Submitted to Bernoulli

High-dimensional general linear hypothesis tests via non-linear spectral shrinkage

HAORAN LI[1,*] ALEXANDER AUE[1,**] and DEBASHIS PAUL[1,†]

[1] *Department of Statistics, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA. E-mail:* [*]hrli@ucdavis.edu; [**]aaue@ucdavis.edu; [†]debpaul@ucdavis.edu

We are interested in testing general linear hypotheses in a high-dimensional multivariate linear regression model. The framework includes many well-studied problems such as two-sample tests for equality of population means, MANOVA and others as special cases. A family of rotation-invariant tests is proposed that involves a flexible spectral shrinkage scheme applied to the sample error covariance matrix. The asymptotic normality of the test statistic under the null hypothesis is derived in the setting where dimensionality is comparable to sample sizes, assuming the existence of certain moments for the observations. The asymptotic power of the proposed test is studied under various local alternatives. The power characteristics are then utilized to propose a data-driven selection of the spectral shrinkage function. As an illustration of the general theory, we construct a family of tests involving ridge-type regularization and suggest possible extensions to more complex regularizers. A simulation study is carried out to examine the numerical performance of the proposed tests.

Keywords: General linear hypothesis, Local alternatives, Ridge shrinkage, Random matrix theory, Spectral shrinkage.

1. Introduction

In multivariate analysis, one of the fundamental inferential problems is to test a hypothesis involving a linear transformation of regression coefficients under a linear model. Suppose $\mathbf{Y}$ is a $p \times N$ matrix of observations modeled as

$$\mathbf{Y} = BX + \Sigma_p^{1/2}\mathbf{Z} , \tag{1.1}$$

where (i) B is a $p \times k$ matrix of regression coefficients; (ii) X is a $k \times N$ design matrix of rank k; (iii) $\mathbf{Z}$ is a $p \times N$ matrix with i.i.d. entries having zero mean and unit variance; and (iv) Σ_p, a $p \times p$ nonnegative definite matrix, is the population covariance matrix of the errors, with $\Sigma_p^{1/2}$ a "square-root" of Σ_p so that $\Sigma_p = \Sigma_p^{1/2}(\Sigma_p^{1/2})^T$. General linear hypotheses involving the linear model (1.1) are of the form

$$H_0\colon BC = 0 \qquad \text{vs.} \qquad H_a\colon BC \neq 0, \tag{1.2}$$

for an arbitrary $k \times q$ "constraints matrix" C, subject to the requirement that BC is estimable. Without loss of generality, C is taken to be of rank q. Throughout, we assume that q and k are fixed, even as observation dimension p and sample size N increase to infinity. Henceforth, $n = N - k$ is used to denote the effective sample size, which is also the degree of freedom associated with the sample error covariance matrix.

With various choices of X and C, the testing formulation incorporates many hypotheses of interest. For example, multivariate analysis of variance (MANOVA) is a special case. When the sample size N is substantially larger than the dimension p of the observations, this problem is well-studied. Anderson (1958) and Muirhead (2009) are among standard references. Various classical inferential procedures involve the matrices

$$\widehat{\mathbf{\Sigma}}_p = \frac{1}{n}\mathbf{Y}(I - X^T(XX^T)^{-1}X)\mathbf{Y}^T, \tag{1.3}$$

$$\widehat{\mathbf{H}}_p = \frac{1}{n}\mathbf{Y}X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1}C^T(XX^T)^{-1}X\mathbf{Y}^T, \tag{1.4}$$

so that $\widehat{\mathbf{\Sigma}}_p$ is the residual covariance of the full model, an estimator of Σ_p, while $\widehat{\mathbf{H}}_p$ is the hypothesis sums of squares and cross products matrix, scaled by n^{-1}. In a one-way MANOVA set-up, $\widehat{\mathbf{\Sigma}}_p$ and $\widehat{\mathbf{H}}_p$ are, respectively, the within-group and between-group sums of squares and products matrices, scaled by n^{-1}. In the rest of the paper, we shall refer to $\widehat{\mathbf{\Sigma}}_p$ as the sample covariance matrix.

The testing problem (1.2) is well-studied in the classical multivariate analysis literature. Three standard test procedures are the likelihood ratio test (LR), Lawley–Hotelling trace test (LH) and Bartlett–Nanda–Pillai trace (BNP) test. They are called *invariant tests*, since under Gaussianity the null distributions of the test statistics are invariant with respect to Σ_p. One common feature is that all test statistics are linear functionals of the spectrum of $\widehat{\mathbf{H}}_p\widehat{\mathbf{\Sigma}}_p^{-1}$. Since this matrix is asymmetric, for convenience, a standard transformation is applied, giving the expressions of the invariant tests as follows. Define

$$Q_n = X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1/2}, \tag{1.5}$$

$$\mathbf{M}_0 = \frac{1}{n}Q_n^T\mathbf{Y}^T\widehat{\mathbf{\Sigma}}_p^{-1}\mathbf{Y}Q_n.$$

The matrix $Q_nQ_n^T$ is the "hat matrix" of the reduced model under the null hypothesis. Note that the non-zero eigenvalues of $\widehat{\mathbf{H}}_p\widehat{\mathbf{\Sigma}}_p^{-1} = n^{-1}\mathbf{Y}Q_nQ_n^T\mathbf{Y}^T\widehat{\mathbf{\Sigma}}_p^{-1}$ are the same as those of $\mathbf{M}_0$. The test statistics for the LR, LH and BNP tests can be expressed as

$$T_0^{\mathrm{LR}} = \sum_{i=1}^q \log\{1 + \lambda_i(\mathbf{M}_0)\},$$

$$T_0^{\mathrm{LH}} = \sum_{i=1}^q \lambda_i(\mathbf{M}_0),$$

$$T_0^{\mathrm{BNP}} = \sum_{i=1}^q \lambda_i(\mathbf{M}_0)/\{1 + \lambda_i(\mathbf{M}_0)\}.$$

The symbol $\lambda_i(\cdot)$ denotes the i-th largest eigenvalue of a symmetric matrix, further using the convention that $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ indicate the largest and smallest eigenvalue, respectively.

