Sample canonical correlation coefficients of high-dimensional random vectors with finite rank correlations

Zongming Ma *[1] and Fan Yang †[1]

[1]Department of Statistics, University of Pennsylvania

February 28, 2021

Abstract

Consider two random vectors $\widetilde{\mathbf{x}} \in \mathbb{R}^p$ and $\widetilde{\mathbf{y}} \in \mathbb{R}^q$ of the forms $\widetilde{\mathbf{x}} = A\mathbf{z} + \mathbf{C}_1^{1/2}\mathbf{x}$ and $\widetilde{\mathbf{y}} = B\mathbf{z} + \mathbf{C}_2^{1/2}\mathbf{y}$, where $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$ and $\mathbf{z} \in \mathbb{R}^r$ are independent random vectors with i.i.d. entries of zero mean and unit variance, $\mathbf{C}_1$ and $\mathbf{C}_2$ are $p \times p$ and $q \times q$ deterministic population covariance matrices, and A and B are $p \times r$ and $q \times r$ deterministic factor loading matrices. With n independent observations of $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}})$, we study the sample canonical correlations between $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$. We consider the high-dimensional setting with finite rank correlations, that is, $p/n \to c_1$ and $q/n \to c_2$ as $n \to \infty$ for some constants $c_1 \in (0,1)$ and $c_2 \in (0, 1-c_1)$, and r is a fixed integer. Let $t_1 \geqslant t_2 \geqslant \cdots \geqslant t_r \geqslant 0$ be the squares of the nontrivial population canonical correlation coefficients between $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$, and let $\widetilde{\lambda}_1 \geqslant \widetilde{\lambda}_2 \geqslant \cdots \geqslant \widetilde{\lambda}_{p\wedge q} \geqslant 0$ be the squares of the sample canonical correlation coefficients. If the entries of $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are i.i.d. Gaussian, then the following dichotomy has been shown in [7] for a fixed threshold $t_c \in (0,1)$: for any $1 \leqslant i \leqslant r$, if $t_i < t_c$, then $\widetilde{\lambda}_i$ converges to the right-edge λ_+ of the limiting eigenvalue spectrum of the sample canonical correlation matrix, and moreover, $n^{2/3}(\widetilde{\lambda}_i - \lambda_+)$ converges weakly to the Tracy-Widom law; if $t_i > t_c$, then $\widetilde{\lambda}_i$ converges to a deterministic limit $\theta_i \in (\lambda_+, 1)$ that is determined by c_1, c_2 and t_i. In this paper, we prove that these results hold universally under the sharp fourth moment conditions on the entries of $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. Moreover, we prove the results in full generality, in the sense that they also hold for near-degenerate t_i's and for t_i's that are close to the threshold t_c. Finally, we also provide almost sharp convergence rates for the sample canonical correlation coefficients under a general a-th moment assumption.

1 Introduction

Since the seminal work by Hotelling [32], the canonical correlation analysis (CCA) has been one of the most classical methods to study the correlations between two random vectors. Given two random vectors $\widetilde{\mathbf{x}} \in \mathbb{R}^p$ and $\widetilde{\mathbf{y}} \in \mathbb{R}^q$, we denote the population covariance and cross-covariance matrices by

$$\widetilde{\Sigma}_{xx} := \mathrm{Cov}(\widetilde{\mathbf{x}}, \widetilde{\mathbf{x}}), \quad \widetilde{\Sigma}_{yy} := \mathrm{Cov}(\widetilde{\mathbf{y}}, \widetilde{\mathbf{y}}), \quad \widetilde{\Sigma}_{xy} = \widetilde{\Sigma}_{yx}^\top := \mathrm{Cov}(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}).$$

It is well-known that the i-th canonical correlation coefficient (CCC) between $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$, denoted by ρ_i, is the square root of the i-th largest eigenvalue t_i of the population canonical correlation (PCC) matrix

$$\widetilde{\mathbf{\Sigma}} := \widetilde{\Sigma}_{xx}^{-1/2}\widetilde{\Sigma}_{xy}\widetilde{\Sigma}_{yy}^{-1}\widetilde{\Sigma}_{yx}\widetilde{\Sigma}_{xx}^{-1/2}.$$

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

†E-mail: fyang75@wharton.upenn.edu. F. Yang was supported by the Wharton Dean's Fund for Postdoctoral Research.

Suppose we observe n independent samples of $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}})$. Then we can study the population CCC's through their sample counterparts. More precisely, we form data matrices $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$ as

$$\widetilde{\mathcal{X}} := n^{-1/2}\left(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \cdots, \widetilde{\mathbf{x}}_n\right), \quad \widetilde{\mathcal{Y}} := n^{-1/2}\left(\widetilde{\mathbf{y}}_1, \widetilde{\mathbf{y}}_2, \cdots, \widetilde{\mathbf{y}}_n\right), \tag{1.1}$$

where $n^{-1/2}$ is a convenient scaling, so that the sample covariance and cross-covariance matrices can be written concisely as

$$\widetilde{S}_{xx} := \frac{1}{n}\sum_{i=1}^{n} \widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}_i^\top = \widetilde{\mathcal{X}}\widetilde{\mathcal{X}}^\top, \quad \widetilde{S}_{yy} := \frac{1}{n}\sum_{i=1}^{n} \widetilde{\mathbf{y}}_i \widetilde{\mathbf{y}}_i^\top = \widetilde{\mathcal{Y}}\widetilde{\mathcal{Y}}^\top, \quad \widetilde{S}_{xy} = \widetilde{S}_{yx}^\top := \frac{1}{n}\sum_{i=1}^{n} \widetilde{\mathbf{x}}_i \widetilde{\mathbf{y}}_i^\top = \widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}^\top.$$

Then the squares of the sample CCC's, $\widetilde{\lambda}_1 \geqslant \widetilde{\lambda}_2 \geqslant \cdots \geqslant \widetilde{\lambda}_{p\wedge q} \geqslant 0$, are defined as the eigenvalues of the *sample canonical correlation* (SCC) matrix

$$\mathcal{C}_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}} := \widetilde{S}_{xx}^{-1/2}\widetilde{S}_{xy}\widetilde{S}_{yy}^{-1}\widetilde{S}_{yx}\widetilde{S}_{xx}^{-1/2}.$$

If $n \to \infty$ while p, q and r are fixed, the SCC matrix converges to the PCC matrix almost surely by law of large number, and hence the sample CCC can be used as a consistent estimator of the population CCC. However, many modern applications, such as statistical learning, wireless communications, medical imaging, financial economics and population genetics, are seeing a rapidly increasing demand in analyzing high-dimensional data, where p and q are comparable to n when n is large. In the high-dimensional setting, the behavior of the SCC matrix can deviate greatly from the PCC matrix due to the so-called "curse of dimensionality".

There have been several works on the theoretical analysis of high-dimensional CCA. We mention some of them that are most related to this paper.

First, we consider the null case where $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ are independent random vectors. When $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ are independent Gaussian vectors, the eigenvalues of the SCC matrix have the same joint distribution as those of a double Wishart matrix [34]. In particular, the joint distribution of the eigenvalues of double Wishart matrices has been studied in the context of the Jacobi ensemble and F-type matrices [31, 34], which implies that the largest few eigenvalues of the SCC matrix satisfy the Tracy-Widom law asymptotically. For general distributed random vectors $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$, the Tracy-Widom fluctuation of the largest eigenvalues of the SCC matrix is proved in [30] under the assumption that the entries of $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ have finite moments up to any order. The moment assumption is later relaxed to the finite fourth moment assumption in [50]. In the Gaussian case, it is shown in [46] that, almost surely, the empirical spectral distribution (ESD) of the SCC matrix converges weakly to a deterministic probability distribution (cf. (2.12)). In the general non-Gaussian case, both the convergence and the linear spectral statistics of the ESD of the SCC matrix have been proved in [52, 53].

Then we consider the case where $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ have finite rank correlations. If $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{y}}$ are random Gaussian vectors, then the asymptotic distributions of the sample canonical correlation coefficients have been derived when one of p and q is fixed as $n \to \infty$ [27]. If p and q are both proportional to n, the asymptotic distributions of the sample CCC's have been established under the Gaussian assumption in [7]. Under certain sparsity assumptions, the theory of high-dimensional sparse CCA and it applications have been discussed in [28, 29]. In [40], the authors derived asymptotic null and non-null distributions of several test statistics for tests of redundancy in high-dimensional CCA. In [35], the authors studied the asymptotic behaviors of the likelihood ratio processes of CCA under the null hypothesis of no spikes and the alternative hypothesis of a single spike.

In this paper, we consider the following signal-plus-noise model for $\widetilde{\mathbf{x}} \in \mathbb{R}^p$ and $\widetilde{\mathbf{y}} \in \mathbb{R}^q$:

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

Here $\mathbf{z} \in \mathbb{R}^r$ is a rank-r signal vector with i.i.d. entries of mean zero and variance one, and A and B are $p \times r$ and $q \times r$ deterministic factor loading matrices, respectively. $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$ are two independent noise vectors with i.i.d. entries of mean zero and variance one, and $\mathbf{C}_1$ and $\mathbf{C}_2$ are $p \times p$ and $q \times q$ deterministic population covariance matrices. Then we can write the data matrices in (1.1) into

$$\widetilde{\mathcal{X}} := AZ + \mathbf{C}_1^{1/2} X, \quad \widetilde{\mathcal{Y}} := BZ + \mathbf{C}_2^{1/2} Y, \tag{1.2}$$

where X, Y and Z are respectively $p \times n$, $q \times n$ and $r \times n$ matrices with i.i.d. entries of mean zero and variance n^{-1}. We consider the high-dimensional setting with a low-rank signal, that is, $p/n \to c_1$ and $q/n \to c_2$ as $n \to \infty$ for some constants $c_1 \in (0,1)$ and $c_2 \in (0, 1-c_1)$, and r is a fixed integer that does not depend on n.

For the model (1.2), the PCC matrix is given by

$$\widetilde{\mathbf{\Sigma}} = (\mathbf{C}_1 + AA^\top)^{-1/2} AB^\top (\mathbf{C}_2 + BB^\top)^{-1} BA^\top (\mathbf{C}_1 + AA^\top)^{-1/2},$$

which is of rank at most r. We order the nontrivial eigenvalues of $\widetilde{\mathbf{\Sigma}}$ as $t_1 \geqslant t_2 \geqslant \cdots \geqslant t_r \geqslant 0$. Under the Gaussian assumption, that is, X, Y and Z are independent random matrices with i.i.d. Gaussian entries, Bao et al. [7] proved that for any $1 \leqslant i \leqslant r$, $\widetilde{\lambda}_i$ exhibits very different behaviors depending on whether t_i is below or above the threshold t_c, where

$$t_c := \sqrt{\frac{c_1 c_2}{(1-c_1)(1-c_2)}}. \tag{1.3}$$

More precisely, if $t_i < t_c$, then the corresponding sample CCC $\widetilde{\lambda}_i$ sticks to the right edge λ_+ of the bulk eigenvalue spectrum (cf. (2.13)) of the SCC matrix, and $n^{2/3}(\widetilde{\lambda}_i - \lambda_+)$ converges weakly to a type-1 Tracy-Widom distribution. On the other hand, if $t_i > t_c$, then it gives rise to an outlier $\widetilde{\lambda}_i$ that lies around a fixed location $\theta_i \in (\lambda_+, 1)$ determined by t_i, c_1 and c_2. Furthermore, $n^{1/2}(\widetilde{\lambda}_i - \theta_i)$ converges weakly to a centered Gaussian. Such an abrupt change of the behavior of $\widetilde{\lambda}_i$ when t_i crosses the threshold t_c is generally called a *BBP transition*, which dates back to the seminal work of Baik, Ben Arous and Péché [5] on spiked sample covariance matrices. The BBP transition has been observed in many random matrix ensembles with finite rank perturbations. Without attempting to be comprehensive, we mention the references [14, 15, 24, 36, 37, 42] on deformed Wigner matrices, [3, 5, 6, 12, 25, 33, 41] on spiked sample covariance matrices, [17, 49, 51] on spiked separable covariance matrices, and [8, 9, 10, 47] on several other deformed random matrix ensembles. In our setting, the SCC matrix $\mathcal{C}_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}}$ can be regarded as a finite rank perturbation of the SCC matrix in the null case with $r = 0$.

A natural question is whether the results in [7] hold universally, that is, whether $\widetilde{\lambda}_i$ satisfies the same properties if we only assume certain moment conditions on the entries of X, Y and Z. In fact, the proof in [7] depends crucially on the rotational invariance of multivariate Gaussian distributions under orthogonal transforms, and it is hard (if possible) to be extended to the data matrices with generally distributed entries. In this paper, we answer the above question definitely, and show the universality of the results in [7]. Moreover, we highlight the following improvements over the results in [7].

- Theorem 2.14 shows that the following results hold assuming only a finite fourth moment condition (actually we require a slightly weaker condition (2.34)): for $1 \leqslant i \leqslant r$, $n^{2/3}(\widetilde{\lambda}_i - \lambda_+)$ converges weakly to Tracy-Widom distribution if $t_i < t_c$, while $t_i \to \theta_i$ in probability if $t_i > t_c$.

- We obtain quantitative versions of all the results under general moment assumptions: Theorem 2.9, provides almost sharp convergences rates for the sample CCC's; Theorem 2.11 provides an almost sharp eigenvalue sticking estimate, which shows that the eigenvalues of the SCC matrix stick to those of the null SCC matrix with $r = 0$.

- Our results hold even when some t_i's are close to the BBP transition threshold t_c, and when there are groups of near-degenerate t_i's—both of these two cases are ruled out in the assumptions of [7].

To complete the theory, we still need to prove the CLT for $\widetilde{\lambda}_i$ when $t_i > t_c$. Due to length constraint, we postpone it to [39], where we will show that $n^{1/2}(\widetilde{\lambda}_i - \theta_i)$ still converges to a center Gaussian but with a limiting variance that is different from the one in the Gaussian case. Instead of using the rotational invariance of multivariate Gaussian distributions, the proofs in this paper are based on a linearization method developed in [50], 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 (cf. (3.3)). Moreover, an optimal local law was proved for the resolvent $G := H^{-1}$ in [50], which is the basis of all the proofs in this paper. Our approach is relatively more flexible and allows us to obtain precise convergence rates for the eigenvalues of the SCC matrix.

This paper is organized as follows. In Section 2, we define the model and state the main results, Theorem 2.9, Theorem 2.11 and Theorem 2.14. In Section 3, we introduce the linearization method and collect some basic tools that will be used in the proof. Then we will give the proofs of Theorem 2.9, Theorem 2.11 and Theorem 2.14 in Sections 4, 5 and 6. Our proofs utilize a result on the eigenvalues of the null SCC matrix, Lemma 2.7, which will be proved in Section 7.

Conventions. For two quantities a_n and b_n depending on n, we use $a_n = \mathrm{O}(b_n)$ to mean that $|a_n| \leqslant C|b_n|$ for a constant $C > 0$, and use $a_n = \mathrm{o}(b_n)$ to indicate that $|a_n| \leqslant c_n|b_n|$ for a positive sequence of numbers $c_n \downarrow 0$ as $n \to \infty$. We will 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 $\|A\|$ to denote its operator norm. For a vector $\mathbf{v}$, we use $\|\mathbf{v}\|$ to denote its Euclidean norm. In this paper, we will write an identity matrix as I or 1 without causing any confusions.

2 The model and main results

2.1 The model

We consider two independent families of data matrices $X = (x_{ij})$ and $Y = (y_{ij})$, which are of dimensions $p \times n$ and $q \times n$, 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 real independent random variables satisfying that

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

To be more general, we do not assume that these random variables are identically distributed. Then we define the following data model with finite rank correlation:

$$\widetilde{\mathcal{X}} := \mathbf{C}_1^{1/2} X + \widetilde{A} Z, \quad \widetilde{\mathcal{Y}} := \mathbf{C}_2^{1/2} Y + \widetilde{B} Z,$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are $p \times p$ and $q \times q$ deterministic positive definite symmetric covariance matrices, $\widetilde{A}$ and $\widetilde{B}$ are $p \times r$ and $q \times r$ deterministic matrices, and $Z = (z_{ij})$ is an $r \times n$ random matrix which leads to the nontrivial correlation between $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}}$. We assume that Z is independent of X and Y, and the entries z_{ij}, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant n$, are independent random variables satisfying

$$\mathbb{E}z_{ij} = 0, \qquad \mathbb{E}|z_{ij}|^2 = n^{-1}. \tag{2.2}$$

In this paper, we study the eigenvalues of the *sample canonical correlation* (SCC) matrix

$$\mathcal{C}_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}} = \left(\widetilde{\mathcal{X}}\widetilde{\mathcal{X}}^\top\right)^{-1/2} \left(\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}^\top\right) \left(\widetilde{\mathcal{Y}}\widetilde{\mathcal{Y}}^\top\right)^{-1} \left(\widetilde{\mathcal{Y}}\widetilde{\mathcal{X}}^\top\right) \left(\widetilde{\mathcal{X}}\widetilde{\mathcal{X}}^\top\right)^{-1/2}.$$

In particular, we are interested in the relations between the eigenvalues of $\mathcal{C}_{\widetilde{\mathcal{X}}\widetilde{\mathcal{Y}}}$ and the *canonical correlation coefficients*—the square roots of the eigenvalues of the *population canonical correlation* (PCC) matrix

$$\widetilde{\mathbf{\Sigma}} := \widetilde{\Sigma}_{xx}^{-1/2} \widetilde{\Sigma}_{xy} \widetilde{\Sigma}_{yy}^{-1} \widetilde{\Sigma}_{yx} \widetilde{\Sigma}_{xx}^{-1/2},$$

where

$$\widetilde{\Sigma}_{xx} := \mathbf{C}_1 + \widetilde{A}\widetilde{A}^\top, \quad \widetilde{\Sigma}_{yy} := \mathbf{C}_2 + \widetilde{B}\widetilde{B}^\top, \quad \widetilde{\Sigma}_{xy} = \widetilde{\Sigma}_{yx}^\top := \widetilde{A}\widetilde{B}^\top.$$

Note that the eigenvalues of both SCC and PCC matrices are unchanged under the non-singular transformations $\widetilde{\mathcal{X}} \to \mathcal{X} := \mathbf{C}_1^{-1/2}\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{Y}} \to \mathcal{Y} := \mathbf{C}_2^{-1/2}\widetilde{\mathcal{Y}}$. Thus it suffices to consider the data matrices

$$\mathcal{X} := X + AZ, \quad \mathcal{Y} := Y + BZ, \quad \text{where } A := \mathbf{C}_1^{-1/2}\widetilde{A}, \quad B := \mathbf{C}_2^{-1/2}\widetilde{B}. \tag{2.3}$$

We assume that A and B have the following singular value decompositions:

$$A = \sum_{i=1}^r a_i\, \mathbf{u}_i^a (\mathbf{v}_i^a)^\top, \quad B = \sum_{i=1}^r b_i\, \mathbf{u}_i^b (\mathbf{v}_i^b)^\top, \tag{2.4}$$

where $\{a_i\}$ and $\{b_i\}$ are the singular values, $\{\mathbf{u}_i^a\}$ and $\{\mathbf{u}_i^b\}$ are the left singular vectors, and $\{\mathbf{v}_i^a\}$ and $\{\mathbf{v}_i^b\}$ are the right singular vectors. We assume that for some constant $C > 0$,

$$0 \leqslant a_r \leqslant \cdots \leqslant a_2 \leqslant a_1 \leqslant C, \quad 0 \leqslant b_r \leqslant \cdots \leqslant b_2 \leqslant b_1 \leqslant C. \tag{2.5}$$

In this paper, we consider the high dimensional setting, that is,

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

For simplicity of notations, we will always abbreviate $c_1(n) \equiv c_1$ and $c_2(n) \equiv c_2$ for the rest of the paper. Without loss of generality, we assume that $c_1 \geqslant c_2$.

We now summarize the main assumptions for future reference. For our purpose, we relax the assumptions (2.1) and (2.2) a little bit. One can refer to Corollary 2.13 for the reason of this extension.

Assumption 2.1. *Fix a small constant $\tau > 0$.*

(i) $X = (x_{ij})$ and $Y = (Y_{ij})$ are two real independent $p \times n$ and $q \times n$ random matrices. Their entries are independent random variables that satisfy the following moment conditions:

$$\max_{i,j} |\mathbb{E}x_{ij}| \leqslant n^{-2-\tau}, \quad \max_{i,j} |\mathbb{E}y_{ij}| \leqslant n^{-2-\tau}, \tag{2.6}$$

$$\max_{i,j} \left|\mathbb{E}|x_{ij}|^2 - n^{-1}\right| \leqslant n^{-2-\tau}, \quad \max_{i,j} \left|\mathbb{E}|y_{ij}|^2 - n^{-1}\right| \leqslant n^{-2-\tau}. \tag{2.7}$$

We remark that (2.6) and (2.7) are (slightly) more general than (2.1).

(ii) $Z = (z_{ij})$ is a real $r \times n$ random matrix that is independent of X and Y, and its entries are independent random variables that satisfy the following moment conditions:

$$\max_{i,j} |\mathbb{E}z_{ij}| \leqslant n^{-1-\tau}, \quad \max_{i,j} \left|\mathbb{E}|z_{ij}|^2 - n^{-1}\right| \leqslant n^{-1-\tau}. \tag{2.8}$$

(iii) We assume that

$$r \leqslant \tau^{-1}, \quad \tau \leqslant c_2 \leqslant c_1, \quad c_1 + c_2 \leqslant 1 - \tau. \tag{2.9}$$

(iv) We consider the data model in (2.3)*, where A and B satisfy* (2.4) *and* (2.5).

In this paper, we will study the SCC matrix

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

the null SCC matrix

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

where

$$S_{xx} := XX^\top, \quad S_{yy} := YY^\top, \quad S_{xy} = S_{yx}^\top := XY^\top \tag{2.10}$$

and the PCC matrix

$$\boldsymbol{\Sigma}_{\mathcal{X}\mathcal{Y}} := \Sigma_{xx}^{-1/2}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1/2}$$

where

$$\Sigma_{xx} = I_p + AA^\top, \quad \Sigma_{yy} = I_q + BB^\top, \quad \Sigma_{xy} = \Sigma_{yx}^\top = AB^\top.$$

Moreover, we will also consider the following matrices:

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

and

$$\mathcal{C}_{YX} = S_{yy}^{-1/2} S_{yx} S_{xx}^{-1} S_{xy} S_{yy}^{-1/2}, \quad \boldsymbol{\Sigma}_{\mathcal{Y}\mathcal{X}} := \Sigma_{yy}^{-1/2}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1/2}.$$

Finally, we define another null SCC matrix $\mathcal{C}^b_{\mathcal{Y}X}$ as

$$\mathcal{C}^b_{\mathcal{Y}X} := (S^b_{yy})^{-1/2} S^b_{yx} S_{xx}^{-1} S^b_{xy} (S^b_{yy})^{-1/2} \quad \text{with} \quad S^b_{yy} := \mathcal{Y}\mathcal{Y}^\top, \quad S^b_{xy} = (S^b_{yx})^\top := X\mathcal{Y}^\top. \tag{2.11}$$

The matrix $\mathcal{C}^b_{X\mathcal{Y}}$ can be defined in the obvious way.

2.2 The main results

In the null case with $r = 0$, we denote the eigenvalues of $\mathcal{C}_{YX}$ by $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_q \geqslant 0$. Then $\mathcal{C}_{XY}$ 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 denote the ESD of $\mathcal{C}_{YX}$ by

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

It is known [46, 52] that, almost surely, F_n converges weakly to a deterministic probability distribution $F(x)$ with density

$$f(x) = \frac{1}{2\pi c_2}\frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{x(1-x)}\mathbf{1}_{\lambda_- \leqslant x \leqslant \lambda_+}, \tag{2.12}$$

where

$$\lambda_\pm := \left(\sqrt{c_1(1-c_2)} \pm \sqrt{c_2(1-c_1)}\right)^2. \tag{2.13}$$

For the model (2.3) with finite rank correlations, we denote the eigenvalues of $\mathcal{C}_{\mathcal{Y}\mathcal{X}}$ by $\widetilde{\lambda}_1 \geqslant \widetilde{\lambda}_2 \geqslant \cdots \geqslant \widetilde{\lambda}_q \geqslant 0$, while $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ has $(p-q)$ more trivial zero eigenvalues $\widetilde{\lambda}_{q+1} = \cdots = \widetilde{\lambda}_p = 0$. We denote the eigenvalues of $\boldsymbol{\Sigma}_{\mathcal{X}\mathcal{Y}}$ by

$$t_1 \geqslant t_2 \geqslant \cdots \geqslant t_r \geqslant t_{r+1} = \cdots = t_q = 0. \tag{2.14}$$

Suppose the entries of X and Y are i.i.d. Gaussian. Then it was proved in [7] that, if $t_i > t_c$ (recall (1.3)), $\widetilde{\lambda}_i - \theta_i \to 0$ almost surely, where

$$\theta_i := t_i\left(1 - c_1 + c_1 t_i^{-1}\right)\left(1 - c_2 + c_2 t_i^{-1}\right); \tag{2.15}$$

if $t_i \leqslant t_c$, $\widetilde{\lambda}_i - \lambda_+ \to 0$ almost surely. Note that, for $t_i > t_c$ we have $\theta_i > \lambda_+$, that is, $\widetilde{\lambda}_i$ will be an outlier that is detached from the support $[\lambda_-, \lambda_+]$ of the limiting distribution $F(x)$.

