 1 + \mathrm{O}(\Lambda) = 1 + \mathrm{O}((\log n)^{-1})$. Hence taking average of the $(1,1)$-th entry of (6.47) over μ, we get that

$$\begin{aligned} \mathbf{1}(\Xi)m_3 = \mathbf{1}(\Xi)\Bigg[&\frac{1-(z-1)m_2(z)}{z^{-1} - (m_1(z) + m_2(z)) + (z-1)m_1(z)m_2(z)} \\ &- (\widetilde{\pi}[Z]\widetilde{\pi})_{11} + \mathrm{O}_\prec\left(\phi_n^2 + \Psi_\Lambda^2\right)\Bigg]. \end{aligned} \tag{6.49}$$

The plugging (6.46) into (6.49), and using (3.1) and the definition of Ξ, we can obtain that

$$\begin{aligned} \mathbf{1}(\Xi)\left[z^{-1}m_3 + m_3\left(\frac{c_1}{m_3} + \frac{c_2}{m_4}\right) + (c_1 - 1)(z-1)\frac{c_2}{m_4} - 1\right] \\ \prec |\langle Z\rangle_1| + |\langle Z\rangle_2| + \|[Z]\| + \phi_n^2 + \Psi_\Lambda^2. \end{aligned} \tag{6.50}$$

Then using (3.16) and rearranging terms, we can obtain (6.37).

Then we prove (6.39) and (6.40). When $\eta \geqslant C_0$, by (6.29) we have

$$\max_{i\in\mathcal{I}_1\cup\mathcal{I}_2}|Z_i| + \max_{\mu\in\mathcal{I}_3}\|Z_{[\mu]}\| + \Lambda_o \prec \phi_n.$$

On the other hand, applying (3.9) and Lemma 3.3 to (6.41) and (6.42), we obtain that

$$\max_{i\in\mathcal{I}_1\cup\mathcal{I}_2} G_{ii}^{-1} + \max_{\mu\in\mathcal{I}_3}\|G_{[\mu\mu]}^{-1}\| = \mathrm{O}(1)$$

with high probability when $\eta \geqslant C_0$. Together with (6.7) and (6.14), we get that

$$|m_1 - m_1^{[\mu]}| + |m_2 - m_2^{[\mu]}| + |m_3 - m_3^{(i)}| + |m_4 - m_4^{(j)}| \prec \Lambda_o^2 \prec \phi_n^2. \tag{6.51}$$

Moreover, we still have

$$\mathbf{1}(C_0 \leqslant \eta \leqslant 2C_0)(|\varepsilon_i| + |\varepsilon_j| + \|\varepsilon_\mu\|) \prec \phi_n. \tag{6.52}$$

Then going through the previous argument on event Ξ, one can see that in order to prove (6.39) and (6.40), it suffices to bound m_3^{-1}, m_4^{-1} and $\|\widetilde{\pi}\|$ from above. In particular, it suffices to prove the following bounds: with high probability,

$$\mathbf{1}(C_0 \leqslant \eta \leqslant 2C_0)\left[|m_3|^{-1} + |m_4|^{-1}\right] \leqslant C \tag{6.53}$$

and

$$\mathbf{1}(C_0 \leqslant \eta \leqslant 2C_0)|z^{-1} - (m_1 + m_2) + (z-1)m_1 m_2|^{-1} \leqslant C \tag{6.54}$$

for some constant $C > 0$.

First using (3.9) and Lemma 3.3, we obtain that for $C_0 \leqslant \eta \leqslant 2C_0$, with high probability,

$$|m_1| + |m_2| + |m_3| + |m_4| \leqslant \frac{C}{C_0}, \tag{6.55}$$

for some constant $C > 0$ that is independent of C_0. Using the spectral decomposition (3.10) and (3.11), it is easy to see that $\operatorname{Im} m_1(z) \geqslant 0$ and $\operatorname{Im} m_2(z) \geqslant 0$. Hence we have

$$\left|\frac{1}{z-1} - m_1\right| \geqslant -\operatorname{Im}\frac{1}{z-1} \geqslant cC_0^{-1}, \quad \left|\frac{1}{z-1} - m_2\right| \geqslant -\operatorname{Im}\frac{1}{z-1} \geqslant cC_0^{-1}, \tag{6.56}$$

for some constant $c > 0$ that is independent of C_0. Thus we obtain that

$$\left|\left(\frac{1}{z-1} - m_1\right)\left(\frac{1}{z-1} - m_2\right) - \frac{1}{z(z-1)^2}\right| \geqslant c^2 C_0^{-2} - \frac{1}{|z||z-1|^2} \geqslant \frac{1}{2}c^2 C_0^{-2}$$

as long as C_0 is taken large enough, which then implies (6.54). Now by (6.48), (6.54) and (6.55), we know that $\widetilde{\pi}(z) = \mathrm{O}(1)$ with high probability. Thus as in (6.49), we can derive from (6.47) and (6.52) that

$$m_3 = \frac{1 - (z-1)m_2(z)}{z^{-1} - (m_1(z) + m_2(z)) + (z-1)m_1(z)m_2(z)} + \mathrm{O}_\prec(\phi_n).$$

Then using (6.56) and (6.54), we obtain that with high probability, $|m_3| \geqslant c$ for some constant $c > 0$. Similarly, we can obtain the same bound for m_4. This gives (6.53). ☐

The following lemma gives the stability of the equation $f_3(u,z) = 0$, where

$$f_3(u,z) := u^2(z) + \left[(2c_1 - 1)z - c_1 + c_2\right]u(z) + c_1(c_1 - 1)z(z-1).$$

Roughly speaking, it states that if $f(m_3(z), z)$ is small and $m_3(\widetilde{z}) - m_{3c}(\widetilde{z})$ is small for $\operatorname{Im}\widetilde{z} \geqslant \operatorname{Im} z$, then $m_3(z) - m_{3c}(z)$ is small. For an arbitrary $z \in S(\varepsilon)$, we define the discrete set

$$L(z) := \{z\} \cup \{z' \in S(c_0, C_0, \varepsilon) : \operatorname{Re} z' = \operatorname{Re} z, \operatorname{Im} z' \in [\operatorname{Im} z, \varepsilon^{-1}] \cap (n^{-100}\mathbb{N})\}.$$

Thus, if $\operatorname{Im} z \geqslant \varepsilon^{-1}$, then $L(z) = \{z\}$; if $\operatorname{Im} z < \varepsilon^{-1}$, then $L(z)$ is a 1-dimensional lattice with spacing n^{-100} plus the point z.

Lemma 6.9. *Fix a constant $\varepsilon > 0$. The self-consistent equation $f_3(u,z) = 0$ is stable on $S(\varepsilon)$ in the following sense. Suppose the z-dependent function δ satisfies $n^{-2} \leqslant \delta(z) \leqslant (\log n)^{-1}$ for $z \in S(\varepsilon)$ 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* $u_3 : S(\varepsilon) \to \mathbb{C}$ *is the Stieltjes transform of a measure* μ_3 *with* $\mu_3(\mathbb{R}) = \mathrm{O}(1)$. *Let* $z \in S(\varepsilon)$ *and suppose that for all* $z' \in L(z)$ *we have*

$$\left|f_3(z', u_3)\right| \leqslant \delta(z'). \tag{6.57}$$

Then we have

$$|u_3(z) - m_{3c}(z)| \leqslant \frac{C\delta}{\sqrt{\kappa + \eta + \delta}}, \tag{6.58}$$

for some constant $C > 0$ *independent of* z *and* n, *where* κ *is defined in (2.41).*

Proof. This lemma can proved with the same method as in e.g. [5, Lemma 4.5] and [24, Appendix A.2]. The only inputs are the form of the function f_3 and Lemma 3.2. ☐

Note that by Lemma 6.9, (6.39), (6.40) and (3.16), we immediately get that

$$\mathbf{1}(C_0 \leqslant \eta \leqslant 2C_0)|m_\alpha(z) - m_{\alpha c}(z)| \prec \phi_n, \quad \alpha = 1, 2, 3, 4. \tag{6.59}$$

From (6.29), we obtain the off-diagonal estimate

$$\mathbf{1}(C_0 \leqslant \eta \leqslant 2C_0)\Lambda_o(z) \prec \phi_n. \tag{6.60}$$

Plugging (6.29) and (6.59) into (6.41), we get that

$$\mathbf{1}(C_0 \leqslant \eta \leqslant 2C_0) \max_{i \in \mathcal{I}_1 \cup \mathcal{I}_2} |G_{ii} - \Pi_{ii}| \prec \phi_n. \tag{6.61}$$

Plugging (6.29) and (6.59) into (6.42), we obtain that for $C_0 \leqslant \eta \leqslant 2C_0$,

$$G_{[\mu\mu]}^{-1}(z) = \pi^{-1}(z) + \mathrm{O}_\prec(\phi_n) \Rightarrow \max_{\mu \in \mathcal{I}_3} \left\|\pi^{-1}\left(G_{[\mu\mu]} - \pi\right)\pi^{-1}\right\| \prec \phi_n. \tag{6.62}$$

Starting from these initial resolvent estimates, using a standard continuity (in z) argument, the self-consistent estimates (6.34)-(6.35), and Lemma 6.7, we can prove the following weak version of (6.3).

Lemma 6.10 (Weak entrywise local law). *For any small constant* $\varepsilon > 0$, *we have*

$$\Lambda(z) \prec \phi_n^{1/2} + (n\eta)^{-1/4}, \tag{6.63}$$

uniformly in $z \in S(\varepsilon)$.