In contemporary statistical research and applications, high-dimensional data whose dimension is at least comparable to the sample size is ubiquitous. In this paper, focus is on the interesting boundary case when dimension and sample sizes are comparable. Primarily due to inconsistency of conventional estimators of model parameters — such as $\widehat{\mathbf{\Sigma}}_p$ —, classical test procedures for the hypothesis (1.2) — such as the LR, LH and BNP tests — perform poorly in such settings. When the dimension p is larger than the degree of freedom n, the invariant tests are not even well-defined because $\widehat{\mathbf{\Sigma}}_p$ is singular. Even when p is strictly less than n, but the ratio $\gamma_n = p/n$ is close to 1, these tests are known to have poor power behavior. Asymptotic results when $\gamma_n \to \gamma \in (0, 1)$ were obtained in Fujikoshi, Himeno and Wakaki (2004) under Gaussianity of the populations, and more recently in Bai, Choi and Fujikoshi (2017) under more general settings that only require the existence of certain moments.

Pioneering work on modifying the classical solutions in high dimension is in Bai et al. (2013), who corrected the scaling of the LR statistic when $n \geqslant p$ but p, k and q are proportional to n. The corrected LR statistic was shown to have significantly more power than its classical counterpart. In contrast, in this paper, we focus on the setting where k and q are fixed even as $n, p \to \infty$ so that $\gamma_n = p/n \to \gamma \in (0, \infty)$. In the multivariate regression problem, this corresponds to a situation where the response is high-dimensional, while the predictor is finite-dimensional. In the MANOVA problem, this framework corresponds to high-dimensional observations belonging to one of a finite number of populations.

To the best of our knowledge, when $n < p$, the linear hypothesis testing problem has been studied in depth only for specific submodels of (1.1), primarily for the important case of two-sample tests for equality of population means. For the latter tests, a widely used idea is to construct modified statistics based on replacing $\widehat{\mathbf{\Sigma}}_p^{-1}$ with an appropriate substitute. This approach was pioneered in Bai and Saranadasa (1996) and further developed in Chen and Qin (2010). Various extensions to one-way MANOVA (Srivastava and Fujikoshi, 2006; Yamada and Himeno, 2015; Srivastava and Fujikoshi, 2006; Hu et al., 2017) and a general multi-sample Behrens–Fisher problem under heteroscedasticity (Zhou, Guo and Zhang, 2017) exist. Other notable works for the two-sample problem include Biswas and Ghosh (2014); Chang et al. (2017); Chen, Li and Zhong (2014); Guo and Chen (2016); Lopes, Jacob and Wainwright (2011); Srivastava, Li and Ruppert (2016); Wang, Peng and Li (2015). A second approach aims to regularize the matrix $\widehat{\mathbf{\Sigma}}_p$ to address the issue of its near-singularity in high dimensions; see Chen et al. (2011) and Li et al. (2016) for ridge-type penalties in two-sample settings. Finally, another alternative line of attack consists of exploiting sparsity; see Cai, Liu and Xia (2014); Cai and Xia (2014).

In this paper, we seek to regularize the spectrum of $\widehat{\mathbf{\Sigma}}_p$ by flexible shrinkage functions. For a symmetric $p \times p$ matrix A and a function $g(\cdot)$ on $\mathbb{R}$, define

$$g(A) = R_A \mathrm{diag}\big(g(\lambda_1(A)), \ldots, g(\lambda_p(A))\big) R_A^T,$$

where R_A is the matrix of eigenvectors associated with the ordered eigenvalues of A. Now, consider any real-valued function $f(\cdot)$ on $\mathbb{R}$ that is analytic over a specific domain associated with the limiting behavior of the eigenvalues of $\widehat{\mathbf{\Sigma}}_p$, as elaborated in Section 2. The proposed statistics are functionals of eigenvalues of the regularized quadratic forms

$$\mathbf{M}(f) = \frac{1}{n} Q_n^T \mathbf{Y}^T f(\widehat{\mathbf{\Sigma}}_p) \mathbf{Y} Q_n.$$

Specifically, we propose regularized versions of LR, LH and BNP test criteria, respectively, namely

$$\begin{aligned}
T^{\mathrm{LR}}(f) &= \sum_{i=1}^{q} \log\{1 + \lambda_i(\mathbf{M}(f))\}, \\
T^{\mathrm{LH}}(f) &= \sum_{i=1}^{q} \lambda_i(\mathbf{M}(f)), \\
T^{\mathrm{BNP}}(f) &= \sum_{i=1}^{q} \lambda_i(\mathbf{M}(f))/\{1 + \lambda_i(\mathbf{M}(f))\}.
\end{aligned}$$

These test statistics are designed to capture possible departures from the null hypothesis, when $\widehat{\mathbf{\Sigma}}_p$ is replaced by $f(\widehat{\mathbf{\Sigma}}_p)$, while suitable choices of the regularizer f allow for getting around the problem of singularity or near-singularity when p is comparable to n.

Notice that $\mathbf{M}(f)$ has the same non-zero eigenvalues as $f(\widehat{\mathbf{\Sigma}}_p)\widehat{\mathbf{H}}_p$. Thus, the proposed test family is a generalization of the classical statistics based on $\widehat{\mathbf{\Sigma}}_p^{-1}\widehat{\mathbf{H}}_p$. Importantly, $\mathbf{M}(f)$ — and consequently the proposed statistics — is *rotation-invariant*, which means if a linear transformation is applied to the observations with an arbitrary orthogonal matrix, the statistic remains unchanged. It is a desirable property when not much additional knowledge about Σ_p and BC is available. It should be noted that the two-sample mean tests by Bai and Saranadasa (1996) and Li et al. (2016), together with their generalization to MANOVA, are special cases of the proposed family with $f(x) = 1$ and $f(x) = 1/(x+\lambda)$, $\lambda > 0$, respectively.

The present work builds on the work by Li et al. (2016). The theoretical analysis also involves an extension of the analytical framework adopted by Pan and Zhou (2011) in their study of the asymptotic behavior of Hotelling's T^2 statistic for non-Gaussian observations. However, the current work goes well beyond the existing literature in several aspects. We highlight these as the key contributions of this manuscript: (a) We propose new families of rotation-invariant tests for general linear hypotheses for multivariate regression problems involving high-dimensional response and fixed-dimensional predictor variables that incorporate a flexible regularization scheme to account for the dimensionality of the observations growing proportional to the sample size. (b) Unlike Li et al. (2016), who assumed sub-Gaussianity, here only the existence of finite fourth moments of the observations is required. (c) Unlike Pan and Zhou (2011), who assumed $\Sigma_p = I_p$, Σ_p is allowed to be fairly arbitrary and subjected only to some standard conditions on the limiting behavior of its spectrum. (d) We carry out a detailed analysis of the power characteristics of the proposed tests. The proposal of a class of local alternatives enables a clear interpretation of the contributions of different parameters in the performance of the test. (e) We develop a data-driven test procedure based on the principle of maximizing asymptotic power under appropriate local alternatives. This principle leads to the definition of a composite test that combines the optimal tests associated with a set of different kinds of local alternatives. The latter formulation is an extension of the data-adaptive test procedure designed by Li et al. (2016) for the two-sample testing problem.

