

Ridge-Regularized Largest Root Test for General High-dimensional Double Wishart Problems

Haoran Li*

*Department of Mathematics and Statistics
Auburn University*

April 23, 2025

Abstract

In multivariate analysis, many core problems involve the eigen-analysis of an F -matrix, $\mathbf{F} = \mathbf{W}_1 \mathbf{W}_2^{-1}$, constructed from two Wishart matrices, \mathbf{W}_1 and \mathbf{W}_2 . These so-called *Double Wishart problems* arise in contexts such as MANOVA, covariance matrix equality testing, and hypothesis testing in multivariate linear regression. A prominent classical approach, Roy's largest root test, relies on the largest eigenvalue of \mathbf{F} for inference. However, in high-dimensional settings, this test becomes impractical due to the singularity or near-singularity of \mathbf{W}_2 . To address this challenge, we propose a ridge-regularization framework by introducing a ridge term to \mathbf{W}_2 . Specifically, we develop a family of ridge-regularized largest root tests, leveraging the largest eigenvalue of $\mathbf{F}_\lambda = \mathbf{W}_1(\mathbf{W}_2 + \lambda I)^{-1}$, where $\lambda > 0$ is the regularization parameter. Under mild assumptions, we establish the asymptotic Tracy-Widom distribution of the largest eigenvalue of \mathbf{F}_λ after appropriate scaling. An efficient method for estimating the scaling parameters is proposed using the Marčenko-Pastur equation, and the consistency of these estimators is proven. The proposed framework is applied to illustrative Double Wishart problems, and simulation studies are conducted to evaluate the numerical performance of the methods. Finally, the proposed method is applied to the *Human Connectome Project* data to test for the presence of associations between volumetric measurements of human brain and behavioral variables.

Keywords: Regularized tests, F -matrix, Tracy-Widom Law, Random Matrix Theory.

*Haoran Li is Assistant Professor, Department of Mathematics and Statistics, Auburn University, 221 Parker Hall, Auburn, AL, 36849. hz10152@auburn.edu

1 Introduction

Double Wishart problems represent a fundamental class of hypothesis testing challenges encountered in various fields, including multivariate data analysis, signal processing, and economics. Notable examples include *multivariate analysis of variance* (MANOVA), testing general linear hypotheses in multivariate linear models (GLHT), testing the equality of covariance matrices (TECM), and *canonical correlation analysis* (CCA). Each of these problems involves a standard null hypothesis, such as equality of group means, zero regression coefficients, equality of covariance matrices, or absence of correlation. These problems are categorized within the same class because they share a common structure: the standard tests for these hypotheses rely on the eigen-analysis of an F -matrix of the form $\mathbf{F} = \mathbf{W}_1 \mathbf{W}_2^{-1}$, where \mathbf{W}_1 and \mathbf{W}_2 are independent Wishart matrices under the assumption of normality in the observations. This formulation gives rise to the term “double Wishart”. The specific structures of \mathbf{W}_1 and \mathbf{W}_2 vary depending on the particular double Wishart problem.

This paper introduces a general framework for the F -matrix, designed to unify the analysis of various double Wishart problems under the null hypothesis. Specifically, we define:

$$\mathbf{F} = \mathbf{W}_1 \mathbf{W}_2^{-1} = \left(\frac{1}{n_1} \Sigma_p^{1/2} \mathbf{Z} P_1 \mathbf{Z}^T \Sigma_p^{1/2} \right) \left(\frac{1}{n_2} \Sigma_p^{1/2} \mathbf{Z} P_2 \mathbf{Z}^T \Sigma_p^{1/2} \right)^{-1}, \quad \text{where} \quad (1)$$

- (a) $\mathbf{Z} = [z_{ij}]$ is a $p \times n_T$ matrix with i.i.d. entries z_{ij} having mean 0 and variance 1;
- (b) Σ_p is an arbitrary $p \times p$ positive definite covariance matrix, and $\Sigma_p^{1/2}$ is its symmetric “square-root”;
- (c) P_1 and P_2 are $n_T \times n_T$ orthogonal projection matrices of rank n_1 and n_2 , respectively,

with P_1 orthogonal to P_2 .

The F -matrices associated with a broad range of double Wishart problems under their standard null hypotheses conform to (1), including MANOVA, GLHT, TECM, and CCA (see Han et al. [2018] for a detailed discussion on CCA). As an illustrative example, we present the details of GLHT.

In the context of GLHT, we consider a setting with p -variate responses and m -variate predictors, observed from n_T independent subjects. The relationship between the responses and predictors is modeled using the linear model:

$$\mathbf{Y} = BX + \Sigma_p^{1/2}\mathbf{Z},$$

where \mathbf{Y} is the $p \times n_T$ response matrix; X is the $m \times n_T$ deterministic predictor matrix; B is the $p \times m$ coefficient matrix; \mathbf{Z} and Σ_p are as in (a) and (b). The primary objective is to test linear hypotheses about the regression coefficients of the form:

$$H_0 : BC = 0 \quad \text{vs.} \quad H_a : BC \neq 0,$$

where C is an $m \times q$ constraint matrix, with each column representing a linear hypothesis on the coefficients. This formulation is broadly applicable across various fields, with MANOVA being a notable special case. Assuming X and C are of full rank, standard tests in this context rely on the hypothesis sum-of-squares matrix and residual sum-of-squares matrix:

$$\begin{aligned} SS_H &= \mathbf{Y}X^T (XX^T)^{-1} C \left[C^T (XX^T)^{-1} C \right]^{-1} C^T (XX^T)^{-1} X\mathbf{Y}^T := \mathbf{Y}P_1\mathbf{Y}^T, \\ SS_{\text{Res}} &= \mathbf{Y} \left(I_{n_T} - X^T (XX^T)^{-1} X \right) \mathbf{Y}^T := \mathbf{Y}P_2\mathbf{Y}^T. \end{aligned} \tag{2}$$

In the context of MANOVA, SS_H and SS_{Res} are known as the between-group and within-

group sum-of-squares matrices. Under this framework, \mathbf{W}_1 and \mathbf{W}_2 are defined as:

$$\mathbf{W}_1 = \frac{1}{n_1} SS_H, \quad \text{and} \quad \mathbf{W}_2 = \frac{1}{n_2} SS_{\text{Res}}, \quad (3)$$

where $n_1 = q$ and $n_2 = n_T - m$ are the degrees of freedom of \mathbf{W}_1 and \mathbf{W}_2 , respectively. It is straightforward to verify that this construction conforms to (1) under $H_0 : BC = 0$. Additionally, note that \mathbf{W}_2 serves as an unbiased estimator of Σ_p under both H_0 and H_a . This property is crucial for the estimation procedure proposed in Section 3. Further details are presented in Section 4.

To test the hypothesis, classical literature commonly employs two main types of techniques, both relying on the eigenvalues of \mathbf{F} . The first approach involves summing all eigenvalues of \mathbf{F} after applying a specific transformation, as in the classical likelihood ratio test statistic, also known as Wilk's lambda. This method is useful for assessing the overall deviation from the null hypothesis.

In this paper, we focus on the second approach, known as *Roy's largest root test*, which relies exclusively on the largest eigenvalue of \mathbf{F} , denoted by ℓ_{\max} . This technique is particularly effective for detecting concentrated alternatives, where the signal is primarily captured by the leading eigen-direction. Specifically, in the context of GLHT, the test rejects the null hypothesis $H_0 : BC = 0$ if $\ell_{\max} > \xi_\alpha$ at a significance level α , where ξ_α is the critical value corresponding to the $(1 - \alpha) \times 100\%$ quantile of the distribution of ℓ_{\max} under the null. Traditionally, ξ_α is approximated using the critical value of an F or χ^2 distribution after appropriate scaling (see Chapter 10 of Muirhead [2009] for further details).

In contemporary statistical research, it is increasingly common for the dimension p to be at least comparable to the sample size n_T or the degrees of freedom n_1 and n_2 . While Roy's largest root test performs well when n_1 and n_2 are much larger than p , its

reliability diminishes significantly in high-dimensional settings. Specifically, when $p > n_2$, the singularity of \mathbf{W}_2 renders the F -matrix ill-defined. Even when p is smaller than n_2 , the presence of small eigenvalues in \mathbf{W}_2 can cause instability in \mathbf{W}_2^{-1} , thereby compromising the power of the test. Addressing these challenges is crucial for ensuring the robustness and accuracy of statistical methods in high-dimensional contexts.

To correct the asymptotic null distribution of Roy’s largest root, Johnstone [2008] demonstrated that the largest root of a double Wishart matrix under normality follows an asymptotic Tracy-Widom distribution of type one after appropriate scaling, given that p , n_1 , and n_2 grow to infinity proportionally with $p < n_2$. This result was later extended to non-Gaussian settings by Han et al. [2016] and Han et al. [2018]. However, when $p > n_2$, to the best of our knowledge, remedies for the classical double Wishart tests have been thoroughly studied only for the likelihood ratio tests and when n_1 is small. One widely used approach is to construct modified statistics by replacing \mathbf{W}_2^{-1} with an appropriate substitute. This method was pioneered by Bai and Saranadasa [1996] and Chen and Qin [2010] for two-sample mean tests and has been applied to MANOVA by Srivastava and Fujikoshi [2006], Yamada and Himeno [2015], and Hu et al. [2017]. Another approach involves regularizing \mathbf{W}_2 to address its near-singularity. Chen et al. [2014] proposed introducing a ridge term to \mathbf{W}_2 for two-sample mean tests, a concept further developed by Li et al. [2020b]. This idea was also extended to GLHT and more flexible forms of regularization by Li et al. [2020a]. Other notable contributions in this area include the works of He et al. [2021], Lopes et al. [2011], and Wang et al. [2015].

This paper aims to develop a high-dimensional largest root test that remains applicable when p , n_1 and n_2 are comparable. To address the challenge posed by the singularity of \mathbf{W}_2 when $p > n_2$, we introduce a ridge regularization framework. Specifically, inspired by

the work of Li et al. [2020b], we propose a family of ridge-regularized F -matrices:

$$\mathbf{F}_\lambda = \mathbf{W}_1 \left(\mathbf{W}_2 + \lambda I_p \right)^{-1} = \left(\frac{1}{n_1} \Sigma_p^{1/2} \mathbf{Z} P_1 \mathbf{Z}^T \Sigma_p^{1/2} \right) \left(\frac{1}{n_2} \Sigma_p^{1/2} \mathbf{Z} P_2 \mathbf{Z}^T \Sigma_p^{1/2} + \lambda I_p \right)^{-1}, \quad (4)$$

where $\lambda > 0$ is a regularization parameter. The largest eigenvalue of \mathbf{F}_λ , denoted by $\ell_{\max}(\mathbf{F}_\lambda)$, serves as the test statistic, designed to detect departures from the null hypothesis.

Importantly, the regularized F -matrix and its largest root are *rotation-invariant*, meaning they remain unchanged under arbitrary orthogonal transformations of the data. This invariance arises from the fact that the regularized matrix $\mathbf{W}_2 + \lambda I_p$ preserves the eigenstructure of \mathbf{W}_2 . This property is particularly advantageous in settings where additional structural knowledge, such as sparsity, is limited. It ensures that the test's performance does not depend on the specific coordinate system in which the data are observed. Consequently, the test remains robust across different data representations, making it broadly applicable in various high-dimensional scenarios.

In this paper, we focus on the regime where p , n_1 , and n_2 are all comparable (see **C1** in Section 2). Under this regime, we establish the asymptotic Tracy-Widom distribution of \mathbf{F}_λ under the null hypothesis after appropriate scaling. We propose consistent estimators for the scaling parameters, which rely exclusively on the eigenvalues of \mathbf{W}_2 . The proposed framework is then applied to representative double Wishart problems.

The main contributions of this work are as follows. First, we extend the existing analysis of high dimensional double Wishart problems to the regime where n_1 is comparable to n_2 while also allowing p exceed n_2 . Second, to the best of our knowledge, this is the first work to investigate largest root-type tests in the presence of statistical regularization. Third, from an *Random Matrix Theory* (RMT) perspective, this work is the first to establish the asymptotic Tracy-Widom distribution for the extreme eigenvalue of a regu-

larized random matrix. Fourth, we propose an estimation procedure for the complicated parameters involved in the asymptotic Tracy-Widom distribution of extreme eigenvalues of a random matrix. These types of parameters frequently appear in the RMT literature, yet their estimation has remained challenging.

The paper is organized as follows. After giving necessary preliminaries in RMT, Section 2 introduces the asymptotic Tracy-Widom distribution for the ridge-regularized largest root test statistics after appropriate scaling. Section 3 details the estimation method for the scaling parameters, which relies solely on the eigenvalues of \mathbf{W}_2 , and establishes the consistency of the proposed approach. Applications to GLHT and TECM are discussed in Section 4, while Section 5 reports simulation results. In Section 6, we apply the proposed procedure to the *Human Connectome Project* data. Finally, Section 7 summarizes the main findings of the paper and outlines future research directions. Technical details and proofs of the main results are provided in the Supplementary Material.

2 Asymptotic theory

After giving necessary preliminaries on Random Matrix Theory, the asymptotic theory of the proposed ridge-regularized largest root is presented in this section. First of all, note that \mathbf{F}_λ is asymmetric, making it technically more convenient to analyze a symmetric dual matrix of the form $\tilde{\mathbf{F}}_\lambda = U_1^T \mathbf{Z}^T \mathbf{G}_\lambda^{-1} \mathbf{Z} U_1$, where

$$\mathbf{G}_\lambda = \mathbf{Z} U_2 U_2^T \mathbf{Z}^T + \lambda \Sigma_p^{-1}.$$

Here, U_1 and U_2 are $p \times n_1$ and $p \times n_2$ scaled orthogonal matrices satisfying $U_1 U_1^T = n_1^{-1} P_1$ and $U_2 U_2^T = n_2^{-1} P_2$. It is straightforward to see that $\tilde{\mathbf{F}}_\lambda$ shares the same nonzero eigenvalues

as \mathbf{F}_λ . Thus, for technical convenience, we focus our analysis on $\tilde{\mathbf{F}}_\lambda$.

Definition 2.1 (Empirical Spectral Distribution). *For any $p \times p$ matrix A with real-valued eigenvalues, denote $\ell_j(A)$ to be the j -th largest eigenvalue of A and specifically $\ell_{\max}(A) = \ell_1(A)$ and $\ell_{\min}(A) = \ell_p(A)$. Define the Empirical Spectral Distribution (ESD) F^A of A by*

$$F^A(\tau) = \frac{1}{p} \sum_{j=1}^p \delta_{\ell_j(A)}(\tau),$$

where $\delta_x(y) = 1$ if $x \leq y$ and 0 otherwise.

Under the framework (1), the following basic assumptions are employed.

- C1** (High-dimensional regime) We assume p, n_T, n_1, n_2 grow to infinity simultaneously in the sense that $p/n_T \rightarrow \gamma_T$, $p/n_1 \rightarrow \gamma_1$, $p/n_2 \rightarrow \gamma_2$, where $\gamma_T, \gamma_1, \gamma_2 \in (0, \infty)$ are constants. Also, $n_1 + n_2 \leq n_T$. Moreover, we assume the convergence rate is such that $p^{2/3}|p/n_1 - \gamma_1| \rightarrow 0$ and $p^{2/3}|p/n_2 - \gamma_2| \rightarrow 0$.
- C2** (Moments conditions) $\mathbb{E}|z_{ij}|^k < \infty$ for any $k \geq 2$.
- C3** (Boundedness of spectrum) $\liminf_{p \rightarrow \infty} \ell_{\min}(\Sigma_p) > 0$ and $\limsup_{p \rightarrow \infty} \ell_{\max}(\Sigma_p) < \infty$.
- C4** (Asymptotic stability of population ESD) There exists a deterministic distribution F^{Σ_∞} with compact support in $(0, \infty)$, such that

$$p^{2/3} \mathcal{D}_W(F^{\Sigma_p}, F^{\Sigma_\infty}) \rightarrow 0, \quad \text{as } p \rightarrow \infty,$$

where $\mathcal{D}_W(F_1, F_2)$ denotes the *Wasserstein distance* between distributions F_1 and F_2 , defined as

$$\mathcal{D}_W(F_1, F_2) = \sup_f \left\{ \left| \int f dF_1 - \int f dF_2 \right| : f \text{ is 1-Lipschitz} \right\}.$$

C5 (Asymptotic stability of edge eigenvalues) Denote the rightmost edge point of the support of F^{Σ_∞} to be σ_{\max} . That is, $\sigma_{\max} = \inf\{x \in \mathbb{R} : F^{\Sigma_\infty}(x) = 1\}$. We assume

$$n^{2/3}|\ell_{\max}(\Sigma_p) - \sigma_{\max}| \rightarrow 0.$$

Additionally, if σ_{\max} is a point of discontinuity of F^{Σ_∞} , we assume $F^{\Sigma_p}(\{\ell_{\max}(\Sigma_p)\}) \rightarrow F^{\Sigma_\infty}(\{\sigma_{\max}\})$.

C6 (Regular edge) We assume F^{Σ_∞} is regular near σ_{\max} in the sense of Definition 2.7 (see Section 2.2).

C1 defines the asymptotic regime where dimensionality p , the sample size n_T , degrees of freedom n_1 , and n_2 grow proportionately. As stated in **C2**, our asymptotic results only require moment conditions. **C3** is a standard assumption in the RMT literature, ensuring that eigenvalue bounds are obtainable. **C4** ensures that the bulk spectrum of Σ_p characterized by its ESD stabilizes at a deterministic limit as p increases. While **C4** offers limited insight into the edge eigenvalue, **C5** ensures that the largest eigenvalue of Σ_p also stabilizes at the rightmost support edge of F^{Σ_∞} . The details of **C6** are given in Section 2.2.

It is worth mentioning that **C5** excludes the possibility that a fixed number of eigenvalues of Σ_p remain outside the support of F^{Σ_∞} . This situation arises in classical factor models. While we conjecture that our analysis can be extended to a class of factor models, such a generalization is beyond the scope of this work. Instead, to assess the performance of the proposed method when these assumptions are violated, we include a factor model setting in our numerical study in Section 5.

2.1 Preliminary results

In this section, we present key preliminary results in RMT, focusing on the asymptotic Marčenko-Pastur law for the ESD of \mathbf{G}_λ and the matrix

$$\widetilde{\mathbf{W}}_2 = U_2^T \mathbf{Z}^T \Sigma_p \mathbf{Z} U_2.$$

The matrix $\widetilde{\mathbf{W}}_2$ is the companion matrix of \mathbf{W}_2 in the sense that $\widetilde{\mathbf{W}}_2$ shares the same nonzero eigenvalues as \mathbf{W}_2 , differing only by $|p - n_2|$ additional zeros in their spectra. Consequently, the ESD of $\widetilde{\mathbf{W}}_2$ and \mathbf{W}_2 can be easily reconstructed from one another. For mathematical convenience, we focus on the asymptotics of $F^{\widetilde{\mathbf{W}}_2}$ rather than $F^{\mathbf{W}_2}$ in the subsequent analysis, as it results in cleaner and more tractable expressions. Results in this section are based on Silverstein and Bai [1995] and Bai and Silverstein [2004].

Recall that the *Stieltjes transform* $\varphi(\cdot)$ of any function F of bounded variation on \mathbb{R} is defined as:

$$\varphi(z) = \int_{-\infty}^{\infty} \frac{dF(x)}{x - z}, \quad z \in \mathbb{C}^+ := \{E + i\eta : \eta > 0\}.$$

Note that φ is a mapping from \mathbb{C}^+ to \mathbb{C}^+ . The Stieltjes transform plays a role in RMT analogous to that of the Fourier transform in classical probability theory. It allows for the reconstruction of a distribution function via an inversion formula and facilitates the study of probability measure convergence through the convergence of its Stieltjes transform. For further details, see Bai and Silverstein [2010].

Given a compactly supported probability measure F and any $\gamma > 0$, the renowned Marčenko-Pastur equation in RMT defined on $z \in \mathbb{C}^+$ is expressed as

$$z = -\frac{1}{\varphi} + \gamma \int \frac{\tau dF(\tau)}{\tau\varphi + 1}. \tag{5}$$

Lemma 2.2. *Under the conditions on F and γ , for any $z \in \mathbb{C}^+$, there exists a unique solution $\varphi \equiv \varphi(z) \in \mathbb{C}^+$ to Eq. (5). The function $\varphi(z)$ is the Stieltjes transform of a probability measure with bounded support in $[0, \infty)$.*

Unfortunately, except in extreme cases where F has a simple structure, there is no explicit closed-form solution for $\varphi_{\gamma, F}(z)$ or the associated probability distribution. For further details of Marčenko-Pastur equation, see Chapter 3 of Bai and Silverstein [2010].

Lemma 2.3 (Marčenko-Pastur Theorem). *Assume that conditions **C1**–**C4** hold. As $p \rightarrow \infty$, the ESD $F^{\widetilde{\mathbf{W}}_2}$ converges pointwise almost surely to \mathcal{W} at all points of continuity of \mathcal{W} , where \mathcal{W} is the probability measure associated with $\varphi_{\gamma_2, F^{\Sigma_\infty}}$.*

Moreover, for any $z \in \mathbb{C}^+$, let $\hat{\varphi}(z)$ be the Stieltjes transform of $F^{\widetilde{\mathbf{W}}_2}$. That is,

$$\hat{\varphi}(z) = \int \frac{dF^{\widetilde{\mathbf{W}}_2}(\tau)}{\tau - z} = \frac{1}{n_2} \sum_{j=1}^{n_2} \frac{1}{\ell_j(\widetilde{\mathbf{W}}_2) - z}.$$

For any closed and bounded region $\mathcal{Z} \subset \mathbb{C}^+$ and any fixed $k \in \mathbb{N}$, we have

$$\sup_{z \in \mathcal{Z}} p^{2/3} |\hat{\varphi}^{(k)}(z) - \varphi^{(k)}(z)| \xrightarrow{P} 0.$$

Following immediately from Lemma 2.2, the Marčenko-Pastur equation (5) holds approximately at $\hat{\varphi}(z)$ in the sense that

$$z + \frac{1}{\hat{\varphi}(z)} - \frac{p}{n_2} \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\tau \hat{\varphi}(z) + 1} \xrightarrow{P} 0, \quad (6)$$

uniformly for $z \in \mathcal{Z}$. This result serves as the foundation for the estimation method proposed in Section 3.

Next, we analyze the limiting ESD of \mathbf{G}_λ . For any fixed $\lambda > 0$, consider the following

equation defined for $z \in \mathbb{C}^+$

$$\phi = \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\tau\{[1 + \gamma_2\phi]^{-1} - z\} + \lambda}. \quad (7)$$

Lemma 2.4. *Let F^{Σ_∞} be any compactly supported probability measure that is non-degenerate to zero. For any given $z \in \mathbb{C}^+$, there exists a unique solution $\phi \equiv \phi(z) \in \mathbb{C}^+$ to Eq. (7). The function $\phi(z)$ is the Stieltjes transform of a probability measure with bounded support in $(0, \infty)$, henceforth denoted by \mathcal{G}_λ .*

Theorem 2.5 (Generalized Marčenko-Pastur Theorem of \mathbf{G}_λ). *Suppose **C1**–**C4** hold. Fix $\lambda > 0$. As $p \rightarrow \infty$, the ESD $F^{\mathbf{G}_\lambda}$ of \mathbf{G}_λ converges pointwise almost surely to \mathcal{G}_λ at all points of continuity of \mathcal{G}_λ .*

While Eq. (7) plays a fundamental role in the subsequent analysis, we present an equivalent reformulation that is more convenient to work with in certain cases. For any $z \in \mathbb{C}^+$, define

$$h = h(z) = z - [1 + \gamma_2\phi(z)]^{-1}.$$

Then, Eq. (7) can be rewritten as

$$z = h + \left(1 + \gamma_2 \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\lambda - \tau h}\right)^{-1}. \quad (8)$$

By Lemma 2.4, it follows that for any $z \in \mathbb{C}^+$, there exists a unique solution $h \in \mathbb{C}^+$ to Eq. (8).

Lastly, we consider the extension of $\phi(z)$ to the real line.

Lemma 2.6. *Under the conditions of Lemma 2.4, the measure \mathcal{G}_λ is compactly supported in $(0, \infty)$. Denote the leftmost edge of the support of \mathcal{G}_λ by $\rho \equiv \rho_\lambda$, i.e., $\rho = \sup\{x \in \mathbb{R} :$*

$\mathcal{G}_\lambda(x) = 0\}$. The following results hold:

(i) We have $\rho > 0$, and for any $x \in [0, \rho)$,

$$\lim_{z \in \mathbb{C}^+ \rightarrow x} \phi(z) := s(x) \quad \text{exists and is real-valued.}$$

Moreover, the following relationship holds:

$$s(x) = \int \frac{d\mathcal{G}_\lambda(\tau)}{\tau - x}, \quad x \in [0, \rho). \quad (9)$$

(ii) From Eq (9), $s(x)$ is analytic on $[0, \rho)$, and its k th derivative satisfies $s^{(k)}(x) > 0$ for any $k \in \mathbb{N}$ and $x \in [0, \rho)$. This implies that $s^{(k)}(x)$ is monotonically increasing on $[0, \rho)$ for all k .

2.2 Asymptotic Tracy-Widom distribution

In this section, we present the asymptotic Tracy-Widom distribution for the proposed ridge-regularized largest root statistic $\ell_{\max}(\mathbf{F}_\lambda)$ (or equivalently $\ell_{\max}(\tilde{\mathbf{F}}_\lambda)$) after appropriate scaling. All quantities introduced in this section may depend on $(\gamma_1, \gamma_2, F^{\Sigma_\infty}, \lambda)$. To simplify notation and reduce complexity, we omit these arguments when there is no risk of ambiguity.

We first impose an edge regularity condition on F^{Σ_∞} . Recall that σ_{\max} is the rightmost edge point of the support of F^{Σ_∞} . For $h \in (-\infty, \lambda/\sigma_{\max})$, define

$$x(h) = h + \left[1 + \gamma_2 \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\lambda - \tau h} \right]^{-1}. \quad (10)$$

It can be viewed as an extension of Eq. (8) when $h \in \mathbb{C}^+$ approaches the real line.

Definition 2.7. We say that F^{Σ_∞} is regular near σ_{\max} if any of the following holds:

- (i) F^{Σ_∞} is discontinuous at σ_{\max} and $\omega_{\max}\gamma_2 \neq 1$, where $\omega_{\max} = F^{\Sigma_\infty}(\{\sigma_{\max}\})$.
- (ii) F^{Σ_∞} is continuous at σ_{\max} , and there exists $h_0 < \lambda/\sigma_{\max}$ such that $x'(h_0) = 0$.

We claim that the regularity conditions are mild in the sense that F^{Σ_p} is discrete for any finite p , and any continuous probability measure can be well-approximated by a discrete one. Under this condition, \mathcal{G}_λ is *well-behaved* near ρ in the following sense: either ρ is a discrete point of \mathcal{G}_λ , or the density $f_{\mathcal{G}}$ of \mathcal{G}_λ exhibits a *square-root behavior* near ρ . The former case occurs when $\omega_{\max}\gamma_2 > 1$; otherwise, the latter applies. Specifically, in the latter case, there exist positive constants C_1 , C_2 , and ϵ such that

$$C_1\sqrt{x-\rho} \leq f_{\mathcal{G}}(x) \leq C_2\sqrt{x-\rho}, \quad \text{for } x \in (\rho, \rho + \epsilon).$$

Similar regularity conditions are widely used in the RMT literature when studying the Tracy-Widom fluctuations of extreme eigenvalues, albeit in various forms. For example, see Definition 2.7 of Knowles and Yin [2017], Condition (1) of El Karoui [2007], and Condition (2.12) of Lee and Schnelli [2016]. Furthermore, the square-root behavior near the edge of the spectrum is broadly observed across various random matrix models. This behavior is closely associated with Tracy-Widom fluctuations and plays a critical role in their characterization.

The following lemma follows from the behavior of \mathcal{G}_λ near ρ .

Lemma 2.8. Suppose the conditions in Lemma 2.6 hold, and further assume that F^{Σ_∞} is regular near σ_{\max} . Then, $s'(x) \rightarrow \infty$ as $x \uparrow \rho$.

Our main theorem is presented as follows.

Theorem 2.9. *Suppose that **C1**–**C6** hold. For any fixed $\lambda > 0$, as $p \rightarrow \infty$, the asymptotic distribution of the largest eigenvalue of \mathbf{F}_λ (or equivalently $\tilde{\mathbf{F}}_\lambda$) is given by*

$$\frac{p^{2/3}}{\Theta_2} \left(\ell_{\max}(\mathbf{F}_\lambda) - \Theta_1 \right) \Longrightarrow \text{TW}_1, \quad (11)$$

where TW_1 denotes the Tracy-Widom distribution of type 1, and \Longrightarrow indicates convergence in distribution. The centering and scaling parameters Θ_1 and Θ_2 are determined as follows. Define β to be the unique solution in $(0, \rho)$ to

$$\beta^2 s'(\beta) = \frac{1}{\gamma_1}. \quad (12)$$

By Lemma 2.8, such a solution β exists and is unique. Then,

$$\Theta_1 = \frac{1}{\beta} \left[1 + \gamma_1 \beta s(\beta) \right], \quad (13)$$

$$\Theta_2 = \left[\frac{\gamma_1^3}{2} s''(\beta) + \frac{\gamma_1^2}{\beta^3} \right]^{1/3}. \quad (14)$$

Remark 2.10. *For detailed information on Tracy-Widom distributions, we refer readers to Tracy and Widom [1994], Tracy and Widom [1996], and Johnstone [2008]. Available software for working with Tracy-Widom laws includes the **R** package **RMTstat** [Johnstone et al., 2022] and the **MATLAB** package **RMLab** [Dieng, 2006].*

Remark 2.11. *Unlike the $O(p)$ scaling observed in classical extreme value theory for i.i.d. random variables with appropriate tail behavior, the normalized largest eigenvalue of \mathbf{F}_λ fluctuates on the $O(p^{2/3})$ scale. An analogous scaling has been established for the largest eigenvalue of various random matrices, such as those in the Jacobi orthogonal ensemble [Johnstone, 2008]. This $O(p^{2/3})$ fluctuation reflects the repulsion between eigenvalues,*

causing them to be more regularly spaced than i.i.d. random variables.

Remark 2.12. When $\lambda = 0$ and $\gamma_2 < 1$, the matrix \mathbf{F}_0 reduces to the classical F -matrix. Han et al. [2018] established the asymptotic Tracy-Widom distribution for the largest eigenvalue of an F -matrix in that setting. Theorem 2.9 can be viewed as a generalization of Han et al. [2018] to $\lambda > 0$ and an arbitrary covariance matrix Σ_p .

Remark 2.13. Theorem 2.9 can be generalized to the joint asymptotic distribution of the first k largest eigenvalues of \mathbf{F}_λ . Specifically, under **C1–C6**, for any fixed integer k and $\lambda > 0$, we have

$$\begin{aligned} & \lim_{p \rightarrow \infty} \mathbb{P} \left(\frac{p^{2/3}}{\Theta_2} \left(\ell_1(\mathbf{F}_\lambda) - \Theta_1 \right) \leq t_1, \dots, \frac{p^{2/3}}{\Theta_2} \left(\ell_k(\mathbf{F}_\lambda) - \Theta_1 \right) \leq t_k \right) \\ &= \lim_{p \rightarrow \infty} \mathbb{P} \left(\frac{p^{2/3}}{\gamma_1^{2/3}} \left(\ell_1^{\text{GOE}} - 2 \right) \leq t_1, \dots, \frac{p^{2/3}}{\gamma_1^{2/3}} \left(\ell_k^{\text{GOE}} - 2 \right) \leq t_k \right), \end{aligned} \quad (15)$$

for any $t_1, \dots, t_k \in \mathbb{R}$. Here, ℓ_k^{GOE} denotes the k -th largest eigenvalue of a $p \times p$ matrix from the Gaussian Orthogonal Ensemble (GOE).