Proof. One can prove this lemma using a continuity argument as in e.g. [5, Section 4.1], [12, Section 5.3] or [13, Section 3.6]. The key inputs are Lemmas 6.7, Lemma 6.8, Lemma 6.9, and the estimates (6.59)-(6.62) in the $C_0 \leqslant \eta \leqslant 2C_0$ case. All the other parts of the proof are essentially the same. ☐

To get the strong entrywise local law as in (6.3), we need stronger bounds on $\langle Z\rangle_1$, $\langle Z\rangle_2$ and $[Z]$ in (6.38). They follow from the following *fluctuation averaging lemma.*

Lemma 6.11 (Fluctuation averaging). *Suppose* Φ *and* Φ_o *are positive,* n*-dependent deterministic functions on* $S(\varepsilon)$ *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* $z \in S(\varepsilon)$ *we have*

$$|\langle Z\rangle_1| + |\langle Z\rangle_2| + \|[Z]\| \prec \Phi_o^2. \tag{6.64}$$

Proof. The bound on $|\langle Z\rangle_1| + |\langle Z\rangle_2|$ can be proved in the same way as [12, Theorem 4.7]. The bound on $\|[Z]\|$ can be proved in the same way as [31, Lemma 4.9]. ☐

Now we give the proof of Proposition 6.1.

Proof of Proposition 6.1. By Lemma 6.10, the event Ξ holds with high probability. Then by Lemma 6.10 and Lemma 6.7, we can take

$$\Phi_o = \phi_n + \sqrt{\frac{\operatorname{Im} m_c + \Phi}{n\eta}} + \frac{1}{n\eta}, \quad \Phi = \phi_n^{1/2} + (n\eta)^{-1/4}, \tag{6.65}$$

in Lemma 6.11. Then (6.37) gives

$$|f_3(z, m_3)| \prec \phi_n^2 + \frac{\operatorname{Im} m_c + \Phi}{n\eta} \lesssim \phi_n^2 + \frac{\operatorname{Im} m_c}{n\eta} + \frac{n^{2\tau}}{(n\eta)^2} + n^{-2\tau}\Phi^2$$

for any fixed constant $\tau > 0$. Using Lemma 6.9, we get

$$\begin{aligned} |m_3 - m_{3c}| &\prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{\operatorname{Im} m_c}{n\eta\sqrt{\kappa+\eta}} + \frac{n^\tau}{n\eta} + n^{-\tau}\Phi \\ &\prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{n^\tau}{n\eta} + n^{-\tau}\Phi, \end{aligned} \tag{6.66}$$

where we used $\operatorname{Im} m_c = \mathrm{O}(\sqrt{\kappa+\eta})$ by (3.2) in the second step. With (6.36), (3.16) and (6.66), we get the same bound for m_α, $\alpha = 1, 2, 3, 4$,

$$|m_\alpha(z) - m_{\alpha c}(z)| \prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{n^\tau}{n\eta} + n^{-\tau}\Phi. \tag{6.67}$$

Plugging (6.67) into (6.41) and using (6.43), we obtain that

$$\max_{i\in\mathcal{I}_1\cup\mathcal{I}_2} |G_{ii} - \Pi_{ii}| \prec \Phi_o + \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{n^\tau}{n\eta} + n^{-\tau}\Phi \prec \phi_n + \Psi(z) + \frac{n^\tau}{n\eta} + n^{-\tau}\Phi.$$

Similarly plugging (6.67) into (6.42) and using (6.43), we obtain that

$$\max_{\mu\in\mathcal{I}_3} \left\|\pi^{-1}(G_{[\mu\mu]} - \pi)\pi^{-1}\right\| \prec \phi_n + \Psi(z) + \frac{n^\tau}{n\eta} + n^{-\tau}\Phi.$$

Finally by (6.28), we have

$$\Lambda_o \prec \phi_n + \Psi(z) + \frac{n^\tau}{n\eta} + n^{-\tau}\Phi.$$

In sum, we obtain a self-improving estimate on Λ:

$$\Lambda \prec \Phi \;\Rightarrow\; \Lambda \prec \phi_n + \Psi(z) + \frac{n^\tau}{n\eta} + n^{-\tau}\Phi.$$

Afer $\mathrm{O}(\tau^{-1})$ many iterations, we obtain that

$$\Lambda \prec \phi_n + \Psi(z) + \frac{n^\tau}{n\eta}.$$

Since τ can be arbitrarily small, we conclude (6.3). □

Finally, we prove the weak averaged local laws in Theorem 2.11.

Proof of (2.44) *and* (2.45). We can repeat the argument at the beginning of the above proof of Proposition 6.1. Taking $\Phi = \Phi_o = \phi_n + \Psi$, (6.37) and Lemma 6.11 give that

$$|f_3(z, m_3)| \prec \phi_n^2 + \Psi^2(z) \lesssim \phi_n^2 + \frac{\operatorname{Im} m_c}{n\eta} + \frac{1}{(n\eta)^2}.$$

Using Lemma 6.9, we get

$$|m_3 - m_{3c}| \prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{\operatorname{Im} m_c}{n\eta\sqrt{\kappa+\eta}} + \frac{1}{n\eta} \prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{1}{n\eta}. \tag{6.68}$$

With (6.36), (3.16) and (6.68), we conclude (2.44).

For $z \in S_{out}(\varepsilon)$, we use Lemma 6.9 again to get that

$$\begin{aligned} |m_3 - m_{3c}| &\prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{\operatorname{Im} m_c}{n\eta\sqrt{\kappa+\eta}} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}} \\ &\prec \min\left\{\phi_n, \frac{\phi_n^2}{\sqrt{\kappa+\eta}}\right\} + \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}, \end{aligned} \tag{6.69}$$

where we used the stronger bound $\operatorname{Im} m_c = \mathrm{O}(\eta/\sqrt{\kappa+\eta})$ by (3.2) in the second step. With (6.36), (3.16) and (6.69), we conclude (2.45). □

6.3 A centralization argument

In this subsection, we discuss how to relax the assumptions (6.1) and (6.2) to the weaker ones in (2.4) and (2.5). First, under the relaxed variance assumption (2.5), the only differences from the previous argument in Section 6.2 are the equations (6.41) and (6.42). More precisely, we now have

$$\mathbb{E}_i(WG^{(i)}W^T)_{ii} = m_{3,4}^{(i)} + \mathrm{O}(n^{-1-\tau}\|G\|_{\max}), \quad \text{if } i \in \mathcal{I}_{1,2},$$

and hence ε_i in (6.41) will contain an extra error $\mathrm{O}(n^{-1-\tau}\|G\|_{\max})$. Similarly, the term ε_μ in (6.42) will also contain this kind of error. This extra error will lead to an negligible term $\mathrm{O}(n^{-1-\tau})$ in all the bounds of Theorem 2.11, and hence does not affect our results.

Then we relax (6.1) to (2.4). For X and Y satisfying the assumptions in Theorem 2.5, we write $X = X_1 + \mathcal{E}_1$ and $Y = Y_1 + \mathcal{E}_2$, where $\mathcal{E}_1 := \mathbb{E}X$ and $\mathcal{E}_2 := \mathbb{E}Y$. Then X_1 and Y_1 are random matrices satisfying the assumptions in Theorem 2.5 and (6.1), and $\mathcal{E}_1$ and $\mathcal{E}_2$ are deterministic matrices such that

$$\max_{i\in\mathcal{I}_1,\mu\in\mathcal{I}_3} |(\mathcal{E}_1)_{i\mu}| + \max_{j\in\mathcal{I}_2,\nu\in\mathcal{I}_4} |(\mathcal{E}_2)_{j\nu}| \leqslant n^{-2-\tau}. \tag{6.70}$$

We denote $G_1(z) := H_1^{-1}(z)$ and $G(z) := [H_1(z) + V]^{-1}$, where

$$H_1(\lambda) := \begin{pmatrix} 0 & \begin{pmatrix} X_1 & 0 \\ 0 & Y_1 \end{pmatrix} \\ \begin{pmatrix} X_1^T & 0 \\ 0 & Y_1^T \end{pmatrix} & \begin{pmatrix} \lambda I_n & \lambda^{1/2} I_n \\ \lambda^{1/2} I_n & \lambda I_n \end{pmatrix}^{-1} \end{pmatrix}, \quad V := \begin{pmatrix} 0 & \begin{pmatrix} \mathcal{E}_1 & 0 \\ 0 & \mathcal{E}_2 \end{pmatrix} \\ \begin{pmatrix} \mathcal{E}_1^T & 0 \\ 0 & \mathcal{E}_2^T \end{pmatrix} & 0 \end{pmatrix}.$$

Lemma 6.12. *If Theorem 2.11 holds for G_1, then it also holds for G.*

Proof. We expand G using the resolvent expansion

$$G = G_1 - G_1VG_1 + (G_1V)^2G_1 - (G_1V)^3G. \tag{6.71}$$

For any deterministic unit vector $\mathbf{v} \in \mathbb{C}^{\mathcal{I}}$, we have

$$\begin{aligned}|\langle \mathbf{v}, G_1VG_1\mathbf{v}\rangle| &\leqslant 2\sum_{\mu\in\mathcal{I}_3\cup\mathcal{I}_4}\Big|\sum_{i\in\mathcal{I}_1\cup\mathcal{I}_2}(G_1)_{\mathbf{v}i}\,V_{i\mu}\Big||\,(G_1)_{\mu\mathbf{v}}\,| \\ &\prec \max_{\mu}\Big(\sum_{i\in\mathcal{I}_1\cup\mathcal{I}_2}|V_{i\mu}|^2\Big)^{1/2}\sum_{\mu\in\mathcal{I}_3\cup\mathcal{I}_4}|\,(G_1)_{\mu\mathbf{v}}\,| \\ &\prec n^{-1-\tau}\Big(\sum_{\mu\in\mathcal{I}_3\cup\mathcal{I}_4}|\,(G_1)_{\mu\mathbf{v}}\,|^2\Big)^{1/2} \prec n^{-1-\tau}\eta^{-1/2},\end{aligned}$$

where in the second step we used (2.42) for G_1 with vectors $\mathbf{v}$ and $\sum_i V_{i\mu}\mathbf{e}_i$, in the third step the Cauchy-Schwarz inequality and (6.70), and in the last step Lemma 3.8 and (2.42) for G_1. Together with a simple application of the polarization identity, we obtain the bound

$$|\langle \mathbf{v}, G_1VG_1\mathbf{w}\rangle| \prec n^{-1-\tau}\eta^{-1/2}, \tag{6.72}$$

for any deterministic unit vectors $\mathbf{v}, \mathbf{w} \in \mathbb{C}^{\mathcal{I}}$. With a similar argument, we obtain that

$$\big|\langle \mathbf{v}, (G_1V)^2G_1\mathbf{w}\rangle\big| \prec n^{-2-2\tau}\eta^{-1}. \tag{6.73}$$