The rest of the paper is organized as follows. Section 2 introduces the asymptotics of the proposed test family both under the null hypothesis and under a class of local alternatives. Using these local alternatives, in Section 3 a data-driven shrinkage selection methodology based on maximizing asymptotic power is developed. In Section 4, an application of the asymptotic theory and the shrinkage selection method is given for the ridge-regularization family. An extension of ridge-regularization to higher orders is also discussed. The results of a simulation study are reported in Section 5. Section 6 contains additional discussion. In the Appendix, proof outlines of the main theorems are presented, while technical details are collected in the Supplementary Material.

2. Asymptotic theory

After giving necessary preliminaries on *Random Matrix Theory* (RMT), the asymptotic theory of the proposed tests under the null hypothesis and under various local alternative models is presented in this section. For any $p \times p$ symmetric matrix A, define the *Empirical Spectral Distribution* (ESD) F^A of A by

$$F^A(\tau) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}_{\{\lambda_i(A) \leqslant \tau\}}.$$

In the following, $\|\cdot\|_{\max}$ stands for the maximum absolute value of the entries of a matrix. The following assumptions are employed.

- **C1** (Moment conditions) The entries z_{ij} of $\mathbf{Z}$ are i.i.d. such that $\mathbb{E}[z_{ij}] = 0$, $\mathbb{E}[z_{ij}^2] = 1$, $\mathbb{E}[z_{ij}^4] < \infty$;
- **C2** (High-dimensional setting) k and q are fixed, while $p, n \to \infty$ such that $\gamma_n = p/n \to \gamma \in (0, \infty)$ and $\sqrt{n}|\gamma_n - \gamma| \to 0$;
- **C3** (Boundedness of spectral norm) Σ_p is non-negative definite and $\limsup_p \lambda_{\max}(\Sigma_p) < \infty$;
- **C4** (Asymptotic stability of ESD) There exists a distribution L^Σ with compact support in $[0, \infty)$, non-degenerate at zero, such that $\sqrt{n}\mathcal{D}_W(F^{\Sigma_p}, L^\Sigma) \to 0$, as $n, p \to \infty$, 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 (Asymptotically full rank) X is of full rank and $n^{-1}XX^T$ converges to a positive definite $k \times k$ matrix. Moreover, $\limsup_{n\to\infty} \|X\|_{\max} < \infty$;

C6 (Asymptotically estimable) $\liminf_{n\to\infty} \lambda_{\min}(C^T(n^{-1}XX^T)^{-1}C) > 0$.

2.1. Preliminaries on random matrix theory

Recall that the Stieltjes transform $m_G(\cdot)$ of any function G of bounded variation on $\mathbb{R}$ is defined by

$$m_G(\mathbb{z}) = \int_{-\infty}^{\infty} \frac{dG(x)}{x - \mathbb{z}}, \qquad \mathbb{z} \in \mathbb{C}^+ := \{u + iv\colon v > 0\}.$$

Minor modifications of a standard RMT result imply that, under Conditions **C1**–**C6**, the ESD $F^{\widehat{\mathbf{\Sigma}}_p}$ converges almost surely to a nonrandom distribution F^∞ at all points of continuity of F^∞. This limit is determined in such a way that for any $z \in \mathbb{C}^+$, the Stieltjes transform $m(\cdot) = m_{F^\infty}(\cdot)$ of F^∞ is the unique solution in $\mathbb{C}^+$ of the equation

$$m(\mathbb{z}) = \int \frac{dL^\Sigma(\tau)}{\tau(1 - \gamma - \gamma\mathbb{z}m(\mathbb{z})) - \mathbb{z}}. \tag{2.1}$$

Equation (2.1) is often referred to as the Marčenko–Pastur equation. Moreover, pointwise almost surely for $\mathbb{z} \in \mathbb{C}^+$, $m_{F^{\widehat{\mathbf{\Sigma}}_p}}(\mathbb{z})$ converges to $m_{F^\infty}(\mathbb{z})$. The convergence holds even when $\mathbb{z} \in \mathbb{R}_-$ (negative reals) with a smooth extension of m_{F^∞} to $\mathbb{R}_-$. Readers may refer to Bai and Silverstein (2004) and Paul and Aue (2014) for more details. From now on, for notational simplicity, we shall write $m_{F^\infty}(\mathbb{z})$ as $m(\mathbb{z})$ and write $m_{F^{\widehat{\mathbf{\Sigma}}_p}}(\mathbb{z})$ as $m_{n,p}(\mathbb{z})$. Note that

$$m_{n,p}(\mathbb{z}) = \frac{1}{p}\mathrm{tr}(\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}$$

and define

$$\Theta(\mathbb{z}, \gamma) = \{1 - \gamma - \gamma\mathbb{z}m(\mathbb{z})\}^{-1}. \tag{2.2}$$

It is known that $(\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}$, for any fixed $\mathbb{z} \in \mathbb{C}^+$, has a *deterministic equivalent* (Bai and Silverstein (2004); Liu, Aue and Paul (2015); Li et al. (2016)), given by

$$\{\Theta^{-1}(\mathbb{z}, \gamma)\Sigma_p - \mathbb{z}I\}^{-1},$$

in the sense that for symmetric matrices A bounded in operator norm, as $n \to \infty$,

$$p^{-1}\mathrm{tr}\Big[(\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}A\Big] - p^{-1}\mathrm{tr}\Big[\{\Theta^{-1}(\mathbb{z}, \gamma)\Sigma_p - \mathbb{z}I\}^{-1}A\Big] \to 0, \quad \text{with probability 1.}$$

Resolvent and deterministic equivalent will be used frequently in this paper. They will appear for example as Cauchy kernels in contour integrals in various places.