3 Estimation

In this section, we develop an estimation procedure for the centering and scaling parameters Θ_1 and Θ_2 . The proposed estimators leverage the eigenvalues of \mathbf{W}_2 and are designed to be computationally efficient while maintaining robustness in high-dimensional settings. This section is organized as follows. Section 3.1 introduces a reformulation of the problem as an optimization task. The proposed algorithm is described in detail in Section 3.2. Section S.1 provides implementation details and discusses practical considerations for the proposed method. Finally, in Section 3.3, we establish the consistency of the proposed estimators. Throughout this section, the parameters γ_1 and γ_2 are estimated as $\hat{\gamma}_1 = p/n_1$

and $\hat{\gamma}_2 = p/n_2$, respectively, without further elaboration.

3.1 Problem formulation

The estimation of Θ_1 and Θ_2 necessitates the precise estimation of the edge ρ and the function $s(x)$ along with its derivatives on the interval $[0, \rho)$. Although $s(x)$ is formally defined as the extension of the Stieltjes transform $\phi(z)$ from \mathbb{C}^+ to the real line as in Lemma 2.6, this definition provides limited practical utility due to the absence of an explicit analytical expression. In this section, we rigorously demonstrate that both ρ and $s(x)$ can be characterized by Eq. (10), thereby establishing a foundation for their accurate estimation.

Lemma 3.1. *Assume that F^{Σ_∞} is regular near σ_{\max} as defined in Definition 2.7. If there exists $h_0 < \lambda/\sigma_{\max}$ such that $x'(h_0) = 0$, then the edge ρ is given by $\rho = x(h_0) > \lambda/\sigma_{\max}$; otherwise, $\rho = \lambda/\sigma_{\max}$.*

Corollary 3.2. *Consider the case where σ_{\max} is a point of discontinuity of F^{Σ_∞} , and recall that $\omega_{\max} = F^{\Sigma_\infty}(\{\sigma_{\max}\})$. The edge ρ is determined as follows:*

- (i) *If $\gamma_2\omega_{\max} < 1$, there exists $h_0 < \lambda/\sigma_{\max}$ such that $x'(h_0) = 0$. In this case, $\rho = x(h_0) > \lambda/\sigma_{\max}$, and ρ is a point of continuity of \mathcal{G}_λ .*
- (ii) *If $\gamma_2\omega_{\max} > 1$, we have $x'(h) > 0$ for any $h \in (0, \lambda/\sigma_{\max})$. We have $\rho = \lambda/\sigma_{\max}$, and ρ is a point of discontinuity of \mathcal{G}_λ .*

Define the auxiliary functions on $h \in (-\infty, \lambda/\sigma_{\max})$ as

$$\mathcal{H}_j(h) = \int \frac{\tau^j dF^{\Sigma_\infty}(\tau)}{(\lambda - \tau h)^j}, \quad j = 1, 2, 3.$$

With the definition, $x(h) = h + [1 + \gamma_2\mathcal{H}_1(h)]^{-1}$ and $x'(h) = 1 - \gamma_2[1 + \gamma_2\mathcal{H}_1(h)]^{-2}\mathcal{H}_2(h)$.

Theorem 3.3. *Assume that F^{Σ_∞} is regular near σ_{\max} as defined in Definition 2.7. Fix any $x_0 \in [0, \rho)$. Then, $s(x_0)$ is determined as*

$$s(x_0) = \frac{1}{\gamma_2} \left(\frac{1}{x_0 - h_0} - 1 \right),$$

where h_0 is the unique solution in $(-\infty, \lambda/\sigma_{\max})$ to

$$\begin{cases} x_0 = h_0 + [1 + \gamma_2 \mathcal{H}_1(h_0)]^{-1}, \\ 1 - \gamma_2 [1 + \gamma_2 \mathcal{H}_1(h_0)]^{-2} \mathcal{H}_2(h_0) > 0. \end{cases} \quad (16)$$

Furthermore, the derivatives $s'(x_0)$ and $s''(x_0)$ are determined as:

$$\begin{aligned} g(x_0) &= x_0 - \frac{1}{1 + \gamma_2 s(x_0)}, \\ \zeta(x_0) &= \frac{\gamma_2}{(1 + \gamma_2 s(x_0))^2}, \end{aligned} \quad (17)$$

$$s'(x_0) = \mathcal{H}_2(g(x_0)) (\zeta(x_0) s'(x_0) + 1),$$

$$s''(x_0) = 2 (\zeta(x_0) s'(x_0) + 1)^2 \mathcal{H}_3(g(x_0)) + \mathcal{H}_2(g(x_0)) \left[\frac{-2(\gamma_2 s'(x_0))^2}{(1 + \gamma_2 s(x_0))^3} + \zeta(x_0) s''(x_0) \right].$$

It is worth noting that the third equation in (17) is linear in $s'(x_0)$, given $s(x_0)$ and $\mathcal{H}_2(\cdot)$.

Similarly, the last equation in (17) is linear in $s''(x_0)$, given $s(x_0)$, $s'(x_0)$, $\mathcal{H}_2(\cdot)$, and $\mathcal{H}_3(\cdot)$.

The following lemma is useful for the numerical computation of $s(x_0)$.

Lemma 3.4. *Under the conditions of Theorem 3.3, the derivative $x'(h) = 1 - \gamma_2 [1 + \gamma_2 \mathcal{H}_1(h)]^{-2} \mathcal{H}_2(h)$ is monotonically decreasing on $(-\infty, \lambda/\sigma_{\max})$.*

This result implies that $x(h)$ is either strictly increasing on $(-\infty, \lambda/\sigma_{\max})$ or exhibits a single concave-down peak within this interval. If $x_0 = h + [1 + \gamma_2 \mathcal{H}_1(h)]^{-1}$ has two distinct roots, Eq. (16) indicates that h_0 corresponds to the smaller root.

Through the results in Theorem 3.3, we can transform the problem of estimating $s(x)$, $s'(x)$, and $s''(x)$ into the task of estimating $\mathcal{H}_j(h)$ for $j = 1, 2, 3$ over $h \in (-\infty, \lambda/\sigma_{\max})$. To this end, we revisit Lemma 2.3 and Eq. (6), which together suggest that:

$$\begin{aligned}\mathcal{H}_1(-\lambda\hat{\varphi}(z)) &= \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\lambda + \tau\lambda\hat{\varphi}(z)} \approx \frac{z}{\lambda\hat{\gamma}_2} + \frac{1}{\lambda\hat{\gamma}_2\hat{\varphi}(z)}, \\ \mathcal{H}_2(-\lambda\hat{\varphi}(z)) &= \int \frac{\tau^2 dF^{\Sigma_\infty}(\tau)}{(\lambda + \tau\lambda\hat{\varphi}(z))^2} \approx \frac{1}{\lambda^2\hat{\gamma}_2\hat{\varphi}^2(z)} - \frac{1}{\lambda^2\hat{\gamma}_2\hat{\varphi}'(z)}.\end{aligned}$$

Here, we are extending the definition of $\mathcal{H}_j(h)$ to the complex domain. The second result is obtained by differentiating both sides of Eq. (6). Notably, $\hat{\varphi}(z)$ solely relies on the eigenvalues of \mathbf{W}_2 as

$$\hat{\varphi}(z) = \frac{1}{n_2} \sum_{j=1}^{n_2} \frac{1}{\ell_j(\mathbf{W}_2) - z}, \quad \text{with } \ell_j(\mathbf{W}_2) = 0 \text{ if } j > p.$$

This observation motivates the following strategy for estimating $\mathcal{H}_j(y)$ for $j = 1, 2, 3$.

For general F^{Σ_∞} , it can be approximated by a weighted sum of point masses, expressed as:

$$F^{\Sigma_\infty}(\tau) \simeq \sum_{k=1}^K w_k \delta_{\sigma_k}(\tau),$$

where $\{\sigma_k\}_{k=1}^K$ is a grid of points chosen to densely cover the support of F^{Σ_∞} , and w_k are weights satisfying $w_k \geq 0$ and $\sum_{k=1}^K w_k = 1$. The grid $\{\sigma_k\}_{k=1}^K$ is user-selected and fixed throughout the estimation procedure, while the weights w_k are model parameters. This approach was introduced in El Karoui [2008] and later generalized in Ledoit and Wolf [2012]. The recommended choice of the grid $\{\sigma_k\}_{k=1}^K$ is discussed in Section S.1. Accordingly, the

integrals can be approximated as:

$$\mathcal{H}_j(h) = \int \frac{\tau^j dF^{\Sigma_\infty}(\tau)}{(\lambda - \tau h)^j} \simeq \sum_{k=1}^K \frac{w_k \sigma_k^j}{(\lambda - \sigma_k h)^j} := \tilde{\mathcal{H}}_j(h, w_1, \dots, w_K).$$

For $z \in \mathbb{C}^+$, define:

$$\hat{Q}_1(z) = \frac{z}{\lambda \hat{\gamma}_2} + \frac{1}{\lambda \hat{\gamma}_2 \hat{\varphi}(z)}, \quad \text{and} \quad \hat{Q}_2(z) = \frac{1}{\lambda^2 \hat{\gamma}_2 \hat{\varphi}^2(z)} - \frac{1}{\lambda^2 \hat{\gamma}_2 \hat{\varphi}'(z)}.$$

Consider a grid of points z in \mathbb{C}^+ , denoted by $\{z_i\}_{i=1}^I$. The recommended choice of $\{z_i\}_{i=1}^I$ is provided in Section S.1 of the Supplementary Materials. We select the weights w_k such that $\hat{Q}_j(z_i)$ approximately matches $\tilde{\mathcal{H}}_j(-\lambda \hat{\varphi}(z_i), w_1, \dots, w_K)$ for all $i = 1, 2, \dots, I$, and $j = 1, 2$. Further details are presented in the following sections.

3.2 The algorithm

Given the grids $\{\sigma_k\}_{k=1}^K$ and $\{z_i\}_{i=1}^I$, define the loss function as

$$L_\infty(w_1, \dots, w_K) = \max_{i=1, \dots, I} \max_{j=1, 2} \{|\Re(e_{ij})|, |\Im(e_{ij})|\}, \quad \text{where}$$

$$e_{ij} = \frac{\hat{Q}_j(z_i) - \tilde{\mathcal{H}}_j(-\lambda \hat{\varphi}(z_i), w_1, \dots, w_K)}{|\hat{Q}_j(z_i)|}, \quad i = 1, \dots, I; j = 1, 2.$$

We choose the weights as

$$(w_1^*, \dots, w_K^*) = \arg \min_{w_j \geq 0, \sum w_j = 1} L_\infty(w_1, \dots, w_K).$$

The main motivation for choosing the L_∞ type loss is that given the point masses $\{\sigma_k\}_{k=1}^K$,

Algorithm 1 Linear Programming Formulation for Selecting Weights

Input: A grid of masses $\sigma_1, \dots, \sigma_K$; a grid of points z_1, \dots, z_I ; eigenvalues ℓ_j 's of \mathbf{W}_2 ; regularization parameter λ .

Calculation: Compute

$$e_{ij} = \frac{\hat{Q}_j(z_i)}{|\hat{Q}_j(z_i)|} - \frac{1}{|\hat{Q}_j(z_i)|} \sum_{k=1}^K \frac{\sigma_k^j w_k}{(\sigma_k \lambda \hat{\varphi}(z_i) + \lambda)^j}, \quad i = 1, \dots, I; j = 1, 2,$$

Optimization: Find the optimal $(w_1, \dots, w_K, \theta)$ through the linear programming

$$\begin{aligned} & \underset{\theta, w_1, \dots, w_K}{\text{minimize}} && \theta \\ & \text{subject to} && -\theta \leq \Re(e_{ij}) \leq \theta, \quad \forall i = 1, \dots, I; j = 1, 2, \\ & && -\theta \leq \Im(e_{ij}) \leq \theta, \quad \forall i = 1, \dots, I; j = 1, 2, \\ & && w_k \geq 0, \quad \forall k = 1, \dots, K, \\ & && \text{and } w_1 + w_2 + \dots + w_K = 1. \end{aligned}$$

Truncation: To mitigate the effect of floating-point errors, truncate the optimal weight w_k to $w_k \times \mathbb{1}(w_k > K^{-1}10^{-d})$ and rescale the truncated weights to be of sum one. The recommended choice of d is $d = 2$.

Output: All positive weights $\hat{w}_1, \dots, \hat{w}_B$, the associated masses $\hat{\sigma}_1, \dots, \hat{\sigma}_B$ and loss θ .

the optimization can be formulated into a linear programming problem as presented in Algorithm 1.

Suppose the optimal positive weights given by Algorithm 1 are $\hat{w}_1, \dots, \hat{w}_B$ and the associated point masses are $\hat{\sigma}_1 > \hat{\sigma}_2 > \dots > \hat{\sigma}_B$. We estimate $\mathcal{H}_j(h)$ by

$$\hat{\mathcal{H}}_j(h) = \sum_{k=1}^B \frac{\hat{w}_k \hat{\sigma}_k^j}{(\lambda - \hat{\sigma}_k h)^j}, \quad j = 1, 2, 3, \quad 0 < h < \lambda/\hat{\sigma}_1.$$

Given $\hat{\mathcal{H}}_j(\cdot)$, $j = 1, 2, 3$, the estimation of $s(x)$ and its derivatives over the domain $[0, \rho)$ can be performed efficiently by leveraging its inherent smoothness. First of all, following Corollary 3.2, the estimator $\hat{\rho}$ of ρ is obtained in Algorithm 2. Secondly, note that (17) defines an ordinary differential equation (ODE) system. We propose to estimate $s(x)$, $s'(x)$, $s''(x)$ by solving this ODE, as detailed in Algorithm 3.

Following Algorithm 2 and Algorithm 3, we estimate β by $\hat{\beta}$, which is the unique solution

Algorithm 2 Estimation of ρ

Input: $\lambda, \hat{\mathcal{H}}_1(\cdot), \hat{\mathcal{H}}_2(\cdot)$, largest point-mass $\hat{\sigma}_1$, the associated weight \hat{w}_1 ;

Case 1: If $\hat{\gamma}_2 \hat{w}_1 \geq 1$, $\hat{\rho} = \lambda / \hat{\sigma}_1$;

Case 2: Otherwise, solve for the unique root \hat{h} in $(-\infty, \lambda / \hat{\sigma}_1)$ to

$$\frac{\hat{\gamma}_2 \hat{\mathcal{H}}_2(h)}{(1 + \hat{\gamma}_2 \hat{\mathcal{H}}_1(h))^2} = 1.$$

The function on the left-hand side is monotonically increasing on $(-\infty, \lambda / \hat{\sigma}_1)$. The equation can be solved using Newton's method with an initial value pre-selected through grid search. Then, $\hat{\rho} = \hat{h} + [1 + \hat{\gamma}_2 \hat{\mathcal{H}}_1(-\hat{h})]^{-1}$.

Output: Estimated edge $\hat{\rho}$.

on $(0, \hat{\rho})$ to $\beta^2 \hat{s}'(\beta) = 1 / \hat{\gamma}_1$. Then, Θ_1 and Θ_2 are estimated as:

$$\hat{\Theta}_1 = \frac{1}{\hat{\beta}} \left[1 + \hat{\gamma}_1 \hat{\beta} \hat{s}(\hat{\beta}) \right] \quad \text{and} \quad \hat{\Theta}_2 = \left[\frac{\hat{\gamma}_1^3}{2} \hat{s}''(\hat{\beta}) + \frac{\hat{\gamma}_1^2}{\hat{\beta}^3} \right]^{1/3}. \quad (18)$$

Algorithm 3 ODE Formulation For Estimation of $s(x)$, $s'(x)$, and $s''(x)$

Input: $\hat{\rho}, \hat{\mathcal{H}}_1(\cdot), \hat{\mathcal{H}}_2(\cdot), \hat{\mathcal{H}}_3(\cdot)$;

Initial: Find s as the unique solution to $\hat{\mathcal{H}}_1\left(\frac{-1}{1+\hat{\gamma}_2 s}\right) = s$ and $\hat{\gamma}_2 \hat{\mathcal{H}}_2\left(\frac{-1}{1+\hat{\gamma}_2 s}\right) < (1 + \hat{\gamma}_2 s)^2$.

ODE: On $x \in [0, \hat{\rho})$, solve the following ODE with the initial $\hat{s}(0) = s$:

$$\begin{aligned} \hat{s}'(x) &= \frac{\hat{\mathcal{H}}_2(\hat{g}(x))}{1 - \hat{\mathcal{H}}_2(\hat{g}(x))\hat{\zeta}(x)}, \\ \hat{s}''(x) &= \frac{2[\hat{\zeta}(x)\hat{s}'(x) + 1]^2 \hat{\mathcal{H}}_3(\hat{g}(x)) - 2[1 + \hat{\gamma}_2 \hat{s}(x)]^{-3} \hat{\mathcal{H}}_2(\hat{g}(x))(\hat{\gamma}_2 \hat{s}'(x))^2}{1 - \hat{\zeta}(x)\hat{\mathcal{H}}_2(\hat{g}(x))}, \\ \hat{g}(x) &= x - \frac{1}{1 + \hat{\gamma}_2 \hat{s}(x)}, \quad \text{and} \quad \hat{\zeta}(x) = \frac{\hat{\gamma}_2}{(1 + \hat{\gamma}_2 \hat{s}(x))^2}. \end{aligned}$$

Output: Estimated functions: $\hat{s}(x)$, $\hat{s}'(x)$, $\hat{s}''(x)$ on $[0, \hat{\rho})$.

The implementation details of the algorithms are presented in Section S.1 of the Supplementary Materials.

In Algorithm 1, the choice of the L_∞ type loss function leads to a linear programming

problem. Alternative convex loss functions, such as the L_2 -norm, can also be considered, transforming the optimization into a general convex problem. However, our numerical studies indicate that the L_2 -loss provides no significant advantage over the L_∞ -loss. Given that linear programming is computationally more efficient, we recommend the use of the L_∞ -loss. A detailed comparison between different loss functions is beyond the scope of the current work.

Additional constraints can be incorporated into Algorithm 1 to further enhance precision. For instance, to improve the accuracy of $\hat{\mathcal{H}}_3(y)$, we may want to ensure that the second-order derivative of $\hat{Q}_1(z)$ matches those of $\tilde{\mathcal{H}}_1(-\lambda\hat{\varphi}(z))$. Since the derivatives are also linear in the weights w_k , it is straightforward to regularize the problem and add such constraints. Our numerical studies suggest that when $\hat{\gamma}_2$ is relatively small (e.g., $\hat{\gamma}_2 \leq 1$), incorporating constraints on the second-order derivative of $\tilde{\mathcal{H}}_1(-\lambda\hat{\varphi}(z))$ can slightly improve the estimation accuracy of $\hat{\Theta}_1$ and $\hat{\Theta}_2$. However, when $\hat{\gamma}_2$ is relatively large (e.g., $\hat{\gamma}_2 \geq 5$), such constraints tend to reduce estimation precision. This is mainly because the accuracy of $\hat{\varphi}''(z)$, which serve as the estimator of $\varphi''(z)$, diminishes when p becomes much larger than n_2 .

Algorithm 1 is inspired by El Karoui [2008] and Ledoit and Wolf [2012], but it diverges in the choice of basis functions and the design of the loss function. While El Karoui [2008] and Ledoit and Wolf [2012] prioritize achieving smoothness in the estimator of F^{Σ_∞} by employing a combination of smooth basis functions and point masses, our focus is fundamentally different. In the context of estimating $s(x)$, the smoothness of F^{Σ_∞} is not a primary concern. Consequently, we adopt a simpler approach by approximating F^{Σ_∞} solely using a mixture of point masses. Our numerical experiments indicate that this method performs comparably to, if not better than, the smooth basis approach, provided the grid

of σ_k points is sufficiently dense. This finding suggests that the additional complexity introduced by smooth basis functions offers no substantial benefit in this specific setting. As a result, we focus exclusively on point-mass mixture models in our estimation procedure. In addition, the loss functions used by El Karoui [2008] and Ledoit and Wolf [2012] control only the discrepancy between \hat{Q}_1 and $\tilde{\mathcal{H}}_1$. In contrast, our method explicitly penalizes the discrepancy between \hat{Q}_2 and $\tilde{\mathcal{H}}_2$, as accurate estimation of \mathcal{H}_2 is of direct interest.

3.3 Consistency

In this section, we show the consistency of the proposed estimation procedure.

Definition 3.5. For a grid of complex values $J = \{t_1, t_2, t_3, \dots\}$, define the corresponding grid size as $\text{size}(J) = \sup_k |t_k - t_{k-1}|$.

Definition 3.6. A grid R of real numbers is said to cover a compact interval $[a, b]$ if there exists at least one $t_k \in R$ with $t_k \leq a$ and at least another $t_{k'} \in J$ with $t_{k'} \geq b$. A sequence of grids $\{R_m, m = 1, 2, 3, \dots\}$ is said to eventually cover a compact interval $[a, b]$ if for every $\epsilon > 0$ there exists M such that R_m covers the compact interval $[a + \epsilon, b - \epsilon]$ for all $m \geq M$.

Theorem 3.7. Suppose that **C1–C6** hold. Assume that $R_p = \{\sigma_{p1}, \sigma_{p2}, \dots\}$, for $p = 1, 2, \dots$, is a sequence of grids on the real line that eventually covers $[0, \sigma_{\max}]$, with $\text{size}(R_p) = o(p^{-2/3})$. Suppose that $J_p = \{z_{p1}, z_{p2}, \dots\}$, for $p = 1, 2, \dots$, is a sequence of grids.

(i) Let \hat{F}_p denote the estimated measure obtained from Algorithm 1 using the inputs

$\{\sigma_k\} = R_p$ and $\{z_i\} = J_p$. If J_p is such that for some $\nu > 0$,

$$\sup_{z \in D(\nu)} \inf_{y \in J_p} |z - \varphi(y)| = o(p^{-2/3}), \quad (19)$$

where $D(\nu)$ is the closed half-disk in \mathbb{C}^+ centered at 0 with radius ν . Then, there exists a constant C such that for any $T \geq 3$

$$\mathcal{D}_W(\hat{F}_p, F^{\Sigma_\infty}) \leq \frac{C(1+\nu)}{\nu^2} p^{-2/3} \exp(T) \delta_p + \frac{C}{T},$$

where δ_p is such that $\delta_p \xrightarrow{P} 0$, as $p \rightarrow \infty$.

(ii) Let h_β be the solution to Eq. (16) when $x_0 = \beta$. If J_p is such that there exists $\varepsilon > 0$ such that $\{\varphi(y), y \in J_p\}$ eventually covers $[-h_\beta/\lambda - \varepsilon, -h_\beta/\lambda + \varepsilon]$. Then, the proposed estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$, obtained using the inputs $\{\sigma_k\} = R_p$ and $\{z_i\} = J_p$, satisfy:

$$p^{2/3} |\hat{\Theta}_1 - \Theta_1| \xrightarrow{P} 0 \quad \text{and} \quad |\hat{\Theta}_2 - \Theta_2| \xrightarrow{P} 0.$$

Lemma 3.8. For any $\gamma_1, \gamma_2, \lambda$, and F^{Σ_∞} satisfying **C1–C6**, there exists $\nu > 0$ and a sequence of grids $\{J_p\}$ such that Condition (19) is satisfied.

Theorem 3.7 establishes two fundamental results. First, it provides a Berry-Esseen type bound on the Wasserstein distance between the estimated population empirical spectral distribution (ESD) and its limiting counterpart. Second, it demonstrates that the proposed estimator achieves $o_P(p^{-2/3})$ consistency, provided that $\{\varphi(y), y \in J_p\}$ is dense around $-h_\beta/\lambda$. We can find a sequence of J_p satisfying the condition when β is not excessively close to ρ . The positioning of β is influenced by the parameters γ_1 and γ_2 . Specifically, β approaches 0, as γ_1 grows. Reversely, for small γ_1 , β is close to ρ .

Lemma 3.9. Fix any γ_2, λ , and F^{Σ_∞} satisfying **C1–C6**. There exists $\gamma_1^{(0)}$ such that for any $\gamma_1 > \gamma_1^{(0)}$, we can find a sequence of grids $\{J_p\}$ such that the condition in (ii) of Theorem 3.7 is satisfied.

Lemma 3.9 indicates that the condition in (ii) of Theorem 3.7 can be satisfied when γ_1 is large. This scenario corresponds to the regime where n_2 is large and n_1 is relatively small compared to p . The numerical simulation results reported in Table S.3.1 and Table S.3.2 support these findings, showing that the estimation precision decreases as p/n_2 increases or as p/n_1 decreases under the given simulation settings. Nonetheless, the overall estimation precision remains well-controlled within a reasonable range for all settings under consideration.

4 Applications

We consider the application of the proposed ridge-regularized largest-root framework to the problem of general linear hypotheses (GLHT) and testing the equality of two covariance matrices against a one-sided alternative (TECM). The latter deferred to Section S.2 of the Supplementary Materials due to limited space.

Recall the context of testing general linear hypotheses under a multivariate linear model in the introduction. We assume that \mathbf{Z} and Σ_p satisfy **C2–C6**. Additionally, we assume

$$\begin{aligned}\liminf_{p \rightarrow \infty} \ell_{\min} (n_T^{-1} X X^T) &> 0, \\ \liminf_{p \rightarrow \infty} \ell_{\min} (C^T (n_T^{-1} X X^T)^{-1} C) &> 0.\end{aligned}$$

The two conditions ensure that X is of full rank and BC is estimable in asymptotic sense.

Recall the specification of \mathbf{W}_1 and \mathbf{W}_2 in Eq. (3). Under Gaussianity and the null hypothesis, \mathbf{W}_1 and \mathbf{W}_2 are independent Wishart matrices: $n_1 \mathbf{W}_1 \sim \text{Wishart}(\Sigma_p, n_1)$ and $n_2 \mathbf{W}_2 \sim \text{Wishart}(\Sigma_p, n_2)$. To further clarify, \mathbf{W}_2 serves as an unbiased estimator of Σ_p , while \mathbf{W}_1 quantifies the departure of the observations from the null hypothesis. Specifically,

\mathbf{W}_1 can be expressed as $\mathbf{W}_1 = n_1^{-1} \widehat{BC} [C^T (X X^T)^{-1} C]^{-1} (\widehat{BC})^T$, where \widehat{BC} is the ordinary least squares estimator of BC . For additional details, we refer readers to Chapter 10 of Muirhead [2009].

Given $\lambda > 0$, the proposed ridge-regularized F-matrix is constructed as

$$\mathbf{F}_\lambda = \mathbf{W}_1 (\mathbf{W}_2 + \lambda I_p)^{-1} = \frac{1}{n_1} SS_H \left(\frac{1}{n_2} SS_{\text{Res}} + \lambda I_p \right)^{-1}.$$

To test the hypothesis $H_0 : BC = 0$ against $H_a : BC \neq 0$, the proposed ridge-regularized largest root test rejects the null hypothesis at a significance level α if

$$\hat{\ell}_{\max}(\mathbf{F}_\lambda) := \frac{p^{2/3}}{\hat{\Theta}_2} \{ \ell_{\max}(\mathbf{F}_\lambda) - \hat{\Theta}_1 \} > \text{TW}_1(1 - \alpha),$$

where $\text{TW}_1(1 - \alpha)$ is the $(1 - \alpha) \times 100\%$ quantile of the Tracy-Widom distribution of type 1 and the estimated parameters $\hat{\Theta}_1(\lambda)$ and $\hat{\Theta}_2(\lambda)$ are obtained following the proposed method in Section 3 using the eigenvalues of $\mathbf{W}_2 = n_2^{-1} SS_{\text{Res}}$.

Lemma 4.1 (GLHT: Asymptotic type I error rate). *Assume that the conditions in this section are satisfied. Suppose that $\hat{\Theta}_1 - \Theta_1 = o_P(p^{-2/3})$ and $\hat{\Theta}_2 - \Theta_2 = o_P(1)$. Then, for any fixed $\lambda > 0$ and $0 < \alpha < 1$, the type I error rate of the proposed decision rule satisfies*

$$\mathbb{P} \left(\hat{\ell}_{\max}(\mathbf{F}_\lambda) > \text{TW}_1(1 - \alpha) \mid BC = 0 \right) \longrightarrow \alpha,$$

as $p, n_1, n_2 \rightarrow \infty$ simultaneously as in **C1**.

Lemma 4.2 (GLHT: Asymptotic power). *Assume the same conditions as in Lemma 4.1.*

Under H_a , suppose BC is such that

$$\frac{1}{n_1} \ell_{\max} (BC(C^T(XX^T)^{-1}C)^{-1}C^TB^T) = O((\log p)^c),$$

for some $c > 0$. Then, for any fixed $\lambda > 0$ and $\alpha \in (0, 1)$, the power given BC of the proposed decision rule is such that as $p, n_1, n_2 \rightarrow \infty$ simultaneously as in **C1**,

$$P\left(\hat{\ell}_{\max}(\mathbf{F}_\lambda) > \mathbb{TW}_1(1 - \alpha) \mid BC\right) \rightarrow 1.$$

5 Simulation studies

In this section, we evaluate the performance of the proposed ridge-regularized F -matrix framework and the associated estimation methods through Monte Carlo experiments. Due to limited space, we only highlight key results here and defer the detailed simulation settings and additional results to Section S.3 of the Supplementary Materials.

5.1 Estimation precision

The numerical performance of the proposed estimation method is evaluated in this section. We implement Algorithms 1, 2, and 3 using the eigenvalues of $\mathbf{W}_2 = n_2^{-1} \Sigma_p^{1/2} \mathbf{Z} P_2 \mathbf{Z}^T \Sigma_p^{1/2}$, where P_2 is a randomly chosen projection matrix of rank n_2 . The estimated curves $\hat{s}(x)$, $\hat{s}'(x)$, and $\hat{s}''(x)$ under a representative setting are shown in Figure 5.1. Results for other settings are not reported, as they exhibit similar patterns. The figure demonstrates that the overall estimation accuracy is high, although it diminishes as x approaches ρ_λ . This suggests that the estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$ remain reliable unless β is near ρ . Since β satisfies $\beta^2 s'(\beta) = 1/\gamma_1$, β approaches ρ only when γ_1 is small, implying n_1 is significantly larger

than p . However, this scenario falls outside the high-dimensional regime where n_1 and p are comparable, as assumed in **C1**.

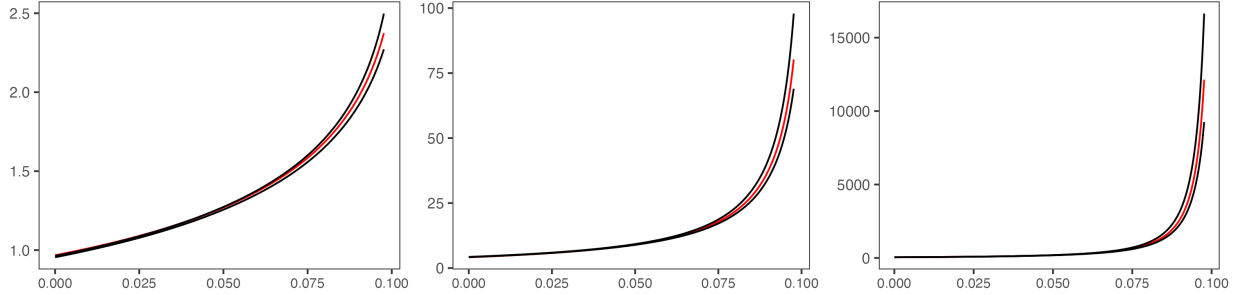


Figure 5.1: The estimated $s(x)$, $s'(x)$ and $s''(x)$ (from left to right) when Σ_p is Poly-Decay, $\hat{\gamma}_2 = 5$, $\lambda = 0.5$. Red: the true functions; Black: 5% and 95% pointwise percentile bands of the estimated functions.