Before stating the main results, we first introduce the following notion of stochastic domination. It was first introduced in [19], and subsequently used in many works on random matrix theory, such as [11, 12, 13, 20, 21, 38]. 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 2.2 (Stochastic domination and high probability event)**.** *(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 an n-dependent parameter set. We say ξ is stochastically dominated by ζ, uniformly in u, if for any small constant $\varepsilon > 0$ and large constant $D > 0$,

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

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 the spectral parameter z). If ξ is complex and 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 sense of operator norm as follows. Let A be a family of matrices and ζ be a family of nonnegative random variables. Then $A = \mathrm{O}_\prec(\zeta)$ means that $\|A\| \prec \zeta$.

(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. Moreover, we say Ξ holds with high probability on an event Ω, if for any constant $D > 0$, $\mathbb{P}(\Omega \backslash \Xi) \leqslant n^{-D}$ for large enough n.

The following lemma collects basic properties of stochastic domination $\prec$, which will be used tacitly throughout this paper.

Lemma 2.3 (Lemma 3.2 in [11])**.** *Let ξ and ζ be families of nonnegative random variables, and let $C > 0$ be any (large) constant.*

(i) Suppose that $\xi(u,v) \prec \zeta(u,v)$ uniformly in $u \in U$ and $v \in V$. If $|V| \leqslant n^C$, 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. If $\xi(u) \prec \Psi(u)$ uniformly in u, then we have $\mathbb{E}\xi(u) \prec \Psi(u)$ uniformly in u.

We introduce the following bounded support condition for the random matrices considered in this paper.

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

$$\max_{i,j} |x_{ij}| \prec \phi_n. \tag{2.16}$$

Whenever (2.16) holds, we say that X has support ϕ_n.

In the rest of this paper, ϕ_n is usually a deterministic parameter satisfying that $n^{-1/2} \leqslant \phi_n \leqslant n^{-c_\phi}$ for some small constant $c_\phi > 0$.

In this paper, we will also consider the case where $|t_i - t_c| = \mathrm{o}(1)$, i.e. the spike t_i is very close to the BBP transition threshold. Suppose that X and Y have bounded support ϕ_n, and Z has bounded support ψ_n. Then we make the following assumption.

Assumption 2.5. *We assume that for some integer $0 \leqslant r_+ \leqslant r$, the following statement holds:*

$$t_i - t_c \geqslant n^{-1/3} + \psi_n + \phi_n \quad \text{if and only if} \quad 1 \leqslant i \leqslant r_+. \tag{2.17}$$

The lower bound is chosen for definiteness, and it can be replaced with any n-dependent parameter that is of the same order.

Before stating our main results on the eigenvalues of the SCC matrix $\mathcal{C}_{\mathcal{Y}\mathcal{X}}$, we describe the behaviors of the eigenvalues of the null SCC matrix $\mathcal{C}^b_{\mathcal{Y}X}$ (recall (2.11)). We denote its eigenvalues by $\lambda_1^b \geqslant \lambda_2^b \geqslant \cdots \geqslant \lambda_q^b$. Then we define the quantiles of the density (2.12), which give the classical locations of the eigenvalues.

Definition 2.6. *The classical location γ_j of the j-th eigenvalue is defined as*

$$\gamma_j := \sup_x \left\{ \int_x^{+\infty} f(x)\mathrm{d}x > \frac{j-1}{q} \right\}, \tag{2.18}$$

where f is defined in (2.12)*. Note that we have $\gamma_1 = \lambda_+$ and $\lambda_+ - \gamma_j \sim (j/n)^{2/3}$ for $j > 1$.*

We have the following eigenvalue rigidity and edge universality result for $\mathcal{C}^b_{\mathcal{Y}X}$. If $B = 0$, i.e. there is no Z term, then the same results have been proved in [50] under the same assumptions.

Lemma 2.7. *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 $c_\phi > 0$, and Z has bounded support ψ_n such that $n^{-1/2} \leqslant \psi_n \leqslant n^{-c_\psi}$ for some constant $c_\psi > 0$. Assume that*

$$\max_{i,j} \mathbb{E}|x_{ij}|^3 = \mathrm{O}(n^{-3/2}), \quad \max_{i,j} \mathbb{E}|y_{ij}|^3 = \mathrm{O}(n^{-3/2}), \quad \max_{i,j} \mathbb{E}|x_{ij}|^4 \prec n^{-2}, \quad \max_{i,j} \mathbb{E}|y_{ij}|^4 \prec n^{-2}. \tag{2.19}$$

Then the eigenvalues of the null SCC matrix $\mathcal{C}^b_{\mathcal{Y}X}$ satisfy the following eigenvalue rigidity estimate: for any constant $\delta > 0$ and all $1 \leqslant j \leqslant (1-\delta)q$,

$$|\lambda_i^b - \gamma_i| \prec i^{-1/3} n^{-2/3}. \tag{2.20}$$

Moreover, we have that for any fixed $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}\left[\left(n^{2/3} \frac{\lambda_i^b - \lambda_+}{c_{TW}} \leqslant s_i \right)_{1\leqslant i\leqslant k} \right] = \lim_{n\to\infty} \mathbb{P}^{GOE} \left[\left(n^{2/3}(\lambda_i - 2) \leqslant s_i \right)_{1\leqslant i\leqslant k} \right], \tag{2.21}$$

for all $(s_1, s_2, \ldots, s_k) \in \mathbb{R}^k$, 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 GOE (Gaussian orthogonal ensemble), which is an $n \times n$ symmetry matrix with independent (up to symmetry) Gaussian entries of mean zero and variance n^{-1}.

Remark 2.8. Taking $k = 1$ in (2.21), we obtain that

$$n^{2/3} \frac{\lambda_1^b - \lambda_+}{c_{TW}} \Rightarrow F_1,$$

where F_1 is the type-1 Tracy-Widom distribution as given by [44, 45]. Moreover, the joint distribution of the largest k eigenvalues of GOE can be written in terms of the Airy kernel for any fixed k [26]. Hence (2.21) gives a complete description of the finite-dimensional correlation functions of the edge eigenvalues of $\mathcal{C}^b_{\mathcal{Y}X}$.

Now we are ready to state our main results on the eigenvalues of the SCC matrix $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$. We denote

$$\Delta_i := |t_i - t_c|, \quad \alpha_+ := \min_{1 \leqslant i \leqslant r} |t_i - t_c|. \tag{2.22}$$

We first describe the convergence of the outlier eigenvalues and the extreme non-outlier eigenvalues.

Theorem 2.9. *Suppose the assumptions of Lemma 2.7 and Assumption 2.5 hold. Then we have the following estimates: for $1 \leqslant i \leqslant r_+$, we have*

$$|\widetilde{\lambda}_i - \theta_i| \prec (\psi_n + \phi_n)\Delta_i + n^{-1/2}\Delta_i^{1/2}; \tag{2.23}$$

for any $r_+ + 1 \leqslant i \leqslant \varpi$, where ϖ is a fixed integer, and any constant $\varepsilon > 0$, we have

$$- n^{-2/3+\varepsilon} < \widetilde{\lambda}_i - \lambda_+ \leqslant n^{\varepsilon}(\psi_n^2 + \phi_n^2 + n^{-2/3}) \quad \textit{with high probability.} \tag{2.24}$$

Remark 2.10. This theorem gives precise large deviation bounds on the locations of the outliers and the first few extreme non-outlier eigenvalues. Consider a small support case with $\phi_n + \psi_n \leqslant n^{-1/3}$ (this holds with probability $1 - \mathrm{o}(1)$ if we assume the existence of 12-th moment). Then (2.23) and (2.24) show that the fluctuation of the i-th eigenvalue changes from the order $(\psi_n + \phi_n)\Delta_i + n^{-1/2}\Delta_i^{1/2}$ to $n^{-2/3}$ when Δ_i crosses the scale $n^{-1/3}$. This implies the occurrence of the BBP transition.

For the non-outlier eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$, they are sticked to the corresponding eigenvalues of $\mathcal{C}^b_{X\mathcal{Y}}$ as given by the following lemma.

Theorem 2.11. *Suppose the assumptions of Lemma 2.7 and Assumption 2.5 hold. Assume that $\alpha_+ \geqslant n^{\varepsilon_0}(\psi_n + \phi_n)$ for some constant $\varepsilon_0 > 0$. Then we have the eigenvalue sticking estimate*

$$|\widetilde{\lambda}_{i+r_+} - \lambda_i^b| \prec n^{-1}\alpha_+^{-1} \tag{2.25}$$

for all $i \leqslant (1-\delta)q$, where $\delta > 0$ is any small constant.

Remark 2.12. Theorem 2.11 establishes the large deviation bounds on the non-outlier eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ with respect to the eigenvalues of $\mathcal{C}^b_{X\mathcal{Y}}$. In particular, when $\alpha_+ \gg n^{-1/3}$, the right-hand side of (2.25) is much smaller than $n^{-2/3}$ for $i = \mathrm{O}(1)$. Together with (2.21) for λ_i^b, (2.25) implies that the largest non-outlier eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ also converge to the Tracy-Widom law as long as the population canonical correlation coefficients t_i are away from the transition threshold t_c at least by $\alpha_+ \gg n^{-1/3}$.

Notice that applying (2.25) to $\mathcal{C}^b_{X\mathcal{Y}}$ and $\mathcal{C}_{XY}$ also gives that the eigenvalue λ_i^b are stick to λ_i for $1 \leqslant i \leqslant (1-\delta)q$. Thus we obtain the following eigenvalue sticking estimate

$$|\widetilde{\lambda}_{i+r_+} - \lambda_i| \prec n^{-1}\alpha_+^{-1}. \tag{2.26}$$

The reason why we state (2.25) instead of (2.26) in Theorem 2.11 will be explained below (5.2).

Using a simple cutoff argument, it is easy to obtain the following corollary under the finite a-th moment assumption for any fixed $a > 4$. Since we did not assume that the entries of X and Y are identically distributed, the means and variances of the truncated entries may be different. This is why we have assumed the slightly more general mean and variance conditions (2.6)–(2.8).

Corollary 2.13. *Assume that $X = (x_{ij})$, $Y = (Y_{ij})$ and $Z = (z_{ij})$ are respectively $p \times n$, $q \times n$ and $r \times n$ real matrices, whose entries are independent random variables satisfying* (2.1), (2.2) *and*

$$\max_{i,j} \mathbb{E}|\sqrt{n}x_{ij}|^a \leqslant C, \quad \max_{i,j} \mathbb{E}|\sqrt{n}y_{ij}|^a \leqslant C, \quad \max_{i,j} \mathbb{E}|\sqrt{n}z_{ij}|^b \leqslant C, \tag{2.27}$$

for some constants $a > 4$, $b > 2$ and $C > 0$. Suppose Assumption 2.1 (iii) and Assumption 2.5 hold with

$$\phi_n = n^{-1/2+2/a}, \quad \psi_n = n^{-1/2+1/b}. \tag{2.28}$$

Then we have that for $1 \leqslant i \leqslant r_+$,

$$|\widetilde{\lambda}_i - \theta_i| \leqslant n^{\varepsilon} \left[(\psi_n + \phi_n)\, \Delta_i + n^{-1/2} \Delta_i^{1/2} \right] \quad \text{with probability } 1 - \mathrm{o}(1), \tag{2.29}$$

for any small constant $\varepsilon > 0$. Moreover, assume that the eigenvalues of $\mathbf{\Sigma}_{\mathcal{X}\mathcal{Y}}$ satisfy that

$$\alpha_+ \geqslant n^{\varepsilon_0}(\psi_n + \phi_n) + n^{-1/3+\varepsilon_0} \tag{2.30}$$

for a constant $\varepsilon_0 > 0$. Then we have that for any fixed $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}\left[\left(n^{2/3} \frac{\widetilde{\lambda}_{i+r_+} - \lambda_+}{c_{TW}} \leqslant s_i \right)_{1 \leqslant i \leqslant k} \right] = \lim_{n\to\infty} \mathbb{P}^{GOE} \left[\left(n^{2/3}(\lambda_i - 2) \leqslant s_i \right)_{1 \leqslant i \leqslant k} \right], \tag{2.31}$$

for all $(s_1, s_2, \ldots, s_k) \in \mathbb{R}^k$.

Proof. For ϕ_n and ψ_n in (2.28), we introduce the truncated matrices $\widetilde{X}$, $\widetilde{Y}$ and $\widetilde{Z}$ defined as

$$\widetilde{X}_{ij} := x_{ij} \mathbf{1}_{|x_{ij}| \leqslant \phi_n n^{\varepsilon}}, \quad \widetilde{Y}_{ij} := y_{ij} \mathbf{1}_{|y_{ij}| \leqslant \phi_n n^{\varepsilon}}, \quad \widetilde{Z}_{ij} := z_{ij} \mathbf{1}_{|z_{ij}| \leqslant \psi_n n^{\varepsilon}}.$$

for a sufficiently small constant $\varepsilon > 0$. Combining the moment conditions in (2.27) with Markov's inequality, we obtain that

$$\mathbb{P}(\widetilde{X} \neq X, \widetilde{Y} \neq Y, \widetilde{Z} \neq Z) = \mathrm{O}\left(n^{-a\varepsilon} + n^{-b\varepsilon} \right), \tag{2.32}$$

by a simple union bound. Using (2.27) and integration by parts, we can also verify that

$$\mathbb{E}\, |x_{ij}|\, 1_{|x_{ij}| > \phi_n n^{\varepsilon}} \leqslant n^{-2-\varepsilon}, \quad \mathbb{E}\, |x_{ij}|^2\, 1_{|x_{ij}| > \phi_n n^{\varepsilon}} \leqslant n^{-2-\varepsilon}, \tag{2.33}$$

For example, for the first estimate in (2.33), we have that

$$\begin{aligned}
&\mathbb{E}\, |\mathbf{1}\left(|x_{ij}| > \phi_n n^{\varepsilon}\right) x_{ij}| = \int_0^{\infty} \mathbb{P}\left(|\mathbf{1}\left(|x_{ij}| > \phi_n n^{\varepsilon}\right) x_{ij}| > s \right) \mathrm{d}s \\
&= \int_0^{\phi_n n^{\varepsilon}} \mathbb{P}\left(|x_{ij}| > \phi_n n^{\varepsilon} \right) \mathrm{d}s + \int_{\phi_n n^{\varepsilon}}^{\infty} \mathbb{P}\left(|x_{ij}| > s \right) \mathrm{d}s \\
&\lesssim \int_0^{\phi_n n^{\varepsilon}} \left(n^{1/2+\varepsilon} \phi_n \right)^{-a} \mathrm{d}s + \int_{\phi_n n^{\varepsilon}}^{\infty} \left(\sqrt{n} s \right)^{-a} \mathrm{d}s \leqslant n^{-\frac{1}{2} - 2\frac{a-1}{a} - (a-1)\varepsilon} \leqslant n^{-2-\varepsilon},
\end{aligned}$$

where in the third step we used (2.27) and Markov's inequality, and in the last step we used $a > 4$. The second estimate of (2.33) can be proved in a similar way. Note that (2.33) implies

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

Moreover, we trivially have

$$\mathbb{E}|\tilde{x}_{ij}|^3 \leqslant \mathbb{E}|x_{ij}|^3 = \mathrm{O}(n^{-3/2}), \quad \mathbb{E}|\tilde{x}_{ij}|^4 \leqslant \mathbb{E}|x_{ij}|^4 = \mathrm{O}(n^{-2}).$$

Similar estimates also hold for the entries of $\widetilde{Y}$. Hence $\widetilde{X}$ and $\widetilde{Y}$ are random matrices satisfying Assumption 2.1 (i) and condition (2.19). For $\widetilde{Z}$, using (2.27) and a similar argument we can check that

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

Hence Z is a random matrix satisfying Assumption 2.1 (ii). Now combining (2.32) with Theorem 2.9, we conclude (2.29). Next combining (2.32) with Theorem 2.11, we obtain that

$$|\widetilde{\lambda}_{r_++i} - \lambda_i^b| \prec n^{-1}\alpha_+^{-1} \leqslant n^{-2/3-\varepsilon_0}, \quad 1 \leqslant i \leqslant k.$$

Together with Lemma 2.7, it concludes (2.31). □

If the entries of X, Y are Z are identically distributed, then we can obtain the following result under the weaker tail condition (2.34). We believe it to be a sharp condition.

Theorem 2.14. *Suppose Assumption 2.1 (iii) and Assumption 2.5 hold. Assume that $x_{ij} = n^{-1/2}\widehat{x}_{ij}$, $y_{ij} = n^{-1/2}\widehat{y}_{ij}$ and $z_{ij} = n^{-1/2}\widehat{z}_{ij}$, where $\{\widehat{x}_{ij}\}$, $\{\widehat{y}_{ij}\}$ and $\{\widehat{z}_{ij}\}$ are three independent families of i.i.d. random variables with mean zero and variance one. Moreover, we assume the tail condition*

$$\lim_{t\to\infty} t^4 \left[\mathbb{P}\left(|\widehat{x}_{11}| \geqslant t\right) + \mathbb{P}\left(|\widehat{y}_{11}| \geqslant t\right)\right] = 0. \tag{2.34}$$

We assume that the eigenvalues of $\mathbf{\Sigma}_{\mathcal{X}\mathcal{Y}}$ converge as $n \to \infty$ with

$$\lim_n t_{r_+} > t_c > \lim_n t_{r_++1}. \tag{2.35}$$

Then both (2.31) *and the following convergence hold:*

$$\widetilde{\lambda}_i - \theta_i \to 0 \quad \text{in probability.} \tag{2.36}$$

Finally, for an outlier eigenvalue, $n^{1/2}(\widetilde{\lambda}_i - \theta_i)$ actually converges to a normal distribution, which has been proved in [7] for the Gaussian case and for well-separated outliers, i.e. the outliers are either exactly degenerate or separated from each other by a distance of order 1. The proof for the general distribution case and for near-degenerate outliers is quite involved, and, considering the length of this paper, we postpone it to another paper [39].

3 Linearization method and local laws

The self-adjoint linearization method has been proved to be useful in studying the local laws of random matrices of Gram type [1, 2, 16, 18, 38, 48, 49]. We now introduce a generalization of this method, which was introduced in [50] to study the null SCC matrix $\mathcal{C}_{XY}$. For the discussion below, we assume that $\mathcal{X}\mathcal{X}^\top$, $\mathcal{Y}\mathcal{Y}^\top$, $XX^\top$ and $YY^\top$ are all non-singular almost surely. (This is trivially true if, say, the entries of X, Y and Z have continuous densities.) Then given $\lambda > 0$, it is an eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ if and only if the following equation holds:

$$\det\left(\left(\mathcal{X}\mathcal{Y}^\top\right)\left(\mathcal{Y}\mathcal{Y}^\top\right)^{-1}\left(\mathcal{Y}\mathcal{X}^\top\right) - \lambda\mathcal{X}\mathcal{X}^\top\right) = 0. \tag{3.1}$$

Using Schur complement, we can easily check that equation (3.1) is equivalent to

$$\det\begin{pmatrix} \lambda\mathcal{X}\mathcal{X}^\top & \lambda^{1/2}\mathcal{X}\mathcal{Y}^\top \\ \lambda^{1/2}\mathcal{Y}\mathcal{X}^\top & \lambda\mathcal{Y}\mathcal{Y}^\top \end{pmatrix} = 0.$$

Using Schur complement again, the above equation is equivalent to

$$\det\begin{pmatrix} 0 & \begin{pmatrix} \mathcal{X} & 0 \\ 0 & \mathcal{Y} \end{pmatrix} \\ \begin{pmatrix} \mathcal{X}^\top & 0 \\ 0 & \mathcal{Y}^\top \end{pmatrix} & \begin{pmatrix} \lambda I_n & \lambda^{1/2} I_n \\ \lambda^{1/2} I_n & \lambda I_n \end{pmatrix}^{-1} \end{pmatrix} = 0 \quad \text{if} \quad \lambda \notin \{0, 1\}. \tag{3.2}$$

Inspired by equation (3.2), we define the following $(p+q+2n)\times(p+q+2n)$ symmetric block matrix

$$H(\lambda) := \begin{pmatrix} 0 & \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \\ \begin{pmatrix} X^\top & 0 \\ 0 & Y^\top \end{pmatrix} & \begin{pmatrix} \lambda I_n & \lambda^{1/2} I_n \\ \lambda^{1/2} I_n & \lambda I_n \end{pmatrix}^{-1} \end{pmatrix}. \tag{3.3}$$

In general, we can extend the argument λ to $z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$ and call it $H(z)$, where we take $z^{1/2}$ to be the branch with positive imaginary part. Then using (2.3) and (2.4) we can write equation (3.2) as

$$\det\left[H(\lambda) + \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{E} \end{pmatrix} \begin{pmatrix} 0 & \mathcal{D} \\ \mathcal{D} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{U}^\top & 0 \\ 0 & \mathbf{E}^\top \end{pmatrix}\right] = 0, \tag{3.4}$$

where $\mathcal{D}$ is a $2r \times 2r$ matrix with

$$\mathcal{D} := \begin{pmatrix} \Sigma_a & 0 \\ 0 & \Sigma_b \end{pmatrix}, \quad \Sigma_a := \operatorname{diag}(a_1, \cdots, a_r), \quad \Sigma_b := \operatorname{diag}(b_1, \cdots, b_r), \tag{3.5}$$

and $\mathbf{U}$ and $\mathbf{E}$ are $(p+q)\times 2r$ and $2n \times 2r$ matrices, respectively, with

$$\mathbf{U} := \begin{pmatrix} (\mathbf{u}_1^a, \cdots, \mathbf{u}_r^a) & 0 \\ 0 & (\mathbf{u}_1^b, \cdots, \mathbf{u}_r^b) \end{pmatrix}, \quad \mathbf{E} := \begin{pmatrix} (Z^\top \mathbf{v}_1^a, \cdots, Z^\top \mathbf{v}_r^a) & 0 \\ 0 & (Z^\top \mathbf{v}_1^b, \cdots, Z^\top \mathbf{v}_r^b) \end{pmatrix}. \tag{3.6}$$

If λ is not an eigenvalue of $\mathcal{C}_{XY}$, then $H(\lambda)$ is non-singular by Schur complement, and (3.4) is equivalent to

$$\det\left[1 + \begin{pmatrix} 0 & \mathcal{D} \\ \mathcal{D} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{U}^\top & 0 \\ 0 & \mathbf{E}^\top \end{pmatrix} \frac{1}{H(\lambda)} \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{E} \end{pmatrix}\right] = 0, \tag{3.7}$$

where we used the identity $\det(1 + M_1 M_2) = \det(1 + M_2 M_1)$ for any matrices M_1 and M_2 of conformable dimensions. Inspired by the discussion above, we define the resolvent (or Green's function) as

$$G(z) := [H(z)]^{-1}, \quad z \in \mathbb{C}_+ \cup \mathbb{R}, \tag{3.8}$$

whenever the inverse exists. Note that although $H(\lambda)$ is not well-defined for $\lambda = 1$, we can still define $G(1) = \lim_{z\to 1} G(z)$ using Schur complement; see (3.14) and (3.15) below. In order to study the eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$, we need to obtain some estimates on the $4r \times 4r$ matrices

$$\begin{pmatrix} \mathbf{U}^\top & 0 \\ 0 & \mathbf{E}^\top \end{pmatrix} G(\lambda) \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{E} \end{pmatrix}.$$

These are provided by the *anisotropic local law* on $G(z)$, which is one of the main results in [50]. We will state it in Theorem 3.7 below.

For the proof of Theorem 2.11, we will also use a different representation of (3.7): if λ is not an eigenvalue of $\mathcal{C}_{X\mathcal{Y}}^b$, then λ is an eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ if and only if

$$\det\left[1 + \begin{pmatrix} 0 & \mathcal{D}_a \\ \mathcal{D}_a & 0 \end{pmatrix} \begin{pmatrix} \mathbf{U}_a^\top & 0 \\ 0 & \mathbf{E}_a^\top \end{pmatrix} G^b(\lambda) \begin{pmatrix} \mathbf{U}_a & 0 \\ 0 & \mathbf{E}_a \end{pmatrix}\right] = 0, \tag{3.9}$$

where

$$G^b(z) := \left[H^b(z)\right]^{-1}, \quad H^b(z) := \begin{pmatrix} 0 & \begin{pmatrix} X & 0 \\ 0 & \mathcal{Y} \end{pmatrix} \\ \begin{pmatrix} X^\top & 0 \\ 0 & \mathcal{Y}^\top \end{pmatrix} & \begin{pmatrix} zI_n & z^{1/2} I_n \\ z^{1/2} I_n & zI_n \end{pmatrix}^{-1} \end{pmatrix}, \tag{3.10}$$

and

$$\mathcal{D}_a := \begin{pmatrix} \Sigma_a & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{U}_a := \begin{pmatrix} (\mathbf{u}_1^a, \cdots, \mathbf{u}_r^a) & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{E}_a := \begin{pmatrix} (Z^\top \mathbf{v}_1^a, \cdots, Z^\top \mathbf{v}_r^a) & 0 \\ 0 & 0 \end{pmatrix}.$$

For simplicity of notations, we introduce the following index sets for our linearized matrices.