Combining this estimate with the rough bound (3.9) for G, we get that

$$\begin{aligned}\big|\langle \mathbf{v}, (G_1V)^3G\mathbf{w}\rangle\big| &= \Big|\sum_{i\in\mathcal{I}_1\cup\mathcal{I}_2,\mu\in\mathcal{I}_3\cup\mathcal{I}_4}\big((G_1V)^2G_1\big)_{\mathbf{v}i}\,V_{i\mu}G_{\mu\mathbf{w}}\Big| \\ &\quad+\Big|\sum_{i\in\mathcal{I}_1\cup\mathcal{I}_2,\mu\in\mathcal{I}_3\cup\mathcal{I}_4}\big((G_1V)^2G_1\big)_{\mathbf{v}\mu}\,V_{\mu i}G_{i\mathbf{w}}\Big| \\ &\prec \eta^{-1}\left[\sum_{\mu}\Big|\sum_{i}\big((G_1V)^2G_1\big)_{\mathbf{v}i}\,V_{i\mu}\Big|^2 + \sum_{i}\Big|\sum_{\mu}\big((G_1V)^2G_1\big)_{\mathbf{v}\mu}\,V_{\mu i}\Big|^2\right]^{1/2} \\ &\prec \big(n^{-2-2\tau}\eta^{-1}\big)\,\eta^{-1}\Big(\sum_{i,\mu}|V_{i\mu}|^2\Big)^{1/2} \leqslant n^{-2-3\tau}\eta^{-1},\end{aligned} \tag{6.74}$$

where we used $\eta \gg n^{-1}$ in the last step. Plugging the estimates (6.72)-(6.74) into (6.71), we conclude that

$$|\langle \mathbf{v}, G\mathbf{w}\rangle - \langle \mathbf{v}, G_1\mathbf{w}\rangle| \prec n^{-1-\tau}\eta^{-1/2}, \tag{6.75}$$

for any deterministic unit vectors $\mathbf{v} \in \mathbb{C}^{\mathcal{I}}$. This concludes (2.42) and (2.44) for $G(z)$.

For (2.45), the bounds (6.73) and (6.74) are already good enough. It remains to show that

$$\left|\frac{1}{n}\sum_{\mathfrak{a}\in\mathcal{I}_\alpha}(G_1VG_1)_{\mathfrak{a}\mathfrak{a}}\right| \prec \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}}, \quad \alpha = 1,2,3,4,. \tag{6.76}$$

Using Lemma 3.8, we obtain that for $\alpha = 1, 2$,

$$\left|\frac{1}{n}\sum_{i\in\mathcal{I}_\alpha}(G_1VG_1)_{ii}\right| \prec n^{-2-\tau}\sum_{j\in\mathcal{I}_1\cup\mathcal{I}_2,\mu\in\mathcal{I}_3\cup\mathcal{I}_4}\frac{1}{n}\sum_{i\in\mathcal{I}_\alpha}|(G_1)_{ij}(G_1)_{\mu i}|$$

$$
\begin{aligned}
&\prec n^{-\tau} \max_{\mu\in\mathcal{I}_3\cup\mathcal{I}_4}\left(\frac{1}{n}+\frac{\operatorname{Im}(\mathcal{G}_L)_{jj}+\operatorname{Im}\left(\mathcal{G}_R\mathcal{U}^T\right)_{\mu\mu}}{n\eta}\right) \prec n^{-1-\tau}+n^{-\tau}\frac{\operatorname{Im} m_c+\phi_n+\Psi(z)}{n\eta}\\
&\prec n^{-\tau}\left(\phi_n^2+\Psi^2(z)\right) \prec \frac{n^{-\tau}}{n(\kappa+\eta)}+\frac{n^{-\tau}}{(n\eta)^2\sqrt{\kappa+\eta}},
\end{aligned}
$$

where in the third step we used (2.42) for G_1 and (6.31). The proof for the $\alpha = 3,4$ case is similar. This concludes (2.45). ☐

7 Proof of Theorem 2.11: the anisotropic local law

To conclude Theorem 2.11, it remains to prove the anisotropic local law (2.42). For any vector $\mathbf{u} \in \mathbb{C}^{\mathcal{I}}$ and $\mu \in \mathcal{I}_3$, we denote $u_{[\mu]} := \begin{pmatrix} u_\mu \\ u_{\overline{\mu}} \end{pmatrix}$. By the entrywise local law (6.3), we have that for deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$,

$$
\begin{aligned}
|\langle \mathbf{u}, (G(z)-\Pi(z))\,\mathbf{v}\rangle| \prec \phi_n+\Psi(z) &+ \Big|\sum_{i\neq j\in\mathcal{I}_1\cup\mathcal{I}_2}\overline{u}_i G_{ij} v_j\Big| + \Big|\sum_{\mu\neq\nu\in\mathcal{I}_3} u_{[\mu]}^* G_{[\mu\nu]} v_{[\nu]}\Big| \\
&+ \Big|\sum_{i\in\mathcal{I}_1\cup\mathcal{I}_2,\mu\in\mathcal{I}_3}\overline{u}_i G_{i,[\mu]} v_{[\mu]}\Big| + \Big|\sum_{i\in\mathcal{I}_1\cup\mathcal{I}_2,\mu\in\mathcal{I}_3} u_{[\mu]}^* G_{[\mu],i} v_i\Big|.
\end{aligned}
\tag{7.1}
$$

Note that applying the entrywise local law naively, one can only get that

$$
|\langle \mathbf{u}, (G(z)-\Pi(z))\,\mathbf{v}\rangle| \prec (\phi_n+\Psi(z))\|\mathbf{u}\|_1\|\mathbf{v}\|_1 \leqslant n(\phi_n+\Psi(z)),
$$

using $\|\mathbf{u}\|_1 \leqslant n^{1/2}\|\mathbf{u}\|_2$ and $\|\mathbf{v}\|_1 \leqslant n^{1/2}\|\mathbf{v}\|_2$. To get (2.42), we need to explore the cancellations (due to the random signs of the G entries) in the four sums on the right hand side of (7.1).

We can simplify the problem a little bit. We first notice that by polarization identity of inner products, it suffices to take $\mathbf{u} = \mathbf{v}$ in (7.1). Moreover, since G is symmetric, the last two terms on the right hand side of (7.1) can be bounded in the same way. Then with Markov's inequality, it suffices to prove the following lemma. With Lemma 6.12, it suffices to assume (6.1) for X and Y.

Lemma 7.1. *Suppose* (6.1) *and* (6.3) *hold. Let* $\mathbf{v} \in \mathbb{C}^{\mathcal{I}}$ *be any deterministic unit vector. Then for any* $a \in \mathbb{N}$*, we have the following bounds:*

$$
\mathbb{E}\Big|\sum_{i\neq j\in\mathcal{I}_1\cup\mathcal{I}_2}\overline{v}_i G_{ij} v_j\Big|^{2a} \prec \Phi^{2a}; \tag{7.2}
$$

$$
\mathbb{E}\Big|\sum_{\mu\neq\nu\in\mathcal{I}_3} v_{[\mu]}^* G_{[\mu\nu]} v_{[\nu]}\Big|^{2a} \prec \Phi^{2a}; \tag{7.3}
$$

$$
\mathbb{E}\Big|\sum_{i\in\mathcal{I}_1\cup\mathcal{I}_2,\mu\in\mathcal{I}_3}\overline{v}_i G_{i,[\mu]} v_{[\mu]}\Big|^{2a} \prec \Phi^{2a}. \tag{7.4}
$$

Here we denote $\Phi := \phi_n + \Psi(z)$ *for simplicity.*

The proof of Lemma 7.1 is based on a polynomialization method developed in [5, section 5]. We first give the proof of (7.2) in Section 7.1, which is the easiest, and then give the proof of (7.3) in Section 7.2, which is the hardest. The proof of (7.4) is an easier version of (7.3), and will be omitted.

7.1 Proof of (7.2)

For the proof of (7.2), we will adopt an argument in [30, Appendix A.4]. Recall that in (2.11), we assumed that the X and Y entries have finite third moments. Together with the bounded support condition, we get

$$\mathbb{E}|x_{i\mu}|^n \prec \phi_n^{n-3} n^{-3/2}, \quad \mathbb{E}|y_{j\nu}|^n \prec \phi_n^{n-3} n^{-3/2}, \quad i \in \mathcal{I}_1, \quad j \in \mathcal{I}_2, \quad \mu \in \mathcal{I}_3, \quad \nu \in \mathcal{I}_4. \tag{7.5}$$

Note that we have a stronger fourth moment assumption in (2.11), but it is not necessary for the proof in this section.