The estimation accuracy of $\hat{\Theta}_1$ and $\hat{\Theta}_2$ is summarized in Table S.3.1 and Table S.3.2 of the Supplementary Materials. Specifically, Table S.3.1 reports the mean and standard deviation of the scaled estimation error $p^{2/3}|\hat{\Theta}_1 - \Theta_1|/\Theta_2$. In nearly all settings, the mean error remains below 0.3. Table S.3.2 provides the mean and standard deviation of $|\hat{\Theta}_2 - \Theta_2|$, with the mean error controlled below 0.05 in most cases. Overall, as p/n_2 increases or as p/n_1 decreases, the estimation accuracy decreases. Notably, the choice of λ does not significantly affect the accuracy. The method performs worst under the Toeplitz setting. The density curves of the estimation errors for Θ_1 are displayed in Figure 5.2 for a representative setting. The empirical distributions exhibit an approximately bell-shaped pattern.

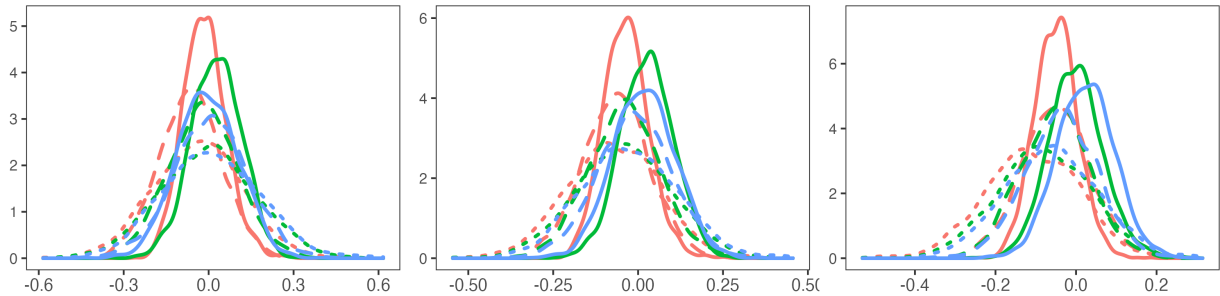


Figure 5.2: Density of $p^{2/3}(\hat{\Theta}_1 - \Theta_1)/\Theta_2$ when Σ_p is Factor-model. From left to right: $\hat{\gamma}_1 = \hat{\gamma}_2$, $\hat{\gamma}_1 = 2\hat{\gamma}_2$, $\hat{\gamma}_1 = 5\hat{\gamma}_2$. Line type: $\hat{\gamma}_2 = 0.5$ (solid), 2 (dashed), 5 (dotted). Color: $\lambda = 0.5$ (red), 1 (green), 1.5 (blue).

5.2 Empirical null distribution

In this section, we examine the empirical distribution of the ridge-regularized largest root under the null hypothesis. We consider the normalized largest root using both the true centering and scaling parameters, as well as the estimated ones. The results are presented in Figures 5.3 for a representative setting. More results can be found in Section S.3 of the Supplementary Materials. These figures provide strong evidence supporting the asymptotic Tracy-Widom distribution of the normalized largest root. Furthermore, the difference between the empirically normalized $\ell_{\max}(\mathbf{F}_\lambda)$ and the version normalized using the true parameters is minimal.

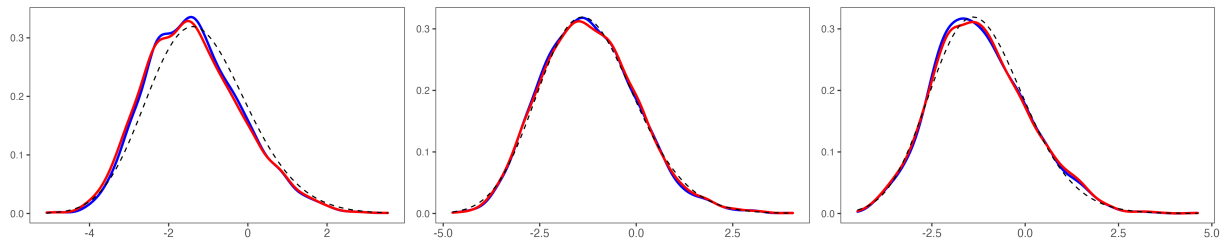


Figure 5.3: Empirical density of $\ell_{\max}(\mathbf{F}_\lambda)$. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1 = 1$, $\lambda = 1$, Σ_p is Factor-model. Red lines: normalized with estimated parameters; Blue lines: normalized with true parameters; Black and dashed: pdf of TW_1 (baseline).

Table 5.1 reports the empirical size at the asymptotic 5% level of the decision rule $\hat{\ell}_{\max}(\mathbf{F}_\lambda) > \text{TW}_1(0.95)$, where $\hat{\ell}_{\max}(\mathbf{F}_\lambda)$ is the normalized largest root with estimated parameters $\hat{\Theta}_1$ and $\hat{\Theta}_2$. The results show that the empirical sizes are slightly conservative but remain well-controlled at the nominal level.

5.3 Empirical power

In this section, we examine the empirical power of the ridge-regularized largest root under the context of GLHT and TECM. The simulation settings are presented in Section S.3 of the Supplementary Materials.

Σ	Size $\times 100\%$		$\hat{\gamma}_2 = 0.5$			$\hat{\gamma}_2 = 2$			$\hat{\gamma}_2 = 5$		
	λ	$\hat{\gamma}_1/\hat{\gamma}_2 =$	1	2	5	1	2	5	1	2	5
Poly-Decay	0.5		3.82	4.48	3.98	4.12	4.28	4.16	4.04	4.24	4.28
	1		3.76	4.40	3.86	4.06	3.74	4.32	4.06	4.16	4.08
	1.5		3.86	4.20	3.80	4.30	3.88	4.40	4.04	4.32	4.08
Toeplitz	0.5		4.50	4.58	4.44	4.34	4.38	4.10	4.24	4.38	4.12
	1		4.26	4.54	4.14	4.22	4.00	4.20	4.20	4.46	4.04
	1.5		4.48	4.80	4.08	4.10	3.90	4.20	4.16	4.38	3.86
Factor	0.5		4.58	4.44	4.16	4.62	4.22	4.10	4.60	4.42	4.54
	1		3.68	3.74	3.76	4.38	4.14	4.26	4.40	4.34	4.56
	1.5		3.78	3.84	3.80	4.02	3.90	4.08	4.58	4.08	4.70

Table 5.1: Empirical sizes at asymptotic level 5% under different settings.

All tests are conducted at level $\alpha = 0.05$. Empirical power curves versus the signal strength parameter ξ are shown in Figure 5.4–5.5 when $\hat{\gamma}_1/\hat{\gamma}_2 = 1$. The results under other settings are reported in Section S.3 of the Supplementary Material. The findings indicate that the proposed methods achieve strong statistical power across the various settings considered. Notably, the methodology demonstrates robustness to different structures of the population covariance matrix, suggesting its broad applicability in practice. Furthermore, the empirical results suggest that the performance of the proposed tests is not overly sensitive to the choice of the regularization parameter λ within an appropriate range.

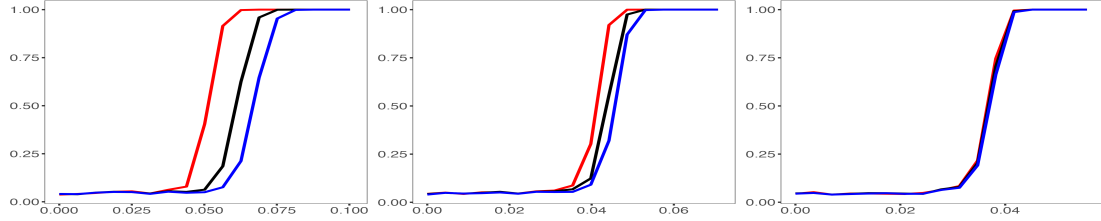


Figure 5.4: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 1$, Σ_p is Poly-Decay. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

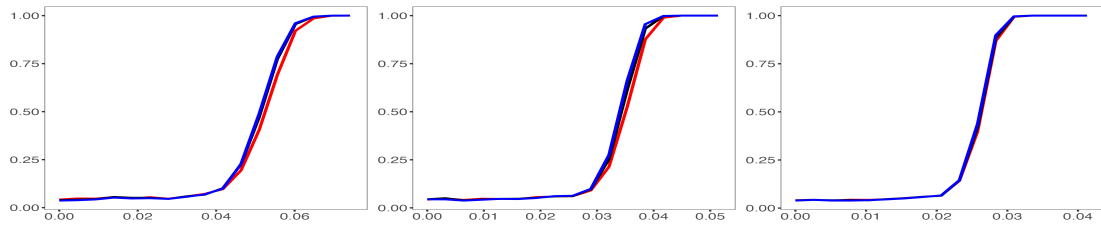


Figure 5.5: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 1$, Σ_p is Toeplitz. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

6 Real data analysis

The proposed method is applied to the Human Connectome Project (HCP) dataset to study the associations between cerebral measurements and behavioral variables. Due to limited space, the analysis is deferred to Section S.4 of the Supplementary Materials.

7 Discussion

In this paper, we proposed a powerful and computationally efficient largest-root test for high-dimensional double Wishart problems. The test is based on a ridge-type regularization of the traditional F -matrix. Leveraging techniques from Random Matrix Theory, we established the asymptotic Tracy-Widom distribution of the largest eigenvalue after appropriate scaling under the null hypothesis, within a regime where the dimension is comparable to the effective sample sizes. To facilitate practical implementation, we developed a computationally efficient procedure for estimating the required scaling parameters. This estimation method relies exclusively on the empirical eigenvalues of the test matrices and does not require additional structural assumptions. The consistency of the proposed estimation procedure was rigorously established. Extensive simulation studies were conducted to evaluate the finite-sample performance of the proposed test. The results demonstrate that the test maintains well-controlled empirical type I error rates and exhibits strong power across a

broad range of settings. Furthermore, the method is shown to be robust to variations in the structure of the population covariance matrix as well as to the distribution of the observations. These properties underscore the practical applicability of the proposed procedure in high-dimensional hypothesis testing scenarios.

Several promising research directions can be pursued to extend the current work. While this study primarily focuses on the asymptotic null distribution of the test statistic, an important theoretical extension is to analyze its asymptotic power under a class of local alternatives. Such an analysis could provide deeper insights into the test’s sensitivity to signal structure and the role of the regularization parameter λ . Moreover, it could facilitate the development of a principled approach for selecting λ in an optimal manner. Given the technical complexity of these extensions and the already intricate nature of the current work, we leave this important analysis for future research.

From a technical perspective, the current analysis assumes regular stability properties for the leading eigenvalues of the population covariance matrix. This assumption excludes certain important models, such as classical factor models, where a finite number of eigenvalues remain well-separated from the bulk. However, our simulation results indicate that the proposed methodology may remain valid under a class of factor models. Establishing rigorous theoretical guarantees in such settings presents an interesting direction for future research.

Another potential avenue for exploration is the integration of the proposed methodology with a variable screening strategy to extend its applicability to ultra-high-dimensional settings. By effectively reducing the dimensionality before applying the test, this approach could enhance computational efficiency while maintaining statistical power in scenarios where p grows significantly faster than the sample sizes.

8 Supplementary material

Supplementary Material includes detailed proofs of the main theoretical results and additional details of the simulation study and real data application.

References

- Z. Bai and H. Saranadasa. Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6:311–329, 1996.
- Z. Bai and J. W. Silverstein. Clt for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553–605, 2004.
- Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer, 2010.
- Z. Bao, G. Pan, and W. Zhou. Local density of the spectrum on the edge for sample covariance matrices with general population. 2013.
- Z. Bao, G. Pan, and Z. Wang. Universality for the largest eigenvalue of sample covariance matrices with general population. *The Annals of Statistics*, 43(1):382–421, 2015.
- A. Bloemendal, L. Erdos, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized wigner matrices. *Electronic Journal of Probability*, 19:33, 2014.
- S. G. Bobkov. Proximity of probability distributions in terms of fourier–stieltjes transforms. *Russian Mathematical Surveys*, 71(6):1021, 2016.
- S. X. Chen and Y.-L. Qin. A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835, 2010.
- S. X. Chen, J. Li, and P.-S. Zhong. Two-sample tests for high dimensional means with thresholding and data transformation. *arXiv preprint arXiv:1410.2848*, 2014.
- M. Dieng. A matlab package for computing tracy-widom distribution and simulating random matrices. *RMLab. 2006*, 2006.

- N. El Karoui. Tracy-widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, 35(2):663–714, 2007.
- N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics*, 36(6):2757–2790, 2008.
- L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized wigner matrices. *Advances in Mathematics*, 229(3):1435–1515, 2012.
- L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. The local semicircle law for a general class of random matrices. *Electronic Journal of Probability*, 18:1–58, 2013.
- X. Han, G. Pan, and B. Zhang. The tracy–widom law for the largest eigenvalue of f type matrices. *The Annals of Statistics*, 44(4):1564–1592, 2016.
- 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.
- Y. He, T. Jiang, J. Wen, and G. Xu. Likelihood ratio test in multivariate linear regression: From low to high dimension. *Statistica Sinica*, 31(3):1215–1238, 2021.
- J. Hu, Z. Bai, C. Wang, and W. Wang. On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics*, 69(2):365–387, 2017.
- I. M. Johnstone. Multivariate analysis and jacobi ensembles: Largest eigenvalue, tracy–widom limits and rates of convergence. *Annals of statistics*, 36(6):2638, 2008.
- I. M. Johnstone, Z. Ma, P. O. Perry, M. Shahram, M. P. O. Perry, and P. Perry. Package ‘rmtstat’. 2022.
- A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probability Theory and Related Fields*, 169:257–352, 2017.
- O. Ledoit and M. Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012.
- J. O. Lee and K. Schnelli. Tracy–widom distribution for the largest eigenvalue of real sample co-

- variance matrices with general population. *The Annals of Applied Probability*, 26(6):3786–3839, 2016.
- H. Li. Analysis of the limiting spectral distribution of large random matrices of the mar\{c\} enko-pastur type. *arXiv preprint arXiv:2408.10068*, 2024.
- H. Li, A. Aue, and D. Paul. High-dimensional general linear hypothesis tests via non-linear spectral shrinkage. *Bernoulli*, 26(4):2541–2571, 2020a.
- H. Li, A. Aue, D. Paul, J. Peng, and P. Wang. An adaptable generalization of hotelling’s t^2 test in high dimension. *The Annals of Statistics*, 48(3):1815–1847, 2020b.
- H. Li, A. Aue, D. Paul, and J. Peng. Testing general linear hypotheses under a high-dimensional multivariate regression model with spiked noise covariance. *Journal of the American Statistical Association*, 119(548):2799–2810, 2024.
- M. Lopes, L. Jacob, and M. J. Wainwright. A more powerful two-sample test in high dimensions using random projection. In *Advances in Neural Information Processing Systems*, pages 1206–1214, 2011.
- R. J. Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009.
- R. R. Nadakuditi and J. W. Silverstein. Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *IEEE Journal of selected topics in Signal Processing*, 4(3):468–480, 2010.
- B. Nadler and I. Johnstone. *On the distribution of Roy’s largest root test in MANOVA and in signal detection in noise*. Department of Statistics, Stanford University, 2011.
- N. S. Pillai and J. Yin. Universality of covariance matrices. *Annals of Applied Probability*, 24(3):935–1001, June 2014.
- J. W. Silverstein and Z. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate analysis*, 54(2):175–192, 1995.
- J. W. Silverstein and S.-I. Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):295–309, 1995.

- M. S. Srivastava and Y. Fujikoshi. Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis*, 97(9):1927–1940, 2006.
- C. A. Tracy and H. Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.
- C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177(3):727–754, 1996.
- L. Wang, B. Peng, and R. Li. A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, 110(512):1658–1669, 2015.
- T. Yamada and T. Himeno. Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. *Journal of Multivariate Analysis*, 139:7–27, 2015.
- L. Zhao, P. R. Krishnaiah, and Z. Bai. On detection of the number of signals when the noise covariance matrix is arbitrary. *Journal of Multivariate Analysis*, 20(1):26–49, 1986.

Supplementary Materials to “Ridge-regularized Largest Root Test for General High-dimensional Double Wishart Problems”

S.1 Implementation details of Algorithm 1, Algorithm 2, and Algorithm 3

Implementation details for Algorithm 1. The recommended choice of the grid $\{\sigma_k\}_{k=1}^K$ is equally spaced on $[\tilde{\ell}_{n_2}, \tilde{\ell}_1]$, with $K \approx 500$. The choice of K strikes a balance between computational efficiency and estimation accuracy. The recommended choice of $\{z_i\}_{i=1}^I$ is such that $\Re(\hat{\varphi}(z_i))$ are evenly spaced on $[\hat{\varphi}(1.05\tilde{\ell}_1), \hat{\varphi}(-\lambda)]$, and $\Im(\hat{\varphi}(z_i)) = 10^{-2} \times \tilde{\ell}_1^{-1}$. We can first select $\hat{\varphi}(z_i)$ and numerically find the corresponding z_i using optimization methods like Newton-Raphson. While more points improve accuracy, we recommend setting $I \approx 500$ to balance computational efficiency and estimation precision. Traditional linear programming solvers can be used for implementation. In our simulation, we utilize the R package **Rglpk**, which implements the GNU Linear Programming Kit. With the recommended settings, the problem can be efficiently handled on a typical PC (≤ 10 seconds).

Implementation details for Algorithm 3. The proposed ODE requires high accuracy on the initial condition $\hat{s}(0)$. We propose using Newton-Raphson method to find $\hat{s}(0)$, starting from the initial point $1/(\lambda\hat{\gamma}_2\hat{\varphi}(-\lambda)) - 1/\hat{\gamma}_2$. The ODE can then be solved by traditional numerical methods such as the fourth-order Runge-Kutta method (RK4). In our simulation studies, the R package **deSolve** is used for implementation.

S.2 Application to testing equality of two covariance matrices against a one-sided alternative

Consider two independent p -variate samples $\{Y_{1j}, 1 \leq j \leq n_{(1)}\}$ and $\{Y_{2j}, 1 \leq j \leq n_{(2)}\}$, following the model:

$$\begin{aligned} Y_{1j} &= \mu_1 + L\xi_j + \Sigma_p^{1/2}Z_{1j}, \quad j = 1, \dots, n_{(1)}, \\ Y_{2j} &= \mu_2 + \Sigma_p^{1/2}Z_{2j}, \quad j = 1, \dots, n_{(2)}, \end{aligned} \tag{S.1}$$

where μ_1 and μ_2 are unknown nonrandom mean vectors; Z_{1j} , Z_{2j} , and ξ_j are independent random vectors with mean 0 and variance 1 for each entry; L is an unknown nonrandom $p \times p$ loading matrix; and Σ_p is an arbitrary positive definite noise covariance matrix. We assume that the entries of Z_{ij} and Σ_p satisfy **C2–C6**.

Under this model, the covariance matrix of the Y_{1j} 's is $\Sigma_{1p} = LL^T + \Sigma_p$, while the covariance matrix of the Y_{2j} 's is $\Sigma_{2p} = \Sigma_p$. The primary interest lies in testing the equality of these two covariance matrices, which is equivalent to testing the existence of LL^T . The problem can thus be formulated as the following hypothesis test:

$$H_0 : LL^T = 0 \quad \text{versus} \quad H_a : LL^T \neq 0.$$

This problem can be seen as a special case of testing the equality of two covariance matrices against a one-sided alternative. The model in (S.1) has been used in previous work, including Zhao et al. [1986], Nadakuditi and Silverstein [2010], and Nadler and Johnstone [2011], to detect signals in the $\{Y_{1j}\}$ sample when a noise-only sample $\{Y_{2j}\}$ is available.

In a high-dimensional setting, where p is comparable to $n_{(1)}$ and $n_{(2)}$, we propose testing the null hypothesis using the largest eigenvalue of the ridge-regularized F -matrix $\mathbf{F}_\lambda = S_1(S_2 + \lambda I_p)^{-1}$, where S_1 and S_2 are the sample covariance matrices of the two samples,

given by:

$$S_1 = \frac{1}{n_1} \sum_{j=1}^{n_{(1)}} (Y_{1j} - \bar{Y}_{1\cdot})(Y_{1j} - \bar{Y}_{1\cdot})^T,$$

$$S_2 = \frac{1}{n_2} \sum_{j=1}^{n_{(2)}} (Y_{2j} - \bar{Y}_{2\cdot})(Y_{2j} - \bar{Y}_{2\cdot})^T.$$

Here, $\bar{Y}_{1\cdot}$ and $\bar{Y}_{2\cdot}$ are the sample means of the respective samples, $n_1 = n_{(1)} - 1$ and $n_2 = n_{(2)} - 1$. It is straightforward to verify that, under the null hypothesis $H_0 : LL^T = 0$, the sample covariance matrices $\mathbf{W}_1 = S_1$ and $\mathbf{W}_2 = S_2$ conform to the unified framework (1). The proposed ridge-regularized largest-root test rejects the null hypothesis if

$$\hat{\ell}_{\max}(\mathbf{F}_\lambda) := \frac{p^{2/3}}{\hat{\Theta}_2} \{\ell_{\max}(\mathbf{F}_\lambda) - \hat{\Theta}_1\} > \mathbb{T}\mathbb{W}_1(1 - \alpha),$$

where the estimated parameters $\hat{\Theta}_1(\lambda)$ and $\hat{\Theta}_2(\lambda)$ are obtained following the method in Section 3 using the eigenvalues of $\mathbf{W}_2 = S_2$.

Lemma S.2.1 (TECM: Asymptotic type I error). *Assume the conditions in this section are satisfied. Suppose that $\hat{\Theta}_1 - \Theta_1 = o_P(p^{-2/3})$ and $\hat{\Theta}_2 - \Theta_2 = o_P(1)$. Then, for any fixed $\lambda > 0$ and $\alpha \in (0, 1)$, the type I error rate for the proposed decision rule is such that*

$$\mathbb{P}\left(\hat{\ell}_{\max}(\mathbf{F}_\lambda) > \mathbb{T}\mathbb{W}_1(1 - \alpha) \mid LL^T = 0\right) \longrightarrow \alpha,$$

as $p, n_1, n_2 \rightarrow \infty$ simultaneously as in **C1**.

Lemma S.2.2 (TECM: Asymptotic power). *Assume the same conditions as in Lemma S.2.1. Under H_a , assume LL^T is such that as $p \rightarrow \infty$,*

$$\ell_{\max}(LL^T) = O((\log p)^c),$$

for some $c > 0$. Then, for any fixed $\lambda > 0$ and $\alpha \in (0, 1)$, the power given LL^T of the

proposed decision rule is such that

$$\mathbb{P}\left(\hat{T}(\lambda) > \text{TW}_1(1 - \alpha) \mid LL^T\right) \rightarrow 1,$$

as $p, n_1, n_2 \rightarrow \infty$ simultaneously as in **C1**.

S.3 Simulation study: additional details

S.3.1 Simulation settings

In this section, we present the simulation settings and additional results of the simulation studies presented in Section 5 of the manuscript. The study considers the following configurations:

- (i) The entries of the error matrix \mathbf{Z} are drawn from normal, t-distributions with 4 degrees of freedom, or Poisson distributions. However, as the results are consistent across these settings, only those for normal errors are reported.
- (ii) With $n_2 = 500$, the dimension p is set to 250, 1000, or 2500, corresponding to $\hat{\gamma}_2 = 0.5$, 2, and 5.
- (iii) Three cases are considered for $\hat{\gamma}_1$: $\hat{\gamma}_1 = \hat{\gamma}_2$, $2\hat{\gamma}_2$, or $5\hat{\gamma}_2$, corresponding to $n_1 = n_2$, $0.5n_2$, or $0.2n_2$.
- (iv) The eigenvalues τ_j of Σ_p are set according to three models:
 - (a) *Poly-Decay*: $\tau_j = 0.01 + (0.1 + p - j)^6$, $j = 1, \dots, p$, representing polynomial decay.
 - (b) *Toeplitz*: Σ_p is a Toeplitz matrix with (i, j) -th entry $0.3^{|i-j|}$, for $i, j = 1, \dots, p$.
 - (c) *Factor-model*: For $j \geq 6$, the eigenvalues match those from the *Poly-Decay* model. The first five eigenvalues are spikes: $\tau_j = (2.2 - 0.2j)\tau_6$, for $j = 1, \dots, 5$.

All models are normalized so that $\text{tr}(\Sigma_p) = p$.

- (v) The regularization parameter λ is set to 0.5, 1, or 1.5. Given that $p^{-1}\text{tr}(\mathbf{W}_2) \approx p^{-1}\text{tr}(\Sigma_p) = 1$, these choices for λ are approximately around the mean of the empirical eigenvalues.

Notably, the *Factor-model* represents a setting in which Condition **C5** is violated. Throughout, Algorithms 1, 2, 3 are implemented with the recommended configurations in Section S.1.

S.3.2 Empirical null distribution: additional details

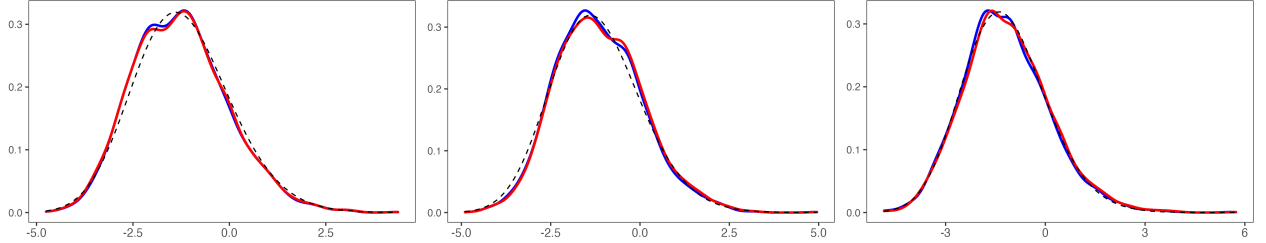


Figure S.3.1: Empirical density of $\ell_{\max}(\mathbf{F}_\lambda)$. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1 = 1$, $\lambda = 0.5$, Σ_p is Poly-Decay. Red lines: normalized with estimated parameters; Blue lines: normalized with true parameters; Black and dashed: pdf of TW_1 (baseline).

S.3.3 Empirical power: additional details

The simulation study under the alternative hypotheses is set as follows.

GLHT Assume the linear model $\mathbf{Y} = B\mathbf{X} + \Sigma_p^{1/2}\mathbf{Z}$ as in the introduction. While n_1, n_2, p, \mathbf{Z} and Σ_p follow configurations (i)-(v), \mathbf{X} is a $n_1 \times (n_1 + n_2)$ matrix whose entries are iid $N(0, 1)$. The coefficient matrix B is assumed to be of low rank with its first column drawn from $N(0, \xi^2)$ and all other columns set to zero. The parameter ξ controls the signal strength. We test $H_0 : B = 0$ against $H_a : B \neq 0$. This setting corresponds to a low-rank signal detection problem.

TECM Under the observation model (S.1), assume n_1, n_2, p, \mathbf{Z} and Σ_p follow configurations (i)-(v) and $\mu_1 = \mu_2 = 0$. The entries of S_j follow the same distribution as those

Σ_p	$\hat{\gamma}_2$	λ	$\hat{\gamma}_1 = \hat{\gamma}_2$	$\hat{\gamma}_1 = 2\hat{\gamma}_2$	$\hat{\gamma}_1 = 5\hat{\gamma}_2$
Poly-Decay	0.5	0.5	0.06 (0.04)	0.05 (0.04)	0.04 (0.03)
		1.0	0.08 (0.06)	0.06 (0.05)	0.05 (0.04)
		1.5	0.10 (0.07)	0.08 (0.06)	0.06 (0.05)
	2.0	0.5	0.10 (0.07)	0.09 (0.06)	0.08 (0.06)
		1.0	0.11 (0.08)	0.09 (0.07)	0.08 (0.06)
		1.5	0.11 (0.09)	0.09 (0.07)	0.07 (0.06)
	5.0	0.5	0.14 (0.10)	0.13 (0.10)	0.13 (0.09)
		1.0	0.14 (0.10)	0.12 (0.09)	0.12 (0.08)
		1.5	0.14 (0.11)	0.12 (0.09)	0.11 (0.08)
Toeplitz	0.5	0.5	0.11 (0.08)	0.09 (0.07)	0.08 (0.06)
		1.0	0.15 (0.11)	0.11 (0.08)	0.08 (0.06)
		1.5	0.19 (0.15)	0.14 (0.11)	0.10 (0.07)
	2.0	0.5	0.16 (0.12)	0.13 (0.09)	0.11 (0.09)
		1.0	0.19 (0.14)	0.13 (0.10)	0.10 (0.08)
		1.5	0.21 (0.16)	0.14 (0.11)	0.10 (0.08)
	5.0	0.5	0.26 (0.19)	0.35 (0.17)	0.15 (0.11)
		1.0	0.24 (0.18)	0.19 (0.13)	0.12 (0.09)
		1.5	0.29 (0.20)	0.18 (0.13)	0.12 (0.09)
Factor	0.5	0.5	0.06 (0.05)	0.06 (0.04)	0.06 (0.04)
		1.0	0.07 (0.06)	0.07 (0.05)	0.05 (0.04)
		1.5	0.09 (0.06)	0.07 (0.06)	0.06 (0.05)
	2.0	0.5	0.10 (0.08)	0.09 (0.07)	0.08 (0.06)
		1.0	0.10 (0.07)	0.09 (0.07)	0.08 (0.06)
		1.5	0.10 (0.08)	0.09 (0.07)	0.08 (0.06)
	5.0	0.5	0.13 (0.10)	0.12 (0.09)	0.13 (0.09)
		1.0	0.13 (0.10)	0.12 (0.09)	0.11 (0.08)
		1.5	0.14 (0.10)	0.12 (0.09)	0.11 (0.08)

Table S.3.1: Mean (standard deviation) of $p^{2/3}|\hat{\Theta}_1 - \Theta_1|/\Theta_2$ under different settings.

Σ_p	$\hat{\gamma}_2$	λ	$\hat{\gamma}_1 = \hat{\gamma}_2$	$\hat{\gamma}_1 = 2\gamma_2$	$\hat{\gamma}_1 = 5\gamma_2$
Poly-Decay	0.5	0.5	0.35 (0.24)	0.27 (0.22)	0.22 (0.17)
		1.0	0.44 (0.36)	0.37 (0.28)	0.31 (0.23)
		1.5	0.67 (0.53)	0.47 (0.37)	0.36 (0.27)
	2.0	0.5	0.31 (0.23)	0.26 (0.21)	0.23 (0.18)
		1.0	0.36 (0.28)	0.28 (0.20)	0.25 (0.19)
		1.5	0.46 (0.36)	0.32 (0.25)	0.25 (0.19)
	5.0	0.5	0.45 (0.33)	0.28 (0.22)	0.22 (0.17)
		1.0	0.49 (0.37)	0.33 (0.25)	0.23 (0.18)
		1.5	0.57 (0.42)	0.38 (0.29)	0.25 (0.19)
Toeplitz	0.5	0.5	0.44 (0.33)	0.33 (0.25)	0.25 (0.19)
		1.0	0.89 (0.67)	0.52 (0.40)	0.35 (0.26)
		1.5	1.49 (1.11)	0.77 (0.57)	0.44 (0.34)
	2.0	0.5	0.66 (0.48)	0.36 (0.27)	0.29 (0.22)
		1.0	0.89 (0.69)	0.45 (0.34)	0.31 (0.23)
		1.5	1.19 (0.93)	0.55 (0.41)	0.35 (0.25)
	5.0	0.5	2.17 (1.26)	1.44 (0.35)	0.63 (0.25)
		1.0	3.08 (1.86)	0.63 (0.44)	0.23 (0.18)
		1.5	2.72 (1.85)	0.52 (0.39)	0.28 (0.21)
Factor	0.5	0.5	0.38 (0.29)	0.47 (0.25)	0.34 (0.25)
		1.0	0.94 (0.46)	0.52 (0.39)	0.38 (0.27)
		1.5	0.88 (0.67)	0.85 (0.52)	0.53 (0.38)
	2.0	0.5	0.34 (0.25)	0.25 (0.19)	0.27 (0.19)
		1.0	0.38 (0.28)	0.32 (0.23)	0.25 (0.19)
		1.5	0.45 (0.34)	0.32 (0.24)	0.29 (0.22)
	5.0	0.5	0.40 (0.30)	0.26 (0.20)	0.20 (0.15)
		1.0	0.46 (0.34)	0.32 (0.24)	0.24 (0.18)
		1.5	0.57 (0.42)	0.33 (0.25)	0.25 (0.19)

Table S.3.2: Mean (standard deviation) $\times 100$ of $|\hat{\Theta}_2 - \Theta_2|/\Theta_2$ under different settings.

of \mathbf{Z} . The loading matrix L is assumed to be of low rank with its first column drawn from $N(0, \xi^2)$ and all other columns set to zero. The parameter ξ controls the signal strength.