Definition 3.1 (Index sets)**.** *We define the index sets*

$$\mathcal{I}_1 := [\![1, p]\!], \quad \mathcal{I}_2 := [\![p+1, p+q]\!], \quad \mathcal{I}_3 := [\![p+q+1, p+q+n]\!], \quad \mathcal{I}_4 := [\![p+q+n+1, p+q+2n]\!].$$

We will consistently use the latin letters $i, j \in \mathcal{I}_1 \cup \mathcal{I}_2$ and greek letters $\mu, \nu \in \mathcal{I}_3 \cup \mathcal{I}_4$. Moreover, we will use the notations $\mathfrak{a}, \mathfrak{b} \in \mathcal{I} := \cup_{i=1}^4 \mathcal{I}_i$.

Then we define the following forms of resolvents that will be used in the proof.

Definition 3.2 (Resolvents)**.** *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 denote the $\mathcal{I}_\alpha \times \mathcal{I}_\alpha$ block of $G(z)$ by $\mathcal{G}_\alpha(z)$ for $\alpha = 1, 2, 3, 4$. Then we define the partial traces as*

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

Recalling the notations in (2.10)*, we define $\mathcal{H} := S_{xx}^{-1/2} S_{xy} S_{yy}^{-1/2}$ and*

$$R_1(z) := (\mathcal{H}\mathcal{H}^\top - z)^{-1}, \quad R_2(z) := (\mathcal{H}^\top \mathcal{H} - z)^{-1}, \quad m(z) := \frac{1}{q} \operatorname{Tr} R_2(z). \tag{3.11}$$

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

$$\operatorname{Tr} R_1 = \operatorname{Tr} R_2 - \frac{p-q}{z} = q m(z) - \frac{p-q}{z}, \tag{3.12}$$

since $\mathcal{C}_{XY} = \mathcal{H}\mathcal{H}^\top$ has $(p-q)$ more zero eigenvalues than $\mathcal{C}_{YX} = \mathcal{H}^\top \mathcal{H}$. Moreover, we define

$$R(z) := \begin{pmatrix} -z & -z^{1/2}\mathcal{H} \\ -z^{1/2}\mathcal{H}^\top & -z \end{pmatrix}^{-1}.$$

Finally, we can define $\mathcal{G}_L^b(z)$, $\mathcal{G}_R^b(z)$, $m_\alpha^b(z)$, $\mathcal{H}^b$, R^b etc. in the obvious way by replacing Y with $\mathcal{Y}$.

By Schur complement formula, we can check that

$$R(z) := \begin{pmatrix} R_1 & -z^{-1/2} R_1 \mathcal{H} \\ -z^{-1/2} \mathcal{H}^\top R_1 & R_2 \end{pmatrix}.$$

Let $\mathcal{H} = \sum_{k=1}^q \sqrt{\lambda_k} \xi_k \zeta_k^\top$ 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

$$R(z) = \sum_{k=1}^q \frac{1}{\lambda_k - z} \begin{pmatrix} \xi_k \xi_k^\top & -z^{-1/2}\sqrt{\lambda_k} \xi_k \zeta_k^\top \\ -z^{-1/2}\sqrt{\lambda_k} \zeta_k \xi_k^\top & \zeta_k \zeta_k^\top \end{pmatrix} - \frac{1}{z} \begin{pmatrix} \sum_{k=q+1}^p \xi_k \xi_k^\top & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.13}$$

On the other hand, applying Schur complement formula to $G(z)$, it is easy to get that

$$\mathcal{G}_L = \begin{pmatrix} S_{xx}^{-1/2} & 0 \\ 0 & S_{yy}^{-1/2} \end{pmatrix} R(z) \begin{pmatrix} S_{xx}^{-1/2} & 0 \\ 0 & S_{yy}^{-1/2} \end{pmatrix}. \tag{3.14}$$

Moreover, the other blocks take the forms

$$\mathcal{G}_R = \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix} + \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix} \begin{pmatrix} X^\top & 0 \\ 0 & Y^\top \end{pmatrix} \mathcal{G}_L \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}, \tag{3.15}$$

and

$$\mathcal{G}_{LR}(z) = -\mathcal{G}_L(z) \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix}, \quad \mathcal{G}_{RL}(z) = -\begin{pmatrix} zI_n & z^{1/2}I_n \\ z^{1/2}I_n & zI_n \end{pmatrix} \begin{pmatrix} X^\top & 0 \\ 0 & Y^\top \end{pmatrix} \mathcal{G}_L(z). \tag{3.16}$$

Expanding the product in (3.15) using (3.14), and calculating the partial traces, one can verify directly that

$$m_3(z) = z + \frac{1}{n}\left(-2zp - z^2 \operatorname{Tr} R_1 + z \operatorname{Tr} R_2\right) = c_2 z(1-z)m(z) + (1 - c_1 - c_2)z, \tag{3.17}$$

and

$$m_4(z) = z + \frac{1}{n}\left(-2zq - z^2 \operatorname{Tr} R_2 + z \operatorname{Tr} R_1\right) = c_2 z(1-z)m(z) - (c_1 - c_2) + (1 - 2c_2)z, \tag{3.18}$$

where we also used (3.12) in the derivations. In particular, we have the identity

$$m_3(z) - m_4(z) = (1-z)(c_1 - c_2). \tag{3.19}$$

We remark that all the above identities also hold for G^b, $\mathcal{G}_L^b(z)$, $\mathcal{G}_R^b(z)$, $m_\alpha^b(z)$ etc. with some obvious changes of notations.

Since S_{xx} and S_{yy} are standard sample covariance matrices, it is well-known that their eigenvalues are all inside the supports of the Marchenko-Pastur laws—$[(1-\sqrt{c_1})^2, (1+\sqrt{c_1})^2]$ and $[(1-\sqrt{c_2})^2, (1+\sqrt{c_2})^2]$—with probability $1 - \mathrm{o}(1)$ [4]. We denote the extreme eigenvalues of S_{xx} and S_{yy} by $\lambda_1(S_{xx}) \geqslant \lambda_p(S_{xx})$ and $\lambda_1(S_{yy}) \geqslant \lambda_q(S_{yy})$. We shall need some estimates on them with stronger probability bounds as given by the following lemma.

Lemma 3.3. *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 $c_\phi > 0$, and Z has bounded support ψ_n such that $n^{-1/2} \leqslant \psi_n \leqslant n^{-c_\psi}$ for some constant $c_\psi > 0$. Then for any constant $\varepsilon > 0$, we have that with high probability,*

$$(1-\sqrt{c_1})^2 - \varepsilon \leqslant \lambda_p(S_{xx}) \leqslant \lambda_1(S_{xx}) \leqslant (1+\sqrt{c_1})^2 + \varepsilon, \tag{3.20}$$

and

$$(1-\sqrt{c_2})^2 - \varepsilon \leqslant \lambda_q(S_{yy}) \leqslant \lambda_1(S_{yy}) \leqslant (1+\sqrt{c_2})^2 + \varepsilon. \tag{3.21}$$

Moreover, there exists a constant $c > 0$ such that with high probability,

$$c \leqslant \lambda_q(S_{yy}^b) \leqslant \lambda_1(S_{yy}^b) \leqslant c^{-1}, \tag{3.22}$$

where $\lambda_1(S_{yy}^b)$ and $\lambda_q(S_{yy}^b)$ are respectively the largest and smallest eigenvalues of S_{yy}^b.

Proof. The estimates (3.20) and (3.21) have been proved in Lemma 3.3 of [50]. To get (3.22), we write

$$S_{yy}^b = (I_q, B)\, WW^\top \begin{pmatrix} I_q \\ B^\top \end{pmatrix}, \quad W := \begin{pmatrix} Y \\ Z \end{pmatrix}.$$

Since $r/n \to 0$, the estimate (3.21) applied to $WW^\top$ gives that with high probability,

$$(1-\sqrt{c_2})^2 - \varepsilon \leqslant \lambda_{q+r}(WW^\top) \leqslant \lambda_1(WW^\top) \leqslant (1+\sqrt{c_2})^2 + \varepsilon.$$

Then using that for any unit vector $\mathbf{v} \in \mathbb{R}^q$, $\mathbf{v}^\top S_{yy}^b \mathbf{v} \sim \mathbf{v}^\top WW^\top \mathbf{v}$, we conclude (3.22). ☐

Let $m_{\alpha c}$ be the asymptotic limits of m_α for $\alpha = 1, 2, 3, 4$. In [50], we have obtained that

$$m_{1c}(z) = \frac{-z + c_1 + c_2 + \sqrt{(z-\lambda_-)(z-\lambda_+)}}{2(1-c_1)z(1-z)} - \frac{c_1}{(1-c_1)z}, \tag{3.23}$$

$$m_{2c}(z) = \frac{-z + c_1 + c_2 + \sqrt{(z-\lambda_-)(z-\lambda_+)}}{2(1-c_2)z(1-z)} - \frac{c_2}{(1-c_2)z}, \tag{3.24}$$

$$m_{3c}(z) = \frac{1}{2}\left[(1-2c_1)z + c_1 - c_2 + \sqrt{(z-\lambda_-)(z-\lambda_+)}\right], \tag{3.25}$$

$$m_{4c}(z) = \frac{1}{2}\left[(1-2c_2)z + c_2 - c_1 + \sqrt{(z-\lambda_-)(z-\lambda_+)}\right], \tag{3.26}$$

where $\lambda_\pm$ are defined in (2.13). On can check that when $z \to 1$, both $m_{1c}(z)$ and $m_{2c}(z)$ have finite limits, and without loss of generality, we still denote them by $m_{1c}(1)$ and $m_{2c}(1)$. By (3.17), we can easily obtain the asymptotic limit of $m(z)$ as

$$m_c(z) = \frac{m_{3c}(z) + (c_1 + c_2 - 1)z}{c_2 z(1-z)} = \frac{1-c_2}{c_2} m_{2c}(z). \tag{3.27}$$

Through direct calculation, one can check easily that $m_{\alpha c}$ satisfy the following equations:

$$m_{1c} = -\frac{c_1}{m_{3c}}, \quad m_{2c} = -\frac{c_2}{m_{4c}}, \quad m_{3c}(z) - m_{4c}(z) = (1-z)(c_1 - c_2). \tag{3.28}$$

Finally, we introduce the function

$$\begin{aligned} h(z) :&= \frac{z^{-1/2} m_{3c}(z)}{1 + (1-z)m_{2c}(z)} = \frac{z^{-1/2} m_{4c}(z)}{1 + (1-z)m_{1c}(z)} \\ &= \frac{z^{1/2}}{2}\left[-z + (2 - c_1 - c_2) + \sqrt{(z-\lambda_-)(z-\lambda_+)}\right]. \end{aligned} \tag{3.29}$$

Now with the functions $m_{\alpha c}$ and h, we can define the matrix limit of $G(z)$ as

$$\Pi(z) := \begin{pmatrix} \begin{pmatrix} c_1^{-1} m_{1c}(z) I_p & 0 \\ 0 & c_2^{-1} m_{2c}(z) I_q \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} m_{3c}(z) I_n & h(z) I_n \\ h(z) I_n & m_{4c}(z) I_n \end{pmatrix} \end{pmatrix}. \tag{3.30}$$

Given $z = E + \mathrm{i}\eta$, we define its distance (along the real axis) to the two edges as

$$\kappa \equiv \kappa_E := \min\{|E - \lambda_-|, |E - \lambda_+|\}. \tag{3.31}$$

We have the following lemma, which can be proved through direct calculations using (3.23)–(3.26).

Lemma 3.4. *Fix any constants $c, C > 0$. If (2.9) holds, then we have the following estimates.*

(1) For $z \in \mathbb{C}_+ \cap \{z : c \leqslant |z| \leqslant C\}$, we have

$$|m_{3c}(z)| \sim 1, \quad 0 \leqslant \operatorname{Im} m_{3c}(z) \sim \begin{cases} \eta/\sqrt{\kappa + \eta}, & \text{if } E \notin [\lambda_-, \lambda_+] \\ \sqrt{\kappa + \eta}, & \text{if } E \in [\lambda_-, \lambda_+] \end{cases}. \tag{3.32}$$

(2) For $z, z_1, z_2 \in \mathbb{C}_+ \cap \{z : c \leqslant |z| \leqslant C\} \cap \{\operatorname{Re} z > \lambda_+\}$, we have

$$|m_{3c}(z) - m_{3c}(\lambda_+)| \sim |z - \lambda_+|^{1/2}, \quad |m_{3c}'(z)| \sim |z - \lambda_+|^{-1/2}, \tag{3.33}$$

and

$$|m_{3c}(z_1) - m_{3c}(z_2)| \sim \frac{|z_1 - z_2|}{\max_{i=1,2} |z_i - \lambda_+|^{1/2}}. \tag{3.34}$$

The above estimates also hold for m_{1c}, m_{2c}, m_{4c} and m_c. Finally, (3.33), (3.34) *and the first estimate in* (3.32) *hold for $h(z)$.*

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

Definition 3.5 (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 denote*

$$\mathcal{A}_{\mathbf{v}\mathbf{w}} := \langle \mathbf{v}, \mathcal{A}\mathbf{w} \rangle, \quad \mathcal{A}_{\mathbf{v}\mathfrak{a}} := \langle \mathbf{v}, \mathcal{A}\mathbf{e}_{\mathfrak{a}} \rangle, \quad \mathcal{A}_{\mathfrak{a}\mathbf{w}} := \langle \mathbf{e}_{\mathfrak{a}}, \mathcal{A}\mathbf{w} \rangle, \tag{3.35}$$

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 $\mathbf{v}^*$ being the conjugate transpose of $\mathbf{v}$. 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 $\begin{pmatrix} \mathbf{v} \\ \mathbf{0}_{q+2n} \end{pmatrix} \in \mathbb{C}^{\mathcal{I}}$.

We define the following spectral domains for the local laws of $G(z)$.

Definition 3.6 (Spectral domains)**.** *For any constant $\varepsilon > 0$, we define the domains*

$$S(\varepsilon) := \left\{ z = E + \mathrm{i}\eta : \varepsilon \leqslant E \leqslant 2, n^{-1+\varepsilon} \leqslant \eta \leqslant \varepsilon^{-1} \right\}. \tag{3.36}$$

and

$$S_{out}(\varepsilon) := S(\varepsilon) \cap \{z = E + \mathrm{i}\eta : E \notin [\lambda_-, \lambda_+], n\eta\sqrt{\kappa + \eta} \geqslant n^{\varepsilon}\}. \tag{3.37}$$

Correspondingly, we shall define the following two domains that are away from $z = 1$: for any fixed $\widetilde{\varepsilon} > 0$,

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

Now we are ready to state the local laws for $G(z)$. For $z = E + \mathrm{i}\eta$, we define the control parameter

$$\Psi(z) := \sqrt{\frac{\operatorname{Im} m_c(z)}{n\eta}} + \frac{1}{n\eta}. \tag{3.38}$$

Theorem 3.7 (Theorem 2.11 and Theorem 2.12 of [50])**.** *Suppose the assumptions of Lemma 2.7 hold. Then for any fixed $\widetilde{\varepsilon}, \varepsilon > 0$, the following estimates hold.*

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

$$|G_{\mathbf{u}\mathbf{v}}(z) - \Pi_{\mathbf{u}\mathbf{v}}(z)| \prec \phi_n + \Psi(z). \tag{3.39}$$

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

$$|m(z) - m_c(z)| \prec (n\eta)^{-1}. \tag{3.40}$$

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

$$|m(z) - m_c(z)| \prec \frac{1}{n(\kappa + \eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa + \eta}}, \tag{3.41}$$

for any $z \in \widetilde{S}_{out}(\varepsilon, \widetilde{\varepsilon})$. The estimates (3.40) *and* (3.41) *also hold for $m_\alpha(z) - m_{\alpha c}(z)$, $\alpha = 1, 2, 3, 4$.*

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

The averaged local law leads to the following rigidity of eigenvalues.

Theorem 3.8 (Theorem 2.5 of [50]). *Suppose the assumptions of Lemma 2.7 hold. For any fixed $\delta > 0$, the following rigidity estimate holds for all $1 \leqslant j \leqslant (1-\delta)q$:*

$$|\lambda_i - \gamma_i| \prec i^{-1/3} n^{-2/3}. \tag{3.42}$$

The anisotropic local law (3.39) and the rigidity estimate (3.42) together give the following delocalization estimates of eigenvectors.

Lemma 3.9 (Lemma 3.9 of [50]). *Suppose* (3.39) *and* (3.42) *hold. Then for any small constant $\delta > 0$ and deterministic unit vectors $\mathbf{u}_\alpha \in \mathbb{C}^{\mathcal{I}_\alpha}$, $\alpha = 1,2,3,4$, the following estimates hold:*

$$\max_{1 \leqslant k \leqslant (1-\delta)q} \left\{ \left| \langle \mathbf{u}_1, S_{xx}^{-1/2} \xi_k \rangle \right|^2 + \left| \langle \mathbf{u}_2, S_{yy}^{-1/2} \zeta_k \rangle \right|^2 \right\} \prec n^{-1}, \tag{3.43}$$

and

$$\max_{1 \leqslant k \leqslant (1-\delta)q} \left\{ \left| \langle \mathbf{u}_3, X^\top S_{xx}^{-1/2} \xi_k \rangle \right|^2 + \left| \langle \mathbf{u}_4, Y^\top S_{yy}^{-1/2} \zeta_k \rangle \right|^2 \right\} \prec n^{-1}. \tag{3.44}$$

Away from the support $[\lambda_-, \lambda_+]$, the anisotropic local law can be strengthened as follows.

Theorem 3.10 (Anisotropic local law outside of the spectrum). *Suppose the assumptions of Lemma 2.7 hold. Fix any constant $\varepsilon > 0$. Then for any*

$$z \in D_{out}(\varepsilon) := \left\{ E + \mathrm{i}\eta : \lambda_+ + n^{-2/3+\varepsilon} \leqslant E \leqslant 2, 0 \leqslant \eta \leqslant 1 \right\}, \tag{3.45}$$

and deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$, the following anisotropic local law holds:

$$|G_{\mathbf{uv}}(z) - \Pi_{\mathbf{uv}}(z)| \prec \phi_n + \sqrt{\frac{\operatorname{Im} m_c(z)}{n\eta}} \asymp \phi_n + n^{-1/2}(\kappa + \eta)^{-1/4}. \tag{3.46}$$

Proof. The second step of (3.46) follows from (3.32). Using (3.39) and $\kappa \geqslant n^{-2/3+\varepsilon}$, we have that (3.46) holds for $z \in S(\varepsilon) \cap D_{out}(\varepsilon)$ with $\eta \geqslant \eta_0 := n^{-1/2}\kappa^{1/4}$. Hence it remains to prove that for $z \in D_{out}(\varepsilon)$ with $0 \leqslant \eta \leqslant \eta_0$, we have

$$|G_{\mathbf{vv}}(X, z) - \Pi_{\mathbf{vv}}(z)| \prec \phi_n + n^{-1/2}\kappa^{-1/4}, \tag{3.47}$$

for any deterministic unit vector $\mathbf{v} \in \mathbb{C}^{\mathcal{I}}$. Note that (3.47) implies (3.46) by the polarization identity

$$\begin{aligned} \langle \mathbf{u}, \mathcal{M}\mathbf{v} \rangle = & \frac{1}{4}\langle (\mathbf{u}+\mathbf{v}), \mathcal{M}(\mathbf{u}+\mathbf{v}) \rangle - \frac{1}{4}\langle (\mathbf{u}-\mathbf{v}), \mathcal{M}(\mathbf{u}-\mathbf{v}) \rangle \\ & + \frac{\mathrm{i}}{4}\langle (\mathrm{i}\mathbf{u}+\mathbf{v}), \mathcal{M}(\mathrm{i}\mathbf{u}+\mathbf{v}) \rangle - \frac{\mathrm{i}}{4}\langle (\mathrm{i}\mathbf{u}-\mathbf{v}), \mathcal{M}(\mathrm{i}\mathbf{u}-\mathbf{v}) \rangle \end{aligned}$$

for any $\mathcal{I} \times \mathcal{I}$ matrix $\mathcal{M}$. Now fix any $z = E + \mathrm{i}\eta \in D_{out}(\varepsilon)$ with $\eta \leqslant \eta_0$. We denote $z_0 := E + \mathrm{i}\eta_0$. Since (3.47) holds at z_0, it suffices to prove that

$$\Pi_{\mathbf{vv}}(z) - \Pi_{\mathbf{vv}}(z_0) \prec n^{-1/2}\kappa^{-1/4}, \tag{3.48}$$

and

$$G_{\mathbf{vv}}(z) - G_{\mathbf{vv}}(z_0) \prec n^{-1/2}\kappa^{-1/4}. \tag{3.49}$$

The estimate (3.48) follows immediately from (3.34). It remains to show (3.49).

We write $\mathbf{v} = \left(\mathbf{v}_1^\top, \mathbf{v}_2^\top, \mathbf{v}_3^\top, \mathbf{v}_4^\top\right)^\top$, where $\mathbf{v}_\alpha \in \mathbb{C}^{\mathcal{I}_\alpha}$, $\alpha = 1,2,3,4$. We claim that

$$\left(\mathbf{v}_1^*, \mathbf{v}_2^*\right)\left[\mathcal{G}_L(z) - \mathcal{G}_L(z_0)\right]\begin{pmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{pmatrix} \prec n^{-1/2}\kappa^{-1/4}. \tag{3.50}$$

Using (3.13) and (3.14), and recalling that with high probability $E - \lambda_k \gtrsim 1$ for $k \geqslant (1-\delta)q$ by rigidity estimate (3.42), we obtain that

$$\begin{aligned} |\langle \mathbf{v}_1, (G(z) - G(z_0))\,\mathbf{v}_1\rangle| \prec & \sum_{k \leqslant (1-\delta)q} \frac{\eta_0 |\langle \mathbf{v}_1, S_{xx}^{-1/2}\xi_k\rangle|^2}{\left[(E-\lambda_k)^2 + \eta^2\right]^{1/2}\left[(E-\lambda_k)^2+\eta_0^2\right]^{1/2}} \\ & + \eta_0 \sum_{k > (1-\delta)q} |\langle \mathbf{v}_1, S_{xx}^{-1/2}\xi_k\rangle|^2. \end{aligned} \tag{3.51}$$

By (3.42), we have that for any $k \geqslant 1$, $E - \lambda_k \gtrsim \kappa \gg \eta_0$ with high probability. Then using (3.43) and (3.20), we can bound (3.51) by

$$\begin{aligned} |\langle \mathbf{v}_1, (G(z) - G(z_0))\,\mathbf{v}_1\rangle| &\prec \eta_0 + \frac{1}{q}\sum_{k=1}^q \frac{\eta_0}{(E-\lambda_k)^2+\eta_0^2} = \eta_0 + \operatorname{Im} m(z_0) \\ &\prec \eta_0 + \frac{1}{n\kappa} + \frac{1}{(n\eta_0)^2\sqrt{\kappa}} + \operatorname{Im} m_c(z_0) \\ &\lesssim \frac{1}{n\kappa} + \frac{1}{(n\eta_0)^2\sqrt{\kappa}} + \frac{\eta_0}{\sqrt{\kappa+\eta_0}} \lesssim n^{-1/2}\kappa^{-1/4}, \end{aligned}$$

where we used the spectral decomposition for $m(z)$ in the second step, (3.41) in the third step, and (3.32) in the fourth step. Similarly, we have

$$\begin{aligned} |\langle \mathbf{v}_1, (G(z) - G(z_0))\,\mathbf{v}_2\rangle| &\prec |1 - (zz_0^{-1})^{1/2}|\,|\langle \mathbf{v}_1, G(z_0)\,\mathbf{v}_2\rangle| + \sum_{k=1}^q \frac{\eta_0 |\langle \mathbf{v}_1, S_{xx}^{-1/2}\xi_k\rangle||\langle \mathbf{v}_2, S_{yy}^{-1/2}\zeta_k\rangle|}{|\lambda_k - z||\lambda_k - z_0|} \\ &\prec \eta_0 + \operatorname{Im} m(z_0) \prec n^{-1/2}\kappa^{-1/4}. \end{aligned}$$

With similar arguments for $\langle \mathbf{v}_2, (G(z) - G(z_0))\,\mathbf{v}_1\rangle$ and $|\langle \mathbf{v}_2, (G(z) - G(z_0))\,\mathbf{v}_2\rangle$, we can conclude (3.50). Finally, using (3.50), (3.15), (3.16) and Lemma 3.9, we can prove (3.49). We omit the details. □

The second moment of $\langle \mathbf{u}, (G(z) - \Pi(z))\mathbf{v}\rangle$ in fact satisfies a stronger bound. It will be used in the proof of Theorem 2.14.