2.2. Asymptotics under the null hypothesis

To begin with, for $k \geqslant 1$, denote by $\mathbf{W} = [w_{ij}]_{i,j=1}^k$ the *Gaussian Orthogonal Ensemble* (**GOE**) defined by (1) $w_{ij} = w_{ji}$; (2) $w_{ii} \sim \mathcal{N}(0,1)$, $w_{ij} \sim \mathcal{N}(0,1/2)$, $i \neq j$; (3) w_{ij}'s are jointly independent for $1 \leqslant i \leqslant j \leqslant k$. Throughout this paper, $f(\cdot)$ is assumed to be analytic in an open interval containing

$$\mathcal{X} := [0,\ \limsup_{p\to\infty} \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2].$$

Let $\mathcal{C}$ to be a closed contour enclosing $\mathcal{X}$ such that $f(\cdot)$ has a complex extension to the interior of $\mathcal{C}$. Further use $\mathcal{C}^2$ to denote $\mathcal{C} \otimes \mathcal{C} = \{(\mathbb{z}_1, \mathbb{z}_2) \colon \mathbb{z}_1, \mathbb{z}_2 \in \mathcal{C}\}$. The asymptotic null distribution is determined in the next theorem.

Theorem 2.1 *Suppose **C1**–**C6** hold. Under the null hypothesis $H_0 \colon BC = 0$,*

$$\sqrt{n}\{\mathbf{M}(f) - \Omega(f,\gamma) I_q\} \Longrightarrow \Delta^{1/2}(f,\gamma)\mathbf{W},$$

where $\Longrightarrow$ denotes weak convergence and $\Omega(f,\gamma)$ and $\Delta(f,\gamma)$ are as follows. With $\Theta(\mathbb{z},\gamma)$ defined in (2.2),

$$\Omega(f,\gamma) = \frac{-1}{2\pi i} \oint_{\mathcal{C}} f(\mathbb{z})(\Theta(\mathbb{z},\gamma) - 1) d\mathbb{z}.$$

For any two analytic functions f_1 and f_2,

$$\Delta(f_1, f_2, \gamma) = \frac{2}{(2\pi i)^2} \oiint_{\mathcal{C}^2} f_1(\mathbb{z}_1) f_2(\mathbb{z}_2) \delta(\mathbb{z}_1, \mathbb{z}_2, \gamma) d\mathbb{z}_1 d\mathbb{z}_2,$$

and $\Delta(f,f,\gamma)$ is written as $\Delta(f,\gamma)$ for simplicity. The kernel $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$ is such that

$$\begin{aligned}
\delta(\mathbb{z}_1, \mathbb{z}_2, \gamma) &= \Theta(\mathbb{z}_1,\gamma)\Theta(\mathbb{z}_2,\gamma)\Big[\frac{\mathbb{z}_1\Theta(\mathbb{z}_1,\gamma) - \mathbb{z}_2\Theta(\mathbb{z}_2,\gamma)}{\mathbb{z}_1 - \mathbb{z}_2} - 1\Big], \\
\delta(\mathbb{z}, \mathbb{z}, \gamma) &= \lim_{\mathbb{z}_2 \to \mathbb{z}} \delta(\mathbb{z}, \mathbb{z}_2, \gamma) = \Theta^2(\mathbb{z},\gamma)\Big[\frac{\partial \mathbb{z}\Theta(\mathbb{z},\gamma)}{\partial \mathbb{z}} - 1\Big] \\
&= \gamma\{1 + \mathbb{z}m(\mathbb{z})\}\Theta^3(\mathbb{z},\gamma) + \gamma\mathbb{z}\{m(\mathbb{z}) + \mathbb{z}m'(\mathbb{z})\}\Theta^4(\mathbb{z},\gamma).
\end{aligned}$$

The contour integral is taken counter-clockwise.

Using knowledge of the eigenvalues of the **GOE** leads to the following statement.

Corollary 2.1 *Let the conditions of Theorem 2.1 be satisfied. Assume that $\Delta(f,\gamma) > 0$ and let*

$$\tilde{\lambda}_i = \frac{\sqrt{n}}{\Delta^{1/2}(f,\gamma)}\{\lambda_i(\mathbf{M}(f)) - \Omega(f,\gamma)\}, \qquad i = 1, \ldots, q.$$

Then, the limiting joint density function of $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_q)$ at $y_1 \geqslant y_2 \geqslant \cdots \geqslant y_q$ is given by

$$\Big(2^{q/2}\prod_{i=1}^q \Gamma(i/2)\Big)^{-1} \prod_{i<j}(y_i - y_j) \exp\Big(-\frac{1}{2}\sum_{i=1}^q y_i^2\Big).$$

Although without closed forms, $\Omega(f,\gamma)$ and $\Delta(f,\gamma)$ do not depend on the choice of $\mathcal{C}$ used to compute the contour integral. With the resolvent as kernel $\mathbf{M}(f)$ can be expressed as the integral of

$$f(\mathbb{z})n^{-1}Q_n^T\mathbf{Y}^T(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}I_p)^{-1}\mathbf{Y}Q_n$$

on any contour $\mathcal{C}$, up to a scaling factor. The quadratic form $n^{-1}Q_n^T\mathbf{Y}^T(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}I_p)^{-1}\mathbf{Y}Q_n$ is then shown to concentrate around $[\Theta(\mathbb{z},\gamma)-1]I_q$, which consequently serves as the integral kernel in $\Omega(f,\gamma)$. The kernel $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$ of $\Delta(f,\gamma)$ is the limit of $\mathbb{E}[n^{-1}\mathrm{tr}\{(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}_1I_p)^{-1}\Sigma_p(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}_2I_p)^{-1}\Sigma_p\}]$.

Remark 2.1 *Two sufficient conditions for $\Delta(f,\gamma)>0$ are*
(1) $f(x)>0$ for $x\in\mathcal{X}$;
(2) $f(x)\geqslant 0$ for $x\in\mathcal{X}$, with $f(x)\neq 0$ for some $x\in\mathcal{X}$, and $\liminf \lambda_{\min}(\Sigma_p)>0$.

It would be convenient if $\Omega(f,\gamma)$ and $\Delta(f,\gamma)$ had closed forms in order to avoid computational inefficiencies. Closed forms are available for special cases as shown in the following lemma.