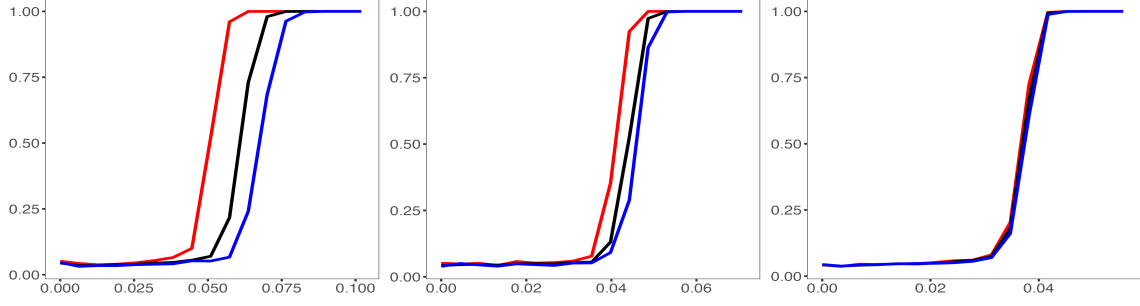


Figure S.3.2: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 1$, Σ_p is Factor-Model. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

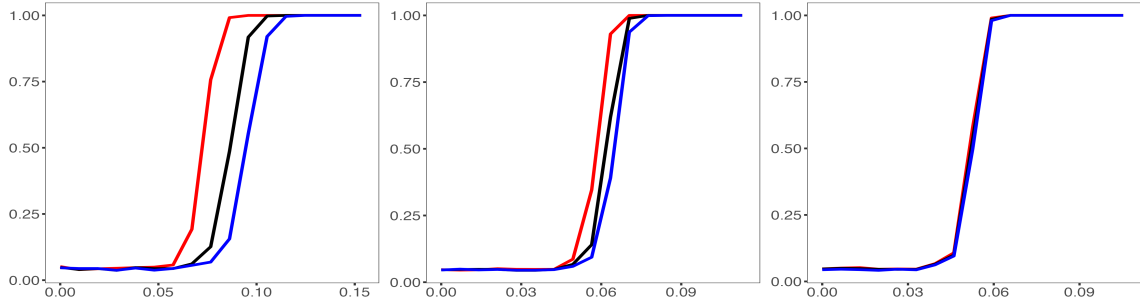


Figure S.3.3: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 1$, Σ_p is Poly-Decay. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

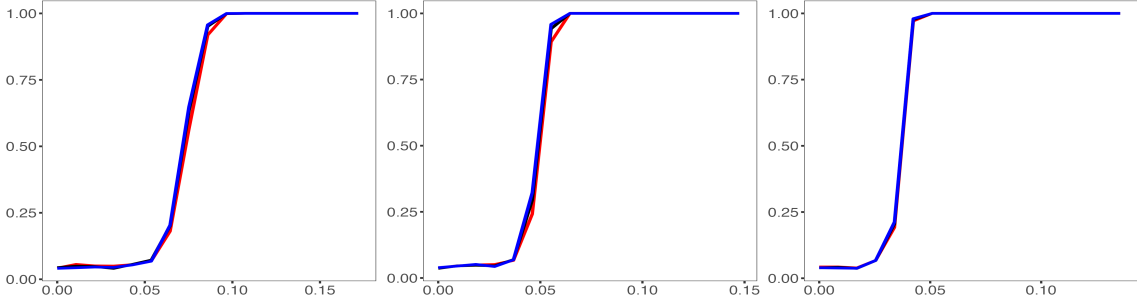


Figure S.3.4: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 1$, Σ_p is Toeplitz. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

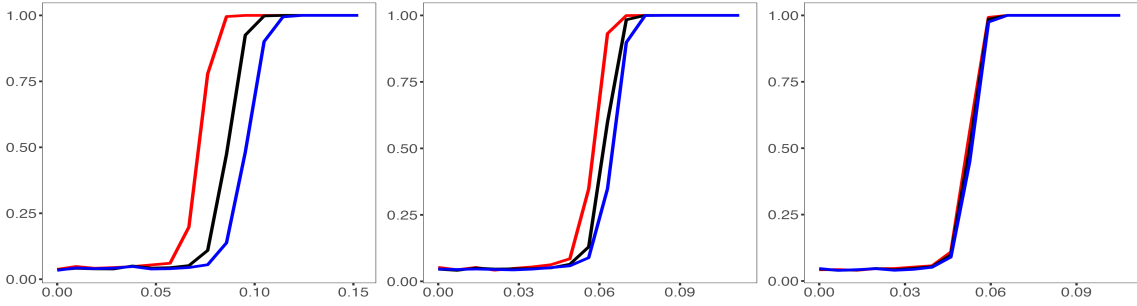


Figure S.3.5: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 1$, Σ_p is Factor-Model. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

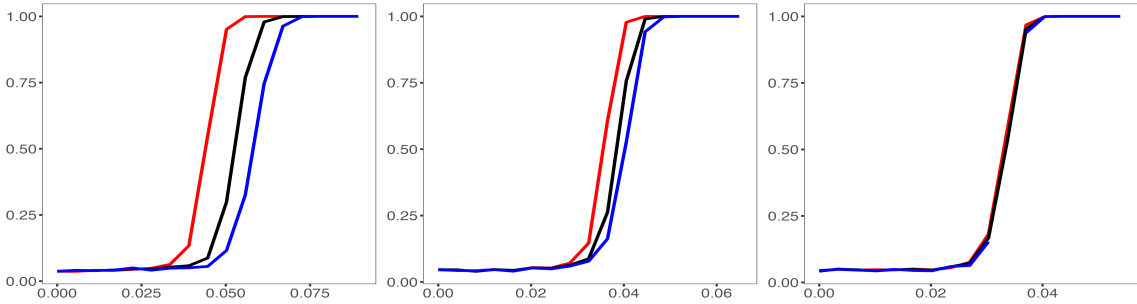


Figure S.3.6: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 2$, Σ_p is Poly-Decay. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

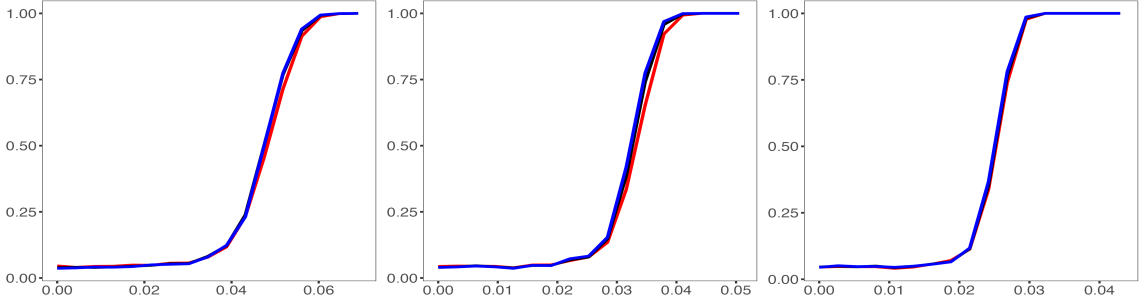


Figure S.3.7: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 2$, Σ_p is Toeplitz. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

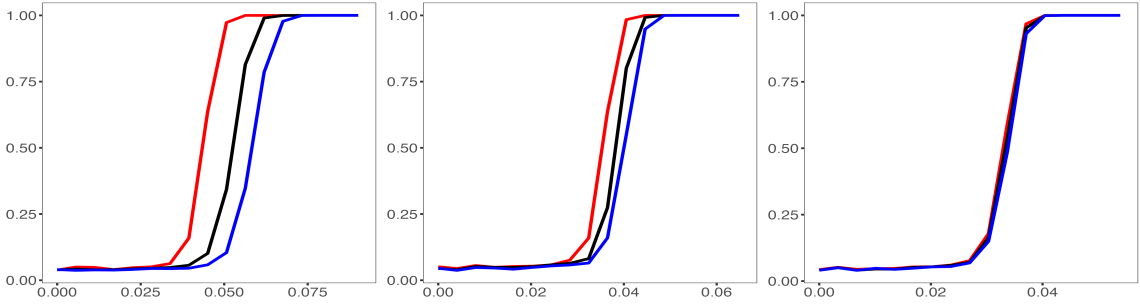


Figure S.3.8: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 2$, Σ_p is Factor-Model. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

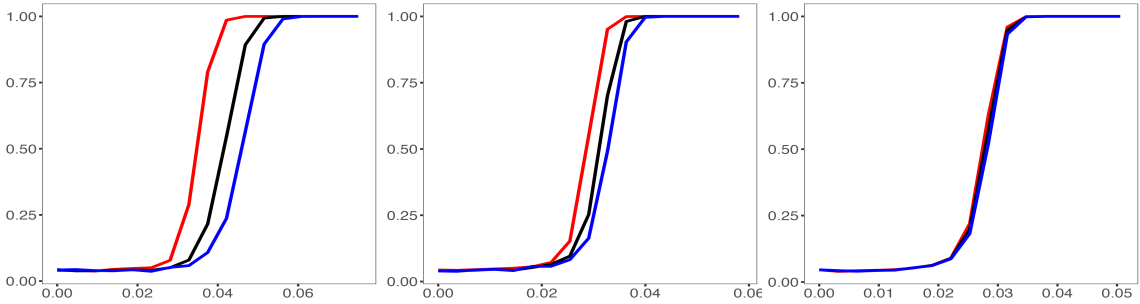


Figure S.3.9: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 5$, Σ_p is Poly-Decay. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

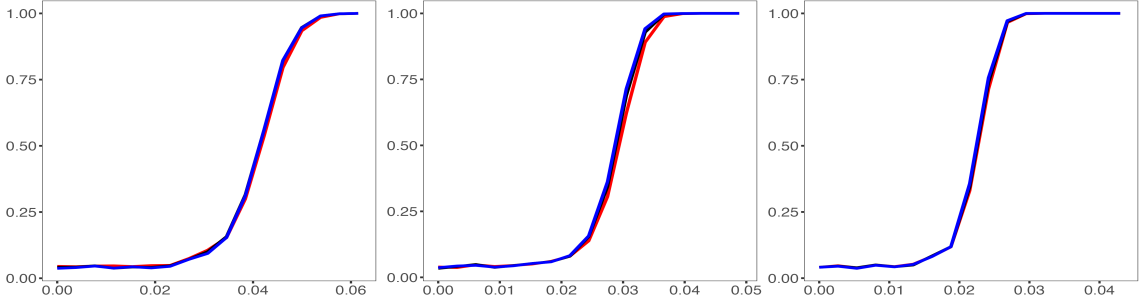


Figure S.3.10: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 5$, Σ_p is Toeplitz. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

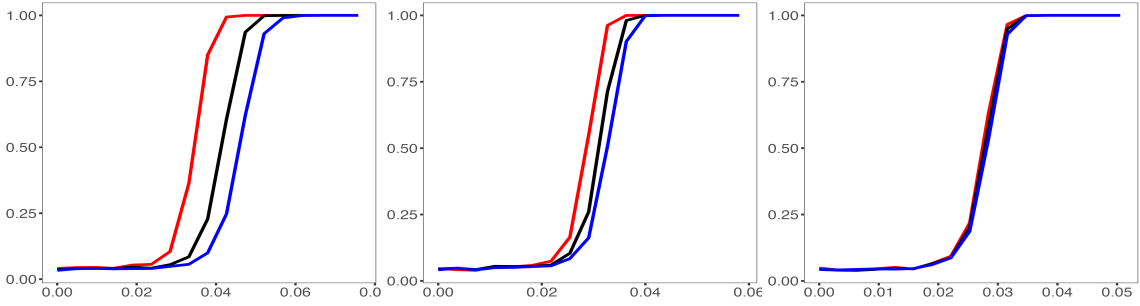


Figure S.3.11: Empirical power under GLHT settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 5$, Σ_p is Factor-Model. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

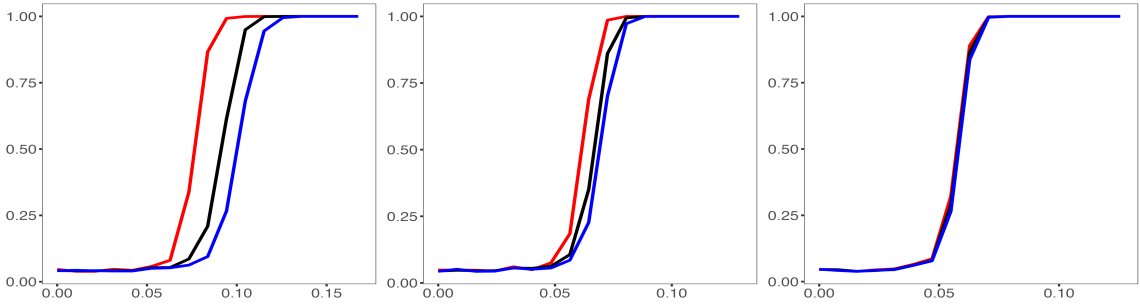


Figure S.3.12: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 2$, Σ_p is Poly-Decay. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

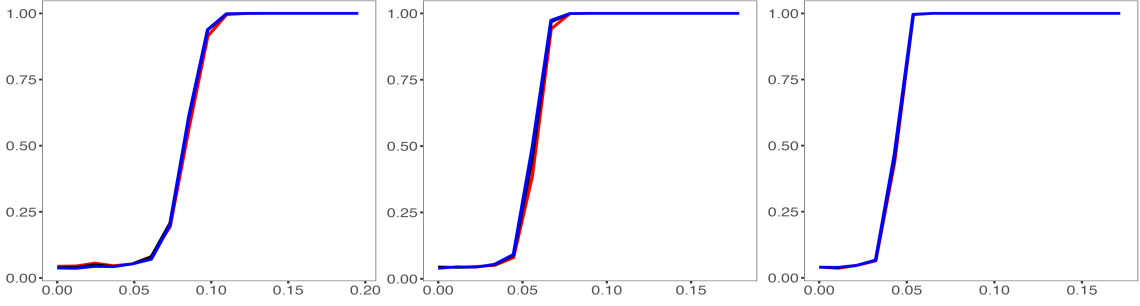


Figure S.3.13: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 2$, Σ_p is Toeplitz. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

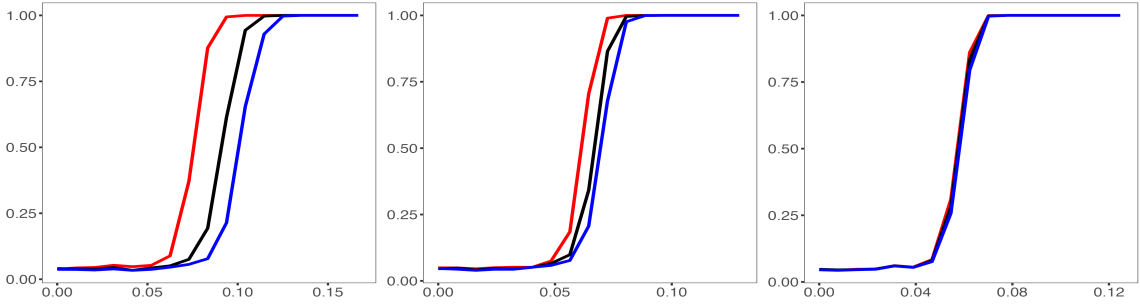


Figure S.3.14: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 2$, Σ_p is Factor-Model. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

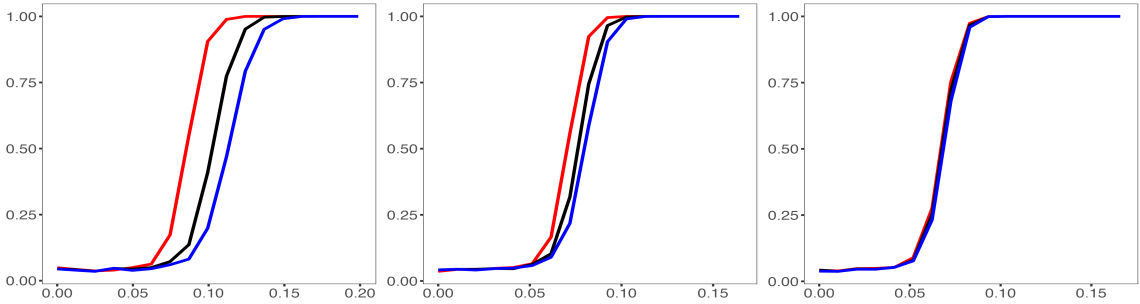


Figure S.3.15: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 5$, Σ_p is Poly-Decay. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

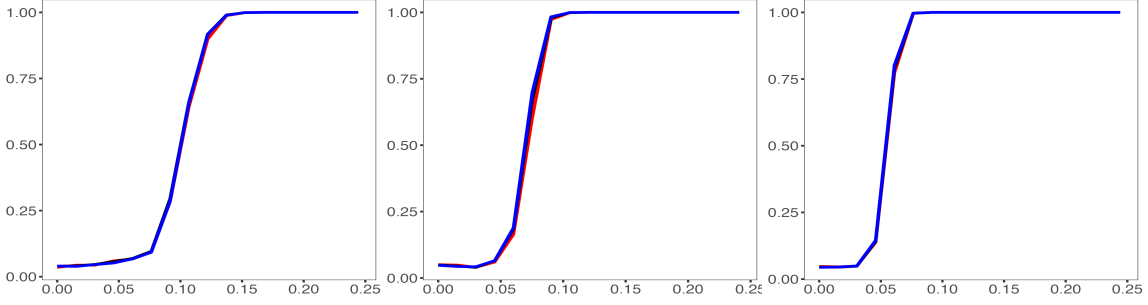


Figure S.3.16: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 5$, Σ_p is Toeplitz. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

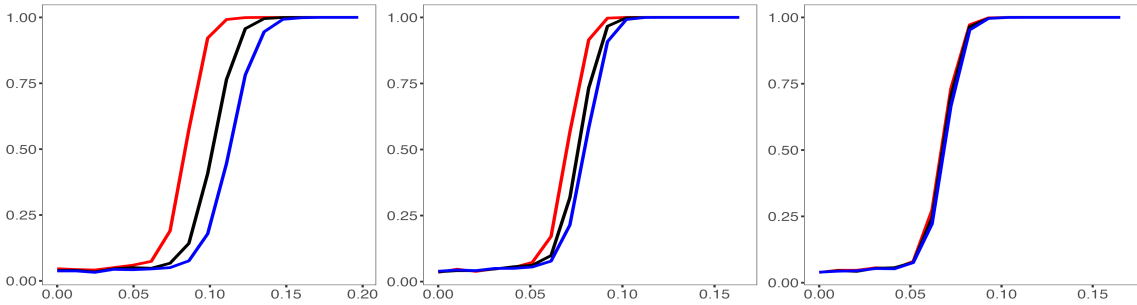


Figure S.3.17: Empirical power under TECM settings. From left to right: $\hat{\gamma}_2 = 0.5, 2, 5$, while $\hat{\gamma}_1/\hat{\gamma}_2 = 5$, Σ_p is Factor-Model. Red lines: $\lambda = 0.5$; Black lines: $\lambda = 1$; Blue lines: $\lambda = 1.5$.

S.4 Results of the real data analysis

The Human Connectome Project (HCP), funded by the National Institutes of Health (NIH), aims to map the neural pathways that underlie human brain function and behavior. A consortium led by Washington University in St. Louis and the University of Minnesota has collected extensive neuroimaging and behavioral data from $n_T = 1,113$ healthy young adults aged 22 to 35. This dataset facilitates detailed analyses of brain circuits and their relationships to behavior and genetics. Among the publicly available data are cerebral volumetric measurements and various behavioral evaluation scores. In this section, we utilize the proposed tests to investigate the associations between cerebral measurements and behavioral variables using the HCP dataset.

The behavioral evaluation scores in the HCP dataset encompass several domains, including cognition, emotion, sensation, and motor functions. The cognition domain assesses various cognitive abilities such as episodic memory, cognitive flexibility, and attention control. The emotion domain includes measures of psychological well-being, social relationships, stress, and self-efficacy. The sensation domain comprises tests related to audition, olfaction, taste, and vision. The motor domain evaluates aspects like cardiovascular endurance, manual dexterity, grip strength, and gait speed. After a pre-screening process to filter out highly correlated variables, we selected $p = 127$ representative behavioral variables to study their associations with cerebral measurements.

Cerebral measurements were obtained for 68 cortical surface regions, assessing surface area, cortical thickness, and gray matter volume. These regions are distributed across 14 cerebral lobes symmetrically located in both hemispheres, resulting in a total of 204 cortical surface variables ($68 \text{ regions} \times 3 \text{ measurement types}$). Additionally, each voxel in the brain’s subcortical structures was assigned one of 44 labels, with the volume of each labeled region measured, yielding 44 subcortical variables. Four labels were excluded from the analysis due to a high proportion of missing values.

Available demographics information includes the age and gender of subjects. Specifi-

cally, the subjects are divided into four age groups, namely 22–25, 26–30, 31–35, and 36+. The data set is roughly balanced with respect to gender as the 1113 subjects consist of 606 females and 507 males.

Specifically, we consider the following multivariate regression model:

$$Y_i = \beta_0 + \beta_1^T \mathbf{L}_{1i} + \cdots + \beta_{14}^T \mathbf{L}_{14,i} + \beta_{15}^T \mathbf{SC}_i + \beta_{16}^T \mathbf{D}_i + \Sigma_p^{1/2} Z_i. \quad (\text{S.2})$$

where (i) Y_i is the vector of 127 behavioral scores of subject i ; (ii) \mathbf{L}_{ji} ($j = 1, \dots, 14$) represents the cerebral measurements vector of the surface regions belonging to lobe j for subject i , with dimensions varying from 3 to 33 depending on the lobe. The variables are of 3 types: surface area, cortical thickness and gray matter volume; (iii) \mathbf{SC}_i contains the volume measurements of 40 subcortical regions of subject i ; (iv) \mathbf{D}_i are age and gender group dummy variables of subject i . This model allows us to assess the associations between various cerebral measurements and behavioral outcomes while accounting for demographic variables. The dimension of the explanatory variables (including the intercept) is $m = 249$, the dimension of response Y_i is $p = 127$ and the sample size is $n_T = 1113$. We are interested in testing the joint significance of all variables associated with each cerebral lobe. Namely, we test $H_0 : \beta_j = 0$ against $H_a : \beta_j \neq 0$, for each $j = 1, \dots, 14$. The proposed test procedure in Section 4 is applied, with $\lambda = 0.5\hat{m}$, \hat{m} or $1.5\hat{m}$, where \hat{m} is the average of eigenvalues of \mathbf{W}_2 . The p-values are reported in Table S.4.1

Among the significant coefficients detected by the proposed tests are left medial temporal lobe, right frontal lobe and left parietal lobe. Similar results are reported in the literature, for example Li et al. [2024]. The left medial temporal lobe, encompassing structures such as the hippocampus and parahippocampal gyrus, plays a pivotal role in memory formation and retrieval. Damage to the lobe has been associated with impairments in verbal memory tasks, indicating its importance in processing verbal material. Functional connectivity within the lobe is crucial for cognitive functions like episodic memory.

The frontal lobes, located directly behind the forehead, are the largest lobes in the

Lobe name	$\lambda/\hat{m} = 0.5$	$\lambda/\hat{m} = 1$	$\lambda/\hat{m} = 1.5$
Left Temporal Lobe Medial Aspect	0.070	0.020	0.009
Right Temporal Lobe Medial Aspect	0.435	0.565	0.637
Left Temporal Lobe Lateral Aspect	0.723	0.766	0.760
Right Temporal Lobe Lateral Aspect	0.085	0.035	0.022
Left Frontal Lobe	0.658	0.585	0.579
Right Frontal Lobe	0.010	0.001	0.001
Left Parietal Lobe	0.108	0.056	0.045
Right Parietal Lobe	0.365	0.337	0.338
Left Occipital Lobe	0.786	0.759	0.725
Right Occipital Lobe	0.204	0.357	0.357
Left Cingulate Cortex	0.252	0.252	0.285
Right Cingulate Cortex	0.153	0.094	0.086
Left Insula	0.164	0.210	0.266
Right Insula	0.185	0.244	0.291

Table S.4.1: P-values: Joint significance tests of 14 lobes.

human brain and are essential for voluntary movement, expressive language, and higher-level executive functions. These executive functions encompass a range of cognitive abilities, including planning, organizing, initiating, self-monitoring, and controlling responses to achieve specific goals. Damage to the frontal lobe can result in impairments in these areas, affecting one's ability to perform tasks that require these skills.

The parietal lobe, situated behind the frontal lobe, plays a pivotal role in integrating sensory information from various parts of the body. It is responsible for processing touch, temperature, pressure, and pain sensations. Additionally, the left parietal lobe is involved in controlling skilled motor actions and is associated with numerical processing and language functions. Damage to this area can lead to difficulties in these domains, affecting one's ability to perform tasks that require these skills.

S.5 Basic definitions and additional properties

We present the proof of technical results in the manuscript. In this section, we first introduce the necessary notations and definitions. Secondly, we introduce additional results on the properties of Marčenko-Pastur Equation (5) and Eq. (7). Throughout this section, Conditions **C1**– **C6** are assumed and will not be explicitly restated.

S.5.1 Basic definitions

Throughout, we use c , c' , C , or C' to denote a general constant that may vary from line to line. We always use $E = E(z)$ and $\eta = \eta(z)$ to denote the real and imaginary part of a complex value z . The argument z is frequently omitted from mathematical expressions.

Definition S.5.1 (Matrix norms). *Let $A = (a_{ij})$ be a complex matrix. We define the following norms*

$$\|A\| = \max_{\|x\|=1} \|x^* A x\|, \quad \|A\|_\infty = \max_{i,j} |a_{ij}|, \quad \|A\|_F = \sqrt{\text{tr}(AA^*)},$$

where A^* is the conjugate transpose of a matrix A . If A is a vector, we write $|A| = \|A\|$ for simplicity.

The following notion of a high-probability bound has been used in a number of works on random matrix theory. It provides a simple way of systematizing and making precise statements of the form “ ξ is bounded with high probability by ζ up to small powers of n ”.

Definition S.5.2 (Stochastic domination).

(i) *Consider two families of nonnegative random variables*

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

where $U^{(n)}$ is a possibly n -dependent parameter set. We say that ξ is stochastically dominated by ζ , uniformly in u , if for all (small) $\varepsilon > 0$ and (large) $D > 0$ we have

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

for large enough $n \geq n_0(\varepsilon, D)$. Throughout this paper the stochastic domination will always be uniform in all parameters (such as matrix indices, deterministic vectors, and spectral parameters z in certain domain that are not explicitly fixed. If ξ is

stochastically dominated by ζ , uniformly in u , we use the notation $\xi < \zeta$. Moreover, if for some complex family ξ we have $|\xi| < \zeta$ we also write $\xi = O_{<}(\zeta)$.

(ii) We extend the definition of $O_{<}(\cdot)$ to matrices in the weak operator sense as follows.

Let A be a family of complex square random matrices and ζ a family of nonnegative random variables. Then we use $A = O_{<}(\zeta)$ to mean $|\langle \mathbf{v}, A\mathbf{w} \rangle| < \zeta |\mathbf{v}| |\mathbf{w}|$ uniformly for all deterministic vectors \mathbf{v} and \mathbf{w} .

(iii) If there exists a positive constant C such that $\xi \leq C\zeta$, then we write $\xi \lesssim \zeta$. If further there exists a positive constant C' such that $\zeta \leq C'\xi$, then we write $\xi \asymp \zeta$.

Definition S.5.3. We say that an event Λ holds with high probability if for any large positive constant D , there exists $n_0(D)$ such that

$$\mathbb{P}(\Lambda^c) \leq n^{-D}, \quad \text{for any } n \geq n_0(D).$$

S.5.2 Additional properties of the Marčenko-Pastur Equation

We present additional properties of the solution to the Marčenko-Pastur Equation (5). Applying Lemma 2.2 with the population ESD being $\mathcal{G}_\lambda(1/\tau)$ (the composition of the inverse operator and \mathcal{G}_λ) and the dimension-to-sample-size ratio being γ_1 , we have that for any $z \in \mathbb{C}^+$, there is a unique solution $q(z)$ in \mathbb{C}^+ to

$$z = -\frac{1}{q} + \gamma_1 \int \frac{d\mathcal{G}_\lambda(\tau)}{\tau + q}. \quad (\text{S.3})$$

Moreover, $q(z)$ is the Stieltjes transform of a probability measure, denoted to be \mathcal{F}_q , compactly supported in $(0, \infty)$.

Recall the definition of $\phi(z)$: $\mathbb{C}^+ \rightarrow \mathbb{C}^+$ as the Stieltjes transform of \mathcal{G}_λ (See Lemma 2.4). That is,

$$\phi(z) = \int \frac{d\mathcal{G}_\lambda(\tau)}{\tau - z}, \quad z \in \mathbb{C}^+.$$

For convenience, we extend the definition of $\phi(z)$ to $z \in \mathbb{C}^- = \{E + i\eta : \eta < 0\}$ by setting $\phi(z) = \overline{\phi(\bar{z})}$ if $\eta < 0$. By doing so, we can concisely write (S.3) as

$$z = -\frac{1}{q(z)} + \gamma_1 \phi(-q(z)).$$

Recall that $s(x)$ is the extension of $\phi(z)$ to the real line. Define the function

$$f(q) = -\frac{1}{q} + \gamma_1 s(-q), \quad \text{for } q \in (-\rho, 0).$$

Recall the definition of β and Θ_1 in Theorem 2.9. The following relationship holds. The critical point of $f(q)$ is $-\beta$ since $f'(-\beta) = 1/\beta^2 - \gamma_1 s'(\beta) = 0$. Moreover,

$$\Theta_1 = f(-\beta) = \frac{1}{\beta} + \gamma_1 s(\beta).$$

Lemma S.5.4. *The following results are known for \mathcal{F}_q and $q(z)$.*

- (1) *The rightmost edge of the support of \mathcal{F}_q is Θ_1 .*
- (2) *Fix any small constant $a > 0$. Then, there exists a constant $C > 0$, depending on a , such that*

$$C^{-1} \leq |q(z)| \leq C, \quad \text{and} \quad \Im q(z) \geq C^{-1}\eta,$$

for all $z \in \mathbb{C}^+$ satisfying $a \leq |z| \leq a^{-1}$.

- (3) *There exists sufficiently small constants $C > 0$ and $C' > 0$ such that for all $z \in \{E + i\eta : |E - \Theta_1| \leq C, \ 0 < \eta < C^{-1}\}$*

$$\inf_{\tau \in S_{\mathcal{G}_\lambda}} |\tau + q(z)| \geq C',$$

where $S_{\mathcal{G}_\lambda}$ is the support of \mathcal{G}_λ . It indicates that the equation (S.3) is non-singular near Θ_1 .

(4) There exists an open ball of Θ_1 (in \mathbb{C}), say \mathfrak{N}_{Θ_1} such that when $z = E + i\eta \in \mathfrak{N}_{\Theta_1} \cap \mathbb{C}^+$,

$$\Im q(z) \asymp \begin{cases} \sqrt{\kappa + \eta} & \text{if } E < \Theta_1, \\ \frac{\eta}{\sqrt{\kappa + \eta}} & \text{if } E > \Theta_1, \end{cases}$$

where $\kappa = |E - \Theta_1|$.