Lemma 3.11. *Suppose the assumptions of Lemma 2.7 hold. Fix any constant $\varepsilon > 0$. For all $z \in S(\varepsilon)$ (recall (3.36)), we have that*

$$\mathbb{E}\,|G_{\mathbf{uv}}(z) - \Pi_{\mathbf{uv}}(z)|^2 \prec \Psi^2(z), \tag{3.52}$$

for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$. Moreover, for all $z \in D_{out}(\varepsilon)$ we have that

$$\mathbb{E}\,|G_{\mathbf{uv}}(z) - \Pi_{\mathbf{uv}}(z)|^2 \prec \frac{1}{n\sqrt{\kappa+\eta}}. \tag{3.53}$$

Proof. The estimate (3.52) has been proved in Lemma 3.10 of [50]. The estimate (3.53) can be proved using almost the same argument, where the only difference is that we replace the anisotropic local law (3.39) with the stronger one (3.46) in the proof. We omit the details. ☐

Finally, we state the local law for $G^b(z)$, which can be derived easily from the local law for $G(z)$ with the following Woodbury matrix identity: for $\mathcal{A}, S, \mathcal{B}, T$ of conformable dimensions,

$$(\mathcal{A} + S\mathcal{B}T)^{-1} = \mathcal{A}^{-1} - \mathcal{A}^{-1}S(\mathcal{B}^{-1} + T\mathcal{A}^{-1}S)^{-1}T\mathcal{A}^{-1}, \tag{3.54}$$

and the following approximate isometry condition of Z:

$$\|ZZ^\top - I_r\| \prec \psi_n. \tag{3.55}$$

The estimate (3.55) can be proved using standard large deviation estimate (cf. Lemma 3.8 of [22]). We define

$$\Pi^b(z) := \Pi(z) - \Pi(z)\begin{pmatrix}\mathbf{U}_b & 0\\ 0 & \mathbf{E}_b\end{pmatrix}\begin{pmatrix}\begin{pmatrix}0 & 0\\ 0 & c_2 m_{2c}^{-1}(z)\Sigma_b\mathcal{M}_b\end{pmatrix} & \begin{pmatrix}0 & 0\\ 0 & \mathcal{M}_b\end{pmatrix}\\ \begin{pmatrix}0 & 0\\ 0 & \mathcal{M}_b\end{pmatrix} & \begin{pmatrix}0 & 0\\ 0 & m_{4c}^{-1}(z)\Sigma_b\mathcal{M}_b\end{pmatrix}\end{pmatrix}\begin{pmatrix}\mathbf{U}_b^\top & 0\\ 0 & \mathbf{E}_b^\top\end{pmatrix}\Pi(z),$$

where

$$\mathcal{M}_b := \frac{\Sigma_b}{1+\Sigma_b^2}, \quad \mathbf{U}_b := \begin{pmatrix}0 & 0\\ 0 & (\mathbf{u}_1^b, \cdots, \mathbf{u}_r^b)\end{pmatrix}, \quad \mathbf{E}_b := \begin{pmatrix}0 & 0\\ 0 & (Z^\top\mathbf{v}_1^b, \cdots, Z^\top\mathbf{v}_r^b)\end{pmatrix}. \tag{3.56}$$

Lemma 3.12 (Local laws for G^b)**.** *Suppose the assumptions of Lemma 2.7 hold. Fix any constant $\varepsilon > 0$ and unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{\mathcal{I}}$ that are independent of X and Y. Then we have that for all $z \in S(\varepsilon)$,*

$$\left|G^b_{\mathbf{uv}}(z) - \Pi^b_{\mathbf{uv}}(z)\right| \prec \psi_n + \phi_n + \Psi(z), \tag{3.57}$$

and for all $z \in D_{out}(\varepsilon)$,

$$\left|G^b_{\mathbf{uv}}(z) - \Pi^b_{\mathbf{uv}}(z)\right| \prec \psi_n + \phi_n + n^{-1/2}(\kappa+\eta)^{-1/4}. \tag{3.58}$$

Moreover (3.57) *and* (2.20) *together imply that for any constant $\delta > 0$,*

$$\max_{1\leqslant k\leqslant(1-\delta)q}\left\{\left|\langle\mathbf{u}_1, S_{xx}^{-1/2}\xi_k^b\rangle\right|^2 + \left|\langle\mathbf{u}_2, (S_{yy}^b)^{-1/2}\zeta_k^b\rangle\right|^2\right\} \prec n^{-1}, \tag{3.59}$$

and

$$\max_{1\leqslant k\leqslant(1-\delta)q}\left\{\left|\langle\mathbf{u}_3, X^\top S_{xx}^{-1/2}\xi_k^b\rangle\right|^2 + \left|\langle\mathbf{u}_4, \mathcal{Y}^\top(S_{yy}^b)^{-1/2}\zeta_k^b\rangle\right|^2\right\} \prec n^{-1}, \tag{3.60}$$

where $\{\xi_k^b\}_{k=1}^p$ are $\{\zeta_k^b\}_{k=1}^q$ are the left and right singular vectors of $\mathcal{H}^b$, respectively, and $\mathbf{u}_\alpha \in \mathbb{C}^{\mathcal{I}_\alpha}$ are unit vectors independent of X and Y.

Proof. Using (3.54), we can write $G^b(z)$ in (3.10) as

$$G^b = G - G\begin{pmatrix}\mathbf{U}_b & 0\\ 0 & \mathbf{E}_b\end{pmatrix}\left[\begin{pmatrix}0 & \mathcal{D}_b^{-1}\\ \mathcal{D}_b^{-1} & 0\end{pmatrix} + \begin{pmatrix}\mathbf{U}_b^\top & 0\\ 0 & \mathbf{E}_b^\top\end{pmatrix}G\begin{pmatrix}\mathbf{U}_b & 0\\ 0 & \mathbf{E}_b\end{pmatrix}\right]^{-1}\begin{pmatrix}\mathbf{U}_b^\top & 0\\ 0 & \mathbf{E}_b^\top\end{pmatrix}G. \tag{3.61}$$

where $\mathcal{D}_b := \begin{pmatrix}0 & 0\\ 0 & \Sigma_b\end{pmatrix}$. Since $\mathcal{D}_b^{-1}$ is not well-defined, the above expression should be understood through

$$\left[\begin{pmatrix}0 & \mathcal{D}_b^{-1}\\ \mathcal{D}_b^{-1} & 0\end{pmatrix} + \begin{pmatrix}\mathbf{U}_b^\top & 0\\ 0 & \mathbf{E}_b^\top\end{pmatrix}G\begin{pmatrix}\mathbf{U}_b & 0\\ 0 & \mathbf{E}_b\end{pmatrix}\right]^{-1} := \left[1 + \begin{pmatrix}0 & \mathcal{D}_b\\ \mathcal{D}_b & 0\end{pmatrix}\begin{pmatrix}\mathbf{U}_b^\top & 0\\ 0 & \mathbf{E}_b^\top\end{pmatrix}G\begin{pmatrix}\mathbf{U}_b & 0\\ 0 & \mathbf{E}_b\end{pmatrix}\right]^{-1}\begin{pmatrix}0 & \mathcal{D}_b\\ \mathcal{D}_b & 0\end{pmatrix}.$$

Combining (3.61) with Theorem 3.7, Theorem 3.10 and (3.55), we can conclude (3.57) and (3.58). The estimates (3.59) and (3.60) follow from (3.57) and (2.20) as in Lemma 3.9, where the details can be found in the proof of Lemma 3.9 of [50]. ☐

4 Proof of Theorem 2.9

In this section, we prove Theorem 2.9 using the local laws, Theorems 3.7 and 3.10, and the eigenvalue rigidity estimate, Theorem 3.8. During the proof, in order to avoid some non-generic events, we assume that

$$\text{the entries } x_{ij},\ y_{ij} \text{ and } z_{ij} \text{ have continuous densities.} \tag{4.1}$$

It can be achieved by adding a small perturbation to X, Y and Z. For example, we can add to each matrix a small Gaussian matrix:

$$X \to X + \delta e^{-n} X_G, \quad Y \to Y + \delta e^{-n} Y_G, \quad Z \to Z + \delta e^{-n} Z_G.$$

These Gaussian components are negligible for our results and can be easily removed by taking $\delta \to 0$. Under (4.1), the matrices $\mathcal{X}\mathcal{X}^\top$, $\mathcal{Y}\mathcal{Y}^\top$, $XX^\top$ and $YY^\top$ are all non-singular almost surely under (4.1). Moreover, almost surely, $\lambda = 1$ is not in the spectrum of $\mathcal{C}_{XY}$ or $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$. Hence by (3.7), $0 < \lambda < 1$ is an eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ if and only if

$$\det \left[1 + \begin{pmatrix} 0 & \mathcal{D} \\ \mathcal{D} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{U}^\top & 0 \\ 0 & \mathbf{E}^\top \end{pmatrix} G(\lambda) \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{E} \end{pmatrix} \right] = 0. \tag{4.2}$$

Now for $\lambda \in D_{out}(\varepsilon)$ (recall (3.45)), using Theorem 3.10 and (3.55), we can write (4.2) as

$$\begin{aligned} 0 &= \det \left[1 + \begin{pmatrix} 0 & \mathcal{D} \\ \mathcal{D} & 0 \end{pmatrix} (\Pi_r(\lambda) + \mathcal{E}_r) \right] \\ &= \det \left[\begin{pmatrix} I_{2r} & \mathcal{D} \begin{pmatrix} m_{3c}(\lambda) I_r & h(\lambda)\mathcal{M}_r \\ h(\lambda)\mathcal{M}_r^\top & m_{4c}(\lambda) I_r \end{pmatrix} \\ \mathcal{D} \begin{pmatrix} c_1^{-1} m_{1c}(\lambda) I_r & 0 \\ 0 & c_2^{-1} m_{2c}(\lambda) I_r \end{pmatrix} & I_{2r} \end{pmatrix} + \begin{pmatrix} 0 & \mathcal{D} \\ \mathcal{D} & 0 \end{pmatrix} \mathcal{E}_{4r} \right]. \end{aligned} \tag{4.3}$$

Here $\mathcal{E}_{4r}$ is a $4r \times 4r$ random matrix satisfying

$$\|\mathcal{E}_{4r}\| \prec \psi_n + \phi_n + n^{-1/2}\kappa_\lambda^{-1/4}, \quad \text{with} \quad \kappa_\lambda = \min\{|\lambda - \lambda_-|, |\lambda - \lambda_+|\}, \tag{4.4}$$

$\mathcal{M}_r$ is an $r \times r$ orthogonal matrix with entries

$$(\mathcal{M}_r)_{ij} := (\mathbf{v}_i^a)^\top \mathbf{v}_j^b, \quad 1 \leqslant i, j \leqslant r,$$

and $\Pi_r(\lambda)$ is defined as

$$\Pi_r(\lambda) := \begin{pmatrix} \begin{pmatrix} c_1^{-1} m_{1c}(\lambda) I_r & 0 \\ 0 & c_2^{-1} m_{2c}(\lambda) I_r \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} m_{3c}(\lambda) I_r & h(\lambda)\mathcal{M}_r \\ h(\lambda)\mathcal{M}_r^\top & m_{4c}(\lambda) I_r \end{pmatrix} \end{pmatrix}.$$

Applying Schur complement formula and using (3.28), we obtain that (4.3) is equivalent to

$$\begin{aligned} &\det \left[\begin{pmatrix} I_r + \Sigma_a^2 & h(\lambda) m_{3c}^{-1}(\lambda) \Sigma_a^2 \mathcal{M}_r \\ h(\lambda) m_{4c}^{-1}(\lambda) \Sigma_b^2 \mathcal{M}_r^\top & I_r + \Sigma_b^2 \end{pmatrix} + \mathcal{D}\mathcal{E}_{2r} \right] = 0 \\ \Leftrightarrow\ &\det \left(\frac{m_{3c}(\lambda) m_{4c}(\lambda)}{h^2(\lambda)} I_r - \frac{\Sigma_a}{(1+\Sigma_a^2)^{1/2}} \mathcal{M}_r \frac{\Sigma_b^2}{1+\Sigma_b^2} \mathcal{M}_r^\top \frac{\Sigma_a}{(1+\Sigma_a^2)^{1/2}} + \mathcal{E}_r \right) = 0, \end{aligned} \tag{4.5}$$

where $\mathcal{E}_{2r}$ and $\mathcal{E}_r$ are $2r \times 2r$ and $r \times r$ random matrices, both of which satisfy the same bound as in (4.4). Note that the matrix

$$\frac{\Sigma_a}{(1+\Sigma_a^2)^{1/2}}\mathcal{M}_r\frac{\Sigma_b^2}{1+\Sigma_b^2}\mathcal{M}_r^\top\frac{\Sigma_a}{(1+\Sigma_a^2)^{1/2}}$$

is the PCC matrix $(1+AA^\top)^{-1/2}AB^\top(1+BB^\top)^{-1}BA^\top(1+AA^\top)^{-1/2}$ in the basis of $\mathbf{u}_i^a$, $1 \leqslant i \leqslant r$. Thus its eigenvalues are exactly the squares of the PCC's, $t_1, t_2, \cdots, t_r$ (recall (2.14)). Thus after a change of basis, (4.5) reduces to

$$\det\left(\frac{m_{3c}(\lambda)m_{4c}(\lambda)}{h^2(\lambda)}I_r - \operatorname{diag}(t_1, \cdots, t_r) + \mathcal{E}_r'(\lambda)\right) = 0, \tag{4.6}$$

where $\mathcal{E}_r'$ also satisfies the bound as in (4.4).

Next we show that if $\mathcal{E}_r' = 0$, then solving equation (4.6) gives the classical locations θ_i defined in (2.15). Using (3.25), (3.26) and (3.29), we can calculate that

$$\begin{aligned} f_c(z) := \frac{m_{3c}(z)m_{4c}(z)}{h^2(z)} &= z[1+(1-z)m_{1c}(z)][1+(1-z)m_{2c}(z)] \\ &= \frac{z-(c_1+c_2-2c_1c_2)+\sqrt{(z-\lambda_-)(z-\lambda_+)}}{2(1-c_1)(1-c_2)}. \end{aligned}$$

We can find the inverse function of $f_c(z)$ for $z \notin [\lambda_-, \lambda_+]$ as

$$g_c(\xi) := \xi\left(1-c_1+c_1\xi^{-1}\right)\left(1-c_2+c_2\xi^{-1}\right).$$

Note that $f_c(\lambda)$ is monotonically increasing in λ for $\lambda > \lambda_+$, so the function $f_c(\lambda) - t_i = 0$ has a solution in (λ_+, ∞) if and only if (recall (1.3))

$$f_c(\lambda_+) < t_i \quad \Leftrightarrow \quad t_c < t_i. \tag{4.7}$$

If (4.7) holds, the classical location of the outlier corresponding to t_i is $\theta_i = g_c(t_i)$, which explains (2.15).

With direct calculation, one can verify the following simple estimates on f_c and g_c.

Lemma 4.1. *Fix a large constant $C > 0$. Let $z, z_1, z_2 \in \mathbb{D} := \{z \in \mathbb{C} : \lambda_+ < \operatorname{Re} z < C, 0 < \operatorname{Im} z \leqslant C\}$ and $\xi, \xi_1, \xi_2 \in f_c(\mathbb{D})$. Then the following estimates hold:*

$$|f_c(z) - f_c(\lambda_+)| \sim |z-\lambda_+|^{1/2}, \quad |f_c'(z)| \sim |z-\lambda_+|^{-1/2}, \tag{4.8}$$

$$|g_c(\xi) - \lambda_+| \sim |\xi - t_c|^2, \quad |g_c'(\xi)| \sim |\xi - t_c|, \tag{4.9}$$

and

$$|f_c(z_1) - f_c(z_2)| \sim \frac{|z_1-z_2|}{\max_{i=1,2}|z_i-\lambda_+|^{1/2}}, \quad |g_c(\xi_1)-g_c(\xi_2)| \sim |\xi_1-\xi_2| \cdot \max_{i=1,2}|\xi_i - t_c|. \tag{4.10}$$

The estimate (4.8) *also holds for z with $\lambda_- + c \leqslant \operatorname{Re} z \leqslant \lambda_+$ and $0 < \operatorname{Im} z \leqslant c^{-1}$ for any constant $c > 0$.*

For the proof of Theorem 2.9, we record the following eigenvalues interlacing result:

$$\widetilde{\lambda}_i \in [\lambda_{i+2r}, \lambda_{i-2r}], \tag{4.11}$$

where we adopt the convention that $\lambda_i = 1$ if $i < 1$ and $\lambda_i = 0$ if $i > q$. For the reader's convenience, we briefly describe why (4.11) holds. We first consider a 1-dimensional perturbation:

$$X_1 := X + \mathbf{u}_1\mathbf{v}_1^\top, \quad \mathbf{u}_1 \in \mathbb{R}^p, \quad \mathbf{v}_1 \in \mathbb{R}^n.$$

Then it is easy to see that $\mathcal{P}_X := X^\top(XX^\top)^{-1}X$ is a projection onto the subspace $\mathcal{W}$ spanned by the rows of X. Similarly, $\mathcal{P}_{X_1} := X_1^\top(X_1X_1^\top)^{-1}X_1$ is a projection onto the subspace $\mathcal{W}_1$ spanned by the rows of X_1. Moreover, $\mathcal{W}$ and $\mathcal{W}_1$ differ at most by a 1-dimensional subspace. Hence by Cauchy interlacing, we have

$$\lambda_i\left(\mathcal{P}_{X_1}\mathcal{P}_Y\mathcal{P}_{X_1}\right) \in [\lambda_{i+1}(\mathcal{P}_X\mathcal{P}_Y\mathcal{P}_X), \lambda_{i-1}(\mathcal{P}_X\mathcal{P}_Y\mathcal{P}_X)], \quad \text{where} \quad \mathcal{P}_Y := Y^\top\frac{1}{YY^\top}Y.$$

Notice that $\mathcal{P}_X\mathcal{P}_Y\mathcal{P}_X$ (resp. $\mathcal{P}_{X_1}\mathcal{P}_Y\mathcal{P}_{X_1}$) have the same nonzero eigenvalues as $\mathcal{C}_{XY}$ (resp. $\mathcal{C}_{X_1Y}$): if $\mathbf{u}$ is an eigenvector of $\mathcal{C}_{XY}$ with eigenvalue λ, then $X^\top(XX^\top)^{-1/2}\,\mathbf{u}$ is an eigenvector of $\mathcal{P}_X\mathcal{P}_Y\mathcal{P}_X$ with the same eigenvalue. Thus we get

$$\lambda_i\left(\mathcal{C}_{X_1Y}\right) \in [\lambda_{i+1}(\mathcal{C}_{XY}), \lambda_{i-1}(\mathcal{C}_{XY})].$$

Repeating this estimate r times for the rank-r perturbation $\mathcal{X}$, we get

$$\lambda_i\left(\mathcal{C}^a_{\mathcal{X}Y}\right) \in [\lambda_{i+r}(\mathcal{C}_{XY}), \lambda_{i-r}(\mathcal{C}_{XY})],$$

where $\mathcal{C}^a_{\mathcal{X}Y}$ is defined by replacing X with $\mathcal{X}$ in $\mathcal{C}_{XY}$. Obviously, the same argument works for the rank-r perturbation of Y, which leads to (4.11).

With (4.6) and (4.11), the rest of the proof for Theorem 2.9 is similar to the ones in [12, Section 4] and [36, Section 6], but these references have only considered the cases with small support $\phi_n \prec n^{-1/2}$. We need to adapt their proofs to our setting with larger ϕ_n and ψ_n.

Proof of Theorem 2.9. For simplicity of presentation, in this proof we abbreviate $\phi_n + \psi_n$ as ϕ_n because these two factors always appear together. By Theorems 3.7, 3.8 and 3.10, for any fixed $\varepsilon > 0$, we can choose a high-probability event Ξ on which the following estimates hold:

$$\mathbf{1}(\Xi)\left\| \begin{pmatrix} \mathbf{U}^\top & 0 \\ 0 & \mathbf{E}^\top \end{pmatrix} G(z) \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{E} \end{pmatrix} - \Pi_r(z) \right\| \leqslant n^{\varepsilon/2}\left(\phi_n + \Psi(z)\right), \quad \text{for } z \in S(\varepsilon), \tag{4.12}$$

$$\mathbf{1}(\Xi)\left\| \begin{pmatrix} \mathbf{U}^\top & 0 \\ 0 & \mathbf{E}^\top \end{pmatrix} G(z) \begin{pmatrix} \mathbf{U} & 0 \\ 0 & \mathbf{E} \end{pmatrix} - \Pi_r(z) \right\| \leqslant n^{\varepsilon/2}\left(\phi_n + n^{-1/2}\kappa^{-1/4}\right), \quad \text{for } z \in D_{out}(\varepsilon), \tag{4.13}$$

and for a fixed large integer ϖ,

$$\mathbf{1}(\Xi)\,|\lambda_i - \lambda_+| \leqslant n^{-2/3+\varepsilon}, \quad \text{for } 1 \leqslant i \leqslant \varpi + 2r. \tag{4.14}$$

We remark that the randomness of X and Y only comes into play to ensure that Ξ holds with high probability. The rest of the proof will be entirely deterministic once restricted to Ξ. In the following proof, we assume that ε is a sufficiently small constant.

We now define the index sets

$$\mathcal{O}_\varepsilon := \left\{ i : t_i - t_c \geqslant n^\varepsilon\phi_n + n^{-1/3+\varepsilon} \right\}. \tag{4.15}$$

Since the constant ε is arbitrary, in order to prove (2.23) and (2.24), it suffices to show that for some constant $C > 0$,

$$\mathbf{1}(\Xi)\left|\widetilde{\lambda}_i - \theta_i\right| \leqslant Cn^{2\varepsilon}\left(\phi_n\Delta_i + n^{-1/2}\Delta_i^{1/2}\right), \tag{4.16}$$

for all $i \in \mathcal{O}_{4\varepsilon}$, and

$$-n^{-2/3+\varepsilon} \leqslant \mathbf{1}(\Xi)\left(\widetilde{\lambda}_i - \lambda_+\right) \leqslant Cn^{8\varepsilon}\phi_n^2 + Cn^{-2/3+12\varepsilon} \tag{4.17}$$

for all $i \in \{1, \cdots, \varpi\}\backslash\mathcal{O}_{4\varepsilon}$.

Step 1: Our first step is to prove that on Ξ, there are no eigenvalues outside the neighborhoods of θ_i's. For each $1 \leqslant i \leqslant r_+$, we define the permissible intervals

$$\mathrm{I}_i \equiv \mathrm{I}_i(\mathbf{t}) := \left[\theta_i - n^{\varepsilon}\left(\phi_n \Delta_i + n^{-1/2}\Delta_i^{1/2}\right), \theta_i + n^{\varepsilon}\left(\phi_n \Delta_i + n^{-1/2}\Delta_i^{1/2}\right)\right], \tag{4.18}$$

where $\mathbf{t}$ represents the canonical correlation coefficients $\mathbf{t} := (t_1, t_2, \cdots, t_r)$. We then define

$$\mathrm{I} \equiv \mathrm{I}(\mathbf{t}) := \mathrm{I}_0 \cup \Big(\bigcup_{i \in \mathcal{O}_\varepsilon} \mathrm{I}_i(\mathbf{t})\Big), \quad \mathrm{I}_0 \equiv \mathrm{I}_0 := \left[0, \lambda_+ + n^{2\varepsilon}\phi_n^2 + n^{-2/3+3\varepsilon}\right]. \tag{4.19}$$

We claim the following result.

Lemma 4.2. *The complement of* $\mathrm{I}(\mathbf{t})$ *contains no eigenvalues of* $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$.

Proof. The main idea is similar to the ones for [36, Proposition 6.5] and [17, Lemma S.4.2]. It suffices to show that for any $1 \leqslant i \leqslant r$, if $x \notin \mathrm{I}(\mathbf{t})$, then

$$|f_c(x) - t_i| \geqslant c\left(n^{\varepsilon}\phi_n + n^{-1/2+\varepsilon}\kappa_x^{-1/4}\right) \tag{4.20}$$

for some constant $c > 0$. Thus (4.6) cannot hold on Ξ by (4.13).

For $x \notin \mathrm{I}_0$, using (4.10) we get

$$f_c(x) - t_c = f_c(x) - f_c(\lambda_+) \geqslant c\kappa_x^{1/2} \geqslant c'\left(n^{\varepsilon}\phi_n + n^{-1/2+\varepsilon}\kappa_x^{-1/4}\right).$$

for some constants $c, c' > 0$. This concludes (4.20) for $i \geqslant r_+$ using $t_i \leqslant t_c + n^{-1/3} + \phi_n$.