We first rewrite the product in (7.2) as

$$\Big|\sum_{i\neq j} \overline{v}_i G_{ij} v_j\Big|^{2a} = \sum_{i_k \neq j_k \in \mathcal{I}_1} \prod_{k=1}^{a} \overline{v}_{i_k} G_{i_k j_k} v_{j_k} \cdot \prod_{k=a+1}^{2a} \overline{\overline{v}_{i_k} G_{i_k j_k} v_{j_k}}. \tag{7.6}$$

To organize the sum over indices, we consider all possible partitions of the indices, such that two indices always take the same value if they are in the same partition, and different values otherwise. We use symbol-to-symbol functions to represent the partitions: a partition Γ denotes a map

$$\Gamma : \{i_1, \cdots, i_{2a}, j_1, \cdots, j_{2a}\} \to L(\Gamma), \quad L(\Gamma) = (b_1, \cdots, b_{n(\Gamma)}),$$

where $\Gamma^{-1}(b_k)$ is an equivalence class of the partition, $n(\Gamma)$ is the number of equivalence classes, and b_k are indices taking values in $\mathcal{I}_1 \cup \mathcal{I}_2$. Then we can write (7.6) as

$$\sum_{\Gamma} \sum_{b_1,\ldots,b_{n(\Gamma)}}^{*} \prod_{k=1}^{a} \overline{v}_{\Gamma(i_k)} G_{\Gamma(i_k)\Gamma(j_k)} v_{\Gamma(j_k)} \cdot \prod_{k=a+1}^{2a} \overline{\overline{v}_{\Gamma(i_k)} G_{\Gamma(i_k)\Gamma(j_k)} v_{\Gamma(j_k)}},$$

where Γ ranges over all the partitions, and $\sum^*$ denotes the summation subject to the condition that $b_1, \ldots, b_{n(\Gamma)}$ 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 a, to prove (7.2) it suffices to show that for any fixed Γ,

$$\mathbb{E} \sum_{b_1,\ldots,b_{n(\Gamma)}}^{*} \prod_{k=1}^{a} \overline{v}_{\Gamma(i_k)} G_{\Gamma(i_k)\Gamma(j_k)} v_{\Gamma(j_k)} \cdot \prod_{k=a+1}^{2a} \overline{\overline{v}_{\Gamma(i_k)} G_{\Gamma(i_k)\Gamma(j_k)} v_{\Gamma(j_k)}} \prec \Phi^{2a}. \tag{7.7}$$

We abbreviate

$$P(b_1, ..., b_{n(\Gamma)}) := \prod_{k=1}^{a} G_{\Gamma(i_k)\Gamma(j_k)} \cdot \prod_{k=a+1}^{2a} \overline{G_{\Gamma(i_k)\Gamma(j_k)}}.$$

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 $k = 1, ..., n(\Gamma)$, we denote $\deg(b_k, P) := |\Gamma^{-1}(b_k)|$, which is the number of times that b_k appears as an index of the G entries in P. We define $h := \#\{k : \deg(b_k, P) = 1\}$, 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 $b_1, ..., b_h$. These indices are the ones that cause the main trouble: the sum of b_k, $1 \leqslant k \leqslant h$, in (7.7) contributes a factor $\sum_{b_k} |v_{b_k}|$, which can be of order $n^{1/2}$ as discussed above. However, we can obtain an extra $n^{-1/2}$ factor from the G entries with indices b_k, $1 \leqslant k \leqslant h$. Heuristically, suppose that $b_1 = i$, there is an entry G_{ij} in P, and all the other G entries are independent of the entries in the i-th row and column of H. Then using (6.5) we get

$$\mathbb{E}_i G_{ij} = G_{jj}^{(i)} \mathbb{E}_i \left[(G_{ii} - m) \left(W G^{(ij)} W^T \right)_{ij} \right], \quad i \neq j.$$

Recalling (6.41), if we replace $G_{ii} - m$ with the leading term $\mathcal{Z}_i$, then

$$\mathbb{E}_i\left[Z_i\left(WG^{(ij)}W^T\right)_{ij}\right] = -G^{(i)}_{\mu\mu}\sum_{\mu}\left(\mathbb{E}_i W^3_{i\mu}\right)(G^{(ij)}W^T)_{\mu j} \prec n^{-1/2}\Phi,$$

where we used (7.5) and the fact that $(G^{(ij)}W^T)_{\mu j}$ has the same order as the $G^{(i)}_{\mu j}$ entry by (6.6). In general, one can expand $G_{ii} - m$ using the Taylor expansion of (6.41). It is easy to see that each term in the expansion contains even number of $W_{i\star}$ entries. Together with the $W_{i\star}$ entry in $(WG^{(ij)}W^T)_{ij}$, there cannot be a perfect pairing of all of them, so we obtain an extra $n^{-1/2}$ factor due to the loss of a half free index. Finally, even without the exact independence, we know that the other $G_{\mathfrak{a}\mathfrak{b}}$ entries only have weak correlations with the entries in the i-th rows and columns of H if $\mathfrak{a}, \mathfrak{b} \neq i$. This fact will be explored using resolvent expansions in Lemma 6.3 as in Definition 7.4 below.

Claim 7.2. *We have*

$$|\mathbb{E}P| \prec n^{-h/2}\Phi^{2a}. \tag{7.8}$$

With this claim, we can complete the proof of (7.2).

Proof of (7.2). Note that by $\|\mathbf{v}\|_2 = 1$ and Cauchy-Schwarz inequality, we have $\sum_i |v_i| \leqslant \sqrt{n}$ and $\sum_i |v_i|^n \leqslant 1$ for $n \geqslant 2$. Then if (7.8) holds, we can bound the left hand side of (7.7) by

$$n^{-h/2}\Phi^{2a}\prod_{k=1}^{n(\Gamma)}\sum_{b_k}|v_{b_k}|^{\deg(b_k,P)} \leqslant n^{-h/2}\Phi^{2a}(\sqrt{n})^h \leqslant C\Phi^{2a},$$

which further concludes (7.2). □

It remains to prove Claim 7.2. We define the S variables as

$$S_{ij} := (WG^{(L)}W^T)_{ij}, \quad i,j \in \mathcal{I}_1 \cup \mathcal{I}_2, \tag{7.9}$$

where $L := \{b_1, ..., b_{n(\Gamma)}\}$. With the entrywise local law (6.3), (6.5) and (6.41), we have that

$$|S_{ij} - c_\alpha^{-1} m_{\alpha c}\delta_{ij}| \prec \Phi, \quad i,j \in \mathcal{I}_\alpha, \quad \alpha = 1,2.$$

Our first step is to keep expanding the G entries in P using the resolvent expansions in Lemma 6.3, until each monomial either consists of S variables only or has sufficiently many off-diagonal terms. To perform the resolvent expansion in a systematic way, we introduce the following notions of *string* and *string operator*.

Definition 7.3 (Strings)**.** *Let $\mathfrak{A}$ be the alphabet containing all symbols that will appear during the expansion:*

$$\mathfrak{A} = \{G^{(J)}_{kl} : J \subset L, k,l \in L\} \cup \{(G^{(J)}_{kk})^{-1} : J \subset L, k \in L\} \cup \{S_{kl} : k,l \in L\}.$$

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 $[\![\varnothing]\!] = 0$. Here we need to distinguish the difference between a string $\mathbf{s}$ and its value $[\![\mathbf{s}]\!]$. For example, "$G^{(L\setminus\{i,j\})}_{ij}$" and "$G^{(L\setminus\{i,j\})}_{ii}G^{(L\setminus\{j\})}_{jj}S_{ij}$" are different strings, but they represent the same random variable by (6.5).

We shall say $G^{(J)}_{kl}$ (resp. $(G^{(J)}_{kk})^{-1}$) is maximally expanded if $J \cup \{k,l\} = L$ (resp. $J \cup \{k\} = L$). Also the S variables are always maximally expanded. A string $\mathbf{s}$ is said to be maximally expanded if all of its symbols are maximally expanded. We shall call $G^{(J)}_{kl}$ and S_{kl} off-diagonal symbols if $k \neq l$, and all the other symbols are diagonal. Note that by the local law (6.3), we have $[\![\mathbf{a}_o]\!] \prec \Phi$ if $\mathbf{a}_o$ is an off-diagonal symbol. 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 in string $\mathbf{s}$, respectively.

Definition 7.4 (String operators)**.** *We define the following operators.*

(i) We define the operator τ_0 acting on a string $\mathbf{s}$ *in the following way. Find the first non-maximally expanded symbol in the* $\mathbf{s}$*, if $G_{ij}^{(J)}$ is found, replace it with $G_{ij}^{(J\cup\{k\})}$ for the first k in $L\backslash(J\cup\{i,j\})$; if $(G_{ii}^{(J)})^{-1}$ is found, replace it with $(G_{ii}^{(J\cup\{k\})})^{-1}$ for the first $k\in L\backslash(J\cup\{i\})$; 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 the operator τ_1 acting on a string $\mathbf{s}$ *in the following way. Find the first non-maximally expanded symbol in the* $\mathbf{s}$*, if $G_{ij}^{(J)}$ is found, replace it with $G_{ik}^{(J)}(G_{kk}^{(J)})^{-1}G_{kj}^{(J)}$ for the first k in $L\backslash(J\cup\{i,j\})$; if $(G_{ii}^{(J)})^{-1}$ is found, replace it with*

$$-G_{ik}^{(J)}G_{ki}^{(J)}(G_{ii}^{(J)})^{-1}(G_{ii}^{(J\cup\{k\})})^{-1}(G_{kk}^{(J)})^{-1}$$

for the first $k\in L\backslash(J\cup\{i\})$; if neither is found, set $\tau_1(\mathbf{s})=\varnothing$ and we say that τ_0 is null for $\mathbf{s}$.