These results are well-known in the RMT literature. See Silverstein and Choi [1995], Bao et al. [2015], Knowles and Yin [2017] and the references therein.

Next, we are interested in the behavior of Eq. (S.3) when a small perturbation is added to z and z is near Θ_1 . For this purpose, we consider the following domains of z . For the positive constants a and a' , we define

$$D = D(a, a', n_1) := \{z = E + i\eta \in \mathbb{C}^+ : |E - \Theta_1| \leq a, n_1^{-1+a'} \leq \eta \leq 1/a'\}. \quad (\text{S.4})$$

In the subsequent analysis, we shall always choose a and a' to be such that Result (3) of Lemma S.5.4 holds for all $z \in D$.

The following lemma indicates that if a small perturbation is added to z when $z \in D$, Eq. (S.3) is stable in the sense that $q(z)$ will only have small changes. The results are a combination of Definition 5.4 and Definition A.2 of Knowles and Yin [2017]. Therefore, the proof is omitted.

Lemma S.5.5 (Strong stability of Eq. (S.3)). *Eq. (S.3) is strongly stable in the following sense. Suppose that $\delta : D \rightarrow (0, \infty)$ satisfies $p^{-2} \leq \delta(z) \leq 1/\log(p)$ for $z \in D$ and that δ is Lipschitz continuous with Lipschitz constant p^2 . Suppose moreover that for each fixed $z \in D$, the function $\eta \rightarrow \delta(E + i\eta)$ is nonincreasing for $\eta > 0$. Suppose that $u : D \rightarrow \mathbb{C}^+$ is the Stieltjes transform of a compactly supported probability measure. Let $z \in D$ and suppose that*

$$|\tilde{z} - z| \leq \delta(z),$$

where $\tilde{z} = -1/u(z) + \gamma_2\phi(-u(z))$. Note that $z = -1/q(z) + \gamma_2\phi(-q(z))$ and $u(z) = q(\tilde{z})$.

If $\Im z < 1$, suppose also that

$$|u - q| \leq \frac{C\delta}{\sqrt{\kappa + \eta} + \sqrt{\delta}} \quad (\text{S.5})$$

holds at $z + ip^{-5}$. Here, $\kappa = |E - \Theta_1|$. Then, Eq. (S.5) holds at z .

It is worth mentioning that if for some $z_0 \in D$ with $\eta(z_0) \geq 1$, $|\tilde{z}_0 - z_0| \leq \delta(z_0)$ is satisfied, then Eq. (S.5) holds for all z in

$$\left\{ z \mid E(z) = E(z_0), \quad \eta(z_0) - \eta(z) \in p^{-4}\mathbb{N} \right\} \cap D,$$

by induction.

S.5.3 Additional properties of the generalized Marčenk-Pastur equation for \mathbf{G}_λ

In this subsection, we present additional properties of $\phi(z)$. These results are analogous to those for $q(z)$ in Section S.5.2. Recall that as in Lemma 2.4, $\phi(z)$ is the solution to Eq. (7), cited here for the readers' convenience

$$\phi(z) = \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\tau\{[1 + \gamma_2\phi(z)]^{-1} - z\} + \lambda}.$$

Recall that as explained in Section 2.1, the equation can be reformulated to

$$z = h(z) + \left[1 + \gamma_2 \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{-\tau h(z) + \lambda} \right]^{-1},$$

where $h(z) = z - [1 + \gamma_2\phi(z)]^{-1}$.

Recall that ρ is defined to be the leftmost edge of the support of \mathcal{G}_λ . The following basic properties of $\phi(z)$ follow immediately from the arguments in Section 5 of Li [2024].

Lemma S.5.6 (Characterization of ρ). (1) If F^{Σ_∞} is continuous at σ_{\max} and Case (ii) of Definition 2.7 is satisfied, then $\rho > \lambda/\sigma_{\max}$ and \mathcal{G}_λ is continuous at ρ . There exists a neighborhood of ρ , say \mathfrak{N} , such that when $E \in \mathfrak{N}$ and $\eta > 0$,

$$\Im\phi(z) \asymp \begin{cases} \sqrt{\kappa + \eta} & \text{if } E > \rho, \\ \frac{\eta}{\sqrt{\kappa + \eta}} & \text{if } E < \rho, \end{cases}$$

where $\kappa = |E - \rho|$.

(2) If F^{Σ_∞} is discrete at σ_{\max} as in Case (i) of Definition 2.7 and additionally $\gamma_2\omega_{\max} < 1$, then the statements in (1) also hold.

(3) If F^{Σ_∞} is discrete at σ_{\max} as in Case (i) of Definition 2.7 and additionally $\gamma_2\omega_{\max} > 1$, then $\rho = \lambda/\sigma_{\max}$. As $z \rightarrow \rho$, $\Im\phi(z)$ diverges.

The expression of $\Im\phi(z)$ near ρ is proved using the argument in the proof of (A.7) of Knowles and Yin [2017] combined with the results in Section 5 of Li [2024]. The details are omitted.

We shall focus on Case (1) and Case (2) of Lemma S.5.6 in the rest of this section. For the positive constants a and a' , we define the following domains. First, we define a domain around ρ with the imaginary part at least $n_2^{-1+a'}$ as

$$Q = Q(a, a', n_2) = \{z \in \mathbb{C}^+ : |E| \leq a^{-1}, \quad E \leq \rho + a, \quad n_2^{-1+a'} \leq \eta \leq 1/a'\}.$$

Secondly, we restrict the real part to be smaller than ρ and define

$$Q_- = Q_-(a, a', n_2) = \{z \in Q(a, a', n_2) : E \leq \rho\}.$$

Thirdly, we define a domain that is away from the support of \mathcal{G}_λ as

$$Q_{\text{away}} = Q_{\text{away}}(a, a', n_2) = \{z \in \mathbb{C}^+ : a \leq \text{dist}(z, \mathcal{I}) \leq a^{-1}, \eta \geq n_2^{-1+a'}\},$$

where $\mathcal{I} = [\rho, (1 + \sqrt{\gamma_2})^2 + \lambda/\liminf \ell_{\min}(\Sigma_p)]$. Here, we are using the fact that the rightmost edge of the support of \mathcal{G}_λ is bounded by $(1 + \sqrt{\gamma_2})^2 + \lambda/\liminf \ell_{\min}(\Sigma_p)$.

The following boundedness property of $\phi(z)$ follow immediately from the arguments in Section 5 of Li [2024].

Lemma S.5.7 (Boundedness of $\phi(z)$). *Suppose that $\rho > \lambda/\sigma_{\max}$. There exists constants a , a' , C and C' such that for all $z \in Q$,*

$$|\phi(z)| \leq C', \quad \Im \phi(z) \geq C\eta, \quad |1 + \gamma_2 \phi(z)| \geq C, \quad \text{and} \quad \inf_{\tau \in S} |-\tau h(z) + \lambda| \geq C,$$

where S is the support of F^{Σ_∞} .

Instead, suppose that $\rho = \lambda/\sigma_{\max}$. Then, for all z such that $|E| \leq a^{-1}$, $E < \rho - a$, $n_2^{-1+a'} \leq \eta \leq 1/a'$, the same bounds hold.

In the subsequent analysis, we shall always choose a and a' to be such that the boundedness of $\phi(z)$ in Lemma S.5.7 holds for $z \in Q$. It implies that Eq. (7) is non-singular when $z \in Q$.

The following lemma is analogous to Lemma S.5.5. It indicates that if a small perturbation is added to z when $z \in Q$, then Eq. (7) is stable in the sense that $\phi(z)$ or $h(z)$ will only have limited changes.

Lemma S.5.8 (Strong stability of Eq. (7)). *Suppose that $\rho > \lambda/\sigma_{\max}$. Eq. (7) is strongly stable in Q in the following sense. Suppose that $\delta : Q \rightarrow (0, \infty)$ satisfies $p^{-2} \leq \delta(z) \leq 1/\log(p)$ for $z \in Q$ and that δ is Lipschitz continuous with Lipschitz constant p^2 . Suppose moreover that for each fixed $z \in Q$, the function $\eta \rightarrow \delta(E + i\eta)$ is nonincreasing for $\eta > 0$. Suppose that $v(z) = z - [1 + \gamma_2 \varrho(z)]^{-1}$ where $\varrho(z)$ is the Stieltjes transform of a compactly supported probability measure. Let $z \in Q$ and suppose that*

$$|\tilde{z} - z| \leq \delta(z),$$

where

$$\tilde{z} = v(z) + \left[1 + \gamma_2 \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{-\tau v(z) + \lambda} \right]^{-1}.$$

Note that $v(z) = h(\tilde{z})$.

If $\Im z < 1$, suppose also that

$$|v - h| \leq \frac{C\delta}{\sqrt{|E - \rho| + \eta} + \sqrt{\delta}} \quad (\text{S.6})$$

holds at $z + ip^{-5}$. Then, Eq. (S.6) holds at z .

Moreover, when z is away from the support of \mathcal{G}_λ . The bound can be improved. The following lemma holds for both the case $\rho > \lambda/\sigma_{\max}$ and $\rho = \lambda/\sigma_{\max}$.

Lemma S.5.9. *Eq. (7) is strongly stable in Q_{away} in the following sense. Suppose that $\delta : Q_{\text{away}} \rightarrow (0, \infty)$ satisfies $\delta(z) \leq 1/\log(p)$ for $z \in Q_{\text{away}}$. Suppose that $v(z) = z - [1 + \gamma_2 \varrho(z)]^{-1}$ where $\varrho(z)$ is the Stieltjes transform of a compactly supported probability measure. Let $z \in Q_{\text{away}}$ and suppose that*

$$|\tilde{z} - z| \leq \delta(z),$$

where

$$\tilde{z} = v(z) + \left[1 + \gamma_2 \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{-\tau v(z) + \lambda} \right]^{-1}.$$

Then,

$$|v(z) - h(z)| \leq C\delta(z),$$

for any $z \in Q_{\text{away}}$.

Corollary S.5.10. *It is worth mentioning that if for some $z_0 \in Q$ with $\eta(z_0) \geq 1$, $|\tilde{z}_0 - z_0| \leq \delta(z_0)$ is satisfied, then Eq. (S.6) holds for all z in*

$$\left\{ z \mid E(z) = E(z_0), \quad \eta(z_0) - \eta(z) \in p^{-4}\mathbb{N} \right\} \cap Q.$$

It follows easily by induction.

Corollary S.5.11. *Under the conditions of Lemma S.5.8, if Eq. (S.6) holds, then we also have*

$$|\varrho - \phi| \leq \frac{c\delta}{\sqrt{\kappa + \eta} + \sqrt{\delta}}.$$

It is obvious since

$$v - h = \frac{\gamma_2(\phi - \varrho)}{(1 + \gamma_2\varrho)(1 + \gamma_2\phi)}$$

and $|1 + \gamma_2\phi(z)| \asymp 1$, $|1 + \gamma_2\varrho(z)| \asymp 1$ when $z \in Q$.

Corollary S.5.12. *Using Lemma S.5.8, we immediately obtain the following result. Suppose that $\varrho(z)$ satisfies*

$$\varrho(z) = \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\tau\{[1 + \gamma_2\varrho(z) + \delta_2(z)]^{-1} - z\} + \lambda} + \delta_1(z), \quad z \in Q.$$

Here, δ_1 and δ_2 are such that $|\delta_1|$ and $|\delta_2|$ satisfy the conditions on δ in Lemma S.5.8. If $\Im z < 1$, suppose also that

$$|\varrho - \phi| \leq \frac{C|\delta_1| + C|\delta_2|}{\sqrt{\kappa + \eta} + \sqrt{|\delta_1| + |\delta_2|}} \tag{S.7}$$

holds at $z + ip^{-5}$. Then, Eq. (S.7) holds at z .

Proof of Lemma S.5.8. Lemma S.5.8 can be shown following very similar arguments to those in the proof of Lemma S.5.5 presented in Section A.2 of Knowles and Yin [2017]. Indeed, the key arguments are Lemma A.3 of Knowles and Yin [2017] which states the regular behavior of the derivatives of the function

$$f(q) = -\frac{1}{q} + \gamma_1 s(-q)$$

near its critical points. To prove Lemma S.5.8, we replace the function with

$$f_{\text{new}}(h) = h + \left[1 + \gamma_2 \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{-\tau h + \lambda} \right]^{-1},$$

Section 5 of Li [2024] indicates that the function has the same behavior at its critical point h_0 (See Lemma 3.1 for the definition of h_0). Moreover, a similar quadratic equation of $(v - h)$ to Eq. (A.13) of Knowles and Yin [2017] can be easily obtained using Taylor's expansion. The rest arguments in Section A.2 of Knowles and Yin [2017] can be used nearly verbatim on the new definition of $f_{\text{new}}(\cdot)$. We therefore omit further details. \square

Proof of Lemma S.5.9. When z is away from the support of \mathcal{G}_λ , as when $z \in Q_{\text{away}}$, $|h'(\tilde{z})|$ is bounded. Therefore,

$$|v(z) - h(z)| = |h(\tilde{z}) - h(z)| \leq \int_z^{\tilde{z}} |h(\zeta)| |d\zeta| \leq C\delta(z).$$

\square

Proof of Corollary S.5.12. The result follows from Lemma S.5.8 and Corollary S.5.11 because the specified ϱ is such that

$$\left| \tilde{v}(z) + \left[1 + \gamma_2 \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{-\tau \tilde{v}(z) + \lambda} \right]^{-1} - z \right| \leq C(|\delta|_1 + |\delta|_2),$$

where $\tilde{v}(z) = z - [1 + \gamma_2 \varrho(z)]^{-1}$. \square

S.6 Properties of \mathbf{G}_λ under Gaussianity

In this section, we mainly study the properties of \mathbf{G}_λ under the following additional conditions:

SC1 \mathbf{Z} is normally distributed.

SC2 $\rho > \lambda/\sigma_{\max}$.

The main results in this section can be extended to the case when $\rho = \lambda/\sigma_{\max}$. The extension is presented in Section S.6.4.

When \mathbf{Z} is normally distributed, \mathbf{G}_λ has the same distribution as

$$\mathbf{G}_\lambda \stackrel{\text{def}}{=} \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T + \lambda\Sigma_p^{-1},$$

where $\tilde{\mathbf{Z}}$ is $p \times n_2$ whose entries are iid $N(0, 1/n_2)$. Results in this section are expected to hold when \mathbf{Z} satisfies **C2**. The extension to non-Gaussianity can be completed following the strategy in Sections 7–10 of Knowles and Yin [2017], which is however beyond our scope.

Importantly, under **C5** and **SC2**, for all sufficiently large p ,

$$(p/n_2)F^{\Sigma_p}(\ell_{\max}(\Sigma_p)) \leq 1 - \epsilon \text{ for some } \epsilon > 0. \quad (\text{S.8})$$

In the subsequent analysis, it is more convenient to consider a finite-sample version of Eq. (7). Specifically, let $\phi_p(z)$ be the unique solution in \mathbb{C}^+ to

$$\phi_p = \int \frac{\tau dF^{\Sigma_p}(\tau)}{\tau\{[1 + (p/n_2)\phi_p]^{-1} - z\} + \lambda}, \quad z \in \mathbb{C}^+. \quad (\text{S.9})$$

Moreover, define

$$x_p(h) = h + \left[1 + (p/n_2) \int \frac{\tau dF^{\Sigma_p}(\tau)}{-\tau h + \lambda} \right]^{-1}.$$

Let ρ_p be the finite-sample version of ρ when $(\gamma_2, F^{\Sigma_\infty})$ is replaced by $(p/n_2, F^{\Sigma_p})$. That

is, $\rho_p = x_p(h_0^{(p)})$ where $h_0^{(p)}$ is the unique solution in $(0, \lambda/\ell_{\max}(\Sigma_p))$ to $x'_p(h_0^{(p)}) = 0$. It is a simple matter to verify that $h_0^{(p)}$ exists and is unique. It is because x'_p is monotonically decreasing and $x'_p(h) < 0$ for all h close to $\lambda/\ell_{\max}(\Sigma_p)$ when p is sufficiently large because of Eq. (S.8)..

It is not hard to verify that $\phi_p(z)$ also satisfies the properties of $\phi(z)$ presented in Section S.6, when all the parameters are replaced by their finite-sample version. We will later refer to these properties on $\phi_p(z)$ without any proof.

Lemma S.6.1. *Suppose that **C1–C6** and **SC1–SC2** hold. We have*

$$p^{2/3}|h_0^{(p)} - h_0| \rightarrow 0, \quad p^{2/3}|\rho_p - \rho| \rightarrow 0.$$

Proof of Lemma S.6.1. Recall that $\rho = x(h_0)$ where $x'(h_0) = 0$ and $h_0 < \lambda/\sigma_{\max}$. For sufficiently large p , $h_0 < \lambda/\ell_{\max}(\Sigma_p) - \epsilon$ for all sufficiently small $\epsilon > 0$. Notice that

$$\rho_p - \rho = x_p(h_0^{(p)}) - x(h_0) = x_p(h_0^{(p)}) - x_p(h_0) + x_p(h_0) - x(h_0) = x'_p(\tilde{h})(h_0^{(p)} - h_0) + (x_p(h_0) - x(h_0)),$$

where \tilde{h} is a point in between h_0 and $h_0^{(p)}$. The second term on the right hand side is clearly $o(p^{-2/3})$ since under **C6**, $\|F^{\Sigma_\infty} - F^{\Sigma_p}\| = o(p^{-2/3})$ and $|p/n_2 - \gamma_2| = o(p^{-2/3})$.

It remains to show $|h_0^{(p)} - h_0| = o(p^{-2/3})$. To this end, observe that for some \tilde{h} in between $h_0^{(p)}$ and h_0 ,

$$0 = x'_p(h_0^{(p)}) = x'_p(h_0) + x''_p(\tilde{h})(h_0^{(p)} - h_0).$$

While $C \leq |x''_p(h)| \leq C^{-1}$ when h is bounded away from $\lambda/\ell_{\max}(\Sigma_p)$, we obtain that $|h_0^{(p)} - h_0| = O(x'_p(h_0))$. While $x'(h_0) = 0$, $x'_p(h_0) = o(p^{-2/3})$ since $\|F^{\Sigma_\infty} - F^{\Sigma_p}\| = o(p^{-2/3})$ and $|p/n_2 - \gamma_2| = o(p^{-2/3})$. It completes the proof of the lemma. \square

In this section, we aim to establish the following technical results. We first introduce a

fundamental control parameter as

$$\Psi(z) = \sqrt{\frac{\Im \phi_p(z)}{n_2 \eta}} + \frac{1}{n_2 \eta}.$$

Recall the definition of the domain Q , Q_- and Q_{away} in Section S.5. We define a finite-sample version of the domains as

$$Q^{(p)} = Q^{(p)}(a, a', n_2) = \{z \in \mathbb{C}^+, \quad |E| \leq a^{-1}, \quad E \leq \rho_p + a, \quad n_2^{-1+a'} \leq \eta \leq 1/a'\}.$$

$$Q_-^{(p)} = Q_-^{(p)}(a, a', n_2) = \{z \in Q^{(p)}(a, a', n_2), \quad E \leq \rho_p - n_2^{-2/3+a'}, \quad n_2^{-2/3} \leq \eta \leq 1/a'\}.$$

$$Q_{\text{away}}^{(p)} = Q_{\text{away}}^{(p)}(a, a', n_2) = \{z \in \mathbb{C}^+ : a \leq \text{dist}(z, \mathcal{I}) \leq a^{-1}, \quad \eta \geq n_2^{-1+a'}\},$$

where $\mathcal{I} = [\rho_p, (1 + \sqrt{p/n_2})^2 + \lambda/\ell_{\min}(\Sigma_p)]$.

Call the resolvent of \mathbf{G}_λ

$$R(z) = (\mathbf{G}_\lambda - zI_p)^{-1}, \quad z \in \mathbb{C}^+.$$

Theorem S.6.2 (Strong local law of $R(z)$). *Suppose that C1–C6 and SC1–SC2 hold.*

There exists constants $a > 0$ such that the following local laws hold for arbitrary $a' > 0$.

(i) *The entrywise local law in $Q^{(p)}$ holds as*

$$R(z) - (\lambda \Sigma_p^{-1} - m_p(z) I_p)^{-1} = O_{<}(\Psi(z)),$$

uniformly in $Q^{(p)}$, where $m_p(z) = z - [1 + (p/n_2)\phi_p(z)]^{-1}$.

(ii) *The averaged local law in $Q^{(p)}$ holds as*

$$\hat{\phi}_p(z) - \phi_p(z) = O_{<} \left(\frac{1}{n_2 \eta} \right),$$

where $\hat{\phi}_p(z) = p^{-1} \text{tr} R(z)$.

(iii) The averaged local law in $Q_-^{(p)}$ holds as

$$\hat{\phi}_p(z) - \phi_p(z) = O_{<} \left(\frac{1}{n_2(\kappa + \eta)} \right),$$

where $\kappa = |\rho_p - E|$.

(iv) The averaged local law in $Q_{\text{away}}^{(p)}$ holds as

$$\hat{\phi}_p(z) - \phi_p(z) = O_{<} \left(\frac{1}{n_2} \right).$$

Once Theorem S.6.2 is ready, it is not hard to deduce the following rigidity of the smallest eigenvalue of \mathbf{G}_λ .

Theorem S.6.3 (Rigidity of the smallest eigenvalue of \mathbf{G}_λ under Gaussianity). *Under the conditions of Theorem S.6.2, for any fixed $\epsilon \in (0, 2/3)$, with high probability, there is no eigenvalue of \mathbf{G}_λ in $(-\infty, \rho_p - n_2^{-2/3+\epsilon})$.*

The following theorem shows the convergence of *linear spectral statistics* of \mathbf{G}_λ .

Theorem S.6.4. *Suppose the conditions of Theorem S.6.2 hold. Consider the interval*

$$\mathcal{I}_\epsilon = [\rho_p - \epsilon, (1 + \sqrt{p/n_2})^2 + \lambda / \liminf \ell_{\min}(\Sigma_p) + \epsilon],$$

where ϵ is any fixed positive constant.

(i) *Then, with high probability, all eigenvalues of \mathbf{G}_λ are contained in \mathcal{I}_ϵ . We denote the event as \mathfrak{M} .*

(ii) *Let f be any function analytic on an open interval containing the interval \mathcal{I}_ϵ . We have*

$$\mathbb{1}(\mathfrak{M}) p^{2/3} \left| \frac{1}{p} \sum_{j=1}^p f(\ell_j(\mathbf{G}_\lambda)) - \int f(\tau) d\mathcal{G}_\lambda(\tau) \right| < \frac{1}{n_2}.$$

S.6.1 Proof of Theorem S.6.2

Throughout the section, we focus on the representation $\mathbf{G}_\lambda = \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T + \lambda\Sigma_p^{-1}$ where $\tilde{\mathbf{Z}}$ is $p \times n_2$ whose entries are iid $N(0, 1/n_2)$. We shall temporarily assume that Σ_p is diagonal with diagonal elements $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p$. The extension to general Σ_p is given in Section S.6.1.5.

In the subsequent analysis, we closely follow the arguments in Sections 4 and 5 of Knowles and Yin [2017]. We therefore omit many details and refer to the corresponding arguments in Knowles and Yin [2017]. We frequently omit the arguments z from our notation.

S.6.1.1 Basic tools

Define the $(p + n_2) \times (p + n_2)$ block matrix

$$\mathbf{J}(z) = \mathbf{K}(z)^{-1}, \quad \mathbf{K}(z) = \begin{pmatrix} \lambda\Sigma_p^{-1} - zI_p & \tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}^T & -I_{n_2} \end{pmatrix}. \quad (\text{S.10})$$

The reason for considering \mathbf{J} and \mathbf{K} is that \mathbf{J} contains the resolvent $R(z)$ as one of its blocks (see Lemma S.6.5 below).

Call

$$\Omega(z) = \begin{pmatrix} (\lambda\Sigma_p^{-1} - m_p(z)I_p)^{-1} & 0 \\ 0 & (z - m_p(z))I_{n_2} \end{pmatrix}.$$

We shall later show that $\mathbf{J} - \Omega = O_{<}(\Psi)$ in $Q^{(p)}$. It actually implies the entrywise local law in $Q^{(p)}$.

We make the following notation. We use A_{st} to denote the (s, t) -th element of a matrix A . The four blocks of $\mathbf{J}(z)$ are called $\mathbf{J}_{(11)}$, $\mathbf{J}_{(12)}$, $\mathbf{J}_{(21)}$ and $\mathbf{J}_{(22)}$. For a set of indices, say $S \subset \{1, 2, \dots, p + n_2\}$, we define the minors $\mathbf{K}^{(S)} = (\mathbf{K}_{st}, s, t \notin S)$. That is, \mathbf{K} but with the rows and columns whose indices specified in S deleted. We also write $\mathbf{J}^{(S)} = (\mathbf{K}^{(S)})^{-1}$. The block formed by the first p rows and columns of $\mathbf{J}^{(S)}$ is called $\mathbf{J}_{(11)}^{(S)}$. The block formed by the last n_2 rows and columns of $\mathbf{J}^{(S)}$ is called $\mathbf{J}_{(22)}^{(S)}$. Similar for $\mathbf{J}_{(12)}^{(S)}$ and $\mathbf{J}_{(21)}^{(S)}$. Similar

notation is made for the matrix Ω .

For $u = 1, 2, \dots, p$, call the u -th row of $\tilde{\mathbf{Z}}$ to be $\tilde{\mathbf{Z}}_u$. For $i = p + 1, \dots, p + n_2$, we use $\tilde{\mathbf{Z}}_{\cdot i}$ to denote the $(i - p)$ -th column of $\tilde{\mathbf{Z}}$. Moreover, if $u \in \{1, \dots, p\}$, let e_u denote the standard unit vector of dimension p in the coordinate direction u . If $i \in \{p + 1, \dots, p + n_2\}$, let e_i denote the standard unit vector of dimension n_2 in the coordinate direction $(i - p)$.

The following lemma is a variant of Lemma 4.4 in Knowles and Yin [2017]. These results follow directly from the Schur complement formula and resolvent identities that have been previously derived in Knowles and Yin [2017] and the references therein. The proof is therefore omitted.

Lemma S.6.5 (Resolvent Identities). *Suppose that Σ_p is diagonal with diagonal elements $\sigma_1, \dots, \sigma_p$.*

(i) *We have*

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{(11)} & \mathbf{J}_{(12)} \\ \mathbf{J}_{(21)} & \mathbf{J}_{(22)} \end{pmatrix} = \begin{pmatrix} R & R\tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}^T R & \tilde{\mathbf{Z}}^T R \tilde{\mathbf{Z}} - I_{n_2} \end{pmatrix}.$$

(ii) *For $u \in \{1, 2, \dots, p\}$, we have*

$$\frac{1}{\mathbf{J}_{uu}} = \lambda/\sigma_u - z - \tilde{\mathbf{Z}}_u \mathbf{J}_{(22)}^{(u)} \tilde{\mathbf{Z}}_u^T.$$

For $u \neq v \in \{1, 2, \dots, p\}$,

$$\mathbf{J}_{uv} = -\mathbf{J}_{uu} \tilde{\mathbf{Z}}_u \mathbf{J}_{(22)}^{(u)} e_v = -\mathbf{J}_{vv} e_u^T \mathbf{J}_{(22)}^{(v)} \tilde{\mathbf{Z}}_v^T = \mathbf{J}_{uu} \mathbf{J}_{vv}^{(u)} \tilde{\mathbf{Z}}_u \mathbf{J}_{(22)}^{(uv)} \tilde{\mathbf{Z}}_v^T.$$

(iii) *For $i \in \{p + 1, \dots, p + n_2\}$, we have*

$$\frac{1}{\mathbf{J}_{ii}} = -1 - \tilde{\mathbf{Z}}_{\cdot i}^T \mathbf{J}_{(11)}^{(i)} \tilde{\mathbf{Z}}_{\cdot i}.$$

For $i \neq j \in \{p + 1, \dots, p + n_2\}$, we have

$$\mathbf{J}_{ij} = -\mathbf{J}_{ii} \tilde{\mathbf{Z}}_{\cdot i}^T \mathbf{J}_{(11)}^{(i)} e_j = -\mathbf{J}_{jj} e_i^T \mathbf{J}_{(11)}^{(j)} \tilde{\mathbf{Z}}_{\cdot j} = \mathbf{J}_{ii} \mathbf{J}_{jj}^{(i)} \tilde{\mathbf{Z}}_{\cdot i}^T \mathbf{J}_{(11)}^{(ij)} \tilde{\mathbf{Z}}_{\cdot j}.$$

(iv) For $u \in \{1, 2, \dots, p\}$ and $i \in \{p+1, \dots, p+n_2\}$,

$$\mathbf{J}_{ui} = -\mathbf{J}_{ii} e_u^T \mathbf{J}_{(11)}^{(i)} \tilde{\mathbf{Z}}_{\cdot i}.$$

(v) If $i, j, k \notin S$ and $i, j \neq k$,

$$\mathbf{J}_{ij}^{(S)} = \mathbf{J}_{ij}^{(S \cup \{k\})} + \frac{\mathbf{J}_{ik}^{(S)} \mathbf{J}_{kj}^{(S)}}{\mathbf{J}_{kk}^{(S)}}.$$

The following result is our fundamental tool for estimating entries of \mathbf{J} .

Lemma S.6.6. Fix $a > 0$. Then, the following estimates hold for any $z \in Q^{(p)}$. We have

$$\|\mathbf{J}\| \leq \frac{C}{\eta^3}, \quad \|\partial z \mathbf{J}\| \leq \frac{C}{\eta^6}.$$

Furthermore, let $\mathbf{w} \in \mathbb{R}^p$ and $\mathbf{v} \in \mathbb{R}^{n_2}$. Then, we have the bounds

$$\begin{aligned} \sum_{u=1}^p |\mathbf{w}^T \mathbf{J}_{(11)} e_u|^2 &= \frac{\Im \mathbf{w}^T \mathbf{J}_{(11)} \mathbf{w}}{\eta}, \\ \sum_{i=p+1}^{p+n_2} |\mathbf{v}^T \mathbf{J}_{(22)} e_i|^2 &\leq \frac{C \|\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^T\|}{\eta} \Im \mathbf{v}^T \mathbf{J}_{(22)} \mathbf{v} + C \mathbf{v}^T \mathbf{v}, \\ \sum_{u=1}^p |\mathbf{v}^T \mathbf{J}_{(21)} e_u|^2 &= \frac{\Im \mathbf{v}^T \mathbf{J}_{(22)} \mathbf{v}}{\eta}, \\ \sum_{i=p+1}^{p+n_2} |\mathbf{w}^T \mathbf{J}_{(12)} e_i|^2 &\leq C \|\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^T\| \sum_{u=1}^p |\mathbf{w}^T \mathbf{J}_{(11)} e_u|^2. \end{aligned}$$

The estimates remain true for $\mathbf{J}^{(S)}$ instead of \mathbf{J} if $S \subset \{1, \dots, p\}$ or $S \subset \{p+1, \dots, p+n_2\}$.

Proof. First, consider the bound on the operator norm of \mathbf{J} .

$$\|\mathbf{J}\| \leq \left\| \begin{pmatrix} (\lambda \Sigma_p^{-1} - z I_p)^{-1} & 0 \\ 0 & I \end{pmatrix} \right\| \left\| \begin{pmatrix} I & (\lambda \Sigma_p^{-1} - z I)^{-1/2} \tilde{\mathbf{Z}} \\ \tilde{\mathbf{Z}}^T (\lambda \Sigma_p^{-1} - z I)^{-1/2} & -I \end{pmatrix} \right\|^{-1} = \|J_1\| \|J_2\|, \text{ say.}$$

For J_1 , we clearly have $\|J_1\| \leq C/(|E - \lambda/\sigma_1| + \eta) \leq C/\eta$, for $z \in Q^{(p)}$. The eigen structure

of a matrix of the form J_2 is analyzed in (4.12-4.15) of Knowles and Yin [2017]. Based on their arguments,

$$\begin{aligned}\|J_2\| &\leq C(1 + \max_u \frac{|s_u| + 1}{|s_u - 1|}) \leq C(1 + \max_u \frac{(C/\eta)\|\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T\| + 1}{|s_u - 1|}) \\ &\leq C + (C/\eta)\|\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T\|\|\lambda\Sigma_p^{-1} - zI\|\|(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T - (\lambda\Sigma_p^{-1} - zI))^{-1}\| \leq C + \frac{C}{\eta^2},\end{aligned}$$

where s_k 's are the eigenvalues of $(\lambda\Sigma_p^{-1} - zI)^{-1/2}\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T(\lambda\Sigma_p^{-1} - zI)^{-1/2}$. The bound on $\|\mathbf{J}\|$ follows.