Next for the case $1 \leqslant i \leqslant r_+$, we take any $x \notin \mathrm{I}_0 \cup \mathrm{I}_i(\mathbf{t})$. We first assume that there exists a constant $\widetilde{c} > 0$ such that $\theta_i \notin [x - \widetilde{c}\kappa_x, x + \widetilde{c}\kappa_x]$. Then since f_c is monotonically increasing on $(\lambda_+, +\infty)$, we have that

$$|f_c(x) - t_i| = |f_c(x) - f_c(\theta_i)| \geqslant |f_c(x) - f_c(x \pm \widetilde{c}\kappa_x)| \geqslant c\kappa_x^{1/2} \geqslant c'\left(n^{\varepsilon}\phi_n + n^{-1/2+\varepsilon}\kappa_x^{-1/4}\right),$$

for some constants $c, c' > 0$, where we used (4.10) in the third step. On the other hand, suppose $\theta_i \in [x - \widetilde{c}\kappa_x, x + \widetilde{c}\kappa_x]$, in which case we have that $\theta_i - \lambda_+ \sim \kappa_x$. With (4.9), we have $\kappa_x \sim \theta_i - \lambda_+ \sim \Delta_i^2$. Then using (4.10) and the definition of $\mathrm{I}_i(\mathbf{t})$, we get that for $x \notin \mathrm{I}_i(\mathbf{t})$,

$$|f_c(x) - t_i| = |f_c(x) - f_c(\theta_i)| \geqslant c\Delta_i^{-1}\left(n^{\varepsilon}\phi_n\Delta_i + n^{-1/2+\varepsilon}\Delta_i^{1/2}\right) \geqslant c'\left(n^{\varepsilon}\phi_n + n^{-1/2+\varepsilon}\kappa_x^{-1/4}\right),$$

for some constants $c, c' > 0$. This concludes (4.20) and hence Lemma 4.2. $\square$

Step 2: Before giving the general proof, for heuristics we consider an easy case where the t_i's are *independent of* n and satisfy that

$$t_1 > t_2 > \cdots > t_{r_+} > \lambda_+. \tag{4.21}$$

We claim that each $\mathrm{I}_i(\mathbf{t})$, $1 \leqslant i \leqslant r_+$, contains precisely one eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$. Fix any $1 \leqslant i \leqslant r_+$ and choose a small n-independent positively oriented closed contour $\mathbf{\Gamma} \subset \mathbb{C}/[0, \lambda_+]$ that encloses θ_i but no other points of the set $\{\theta_i : 1 \leqslant i \leqslant r_+\}$. Define two functions

$$f_1(z) := \det\left(f_c(z)I_r - \operatorname{diag}(t_1, \cdots, t_r)\right), \quad f_2(z) = \det\left(f_c(z)I_r - \operatorname{diag}(t_1, \cdots, t_r) + \mathcal{E}_r'(z)\right). \tag{4.22}$$

The functions f_1, f_2 are holomorphic on and inside $\mathbf{\Gamma}$ when n is sufficiently large, because $\mathbf{\Gamma}$ does not enclose any pole of $G(z)$ by (4.14). Moreover, by the construction of $\mathbf{\Gamma}$, the function f_1 has precisely one zero inside $\mathbf{\Gamma}$ at θ_i. By (4.13), we have

$$\min_{z \in \mathbf{\Gamma}} |f_1(z)| \gtrsim 1, \quad \max_{z \in \mathbf{\Gamma}} |f_1(z) - f_2(z)| = \mathrm{o}(1).$$

The claim then follows from Rouché's theorem.

Step 3: In order to extend the argument in Step 2 to an arbitrary n-dependent configuration $\mathbf{t}_n$, we need to deal with the case where some of the intervals I_i and I_j, $i \neq j$, have non-empty overlaps. For any $\varepsilon > 0$, we denote $r_\varepsilon := |\mathcal{O}_\varepsilon|$. In this step, we prove the following claim for the first $r_{4\varepsilon}$ eigenvalues.

Claim 4.3. *On event Ξ, the estimate* (4.16) *holds for $i \in \mathcal{O}_{4\varepsilon}$.*

Proof. Let $\mathcal{B}$ denote the finest partition of $\{1, \cdots, r_+\}$ in the sense that i and j belong to the same block of $\mathcal{B}$ whenever $\mathrm{I}_i \cap \mathrm{I}_j \neq \varnothing$. We now fix any $1 \leqslant i \leqslant r_{4\varepsilon}$, and denote by B_i the block of $\mathcal{B}$ that contains i. Our first task is to estimate $\theta_{j-1} - \theta_j$ for $j, j-1 \in B_i$. We claim that there exists a constant $C_1 > 0$ such that

$$\theta_{j-1} - \theta_j \leqslant C_1 \left(n^\varepsilon \phi_n \Delta_j + n^{-1/2+\varepsilon} \Delta_j^{1/2} \right), \quad \text{if } \ j \in B_i \text{ and } j-1 \in B_i. \tag{4.23}$$

First we assume that $j \in \mathcal{O}_{3\varepsilon}$. We pick any $x \in \mathrm{I}_j \cap \mathrm{I}_{j-1}$ such that $\theta_j \leqslant x \leqslant \theta_{j-1}$. Then using (4.8) and (4.10) we obtain that

$$|f_c(x) - t_j| = |f_c(x) - f_c(\theta_j)| \leqslant C \left(n^\varepsilon \phi_n + n^{-1/2+\varepsilon} \Delta_j^{-1/2} \right) \ll \Delta_j,$$

using $\Delta_j \geqslant n^{3\varepsilon} \phi_n + n^{-1/3+3\varepsilon}$ for $j \in \mathcal{O}_{3\varepsilon}$. Thus we get that $|f_c(x) - t_c| = (1 + \mathrm{o}(1)) \Delta_j$. Similarly, we can show that $|f_c(x) - t_c| = (1 + \mathrm{o}(1)) \Delta_{j-1}$. This gives (4.23) due to the choice of x and the definition of I_j and I_{j-1}. In addition we also get that

$$\Delta_j = (1 + \mathrm{o}(1)) \Delta_{j-1}, \quad \text{if } \ j \in B_i \text{ and } j-1 \in B_i. \tag{4.24}$$

It remains to verify that $j \in \mathcal{O}_{3\varepsilon}$ for all $j \in B_i$. Let j_0 be the smallest integer such that $\theta_{j_0} \notin B_i$. Since $|B_i| \leqslant r$, by (4.23) we have that

$$\theta_{j_0-1} > \theta_i - C \left(n^\varepsilon \phi_n \Delta_i + n^{-1/2+\varepsilon} \Delta_i^{1/2} \right)$$

for some constant $C > 0$. Then using $i \in \mathcal{O}_{4\varepsilon}$, $j_0 \notin \mathcal{O}_{3\varepsilon}$ and (4.9), we can check that

$$\theta_{j_0-1} - \theta_{j_0} \gg \left(n^\varepsilon \phi_n \Delta_{j_0-1} + n^{-1/2+\varepsilon} \Delta_{j_0-1}^{1/2} \right) + \left(n^\varepsilon \phi_n \Delta_{j_0} + n^{-1/2+\varepsilon} \Delta_{j_0}^{1/2} \right),$$

which contradicts the definition of B_i. This concludes (4.23).

Now with (4.23), (4.24) and $|B_i| \leqslant r$, we obtain that

$$d_i := \operatorname{diam} \left(\bigcup_{j \in B_i} \mathrm{I}_j \right) \leqslant C_r \left(n^\varepsilon \phi_n \Delta_i + n^{-1/2+\varepsilon} \Delta_i^{1/2} \right). \tag{4.25}$$

for some constant $C_r > 0$ depending on r and C_1 only. On the other hand, by (4.9) we have that

$$\theta_i - \lambda_+ - d_i \geqslant c\Delta_i^2 - C_r \left(n^\varepsilon \phi_n \Delta_i + n^{-1/2+\varepsilon} \Delta_i^{1/2} \right) \gg n^{2\varepsilon} \phi_n^2 + n^{-2/3+3\varepsilon},$$

where we used $\Delta_i \geqslant n^{4\varepsilon} \phi_n + n^{-1/3+4\varepsilon}$ for $i \in \mathcal{O}_{4\varepsilon}$ in the second step. Hence there is a gap between the right edge of I_0 and the left edge of $\bigcup_{j \in B_i} \mathrm{I}_j$.

Let x_i and y_i be the left and right end points of the interval $\bigcup_{j \in B_i} \mathrm{I}_j$. Then we pick the contour

$$\mathbf{\Gamma}_i := \{z = x_i + \mathrm{i}\eta : -d_i \leqslant \eta \leqslant d_i\} \cup \{z = y_i + \mathrm{i}\eta : -d_i \leqslant \eta \leqslant d_i\} \cup \{z = E \pm \mathrm{i}d_i : x_i \leqslant E \leqslant y_i\},$$

which lies in the half plane on the right of I_0, and only includes θ_j's with $j \in B_i$, but no other points of the set $\{\theta_i : 1 \leqslant i \leqslant r_+\}$. We again consider the functions in (4.22). We know that $f_1(z)$ has exactly $|B_i|$ eigenvalues at θ_j, $j \in B_i$. Moreover, with the arguments in Lemma 4.2, one can show that

$$\|\mathcal{E}(z)\| = \mathrm{o}(1) \quad \text{for} \quad z \in \mathbf{\Gamma}_i, \quad \mathcal{E}(z) := [f_c(z)I_r - \mathrm{diag}\,(t_1, \cdots, t_r)]^{-1}\,\mathcal{E}_r'(z).$$

Thus we have

$$|f_2(z) - f_1(z)| = |f_1(z)|\,|\det(1 + \mathcal{E}(z)) - 1| < |f_1(z)| \quad \text{for} \quad z \in \mathbf{\Gamma}_i.$$

Then by Rouché's theorem, $f_2(z)$ has exactly $|B_i|$ eigenvalues in $\bigcup_{j\in B_i} \mathrm{I}_j$. Together with Lemma 4.2 and a simple eigenvalues counting argument, we get that $\widetilde{\lambda}_i \in \bigcup_{j \in B_i} \mathrm{I}_j$, and hence

$$|\widetilde{\lambda}_i - \theta_i| \leqslant d_i, \quad i \in \mathcal{O}_{4\varepsilon}.$$

This concludes Claim 4.3 by (4.25). □

Step 4: Finally, we consider the eigenvalues $\widetilde{\lambda}_i$ with $i \notin \mathcal{O}_{4\varepsilon}$. First by (4.14) and (4.11), we have that

$$\lambda_i(s) \geqslant \lambda_+ - n^{-2/3+\varepsilon}, \quad i \leqslant \varpi. \tag{4.26}$$

For the upper bound, we consider the intervals as in (4.18) and

$$\widehat{\mathrm{I}}_0 := \left[0, \lambda_+ + \widetilde{C}_1\left(n^{8\varepsilon}\phi_n^2 + n^{-2/3+12\varepsilon}\right)\right],$$

for a large constant $\widetilde{C}_1 > 0$. Then we define a partition $\mathcal{B}$ as in Step 3, where B_0 is the block of $\mathcal{B}$ that contains i. With the same arguments as in the proof of Claim 4.3, we can prove that

$$\widehat{\mathrm{I}}_0 \cup \left(\bigcup_{j\in B_0} \mathrm{I}_j\right) \subset \left[0, \lambda_+ + C_2\left(n^{8\varepsilon}\phi_n^2 + n^{-2/3+12\varepsilon}\right)\right] \tag{4.27}$$

for some constant $C_2 > 0$. Moreover, for any $j \notin B_0$, we have that $j \in \mathcal{O}_{4\varepsilon}$ by (4.9) as long as $\widetilde{C}_1$ is chosen large enough. Thus with Lemma 4.2, the result of Step 3 and a simple eigenvalues counting argument, we get that

$$\widetilde{\lambda}_i \in \widehat{\mathrm{I}}_0 \cup \left(\bigcup_{j\in B_0} \mathrm{I}_j\right), \quad i \notin \mathcal{O}_{4\varepsilon}.$$

This concludes (4.17) by (4.27), and hence completes the proof of Theorem 2.9. □

5 Proof of Theorem 2.11

As in Section 4, from equation (3.9), we can derive a similar equation as (4.6). More precisely, suppose λ is not an eigenvalue of $\mathcal{C}^b_{\mathcal{X}\mathcal{Y}}$ and the following local law holds for $G^b(\lambda)$:

$$\begin{pmatrix} \mathbf{U}_a^\top & 0 \\ 0 & \mathbf{E}_a^\top \end{pmatrix} G^b(\lambda) \begin{pmatrix} \mathbf{U}_a & 0 \\ 0 & \mathbf{E}_a \end{pmatrix} - \Pi_r^b(\lambda) = \mathrm{O}\left(\Phi_n\right) \quad \text{with high probability,}$$

where Φ_n is a deterministic parameter satisfying $0 < \Phi_n \leqslant n^{-\varepsilon}$ for a constant $\varepsilon > 0$, and

$$\Pi_r^b(\lambda) := \begin{pmatrix} \begin{pmatrix} c_1^{-1} m_{1c}(\lambda) I_r & 0 \\ 0 & 0 \end{pmatrix} & 0 \\ 0 & \begin{pmatrix} m_{3c}(\lambda) I_r - \frac{h^2(\lambda)}{m_{4c}(\lambda)} \mathcal{M}_r \frac{\Sigma_b^2}{1+\Sigma_b^2} \mathcal{M}_r^\top & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix}.$$

Then λ is an eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ if and only if

$$\det\left(f_c(\lambda)I_r - \operatorname{diag}(t_1, \cdots, t_r) + \mathcal{E}_r(\lambda)\right) = 0, \tag{5.1}$$

with $\mathcal{E}_r$ satisfying $\|\mathcal{E}_r\| \lesssim \psi_n + \Phi_n$ with high probability. Moreover, similar to (4.11), we have the following eigenvalues interlacing,

$$\widetilde{\lambda}_i \in [\lambda^b_{i+r}, \lambda^b_{i-r}], \tag{5.2}$$

where we adopt the convention that $\lambda^b_i = 1$ if $i < 1$ and $\lambda^b_i = 0$ if $i > q$. This is the main reason why we use $\mathcal{C}^b_{X\mathcal{Y}}$ and $G^b(z)$ instead of $\mathcal{C}_{XY}$ and $G(z)$ for the proof of Theorem 2.11—the interlacing result (4.11) is not strong enough by using a rank-$(2r)$ perturbation.

Proof of Theorem 2.11. Again, in this proof, we abbreviate $\phi_n + \psi_n$ as ϕ_n. By (2.20), Theorem 2.9, Lemma 3.3, (3.55) and Lemma 3.12, for any small constant $\varepsilon > 0$ and fixed integer $\varpi \in \mathbb{N}$, we can choose a high-probability event Ξ on which the following estimates hold:

$$\mathbf{1}(\Xi)\left|\lambda^b_i - \lambda_+\right| \leqslant n^{-2/3+\varepsilon/2}, \quad \text{for } 1 \leqslant i \leqslant \varpi; \tag{5.3}$$

$$\mathbf{1}(\Xi)|\lambda^b_i - \gamma_i| \leqslant i^{-1/3} n^{-2/3+\varepsilon/2}, \quad \text{for } 1 \leqslant i \leqslant (1-\delta)q; \tag{5.4}$$

$$\mathbf{1}(\Xi)\left(\widetilde{\lambda}_i - \theta_i\right) \leqslant n^{\varepsilon}\phi_n\Delta_i + n^{-1/2+\varepsilon}\Delta_i^{1/2}, \quad \text{for } i \leqslant r_+; \tag{5.5}$$

$$-\mathbf{1}(\Xi)n^{-2/3+\varepsilon/2} \leqslant \mathbf{1}(\Xi)\left(\widetilde{\lambda}_i - \lambda_+\right) \leqslant n^{\varepsilon/2}\phi_n^2 + n^{-2/3+\varepsilon/2}, \quad \text{for } r_+ + 1 \leqslant i \leqslant \varpi; \tag{5.6}$$

$$c_0 \leqslant \min\{\lambda_p(S_{xx}), \lambda_q(S^b_{yy})\} \leqslant \max\{\lambda_1(S_{xx}), \lambda_1(S^b_{yy})\} \leqslant c_0^{-1}; \tag{5.7}$$

$$\mathbf{1}(\Xi)\|ZZ^\top - I_r\| \leqslant n^{\varepsilon/20}\phi_n; \tag{5.8}$$

$$\mathbf{1}(\Xi)|m^b(z) - m_c(z)| \leqslant n^{\varepsilon/4}\left(\phi_n + \Psi(z)\right), \quad \text{for } z \in S(\varepsilon); \tag{5.9}$$

$$\mathbf{1}(\Xi)\left\|\begin{pmatrix}\mathbf{U}_a^\top & 0\\ 0 & \mathbf{E}_a^\top\end{pmatrix} G^b(z)\begin{pmatrix}\mathbf{U}_a & 0\\ 0 & \mathbf{E}_a\end{pmatrix} - \Pi^b_r(z)\right\| \leqslant n^{\varepsilon/2}\left(\phi_n + \Psi(z)\right), \quad \text{for } z \in S(\varepsilon); \tag{5.10}$$

$$\mathbf{1}(\Xi)\left\|\begin{pmatrix}\mathbf{U}_a^\top & 0\\ 0 & \mathbf{E}_a^\top\end{pmatrix} G^b(z)\begin{pmatrix}\mathbf{U}_a & 0\\ 0 & \mathbf{E}_a\end{pmatrix} - \Pi^b_r(z)\right\| \leqslant n^{\varepsilon/2}\left(\phi_n + n^{-1/2}\kappa^{-1/4}\right), \quad \text{for } z \in D_{out}(\varepsilon); \tag{5.11}$$

$$\mathbf{1}(\Xi)\max_{1\leqslant k\leqslant(1-\delta)q}\left\{\left|\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi^b_k\rangle\right|^2 + \left|\langle \mathbf{u}_2, (S^b_{yy})^{-1/2}\zeta^b_k\rangle\right|^2\right\} \leqslant n^{-1+\varepsilon/20}, \quad \mathbf{u}_1 \in \mathbb{C}^{\mathcal{I}_1},\ \mathbf{u}_2 \in \mathbb{C}^{\mathcal{I}_2}; \tag{5.12}$$

$$\mathbf{1}(\Xi)\max_{1\leqslant k\leqslant(1-\delta)q}\left\{\left|\langle \mathbf{u}_3, X^\top S_{xx}^{-1/2}\xi^b_k\rangle\right|^2 + \left|\langle \mathbf{u}_4, \mathcal{Y}^\top(S^b_{yy})^{-1/2}\zeta^b_k\rangle\right|^2\right\} \leqslant n^{-1+\varepsilon/20}, \quad \mathbf{u}_3 \in \mathbb{C}^{\mathcal{I}_3},\ \mathbf{u}_4 \in \mathbb{C}^{\mathcal{I}_4}. \tag{5.13}$$

Here c_0 is a small enough constant, and the vectors $\mathbf{u}_\alpha$, $\alpha = 1, 2, 3, 4$, belong to a set of vectors Γ that is independent of X and Y, has cardinality $n^{\mathrm{O}(1)}$, and includes all the unit vectors that will be used in the proof. Again the randomness of X, Y and Z only comes into play to ensure that Ξ holds with high probability, and the rest of the proof will be entirely deterministic.

Step 1: As in the proof of Theorem 2.9, we first find a permissible region. For any i, we define the set

$$\Omega_i := \left\{x \in [\lambda^b_{i+r+1}, \lambda_+ + n^{2\varepsilon}\phi_n^2 + n^{-2/3+2\varepsilon}] : \operatorname{dist}\left(x, \operatorname{Spec}(\mathcal{C}^b_{X\mathcal{Y}})\right) > n^{-1+\varepsilon}\alpha_+^{-1}\right\}, \tag{5.14}$$

where $\operatorname{Spec}(\mathcal{C}^b_{X\mathcal{Y}})$ stands for the eigenvalue spectrum of $\mathcal{C}^b_{X\mathcal{Y}}$.

Lemma 5.1. *There exists a constant $C_1 > 0$ such that for $\alpha_+ \geqslant C_1 \left(n^{\varepsilon}\phi_n + n^{-1/3+\varepsilon}\right)$ and $i \leqslant n^{1-2\varepsilon}\alpha_+^3$, the set Ω_i contains no eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$.*

Proof. In the proof, we always use the following spectral parameters

$$\eta_x := n^{-1+\varepsilon}\alpha_+^{-1}, \quad z_x = x + \mathrm{i}\eta_x. \tag{5.15}$$

Suppose $x \in \Omega_i$. We first claim that for any deterministic unit vectors $\mathbf{u}, \mathbf{v} \in \Gamma$, we have

$$|G^b_{\mathbf{u}\,\mathbf{v}}(z_x) - G^b_{\mathbf{u}\,\mathbf{v}}(x)| \leqslant Cn^{\varepsilon/20}\,\mathrm{Im}\, m^b(z_x) + Cn^{\varepsilon/20}\eta_x, \quad x \in \Omega_i. \tag{5.16}$$

We use a similar argument as in the proof of Theorem 3.10. To illustrate the idea, for $\mathbf{v} = \left(\mathbf{v}_1^\top, \mathbf{v}_2^\top, \mathbf{v}_3^\top, \mathbf{v}_4^\top\right)^\top$ and $\mathbf{u} = \left(\mathbf{u}_1^\top, \mathbf{u}_2^\top, \mathbf{u}_3^\top, \mathbf{u}_4^\top\right)^\top$ with $\mathbf{u}_\alpha, \mathbf{v}_\alpha \in \mathbb{C}^{\mathcal{I}_\alpha}$, we calculate $G^b_{\mathbf{u}_1\,\mathbf{v}_1}(z_x) - G^b_{\mathbf{u}_1\,\mathbf{v}_1}(x)$ as an example. As in (3.51), we have

$$\begin{aligned}
|G^b_{\mathbf{u}_1\mathbf{v}_1}(z_x) - G^b_{\mathbf{u}_1\mathbf{v}_1}(x)| &\lesssim \sum_{k\leqslant(1-\delta)q} \frac{\eta_x|\langle \mathbf{v}_1, S_{xx}^{-1/2}\xi_k^b\rangle||\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_k^b\rangle|}{|\lambda_k^b - x|\left[(\lambda_k^b - x)^2 + \eta_x^2\right]^{1/2}} + \eta_x \sum_{k>(1-\delta)q} |\langle \mathbf{v}_1, S_{xx}^{-1/2}\xi_k^b\rangle||\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_k^b\rangle| \\
&\lesssim n^{-1+\varepsilon/20}\sum_{k=1}^{q} \frac{\eta_x}{(\lambda_k^b - x)^2 + \eta_x^2} + \eta_x \lesssim n^{\varepsilon/20}\,\mathrm{Im}\, m^b(z_x) + \eta_x,
\end{aligned}$$

where in the second step we used (5.7), (5.12) and $|\lambda_k^b - x| \geqslant \eta_x$ for $x \in \Omega_i$, and in the last step we used the spectral decomposition of $m^b(z_x)$. The proofs for the rest of the cases $G^b_{\mathbf{u}_\alpha\,\mathbf{v}_\beta}(z_x) - G^b_{\mathbf{u}_\alpha\,\mathbf{v}_\beta}(x)$, $\alpha, \beta = 1, 2, 3, 4$, are similar, so we omit the details.

Recall that $x \in \Omega_i$ is an eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ if and only if (5.1) holds, where $\mathcal{E}_r$ satisfy the following bound by (5.16), (5.9) and (5.10):

$$\|\mathcal{E}_r(x)\| \leqslant C\left(n^{\varepsilon/20}\,\mathrm{Im}\, m_c(z_x) + n^{\varepsilon/20}\eta_x + n^{\varepsilon/2}\phi_n + n^{\varepsilon/2}\Psi(z_x)\right)$$

for some constant $C > 0$. With (3.32) and the definition of $\Psi(z_x)$ in (3.38), we can further bound that

$$\|\mathcal{E}_r(x)\| \leqslant C'\left(n^{\varepsilon/2}\phi_n + n^{\varepsilon/2}\,\mathrm{Im}\, m_c(z_x) + \frac{n^{\varepsilon/2}}{n\eta_x}\right)$$

for some constant $C' > 0$. Now to prove the lemma, it suffices to show that for any $1 \leqslant j \leqslant r$,

$$|f_c(x) - t_j| > C'\left(n^{\varepsilon/2}\phi_n + n^{\varepsilon/2}\,\mathrm{Im}\, m_c(z_x) + \frac{n^{\varepsilon/2}}{n\eta_x}\right), \quad x \in \Omega_i. \tag{5.17}$$

Since $i \leqslant n^{1-2\varepsilon}\alpha_+^3$, by (5.4) we have

$$-\left(n^{2\varepsilon}\phi_n^2 + n^{-2/3+2\varepsilon}\right) \leqslant \lambda_+ - x \lesssim (i/n)^{2/3} + i^{-1/3}n^{-2/3+\varepsilon/2} \lesssim n^{-4\varepsilon/3}\alpha_+^2, \quad x \in \Omega_i, \tag{5.18}$$

where we also used $\gamma_i \sim (i/n)^{2/3}$ and $\alpha_+ \geqslant n^{-1/3+\varepsilon}$. Then by (4.8), we have

$$|f_c(x) - t_c| = |f_c(x) - f_c(\lambda_+)| \leqslant Cn^{-2\varepsilon/3}\alpha_+, \quad x \in \Omega_i \cap \{x : x \leqslant \lambda_+\}.$$

and

$$|f_c(x) - t_c| = |f_c(x) - f_c(\lambda_+)| \leqslant C\left(n^{\varepsilon}\phi_n + n^{-1/3+\varepsilon}\right), \quad x \in \Omega_i \cap \{x : x > \lambda_+\},$$

for some constant $C > 0$ that does not depend on C_1. Hence as long as C_1 is chosen large enough, we have

$$|f_c(x) - t_c| \leqslant \frac{1}{4}\alpha_+ \quad \Rightarrow \quad |f_c(x) - t_j| \geqslant \frac{3}{4}\alpha_+, \tag{5.19}$$

where we used the definition of α_+ in (2.22). On the other hand, with (3.32), (5.15) and (5.18) we can verify that

$$C'\left(n^{\varepsilon/2}\phi_n + n^{\varepsilon/2}\operatorname{Im} m_c(z_x) + \frac{n^{\varepsilon/2}}{n\eta_x}\right) \leqslant C''\left(n^{\varepsilon/2}\phi_n + n^{\varepsilon/2}\sqrt{\kappa_x + \eta_x} + n^{-\varepsilon/2}\alpha_+\right) \ll \alpha_+$$

for $x \in \Omega_i \cap \{x : x \leqslant \lambda_+\}$, and

$$C'\left(n^{\varepsilon/2}\phi_n + n^{\varepsilon/2}\operatorname{Im} m_c(z_x) + \frac{n^{\varepsilon/2}}{n\eta_x}\right) \leqslant C''\left(n^{\varepsilon/2}\phi_n + n^{\varepsilon/2}\frac{\eta_x}{\sqrt{\kappa_x + \eta_x}} + n^{-\varepsilon/2}\alpha_+\right) \ll \alpha_+$$

for $x \in \Omega_i \cap \{x : x > \lambda_+\}$. Together with (5.19), we see that (5.17) holds. This concludes the proof. □

Step 2: In this step, we perform a counting argument for a special case as in the following lemma. We postpone its proof until we finish the proof of Theorem 2.11.