(iii) Define the operator ρ acting on a string $\mathbf{s}$ *in the following way. Replace each maximally expanded off-diagonal $G_{ij}^{(L\backslash\{i,j\})}$ in* $\mathbf{s}$ *with $G_{ii}^{(L\backslash\{i,j\})}G_{jj}^{(L\backslash\{j\})}S_{ij}$.*

By Lemma 6.3, it is clear that for any string $\mathbf{s}$,

$$[\![\tau_0(\mathbf{s})]\!]+[\![\tau_1(\mathbf{s})]\!]=[\![\mathbf{s}]\!],\quad [\![\rho(\mathbf{s})]\!]=[\![\mathbf{s}]\!]. \tag{7.10}$$

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 $\mathbf{s}_1:=\rho(\tau_1(\mathbf{s}))$. Then by (7.10) we have

$$\sum_{|w|=m}[\![\mathbf{s}_w]\!]=[\![\mathbf{s}]\!], \tag{7.11}$$

where $w=w(1)w(2)\ldots w(m)$ with $w(i)\in\{0,1\}$ ranges over all binary sequences w with length $|w|=m$, and we used the notation

$$\mathbf{s}_w:=\rho^{w(m)}\tau_{w(m)}\ldots\rho^{w(2)}\tau_{w(2)}\rho^{w(1)}\tau_{w(1)}(\mathbf{s}),\quad \text{where } \rho^0\equiv 1.$$

Lemma 7.5 (Lemma 5.9 of [5])**.** *Consider the string $\mathbf{s}=$ "$P(b_1,...,b_{n(\Gamma)})$". Fix any $l_0\in\mathbb{N}$. There exists a constant $K(a,l_0)\in\mathbb{N}$ depending on a and l_0 only such that the following property holds. For any binary sequence w with $|w|=K(a,l_0)$ and $\mathbf{s}_w\neq\varnothing$, either $\mathcal{F}_{\text{off}}(\mathbf{s}_w)\geqslant l_0$ or $\mathbf{s}_w$ is maximally expanded.*

Let $\omega>0$ be a constant such that $\Phi\leqslant n^{-\omega/2}$. If we choose $l_0=\lceil(h\omega^{-1}+2a)\rceil$, then

$$\sum_{|w|=K(a,l_0)}[\![\mathbf{s}_w]\!]\cdot\mathbf{1}(\mathcal{F}_{\text{off}}(\mathbf{s}_w)\geqslant l_0)\prec 2^{K(a,l_0)}\Phi^{l_0}\prec n^{-h/2}\Phi^{2a}. \tag{7.12}$$

Then by Lemma 7.5, to prove Claim 7.2 it suffices to show that

$$|\mathbb{E}[\![\mathbf{s}_w]\!]|\prec n^{-h/2}\Phi^{2a} \tag{7.13}$$

for any maximally expanded string $\mathbf{s}_w$ with $|w|=K(a,l_0)$. Note that the maximally expanded string $\mathbf{s}_w$ thus obtained consists only of S symbols and diagonal G symbols $G_{ii}^{(L\backslash\{i\})}$ and $(G_{ii}^{(L\backslash\{i\})})^{-1}$. By (6.4), we can replace $(G_{kk}^{(l)})^{-1}$ with $(G_{ii}^{(L\backslash\{i\})})^{-1}=-S_{ii}-zn^{-10}$. Then as in (6.41), for $i\in\mathcal{I}_\alpha$, $\alpha=1,2$, we can expand $G_{ii}^{(L\backslash\{i\})}$ as,

$$G_{ii}^{(L\backslash\{i\})}=\frac{1}{-m_{(\alpha+2)c}+\left(m_{(\alpha+2)c}-S_{ii}-zn^{-10}\right)}$$

$$= \frac{-1}{m_{(\alpha+2)c}} \sum_{k=0}^{K(a,l_0)} \left(\frac{m_{(\alpha+2)c} - S_{ii} - zn^{-10}}{m_{(\alpha+2)c}} \right)^k + \mathrm{O}_\prec(n^{-h/2}\Phi^{2a}).$$

We apply the above expansions to the G symbols in $\mathbf{s}_w$, disregard the sufficiently small tails, and denote the resulting polynomial (in terms of the symbols S_{ij}) by P_w. Then P_w can be written as a finite sum of maximally expanded strings (or monomials) consisting of the S symbols only. Moreover, the number of such monomials depends only on a and l_0. Hence it suffices to show that for any such monomial M_w, we have

$$|\mathbb{E}[\![M_w]\!]| \prec n^{-h/2}\Phi^{2a}. \tag{7.14}$$

Recall that in the initial string P, we assume the following setting

$$\sum_{k=1}^{n} \deg(b_k, P) = 4a, \quad \text{and} \quad \deg(b_k, P) = 1, \quad \text{for } k = 1, ..., h. \tag{7.15}$$

Now in M_w, let $\deg_o(b_k, M_w)$ denotes the number of times that b_k appears as an index of the *off-diagonal* S variables in M_w. Then it is easy to verify the following relations:

$$\deg_o(b_k, M_w) \geqslant \deg(b_k, P), \quad \deg_o(b_k, M_w) = \deg(b_k, P) \mod 2, \tag{7.16}$$

where the first inequality is trivial, and the second identity follows from the simple fact that none of the above expansions changes the parity of the index b_k.

Suppose M_w takes the form

$$\begin{aligned} M_w = \prod_{j=1}^{K_w} S_{b_{k_j} b_{l_j}} &= \sum_{\mu_j, \nu_j \in \mathcal{I}_2} \prod_{j=1}^{K_w} W_{b_{k_j}\mu_j} G^{(L)}_{\mu_j \nu_j} W^T_{\nu_j b_{l_j}} \\ &= \sum_{\widetilde{\Gamma}} \sum^{*}_{\widetilde{b}_1, ..., \widetilde{b}_{n(\widetilde{\Gamma})}} \prod_{j=1}^{K_w} W_{b_{k_j}\widetilde{\Gamma}(\mu_j)} G^{(L)}_{\widetilde{\Gamma}(\mu_j)\widetilde{\Gamma}(\nu_j)} W_{b_{l_j}\widetilde{\Gamma}(\nu_j)} \end{aligned}$$

where K_w is the number of S-variables in M_ω, $\widetilde{\Gamma}$ ranges over all partitions of the set of the labels $\{\mu_1, ..., \mu_{K_w}, \nu_1, ..., \nu_{K_w}\}$, $\{\widetilde{b}_1, ..., \widetilde{b}_{n(\widetilde{\Gamma})}\}$ denotes the set of distinct equivalence classes for a particular $\widetilde{\Gamma}$, and $\sum^*$ denotes the summation subject to the condition that $\widetilde{b}_k$'s all take distinct values. Here again $\widetilde{\Gamma}(\cdot)$ is regarded as a symbolic mapping from the set of labels to the set of equivalence classes. Note that the number of partitions depends only on K_w. For a fixed partition $\widetilde{\Gamma}$, we denote

$$R(\widetilde{b}_1, ..., \widetilde{b}_{n(\widetilde{\Gamma})}; \widetilde{\Gamma}) := \prod_{j=1}^{K_\omega} W_{b_{k_j}\widetilde{\Gamma}(\mu_j)} G^{(L)}_{\widetilde{\Gamma}(\mu_j)\widetilde{\Gamma}(\nu_j)} W_{b_{l_j}\widetilde{\Gamma}(\nu_j)}.$$

Then to prove (7.14), it suffices to show that

$$\left| \mathbb{E} R(\widetilde{b}_1, ..., \widetilde{b}_{n(\widetilde{\Gamma})}; \widetilde{\Gamma}) \right| \prec n^{-n(\widetilde{\Gamma}) - h/2}\Phi^{2a}. \tag{7.17}$$

for any partition $\widetilde{\Gamma}$.

To facilitate the description of the proof, we introduce the graphical notations. We use a connected graph (V, E) to represent R, where the vertex set V consists of black vertices $b_1, \ldots, b_{n(\Gamma)}$ and white vertices

$\widetilde{b}_1, \ldots, \widetilde{b}_{n(\widetilde{\Gamma})}$, and the edge set E consists of (k, α) edges representing $W_{b_k \widetilde{b}_\alpha}$ and (α, β) edges representing $G_{\widetilde{b}_\alpha \widetilde{b}_\beta}$. We denote

$$e_{k\alpha} := \text{number of } (k,\alpha) \text{ edges in } R, \quad d_\alpha := \text{number of } (\alpha,\alpha) \text{ edges in } R,$$

and

$$e^{(o)}_{k\alpha} := \text{number of } (k,\alpha) \text{ edges that are from off-diagonal } S \text{ in } M_w.$$

Due to the mean zero condition (6.1), to attain a nonzero expectation we must have

$$e_{k\alpha} = 0 \quad \text{or} \quad e_{k\alpha} \geqslant 2 \quad \text{for all } k, \alpha. \tag{7.18}$$

We also have that, by definition,

$$\sum_\alpha e^{(o)}_{k\alpha} = \deg_o(b_k, M_w) \tag{7.19}$$

By (7.15), (7.16) and (7.18), 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) edges such that $e_{k\alpha} \geqslant 2$ and $(k, \alpha) \notin H$. Denote

$$s_\alpha := \sum_{k=1}^{n(\Gamma)} e_{k\alpha}, \quad h_{k\alpha} := \mathbf{1}_{(k,\alpha)\in H}, \quad h_\alpha := \sum_{k=1}^{n(\Gamma)} h_{k\alpha}, \quad f_\alpha := \sum_{k=1}^{n(\Gamma)} \mathbf{1}_{(k,\alpha)\in F},$$

for all $k = 1, ..., n(\Gamma)$ and $\alpha = 1, ..., n(\widetilde{\Gamma})$. From the above definitions, it is easy to see that $s_\alpha \geqslant 2$ and $h_\alpha + f_\alpha > 0$ (since the classes $\widetilde{b}_\alpha$ are nontrivial), $s_\alpha \geqslant 2d_\alpha$ (since one (α, α) edge corresponds to two (k, α) edges), and

$$\sum_\alpha h_{k\alpha} = \mathbf{1}(1 \leqslant k \leqslant h), \quad \sum_\alpha h_\alpha = h. \tag{7.20}$$