The bound on the operator norm of $\partial_z \mathbf{J}$ follows from the fact that $\|\partial_z \mathbf{J}\| = \|\mathbf{J}(\partial_z \mathbf{K})\mathbf{J}\| \leq \frac{C}{\eta^6}$.

Next, using Result (i) of Lemma S.6.5,

$$\sum_{u=1}^p |\mathbf{w}^T \mathbf{J}_{(11)} e_u|^2 = \sum_{u=1}^p \mathbf{w}^T \mathbf{J}_{(11)} e_u e_u^T \bar{\mathbf{J}}_{(11)} \mathbf{w} = \mathbf{w}^T \mathbf{J}_{(11)} \bar{\mathbf{J}}_{(11)} \mathbf{w} = \mathbf{w}^T \frac{\Im \mathbf{J}_{(11)}}{\eta} \mathbf{w}.$$

Moreover,

$$\begin{aligned}\sum_{i=p+1}^{p+n_2} |\mathbf{v}^T \mathbf{J}_{(22)} e_i|^2 &= \sum_{i=p+1}^{p+n_2} |\mathbf{v}^T \tilde{\mathbf{Z}}^T R \tilde{\mathbf{Z}} e_i - \mathbf{v}^T e_i|^2 \leq 2\mathbf{v}^T \tilde{\mathbf{Z}}^T \mathbf{J}_{(11)} \tilde{\mathbf{Z}} \tilde{\mathbf{Z}}^T \bar{\mathbf{J}}_{(11)} \tilde{\mathbf{Z}} \mathbf{v} + 2\mathbf{v}^T \mathbf{v} \\ &\leq C\|\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T\|\mathbf{v}^T \tilde{\mathbf{Z}}^T \mathbf{J}_{(11)} \bar{\mathbf{J}}_{(11)} \tilde{\mathbf{Z}} \mathbf{v} + C\mathbf{v}^T \mathbf{v} \leq C\|\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T\| \frac{\Im \mathbf{v}^T \mathbf{J}_{(22)} \mathbf{v}}{\eta} + C\mathbf{v}^T \mathbf{v}.\end{aligned}$$

The rest two results on $\mathbf{J}_{(12)}$ and $\mathbf{J}_{(21)}$ can be proved similarly using Result (i) of Lemma S.6.5. We omit the details. \square

Lemma S.6.7. *Under $\mathbf{C1}$, $\|\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T\| \leq (1 + \sqrt{p/n_2})^2 + \epsilon$ with high probability, for any fixed $\epsilon > 0$.*

The lemma follows from Theorem 2.10 of Bloemendal et al. [2014].

S.6.1.2 Weak entrywise law

We first show a weak entrywise local law of \mathbf{J} in $Q^{(p)}$.

Proposition S.6.8. *Suppose that the assumptions of Theorem S.6.2 hold. Define*

$$\Lambda = \max_{1 \leq s, t \leq p+n_2} |(\mathbf{J} - \Omega)_{st}| \quad \text{and} \quad \Lambda_o = \max_{s \neq t} |(\mathbf{J} - \Omega)_{st}|.$$

Then, $\Lambda < (n_2\eta)^{-1/4}$ uniformly in $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$.

Define the averaged control parameter

$$\mathfrak{B} = \mathfrak{B}_p + \mathfrak{B}_n, \quad \mathfrak{B}_p = \left| \frac{1}{p} \sum_{i=u}^p (\mathbf{J}_{uu} - \Omega_{uu}) \right|, \quad \mathfrak{B}_n = \left| \frac{1}{n_2} \sum_{i=p+1}^{p+n_2} (\mathbf{J}_{ii} - \Omega_{ii}) \right|.$$

Note that $\mathfrak{B}_p = |\hat{\phi}_p - \phi_p|$. We have the trivial bound $\mathfrak{B} \leq C\Lambda$.

For $s \in \{1, 2, \dots, p+n_2\}$, we introduce the conditional expectation

$$\mathbb{E}_s[\cdot] = \mathbb{E}[\cdot \mid \mathbf{K}^{(s)}].$$

Using Lemma S.6.5, we get that, for $u \in \{1, \dots, p\}$,

$$\frac{1}{\mathbf{J}_{uu}} = \lambda/\sigma_u - z - \frac{1}{n_2} \text{tr} \mathbf{J}_{(22)}^{(u)} - W_u, \quad W_u = (1 - \mathbb{E}_u)(\tilde{\mathbf{Z}}_u \mathbf{J}_{(22)}^{(u)} \tilde{\mathbf{Z}}_u^T). \quad (\text{S.11})$$

For $i \in \{p+1, \dots, p+n_2\}$,

$$\frac{1}{\mathbf{J}_{ii}} = -1 - \frac{1}{n_2} \text{tr} \mathbf{J}_{(11)}^{(i)} - W_i, \quad W_i = (1 - \mathbb{E}_i)(\tilde{\mathbf{Z}}_i^T \mathbf{J}_{(11)}^{(i)} \tilde{\mathbf{Z}}_i). \quad (\text{S.12})$$

We define the z -dependent event $\Xi = \{\Lambda \leq 1/(\log n_2)\}$ and the control parameter

$$\Psi_{\mathfrak{B}} = \sqrt{\frac{\Im \phi_p + \mathfrak{B}}{n_2\eta}}.$$

The following estimate is analogous to Lemma 5.2 of Knowles and Yin [2017].

Lemma S.6.9. *Suppose that the assumptions of Theorem S.6.2 hold. Then, uniformly for*

$s \in \{1, \dots, p + n_2\}$ and $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$, we have

$$\mathbb{1}(\Xi)(|W_s| + \Lambda_o) < \Psi_{\mathfrak{B}},$$

$$\mathbb{1}(\eta \geq 1)(|W_s| + \Lambda_o) < \Psi_{\mathfrak{B}}.$$

Proof of Lemma S.6.9. The proof relies on the identities from Lemma S.6.5 and large deviation estimates like Lemma 3.1 of Bloemendal et al. [2014]. The arguments in the proof are similar to those in Lemma 5.2 of Knowles and Yin [2017].

From (1) of Lemma S.5.7, when $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$, $|\Omega_{ss}| \asymp 1$. Therefore,

$$\mathbb{1}(\Xi)\mathbf{J}_{ss} \asymp 1.$$

Using Result (v) of Lemma S.6.5 and a simple induction argument, it is not hard to conclude that

$$\mathbb{1}(\Xi)\mathbf{J}_{ss}^{(S)} \asymp 1,$$

for any $S \subset \{1, \dots, p + n_2\}$ and $s \notin S$, satisfying $|S| \leq C$.

Let us first estimate Λ_o . When $u \neq v \in \{1, \dots, p\}$, using Result (ii) of Lemma S.6.5 and a large deviation estimation (Lemma 3.1 of Bloemendal et al. [2014]),

$$\mathbb{1}(\Xi)|\mathbf{J}_{uv}| \leq \mathbb{1}(\Xi)|\mathbf{J}_{uu}\mathbf{J}_{vv}^{(u)}| \left| \sum_{i,j=p+1}^{p+n_2} \tilde{Z}_{ui}\tilde{Z}_{vj}\mathbf{J}_{i,j}^{(uv)} \right| < \mathbb{1}(\Xi)C \left(\frac{1}{n_2^2} \sum_{i,j=p+1}^{p+n_2} |\mathbf{J}_{ij}^{(uv)}|^2 \right)^{1/2}. \quad (\text{S.13})$$

The term in parentheses is

$$\begin{aligned} \mathbb{1}(\Xi)\frac{1}{n_2^2} \sum_{i,j=p+1}^{p+n_2} |\mathbf{J}_{ij}^{(uv)}|^2 &= \mathbb{1}(\Xi)\frac{1}{n_2^2} \sum_{i,j=p+1}^{p+n_2} |e_i^T \mathbf{J}_{22}^{(uv)} e_j|^2 < \mathbb{1}(\Xi)\frac{1}{n_2^2 \eta} \|\tilde{Z}\tilde{Z}^T\| \sum_{i=p+1}^{p+n_2} \Im \mathbf{J}_{ii}^{(uv)} + \mathbb{1}(\Xi)\frac{1}{n_2} \\ &< \mathbb{1}(\Xi)\frac{1}{n_2^2 \eta} \|\tilde{Z}\tilde{Z}^T\| \sum_{i=p+1}^{p+n_2} \Im \mathbf{J}_{ii} + \mathbb{1}(\Xi)\frac{\Lambda_o^2}{n_2 \eta} + \mathbb{1}(\Xi)\frac{1}{n_2} < \frac{\Im \phi_p + \mathfrak{B} + \Lambda_o^2}{n_2 \eta} + \frac{1}{n_2} < \frac{\Im \phi_p + \mathfrak{B} + \Lambda_o^2}{n_2 \eta}. \end{aligned}$$

Here, we are using Result (v) of Lemma S.6.5, Lemma S.6.7 and Lemma S.6.6. The last

step is due to the fact that $1 < \Im \phi_p / \eta$.

A similar result for \mathbf{J}_{ij} , \mathbf{J}_{ui} , and \mathbf{J}_{iu} , $i \in \{p+1, p+n_2\}$ and $u \in \{1, p\}$ can be obtained using Result (iii) and (iv) of Lemma S.6.5 and a large deviation estimation. We omit the details as the arguments are very similar.

All together, we conclude that

$$\mathbb{1}(\Xi) \Lambda_o < \mathbb{1}(\Xi) \sqrt{\frac{\Im \phi_p + \mathfrak{B} + \Lambda_o^2}{n_2 \eta}}.$$

Since $1/\sqrt{n_2 \eta} \rightarrow 0$ when $z \in Q^{(p)}$, it is easy to deduce that $\mathbb{1}(\Xi) \Lambda_o < \Psi_{\mathfrak{B}}$.

An analogous argument for $|W_s|$ (see e.g. Lemma 5.2 of Erdős et al. [2013]) completes the proof of the statement in Proposition S.6.8 when Ξ holds.

As for the case when $\eta \geq 1$, we proceed similarly. For $\eta \geq 1$, we proceed as above to get $|W_s| < n_2^{-1/2}$, where we used that $\|\mathbf{J}\| \leq C/\eta \leq C$ by Lemma S.6.6. Similarly, as in Eq. (S.13), we get

$$\mathbf{J}_{uv} \leq |\mathbf{J}_{uu} \mathbf{J}_{vv}^{(u)}| \left| \sum_{i,j=p+1}^{p+n_2} \tilde{Z}_{ui} \tilde{Z}_{vj} \mathbf{J}_{i,j}^{(uv)} \right| < |\mathbf{J}_{uu} \mathbf{J}_{vv}^{(u)}| \frac{1}{\sqrt{n_2}},$$

where in the last step we used that $|\Im \mathbf{J}_{ii}^{(uv)}| \leq C$ because $\|\mathbf{J}\| \leq C$. Moreover, $|\mathbf{J}_{uu} \mathbf{J}_{vv}^{(u)}| \leq C$.

This completes the proof of Lemma S.6.9. \square

In the following, we aim to show that $\hat{\phi}_p$ satisfies a perturbed version of Eq. (S.9) when Ξ holds or when $\eta \geq 1$. From Eq. (S.11), Eq. (S.12), Lemma S.6.9, Result (v) of Lemma S.6.5, we get that uniformly for $u \in \{1, \dots, p\}$, $i \in \{p+1, \dots, p+n_2\}$, $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$,

$$\mathbb{1}(\Xi) \frac{1}{\mathbf{J}_{uu}} = \mathbb{1}(\Xi) \left(\lambda / \sigma_u - z - \frac{1}{n_2} \sum_{i=p+1}^{p+n_2} \mathbf{J}_{ii} - W_u + O_{<}(\Psi_{\mathfrak{B}}^2) \right), \quad (\text{S.14})$$

$$-\mathbb{1}(\Xi) \frac{1}{\mathbf{J}_{ii}} = \mathbb{1}(\Xi) \left(1 + \frac{1}{n_2} \sum_{u=1}^p \mathbf{J}_{uu} + W_i + O_{<}(\Psi_{\mathfrak{B}}^2) \right). \quad (\text{S.15})$$

Invert Eq. (S.15), get the Taylor's expansion of the right-hand side to the order of W_i^2 ,

and average the equation over $i = p + 1, \dots, p + n_2$. We obtain

$$-\mathbb{1}(\Xi) \frac{1}{n_2} \sum_{i=p+1}^{p+n_2} \mathbf{J}_{ii} = \mathbb{1}(\Xi) \frac{1}{1 + \frac{1}{n_2} \sum_{u=1}^p \mathbf{J}_{uu}} + \mathbb{1}(\Xi) \delta_1, \quad (\text{S.16})$$

where the residual δ_1 is

$$\mathbb{1}(\Xi) \delta_1 = \mathbb{1}(\Xi) [W]_n + \mathbb{1}(\Xi) \frac{1}{n_2} \sum_{i=p+1}^{p+n_2} O_{<}(|W_i|^2) + \mathbb{1}(\Xi) O_{<}(\Psi_{\mathfrak{B}}^2) = \mathbb{1}(\Xi) [W]_n + O_{<}(\Psi_{\mathfrak{B}}^2),$$

$$[W]_n = \frac{1}{(1 + \frac{1}{n_2} \sum_{u=1}^p \mathbf{J}_{uu})^2} \frac{1}{n_2} \sum_{i=p+1}^{p+n_2} W_i.$$

Here, we are using Lemma S.6.9.

Similarly, from Eq. (S.14),

$$\mathbb{1}(\Xi) \frac{1}{p} \sum_{u=1}^p \mathbf{J}_{uu} = \mathbb{1}(\Xi) \frac{1}{p} \sum_{u=1}^p \frac{1}{\lambda/\sigma_u - z - \frac{1}{n_2} \sum_{i=p+1}^{p+n_2} \mathbf{J}_{ii}} + \delta_2,$$

where

$$\mathbb{1}(\Xi) \delta_2 = \mathbb{1}(\Xi) [W]_p + \mathbb{1}(\Xi) \frac{1}{p} \sum_{u=1}^p O_{<}(|W_u|^2) + \mathbb{1}(\Xi) O_{<}(\Psi_{\mathfrak{B}}^2) = \mathbb{1}(\Xi) [W]_p + O_{<}(\Psi_{\mathfrak{B}}^2),$$

$$[W]_p = \frac{1}{p} \sum_{u=1}^p \frac{-1}{(\lambda/\sigma_u - z - \frac{1}{n_2} \sum_{i=p+1}^{p+n_2} \mathbf{J}_{ii})^2} W_u.$$

All together, we obtain

$$\mathbb{1}(\Xi) \hat{\phi}_p = \mathbb{1}(\Xi) \int \frac{dF^{\Sigma_p}(\tau)}{\lambda/\tau - z - \frac{1}{1 + (p/n_2) \hat{\phi}_p}} + \mathbb{1}(\Xi) \delta_2,$$

which is a perturbed version of Eq. (S.9). Note that Lemma S.6.9 guarantees that $\mathbb{1}(\Xi) \delta_1 < \Psi_{\mathfrak{B}}$ and $\mathbb{1}(\Xi) \delta_2 < \Psi_{\mathfrak{B}}$. This estimate is actually not tight. We shall later improve the estimate to $O_{<}(\Psi_{\mathfrak{B}}^2)$.

On the other hand, it is not hard to identify that the derivation actually only relies on

fact that $\mathbb{1}(\Xi)(|W|_s + \Lambda_o) < \Psi_{\mathfrak{B}}$ from Lemma S.6.9. Since the same control holds when $\eta \geq 1$. We obtain that

$$\mathbb{1}(\eta \geq 1)\hat{\phi}_p = \mathbb{1}(\eta \geq 1) \int \frac{dF^{\Sigma_p}(\tau)}{\lambda/\tau - z - \frac{1}{1 + (p/n_2)\hat{\phi}_p}} + \mathbb{1}(\eta \geq 1)\delta_2,$$

and $\mathbb{1}(\eta \geq 1)\delta_1 < \Psi_{\mathfrak{B}}$, $\mathbb{1}(\Xi)\delta_2 < \Psi_{\mathfrak{B}}$.

We apply Lemma S.5.8, Lemma S.5.9, and Corollary S.5.12 to control the difference between $\hat{\phi}_p$ and ϕ_p when $\eta \geq 1$. In particular, we have the following results.

Lemma S.6.10. *Suppose that the conditions of Theorem S.6.3 hold. Then, we have $\Lambda < n_2^{-1/2}$ uniformly in $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$ satisfying $\eta \geq 1$.*

Proof. Applying Lemma S.6.9, $\Lambda_o < \Psi_{\mathfrak{B}} < n_2^{-1/2}$. Here, we are using the fact that $\Im\phi_p + \mathfrak{B} = O(1)$, when $\eta \geq 1$. We only need to show the diagonal elements are such that

$$\begin{aligned} \mathbf{J}_{uu} - \frac{1}{\lambda/\sigma_u - m_p} &< n_2^{-1/2}, \quad u = 1, 2, \dots, p, \\ -\mathbf{J}_{ii} - \frac{1}{1 + p/n_2\phi_p} &< n_2^{-1/2}, \quad i = p+1, p+2, \dots, p+n_2. \end{aligned}$$

Applying Lemma S.5.9 or Corollary S.5.12, when $\eta \geq 1$,

$$|\hat{\phi}_p - \phi_p| \leq C(|\delta_1| + |\delta_2|) < \Psi_{\mathfrak{B}} < n_2^{-1/2}.$$

Then, by Eq. (S.15)

$$-\frac{1}{\mathbf{J}_{ii}} = 1 + \phi_p + (\hat{\phi}_p - \phi_p) + W_i + O_{<}(\Psi_{\mathfrak{B}}^2) = 1 + \phi_p + O_{<}(n_2^{-1/2}).$$

Therefore, $-\mathbf{J}_{ii} - \frac{1}{1+\phi_p} < n_2^{-1/2}$. The estimate of \mathbf{J}_{uu} is obtained similarly. \square

To show Proposition S.6.8, we apply the stochastic continuity argument, exactly as Section 4 of Bloemendal et al. [2014]. The argument allows us to propagate smallness of Λ

from $\eta \geq 1$ to the whole domain $Q^{(p)} \cup Q_{\text{away}}^{(p)}$. Note that in the domain, $\eta \geq n_2^{-1+a'}$. We choose $\varepsilon < a'/4$ and an arbitrary $D > 0$. Let

$$L(z) = \{z\} \cup \{s \in Q^{(p)} \cup Q_{\text{away}}^{(p)} : \Re s = \Re z, \quad \Im s \in [\Im z, 1] \cap [n_2^{-5}\mathbb{N}]\}.$$

We introduce the random function

$$v(z) := \max_{s \in L(z)} \Lambda(s)(n_2 \Im s)^{1/4}$$

Our goal is to prove that with high probability there is a gap in the range of v , i.e.

$$\mathbb{P} \left(v(z) \leq n_2^\varepsilon, \quad v(z) > n_2^{\varepsilon/2} \right) \leq N^{-D+5} \quad (\text{S.17})$$

for all $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$ and large enough n_2 . This equation says that with high probability the range of v has a gap: it cannot take values in the interval $(n_2^{\varepsilon/2}, n_2^\varepsilon]$.

Since we are dealing with random variables, one has to keep track of probabilities of exceptional events. To that end, we only work on a discrete set of values of η , which allows us to control the exceptional probabilities by a simple union bound.

Next, we prove Eq. (S.17). First, Lemma S.6.9 yields $\mathbb{1}(\Xi)(|\delta_1| + |\delta_2|) < \Psi_{\mathfrak{B}} \leq (n_2 \eta)^{-1/2}$. Since $\{v(z) \leq n_2^\varepsilon\} \subset \Xi(z) \cap \Xi(s)$ for all $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$ and $s \in L(z)$, we find that for all $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$ and $s \in L(z)$ that

$$\mathbb{P} \left(v(z) \leq n_2^\varepsilon, \quad (|\delta_1(s)| + |\delta_2(s)|)(n_2 \Im s)^{1/2} > n_2^{\varepsilon/2} \right) \leq n_2^{-D}$$

for large enough n_2 (independent of z and w). Using a union bound, we therefore get

$$\mathbb{P} \left(v(z) \leq n_2^\varepsilon, \quad \max_{s \in L(z)} (|\delta_1| + |\delta_2|)(n_2 \Im s)^{1/2} > n_2^{\varepsilon/2} \right) \leq n_2^{-D+5}$$

Next, applying Lemma S.5.8 and Lemma S.5.9 with the perturbation $n_2^{\varepsilon/2}(n_2 \Im s)^{-1/2}$, we

get

$$\mathbb{P} \left(v(z) \leq n_2^\varepsilon, \max_{s \in L(z)} \mathfrak{B}_p(s)(n_2 \Im s)^{1/4} > n_2^{\varepsilon/4} \right) \leq n_2^{-D+5}.$$

Moreover, since $-\mathbf{J}_{ii} - (1 + (p/n_2)\phi_p)^{-1} = O_{<}(\mathfrak{B}_p) + O_{<}(\Psi_{\mathfrak{B}})$, we have

$$\mathbb{P} \left(v(z) \leq n_2^\varepsilon, \max_{s \in L(z)} \mathfrak{B}(s)(n_2 \Im s)^{1/4} > 2n_2^{\varepsilon/4} \right) \leq n_2^{-D+5}.$$

Then, Eq. (S.17) follows from the bound

$$\begin{aligned} & \mathbb{P} \left(v(z) \leq n_2^\varepsilon, v(s) \geq n_2^{\varepsilon/2} \right) \\ & \leq \mathbb{P} \left(v(z) \leq n_2^\varepsilon, \max_{s \in L(z)} \Lambda(s)(n_2 \Im s)^{1/4} \geq n_2^{\varepsilon/2} \right) \\ & \leq \mathbb{P} \left(v(z) \leq n_2^\varepsilon, \max_{s \in L(z)} \Psi_{\mathfrak{B}}(n_2 \Im s)^{1/4} \geq n_2^{\varepsilon/2} \right) \\ & \leq \mathbb{P} \left(v(z) \leq n_2^\varepsilon, \max_{s \in L(z)} \mathfrak{B}(n \Im s)^{-1} (n_2 \Im s)^{1/2} \geq (1/2)n_2^\varepsilon \right) + n^{-D+5} \\ & \leq \mathbb{P} \left(v(z) \leq n_2^\varepsilon, \max_{s \in L(z)} \mathfrak{B}(n_2 \Im s)^{1/4} \geq n_2^\varepsilon (n_2 \Im s)^{3/4}/2 \right) + n^{-D+5} \\ & \leq \mathbb{P} \left(v(z) \leq n_2^\varepsilon, \max_{s \in L(z)} \mathfrak{B}(n_2 \Im s)^{1/4} \geq n_2^\varepsilon (n_2)^{3a'/4}/2 \right) + n^{-D+5} \\ & \leq 2n_2^{-D+5}. \end{aligned}$$

Here, we are using the fact that for all $s \in L(z)$

$$\frac{\Im \phi_p(s)}{n_2 \Im s} (n_2 \Im s)^{1/2} = \frac{\Im \phi_p(s)}{\sqrt{n_2 \Im s}} < 1.$$

Therefore,

$$\mathbb{P} \left(v(z) \leq n_2^\varepsilon, \max_{s \in L(z)} \Im \phi_p(s)(n \Im s)^{-1} (n_2 \Im s)^{1/2} \geq (1/2)n_2^\varepsilon \right) \leq n_2^{-D+5}.$$

We conclude the proof of Proposition S.6.8 by combining Eq. (S.17) and Lemma (S.6.10) with a continuity argument. We choose a lattice $\Delta \subset Q^{(p)} \cup Q_{\text{away}}^{(p)}$ such that $|\Delta| \leq n_2^{10}$ and

for each $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$ there exists a $s \in \Delta$ satisfying $|z - s| \leq N^{-4}$. Then, Eq. (S.17) combined with a union bound yields

$$\mathbb{P} \left(\exists s \in \Delta : v(s) \in \left(n_2^{\varepsilon/2}, n_2^\varepsilon \right] \right) \leq n_2^{-D+15}$$

From the definitions of Λ and $Q^{(p)} \cup Q_{\text{away}}^{(p)}$, we find that v is Lipschitz continuous, with Lipschitz constant n_2^2 .

$$\mathbb{P} \left(\exists z \in Q^{(p)} \cup Q_{\text{away}}^{(p)} : v(z) \in \left(2n_2^{\varepsilon/2}, n_2^\varepsilon \right] \right) \leq n_2^{-D+15}$$

By Lemma S.6.10, we have

$$\mathbb{P} \left(v(z) > n_2^{\varepsilon/2} \right) \leq n_2^{-D+15}$$

for any $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$ with $\eta \geq 1$. Therefore,

$$\mathbb{P} \left(\max_{z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}} v(z) > 2n_2^{\varepsilon/2} \right) \leq 2n_2^{-D+15}.$$

It completes the proof of Proposition S.6.8.

S.6.1.3 The averaged local law

We use Proposition S.6.8 and improved estimates for the averaged quantities $[W]_p$ and $[W]_n$ to deduce the averaged local laws in $Q^{(p)}$, $Q_-^{(p)}$, and $Q_{\text{away}}^{(p)}$, as in Theorem S.6.2. Indeed, we shall show that

$$\mathbb{1}(\Xi)[W]_p < \Psi_{\mathfrak{B}}^2 \quad \text{and} \quad \mathbb{1}(\Xi)[W]_n < \Psi_{\mathfrak{B}}^2.$$

If so, the estimate of δ_1 and δ_2 is improved to $\mathbb{1}(\Xi)|\delta_1| < \Psi_{\mathfrak{B}}^2$ and $\mathbb{1}(\Xi)|\delta_2| < \Psi_{\mathfrak{B}}^2$.

Lemma S.6.11 (Fluctuation averaging). *Suppose that the assumptions of Theorem S.6.2 hold. Suppose that Υ is a positive, n_2 -dependent, deterministic function on $Q^{(p)} \cup Q_{\text{away}}^{(p)}$*

satisfying $n_2^{-1/2} \leq \Upsilon \leq n_2^{-c}$ for some constant $c > 0$. Suppose moreover that $\Lambda < n_2^{-c}$ and $\Lambda_o < \Upsilon$ on $Q^{(p)} \cup Q_{\text{away}}^{(p)}$. Then on $Q^{(p)} \cup Q_{\text{away}}^{(p)}$ we have

$$\frac{1}{p} \sum_{u=1}^p \frac{1}{(\lambda/\sigma_u - z - \frac{1}{n_2} \sum_{i=p+1}^{p+n_2} \mathbf{J}_{ii})^2} (1 - \mathbb{E}_u) \frac{1}{\mathbf{J}_{uu}} = O_{<}(\Upsilon^2)$$

and

$$\frac{1}{n_2} \sum_{i=p+1}^{p+n_2} (1 - \mathbb{E}_i) \frac{1}{\mathbf{J}_{ii}} = O_{<}(\Upsilon^2).$$

Th results are an extension of Lemma 5.6 of Knowles and Yin [2017], Lemma 4.9 of Bloemendal et al. [2014], Theorem 4.7 of Erdős et al. [2013]. They can be proved using exactly the same method. We therefore omit the details. The readers may refer to the proof of Theorem 4.7 of Erdős et al. [2013] for details.

Note that Proposition S.6.8 implies that $1 - \mathbb{1}(\Xi) < 0$, that is Ξ holds with high probability. In the following analysis, we shall always assume Ξ holds. Combining with Lemma S.6.9, we have

$$\Lambda < \frac{1}{(n_2 \eta)^{1/4}}, \quad \Lambda_o < \Psi_{\mathfrak{B}},$$

on $Q^{(p)} \cup Q_{\text{away}}^{(p)}$. Therefore, the condition of Lemma S.6.11 is satisfied and we obtain

$$|[W]_p| + |[W]_n| < \Psi_{\mathfrak{B}}^2,$$

on $Q^{(p)} \cup Q_{\text{away}}^{(p)}$. Recall that $\delta_1 = [W]_n + O_{<}(\Psi_{\mathfrak{B}}^2)$ and $\delta_2 = [W]_p + O_{<}(\Psi_{\mathfrak{B}}^2)$. We immediately get

$$(|\delta_1| + |\delta_2|) < \Psi_{\mathfrak{B}}^2,$$

uniformly in $Q^{(p)} \cup Q_{\text{away}}^{(p)}$.

The following lemma holds immediately from the definition of $\Psi_{\mathfrak{B}}$.

Lemma S.6.12. *Let $c > 0$ and suppose that we have $\mathfrak{B} < (n_2 \eta)^{-c}$ uniformly in $z \in$*

$Q^{(p)} \cup Q_{\text{away}}^{(p)}$. Then, we have

$$|\delta_1| + |\delta_2| < \frac{\Im \phi_p + (n_2 \eta)^{-c}}{n_2 \eta}$$

uniformly in $z \in Q^{(p)} \cup Q_{\text{away}}^{(p)}$.

We first consider the averaged local law in $Q^{(p)}$. Suppose $\mathfrak{B} < (n_2 \eta)^{-c}$. Then, uniformly in $z \in Q^{(p)}$,

$$|\delta_1| + |\delta_2| < \frac{\Im \phi_p + (n_2 \eta)^{-c}}{n_2 \eta}.$$

We invoke Lemma S.5.8 to get

$$\begin{aligned} \mathfrak{B}_p = |\hat{\phi}_p - \phi_p| &< \frac{\frac{\Im \phi_p + (n_2 \eta)^{-c}}{n_2 \eta}}{\sqrt{\kappa + \eta} + \sqrt{\frac{\Im \phi_p + (n_2 \eta)^{-c}}{n_2 \eta}}} \leq \frac{\Im \phi_p}{n_2 \eta} \frac{1}{\sqrt{\kappa + \eta}} + \sqrt{\frac{(n_2 \eta)^{-c}}{n_2 \eta}} \\ &\leq \frac{C}{n_2 \eta} + (n_2 \eta)^{-1/2-c/2} < (n_2 \eta)^{-1/2-c/2}. \end{aligned} \quad (\text{S.18})$$

In the fourth step, we are using the expression of $\Im \phi_p$ as shown in Lemma S.5.6. As for \mathfrak{B}_n , plugging the estimate of $\hat{\phi}_p$ back to Eq. (S.16), we also have

$$\mathfrak{B}_n < \mathfrak{B}_p + \Psi_{\mathfrak{B}}^2 < (n_2 \eta)^{-1/2-c/2} + \frac{\Im \phi_p + (n_2 \eta)^{-c}}{n_2 \eta} < (n_2 \eta)^{-1/2-c/2}.$$

All together, we have shown that

$$\mathfrak{B} < (n_2 \eta)^{-c} \implies \mathfrak{B} < (n_2 \eta)^{-1/2-c/2}. \quad (\text{S.19})$$

From Proposition S.6.8, we know that $\mathfrak{B} \leq C\Lambda < (n_2 \eta)^{-1/4}$. Therefore, for any $\varepsilon > 0$, we iterate Eq. (S.19) finite number of times to get,

$$\mathfrak{B} < (n_2 \eta)^{-1+\varepsilon}.$$

It completes the proof of the averaged local law in $Q^{(p)}$ in Theorem S.6.2 under the additional assumption that Σ_p is diagonal.