Lemma 2.1 *When $f(x,\ell)=(x-\ell)^{-1}$ with $\ell\in\mathbb{R}^-$, the contour integrals in Theorem 2.1 have closed forms, namely, for j, j_1, $j_2=0,1,2,\ldots$,*

$$\frac{-1}{2\pi i}\oint_{\mathcal{C}}\frac{\partial^j f(\mathbb{z},\ell)}{\partial\ell^j}(\Theta(\mathbb{z},\gamma)-1)d\mathbb{z}=\frac{\partial^j(\Theta(\ell,\gamma)-1)}{\partial\ell^j},$$

$$\frac{1}{(2\pi i)^2}\oiint_{\mathcal{C}^2}\frac{\partial^{j_1}f(\mathbb{z}_1,\ell_1)}{\partial\ell_1^{j_1}}\frac{\partial^{j_2}f(\mathbb{z}_2,\ell_2)}{\partial\ell_2^{j_2}}\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)d\mathbb{z}_1d\mathbb{z}_2=\frac{\partial^{j_1+j_2}\delta(\ell_1,\ell_2,\gamma)}{\partial\ell_1^{j_1}\partial\ell_2^{j_2}}.$$

The results continue to hold when $\ell\in\mathbb{C}\backslash\mathcal{X}$.

Lemma 2.1 indicates that it is possible to have convenient and accurate estimators of the asymptotic mean and variance of $\mathbf{M}(f)$ under ridge-regularization. The result easily generalizes to the setting when $f(x)$ is a linear combination of functions of the form $(x-\ell_j)^{-1}$, for any finite collection of ℓ_j's. We elaborate on this in Section 4.

To conduct the tests, consistent estimators of $\Omega(f,\gamma)$ and $\Delta(f,\gamma)$ are needed.

Lemma 2.2 *Let $\widehat{\Theta}(\mathbb{z},\gamma_n)$ and $\widehat{\delta}(\mathbb{z}_1,\mathbb{z}_2,\gamma_n)$ be the plug-in estimators of $\Theta(\mathbb{z},\gamma)$ and $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$, with $(m(\mathbb{z}),\gamma)$ estimated by $(m_{n,p}(\mathbb{z}),\gamma_n)$. For general f, f_1, f_2, we can estimate $\Omega(f,\gamma)$ and $\Delta(f_1,f_2,\gamma)$ by replacing $\Theta(\mathbb{z},\gamma)$ and $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$ with $\widehat{\Theta}(\mathbb{z},\gamma_n)$ and $\widehat{\delta}(\mathbb{z}_1,\mathbb{z}_2,\gamma_n)$. Denote the resulting estimators by $\widehat{\Omega}(f,\gamma_n)$ and $\widehat{\Delta}(f_1,f_2,\gamma_n)$. Then,*

$$\sqrt{n}|\widehat{\Omega}(f,\gamma_n)-\Omega(f,\gamma)|\xrightarrow{P}0,$$

$$\sqrt{n}|\widehat{\Delta}(f_1,f_2,\gamma_n)-\Delta(f_1,f_2,\gamma)|\xrightarrow{P}0,$$

where $\xrightarrow{P}$ indicates convergence in probability. Again, we write $\widehat{\Delta}(f,f,\gamma_n)$ as $\widehat{\Delta}(f,\gamma_n)$.

For the special case of $f^{(j)}(x,\ell) = \partial^j (x-\ell)^{-1}/\partial \ell^j$, $j = 0,1,2,\dots$ and $\ell \in \mathbb{C}\backslash\mathcal{X}$, using Lemma 2.1, natural estimators in closed forms are

$$\widehat{\Omega}(f^{(j)}(x,\ell),\gamma_n) = \frac{\partial^j(\widehat{\Theta}(\ell,\gamma_n)-1)}{\partial \ell^j},$$
$$\widehat{\Delta}(f^{(j_1)}(x,\ell_1), f^{(j_2)}(x,\ell_2),\gamma_n) = \frac{\partial^{j_1+j_2} 2\widehat{\delta}(\ell_1,\ell_2,\gamma_n)}{\partial \ell_1^{j_1}\partial \ell_2^{j_2}}.$$

In particular, for $j, j_1, j_2 = 0$,

$$\widehat{\Omega}(f(x,\ell),\gamma_n) = \widehat{\Theta}(\ell,\gamma_n) - 1,$$
$$\widehat{\Delta}(f(x,\ell_1), f(x,\ell_2),\gamma_n) = 2\widehat{\delta}(\ell_1,\ell_2,\gamma_n).$$

The estimators are consistent, for any fixed j and ℓ. Given the eigenvalues of $\widehat{\mathbf{\Sigma}}_p$, the computational complexity of calculating the above estimators is $O(p)$.

Recall the definitions of $T^{\mathrm{LR}}(f), T^{\mathrm{LH}}(f)$ and $T^{\mathrm{BNP}}(f)$ from Section 1.

Theorem 2.2 *Suppose **C1**–**C6** hold and $\Delta(f,\gamma) > 0$. Under the null hypothesis H_0: $BC = 0$,*

$$\widehat{T}^{\mathrm{LR}}(f) := \frac{\sqrt{n}\{1+\widehat{\Omega}(f,\gamma_n)\}}{q^{1/2}\widehat{\Delta}^{1/2}(f,\gamma_n)}[T^{\mathrm{LR}}(f) - q\log\{1+\widehat{\Omega}(f,\gamma_n)\}] \Longrightarrow \mathcal{N}(0,1),$$
$$\widehat{T}^{\mathrm{LH}}(f) := \frac{\sqrt{n}}{q^{1/2}\widehat{\Delta}^{1/2}(f,\gamma_n)}\{T^{\mathrm{LH}}(f) - q\widehat{\Omega}(f,\gamma_n)\} \Longrightarrow \mathcal{N}(0,1),$$
$$\widehat{T}^{\mathrm{BNP}}(f) := \frac{\sqrt{n}\{1+\widehat{\Omega}(f,\gamma_n)\}^2}{q^{1/2}\widehat{\Delta}^{1/2}(f,\gamma_n)}\Big\{T^{\mathrm{BNP}}(f) - q\frac{\widehat{\Omega}(f,\gamma_n)}{1+\widehat{\Omega}(f,\gamma_n)}\Big\} \Longrightarrow \mathcal{N}(0,1).$$

For any of the three tests, the null hypothesis is rejected at asymptotic level α, if $\widehat{T}(f) > \xi_\alpha$, where ξ_α is the $1-\alpha$ quantile of the standard normal distribution.