Lemma 5.2. *Given $0 \leqslant r_+ \leqslant r$, we choose a matrix $A \equiv A(0)$ of rank r_+ such that the eigenvalues configuration $\mathbf{t} \equiv \mathbf{t}(0) := (t_1, t_2, \cdots, t_r)$ of the PCC matrix satisfies that*

$$(t_{r_+} - t_c) \wedge (t_c - t_{r_+ + 1}) \wedge \min_{1 \leqslant i \leqslant r_+ - 1} (t_i - t_{i+1}) \gtrsim 1. \tag{5.20}$$

Then for $i \leqslant n^{1-4\varepsilon}\alpha_+^3(0)$, we have

$$|\widetilde{\lambda}_{i+r_+} - \lambda_i^b| \leqslant n^{-1+2\varepsilon}\alpha_+^{-1}(0), \tag{5.21}$$

where $\alpha_+(0)$ is defined as in (2.22) for $\mathbf{t}(0)$. (The meaning of the argument 0 will be clear in Step 3 below.)

Step 3: In this step we employ a continuity argument as in [36, Section 6.5] and [17, Section S.4.2]. We choose a continuous (n-dependent) path $A(s)$ for $0 \leqslant s \leqslant 1$, such that $A(1) = A$ is the matrix in Theorem 2.11, and $A(0)$ gives an eigenvalues configuration $\mathbf{t}(0)$ satisfying (5.20). Correspondingly, we have continuous paths of the configurations $\mathbf{t}(s)$ and the sample eigenvalues $\{\widetilde{\lambda}_i(s)\}_{i=1}^n$. We can choose $A(s)$ such that

$$\inf_{s \in [0,1]} \alpha_+(s) \gtrsim \alpha_+ \equiv \alpha_+(1),$$

where $\alpha_+(s)$ is defined as in (2.22) for the eigenvalues configuration $\mathbf{t}(s)$.

In this step we consider the case where $\alpha_+ \geqslant C_1\left(n^{\varepsilon}\phi_n + n^{-1/3+\varepsilon}\right)$ and $i \leqslant n^{1-4\varepsilon}\alpha_+^3$. Without loss of generality, we rename $\alpha_+ := \inf_{s\in[0,1]} \alpha_+(s)$. Define

$$\widetilde{I}_0 := \left\{x \in [0, \lambda_+ + n^{2\varepsilon}\phi_n^2 + n^{-2/3+2\varepsilon}] : \operatorname{dist}\left(x, \operatorname{Spec}(\mathcal{C}^b_{\mathcal{X}\mathcal{Y}})\right) \leqslant n^{-1+\varepsilon}\alpha_+^{-1}\right\}.$$

Note that $\widetilde{I}_0$ is a union of connected intervals. Due to the interlacing (5.2), we have

$$\lambda^b_{i+r} \leqslant \widetilde{\lambda}_i(s) \leqslant \lambda^b_{i-r}, \quad s \in [0,1]. \tag{5.22}$$

By Lemma 5.1 and Lemma 5.2, we know

$$|\widetilde{\lambda}_{i+r_+}(0) - \lambda_i^b| \leqslant n^{-1+2\varepsilon}\alpha_+^{-1},$$

and

$$\mathrm{dist}\left(\widetilde{\lambda}_{i+r_+}(s), \mathrm{Spec}(\mathcal{C}^b_{\mathcal{X}\mathcal{Y}})\right) \leqslant n^{-1+\varepsilon}\alpha_+^{-1}, \quad s\in[0,1]. \tag{5.23}$$

In addition, by continuity of eigenvalues with respect to s, we know that $\widetilde{\lambda}_{i+r_+}(s)$ is in the same connected component of $\widetilde{I}_0$ as $\widetilde{\lambda}_{i+r_+}(0)$. For any i, let B_i be the set of j such that λ_i^b and λ_j^b are in the same connected component of $\widetilde{I}_0$. Then we conclude that

$$\widetilde{\lambda}_{i+r_+}(s) \in \bigcup_{j\in B_i: |i+r_+-j|\leqslant r}\left[\lambda_j^b - n^{-1+2\varepsilon}\alpha_+^{-1}, \lambda_j^b + n^{-1+2\varepsilon}\alpha_+^{-1}\right].$$

This gives that

$$\left|\widetilde{\lambda}_{i+r_+}(s) - \lambda_i^b\right| \leqslant 2rn^{-1+2\varepsilon}\alpha_+^{-1}, \quad s\in[0,1]. \tag{5.24}$$

Step 4: Finally we consider the cases $\alpha_+ < C_1\left(n^\varepsilon\phi_n + n^{-1/3+\varepsilon}\right)$, or $i > n^{1-4\varepsilon}\alpha_+^3$. Suppose first that $\alpha_+ < C_1\left(n^\varepsilon\phi_n + n^{-1/3+\varepsilon}\right)$. Then by the assumption of Theorem 2.11, if ε is small enough such that $\varepsilon < \varepsilon_0$, we must have

$$\phi_n \leqslant n^{-1/3}, \quad \text{and} \quad \alpha_+ \lesssim n^{-1/3+\varepsilon}. \tag{5.25}$$

Now using (5.25), (5.2), (5.4) and (5.6), we find that

$$|\widetilde{\lambda}_{i+r_+} - \lambda_i^b| \lesssim n^{-2/3+\varepsilon} \lesssim n^{-1+2\varepsilon}\alpha_+^{-1}.$$

On the other hand, suppose $i > n^{1-4\varepsilon}\alpha_+^3$. If $i \leqslant r$, then we have $\alpha_+ \lesssim n^{-1/3+4\varepsilon/3}$, and with the same argument as above, we get

$$|\widetilde{\lambda}_{i+r_+} - \lambda_i^b| \leqslant Cn^{-2/3+\varepsilon} \leqslant n^{-1+3\varepsilon}\alpha_+^{-1}.$$

Otherwise, using (5.2) and (5.4) we get

$$|\widetilde{\lambda}_{i+r_+} - \lambda_i^b| \leqslant Ci^{-1/3}n^{-2/3+\varepsilon/2} \leqslant n^{-1+2\varepsilon}\alpha_+^{-1}.$$

Combining the above three estimates with (5.24), we conclude (2.25), since $\varepsilon > 0$ can be arbitrarily small. □

For the proof of Lemma 5.2, we shall use an argument that extends the one in the proof of [36, Proposition 6.8]. However, the proof in [7, Section 7] may also work, where the authors proved essentially the same result but only for $i \leqslant \varpi$ for some fixed integer ϖ.

Proof of Lemma 5.2. Note that in this lemma, we have $\alpha_+ \equiv \alpha_+(0) \sim 1$. In the first step, we group together the eigenvalues λ_i that are close to each other. More precisely, let $\mathcal{B} = \{B_k\}$ be the finest partition of $\{1, \cdots, q\}$ such that $i < j$ belong to the same block of $\mathcal{B}$ if

$$|\lambda_i^b - \lambda_j^b| \leqslant n^{-1+7\varepsilon/6}\alpha_+^{-1}.$$

Note that each block B_k of $\mathcal{B}$ consists of a sequence of consecutive integers. We order the blocks in the descending order, that is, if $k < l$ then $\lambda_{i_k}^b > \lambda_{i_l}^b$ for all $i_k \in B_k$ and $i_l \in B_l$.

We first derive a bound on the sizes of the blocks. We define k^* such that $n_0 := \lceil n^{1-4\varepsilon}\alpha_+^3\rceil \in B_{k^*}$. For any $k \leqslant k^*$, we take $i < j$ such that i and j both belong to the block B_k. Then by (5.2) and (5.4), we have that for some constants $c, C > 0$,

$$c\left[\left(\frac{j}{n}\right)^{2/3} - \left(\frac{i}{n}\right)^{2/3}\right] - Ci^{-1/3}n^{-2/3+\varepsilon/2} \leqslant \lambda_i^b - \lambda_j^b \leqslant C(j-i)n^{-1+7\varepsilon/6}\alpha_+^{-1}.$$

Now using $j^{2/3} - i^{2/3} \geqslant j^{-1/3}(j-i)$, we obtain that

$$\left(j^{-1/3} - Cn^{-1/3+7\varepsilon/6}\alpha_+^{-1}\right)(j-i) \leqslant Ci^{-1/3}n^{\varepsilon/2}.$$

From this estimate we conclude that if i and j satisfy

$$1 \leqslant i \leqslant j \leqslant n^{1-15\varepsilon/4}, \tag{5.26}$$

then

$$j - i \leqslant C(j/i)^{1/3}n^{\varepsilon/2}. \tag{5.27}$$

Now we claim that

$$|B_k| \leqslant Cn^{3\varepsilon/4} \quad \text{for} \quad k = 1, \cdots, k^*, \tag{5.28}$$

and for any given $i_k \in B_k$,

$$|\lambda_i^b - \gamma_{i_k}| \leqslant i^{-1/3}n^{-2/3+\varepsilon} \quad \text{for all} \quad i \in B_k. \tag{5.29}$$

To prove (5.28) and (5.29), we denote $\alpha_k := \max_{i\in B_k} i$ and $\beta_k := \min_{i\in B_k} i$. If $i \in B_k$ satisfies $i \geqslant \alpha_k/2$, then (5.27) gives that $\alpha_k - i \leqslant Cn^{\varepsilon/2}$, with which we obtain that

$$|\gamma_i - \gamma_{\alpha_k}| \leqslant Ci^{-1/3}n^{-2/3+\varepsilon/2}.$$

On the other hand, if $i \in B_k$ satisfies $i \leqslant \alpha_k/2$, then (5.27) gives that $\alpha_k - i \leqslant \alpha_k \leqslant Cn^{3\varepsilon/4}$. Thus we get

$$|\gamma_i - \gamma_{\alpha_k}| \leqslant |\gamma_1 - \gamma_{\alpha_k}| \leqslant Cn^{-2/3+\varepsilon/2} \leqslant Ci^{-1/3}n^{-2/3+3\varepsilon/4}.$$

Together with (5.4), we obtain that

$$|\lambda_i^b - \gamma_{i_k}| \leqslant |\lambda_i^b - \gamma_i| + |\gamma_i - \gamma_{\alpha_k}| + |\gamma_{\alpha_k} - \gamma_{i_k}| \leqslant Ci^{-1/3}n^{-2/3+3\varepsilon/4} \leqslant i^{-1/3}n^{-2/3+\varepsilon}.$$

From the above proof, we see that (5.28) and (5.29) as long as (5.26) holds. We still need to prove (5.26) for $i, j \in B_{k^*}$. In fact, if there is $j \in B_{k^*}$ such that $j \geqslant n^{1-15\varepsilon/4}$, then we can find $j' \in B_{k^*}$ such that $n^{\varepsilon} \leqslant j' - n_0 \leqslant 2n^{\varepsilon}$, which contradicts (5.27) and (5.28).

We are now ready to complete the proof. For any $1 \leqslant k \leqslant k^*$, we denote

$$\mathfrak{a}_k := \min_{i\in B_k} \lambda_i^b = \lambda_{\alpha_k}^b, \quad \mathfrak{b}_k := \max_{i\in B_k} \lambda_i^b = \lambda_{\beta_k}^b. \tag{5.30}$$

We introduce a continuous path as

$$x_s^k := (1-s)\left(\mathfrak{a}_k - \delta_n/3\right) + s\left(\mathfrak{b}_k + \delta_n/3\right), \quad s \in [0,1], \tag{5.31}$$

where $\delta_n := n^{-1+7\varepsilon/6}\alpha_+^{-1}$. The interval $[x_0^k, x_1^k]$ contains precisely the eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}^b$ that are in B_k, and the endpoints x_0^k and x_1^k are at distances at least $\delta_n/3$ from any eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}^b$. Then we have the following proposition. We postpone its proof until we finish the proof of Lemma 5.2.

Proposition 5.3. *Almost surely, there are at least $|B_k|$ eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ in $[x_0^k, x_1^k]$.*

Here "almost surely" in the statement is due to the assumption (4.1): in the proof we discard a measure zero non-generic event. We postpone its proof until we complete the proof of Lemma 5.2.

We now use a standard interlacing argument to show that $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ has at most $|B_k|$ eigenvalues in $[x_0^k, x_1^k]$. By (5.2), there are at most $|B_1| + r_+$ eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ in $[x_0^1, \infty)$ (recall that the rank of $A(0)$ is r_+). Moreover, with the argument in Section 4, we can prove that (5.5) holds in the case $A \equiv A(0)$, i.e. there are exactly r_+ outliers. Then together with Proposition 5.3, we obtain that there are exactly $|B_1|$ eigenvalues of

$\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ in $[x_0^1, x_1^1]$. Repeating this argument, we can show that $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ has exact $|B_k|$ eigenvalues in $[x_0^k, x_1^k]$ for all $k = 2, \cdots, k^*$. Moreover, using (5.28) we find that for any $i \in B_k$,

$$\sup\left\{|x - \lambda_i^b| : x \in [x_0^k, x_1^k]\right\} \leqslant Cn^{3\varepsilon/4}\left(n^{-1+7\varepsilon/6}\alpha_+^{-1}\right) \leqslant n^{-1+2\varepsilon}\alpha_+^{-1},$$

which concludes Lemma 5.2. □

Finally we give the proof of Proposition 5.3.

Proof of Proposition 5.3. For the spectral decomposition of $R^b(z)$ (which takes a similar form as (3.13)), we define

$$P_{B_k}R^b(z) := \sum_{l \in B_k} \frac{1}{\lambda_l^b - z}\begin{pmatrix} \xi_l^b(\xi_l^b)^\top & -z^{-1/2}(\lambda_l^b)^{1/2}\xi_l^b(\zeta_l^b)^\top \\ -z^{-1/2}(\lambda_l^b)^{1/2}\zeta_l^b(\xi_l^b)^\top & \zeta_l^b(\zeta_l^b)^\top \end{pmatrix}, \tag{5.32}$$

and $P_{B_k^c}R^b(z) := R^b(z) - P_{B_k}R^b(z)$. We define $P_{B_k^c}G^b$ by replacing R with $P_{B_k^c}R^b$, and Y with $\mathcal{Y}$ in (3.14), (3.15) and (3.16). Then we define $P_{B_k}G^b(z) := G^b(z) - P_{B_k^c}G^b(z)$. Let $x \in [x_0^k, x_1^k]$ and denote $z_x := x + \mathrm{i}\eta_x$ with $\eta_x := n^{-1+7\varepsilon/6}\alpha_+^{-1}$. We claim that

$$\left\|\begin{pmatrix} \mathbf{U}_a^\top & 0 \\ 0 & \mathbf{E}_a^\top \end{pmatrix}\left[P_{B_k^c}G^b(z_x) - P_{B_k^c}G^b(x)\right]\begin{pmatrix} \mathbf{U}_a & 0 \\ 0 & \mathbf{E}_a \end{pmatrix}\right\| \lesssim Cn^{\varepsilon/20}\,\mathrm{Im}\, m^b(z_x) + Cn^{\varepsilon/20}\eta_x. \tag{5.33}$$

The proof is very similar to the one for (5.16). For example, for deterministic unit vectors $\mathbf{u}_1, \mathbf{v}_1 \in \mathcal{I}_1$, using (3.14), (5.7) and (5.12) we get

$$\begin{aligned} &\left|P_{B_k^c}G^b_{\mathbf{u}_1\mathbf{v}_1}(z_x) - P_{B_k^c}G^b_{\mathbf{u}_1\mathbf{v}_1}(x)\right| \\ &\lesssim \sum_{l \notin B_k, l \leqslant (1-\delta)q} \frac{\eta_x|\langle \mathbf{v}_1, S_{xx}^{-1/2}\xi_l^b\rangle||\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_l^b\rangle|}{|\lambda_l^b - x|\left[(\lambda_l^b - x)^2 + \eta_x^2\right]^{1/2}} + \eta_x \sum_{l > (1-\delta)q} |\langle \mathbf{v}_1, S_{xx}^{-1/2}\xi_l^b\rangle||\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_l^b\rangle| \\ &\lesssim n^{-1+\varepsilon/20}\sum_{l=1}^{q} \frac{\eta_x}{(\lambda_l^b - x)^2 + \eta_x^2} + \eta_x \lesssim n^{\varepsilon/20}\,\mathrm{Im}\, m^b(z_x) + \eta_x, \end{aligned}$$

where in the second step we used $|\lambda_l^b - x| \gtrsim \eta_x$ for $l \notin B_k$. The proofs for the rest of the cases $(G^b_{\mathbf{u}_\alpha\mathbf{v}_\beta}(z_x) - G^b_{\mathbf{u}_\alpha\mathbf{v}_\beta}(x))$, $\alpha, \beta = 1, 2, 3, 4$, are similar, so we omit the details.

Then we claim that

$$\left|P_{B_k}G^b_{\mathbf{u}\mathbf{v}}(z_x)\right| + \left|P_{B_k}G^b_{\mathbf{u}\mathbf{v}}(x_0^k)\right| \leqslant n^{-\varepsilon/3}. \tag{5.34}$$

For example, for the z_x term we have

$$\left|\sum_{l \in B_k} \frac{\langle \mathbf{u}_1, S_{xx}^{-1/2}\xi_l^b\rangle\langle \xi_l^b S_{xx}^{-1/2}, \mathbf{v}_1\rangle}{\lambda_l^b - z_x}\right| \leqslant Cn^{3\varepsilon/4}\eta_x^{-1}n^{-1+\varepsilon/20} \ll n^{-\varepsilon/3},$$

where we used (5.12) and (5.28). The proofs for the rest of the blocks $P_{B_k}G^b_{\mathbf{u}_\alpha\mathbf{v}_\beta}(z_x)$, $\alpha, \beta = 1, 2, 3, 4$, are similar. For $z = x_0^k$, the proof is the same except that we need to use $|\lambda_l^b - x_0^k| \gtrsim n^{-1+7\varepsilon/6}\alpha_+^{-1}$ for $l \in B_k$.

We remove the zero singular values of A and redefine that

$$\Sigma_a := \mathrm{diag}(a_1, \cdots, a_{r_+}), \quad U_a = \left(\mathbf{u}_1^a, \cdots, \mathbf{u}_{r_+}^a\right), \quad E_a = \left(Z^\top\mathbf{v}_1^a, \cdots, Z^\top\mathbf{v}_{r_+}^a\right).$$

Then inspired by (3.9), for $x \notin \mathrm{spec}(\mathcal{C}^b_{X\mathcal{Y}})$ we define

$$\mathcal{M}(x) := \begin{pmatrix} 0 & \Sigma_a^{-1} \\ \Sigma_a^{-1} & 0 \end{pmatrix} + \begin{pmatrix} U_a^\top & 0 \\ 0 & E_a^\top \end{pmatrix}\begin{pmatrix} \mathcal{G}_1^b(x) & \mathcal{G}_{13}^b(x) \\ \mathcal{G}_{31}^b(x) & \mathcal{G}_3^b(x) \end{pmatrix}\begin{pmatrix} U_a & 0 \\ 0 & E_a \end{pmatrix},$$

where we recall that $\mathcal{G}^b_\alpha$ is the $\mathcal{I}_\alpha \times \mathcal{I}_\alpha$ block of G^b (cf. Definition 3.2), and we use $\mathcal{G}^b_{\alpha\beta}$ to denote the $\mathcal{I}_\alpha \times \mathcal{I}_\beta$ block of G^b. Then we know that almost surely, $x \in \mathrm{Spec}(\mathcal{C}_{\mathcal{X}\mathcal{Y}}) \setminus \mathrm{Spec}(\mathcal{C}^b_{X\mathcal{Y}})$ if and only if $\mathcal{M}(x)$ is singular. To simplify the notation, we shall denote

$$\left[G^b(z)\right]_{1,3} := \begin{pmatrix} \mathcal{G}^b_1(z) & \mathcal{G}^b_{13}(z) \\ \mathcal{G}^b_{31}(z) & \mathcal{G}^b_3(z) \end{pmatrix}.$$

Now using (5.9), (5.10), (5.33) and (5.34), we obtain that

$$\begin{aligned}\mathcal{M}(x) &= \begin{pmatrix} 0 & \Sigma_a^{-1} \\ \Sigma_a^{-1} & 0 \end{pmatrix} + \begin{pmatrix} U_a^\top & 0 \\ 0 & E_a^\top \end{pmatrix} \left[P_{B_k} G^b(x) + P_{B_k^c}\left(G^b(x) - G^b(z_x)\right) + G^b(z_x) - P_{B_k} G^b(z_x)\right]_{1,3} \begin{pmatrix} U_a & 0 \\ 0 & E_a \end{pmatrix} \\ &= \begin{pmatrix} 0 & \Sigma_a^{-1} \\ \Sigma_a^{-1} & 0 \end{pmatrix} + \begin{pmatrix} U_a^\top & 0 \\ 0 & E_a^\top \end{pmatrix} \left[P_{B_k} G^b(x)\right]_{1,3} \begin{pmatrix} U_a & 0 \\ 0 & E_a \end{pmatrix} + \left[\Pi^b_r(z_x)\right]_{1,3} + R_0(x) \\ &= \begin{pmatrix} 0 & \Sigma_a^{-1} \\ \Sigma_a^{-1} & 0 \end{pmatrix} + \begin{pmatrix} U_a^\top & 0 \\ 0 & E_a^\top \end{pmatrix} \left[P_{B_k} G^b(x)\right]_{1,3} \begin{pmatrix} U_a & 0 \\ 0 & E_a \end{pmatrix} + \left[\Pi^b_r(\lambda_+)\right]_{1,3} + R(x), \end{aligned} \tag{5.35}$$

where

$$\left[\Pi^b_r(z)\right]_{1,3} := \begin{pmatrix} c_1^{-1} m_{1c}(z) I_r & 0 \\ 0 & m_{3c}(z) I_r - \frac{h^2(z)}{m_{4c}(z)} \mathcal{M}_r \frac{\Sigma_b^2}{1+\Sigma_b^2} \mathcal{M}_r^\top \end{pmatrix},$$

and R_0 and R_1 are two matrices satisfying that

$$\|R_0(x)\| = \mathrm{O}\left(n^{\varepsilon/20}\eta_x + n^{\varepsilon/20}\,\mathrm{Im}\, m_c(z_x) + n^{\varepsilon/2}\Psi(z_x) + n^{\varepsilon/2}\phi_n + n^{-\varepsilon/3}\right) = \mathrm{O}\left(n^{-\varepsilon/3}\right),$$

and

$$\|R(x)\| = \left\|R_0(x) + \mathrm{O}(\sqrt{\kappa_x + \eta_x})\right\| = \mathrm{O}\left(n^{-\varepsilon/3}\right).$$

In bounding the $\|R_0(x)\|$ and $\|R(x)\|$, we also used Lemma 3.4, (3.38) and that

$$\kappa_x \leqslant \max\left\{|\lambda_+ - x_0^k|, |\lambda_+ - x_1^k|\right\} \lesssim (n^{1-15\varepsilon/4}/n)^{2/3} + n^{-2/3+\varepsilon} + n^{-1+7\varepsilon/6}\alpha_+^{-1} \ll n^{-\varepsilon/3},$$

where in the second step we used (5.26), (5.29) and the definitions in (5.31). Moreover, $R(x)$ is real symmetric (because all the other terms in (5.35) are real symmetric), and continuous in x on the extended real line $\overline{\mathbb{R}}$.

The rest of the proof follows from a continuity argument, which is exactly the same as the proof in [36, Section 6.4]. Instead of writing down all the details, we shall give an almost rigorous argument to show how equation (5.35) implies Proposition 5.3.

First, we claim that $\mathcal{M}(x)$ has some negative singular values when $x = x_0^k$. By (5.34), (5.35) gives that

$$\mathcal{M}(x_0^k) = \begin{pmatrix} 0 & \Sigma_a^{-1} \\ \Sigma_a^{-1} & 0 \end{pmatrix} + \left[\Pi^b_r(\lambda_+)\right]_{1,3} + \mathrm{O}(n^{-\varepsilon/3}).$$

Let $\mathbf{v}_i$ be an eigenvector of

$$\frac{\Sigma_a}{(1+\Sigma_a^2)^{1/2}} \mathcal{M}_r \frac{\Sigma_b^2}{1+\Sigma_b^2} \mathcal{M}_r^\top \frac{\Sigma_a}{(1+\Sigma_a^2)^{1/2}}$$

with eigenvalue t_i. Then for $\mathbf{u}_i =: \begin{pmatrix} m_{3c}(\lambda_+)(1+\Sigma_a^2)^{-1/2}\,\mathbf{v}_i \\ \Sigma_a (1+\Sigma_a^2)^{-1/2}\,\mathbf{v}_i \end{pmatrix}$, we can verify that

$$\mathbf{u}_i^\top \mathcal{M}(x_0^k)\, \mathbf{u}_i = \frac{h^2(\lambda_+)}{m_{4c}(\lambda_+)} \left(f_c(\lambda_+) - t_i\right) \|\, \mathbf{v}_i \,\|^2 + \mathrm{O}(n^{-e/3}) \|\, \mathbf{v}_i \,\|^2 < 0,$$

where we used $m_{4c}(\lambda_+) > 0$, $t_i > t_c = f_c(\lambda_+)$ and $t_i - t_c \sim 1$ for $1 \leqslant i \leqslant r_+$.