Next, we consider the averaged local law in $Q_-^{(p)}$. The follows exactly from the same arguments but with the specific form of $\Im\phi_p$ when $E \leq \rho - n^{-2/3+a'}$. When E is restricted to $E \leq \rho - n_2^{-2/3+a'}$ and $\eta \geq n_2^{-2/3}$, we have

$$\Im\phi_p \asymp \frac{\eta}{\sqrt{\kappa + \eta}},$$

where $\kappa = \rho - E$. Assume that $\mathfrak{B} < (n_2\eta)^{-c}$. When $z \in Q_-^{(p)}$, Eq. (S.18) can be improved to

$$\mathfrak{B}_p < \frac{\Im\phi_p}{n_2\eta\sqrt{\kappa + \eta}} + \frac{1}{(n_2\eta)^c} \frac{1}{n_2\eta\sqrt{\kappa + \eta}} = \frac{1}{n_2(\kappa + \eta)} + \frac{1}{(n_2\eta)^c} \frac{1}{n_2\eta\sqrt{\kappa + \eta}} < \frac{1}{n_2(\kappa + \eta)} + \frac{1}{(n_2\eta)^{c+a'/2}}.$$

plugging the estimate of $\hat{\phi}_p$ back to Eq. (S.16), we also have

$$\mathfrak{B}_n < \mathfrak{B}_p + \frac{\Im\phi_p + (n_2\eta)^{-c}}{n_2\eta} < \frac{1}{n_2(\kappa + \eta)} + \frac{1}{(n_2\eta)^{c+a'/2}}.$$

All together, we get

$$\mathfrak{B} < \frac{1}{(n_2\eta)^c} \implies \mathfrak{B} < \frac{1}{n_2(\kappa + \eta)} + \frac{1}{(n_2\eta)^{c+a'/2}}.$$

Again, from Propostion S.6.8, $\mathfrak{B} < (n_2\eta)^{-1/4}$. We iterate the equation finite number of times to get,

$$\mathfrak{B} < \frac{1}{n_2(\kappa + \eta)},$$

uniformly on $Q_-^{(p)}$. It completes the proof of the averaged local law in $Q_-^{(p)}$ in Theorem S.6.2 under the additional assumption that Σ_p is diagonal.

Lastly, we consider the averaged local law in $Q_{\text{away}}^{(p)}$. It follows from similar arguments as the previous two cases. On $Q_{\text{away}}^{(p)}$, we invoke Lemma S.5.9 to control $|\hat{\phi}_p - \phi_p|$. Assuming

$\mathfrak{B} < (n_2\eta)^{-c}$, then

$$\mathfrak{B}_p = |\hat{\phi}_p - \phi_p| \leq C|\delta_1| + C|\delta_2| < \frac{\Im\phi_p + (n_2\eta)^{-c}}{n_2\eta} < \frac{1}{n_2} + (n_2\eta)^{-c-1} < \frac{1}{n_2} + (n_2\eta)^{-c-1}.$$

Here, we are using the fact that $\Im\phi_p/\eta \asymp 1$ when z is away from the support of \mathcal{G}_λ . The rest arguments are very similar to those in the previous two cases. We omit further details.

We completed the proof of the averaged local laws when Σ_p is diagonal. The diagonal assumption will be removed in Section S.6.1.5.

S.6.1.4 The entrywise local law

It remains to prove the entrywise local law. Using Eq. (S.11), Result (v) of Lemma S.6.5, Lemma S.6.9 and the averaged local law, we obtain

$$\mathbf{J}_{uu} - \frac{1}{\lambda/\sigma_u - z + \phi_p} < \Psi_{\mathfrak{B}} < \Psi,$$

uniformly for $u = 1, \dots, p$ and $z \in Q^{(p)}$. Similarly, using Eq. (S.12),

$$\mathbf{J}_{ii} + \frac{1}{1 + (p/n_2)\phi_p} < \Psi,$$

uniformly for $i = p+1, \dots, p+n_2$ and $z \in Q^{(p)}$. Moreover, since $\Lambda_o < \Psi_{\mathfrak{B}} < \Psi$, the entrywise local law follows.

S.6.1.5 Extension to non-diagonal Σ_p

In this section, we show that Theorem S.6.2 and Theorem S.6.3 when the population covariance matrix is non-diagonal. In order to utilize the notation in the previous sections, we denote the population covariance as $\tilde{\Sigma}_p$ and continue to use Σ_p to denote the diagonal matrix containing the eigenvalues of $\tilde{\Sigma}_p$. The eigendecomposition of $\tilde{\Sigma}_p$ is denoted as

$$\tilde{\Sigma}_p = H\Sigma_p H^T,$$

where H is the eigenvector matrix of $\tilde{\Sigma}_p$. In this section, we show that if Theorem S.6.2 holds with Σ_p , it also holds with $\tilde{\Sigma}_p$. All definitions in Section S.6.1.1 are kept.

Define

$$\begin{aligned}\tilde{\mathbf{K}} &= \begin{pmatrix} H & 0 \\ 0 & I_{n_2} \end{pmatrix} \mathbf{K} \begin{pmatrix} H^T & 0 \\ 0 & I_{n_2} \end{pmatrix}, \\ \tilde{\mathbf{J}} &= \begin{pmatrix} H & 0 \\ 0 & I_{n_2} \end{pmatrix} \mathbf{J} \begin{pmatrix} H^T & 0 \\ 0 & I_{n_2} \end{pmatrix}, \\ \tilde{\Omega} &= \begin{pmatrix} H & 0 \\ 0 & I_{n_2} \end{pmatrix} \Omega \begin{pmatrix} H^T & 0 \\ 0 & I_{n_2} \end{pmatrix}.\end{aligned}$$

Then, $\tilde{\mathbf{K}}$, $\tilde{\mathbf{K}}$ and $\tilde{\Omega}$ are equal to \mathbf{K} , \mathbf{J} and Ω when Σ_p is replaced by $\tilde{\Sigma}_p$.

The averaged local laws in $Q^{(p)}$, $Q_-^{(p)}$ and $Q_{\text{away}}^{(p)}$ still hold when \mathbf{J} and Ω are replaced by $\tilde{\mathbf{J}}$ and $\tilde{\Omega}$, since

$$\frac{1}{p} \sum_{u=1}^p \tilde{\mathbf{J}}_{uu} = \frac{1}{p} \text{tr}(\tilde{\mathbf{J}}_{(11)}) = \frac{1}{p} \text{tr}(\mathbf{J}_{(11)} H^T H) = \hat{\phi}_p.$$

It remains to show the entrywise local law. We shall actually show

$$H(\mathbf{J}_{(11)} - \Omega_{(11)})H^T = O_{<}(\Psi),$$

uniformly in $Q^{(p)}$. In components, this reads

$$\left| \sum_{u,v=1}^p H_{iu} \left(\mathbf{J}_{uv} - \mathbb{1}(u=v) \frac{1}{\lambda/\sigma_u - m_p} \right) H_{vj} \right| < \Psi,$$

for all $i, j \in \{1, 2, \dots, p\}$.

The proof is based on the polynomialization method developed in Section 5 of Bloemendal et al. [2014]. The argument is very similar to that of Bloemendal et al. [2014], and we only outline the differences.

By the assumption

$$\mathbf{J}_{(11)} - (\lambda \Sigma_p^{-1} - m_p I_p)^{-1} = O_{<}(\Psi)$$

and orthogonality of H , we have

$$\begin{aligned} & \sum_{u,v=1}^p H_{iu} \left(\mathbf{J}_{uv} - \mathbb{1}(u=v) \frac{1}{\lambda/\sigma_u - m_p} \right) H_{vj} \\ &= \sum_{u=1}^p H_{iu} \left(\mathbf{J}_{uu} - \frac{1}{\lambda/\sigma_u - m_p} \right) H_{uj} + \sum_{u \neq v} H_{iu} \mathbf{J}_{uv} H_{vj} = O_{<}(\Psi) + \mathcal{Z} \end{aligned}$$

where we defined $\mathcal{Z} := \sum_{u \neq v} H_{iu} \mathbf{J}_{uv} H_{vj}$. We need to prove that $|\mathcal{Z}| < \Psi$, which, following Section 5 of Bloemendal et al. [2014] we do by estimating the moment $\mathbb{E}|\mathcal{Z}|^k$ for fixed $k \in 2\mathbb{N}$. The argument from Section 5 of Bloemendal et al. [2014] may be taken over with minor changes.

Use Result (ii) and (v) of Lemma S.6.5,

$$\sum_{u \neq v \notin S} H_{iu} \mathbf{J}_{uv}^{(S)} H_{vj} = \sum_{u \neq v \notin S} H_{iu} H_{vj} \mathbf{J}_{uu}^{(S)} \mathbf{J}_{vv}^{(S \cup \{u\})} \left(\tilde{\mathbf{Z}}_u \mathbf{J}_{(22)}^{(S \cup \{uv\})} \tilde{\mathbf{Z}}_v^T \right)$$

for any $S \subset \{1, 2, \dots, p\}$ and

$$\begin{aligned} \mathbf{J}_{uu}^{(S)} &= \frac{1}{\lambda/\sigma_u - z - \tilde{\mathbf{Z}}_u \mathbf{J}_{(22)}^{(S \cup \{u\})} \tilde{\mathbf{Z}}_u^T} \\ &= \sum_{\ell=0}^{L-1} \frac{1}{(\lambda/\sigma_u - z + [1 + (p/n_2)\phi_p]^{-1})^{\ell+1}} \left(\left(\tilde{\mathbf{Z}}_u \mathbf{J}_{(22)}^{(S \cup \{u\})} \tilde{\mathbf{Z}}_u^T \right) - \frac{1}{1 + (p/n_2)\phi_p} \right)^\ell + O_{<}(\Psi^L). \end{aligned}$$

We omit further details.

S.6.2 Proof of Theorem S.6.3

The rigidity of the smallest eigenvalue of \mathbf{G}_λ can be proved using the averaged local law in $Q_-^{(p)}$ as in Theorem S.6.2. In the following, $\epsilon \in (0, 2/3)$ is fixed. Notice that $|\rho_p - \rho| = o(p^{-2/3})$. We only need to show that with high probability there is no eigenvalue in $(-\infty, \rho_p - n_2^{-2/3+\epsilon})$. Indeed, it suffices to consider the interval $[\lambda/\ell_{\max}(\Sigma_p), \rho_p - n_2^{-2/3+\epsilon}]$, since $\mathbf{G}_\lambda = \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T + \lambda\Sigma_p^{-1}$.

Recall the definition of $Q^{(p)}$ and $Q_-^{(p)}$. We consider a and a' to be sufficiently small so

that $\{z : E \in \lambda/\ell_{\max}(\Sigma_p), \rho_p - n_2^{-2/3+\epsilon}], n_2^{-2/3} \leq \eta \leq 1\}$ is contained in $Q_-^{(p)}$. Consider the set $Q^* = Q_-^{(p)} \cap \{z \mid \eta = n_2^{-1/2-\epsilon/4} |E - \rho|^{1/4}\}$.

To show that there is no eigenvalue in $[\lambda/\ell_{\max}(\Sigma_p), \rho - n_2^{-2/3+\epsilon})$, we only need to show that uniformly in Q^*

$$\Im \hat{\phi}_p(z) < n_2^{-\epsilon/2} (n_2 \eta)^{-1}.$$

It is because that $\hat{\phi}_p$ is the Stieltjes transform of \mathbf{G}_λ . When $z = \ell(\mathbf{G}_\lambda) + i\eta$ where $\ell(\mathbf{G}_\lambda)$ is an eigenvalue of \mathbf{G}_λ , $\Im \hat{\phi}_p(z) \asymp \frac{1}{n_2 \eta}$.

Recall that in Lemma S.5.6 for $z \in Q^*$,

$$\Im \phi_p \asymp \eta(|E - \rho| + \eta)^{-1/2}.$$

We conclude that it suffices to prove

$$|\hat{\phi}_p - \phi_p| < \frac{1}{n_2(|E - \rho| + \eta)},$$

which is ensured by the averaged local law in $Q_-^{(p)}$ in Theorem S.6.2.

S.6.3 Proof of Theorem S.6.4

Theorem S.6.3 indicates that the smallest eigenvalue of \mathbf{G}_λ is larger than $\rho - \epsilon$ with high probability. Lemma S.6.7 indicates that the largest eigenvalue of \mathbf{G}_λ is smaller than $x_{\text{right}} = (1 + \sqrt{\gamma_2})^2 + \lambda/\liminf \ell_{\min}(\Sigma_p) + \epsilon$. Therefore, the statement in (i) of Theorem S.6.4 holds.

The convergence of $p^{-1} \sum_{j=1}^p f(\ell_j(\mathbf{G}_\lambda))$ can be proved using the strategy of Bai and Silverstein [2004]. We consider a contour in the complex plane that encloses the interval $[\rho - \epsilon, (1 + \sqrt{\gamma_2})^2 + \lambda/\liminf \ell_{\min}(\Sigma_p) + \epsilon]$. Without loss of generality, we choose the rectangle, denoted to be \mathcal{R} , with the four vertices

$$z_1 = \rho - 2\epsilon + \epsilon'i, \quad z_2 = \rho - 2\epsilon - \epsilon'i, \quad z_3 = x_{\text{right}} + \epsilon + \epsilon'i, \quad z_4 = x_{\text{right}} + \epsilon - \epsilon'i.$$

In the following, we choose ϵ and ϵ' to be sufficiently small so that the function f is analytic on the rectangle.

Clearly, with high probability the rectangle encloses all eigenvalues of \mathbf{G}_λ . When it is true,

$$\frac{1}{p} \sum_{j=1}^p f(\ell_j(\mathbf{G}_\lambda)) = \int f(\tau) dF^{\mathbf{G}_\lambda}(\tau) = \frac{1}{2\pi i} \oint_{\mathcal{R}} f(z) \hat{\phi}_p(z) dz.$$

Here, the contour integral is taken over the rectangle in the positive direction in the complex plane. In the integral, we extend the definition of $\hat{\phi}_p$ to \mathbb{C}^- by setting $\hat{\phi}_p(z) = \overline{\hat{\phi}_p(\bar{z})}$ if $\eta(z) < 0$.

Fix any $c \in (0, 1)$. We truncate \mathcal{R} at $\eta = n_2^{-1+c/2}$. Specifically, we define $\hat{\mathcal{R}} = \mathcal{R} \cap \{z : \eta \geq n_2^{-1+c/2}\}$. Then,

$$n_2^{1-c} \mathbb{1}(\mathfrak{M}) \left| \oint_{\mathcal{R} \setminus \hat{\mathcal{R}}} f(z) \hat{\phi}_p(z) dz \right| \leq C n_2^{1-c} n_2^{-1+c/2} \rightarrow 0.$$

It follows

$$\left| \oint_{\mathcal{R} \setminus \hat{\mathcal{R}}} f(z) \hat{\phi}_p(z) dz \right| < \frac{1}{n_2}.$$

Therefore, to show

$$\left| \frac{1}{p} \sum_{j=1}^p f(\ell_j(\mathbf{G}_\lambda)) - \int f(\tau) d\mathcal{G}_\lambda(\tau) \right| < \frac{1}{n_2},$$

it suffices to show

$$\oint_{\hat{\mathcal{R}}} |f(z)| \left| \hat{\phi}_p(z) - \phi_p(z) \right| dz < \frac{1}{n_2}.$$

It is a direct consequence from the averaged local law on $Q_{\text{away}}^{(p)}$ in Theorem S.6.2. It completes the proof.

S.6.4 Discrete case

In Theorem S.6.2, Theorem S.6.3, Theorem S.6.4, we focus on the case when $\rho > \lambda/\sigma_{\max}$ as in **SC2**. In this section, we consider the case when $\rho = \lambda/\sigma_{\max}$. We show analogous results as Theorem S.6.2, Theorem S.6.3 and Theorem S.6.4.

Theorem S.6.13. *Suppose that **C1–C6** and **SC1** hold. Suppose $\rho = \lambda/\sigma_{\max}$. Then, with probability 1, the smallest eigenvalue of \mathbf{G}_λ is $\lambda/\ell_{\max}(\Sigma_p)$.*

Theorem S.6.14. *Suppose the conditions in Theorem S.6.13 hold. Then, uniformly in $Q_{\text{away}}^{(p)}$*

$$\hat{\phi}_p(z) - \phi_p(z) = O_{<}(\frac{1}{n_2}).$$

Theorem S.6.15. *Suppose the conditions in Theorem S.6.13 hold. The results in Theorem S.6.4 still hold.*

Proof of Theorem S.6.13. Under Definition 2.7, for all sufficiently large p , we have

$$pF^{\Sigma_p}(\{\ell_{\max}(\Sigma_p)\}) > n_2.$$

Note that $pF^{\Sigma_p}(\{\ell_{\max}(\Sigma_p)\})$ is the rank of the eigen-subspace associated with the smallest eigenvalue of $\lambda\Sigma_p^{-1}$, that is $\ell_{\max}(\Sigma_p)$. The rank of $\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$ is n_2 with probability 1, since the condition implies that $p > n_2$. Therefore, using Wey's inequality, we obtain that there are at least 1 eigenvalue of $\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T$ less or equal to $\lambda/\ell_{\max}(\Sigma_p)$, which is the smallest one since $\ell_{\min}(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T + \lambda\Sigma_p^{-1}) \geq \lambda/\ell_{\max}(\Sigma_p)$. Therefore, $\ell_{\min} = \lambda/\ell_{\max}(\Sigma_p)$. \square

Proof of Theorem S.6.14. The proof of Theorem S.6.14 follows from the same arguments as those in the proof of Theorem S.6.2. We only need to replace the bound in Lemma S.5.8 by that in Lemma S.5.9. \square

Proof. The proof of Theorem S.6.15 follows from the same arguments as those in the proof of Theorem S.6.4. In the last step, the bound on $|\hat{\phi}_p - \phi_p|$ is given by Theorem S.6.14. \square

S.7 Proof of Theorem 2.9

We start the proof by formulating the resolvent $(U_1^T \mathbf{Z}^T \mathbf{G}_\lambda^{-1} \mathbf{Z} U_1 - z I_{n_1})^{-1}$ of $\tilde{\mathbf{F}}_\lambda$, for $z \in \mathbb{C}^+$. Following the strategy in Knowles and Yin [2017] and Han et al. [2018], we construct a linearizing block matrix as

$$\mathbf{H}(z) = \mathbf{H}(z, \mathbf{Z}) = \begin{pmatrix} -z I_{n_1} & U_1^T \mathbf{Z}^T & 0 \\ \mathbf{Z} U_1 & -\lambda \Sigma_p^{-1} \mathbf{Z} U_2 \\ 0 & U_2^T \mathbf{Z}^T & I_{n_2} \end{pmatrix}.$$

By the Schur complement formula, it turns out that the upper-left block of the 3×3 block matrix $\mathbf{H}^{-1}(z)$, called $\mathbf{L}(z)$, is the resolvent of $\tilde{\mathbf{F}}_\lambda$, that is

$$\mathbf{L}(z) = \Delta^T \mathbf{H}^{-1}(z) \Delta = (U_1^T \mathbf{Z}^T \mathbf{G}_\lambda^{-1} \mathbf{Z} U_1 - z I_{n_1})^{-1}, \quad \text{where } \Delta = \begin{bmatrix} I_{n_1} \\ 0_{(p+n_2) \times n_1} \end{bmatrix}.$$

It then suffices to focus on $\mathbf{H}^{-1}(z)$ instead of $\tilde{\mathbf{F}}_\lambda$ directly. The reason is that $\mathbf{H}(z)$ is linear in \mathbf{Z} .

Recall that Θ_1 is the right most edge of the support of F_q as in Lemma S.5.4. We introduce the fundamental control parameter

$$\Phi(z) = \sqrt{\frac{\Im q(z)}{n_1 \eta}} + \frac{1}{n_1 \eta}.$$

Recall the definition of the domain D in Eq. (S.4) that is around Θ_1 as

$$D = D(a, a', n_1) := \{z = E + i\eta \in \mathbb{C}^+ : |E - \Theta_1| \leq a, n_1^{-1+a'} \leq \eta \leq 1/a'\}.$$

The core of the proof is to build the following strong local law of the resolvent $\mathbf{L}(z)$ in D .

Theorem S.7.1. *Suppose that C1-C6 hold, Then, there exists $a > 0$ such that the following results hold for arbitrary $a' > 0$.*

(i) The entrywise local law holds as

$$\mathbf{L}(z) - q(z)I_{n_1} = O_{<}(\Phi(z)),$$

uniformly $z \in D$.

(ii) The averaged local law holds as

$$|\overline{L}(z) - q(z)| < \frac{1}{n_1\eta},$$

uniformly $z \in D$. Here,

$$\overline{L}(z) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{L}_{ii}(z)$$

and $\mathbf{L}_{ij}(z)$ is the (i, j) -th element of $\mathbf{L}(z)$.

Once Theorem S.6.2 is ready, we can show the edge universality of $\tilde{\mathbf{F}}_\lambda$, which means that the limiting distribution of

$$\frac{p^{2/3}}{\Theta_2}(\ell_{\max}(\tilde{\mathbf{F}}_\lambda) - \Theta_1)$$

is not affected by the distribution of \mathbf{Z} as long as Condition **C2** is satisfied. Similar to Theorem 6.3 of Erdős et al. [2012], to show the edge universality, it suffices to show the following Green function comparison theorem.

Theorem S.7.2. *Let $\epsilon > 0$, $\eta = n^{-2/3-\epsilon}$, $E_1, E_2 \in \mathbb{R}$ satisfy $E_1 < E_2$ and*

$$|E_1 - \Theta_1|, |E_2 - \Theta_1| \leq n^{-2/3+\epsilon}$$

Set $K : \mathbb{R} \rightarrow \mathbb{R}$ to be a smooth function such that

$$\max_x |K^{(l)}(x)| \leq C, \quad l = 1, 2, 3, 4, 5$$

for some constant C . Then there exists a constant $c > 0$ such that for large enough n and small enough ϵ , we have

$$\left| \mathbb{E}K \left(n \int_{E_1}^{E_2} \Im \bar{L}_{\mathbf{Z}}(x + i\eta) dx \right) - \mathbb{E}K \left(n \int_{E_1}^{E_2} \Im \bar{L}_{\mathbf{Z}^0}(x + i\eta) dx \right) \right| \leq n^{-c},$$

where $\bar{L}_{\mathbf{Z}}(z)$ is $(1/n_1) \sum_{i=1}^{n_1} \mathbf{L}_{ii}(z)$ when \mathbf{Z} satisfies the moment condition **C2** and $\bar{L}_{\mathbf{Z}^0}(z)$ is similar to $\bar{L}_{\mathbf{Z}}(z)$ but with \mathbf{Z} replaced by a Gaussian matrix \mathbf{Z}^0 .

It is a standard procedure in RMT to connect Theorem S.7.2 to the edge universality of $\tilde{\mathbf{F}}_{\lambda}$. In particular, we can fix $E^* < p^{-2/3}$ such that $\ell_{\max}(\tilde{\mathbf{F}}_{\lambda}) \leq \Theta_1 + E^*$ with high probability. Choosing $|E - \Theta_1| < p^{-2/3}$, $\eta = n_1^{-2/3-9\epsilon}$ and $l = \frac{1}{2}n_1^{-2/3-\epsilon}$, then for some sufficiently small constant ϵ and sufficiently large constant M , there exists a constant $n_0(\epsilon, M)$ such that

$$\mathbb{E}K \left(\frac{n_1}{\pi} \int_{E-l}^{\Theta_1+E^*} \Im \bar{L}_{\mathbf{Z}}(x + i\eta) dx \right) \leq \mathbb{P} \left(\ell_{\max}(\tilde{\mathbf{F}}_{\lambda}) \leq E \right) \leq \mathbb{E}K \left(\frac{n_1}{\pi} \int_{E+l}^{\Theta_1+E^*} \Im \bar{L}_{\mathbf{Z}}(x + i\eta) dx \right) + n_1^{-M}, \quad (\text{S.20})$$

$$\mathbb{E}K \left(\frac{n_1}{\pi} \int_{E-l}^{\Theta_1+E^*} \Im \bar{L}_{\mathbf{Z}^0}(x + i\eta) dx \right) \leq \mathbb{P} \left(\ell_{\max}(\tilde{\mathbf{F}}_{\lambda}^0) \leq E \right) \leq \mathbb{E}K \left(\frac{n_1}{\pi} \int_{E+l}^{\Theta_1+E^*} \Im \bar{L}_{\mathbf{Z}^0}(x + i\eta) dx \right) + n_1^{-M}, \quad (\text{S.21})$$

where $\tilde{\mathbf{F}}_{\lambda}^0$ is as $\tilde{\mathbf{F}}_{\lambda}$ but when \mathbf{Z} is additionally Gaussian, $n_1 \geq n_0(\epsilon, M)$, and K is a smooth cutoff function satisfying the condition in Theorem S.7.2. We omit the proof of (S.20) and (S.21) because it is a standard procedure in RMT. Similar works include Corollary 5.1 of Lemma 6.1 of Erdős et al. [2012], Bao et al. [2013], and Lemma 4.1 of Pillai and Yin [2014].

Following (S.20), (S.21) and Theorem S.7.2, we can show

$$\mathbb{P}(\ell_{\max}(\tilde{\mathbf{F}}_{\lambda}) \leq E) \geq \mathbb{P}(\ell_{\max}(\tilde{\mathbf{F}}_{\lambda}^0) \leq E - 2l) - n_1^{-M},$$

$$\mathbb{P}(\ell_{\max}(\tilde{\mathbf{F}}_{\lambda}) \leq E) \leq \mathbb{P}(\ell_{\max}(\tilde{\mathbf{F}}_{\lambda}^0) \leq E + 2l) + n_1^{-M}.$$

It then follows that

$$\mathbb{P}\left(\frac{p^{2/3}}{\Theta_2}(\ell_{\max}(\tilde{\mathbf{F}}_\lambda) - \Theta_1) \leq s\right) - \mathbb{P}\left(\frac{p^{2/3}}{\Theta_2}(\ell_{\max}(\tilde{\mathbf{F}}_\lambda^0) - \Theta_1) \leq s\right) \longrightarrow 0. \quad (\text{S.22})$$

The remaining tasks are as follows.

Step 1: Prove Theorem S.7.1 under Gaussianity.

Step 2: Prove the asymptotic Tracy-Widom distribution of the normalized largest root as in Theorem 2.9 under Gaussianity. That is, to prove

$$\mathbb{P}\left(\frac{p^{2/3}}{\Theta_2}(\ell_{\max}(\tilde{\mathbf{F}}_\lambda^0) - \Theta_1) \leq s\right) \rightarrow \text{TW}_1(s),$$

for any $s \in \mathbb{R}$.

Step 3: Prove Theorem S.7.1 under the general condition **C2**.

Step 4: Establish edge universality by proving Theorem S.7.2 under the general condition **C2**. It then follows from **Step 2** and Eq. (S.22) that the asymptotic Tracy-Widom distribution of the normalized largest root holds under the general condition.

The technical arguments in the remaining proof are routine in the literature on “local laws” of random matrices. Similar works include Bloemendal et al. [2014], Knowles and Yin [2017], Han et al. [2016], and Han et al. [2018]. Indeed, most arguments in the proof are variants of those in these prior works with limited changes. Accordingly, in the following sections, we provide an outline of the proof and refer to these prior works for detailed arguments.

Step 1: Proof of Theorem S.7.1 under Gaussianity

We first show that Theorem S.7.1 holds under the additional assumption that \mathbf{Z} is Gaussian. Note that $\mathbf{Z}U_1$ and $\mathbf{Z}U_2$ are independent under Gaussianity. We can equivalently consider the matrix

$$\tilde{\mathbf{Z}}_1^T \mathbf{G}_\lambda^{-1} \tilde{\mathbf{Z}}_1, \quad \text{and} \quad \mathbf{G}_\lambda = \tilde{\mathbf{Z}}_2 \tilde{\mathbf{Z}}_2^T + \lambda \Sigma_p^{-1},$$

where $\tilde{\mathbf{Z}}_1$ is a $p \times n_1$ matrix whose entries are iid $N(0, 1/n_1)$, $\tilde{\mathbf{Z}}_2$ is a $p \times n_2$ matrix whose entries are iid $N(0, 1/n_2)$, and $\tilde{\mathbf{Z}}_1$ and $\tilde{\mathbf{Z}}_2$ are independent.

The main strategy in the proof is as follows. Fixing \mathbf{G}_λ , Theorem 3.14 of Knowles and Yin [2017] can be applied to $\tilde{\mathbf{Z}}^T \mathbf{G}_\lambda^{-1} \tilde{\mathbf{Z}}$ with \mathbf{G}_λ^{-1} be regarded as “ Σ ” in their work. The local laws of $\tilde{\mathbf{F}}_\lambda$ is then obtained as in Theorem S.7.1 but with $q(z)$ replaced by a finite-sample version. In particular, we fix \mathbf{G}_λ and define a finite-sample version of Eq. (S.3) as

$$z = \hat{f}(q), \quad \hat{f}(q) = -\frac{1}{q} + \frac{1}{n_1} \sum_{j=1}^p \frac{1}{g_j + q},$$

where $g_1 \geq g_2 \geq \dots \geq g_p \geq \lambda/\sigma_{\max} > 0$ are eigenvalues of \mathbf{G}_λ . Again, applying Lemma 2.2, for any $z \in \mathbb{C}^+$, there exists a unique solution $\hat{q}_p(z) \in \mathbb{C}^+$ to the above equation. Secondly, we apply the results in Section S.6 to show that \hat{q}_p converges uniformly to $q(z)$ in the domain D .

Let $\hat{\beta}_p \in (0, g_p)$ be such that $\hat{f}'(-\hat{\beta}_p) = 0$. That is,

$$\hat{f}'(-\hat{\beta}_p) = \frac{1}{\hat{\beta}_p^2} - \frac{1}{n_1} \sum_{j=1}^p \frac{1}{(g_j - \hat{\beta}_p)^2} = 0.$$

Clearly, for any fixed discrete points g_j 's such that $g_p > 0$, $\hat{\beta}_p$ exists and is unique. Let $\hat{\Theta}_{p1} = \hat{f}(-\hat{\beta}_p)$ and

$$\hat{\Phi}(z) = \sqrt{\frac{\Im \hat{q}_p(z)}{n_1 \eta}} + \frac{1}{n_1 \eta}.$$

Define the following domain

$$\hat{D} = \hat{D}(a, a', n_1) := \{z = E + i\eta \in \mathbb{C}^+ : |E - \hat{\Theta}_{p1}| \leq a, \quad n_1^{-1+a'} \leq \eta \leq 1/a'\}.$$

The following results hold by applying Theorem 3.14 of Knowles and Yin [2017].

Theorem S.7.3. *Suppose that C1–C6 hold. Fix \mathbf{G}_λ and let $\mathbf{L}(z)$ be the resolvent of $\tilde{\mathbf{Z}}_1^T \mathbf{G}_\lambda^{-1} \tilde{\mathbf{Z}}_1$ at z . Moreover, suppose that*

$$|\hat{\beta}_p - g_p| > c, \tag{S.23}$$

for all sufficiently large p and a sufficiently small constant c . Then, there exists constants $a, a' > 0$ such that the following local laws hold.

(i) *The entrywise local law in \hat{D} holds as*

$$\mathbf{L}(z) - \hat{q}_p(z)I_{n_1} = O_{<}(\hat{\Phi}(z)),$$

uniformly in $\hat{D}(a, a', n_1)$.

(ii) *The averaged local law in \hat{D} holds as*

$$|\bar{L}(z) - \hat{q}_p(z)| < \frac{1}{n_1 \eta},$$

uniformly in $\hat{D}(a, a', n_1)$.

Given Theorem S.7.3, to show Theorem S.7.1, it suffices to show the following theorem.

Theorem S.7.4. *Suppose that C1–C6 hold. \mathbf{G}_λ satisfies the following properties.*

(i) *There exists $c > 0$ such that Eq. (S.23) holds with high probability;*

(ii) *$|\hat{\Theta}_{p1} - \Theta_1| < n_2^{-1}$;*

(iii) There exists $a > 0$ such that for arbitrary $a' > 0$,

$$|\hat{q}_p(z) - q(z)| < \frac{1}{n_1 \eta},$$

uniformly in $\hat{D}(a, a', n_1)$.