2.3. Asymptotic power under local alternatives

This subsection deals with the behavior of the proposed family of tests under a host of local alternatives. We start with deterministic alternatives, a framework commonly used in the literature to study the asymptotic power of inferential procedures. Next, we consider a Bayesian framework, using a class of priors that characterize the structure of the alternatives. Because the results to follow simultaneously hold for $\widehat{T}^{\mathrm{LR}}(f)$, $\widehat{T}^{\mathrm{LH}}(f)$ and $\widehat{T}^{\mathrm{BNP}}(f)$, the unifying notation $\widehat{T}(f)$ will be used to refer to each of the test statistics.

2.3.1. Deterministic local alternatives

Consider a sequence of BC such that, as $n, p \to \infty$,

$$\sqrt{n}C^T B^T\{\Theta^{-1}(z,\gamma)\Sigma_p - zI\}^{-1}BC \longrightarrow D(z,\gamma) \qquad \text{pointwise}, \tag{2.3}$$

on an open subset of $\mathbb{C}$ containing $\mathcal{X}$.

Observe that $\mathbf{Y}Q_n = BC[C^T(n^{-1}XX^T)^{-1}C]^{-1/2} + \Sigma_p^{1/2}\mathbf{Z}Q_n$ and define

$$\mathcal{H}(D,f) = T^{-1/2}\Big[\frac{-1}{2\pi i}\oint_{\mathcal{C}} f(\mathbb{z})D(\mathbb{z},\gamma)d\mathbb{z}\Big]T^{-1/2}, \tag{2.4}$$

where

$$T = \lim_{n\to\infty} C^T(n^{-1}XX^T)^{-1}C. \tag{2.5}$$

Note that T exists and is non-singular under **C5** and **C6**. If further $f(x) \geqslant 0$ for any $x \in \mathcal{X}$, $\mathcal{H}(D,f)$ is non-negative definite.

Theorem 2.3 *Suppose **C1**–**C6** and (2.3) hold, and $\Delta(f,\gamma) > 0$. Then, as $n \to \infty$,*

$$\frac{\sqrt{n}}{\Delta^{1/2}(f,\gamma)}\{\mathbf{M}(f) - \Omega(f,\gamma)I_q\} \Longrightarrow \mathbf{W} + \frac{\mathcal{H}(D,f)}{\Delta^{1/2}(f,\gamma)}.$$

Denote the power functions of $\widehat{T}(f)$ at asymptotic level α, conditional on BC, by

$$\Upsilon(BC,f) = \mathbb{P}(\widehat{T}(f) > \xi_\alpha \mid BC).$$

The asymptotic behavior of the power functions is described in the following corollary.

Corollary 2.2 *Under the assumptions of Theorem 2.3, as $n \to \infty$,*

$$\Upsilon(BC,f) \longrightarrow \Phi\Big(-\xi_\alpha + \frac{\mathrm{tr}(\mathcal{H}(D,f))}{q^{1/2}\Delta^{1/2}(f,\gamma)}\Big),$$

where Φ is the standard normal CDF.

Remark 2.2 *Corollary 2.2 indicates the three proposed statistics have identical asymptotic powers under the assumed local alternatives. This is because the first-order Taylor expansions of x, $\log(1+x)$ and $x/(1+x)$ coincide at 0. However, the respective empirical powers may differ considerably for moderate sample sizes.*

The following remark provides a sufficient condition under which (2.3) is satisfied. Denoting the columns of BC by $[\mu_1,\dots,\mu_q]$, it follows that

$$\sqrt{n}C^TB^T\{\Theta^{-1}(\mathbb{z},\gamma)\Sigma_p - \mathbb{z}I\}^{-1}BC = \sqrt{n}\Big[\mu_i^T\{\Theta^{-1}(\mathbb{z},\gamma)\Sigma_p - \mathbb{z}I_p\}^{-1}\mu_j\Big]_{i,j=1}^q.$$

Remark 2.3 *(a) Let $\mathbf{E}_{m,p}$ denote the eigen-projection associated with $\lambda_{m,p} = \lambda_m(\Sigma_p)$. Suppose that there exists a sequence (in p) of mappings $[\mathfrak{B}_{ij;p}]_{i,j=1}^q$ from $[0,\infty)^{q^2}$ to $[0,\infty)^{q^2}$, satisfying $\mathfrak{B}_{ij;p}(\lambda_{m,p}) = \sqrt{n}p\mu_i^T\mathbf{E}_{m,p}\mu_j$, $m = 1,\dots,p$, and a mapping $[\mathfrak{B}_{ij;\infty}]_{i,j=1}^q$ continuous on $[0,\infty)^{q^2}$ such that, as $p \to \infty$ and for $1 \leqslant i,j \leqslant q$,*

$$\int |\mathfrak{B}_{ij;p}(x) - \mathfrak{B}_{ij;\infty}(x)|dF^{\Sigma_p}(x) \to 0.$$

Then, under ***C4****, it follows that (2.3) holds with* $D(\mathbb{z},\gamma) = [d_{ij}(\mathbb{z},\gamma)]_{i,j=1}^q$ *and*

$$d_{ij}(\mathbb{z},\gamma) = \int \frac{\mathfrak{B}_{ij;\infty}(x)dL^\Sigma(x)}{x\Theta^{-1}(\mathbb{z},\gamma) - \mathbb{z}} = \int \frac{\mathfrak{B}_{ij;\infty}(x)dL^\Sigma(x)}{x\{1-\gamma-\gamma\mathbb{z}m(\mathbb{z})\} - \mathbb{z}}.$$

(b) If $\Sigma_p = I_p$*, then (2.3) is satisfied if* $\sqrt{n}\mu_i^T\mu_j \to \mathcal{K}_{ij}$*, for some constants* $\mathcal{K}_{ij}$*,* $1 \leqslant i,j \leqslant q$*. In this case,* $D(\mathbb{z},\gamma) = (\Theta^{-1}(\mathbb{z},\gamma) - \mathbb{z})^{-1}[\mathcal{K}_{ij}]_{i,j=1}^q$*.*

2.3.2. Probabilistic local alternatives

While deterministic local alternatives provide useful information, they are somewhat restrictive for the purpose of a systematic investigation of the power characteristics. Therefore, probabilistic alternatives are considered in the form of a sequence of prior distributions for BC. This has the added advantage of providing flexibility for incorporating structural information about the regression parameters and the constraints matrices. The proposed formulation of probabilistic alternatives can be seen as an extension of the proposal adopted by Li et al. (2016) in the context of two-sample tests for equality of means. One challenge associated with formulating meaningful alternatives to the hypothesis (1.2), when compared to the two-sample testing problem, is that there are many more plausible ways in which the null hypothesis can be violated. Considering this, we propose a class of alternatives, that on one hand can incorporate a multitude of structures of the parameter BC, while on the other hand retains analytical tractability in terms of providing interpretable expressions for the local asymptotic power.