Since there are totally $\frac{1}{2} \sum_\alpha s_\alpha - d_\alpha$ off-diagonal G edges in R, by (6.3) and (7.5) we have

$$\begin{aligned}
|\mathbb{E}R| &\prec \prod_{\alpha=1}^{n(\widetilde{\Gamma})} \Big(\Phi^{\frac{1}{2}s_\alpha - d_\alpha} \prod_{k=1}^{n(\Gamma)} \mathbb{E}|W_{b_k \widetilde{b}_\alpha}|^{e_{k\alpha}} \Big) \\
&\prec \prod_{\alpha=1}^{n(\widetilde{\Gamma})} \Phi^{\frac{1}{2}s_\alpha - d_\alpha} \Big(\prod_{(k,\alpha)\in H} \phi_n^{e_{k\alpha}-3} n^{-3/2} \Big) \Big(\prod_{(k,\alpha)\in F} \phi_n^{e_{k\alpha}-2} n^{-1} \Big) =: \prod_{\alpha=1}^{m} R_\alpha.
\end{aligned}$$

Now we consider the following four cases for R_α. The arguments essentially are the same as the ones in [30, Appendix A.4], and we repeat them for reader's convenience.

Case 1: $d_\alpha = 0$. In this case we have

$$R_\alpha \prec \Phi^{s_\alpha/2} n^{-(h_\alpha + f_\alpha) - h_\alpha/2} \prec \Phi^{s_\alpha/2} n^{-1-h_\alpha/2} \prec \Phi^{\sum_{k=1}^h h_{k\alpha}/2 + \sum_{k=h+1}^n e^{(o)}_{k\alpha}/2} n^{-1-h_\alpha/2}$$

where in the second step we used $h_\alpha + f_\alpha > 0$, and in the third step we used

$$s_\alpha \geqslant \sum_k e^{(o)}_{k\alpha} \geqslant \sum_{k=1}^h h_{k\alpha} + \sum_{k=h+1}^n e^{(o)}_{k\alpha},$$

where we used that $e^{(o)}_{k\alpha} \geqslant h_{k\alpha}$ for $1 \leqslant k \leqslant h$ (recall that if $(k, \alpha) \in H$, then $e_{k\alpha_k}$ is odd and hence one of the edges must come from the off-diagonal S).

Case 2: $d_\alpha \neq 0$, $h_\alpha = 1$ 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

$$R_\alpha \prec \Phi^{\frac{1}{2}s_\alpha - d_\alpha} n^{-(h_\alpha+f_\alpha)-h_\alpha/2} \prec \Phi^{1/2} n^{-1-h_\alpha/2} = \Phi^{\sum_{k=1}^h h_{k\alpha}/2 + \sum_{k=h+1}^n e_{k\alpha}^{(o)}/2} n^{-1-h_\alpha/2},$$

where in the last step we used $1 = \sum_{k=1}^h h_{k\alpha} + \sum_{k=h+1}^n e_{k\alpha}^{(o)}$, since all the summands except one $h_{k\alpha}$ are 0.

Case 3: $d_\alpha \neq 0$, $h_\alpha = 0$ 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, which implies $s_\alpha - 2d_\alpha = e_{k\alpha}^{(o)}$. Then we can bound

$$R_\alpha \prec \Phi^{\frac{1}{2}s_\alpha - d_\alpha} n^{-(h_\alpha+f_\alpha)-h_\alpha/2} = \Phi^{\sum_k e_{k\alpha}^{(o)}/2} n^{-1-h_\alpha/2} \prec \Phi^{\sum_{k=1}^h h_{k\alpha}/2 + \sum_{k=h+1}^n e_{k\alpha}^{(o)}/2} n^{-1-h_\alpha/2}$$

where we used $e_{k\alpha}^{(o)} \geqslant h_{k\alpha}$ for $1 \leqslant k \leqslant h$ in the last step.

Case 4: $d_\alpha \neq 0$ and $h_\alpha + f_\alpha \geqslant 2$. Then using $s_\alpha \geqslant 2d_\alpha$, $\phi_n \prec \Phi$ and $n^{-1/2} \prec \Phi$, we get that

$$\begin{aligned} R_\alpha &\prec \prod_{(k,\alpha)\in H} \Phi^{e_{k\alpha}-3} n^{-3/2} \prod_{(k,\alpha)\in F} \Phi^{e_{k\alpha}-2} n^{-1} \prec \prod_{(k,\alpha)\in H} \Phi^{e_{k\alpha}-2} n^{-1} \prod_{(k,\alpha)\in F} \Phi^{e_{k\alpha}-1} n^{-1/2} \\ &\leqslant \Phi^{\sum_{k=1}^h h_{k\alpha}/2 + \sum_{k=h+1}^n e_{k\alpha}^{(o)}/2} n^{-1-h_\alpha/2} \end{aligned}$$

where in the last step we used $e_{k\alpha} \geqslant h_{k\alpha} + 2$ for $(k, \alpha) \in H$ and $e_{k\alpha} \geqslant 2$ for $(k, \alpha) \in F$.

Combining the above four cases, we obtain that

$$|\mathbb{E}R| = \prod_{\alpha=1}^{n(\tilde{\Gamma})} R_\alpha \prec n^{-n(\tilde{\Gamma})} n^{-\frac{1}{2}\sum_\alpha h_\alpha} \Phi^{\sum_\alpha \left(\sum_{k=1}^h h_{k\alpha}/2 + \sum_{k=h+1}^n e_{k\alpha}^{(o)}/2\right)}.$$

Since $\sum_\alpha h_\alpha = h$, to prove (7.17) it remains to show that

$$\sum_\alpha \left(\sum_{k=1}^h h_{k\alpha} + \sum_{k=h+1}^{n(\Gamma)} e_{k\alpha}^{(o)} \right) \geqslant 4a. \tag{7.21}$$

For $k = 1, ..., h$, using (7.20) and (7.15) we get that

$$\sum_{\alpha=1}^m h_{k\alpha} = 1 = \deg(b_k, P).$$

For $k = h + 1, ..., n$, using (7.19) and (7.16) we get that

$$\sum_{\alpha=1}^m e_{k\alpha}^{(o)} = \deg_o(b_k, Q) \geqslant \deg(b_k, P).$$

With (7.15), we then conclude (7.21), which concludes the proof of Claim 7.2.

7.2 Proof of (7.3)

The proof of this sections adopts the arguments in [31, Section 5]. We expand the left-hand side in (7.3) as

$$\begin{aligned} \left| \sum_{\mu\neq\nu} v_{[\mu]}^* G_{[\mu\nu]} v_{[\nu]} \right|^{2a} = \sum_\Gamma \sum_{b_1,\dots,b_{n(\Gamma)}}^{*} \prod_{k=1}^a v_{[\Gamma(\mu_k)]}^* G_{[\Gamma(\mu_k)\Gamma(\nu_k)]} v_{[\Gamma(\nu_k)]} \\ \times \prod_{k=a+1}^{2a} \overline{v_{[\Gamma(\mu_k)]}^* G_{[\Gamma(\mu_k)\Gamma(\nu_k)]} v_{[\Gamma(\nu_k)]}}, \end{aligned} \tag{7.22}$$

where we again define partition of indices

$$\Gamma : \{\mu_1, \cdots, \mu_{2a}, \nu_1, \cdots, \nu_{2a}\} \to L(\Gamma), \quad L(\Gamma) = (b_1, \cdots, b_{n(\Gamma)}),$$

$\Gamma^{-1}(b_k)$ are equivalence classes of the partition, $n(\Gamma)$ is the number of equivalence classes, b_k are indices taking values in $\mathcal{I}_3$, and $\sum^*$ denotes the summation subject to the condition that $b_1, \ldots, b_{n(\Gamma)}$ all take distinct values and $\Gamma(\mu_k) \neq \Gamma(\nu_k)$ for all k. Since the number of such partitions Γ is finite and depends only on a, to prove (7.3) it suffices to show that for any fixed Γ,

$$\sum_{b_1,...,b_{n(\Gamma)}}^{*} \mathbb{E}\Delta(b_1, ..., b_{n(\Gamma)}) \prec \Phi^{2a}, \tag{7.23}$$

where we abbreviated

$$\Delta(\Gamma) := \prod_{k=1}^{a} v^*_{[\Gamma(i_k)]} G_{[\Gamma(i_k)\Gamma(j_k)]} v_{[\Gamma(j_k)]} \cdot \prod_{k=a+1}^{2a} \overline{v^*_{[\Gamma(i_k)]} G_{[\Gamma(i_k)\Gamma(j_k)]} v_{[\Gamma(j_k)]}}. \tag{7.24}$$

For simplicity, we again 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 any $b_k \in L$, we can define a corresponding $\mathcal{I}_4$-valued variable $\bar{b}_k$ in the obvious way, and we denote

$$[L] := \{b_1, ..., b_n, \bar{b}_1, ..., \bar{b}_n\}. \tag{7.25}$$

We shall abbreviate $G^{([J])} \equiv G^{[J]}$ for any index set $J \subset \mathcal{I}_3$. Then we define S groups as

$$S_{[\mu\nu]} := \begin{pmatrix} (X^T G^{[L]} X)_{\mu\nu} & (X^T G^{[L]} Y)_{\mu\bar{\nu}} \\ (Y^T G^{[L]} X)_{\bar{\mu}\nu} & (Y^T G^{[L]} Y)_{\bar{\mu}\bar{\nu}} \end{pmatrix}.$$

We can define strings as in Definition 7.3 with bigger alphabet which includes the new G and S groups. Given any index set $J \subset L$, we shall say $G^{[J]}_{[\mu\nu]}$ (resp. $(G^{[J]}_{[\mu\mu]})^{-1}$) is maximally expanded if $J \cup \{\mu, \nu\} = L$ (resp. $J \cup \{\mu\} = L$). Also the S groups are always maximally expanded. We shall call $G^{[J]}_{[\mu\nu]}$ and $S_{[\mu\nu]}$ off-diagonal symbols if $\mu \neq \nu$, and all the other symbols are diagonal. Note that by the local law (6.3) and (6.22), we have that for off-diagonal group $\mathbf{a}_o$, $\pi^{-1} [\![\mathbf{a}_o]\!] \pi^{-1} \prec \Phi$. We use $\mathcal{F}_{\text{off}}(\mathbf{s})$ to denote the number of off-diagonal symbols in the string $\mathbf{s}$. We can define string operators as in Definition 7.4 using the resolvent expansions in Lemma 6.4, and we can perform the resolvent expansions systematically using string operators as in (7.11). We omit the detailed definitions here. Instead, we directly give the following result, which has been proved in [31].