The rest of this subsection is devoted to the proof of Theorem S.7.4. We separately consider two cases: $\rho > \lambda/\sigma_{\max}$ and $\rho = \lambda/\sigma_{\max}$. For both cases, since $s'(x) \rightarrow \infty$ as $x \rightarrow \rho$, we have that $\beta < \rho$. Recall the definition of $f(q)$ and the relationship between β , Θ_1 and $f(q)$ as introduced in Section S.5.2.

We first consider the case when $\rho > \lambda/\sigma_{\max}$. Theorem S.6.3 in Section S.6 indicates that

$$|g_p - \rho| < n_2^{-2/3}.$$

In order to show $|\hat{\beta}_p - g_p| \geq c$ with high probability for some constant c , it suffices to show that $|\hat{\beta}_p - \beta| < n_2^{-1}$ in probability. For $q \in \mathbb{C}^+ \cup (-\rho, 0)$, we consider $f'(q)$, $f''(q)$, and $\hat{f}'(q)$

$$\begin{aligned} f'(q) &= \frac{1}{q^2} - \gamma_1 \int \frac{d\mathcal{G}_\lambda(\tau)}{(\tau + q)^2}, \\ f''(q) &= -\frac{2}{q^3} + 2\gamma_1 \int \frac{d\mathcal{G}_\lambda(\tau)}{(\tau + q)^3}, \\ \hat{f}'(q) &= \frac{1}{q^2} - \frac{1}{n_1} \sum_{j=1}^p \frac{1}{(g_p + q)^2}. \end{aligned}$$

Since $\beta \in (0, \rho)$, there exists a domain $\mathfrak{D}_q = \{z : |E + \beta| \leq C, \ 0 \leq \eta \leq C'\}$ for some $C, C' > 0$ such that

$$|f''(q)| \asymp 1, \quad q \in \mathfrak{D}_q.$$

We notice that in $(-g_p, 0)$, $\hat{f}'(q) > 0$. Therefore, there exists at most one critical point of $\hat{f}'(q)$ on $(-g_p, 0)$. To show that $|\hat{\beta}_p - \beta| < n_2^{-1}$, we only need to show that for arbitrary small $\epsilon > 0$, there exists a root to $\hat{f}(q) = 0$ in $(-\beta - n_1^{-1+\epsilon}, -\beta + n_1^{-1+\epsilon})$ with

high probability. Clearly,

$$\Re f'(-\beta \pm n_2^{-1+\epsilon} + in_2^{-1+\epsilon/2}) \asymp \pm n_2^{-1+\epsilon},$$

$$\Im f'(-\beta \pm n_2^{-1+\epsilon} + in_2^{-1+\epsilon/2}) \asymp n_2^{-1+\epsilon/2},$$

Due to the analyticity of the function $f'(q)$ and $\hat{f}'(q)$ when $q \in \mathfrak{D}_q$, Theorem S.6.4 suggests that

$$\mathbb{1}(g_p \geq (\rho + \beta)/2) \left| \hat{f}(-\beta \pm n_2^{-1+\epsilon} + in_2^{-1+\epsilon/2}) - f(-\beta \pm n_2^{-1+\epsilon} + in_2^{-1+\epsilon/2}) \right| < \frac{1}{n_2}.$$

It follows that when $g_p \geq (\rho + \beta)/2$, with high probability

$$\Re \hat{f}'(-\beta \pm n_2^{-1+\epsilon} + in_2^{-1+\epsilon/2}) \asymp \pm n_2^{-1+\epsilon},$$

$$\Im \hat{f}'(-\beta \pm n_2^{-1+\epsilon} + in_2^{-1+\epsilon/2}) \asymp \pm n_2^{-1+\epsilon/2}.$$

Moreover, since $g_p \geq (\rho + \beta)/2$, $|\hat{f}''(q)| \asymp 1$ when $q \in \mathfrak{D}_q$, we conclude that with high probability

$$\hat{f}'(-\beta \pm n_2^{-1+\epsilon}) \asymp \pm n_2^{-1+\epsilon}.$$

It implies that there exists a critical point of $\hat{f}'(q)$ in between $-\beta \pm n_2^{-1+\epsilon}$ with high probability. It completes the proof of Result (i).

It follows immediately that when $g_p \geq (\rho + \beta)/2$,

$$\hat{\Theta}_{p1} - \Theta_1 = \hat{f}(-\hat{\beta}_p) - f(-\beta) = (\hat{f}(-\hat{\beta}_p) - \hat{f}(-\beta)) + (\hat{f}(-\beta) - f(-\beta)).$$

The first term is such that

$$\hat{f}(-\hat{\beta}_p) - \hat{f}(-\beta) = O_{<}(\hat{\beta}_p - \beta) = O_{<}(\frac{1}{n_2}).$$

Here, we are using the fact that $\hat{f}'(q)$ is bounded within \mathfrak{D}_q . For the second term, for arbitrary $\epsilon > 0$

$$\hat{f}(-\beta) - f(-\beta) = \hat{f}(-\beta + in_2^{-1+\epsilon/2}) - f(-\beta + in_2^{-1+\epsilon/2}) + O_{<}(n_2^{-1+\epsilon/2}) = O_{<}(1/n_2) + O_{<}(n_2^{-1+\epsilon/2}).$$

Here, we are using Theorem S.6.4 and the fact that $f'(q) \asymp$ and $\hat{f}(q) \asymp 1$ when $g_p \geq (\rho + \beta)/2$. It follows that $|\hat{\Theta}_{p1} - \Theta_1| < 1/n_2$.

Lastly, we show $|\hat{q}_p(z) - q(z)| < \frac{1}{n_1\eta}$ uniformly in $\hat{D}(a, a', n_1)$. Recall that

$$z = -\frac{1}{\hat{q}_p(z)} + (p/n_1) \int \frac{dF^{\mathbf{G}_\lambda}(\tau)}{\tau + \hat{q}_p(z)}.$$

In the following, we always assume the all eigenvalues of \mathbf{G}_λ are contained in \mathcal{I}_ϵ for some $\epsilon > 0$ (See Theorem S.6.4 for the definition of \mathcal{I}_ϵ). Fix $z = E + i\eta$ such that $\eta \geq 1$. Applying Lemma S.5.4, we obtain that

$$\Im \hat{q}_p(z) \geq C,$$

for some constant $C > 0$. The constant C only depends on F^{Σ_∞} , γ_1 and γ_2 . Therefore, $-\hat{q}_p(z)$ is away from the support of \mathcal{G}_λ . Applying S.6.4,

$$\left| \int \frac{dF^{\mathbf{G}_\lambda}(\tau)}{\tau + \hat{q}_p(z)} - \int \frac{d\mathcal{G}_\lambda(\tau)}{\tau + \hat{q}_p(z)} \right| < \frac{1}{n_2}.$$

It follows that

$$z + O_{<}(n_1^{-1}) = -\frac{1}{\hat{q}_p(z)} + (p/n_1) \int \frac{d\mathcal{G}_\lambda(\tau)}{\tau + \hat{q}_p(z)}.$$

Applying Lemma S.5.5, we obtain

$$|\hat{q}_p(z) - q(z)| < \frac{1}{n_2}.$$

We then use a stochastic continuity argument and Lemma S.5.5 to propagate the smallness of $|\hat{q}_p(z) - q(z)|$ from $\eta \geq 1$ to all $z \in \hat{D}$. The argument is very similar to those in the proof

of Proposition S.6.8. We therefore omit the details. All together, we conclude that

$$|\hat{q}_p(z) - q(z)| < \frac{1}{n_2\eta}.$$

As for the case when $\rho = \lambda/\sigma_{\max}$, from Theorem S.6.13, Result (i) clearly holds. Result (ii) and (iii) follow from the same arguments as in the case of $\rho = \lambda/\sigma_{\max}$. We only need to replace Theorem S.6.4 with Theorem S.6.15.

The proof of Theorem S.7.4 is completed.

Step 2: Asymptotic Tracy-Widom distribution under Gaussianity

The asymptotic Tracy-Widom law of the normalized largest root follows directly from Theorem 2.4 of Lee and Schnelli [2016]. Indeed, applying their results, we have

Theorem S.7.5. *Suppose that C1–C6 hold. Fix \mathbf{G}_λ . Suppose that \mathbf{G}_λ is such that Eq. (S.23) holds for all sufficiently large p . Then,*

$$\frac{p^{2/3}}{\hat{\Theta}_{p2}} \left(\ell_{\max}(\tilde{F}_\lambda) - \hat{\Theta}_{p1} \right) \Longrightarrow \text{TW}_1,$$

where $\hat{\Theta}_{p1}$ is defined in Step 1 and $\hat{\Theta}_{p2}$ is

$$\hat{\Theta}_{p2} = \left[(p/n_1)^3 \int \frac{dF^{\mathbf{G}_\lambda}(\tau)}{(\tau - \hat{\beta}_p)^3} + \frac{(p/n_1)^2}{\hat{\beta}_p^3} \right]^{1/3}.$$

In Theorem S.7.4, we showed that Eq. (S.23) holds with high probability and $|\hat{\Theta}_{p1} - \Theta_1| < n_1^{-1}$. Following similar arguments, it is not hard to show $\hat{\Theta}_{p2} - \Theta_2 \rightarrow 0$ in probability. Therefore, the asymptotic Tracy-Widom law of the normalized largest root

$$\frac{p^{2/3}}{\Theta_2} \left(\ell_{\max}(\tilde{F}_\lambda) - \Theta_1 \right) \Longrightarrow \text{TW}_1,$$

follows directly from the Slutsky's theorem.

Step 3: Proof of Theorem S.7.1 under general condition C2

Theorem S.7.1 holds under the moments condition **C2**. It can be shown by closely following the arguments in Sections 9.1, 9.2 and 10 of Han et al. [2018], originally developed by Knowles and Yin [2017]. Only minor changes are needed. We only give an outline of the arguments and highlight the difference from those in Han et al. [2018].

Consider the entrywise local law. It suffices to show that for any orthogonal $n_1 \times n_1$ matrix \mathbf{B}_1 and \mathbf{B}_2 ,

$$\|\mathbf{B}_1(\mathbf{L} - q(z)I_{n_1})\mathbf{B}_2^*\|_\infty < \Phi(z), \quad (\text{S.24})$$

for all $z \in S$, where S is an ϵ -net of D with $\epsilon = n_1^{-10}$. Here, \mathbf{B}_2^* is the conjugate transpose of \mathbf{B}_2 . Setting δ to be a sufficiently small positive constant such that $n_1^{24\delta}\Phi \ll 1$, for any given $\eta \geq \frac{1}{n_1}$, we define a series of numbers $\eta_0 \leq \eta_1 \leq \eta_2 \leq \dots \leq \eta_L$, based on η where

$$L := L(\eta) = \max\{L \in \mathbb{N} : \eta n_1^{l\delta} < n_1^{-\delta}\}.$$

So

$$\eta_l := \eta n_1^{l\delta}, \quad (l = 0, 1, \dots, L-1), \quad \eta_L := 1.$$

We work on the net S satisfying the condition that $E + i\eta_l \in S$, $l = 0, \dots, L$, from now on. We define $S_m := \{z \in S : \Im z \geq n_1^{-\delta m}\}$ corresponding to the following events:

$$A_m = \{\|\mathbf{B}_1(\mathbf{L}(z) - q(z)I_{n_1})\mathbf{B}_2^*\mathcal{T}_p(\mathbf{Z})\|_\infty < 1, \text{ for any } z \in S_m\} \quad (\text{S.25})$$

and

$$C_m = \{\|\mathbf{B}_1(\mathbf{L}(z) - q(z)I_{n_1})\mathbf{B}_2^*\mathcal{T}_p(\mathbf{Z})\|_\infty < \Phi, \text{ for any } z \in S_m\}. \quad (\text{S.26})$$

Here,

$$\mathcal{T}_p(\mathbf{Z}) = \begin{cases} 1 & \text{if } \|\mathbf{Z}\mathbf{Z}^T\|_F \leq C_0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that there exists sufficiently large C_0 such that $\mathcal{T}_p(\mathbf{Z}) = 1$ with high probability due to Lemma S.6.7. In the following, we select C_0 to be such a constant.

We start the induction by considering the event A_0 first. In fact, it is not hard to prove that event A_0 holds. By the assumption that $n_1^{24\delta}\Phi \ll 1$, it is easy to see that the event C_m implies the event A_m . We will prove the event (A_{m-1}) implies the event (C_m) for all $1 \leq m \leq \delta^{-1}$ in the sequel, which ensures that Eq. (S.24) holds on the set S uniformly.

Define

$$F_{st}(\mathbf{Z}, z) = ((\mathbf{B}_1 \mathbf{L}(z) \mathbf{B}_2^*)_{st} - q(z) (\mathbf{B}_1 \mathbf{B}_2^*)_{st}) \mathcal{T}_p(\mathbf{Z}), \quad s, t \in \{1, \dots, n_1\}.$$

By Markov's inequality, it suffices to calculate the upper bound of the higher moments of $F_{st}(\mathbf{Z}, z)$. Indeed, we only need to prove the following lemma.

Lemma S.7.6. *Let k be an positive even constant and $m \leq \delta^{-1}$. Suppose Eq. (S.25) for all $z \in S_{m-1}$. Then we have*

$$\mathbb{E} |F_{st}^k(\mathbf{Z}, z)| < (n_1^{24\delta}\Phi)^k$$

for all $1 \leq s, t \leq n_1$ and $z \in S_m$.

The proof of Lemma S.7.6 closely follows that of Lemma 3 in Han et al. [2018]. Similar results can be found in Lemma 7.5 of Knowles and Yin [2017] and Lemma 5 of Han et al. [2016]. In fact, when $\lambda = 0$, the central block of $\mathbf{H}(z)$ becomes zero, reducing $\mathbf{H}(z)$ to the linearizing block matrix used in Han et al. [2018] (also denoted as \mathbf{H}). It is straightforward to verify that all arguments in the proof of Lemma 3 in Han et al. [2018] do not depend on the central block. Specifically, the proof holds for any general matrix \mathbf{H} where all blocks are linear in \mathbf{Z} and satisfy certain operator norm bounds.

The proof strategy involves controlling the difference between $\mathbb{E} |F_{st}^k(\mathbf{Z}, z)|$ and $\mathbb{E} |F_{st}^k(\mathbf{Z}^{\text{Gauss}}, z)|$, where $\mathbf{Z}^{\text{Gauss}}$ has the same structure as \mathbf{Z} but with Gaussian elements. In Step 1, we established that:

$$\mathbb{E} |F_{st}^k(\mathbf{Z}^{\text{Gauss}}, z)| < (n_1^{24\delta}\Phi)^k.$$

Knowles and Yin [2017] introduced an interpolation method to control this difference. The arguments were later summarized and simplified by Han et al. [2018] in Sections 9.1.1, 9.1.2, and 9.2 of their paper. These arguments can be applied almost verbatim, with only minor modifications needed to adjust the mathematical expressions to their counterparts in our paper.

A minor adaptation is required in the arguments presented in Section 9.2 of Han et al. [2018]. While their focus was on controlling all blocks of \mathbf{H}^{-1} , our context requires consideration of only the upper left corner, $\mathbf{L} = \Delta^T \mathbf{H}^{-1} \Delta$. Thus, the arguments concerning the remaining blocks can be omitted. Further details are omitted for brevity.

Regarding the averaged local law, it can be established using the arguments in Section 10 of Han et al. [2018], with notation adjustments to align with our framework.

Step 4: Proof of the Green function comparison theorem

The theorem can be proved following the same arguments as Section 12 of Han et al. [2018] with notation adjustments to align with our framework.

S.8 Proof of Theorem 3.7, Lemma 3.8, and Lemma

3.9

First, consider the proof of Theorem 3.7. Let $F_{(p)}^{\Sigma_\infty}$ be the discretization of F^{Σ_∞} onto the points in R_p . Specifically, we assume that $F_{(p)}^{\Sigma_\infty}$ is the distribution that places masses $F^{\Sigma_\infty}((\sigma_{p,j-1}, \sigma_{pj}])$ at the point σ_{pj} , $j = 1, 2, \dots$, and $\sigma_{p0} = -\infty$. Since $\text{size}(R_p) = o(p^{-2/3})$, clearly

$$\mathcal{D}_W(F_{(p)}^{\Sigma_\infty}, F^{\Sigma_\infty}) = o(p^{-2/3}).$$

Recall the errors in Algorithm 1 as

$$e_{ij} = \frac{\hat{Q}_j(z_i)}{|\hat{Q}_j(z_i)|} - \frac{1}{|\hat{Q}_j(z_i)|} \sum_{k=1}^K \frac{\sigma_k^j w_k}{(\sigma_k \lambda \hat{\varphi}(z_i) + \lambda)^j}.$$

Using Lemma 2.2, uniformly for $y \in J_p$, $\hat{\varphi}(y) - \varphi(y) = o_P(n^{-2/3})$ and $\hat{\varphi}'(y) - \varphi'(y) = o_P(n^{-2/3})$. Therefore, together with Marčenko-Pastur equation (5),

$$\sup_{y \in J_p} \left| \frac{\hat{Q}(y)}{|\hat{Q}(y)|} - \frac{1}{|\hat{Q}(y)|} \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\tau \lambda \hat{\varphi}(y) + \lambda} \right| = o_P(p^{-2/3}).$$

It follows then

$$\sup_{y \in J_p} \left| \frac{\hat{Q}(y)}{|\hat{Q}(y)|} - \frac{1}{|\hat{Q}(y)|} \int \frac{\tau dF_{(p)}^{\Sigma_\infty}(\tau)}{\tau \lambda \hat{\varphi}(y) + \lambda} \right| = o_P(p^{-2/3}).$$

Then, we can conclude that the estimated measure \hat{F}_p is such that

$$\sup_{y \in J_p} \left| \int \frac{\tau d\hat{F}_p(\tau)}{\tau \varphi(y) + 1} - \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\tau \varphi(y) + 1} \right| = o_P(p^{-2/3}).$$

Under the assumption $\{\varphi(y), y \in J_p\}$ is dense in $D(\nu)$. Therefore,

$$\sup_{z \in D(\nu)} \left| \int \frac{\tau d\hat{F}_p(\tau)}{\tau z + 1} - \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\tau z + 1} \right| = o_P(p^{-2/3}). \quad (\text{S.27})$$

Similarly, we can get

$$\sup_{z \in D(\nu)} \left| \int \frac{\tau^2 d\hat{F}_p(\tau)}{(\tau z + 1)^2} - \int \frac{\tau^2 dF^{\Sigma_\infty}(\tau)}{(\tau z + 1)^2} \right| = o_P(p^{-2/3}). \quad (\text{S.28})$$

Notice that the functions are analytic in z . Consider the power series expansion

$$\begin{aligned} \int \frac{\tau d\hat{F}_p(\tau)}{\tau z + 1} - \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{\tau z + 1} &= \sum_{m=0}^{\infty} (-1)^m z^m \left(\int \tau^{m+1} (d\hat{F}_p(\tau) - dF^{\Sigma_\infty}(\tau)) \right) \\ &= \sum_{m=0}^{\infty} (-1)^m z^m (\mu_{m+1}(\hat{F}_p) - \mu_{m+1}(F^{\Sigma_\infty})). \end{aligned}$$

Here, $\mu_m(\cdot)$ is the m th moment of a distribution. Using the Cauchy's estimates on the power series coefficients, we have

$$\delta_p = \sup_{m \geq 1} p^{2/3} \nu^m |\mu_m(\hat{F}_p) - \mu_m(F^{\Sigma_\infty})| = o_P(1).$$

Then, it follows that the characteristic functions of \hat{F}_p and F^{Σ_∞} are such that for $t \in \mathbb{R}$,

$$\left| \frac{\mathbb{E}_{\hat{F}_p} e^{itX} - \mathbb{E}_{F^{\Sigma_\infty}} e^{itX}}{t} \right| \leq C \nu^{-1} \delta_p p^{-2/3} e^{t/\nu},$$

$$\left| \frac{d}{dt} \frac{\mathbb{E}_{\hat{F}_p} e^{itX} - \mathbb{E}_{F^{\Sigma_\infty}} e^{itX}}{t} \right| \leq C \nu^{-2} \delta_p p^{-2/3} e^{t/\nu}.$$

Following from Corollary 8.3 of Bobkov [2016], for all $T \geq 3$

$$\mathcal{D}_W(\hat{F}_p, F^{\Sigma_\infty}) \leq \left(\int_{-T}^T \left| \frac{\mathbb{E}_{\hat{F}_p} e^{itX} - \mathbb{E}_{F^{\Sigma_\infty}} e^{itX}}{t} \right|^2 dt \right)^{1/2} + \left(\left| \frac{d}{dt} \frac{\mathbb{E}_{\hat{F}_p} e^{itX} - \mathbb{E}_{F^{\Sigma_\infty}} e^{itX}}{t} \right|^2 dt \right)^{1/2} + \frac{C}{T}.$$

It follows then

$$\mathcal{D}_W(\hat{F}_p, F^{\Sigma_\infty}) \leq \frac{C(1+\nu)}{\nu^2} p^{-2/3} \exp(T) \delta_p + \frac{C}{T}.$$

Note that setting $T = \log(p^{1/3})$, we have

$$\mathcal{D}_W(\hat{F}_p, F^{\Sigma_\infty}) = o_P(1).$$

Next, we consider the scenario when J_p is such that $\{\varphi(y) : y \in J_p\}$ eventually covers $[-h_\beta/\lambda - \varepsilon, -h_\beta/\lambda + \varepsilon]$. Again, due to the fact that uniformly for $y \in J_p$, $\hat{\varphi}(y_p) - \varphi(y_p) = o_P(p^{-2/3})$, $\hat{\varphi}'(y_p) - \varphi'(y_p) = o_P(p^{-2/3})$, Eq. (S.27) and Eq. (S.28),

$$\sup_{h \in [h_\beta - \lambda\varepsilon, h_\beta + \lambda\varepsilon]} \left| \int \frac{\tau d\hat{F}_p(\tau)}{-h\tau + \lambda} - \int \frac{\tau dF^{\Sigma_\infty}(\tau)}{-h\tau + \lambda} \right| = o_P(p^{-2/3}).$$

$$\sup_{h \in [h_\beta - \lambda\varepsilon, h_\beta + \lambda\varepsilon]} \left| \int \frac{\tau^2 d\hat{F}_p(\tau)}{(-h\tau + \lambda)^2} - \int \frac{\tau^2 dF^{\Sigma_\infty}(\tau)}{(-h\tau + \lambda)^2} \right| = o_P(p^{-2/3}).$$

It indicates that uniformly in $h \in [h_\beta - \lambda\varepsilon, h_\beta + \lambda\varepsilon]$,

$$\hat{\mathcal{H}}_1(h) - \mathcal{H}_1(h) = o_P(p^{-2/3}), \quad \hat{\mathcal{H}}_2(h) - \mathcal{H}_2(h) = o_P(p^{-2/3}) \quad \text{and} \quad \hat{\mathcal{H}}_3(h) - \mathcal{H}_3(h) = o_P(1).$$

Here, the bound on $\hat{\mathcal{H}}_3(h) - \mathcal{H}_3(h)$ is due to $\mathcal{D}_W(\hat{F}_p, F^{\Sigma_\infty})$. As x , $s(x)$, $s'(x)$, and $s''(x)$ are smooth in $\mathcal{H}_j(h)$, $j = 1, 2, 3$ as shown in Theorem 3.3, we get

$$\hat{\beta} - \beta = o_P(p^{-2/3}), \quad \hat{s}(\hat{\beta}) - s(\beta) = o_P(p^{-2/3}), \quad \hat{s}'(\hat{\beta}) - s'(\beta) = o_P(p^{-2/3}), \quad \hat{s}''(\hat{\beta}) - s''(\beta) = o_P(1).$$

Consequently, $\hat{\Theta}_1 - \Theta_1 = o_P(p^{-2/3})$ and $\hat{\Theta}_2 - \Theta_2 = o_P(1)$.

As for Lemma 3.8, recall that $\varphi(z)$ is the Stieltjes transform of \mathcal{W} (Lemma 2.2). Then,

$$\varphi(z) = \int \frac{d\mathcal{W}}{\tau - z}.$$

It follows that $\{\varphi(z) : z \in \mathbb{C}^+ \text{ and } \text{dist}(z, \text{supp}(\mathcal{W})) \geq c\}$ contains the half disk $\mathcal{D}(\nu)$ for some $\nu > 0$. We can then select J_p such that $\varphi(y)$ is dense in $\mathcal{D}(\nu)$ when $y \in J_p$.

As for Lemma 3.9, as $\gamma_1 \rightarrow \infty$, $\beta \rightarrow 0$ and $h_\beta/\lambda \rightarrow -\varphi(-\lambda)$. Therefore, select J_p to be a grid containing a subsequence dense in $[-3/2\lambda, -\lambda/2]$. Then, $\{\varphi(y) : y \in J_p\}$ will eventually covers a neighborhood of $-h_\beta/\lambda$.

S.9 Proof of Lemma 4.1, Lemma 4.2, Lemma S.2.1, and Lemma S.2.2

Lemma 4.1 follows directly from Theorem 2.9 and Slutsky's theorem. In particular, under the assumption that $\hat{\theta} = \Theta_1 + o_P(p^{-2/3})$ and $\hat{\Theta}_2 = \Theta_2 + o_P(1)$,

$$\hat{\ell}_{\max}(\mathbf{F}_\lambda) \implies \mathbb{TW}_1.$$

Consider the proof of Lemma 4.2. Under $H_a : BC \neq 0$, we can decompose \mathbf{F}_λ as

$$\begin{aligned} \mathbf{F}_\lambda &= \mathbf{F}_\lambda^{(0)} + \frac{1}{n_1} BC(C^T(XX^T)^{-1}C)^{-1}C^T B^T(\mathbf{W}_2 + \lambda I_p)^{-1} \\ &\quad + \frac{1}{n_1} (BX P_1 Z^T \Sigma_p^{1/2} + \Sigma_p^{1/2} Z P_1 X^T B^T) (\mathbf{W}_2 + \lambda I_p)^{-1}, \end{aligned}$$

where $\mathbf{F}_\lambda^{(0)}$ is the regularized F -matrix under the null hypothesis $H_0 : BC = 0$. By Theorem 2.9,

$$\ell_{\max}(\mathbf{F}_\lambda^{(0)}) = O_P(1).$$

Under the assumption

$$\frac{1}{n_1} \ell_{\max}(BC(C^T(XX^T)^{-1}C)^{-1}C^T B^T) = O((\log p)^c),$$

we have

$$\frac{1}{n_1} \ell_{\max}(BC(C^T(XX^T)^{-1}C)^{-1}C^T B^T(\mathbf{W}_2 + \lambda I_p)^{-1}) = O_P((\log p)^c).$$

Moreover,

$$\left\| \frac{1}{n_1} B X P_1 Z^T \Sigma_p^{1/2} \right\|_2 \leq \left\| \frac{1}{n_1} BC(C^T(XX^T)^{-1}C)^{-1}C^T B^T \right\|_2^{1/2} \left\| \frac{1}{n_1} \Sigma_p^{1/2} Z Z^T \Sigma_p^{1/2} \right\|_2^{1/2} = O_P((\log p)^{c/2}).$$

Here, we are using the fact that $\|n_1^{-1}\Sigma_p^{1/2}ZZ^T\Sigma_p^{1/2}\|_2 = O_P(1)$ due to Lemma S.6.7.

All together, $\ell_{\max}(\mathbf{F}_\lambda) = O_P((\log p)^c)$ and

$$P\left(\hat{\ell}_{\max}(\mathbf{F}_\lambda) > \mathbb{T}\mathbb{W}_1(1 - \alpha) \mid BC\right) \rightarrow 1.$$

It completes the proof of Lemma 4.2.

Lemma S.2.1 again follows directly from Theorem 2.9 and Slutsky's theorem. Lastly, consider the proof of Lemma S.2.2. Under $H_a : LL^T \neq 0$, we can decompose \mathbf{W}_1 as

$$\mathbf{W}_1 = \frac{1}{n_1}L\xi P_1\xi^T L^T + \frac{1}{n_1}\Sigma_p^{1/2}Z_1P_1Z_1^T\Sigma_p^{1/2} + \frac{1}{n_1}L\xi P_1Z_1^T\Sigma_p^{1/2} + \frac{1}{n_1}\Sigma_p^{1/2}Z_1P_1\xi^T L^T,$$

where $\xi = (\xi_1, \dots, \xi_{n_{(1)}})$, $Z_1 = (Z_{11}, \dots, Z_{1n_{(1)}})$ and P_1 is a projection matrix of rank n_1 .

Notice that

$$\begin{aligned} \left\| \frac{1}{n_1(\log p)^c} L\xi P_1\xi^T L^T (\mathbf{W}_2 + \lambda I_p)^{-1} \right\|_2 &= O_P(1), \\ \left\| \frac{1}{n_1}\Sigma_p^{1/2}Z_1P_1Z_1^T (\mathbf{W}_2 + \lambda I_p)^{-1} \right\|_2 &= O_P(1), \\ \left\| \frac{1}{n_1(\log p)^{c/2}} L\xi P_1Z_1^T\Sigma_p^{1/2} (\mathbf{W}_2 + \lambda I_p)^{-1} \right\|_2 &= O_P(1). \end{aligned}$$

We conclude that

$$\frac{1}{(\log p)^c} \|\mathbf{W}_1(\mathbf{W}_2 + \lambda I_p)^{-1}\|_2 = O_P(1).$$

It follows that

$$P\left(\hat{\ell}_{\max}(\mathbf{F}_\lambda) > \mathbb{T}\mathbb{W}_1(1 - \alpha) \mid LL^T\right) \rightarrow 1.$$

S.10 Proof of other technical results

Lemma 2.8, Lemma 3.1 and Corollary 3.2 can be proved using Theorem 3.2, Theorem 4.1, and Theorem 5.1 of Li [2024]. Specifically, consider first the case when F^{Σ_∞} is discrete at

σ_{\max} and $\omega_{\max}\gamma_2 > 1$. Using Theorem 3.2 of Li [2024], \mathcal{G}_λ is discrete at ρ . Therefore,

$$s'(x) = \frac{C}{(\rho - x)^2} + C \int_{\tau > \rho} \frac{d\mathcal{G}_\lambda(\tau)}{\tau - x} \rightarrow \infty, \quad \text{as } x \uparrow \rho.$$

When $\omega_{\max}\gamma_2 < 1$, it is straightforward to verify that there exists a $c > 0$ such that $x'(h) < 0$ when $h \in [\lambda/\sigma_{\max} - c, \lambda/\sigma_{\max})$. On the other, as $h \rightarrow -\infty$, $x'(h) \rightarrow 1$. Therefore, there exists $h_0 < \lambda/\sigma_{\max}$ such that $x'(h_0) = 0$. Using Theorem 4.1 of Li [2024], $\rho = x(h_0)$. Using Theorem 5.1 of Li [2024], there exists constants C_1, C_2 and ϵ such that

$$C_1\sqrt{x - \rho} \leq f_{\mathcal{G}_\lambda}(x) \leq C_2\sqrt{x - \rho}, \quad x \in (\rho, \rho + \epsilon)$$

where $f_{\mathcal{G}_\lambda}$ is the density function of \mathcal{G}_λ . It follows that

$$s'(x) \geq \int_{\rho}^{\rho + \epsilon} \frac{f_{\mathcal{G}_\lambda}(\tau) d\tau}{(\tau - x)^2} \rightarrow \infty, \quad \text{as } x \uparrow \rho.$$

When F^{Σ_∞} is continuous at σ_{\max} and (ii) of Definition 2.7 holds, Theorem 5.1 of Li [2024] indicates that the square-root behavior of $f_{\mathcal{G}_\lambda}(x)$ still holds. Therefore, $s'(x) \rightarrow \infty$ as $x \uparrow \rho$ due to the same arguments.

For the proof of Theorem 3.3, Eq. (16) follows directly from an application of Theorem 4.1 of Li [2024]. Eq. (17) is obtained by differentiating both sides of Eq. (16) with respect to x_0 .