Assume the following prior model of BC with separable covariance

$$BC = n^{-1/4}p^{-1/2}\mathcal{R}\boldsymbol{\mathcal{V}}\mathcal{S}^T, \tag{2.6}$$

where $\boldsymbol{\mathcal{V}}$ is a $p \times m$ stochastic matrix ($m \geqslant 1$ fixed) with independent elements ν_{ij} such that $\mathbb{E}[\nu_{ij}] = 0$, $\mathbb{E}[|\nu_{ij}|^2] = 1$ and $\max_{ij}\mathbb{E}[|\nu_{ij}|^4] \leqslant p^{c_\nu}$ for some $c_\nu \in (0,1)$; $\mathcal{R}$ is a $p \times p$ deterministic matrix and $\mathcal{S}$ is a fixed $q \times m$ matrix. Moreover, let $\|\mathcal{R}\|_2 \leqslant \mathcal{K}_1 < \infty$ and suppose there is a nonrandom function $h(\mathbb{z},\gamma)$ such that, as $p \to \infty$, on an open subset of $\mathbb{C}$ containing $\mathcal{X}$,

$$p^{-1}\mathrm{tr}\{(\Theta^{-1}(\mathbb{z},\gamma)\Sigma_p - \mathbb{z}I)^{-1}\mathcal{R}\mathcal{R}^T\} \to h(\mathbb{z},\gamma) \qquad \text{pointwise.} \tag{2.7}$$

Recalling that $(\Theta^{-1}(\mathbb{z},\gamma)\Sigma_p - \mathbb{z}I)^{-1}$ is the deterministic equivalent of the resolvent $(\widehat{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}$, existence of the limit (2.7) also implies that $p^{-1}\mathrm{tr}\{(\widehat{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\mathcal{R}\mathcal{R}^T\}$ converges pointwise in probability to $h(\mathbb{z},\gamma)$. Notice also that $p^{-1}\mathrm{tr}\{(\widehat{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\mathcal{R}\mathcal{R}^T\}$ is the Stieltjes transform of a measure supported on the eigenvalues of $\widehat{\boldsymbol{\Sigma}}_p$.

Model (2.6) leads to a fairly broad covariance design for multi-dimensional random elements, encompassing structures commonly encountered in many application domains, especially in spatio-temporal statistics. We give some representative examples by considering various functional forms of the matrix $\mathcal{S}$. Denote by μ_j the columns of BC and by V_j the columns of $\boldsymbol{\mathcal{V}}$.

Example 2.1 In all that follows j takes values in $1,\dots,q$.

(a) *Independent*: $\mu_j = n^{-1/4}p^{-1/2}\mathcal{R}V_j$;
(b) *Longitudinal*: $\mu_j = n^{-1/4}p^{-1/2}\mathcal{R}(V_1 + V_2 j + \cdots + V_m j^{m-1})$;
(c) *Moving average*: $\mu_j = n^{-1/4}p^{-1/2}\mathcal{R}[V_{j+t} + \theta_1 V_{j+t-1} + \cdots + \theta_t V_j]$ for constants $\theta_1, \ldots, \theta_t$.

Taking the MANOVA problem to illustrate, suppose that the columns of B represent group mean vectors, and suppose C is the matrix that determines successive contrasts among them. Then, μ_j is the difference between the means of group j and group $j+1$. Parts (a)–(c) of Example 2.1 correspond then to $\mu_1, \ldots, \mu_q$ respectively following an independent, a longitudinal and a moving average process. The row-wise covariance structure is assumed to be such that each μ_j has a covariance matrix proportional to $n^{-1/2}p^{-1}\mathcal{R}\mathcal{R}^T$. The factor $n^{-1/2}p^{-1}$ provides the scaling for the tests to have non-trivial local power.

A sufficient condition that leads to (2.7), similar to Remark 2.3, is to postulate the existence of functions $\tilde{\mathfrak{B}}_p$ satisfying $\tilde{\mathfrak{B}}_p(\lambda_{j,p}) = \mathrm{tr}\{\mathbf{E}_{j,p}\mathcal{R}\mathcal{R}^T\}$, $j = 1, \ldots, p$, and

$$\int |\tilde{\mathfrak{B}}_p(x) - \tilde{\mathfrak{B}}_\infty(x)| dF^{\Sigma_p}(x) \to 0$$

for some function $\tilde{\mathfrak{B}}_\infty$ continuous on $[0, \infty)$, where $\lambda_{j,p}$ is the jth eigenvalue of Σ_p and $\mathbf{E}_{j,p}$ is the eigen-projection associated with $\lambda_{j,p}$. Then

$$h(\mathbb{z}, \gamma) = \int \frac{\tilde{\mathfrak{B}}_\infty(x) dL^\Sigma(x)}{x\{1 - \gamma - \gamma \mathbb{z} m(\mathbb{z})\} - \mathbb{z}}. \tag{2.8}$$

Equations (2.7) and (2.8) indicate that $h(\mathbb{z}, \gamma)$ effectively captures the distribution of the total spectral mass of $\mathcal{R}\mathcal{R}^T$ across the spectral coordinates of Σ_p, also taking into account the dimensionality effect through the aspect ratio γ. Later, we shall discuss specific classes of the matrices $\mathcal{R}$ that lead to analytically tractable expressions for $h(\mathbb{z}, \gamma)$, with the structure of $\mathcal{R}$ linking the parameter BC under the alternative through (2.6) to the structure of Σ_p.

Another important feature of the probabilistic model is that it incorporates both dense and sparse alternatives through different specifications of the innovation variables ν_{ij}. We consider two special cases.

1. *Dense alternative:* $\nu_{ij} \sim \mathcal{N}(0, 1)$;
2. *Sparse alternative:* $\nu_{ij} \sim G_\eta$, for some $\eta \in (0, 1)$, where G_η is the discrete probability distribution assigning mass $1 - p^{-\eta}$ to 0 and mass $(1/2)p^{-\eta}$ to the points $\pm p^{\eta/2}$.