Next we claim that for $l \in B_k$, almost surely, $\mathcal{M}(x)$ is positive definite when $x \to \lambda_l^b-$ and negative definite when $x \to \lambda_l^b+$. To see why this holds, we pick any unit vector $\mathbf{v} = \left(\mathbf{v}_1^\top, \mathbf{v}_2^\top\right)^\top$, $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^{r_+}$, and denote $\widetilde{\mathbf{v}} = \left(\mathbf{v}_1^\top, \mathbf{0}_{r_+}, \mathbf{v}_2^\top, \mathbf{0}_{r_+}\right)^\top$. Then

$$\begin{aligned}\mathbf{v}^\top \mathcal{M}(x)\, \mathbf{v} &= \mathrm{O}(1) + \widetilde{\mathbf{v}}^\top \begin{pmatrix} \mathbf{U}_a^\top & 0 \\ 0 & \mathbf{E}_a^\top \end{pmatrix} P_{B_k} G^b(x) \begin{pmatrix} \mathbf{U}_a & 0 \\ 0 & \mathbf{E}_a \end{pmatrix} \widetilde{\mathbf{v}} \\ &= \mathrm{O}(1) + \widetilde{\mathbf{w}}^\top \begin{pmatrix} P_{B_k}\mathcal{G}_L^b(x) & -P_{B_k}\mathcal{G}_L^b(x) \\ -P_{B_k}\mathcal{G}_L^b(x) & P_{B_k}\mathcal{G}_L^b(x) \end{pmatrix} \widetilde{\mathbf{w}} = \mathrm{O}(1) + \mathbf{w}^\top P_{B_k} R^b(x)\, \mathbf{w}, \end{aligned} \tag{5.36}$$

where in the second step we used similar identities for G^b as in (3.15) and (3.16) with

$$\widetilde{\mathbf{w}} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{pmatrix} := \begin{pmatrix} I_{p+q} & 0 \\ 0 & \begin{pmatrix} X & 0 \\ 0 & \mathcal{Y} \end{pmatrix} \begin{pmatrix} xI_n & x^{1/2} I_n \\ x^{1/2} I_n & xI_n \end{pmatrix} \end{pmatrix} \begin{pmatrix} \mathbf{U}_a & 0 \\ 0 & \mathbf{E}_a \end{pmatrix} \widetilde{\mathbf{v}}, \quad \mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^{p+q},$$

and in the third step we used (3.14) with

$$\mathbf{w} := \begin{pmatrix} S_{xx}^{-1/2} & 0 \\ 0 & (S_{yy}^b)^{-1/2} \end{pmatrix} (\mathbf{w}_1 - \mathbf{w}_2),$$

Using the spectral decomposition (5.32), we can write

$$P_{B_k} R^b(x) = \frac{1}{2} \sum_{l \in B_k} \left[\frac{x^{-1/2}}{(\lambda_l^b)^{1/2} - x^{1/2}} \begin{pmatrix} \xi_l^b \\ -\zeta_l^b \end{pmatrix} \left((\xi_l^b)^\top, -(\zeta_l^b)^\top\right) - \frac{x^{-1/2}}{(\lambda_l^b)^{1/2} + x^{1/2}} \begin{pmatrix} \xi_l^b \\ \zeta_l^b \end{pmatrix} \left((\xi_l^b)^\top, (\zeta_l^b)^\top\right) \right]. \tag{5.37}$$

In particular, it has poles at $x = \lambda_l^b$ for $l \in B_k$. Combining (5.36) and (5.37), we conclude the claim.

With the above two claims and a simple continuity argument, we see that there exists $x \in (x_0^k, \lambda_{\alpha_k}^b)$ (recall (5.30)) such that $\mathcal{M}(x)$ is singular. Moreover, for any $l, l-1 \in B_k$, there exists $x \in (\lambda_l^b, \lambda_{l-1}^b)$ such that $\mathcal{M}(x)$ is singular. This gives at least $|B_k|$ eigenvalues of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}$ inside $[x_0^k, x_1^k]$ and hence completes the proof. Writing down a rigorous continuity argument involves discussion on some non-generic measure zero events, and we refer the reader to [36, Section 6.4] for more details. □

6 Proof of Theorem 2.14

For the proof of Theorem 2.14, we adopt a similar argument as the one for Theorem 2.7 in [50]. However, our setting here is much more complicated. First, we introduce a cutoff on the matrix entries of X and Y at the level $n^{-\varepsilon}$ for a sufficiently small constant $\varepsilon > 0$:

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

Using (2.34), we can check with integration by parts that for any small constant $\delta > 0$,

$$\alpha_n^{(1)} \leqslant \delta n^{-2+4\varepsilon}, \quad |\beta_n^{(1)}| \leqslant \delta n^{-3/2+3\varepsilon}. \tag{6.1}$$

Now we define independent random variables $\widehat{x}_{ij}^s$, $\widehat{x}_{ij}^l$, $c_{ij}^{(1)}$, $1 \leqslant i \leqslant p$, $1 \leqslant j \leqslant n$, as follows.

Definition 6.1. *We define $\widehat{x}^s_{ij}$ as a random variable that has law $\rho^{(1)}_s$ defined through*

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

for any event Ω, where $\rho^{(1)}(\mathrm{d}x)$ is the law of $\widehat{x}_{ij}$. We define $\widehat{x}^l_{ij}$ as a random variable that has law $\rho^{(1)}_l$ defined through

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

for any event Ω. Finally, $c^{(1)}_{ij}$ is a Bernoulli 0-1 random variable with $\mathbb{P}(c^{(1)}_{ij} = 1) = \alpha^{(1)}_n$ and $\mathbb{P}(c^{(1)}_{ij} = 0) = 1 - \alpha^{(1)}_n$.

In the above definition, $\rho^{(1)}_s$ and $\rho^{(1)}_l$ are defined in a way such that $\widehat{x}^s_{ij}$ and $\widehat{x}^l_{ij}$ are both centered. Now let X^s, X^l and X^c be independent random matrices such that $x^s_{ij} = n^{-1/2}\widehat{x}^s_{ij}$, $x^l_{ij} = n^{-1/2}\widehat{x}^l_{ij}$ and $x^c_{ij} = c^{(1)}_{ij}$. Then we can easily check that

$$x_{ij} \overset{d}{=} x^s_{ij}\left(1 - x^c_{ij}\right) + x^l_{ij} x^c_{ij} - \frac{1}{\sqrt{n}} \frac{\beta^{(1)}_n}{1-\alpha^{(1)}_n}, \tag{6.2}$$

where $\overset{d}{=}$ means that the two random variables have the same distribution. Similarly, we decompose Y as

$$y_{ij} \overset{d}{=} y^s_{ij}\left(1 - y^c_{ij}\right) + y^l_{ij} y^c_{ij} - \frac{1}{\sqrt{n}} \frac{\beta^{(2)}_n}{1-\alpha^{(2)}_n}, \tag{6.3}$$

where the entries y^s_{ij}, y^l_{ij} and y^c_{ij} of the independent random matrices Y^s, Y^l and Y^c are defined in similar ways using

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

Notice that the deterministic matrix $\mathcal{M}_1$ with

$$(\mathcal{M}_1)_{ij} = -\frac{1}{\sqrt{n}} \frac{\beta^{(1)}_n}{1-\alpha^{(1)}_n}, \quad 1 \leqslant i \leqslant p, \ 1 \leqslant j \leqslant n,$$

has operator norm $\mathrm{O}(n^{-1+3\varepsilon})$, which, by Weyl's inequality, perturbs the singular values of X at most by $\mathrm{O}(n^{-1+3\varepsilon})$. Such a small error is always negligible for our result, so we will omit the constant term in (6.2) throughout the proof. Similarly, we will also omit the constant term in (6.3). Finally, we decompose $Z = Z^s + Z^l$, where

$$Z^s_{ij} = \mathbf{1}(|Z_{ij}| \leqslant n^{-\varepsilon}) Z_{ij} + \beta^{(3)}_n, \quad Z^l_{ij} = \mathbf{1}(|Z_{ij}| > n^{-\varepsilon}) Z_{ij} - \beta^{(3)}_n, \quad \beta^{(3)}_n := \mathbb{E}[\mathbf{1}(|Z_{ij}| > n^{-\varepsilon}) Z_{ij}].$$

Using (2.2) and integration by parts, one can verify that

$$\beta^{(3)}_n = \mathrm{O}(n^{-1+\varepsilon}).$$

The deterministic vector $(\beta^{(3)}_n, \cdots, \beta^{(3)}_n)^\top \in \mathbb{R}^n$ has Euclidean norm $\mathrm{O}(n^{-1/2+\varepsilon})$, and we can easily check that it is also negligible for the following arguments. Hence for simplicity of notations, we will omit it throughout the proof.

Remark 6.2. The purpose of the above decomposition (in distribution) is to write (X,Y,Z) into well-behaved random matrices (X^s,Y^s,Z^s) with bounded support $q=\mathrm{O}(n^{-\varepsilon})$ plus a perturbation matrix. For example, for X, the perturbation is of the form $(X^l-X^s)\circ X^c$ up to a negligible deterministic term. Here the matrix X^c gives the locations of the nonzero entries, and its rank is at most $n^{5\varepsilon}$ with high probability; see (6.8) below. The matrix X^l contains the large entries above the cutoff, but the tail condition (2.34) guarantees that the sizes of these entries are of order $\mathrm{o}(1)$ in probability; see (6.13). Hence the perturbation is of low rank and has small signal strengths. We expect that, as in the famous BBP transition [5], the effect of this weak perturbation on the largest few eigenvalues is negligible.

With (2.34) and integration by parts, we can obtain that

$$\mathbb{E}\widehat{x}^s_{11}=0,\quad \mathbb{E}|\widehat{x}^s_{11}|^2=1-\mathrm{O}(n^{-1+2\varepsilon}),\quad \mathbb{E}|\widehat{x}^s_{11}|^3=\mathrm{O}(1),\quad \mathbb{E}|\widehat{x}^s_{11}|^4=\mathrm{O}(\log n). \tag{6.4}$$

Similar estimates hold for the $\widehat{y}^s_{11}$ variable. Hence $X_1:=(\mathbb{E}|\widehat{x}^s_{11}|^2)^{-1/2}X^s$ 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 Lemma 2.7, Theorem 2.9 and Theorem 2.11 with $\phi_n=\psi_n=\mathrm{O}(n^{-\varepsilon})$. Moreover, the small error $\mathrm{O}(n^{-1+2\varepsilon})$ in $\mathbb{E}|\widehat{x}^s_{11}|^2$ and $\mathbb{E}|\widehat{y}^s_{11}|^2$ can be neglected for our purpose. For Z, using $\lim_{t\to\infty}\mathbb{E}\left[|\widehat{z}_{11}|^2\mathbf{1}(|\widehat{z}_{11}|>t)\right]=0$, we get that

$$\mathbb{E}|z^s_{11}|^2=1-\mathrm{o}(1),\quad \mathbb{E}|z^l_{11}|^2=\mathrm{o}(1),$$

where we denote $\widehat{z}^s_{11}:=\sqrt{n}Z^s_{11}$ and $\widehat{z}^l_{11}:=\sqrt{n}Z^l_{11}$. Then $Z_1:=(\mathbb{E}|\widehat{z}^s_{11}|^2)^{-1/2}Z^s$ satisfy the assumptions for Z in Lemma 2.7, Theorem 2.9 and Theorem 2.11. Note that the scaling of Z^s amounts to a rescaling of A and B: $A\to A_1=(\mathbb{E}|\widehat{z}^s_{11}|^2)^{1/2}A$ and $B\to B_1=(\mathbb{E}|\widehat{z}^s_{11}|^2)^{1/2}B$ so that $A_1Z_1=AZ^s$ and $B_1Z_1=BZ^s$. In particular, we have that

$$\text{the } t_i\text{'s in (2.14) are only perturbed by an amount of } \mathrm{o}(1). \tag{6.5}$$

We denote by $\mathcal{C}^s_{\mathcal{X}\mathcal{Y}}$ and $\mathcal{C}^s_{XY}$ the SCC matrices obtained by replacing (X,Y,Z) with (X^s,Y^s,Z^s) in the definitions. Let $\widetilde{\lambda}^s_i$ and λ^s_i be their eigenvalues, respectively. Then by Theorem 2.9 and (6.5), for any $1\leqslant i\leqslant r_+$ we have that

$$|\widetilde{\lambda}^s_i-\theta_i|=\mathrm{o}(1)\quad\text{with high probability}, \tag{6.6}$$

and by Lemma 2.7, we have that

$$\lim_{n\to\infty}\mathbb{P}\left(n^{2/3}\frac{\lambda^s_1-\lambda_+}{c_{TW}}\leqslant s_1\right)=\lim_{n\to\infty}\mathbb{P}^{GOE}\left(n^{2/3}(\lambda_1-2)\leqslant s_1\right). \tag{6.7}$$

Throughout the following proof, we only consider the largest non-outlier eigenvalue. The extension to the case with multiple largest non-outlier eigenvalues is simple. We write the right-hand sides of (6.2) and (6.3) as

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

and

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

We define the matrices $\mathcal{E}^{(1)}:=(\Delta^{(1)}_{ij}x^c_{ij}:1\leqslant i\leqslant p,1\leqslant j\leqslant n)$ and $\mathcal{E}^{(2)}:=(\Delta^{(2)}_{ij}y^c_{ij}:1\leqslant i\leqslant q,1\leqslant j\leqslant n)$. It suffices to show that the effect of $\mathcal{E}^{(1)}$, $\mathcal{E}^{(2)}$ and Z^l on the eigenvalues $\widetilde{\lambda}_i$, $1\leqslant i\leqslant r_+$ and $\widetilde{\lambda}_{r_++1}$ is negligible.

Define the event

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

By a Chernoff bound, we get that

$$\mathbb{P}\left(\left\{\#\{(i,j): x^c_{ij}=1\} \leqslant n^{5\varepsilon}\right\}\right) \geqslant 1-\exp(-n^{\varepsilon}). \tag{6.8}$$

If the number n_0 of the nonzero elements in X^c satisfies $n_0 \leqslant n^{5\varepsilon}$, then we can check that

$$\mathbb{P}\left(\exists\, i=k, j\neq l \text{ or } i\neq k, j=l \text{ such that } x^c_{ij}=x^c_{kl}=1 \,\middle|\, \#\{(i,j): x^c_{ij}=1\}=n_0\right) = \mathrm{O}(n_0^2 n^{-1}). \tag{6.9}$$

Combining the estimates (6.8) and (6.9), we get that

$$\mathbb{P}(\mathscr{A}) \geqslant 1-\mathrm{O}(n^{-1+10\varepsilon}). \tag{6.10}$$

Similarly, for the event

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

we have

$$\mathbb{P}(\mathscr{B}) \geqslant 1-\mathrm{O}(n^{-1+10\varepsilon}), \tag{6.11}$$

if the nonzero elements in Y^c is at most $n^{5\varepsilon}$. On the other hand, using condition (2.34) and Markov's inequality, we get

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

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

$$\mathbb{P}\left(\max_{i,j} |\mathcal{E}^{(1)}_{ij}| \geqslant \omega\right) + \mathbb{P}\left(\max_{i,j} |\mathcal{E}^{(2)}_{ij}| \geqslant \omega\right) = \mathrm{o}(1). \tag{6.12}$$

Define the event

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

Combining (6.10), (6.11) and (6.12), we get

$$\mathbb{P}(\mathscr{A}\cap\mathscr{B}\cap\mathscr{C}_1) = 1-\mathrm{o}(1). \tag{6.13}$$

We also define the event

$$\mathscr{C}_2 := \left\{\|(Z^s)^\top Z^s - I_r\| \leqslant w, \|(Z^l)^\top Z^l\| \leqslant w^2, \|(Z^s)^\top Z^l\| \leqslant w\right\}. \tag{6.14}$$

By strong law of large number, we have $\mathbb{P}(\mathscr{C}_2) = 1-\mathrm{o}(1)$.

Recalling (3.2), we only need to study the zeros of $\det[\widetilde{H}_1(\lambda)]$ on event $\mathscr{A}\cap\mathscr{B}\cap\mathscr{C}_1\cap\mathscr{C}_2$. Here we define $\widetilde{H}_t(\lambda)$, $t\in[0,1]$, as

$$\widetilde{H}_t(\lambda) := \widetilde{H}^s(\lambda) + t\begin{pmatrix} 0 & \begin{pmatrix} \mathcal{E}^{(1)}+AZ^l & 0 \\ 0 & \mathcal{E}^{(2)}+BZ^l \end{pmatrix} \\ \begin{pmatrix} (\mathcal{E}^{(1)}+AZ^l)^\top & 0 \\ 0 & (\mathcal{E}^{(2)}+BZ^l)^\top \end{pmatrix} & 0 \end{pmatrix},$$

where

$$\widetilde{H}^s(\lambda) := H^s(\lambda) + \begin{pmatrix} 0 & \begin{pmatrix} AZ^s & 0 \\ 0 & BZ^s \end{pmatrix} \\ \begin{pmatrix} (AZ^s)^\top & 0 \\ 0 & (BZ^s)^\top \end{pmatrix} & 0 \end{pmatrix},$$

with

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

We would like to extend (6.6) and (6.7) at $t = 0$ all the way to the $t = 1$ case using a continuity argument. Correspondingly, for any $t \in [0, 1]$, we define the PCC matrix $\mathcal{C}_{\mathcal{X}\mathcal{Y}}(t)$ for $\mathcal{X}(t) := X^s + t\widetilde{\mathcal{E}}^{(1)} + A(Z^s + tZ^l)$ and $\mathcal{Y}(t) := Y^s + t\widetilde{\mathcal{E}}^{(2)} + B(Z^s + tZ^l)$, and denote its eigenvalues as $\widetilde{\lambda}_i(t)$. Note that $\widetilde{\lambda}_i = \widetilde{\lambda}_i(1)$ are the eigenvalues we are interested in, and the eigenvalues $\widetilde{\lambda}_i^s = \widetilde{\lambda}_i(0)$ satisfy (6.6) and (6.7). Moreover, $\widetilde{\lambda}_i(t)$ is continuous with respect to t on the extended real line $\overline{\mathbb{R}}$.

Proof of (2.36). Fix any $1 \leqslant i \leqslant r_+$, we pick a sufficiently small constant $\delta > 0$ such that the following properties hold for large enough n: (i) the interval $J_i := [\theta_i - \delta, \theta_i + \delta]$ only contains θ_j's that converge to the same limit as θ_i when $n \to \infty$, (ii) J_i is away from all the other θ_j's at least by δ, and (iii) J_i is away from λ_+ at least by δ. By (6.6), we know $\widetilde{\lambda}_i(0) \in J_i$ with high probability. Now for $\mu := \theta_i \pm \delta$, we claim that

$$\mathbb{P}\left(\det \widetilde{H}_t(\mu) \neq 0 \text{ for all } 0 \leqslant t \leqslant 1\right) = 1 - \mathrm{o}(1). \tag{6.15}$$

If (6.15) holds, then μ is not an eigenvalue of $\mathcal{C}_{\mathcal{X}\mathcal{Y}}(t)$ for all $t \in [0, 1]$ with probability $1 - \mathrm{o}(1)$. By continuity of $\widetilde{\lambda}_i(t)$ with respect to t, we have $\widetilde{\lambda}_i = \widetilde{\lambda}_i(1) \in J_i$ with probability $1 - \mathrm{o}(1)$, that is,

$$\mathbb{P}(|\widetilde{\lambda}_i - \theta_i| \leqslant \delta) = 1 - \mathrm{o}(1).$$

This concludes (2.36) since δ can be arbitrarily small.

For the proof of (6.15), we will condition on $\mathscr{A} \cap \mathscr{B}$ and the event $\mathscr{C}_{n_x n_y}$ that X^c and Y^c have n_x and n_y nonzero entries with $\max\{n_x, n_y\} \leqslant n^{5\varepsilon}$. Moreover, we assume that the positions of the n_x nonzero entries of X^c are $(\sigma_x(1), \pi_x(1)), (\sigma_x(2), \pi_x(2)), \cdots, (\sigma_x(n_x), \pi_x(n_x))$, and the positions of the n_y nonzero entries of Y^c are $(\sigma_y(1), \pi_y(1)), (\sigma_y(2), \pi_y(2)), \cdots, (\sigma_y(n_y), \pi_y(n_x))$. Here $\sigma_x : \{1, \cdots, n_x\} \to \{1, \cdots, p\}$, $\pi_x : \{1, \cdots, n_x\} \to \{1, \cdots, n\}$, $\sigma_y : \{1, \cdots, n_y\} \to \{1, \cdots, q\}$ and $\pi_y : \{1, \cdots, n_y\} \to \{1, \cdots, n\}$ are uniform random injective functions. Then we can rewrite that

$$\widetilde{H}_t(\mu) = H^s(\mu) + O_t \begin{pmatrix} 0 & \begin{pmatrix} \mathcal{D} & 0 \\ 0 & t\mathcal{D}_e \end{pmatrix} \\ \begin{pmatrix} \mathcal{D} & 0 \\ 0 & t\mathcal{D}_e \end{pmatrix} & 0 \end{pmatrix} O_t^\top, \quad O_t := \begin{pmatrix} (\mathbf{U}, \mathbf{F}_1) & 0 \\ 0 & (\mathbf{E}_t, \mathbf{F}_2) \end{pmatrix},$$

where $\mathcal{D}$ and $\mathbf{U}$ have been defined in (3.5) and (3.6); $\mathcal{D}_e := \begin{pmatrix} \Sigma_e^{(1)} & 0 \\ 0 & \Sigma_e^{(2)} \end{pmatrix}$ with

$$\Sigma_e^{(1)} := \operatorname{diag}\left(\mathcal{E}^{(1)}_{\sigma_x(1)\pi_x(1)}, \cdots, \mathcal{E}^{(1)}_{\sigma_x(n_x)\pi_x(n_x)}\right), \quad \Sigma_e^{(2)} := \operatorname{diag}\left(\mathcal{E}^{(2)}_{\sigma_y(1)\pi_y(1)}, \cdots, \mathcal{E}^{(2)}_{\sigma_y(n_y)\pi_y(n_y)}\right);$$

$$\mathbf{E}_t := \begin{pmatrix} (Z_t^\top \mathbf{v}_1^a, \cdots, Z_t^\top \mathbf{v}_r^a) & 0 \\ 0 & (Z_t^\top \mathbf{v}_1^b, \cdots, Z_t^\top \mathbf{v}_r^b) \end{pmatrix}, \quad \text{with} \quad Z_t := Z^s + tZ^l;$$

$$\mathbf{F}_1 := \begin{pmatrix} \left(\mathbf{e}^{(p)}_{\sigma_x(1)}, \cdots, \mathbf{e}^{(p)}_{\sigma_x(n_x)}\right) & 0 \\ 0 & \left(\mathbf{e}^{(q)}_{\sigma_y(1)}, \cdots, \mathbf{e}^{(q)}_{\sigma_y(n_y)}\right) \end{pmatrix};$$

$$\mathbf{F}_2 := \begin{pmatrix} \left(\mathbf{e}^{(n)}_{\pi_x(1)}, \cdots, \mathbf{e}^{(n)}_{\pi_x(n_x)}\right) & 0 \\ 0 & \left(\mathbf{e}^{(n)}_{\pi_y(1)}, \cdots, \mathbf{e}^{(n)}_{\pi_y(n_y)}\right) \end{pmatrix}.$$

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

Applying the identity $\det(1 + \mathcal{AB}) = \det(1 + \mathcal{BA})$, we obtain that

$$\det \widetilde{H}_t(\mu) = \det\left[G^s(\mu)\right] \cdot \det\left[1 + \widetilde{F}_t(\mu) + \mathcal{E}_t(\mu)\right], \tag{6.16}$$

where

$$\widetilde{F}_t(\mu) := \begin{pmatrix} 0 & \begin{pmatrix} \mathcal{D} & 0 \\ 0 & t\mathcal{D}_e \end{pmatrix} \\ \begin{pmatrix} \mathcal{D} & 0 \\ 0 & t\mathcal{D}_e \end{pmatrix} & 0 \end{pmatrix} O_t^\top \Pi(\mu) O_t,$$

and

$$\mathcal{E}_t(\mu) := \begin{pmatrix} 0 & \begin{pmatrix} \mathcal{D} & 0 \\ 0 & t\mathcal{D}_e \end{pmatrix} \\ \begin{pmatrix} \mathcal{D} & 0 \\ 0 & t\mathcal{D}_e \end{pmatrix} & 0 \end{pmatrix} O_t^\top \left[G^s(\mu) - \Pi(\mu)\right] O_t.$$

Note that O_t is deterministic conditioning on Z. Hence by Lemma 3.11, we have that (recall (6.14))

$$\mathbb{E}\left[\left|\left[O_t^\top \left(G^s(\mu) - \Pi(\mu)\right) O_t\right]_{ij}\right|^2 \middle| \mathscr{C}_{n_x n_y}, Z, \mathscr{C}_2\right] \prec n^{-1}, \quad 1 \leqslant i,j \leqslant 2r + n_x + n_y.$$

Applying Markov's inequality to this estimate and using a simple union bound, we get that

$$\max_{1\leqslant i,j\leqslant 2r+n_x+n_y} \left|\left[O_t^\top \left(G^s(\mu) - \Pi(\mu)\right) O_t\right]_{ij}\right| \leqslant n^{-1/4} \quad \text{with probability } 1 - \mathrm{O}(n^{-1/2+11\varepsilon}), \tag{6.17}$$

conditioning on $\mathscr{C}_{n_x n_y}$, Z and $\mathscr{C}_2$. Next we claim that on $\mathscr{C}_1 \cap \mathscr{C}_2$,

$$\sup_{0\leqslant t\leqslant 1} \left\|\widetilde{F}_t(\mu) - \widetilde{F}_0(\mu)\right\| \leqslant C\omega, \tag{6.18}$$

for some constant $C > 0$ that is independent of ω. In fact, expanding $\widetilde{F}_t(\mu)$ and using that $\|\Pi(\mu)\| = \mathrm{O}(1)$, $\|t\Sigma^{(1)}_e\| \leqslant \omega$, $\|t\Sigma^{(2)}_e\| \leqslant \omega$ and $\| \mathbf{E}_t - \mathbf{E}_0 \| = \mathrm{O}(\omega)$ on $\mathscr{C}_1 \cap \mathscr{C}_2$, we can easily obtain (6.18). Then combining (6.17) and (6.18) we get that on event $\mathscr{A} \cap \mathscr{B} \cap \mathscr{C}_1 \cap \mathscr{C}_2$,

$$\det\left(1 + \widetilde{F}_t(\mu) + \mathcal{E}_t(\mu)\right) = \det\left(1 + \widetilde{F}_0(\mu) + \mathrm{O}(\omega)\right) \quad \text{for all } t \in [0,1], \tag{6.19}$$

with probability $1 - \mathrm{o}(1)$. When $t = 0$, the discussion at the beginning of Section 4 (i.e. the argument leading to (4.6)) gives that at $\mu = \theta_i \pm \delta$, $\|(1 + \widetilde{F}_0(\mu))^{-1}\| \leqslant C_\delta$ with high probability for some constant $C_\delta > 0$. Thus by (6.19), as long as ω is sufficiently small, we have that with probability $1 - \mathrm{o}(1)$, $\det(1 + \widetilde{F}_t(\mu) + \mathcal{E}_t(\mu)) \neq 0$ for all $t \in [0,1]$. This concludes (6.15), which further concludes (2.36). ☐

Proof of (2.31) *for Theorem 2.14.* Similar to (6.15), we claim that

$$\mathbb{P}\left(\det \widetilde{H}_t(\mu) \neq 0 \text{ for all } 0 \leqslant t \leqslant 1\right) = 1 - \mathrm{o}(1), \quad \text{for} \quad \mu = \lambda_1(0) \pm n^{-3/4} \equiv \lambda^s_1 \pm n^{-3/4}. \tag{6.20}$$

Recall that at $t = 0$, by Theorem 2.11, we have $|\widetilde{\lambda}_{1+r_+}(0) - \lambda_1^b(0)| \prec n^{-1}$. Applying Theorem 2.11 again gives $|\lambda_1^b(0) - \lambda_1(0)| \prec n^{-1}$. Thus we have that

$$\widetilde{\lambda}_{1+r_+}(0) \in [\lambda_1^s - n^{-3/4}, \lambda_1^s + n^{-3/4}] \quad \text{with high probability.}$$

If (6.20) holds, then by continuity of $\widetilde{\lambda}_{1+r_+}(t)$ with respect to t, we get $\widetilde{\lambda}_{1+r_+} \equiv \widetilde{\lambda}_{1+r_+}(1) \in [\lambda_1^s - n^{-3/4}, \lambda_1^s + n^{-3/4}]$ with probability $1 - \mathrm{o}(1)$, which concludes the proof together with (6.7).