Lemma 7.6 (Lemma 5.9 of [31])**.** *Consider the string* $\mathbf{s} =$ *"*$\Delta(\Gamma)$*". Fix any* $l_0 \in \mathbb{N}$*. There exists a constant* $K(a, l_0) \in \mathbb{N}$ *depending on* a *and* l_0 *only such that the following property holds. For any binary sequence* w *with* $|w| = K(a, l_0)$ *and* $\mathbf{s}_w \neq \varnothing$*, either* $\mathcal{F}_{\text{off}}(\mathbf{s}_w) \geq l_0$ *or* $\mathbf{s}_w$ *is maximally expanded.*

As in (7.12), if we take $l_0 = \lceil (h\omega^{-1} + 2a) \rceil$, then

$$\sum_{|w|=K(a,l_0)} [\![\mathbf{s}_w]\!] \cdot \mathbf{1}(\mathcal{F}_{\text{off}}(\mathbf{s}_w) \geq l_0) \prec n^{h/2} \cdot n^{-h/2} \Phi^{2a} = \Phi^{2a},$$

where the $n^{h/2}$ comes from the vector $\mathbf{v}$, since we have included $v_{\Gamma(\cdot)}$ into $\Delta(\Gamma)$. It remains to handle the maximally expanded strings. With (6.10) and (6.42), we can write

$$(G^{[L\setminus\{\mu\}]}_{[\mu\mu]})^{-1} = \pi^{-1}(z) - \left[S_{[\mu\mu]} - \begin{pmatrix} m_{1c} & 0 \\ 0 & m_{2c} \end{pmatrix} \right],$$

where by local law (6.3) and (6.27), we have

$$\left\| S_{[\mu\mu]} - \begin{pmatrix} m_{1c} & 0 \\ 0 & m_{2c} \end{pmatrix} \right\| \prec \Phi.$$

Then we can Taylor expand $G^{[L\setminus\{\mu\}]}_{[\mu\mu]}$ in terms of $S_{[\mu\mu]}$, and replace the diagonal maximally expanded G groups with diagonal S groups. Finally, as in (7.14), one can see that in order to prove (7.23) it suffices to show that for any such monomial $M_w(\Delta(\Gamma))$ consisting of S variables only, we have

$$\sum_{b_1,\cdots,b_n(\Gamma)}^{*} |\mathbb{E}[\![M_w(\Delta(\Gamma))]\!]| \prec \Phi^{2a}. \tag{7.26}$$

Now we decompose $S_{[\mu\nu]}$ in M_w as

$$S_{[\mu\nu]} = S_{\mu\nu}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + S_{\mu\overline{\nu}}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + S_{\overline{\mu}\nu}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + S_{\overline{\mu}\overline{\nu}}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{7.27}$$

where we define the following symbols:

$$\begin{aligned} S_{\mu\nu} &:= \left(X^T G^{[L]} X\right)_{\mu\nu}, \quad S_{\mu\overline{\nu}} := \left(X^T G^{[L]} Y\right)_{\mu\overline{\nu}}, \\ S_{\overline{\mu}\nu} &:= \left(Y^T G^{[L]} X\right)_{\overline{\mu}\nu}, \quad S_{\overline{\mu}\overline{\nu}} := \left(Y^T G^{[L]} Y\right)_{\overline{\mu}\overline{\nu}}. \end{aligned} \tag{7.28}$$

We expand the $S_{[ij]}$'s in $M_w(\Delta(\Gamma))$ using (7.27), and write $M_w(\Delta(\Gamma))$ as a sum of monomials

$$M_w(\Delta(\Gamma)) = \sum_{\gamma} P_w(\Gamma,\gamma)\mathcal{M}_\gamma, \tag{7.29}$$

where γ is an index to label these monomials, $P_w(\Gamma,\gamma)$ denotes a scalar monomial in terms of $S_{\mu\nu}$ variables only, and $\mathcal{M}_\gamma$ contains the factor depending only on the entries of $\mathbf{v}$. We make a simple observation that

$$\mathcal{F}_{\text{off}}\left(P_w(\Gamma)\right) = \mathcal{F}_{\text{off}}\left(M_w(\Delta(\Gamma))\right) \geqslant \mathcal{F}_{\text{off}}\left(\Delta(\Gamma)\right) = 2a. \tag{7.30}$$

As in the proof of (7.2), we define $\deg(b_k, \Delta(\Gamma)) := |\Gamma^{-1}(b_k)|$ and $h := \#\{k : \deg(b_k, P) = 1\}$. We have already seen that the indices with degree 1 cause the main trouble. Since the number of summands in (7.29) is of order $\mathrm{O}(1)$, to prove (7.26) it suffices to prove that for any fixed monomial $P_w(\Gamma)$ in (7.29),

$$|\mathbb{E}[\![P_w(\Gamma)]\!]| \prec n^{-h/2}\Phi^{2a}. \tag{7.31}$$

To prove (7.31), we need to keep track of the "single" indices in $[L]$ during the expansion. In P_w, let $\deg_o(b_k, P_w)$ denotes the number of times that b_k or $\overline{b}_k$ appears as an index of the *off-diagonal* S variables in P_w. Again we have the following simple relations:

$$\deg_o(b_k, P_w) \geqslant \deg(b_k, \Delta(\Gamma)), \quad \deg_o(b_k, P_w) = \deg(b_k, \Delta(\Gamma)) \mod 2. \tag{7.32}$$

We expand the S variables in P_w using (7.28), and call the resulting string Q_w. We now introduce graphs to conclude the proof of (7.31). We use a connected graph to represent the string Q_w, call it by $\mathfrak{G}_Q$. The indices in $[L]$ are represented by black nodes in $\mathfrak{G}_Q$, and the i, j summation indices in the S variables are represented by white nodes. The X or Y variables are represented by wavy edges, and G are represented by

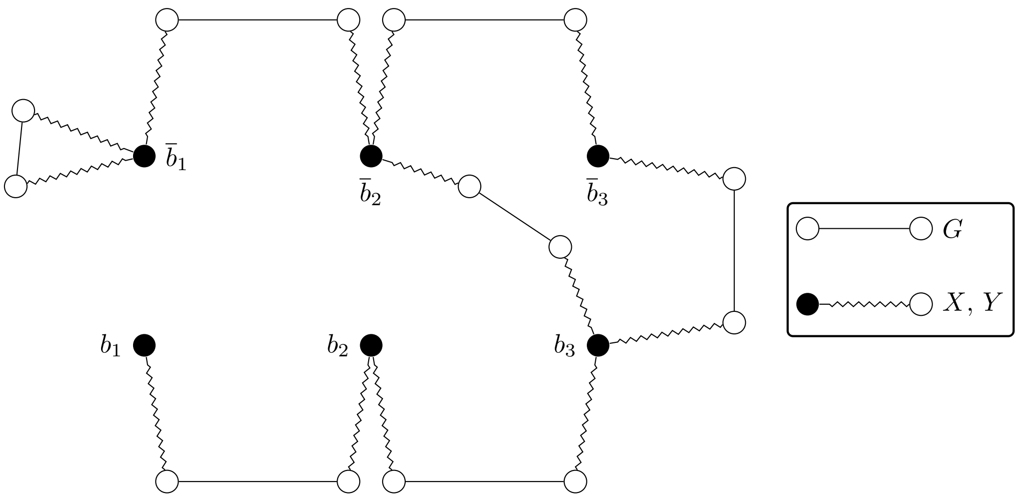

Figure 1: An example of the graph $\mathfrak{G}_Q$.

solid lines. In Figure 1, we give an example of the graph $\mathfrak{G}_Q$. Note that in graph $\mathfrak{G}_Q$, the G edges and X, Y edges are mutually independent, since the G variables are maximally expanded.

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 $\widetilde{\Gamma}$ be a fixed partition and denote its blocks by $\widetilde{w}_1, ..., \widetilde{w}_{n(\widetilde{\Gamma})}$. If two white nodes of some **off-diagonal** S variable happen to lie in the same block, then we merge the two nodes into one white node and call the resulting graph $\mathfrak{G}_{Q1}$. Note that we do not merge the white nodes for diagonal S variables. Let $n_d^{(o)}$ be the number of diagonal G edges in the *off-diagonal* S variables. We trivially have

$$\# \text{ of white nodes} = -n_d^{(o)} + \sum_{k=1}^{n} \left[\deg(b_k) + \deg(\bar{b}_k) \right], \tag{7.33}$$

where the degrees of the nodes are defined in the usual graphical sense.