In the following proof, we choose $z = \lambda_+ + \mathrm{i}n^{-2/3}$. As in (6.16), we need to study

$$\det \left[1 + \widetilde{F}_t(z) + \mathcal{E}_t(z) + \begin{pmatrix} 0 & \begin{pmatrix} \mathcal{D} & 0 \\ 0 & t\mathcal{D}_e \end{pmatrix} \\ \begin{pmatrix} \mathcal{D} & 0 \\ 0 & t\mathcal{D}_e \end{pmatrix} & 0 \end{pmatrix} O_t^\top \left[G^s(\mu) - G^s(z) \right] O_t \right],$$

where we used the simple identity

$$O_t^\top G^s(\mu) O_t = O_t^\top \left[G^s(\mu) - G^s(z) \right] O_t + O_t^\top G^s(z) O_t.$$

Repeating the proof below (6.16), we can show that with probability $1 - \mathrm{o}(1)$,

$$1 + \widetilde{F}_t(z) + \mathcal{E}_t(z) = 1 + \widetilde{F}_0(z) + \mathrm{O}(\omega) \quad \text{for all} \quad t \in [0, 1], \tag{6.21}$$

and $\|(1 + \widetilde{F}_0(z))^{-1}\| \leqslant C$ with high probability for some constant $C > 0$ that is independent of ω. Moreover, we have that

$$\|O_t^\top \left[G^s(\mu) - G^s(z) \right] O_t\| \leqslant n^{-1/6} \quad \text{with probability } 1 - \mathrm{o}(1), \tag{6.22}$$

which is proved as (5.16) in [50]. Combining (6.21) and (6.22), we get that with probability $1 - \mathrm{o}(1)$,

$$\det \left(1 + \widetilde{F}_t(\mu) + \mathcal{E}_t(\mu) \right) = \det \left(1 + \widetilde{F}_0(z) + \mathrm{O}(\omega) \right) \neq 0 \quad \text{for all} \quad t \in [0, 1],$$

as long as ω is sufficiently small. This concludes (6.20), which completes the proof of (2.31) for the $k = 1$ case. It is easy to extend the above proof to the $k > 1$ case, and we omit the details. □

7 Proof of Lemma 2.7

Finally, in this section, we present the proof of Lemma 2.7. It has been proved in [50] when $B = 0$, and we need to show that adding the BZ term to Y does not affect the results. By Theorem 2.5 of [50], (2.21) holds for λ_i, the eigenvalues of $\mathcal{C}_{XY}$. On the other hand, by Theorem 2.11 we have

$$|\lambda_i^b - \lambda_i| \prec n^{-1} \alpha_+^{-1} \lesssim n^{-1},$$

where in the second step we used that $t_i = 0$ for all $1 \leqslant i \leqslant r$ and hence $\alpha_+ = t_c \sim 1$. This shows that (2.21) also holds for λ_i^b.

However, since we need to use (2.20) in the proof of Theorem 2.11, we cannot use (2.25) and (3.42) to conclude (2.20). Instead, we need a separate argument. We first prove an averaged local law for $G^b(z)$ as in (3.40) and (3.41), using the following resolvent estimates.

Lemma 7.1 (Lemma 3.8 of [50]). *For any deterministic unit $\mathbf{v}_\beta \in \mathbb{C}^{\mathcal{I}_\beta}$, $\beta = 3, 4$, we have that*

$$\sum_{\mathfrak{a} \in \mathcal{I}} \left| G_{\mathfrak{a}\mathbf{v}_\beta} \right|^2 \prec 1 + \frac{\left| \operatorname{Im}(\mathcal{U}\mathcal{G}_R)_{\mathbf{v}_\beta \mathbf{v}_\beta} \right|}{\eta}, \quad \sum_{\mathfrak{a} \in \mathcal{I}} \left| G_{\mathbf{v}_\beta \mathfrak{a}} \right|^2 \prec 1 + \frac{\left| \operatorname{Im}(\mathcal{G}_R \mathcal{U}^\top)_{\mathbf{v}_\beta \mathbf{v}_\beta} \right|}{\eta}, \tag{7.1}$$

where

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

Now we calculate $m_3^b(z) = n^{-1}\sum_{\mu\in\mathcal{I}_3} G^b_{\mu\mu}(z)$ using (3.61). By the anisotropic local law (3.57), we have that with high probability,

$$\left\| \left[1 + \begin{pmatrix} 0 & \mathcal{D}_b \\ \mathcal{D}_b & 0 \end{pmatrix}\begin{pmatrix} \mathbf{U}_b^\top & 0 \\ 0 & \mathbf{E}_b^\top \end{pmatrix} G(z) \begin{pmatrix} \mathbf{U}_b & 0 \\ 0 & \mathbf{E}_b \end{pmatrix}\right]^{-1}\begin{pmatrix} 0 & \mathcal{D}_b \\ \mathcal{D}_b & 0 \end{pmatrix}\right\| = \mathrm{O}(1).$$

Hence by (3.61), we obtain that (recall (3.56))

$$|m_3^b(z) - m_3(z)| \prec \max_{1\leqslant k\leqslant r}\sum_{\mu\in\mathcal{I}_3}\left(|G_{\mu\mathbf{u}_k^b}(z)|^2 + |G_{\mu\tilde{\mathbf{v}}_k^b}(z)|^2\right),$$

where we abbreviated that $\tilde{\mathbf{v}}_k^b := Z^\top \mathbf{v}_k^b$. Note that $\tilde{\mathbf{v}}_k^b$ are approximately orthonormal vectors by (3.55). Then using (7.1), we obtain that for $z \in \widetilde{S}(\varepsilon, \widetilde{\varepsilon})$,

$$\begin{aligned}|m_3^b(z) - m_3(z)| &\prec \frac{1}{n} + \max_{1\leqslant k\leqslant r}\frac{|\operatorname{Im}(\mathcal{U}\mathcal{G}_R)_{\mathbf{u}_k^b\mathbf{u}_k^b}| + |\operatorname{Im}(\mathcal{U}\mathcal{G}_R)_{\tilde{\mathbf{v}}_k^b\tilde{\mathbf{v}}_k^b}|}{n\eta} \\ &\prec \frac{1}{n} + \max_{1\leqslant k\leqslant r}\frac{\eta + \operatorname{Im} m_c(z) + \Psi(z) + \psi_n + \phi_n}{n\eta} \lesssim \Psi^2(z) + \frac{\psi_n + \phi_n}{n\eta},\end{aligned} \tag{7.2}$$

where in the second step we used the local law (3.57) and that

$$\left|\operatorname{Im}\left(\mathcal{U}\Pi_R^b(z)\right)_{\mathbf{u}_k^b\mathbf{u}_k^b}\right| + \left|\operatorname{Im}\left(\mathcal{U}\Pi_R^b\right)_{\tilde{\mathbf{v}}_k^b\tilde{\mathbf{v}}_k^b}\right| \lesssim \operatorname{Im} m_c(z) + \eta.$$

Here $\Pi_R^b(z)$ denotes the $(\mathcal{I}_3 \cup \mathcal{I}_4) \times (\mathcal{I}_3 \cup \mathcal{I}_4)$ block of Π^b. Combining (7.2) with the averaged local laws (3.40)–(3.41) for $m_3(z)$, and equation (3.17) for $m_3^b(z)$ and $m^b(z)$, we obtain the following local laws: for any fixed $\varepsilon, \widetilde{\varepsilon} > 0$,

$$|m^b(z) - m_c(z)| \prec (n\eta)^{-1} \tag{7.3}$$

uniformly in $z \in \widetilde{S}(\varepsilon, \widetilde{\varepsilon})$, and

$$|m^b(z) - m_c(z)| \prec \frac{\psi_n + \phi_n}{n\eta} + \frac{1}{n(\kappa+\eta)} + \frac{1}{(n\eta)^2\sqrt{\kappa+\eta}} \tag{7.4}$$

uniformly in $z \in \widetilde{S}_{out}(\varepsilon, \widetilde{\varepsilon})$.

Next we introduce the following regularized resolvents.

Definition 7.2 (Regularized resolvents)**.** *For $z = E + \mathrm{i}\eta \in \mathbb{C}_+$, we define the regularized resolvent $\widehat{G}(z)$ as*

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

Moreover, 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 the resolvents $\widehat{R}(z)$, $\widehat{G}^b(z)$ and $\widehat{R}^b(z)$ etc. can be defined in the obvious way as in Definition 3.2.

By Schur complement formula, we can obtain similar expressions for $\widehat{\mathcal{G}}_L$, $\widehat{\mathcal{G}}_R$ and $\widehat{\mathcal{G}}_{LR}$ as in (3.14)–(3.16). The main reason for introducing the regularized resolvents is that they satisfy the following deterministic bounds: for some constant $C > 0$,

$$\left\|\widehat{G}(z)\right\| \leqslant \frac{Cn^{10}}{\eta}, \quad \left\|\widehat{G}^b(z)\right\| \leqslant \frac{Cn^{10}}{\eta}. \tag{7.5}$$

This estimate has been proved in Lemma 3.6 of [50]. With a standard perturbation argument, we can control the difference between $\widehat{G}(z)$ and $G(z)$ as in the following claim.

Claim 7.3. *Suppose there exists a high probability event Ξ on which $\|G(z)\|_{\max} = \mathrm{O}(1)$ for z in some subset, where $\|G\|_{\max} := \max_{i,j} |G_{ij}|$ denotes the max norm. Then we have that*

$$\|G(z) - \widehat{G}(z)\|_{\max} \leqslant n^{-8} \quad on \quad \Xi. \tag{7.6}$$

The same bound also holds for $\|G^b(z) - \widehat{G}^b(z)\|_{\max}$ on event $\{\|G^b(z)\|_{\max} = \mathrm{O}(1)\}$ or $\{\|\widehat{G}^b(z)\|_{\max} = \mathrm{O}(1)\}$.

Proof. For $t \in [0, 1]$, we define

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

Taking derivative with respect to t, we immediately get that

$$\partial_t G_t(z) = zn^{-10} G_t(z)\begin{pmatrix} I_{p+q} & 0 \\ 0 & 0 \end{pmatrix} G_t(z). \tag{7.7}$$

Thus applying Gronwall's inequality to

$$\|G_t(z)\|_{\max} \leqslant \|G(z)\|_{\max} + Cn^{-9}\int_0^t \|G_s(z)\|_{\max}^2 \mathrm{d}s,$$

we obtain that $\|G_t(z)\|_{\max} \leqslant C$ for all $0 \leqslant t \leqslant 1$ on Ξ. Then using (7.7) again, we get (7.6). ☐

Note that the bound (7.6) is purely deterministic on Ξ, so we do not lose any probability here. Moreover, such a small error n^{-8} will not affect any of our results.

Proof of (2.20). With the same arguments as the ones for [22, Theorems 2.12 and 2.13], [23, Theorem 2.2] and [43, Theorem 3.3], from the averaged local law (7.3) we can derive that for any small constants $\delta, \varepsilon > 0$, (2.20) holds for all $n^{\varepsilon} \leqslant i \leqslant (1-\delta)q$. To conclude (2.20) for the first n^{ε} eigenvalues, we still need to prove an upper bound on them. More precisely, it suffices to show that for any small constant $\varepsilon > 0$,

$$\lambda_1^b \leqslant \lambda_+ + n^{-2/3+\varepsilon}, \quad w.h.p. \tag{7.8}$$

Combining this estimate with the rigidity estimate for $\lambda^b_{n^{\varepsilon}}$, we can conclude that (2.20) holds all $1 \leqslant i < (1-\delta)q$ since ε can be arbitrarily small.

First, using the averaged local law (7.4), we can obtain that for any small constants $c, \varepsilon > 0$,

$$\#\{i : \lambda_i^b \in [\lambda_+ + n^{-2/3+\varepsilon}, 1-c]\} = 0, \quad w.h.p. \tag{7.9}$$

The proof is standard and similar to the one for (4.7) of [50], so we omit the details. It remains to prove that

$$\#\{i : \lambda_i^b \in [1-c, 1]\} = 0, \quad w.h.p., \tag{7.10}$$

for a sufficiently small constant $c > 0$.

For $t \in [0,1]$, we define a continuous path of interpolated random matrices between Y and $Y + BZ$ as

$$\mathcal{Y}_t := Y + tBZ, \quad t \in [0,1].$$

By replacing $\mathcal{Y}$ with $\mathcal{Y}_t$ in (3.10) and Definition 7.2, we can define $H^b_t(z)$, $G^b_t(z)$, $\widehat{H}^b_t(z)$ and $\widehat{G}^b_t(z)$ correspondingly. First, we claim the following result.

Claim 7.4. *With high probability, we have that*

$$\|G^b_t(1-c)\|_{\max} < \infty \quad \text{for all} \quad t \in [0,1]. \tag{7.11}$$

We postpone the proof of this claim until we complete the proof of (2.20). Let $\lambda^b_1(t) \geqslant \lambda^b_2(t) \geqslant \cdots \geqslant \lambda^b_q(t)$ be the eigenvalues of $\mathcal{C}_{X\mathcal{Y}_t}$. For any $1 \leqslant i \leqslant q$, $\lambda^b_i(t) : [0,1] \to \mathbb{R}$ is a continuous function with respect to t on the extended real line $\overline{\mathbb{R}}$. By (3.42), the eigenvalues $\lambda^b_i(0)$ of $\mathcal{C}_{XY}$ are all inside $[0, \lambda_+ + n^{-2/3+\varepsilon}]$ with high probability. If (7.11) holds, then we have that

$$m^b_t(1-c) = \frac{1}{q}\sum_{i=1}^{q} \frac{1}{\lambda^b_i(t) - (1-c)} \quad \text{is finite for all } t \in [0,1].$$

It means that the eigenvalue $\lambda^b_1(t)$ does not cross the point $E = 1 - c$ for all $t \in [0,1]$. Thus we conclude (7.10), which further concludes (7.8) together with (7.9). □

Finally, it remains to prove Claim 7.4.

Proof of Claim 7.4. Take a discrete net of t, $t_k = kn^{-50}$, for $0 \leqslant k \leqslant n^{50}$. First, we claim that there exists a high probability event Ξ_1 so that

$$\mathbf{1}(\Xi_1) \max_{0\leqslant k\leqslant n^{50}} \|\widehat{G}^b_{t_k}(E + \mathrm{i}n^{-10})\|_{\max} \leqslant C \quad \text{for} \quad E := 1-c, \tag{7.12}$$

for some large constant $C > 0$. In fact, notice that $\mathcal{Y}_t$ also satisfies the assumptions for $\mathcal{Y}$ in Lemma 2.7. Hence using (7.9), we obtain that for any t_k, the eigenvalues $\lambda^b_i(t_k)$ are inside $[0, \lambda_+ + n^{-2/3+\varepsilon}] \cup [1-c/2, 1]$ with high probability. By taking a union bound, we get that

$$\min_{0\leqslant k\leqslant n^{50}} \min_{1\leqslant i\leqslant q} |E - \lambda^b_i(t_k)| \gtrsim 1 \quad \text{w.h.p.} \tag{7.13}$$

Applying the spectral decomposition (3.13) to R^b, we obtain from (7.13) that

$$\max_{0\leqslant k\leqslant n^{50}} \left\|R^b_{t_k}(z)\right\| \leqslant C, \quad \text{for} \quad z = E + \mathrm{i}n^{-10}.$$

Combining this bound with (3.14)–(3.16), and using Lemma 3.3, we get that

$$\max_{0\leqslant k\leqslant n^{50}} \left\|G^b_{t_k}(z)\right\| \leqslant C, \quad \text{w.h.p.}$$

Finally, applying Claim 7.3, we get (7.12) for $\widehat{G}^b$.

Now given (7.12), using the deterministic bound (7.5) for $\widehat{G}^b$, we get that on Ξ_1,

$$\begin{aligned}\left\|\widehat{G}^b_t(E + \mathrm{i}n^{-10}) - \widehat{G}^b_{t_k}(E + \mathrm{i}n^{-10})\right\|_{\max} &\lesssim n^{-50}\|\widehat{G}^b_t(E + \mathrm{i}n^{-10})\| \cdot \|Z\| \cdot \|\widehat{G}^b_{t_k}(E + \mathrm{i}n^{-10})\| \\ &\lesssim n^{-50} \cdot \left(n^{20}\right)^2 \cdot \|Z\| \lesssim n^{-10}\|Z\|,\end{aligned}$$

for any $t_{k-1} \leqslant t \leqslant t_k$. By the bounded support condition of Z, we have that $\|Z\| = \mathrm{O}(\sqrt{n})$ on a high probability event Ξ_2. Thus we have that on the high probability event $\Xi_1 \cap \Xi_2$,

$$\left\| \widehat{G}^b_t(E + \mathrm{i}n^{-10}) - \widehat{G}^b_{t_k}(E + \mathrm{i}n^{-10}) \right\|_{\max} \lesssim n^{-50} \cdot \left(n^{20}\right)^2 \cdot \sqrt{n} \leqslant n^{-9},$$

which gives that

$$\mathbf{1}(\Xi_1 \cap \Xi_2) \max_{0 \leqslant t \leqslant 1} \|\widehat{G}^b_t(E + \mathrm{i}n^{-10})\|_{\max} \leqslant C.$$

Finally, using the same perturbation argument as in the proof of Claim 7.3, we can remove both the $\mathrm{i}n^{-10}$ and the regularization in $\widehat{G}$, which gives (7.11) on $\Xi_1 \cap \Xi_2$. □

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. Bai and J. Yao. Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. H. Poincaré Probab. Statist.*, 44(3):447–474, 2008.

[4] 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.

[5] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[6] J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382 – 1408, 2006.

[7] 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.

[8] S. T. Belinschi, H. Bercovici, M. Capitaine, and M. Février. Outliers in the spectrum of large deformed unitarily invariant models. *Ann. Probab.*, 45(6A):3571–3625, 2017.

[9] F. Benaych-Georges, A. Guionnet, and M. Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.*, 16:1621–1662, 2011.

[10] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494 – 521, 2011.

[11] 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.

[12] 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.

[13] P. Bourgade, H.-T. Yau, and J. Yin. Local circular law for random matrices. *Probab. Theory Relat. Fields*, 159:545–595, 2014.

[14] M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009.

[15] M. Capitaine, C. Donati-Martin, and D. Féral. Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 48(1):107–133, 2012.

[16] X. Ding and F. Yang. Edge statistics of large dimensional deformed rectangular matrices. *arXiv:2009.00389*, 2020.

[17] X. Ding and F. Yang. Spiked separable covariance matrices and principal components. *Annals of Statistics (in press)*, 2020.

[18] X. Ding and F. Yang. Tracy-widom distribution for the edge eigenvalues of Gram type random matrices. *arXiv:2008.04166*, 2020.

[19] L. Erdős, A. Knowles, and H.-T. Yau. Averaging fluctuations in resolvents of random band matrices. *Ann. Henri Poincaré*, 14:1837–1926, 2013.

[20] 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.

[21] 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.

[22] 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.

[23] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics*, 229:1435 – 1515, 2012.

[24] D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics*, 272(1):185–228, 2007.

[25] D. Féral and S. Péché. The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case. *Journal of Mathematical Physics*, 50(7):073302, 2009.

[26] P. Forrester. The spectrum edge of random matrix ensembles. *Nucl. Phys. B*, 402(3):709 – 728, 1993.

[27] 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.

[28] C. Gao, Z. Ma, Z. Ren, and H. H. Zhou. Minimax estimation in sparse canonical correlation analysis. *Ann. Statist.*, 43(5):2168–2197, 2015.

[29] C. Gao, Z. Ma, and H. H. Zhou. Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.*, 45(5):2074–2101, 2017.

[30] 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.

[31] 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.

[32] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3-4):321–377, 1936.

[33] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29:295–327, 2001.

[34] I. M. Johnstone. Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. *Ann. Statist.*, 36(6):2638–2716, 2008.

[35] I. M. Johnstone and A. Onatski. Testing in high-dimensional spiked models. *Ann. Statist.*, 48(3):1231–1254, 2020.

[36] A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.*, 66:1663–1749, 2013.

[37] A. Knowles and J. Yin. The outliers of a deformed Wigner matrix. *Ann. Probab.*, 42(5):1980–2031, 2014.

[38] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, pages 1–96, 2016.

[39] Z. Ma and F. Yang. Limiting distribution of the sample canonical correlation coefficients of high-dimensional random vectors. *In preparation*, 2021.

[40] R. Oda, H. Yanagihara, and Y. Fujikoshi. Asymptotic null and non-null distributions of test statistics for redundancy in high-dimensional canonical correlation analysis. *Random Matrices: Theory and Applications*, 08(01):1950001, 2019.

[41] D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.

[42] S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134(1):174–174, 2006.

[43] N. S. Pillai and J. Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24:935–1001, 2014.

[44] C. A. Tracy and H. Widom. Level-spacing distributions and the airy kernel. *Comm. Math. Phys.*, 159:151–174, 1994.

[45] C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177:727–754, 1996.

[46] K. W. Wachter. The limiting empirical measure of multiple discriminant ratios. *Ann. Statist.*, 8(5):937–957, 1980.

[47] Q. Wang and J. Yao. Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *Ann. Statist.*, 45(1):415–460, 2017.

[48] 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.

[49] F. Yang. Edge universality of separable covariance matrices. *Electron. J. Probab.*, 24:57 pp., 2019.

[50] F. Yang. Sample canonical correlation coefficients of high-dimensional random vectors: local law and Tracy-Widom limit. *arXiv:2002.09643*, 2020.

[51] F. Yang, S. Liu, E. Dobriban, and D. P. Woodruff. How to reduce dimension with PCA and random projections? *arXiv:2005.00511.*

[52] Y. Yang and G. Pan. The convergence of the empirical distribution of canonical correlation coefficients. *Electron. J. Probab.*, 17:13 pp., 2012.

[53] 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.