We define the subset of single indices

$$\mathcal{V} := \{ b_k \in L | \ \deg(b_k, \Delta(\Omega)) = 1 \}, \quad |\mathcal{V}| = h. \tag{7.34}$$

By (7.32), there are at least h black nodes with odd deg in $[\mathcal{V}]$. WLOG, we may assume these nodes are $b_1, ..., b_h$. To have nonzero expectation, for each $k = 1, ..., h$, there must exist a block $\widetilde{w}_{i_k}$ connecting to b_k which contains at least 3 white nodes. Then we denote by $A(b_k) \subseteq \widetilde{w}_{i_k}$ the set of the adjacent white nodes to b_k in $\widetilde{w}_{i_k}$ (the block $\widetilde{w}_{i_k}$ may also contain white nodes that do not connect to b_k, hence in general $A(b_k)$ may not be equal to $\widetilde{w}_{i_k}$). We call a white node that is not connected to loops of solid G edges a **normal white node**, i.e. normal white nodes are white nodes that have not been merged before. The other white nodes are called merged white nodes. Then we define

$$\mathcal{V}_0 := \{ b_k | \ A(b_k) \text{ has no normal white nodes}, 1 \leqslant k \leqslant h \},$$

and

$$\mathcal{V}_1 := \{ b_k | \ A(b_k) \text{ has at least one normal white node}, 1 \leqslant k \leqslant h \}.$$

The following lemma gives the key estimates we need.

Lemma 7.7. *For any partition of white nodes* $\widetilde{\Gamma}$,

$$2n(\widetilde{\Gamma}) \leqslant -|\mathcal{V}_1| - |\mathcal{V}_0|/2 - n_d^{(o)} + \sum_{k=1}^{n(\widetilde{\Gamma})} \left[\deg(b_k) + \deg(\bar{b}_k)\right], \tag{7.35}$$

and

$$n_o \geqslant 2a + |\mathcal{V}_0|, \tag{7.36}$$

where n_o *is the total number of off-diagonal* S *variables in* Q_w*.*

Proof. WLOG, let $\widetilde{w}_1, ..., \widetilde{w}_d$ be the distinct blocks among the blocks $\widetilde{w}_{i_k}$, $k = 1, \cdots, h$. A merged white node is connected to two black nodes and a normal white node is connected to one black node. Hence a merged 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 $\widetilde{w}_i$ contains exactly one $A(b_k)$, then

$$|\widetilde{w}_i| \geqslant 3 \geqslant 2 + \mathbf{1}_{\mathcal{V}_1}(b_k) + \frac{\mathbf{1}_{\mathcal{V}_0}(b_k)}{2}.$$

If $\widetilde{w}_i$ contains at least two $A(b_k)$, then

$$\begin{aligned} |\widetilde{w}_i| &\geqslant \sum_{b_k : A(b_k) \subseteq \widetilde{w}_i} \left(2 \cdot \mathbf{1}_{\mathcal{V}_1}(b_k) + \frac{3}{2} \cdot \mathbf{1}_{\mathcal{V}_0}(b_k)\right) \\ &\geqslant 2 + \sum_{b_k : A(b_k) \subseteq W_i} \left(\mathbf{1}_{\mathcal{V}_1}(b_k) + \frac{\mathbf{1}_{\mathcal{V}_0}(b_k)}{2}\right). \end{aligned} \tag{7.37}$$

Here the first inequality holds due to the following reasoning. For each black node b_k with $A(b_k) \subseteq \widetilde{w}_i$, 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 merged white node (since it is shared by two different black nodes). If $b_k \in \mathcal{V}_0$, there are at least three merged white nodes in $A(b_k)$ with total weight $\geqslant 3/2$. If $b_k \in \mathcal{V}_1$, there are at least one normal white node and two other white nodes in $A(b_k)$ with total weight $\geqslant 2$.

Then summing (7.37) over i, we get that

$$\sum_{i=1}^{d} |\widetilde{w}_i| \geqslant 2d + |\mathcal{V}_1| + \frac{|\mathcal{V}_0|}{2}.$$

For the other $n(\widetilde{\Gamma}) - d$ blocks, each of them contains at least two white nodes, so we get that

$$2n(\widetilde{\Gamma}) + |\mathcal{V}_1| + \frac{|\mathcal{V}_0|}{2} \leqslant \sum_{i=1}^{d} |\widetilde{w}_i| + 2(n(\widetilde{\Gamma}) - d) \leqslant -n_d^{(o)} + \sum_{k=1}^{n(\widetilde{\Gamma})} \left[\deg(b_k) + \deg(\bar{b}_k)\right],$$

where we used (7.33) in the last step. Rearranging terms gives (7.35).

For $b_k \in \mathcal{V}_0$, $A(b_k)$ contains at least three white nodes from off-diagonal R-groups. Hence we have $\deg(b_k, Q_w) \geqslant 3$, compared with $\deg(b_k, \Delta(\Gamma)) = 1$. This means that we have applied (6.14) or (6.15) with respect to $\mu = b_k$ at least once and picked the second terms at some step of the expansions (which corresponds to the τ_1 operation in Definition 7.4). Each such operation increases the off-diagonal S variables at least by 1, which gives (7.36). ☐

Now we prove (7.31). By (7.5) and the discussion below (7.34) for the single indices $b_1, \cdots, b_h$, we get

$$\begin{aligned} |\mathbb{E}[\![P_w]\!]| = |\mathbb{E}[\![Q_w]\!]| &\prec \sum_{\widetilde{\Gamma}} \sum_{\widetilde{w}_1,\ldots,\widetilde{w}_{n(\widetilde{\Gamma})}}^{*} \Phi^{n_o - n_d^{(o)}} n^{-n(\widetilde{\Gamma})-h/2} \phi_n^{\sum_{k=1}^{n(\widetilde{\Gamma})}\left[\deg(b_k)+\deg\left(\bar{b}_k\right)\right]-2n(\widetilde{\Gamma})-h} \\ &\prec n^{-h/2}\Phi^{\sum_{k=1}^{n(\widetilde{\Gamma})}\left[\deg(b_k)+\deg\left(\bar{b}_k\right)\right]-2n(\widetilde{\Gamma})-h+n_o-n_d^{(o)}} \\ &\prec n^{-h/2}\Phi^{-|\mathcal{V}_0|/2+n_o} \prec n^{-h/2}\Phi^{2a}, \end{aligned}$$

where in the third step we used (7.35) and $|\mathcal{V}_0| + |\mathcal{V}_1| = h$, and last step (7.36). Thus we have proved (7.31), which concludes the proof of (7.3).

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

[4] Z. Bao, J. Hu, G. Pan, and W. Zhou. Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case. *Ann. Statist.*, 47(1):612–640, 2019.

[5] 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.

[6] 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.

[7] P. Bourgade, H.-T. Yau, and J. Yin. Local circular law for random matrices. *Probab. Theory Relat. Fields*, 159:545–595, 2014.

[8] X. Ding and F. Yang. Spiked separable covariance matrices and principal components. *arXiv:1905.13060*.

[9] 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.

[10] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14:1837–1926, 2013.

[11] 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.

[12] 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.

[13] 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.

[14] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics*, 229:1435 – 1515, 2012.

[15] P. Forrester. The spectrum edge of random matrix ensembles. *Nucl. Phys. B*, 402(3):709 – 728, 1993.

[16] Y. Fujikoshi. High-dimensional asymptotic distributions of characteristic roots in multivariate linear models and canonical correlation analysis. *Hiroshima Math. J.*, 47(3):249–271, 2017.

[17] C. Gao, Z. Ma, Z. Ren, and H. H. Zhou. Minimax estimation in sparse canonical correlation analysis. *Ann. Statist.*, 43(5):2168–2197, 2015.

[18] C. Gao, Z. Ma, and H. H. Zhou. Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.*, 45(5):2074–2101, 2017.

[19] X. Han, G. Pan, and Q. Yang. A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications. *Bernoulli*, 24(4B):3447–3468, 2018.

[20] X. Han, G. Pan, and B. Zhang. The Tracy-Widom law for the largest eigenvalue of F type matrices. *Ann. Statist.*, 44(4):1564–1592, 2016.

[21] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3-4):321–377, 1936.

[22] I. M. Johnstone. Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.*, 36(6):2638–2716, 2008.

[23] I. M. Johnstone and A. Onatski. Testing in high-dimensional spiked models. *Ann. Statist.*, 48(3):1231–1254, 2020.

[24] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, pages 1–96, 2016.

[25] J. O. Lee and J. Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.*, 163:117–173, 2014.

[26] Z. Ma and F. Yang. CLT for the canonical correlation coefficients of high-dimensional random vectors with finite rank correlations. *In preparation*, 2020.

[27] Z. Ma and F. Yang. Universality of the canonical correlation coefficients of high-dimensional random vectors with finite rank correlations. *In preparation*, 2020.

[28] N. S. Pillai and J. Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24:935–1001, 2014.

[29] K. W. Wachter. The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.*, 8(5):937–957, 1980.

[30] H. Xi, F. Yang, and J. Yin. Convergence of eigenvector empirical spectral distribution of sample covariance matrices. *arXiv:1705.03954*.

[31] 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.

[32] F. Yang. Edge universality of separable covariance matrices. *Electron. J. Probab.*, 24:57 pp., 2019.

[33] Y. Yang and G. Pan. The convergence of the empirical distribution of canonical correlation coefficients. *Electron. J. Probab.*, 17:13 pp., 2012.

[34] Y. Yang and G. Pan. Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.*, 43(2):467–500, 04 2015.