Note that the usual notion of sparsity corresponds to the setting where in addition, $\mathcal{R} = I_p$. More generally, the second specification above formulates a prior model for BC that is sparse in the coordinate system determined by $\mathcal{R}$. In particular, if $\mathcal{R}\mathcal{R}^T$ is a polynomial in Σ_p (see Section 3.2 for a discussion), BC can be seen as sparse in the spectral coordinates of Σ_p.

Theorem 2.4 *Suppose that **C1**–**C6** hold and $\Delta(f, \gamma) > 0$. Also suppose that, under H_a, BC has a prior distribution given by (2.6). Then, the power function of each of the three* *test statistics satisfies*

$$\Upsilon(BC, f) \xrightarrow{L_1} \Phi\Big(-\xi_\alpha + \frac{\mathrm{tr}(\mathcal{S}\mathcal{S}^T T^{-1})}{q^{1/2}\Delta^{1/2}(f,\gamma)} \oint_{\mathcal{C}} \frac{-1}{2\pi i} f(\mathbb{z}) h(\mathbb{z},\gamma) d\mathbb{z} \Big), \tag{2.9}$$

as $n \to \infty$, where T is as in (2.5) and $\xrightarrow{L_1}$ indicates L_1-convergence (with respect to the prior measure of BC).

Remark 2.4 *Even if the quantity $h_p(\mathbb{z},\gamma) = p^{-1}\mathrm{tr}\{(\Theta^{-1}(\mathbb{z},\gamma)\Sigma_p - \mathbb{z}I)^{-1}\mathcal{R}\mathcal{R}^T\}$ does not converge, it can be verified that the difference between the left- and right-hand sides of (2.9) still converges to zero in L_1 if $h(\mathbb{z},\gamma)$ is replaced by $h_p(\mathbb{z},\gamma)$.*

Observe that the matrices $\mathcal{R}$ and $\mathcal{S}$ decouple in the expression (2.9) for the asymptotic power. Dependence on the unknown error covariance matrix Σ_p, besides $\Delta^{1/2}(f,\gamma)$, is only through the function $h(\mathbb{z},\nu)$, which incorporates the structure of the matrix $\mathcal{R}\mathcal{R}^T$. It is also noticeable that distributional characteristics of the variables ν_{ij} do not affect the asymptotic power. Indeed, the proposed tests have the same local asymptotic power under both sparse and dense alternatives.

3. Data-driven selection of shrinkage

In this section, we introduce a data-driven procedure to select the "optimal" f from a parametric family $\mathfrak{F}$ of shrinkage functions. The strategy is to maximize the local power function $\Upsilon(BC, f)$ over f, given a class of probabilistic local alternatives as in (2.6). In designing the classes of alternatives, we focus our attention only on the specification of $\mathcal{R}$. This is because, as the expression (2.9) shows, the dependence on the matrix $\mathcal{S}$ is only through a multiplier involving a "known" matrix T, while the effect of the unknown covariance Σ_p (and its interaction with $\mathcal{R}$) manifests itself through the function $h(\mathbb{z},\gamma)$. Another reason for focusing on $\mathcal{R}$ is that the choice of $\mathcal{S}$ is closely related to the specific type of linear model being considered, while the choice of $\mathcal{R}$ is associated with the structure of the error distribution.

We present some settings of BC for which $h(\mathbb{z},\gamma)$ can be computed explicitly. We also verify that the standardized test statistic with the data-driven selection of f is still asymptotically standard normal under suitable conditions. Hence, the Type 1 error rate of the tests is asymptotically not inflated, although the same data is used for both shrinkage selection and testing. Lastly, we present a composite test procedure that combines the optimal tests corresponding to different prior models of BC and thereby improves adaptivity to various kinds of alternatives.

3.1. Shrinkage family

Suppose the family of shrinkage functions is such that

$$\mathfrak{F} = \{f_\ell \colon \ell \in \mathcal{L}\},$$

(i) $\mathcal{L}$ is a compact subset of $\mathbb{R}^r$, $r \in \mathbb{N}^+$;
(ii) There is a closed, connected subset $\mathcal{Z}$ of $\mathbb{C}$ such that $\mathcal{X} = [0, \ \limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2] \subset \mathcal{Z}$, and the third-order partial derivatives of f_ℓ with respect to ℓ are continuous on $\mathcal{L} \otimes \mathcal{Z}$;
(iii) The gradient $\nabla_\ell f_\ell$ and the Hessian $\nabla_\ell^2 f_\ell$ of f_ℓ with respect to ℓ have analytic extensions to $\mathcal{Z}$ for all $\ell \in \mathcal{L}$;
(iv) $\inf_{\ell \in \mathcal{L}} \Delta(f_\ell, \gamma) > 0$.

Under the probabilistic prior model (2.6) with $h(\mathbb{z}, \gamma)$ in (2.7) given, define

$$\Xi(\ell, h, \gamma) = \frac{-1}{2\pi i \Delta^{1/2}(f_\ell, \gamma)} \oint_{\mathcal{C}} f_\ell(\mathbb{z}) h(\mathbb{z}, \gamma) d\mathbb{z}.$$

Theorem 2.4 suggests that ℓ should be chosen such that $\Xi(\ell, h, \gamma)$ is maximized, that is,

$$\ell_{opt} = \arg\max_{\ell \in \mathcal{L}} \Xi(\ell, h, \gamma).$$

The test with the selected shrinkage will then be the locally most powerful test under the alternatives specified by (2.6) and (2.7) for any given choice of $\mathcal{S}$. Since $\Xi(\ell, h, \gamma)$ is continuous with respect to ℓ under condition (i)–(iv), ℓ_{opt} exists. Importantly, $\Xi(\ell, h, \gamma)$ does not rely on $\mathcal{S}$. In other words, different column-wise covariance structures of BC are uniform in terms of selecting the optimal shrinkage. This significantly simplifies the selection procedure.

Recall that $h(\mathbb{z}, \ell)$ is the limit of $p^{-1}\mathrm{tr}\{(\Theta^{-1}(\mathbb{z}, \gamma)\Sigma_p - \mathbb{z}I)^{-1}\mathcal{R}\mathcal{R}^T\}$. We next present two possible settings of $\mathcal{R}\mathcal{R}^T$ under which $h(\mathbb{z}, \gamma)$ and consequently $\Xi(\ell, h, \gamma)$ can be accurately estimated:

(1) Suppose $\mathcal{R}\mathcal{R}^T$ is specified. Then, $h(\mathbb{z}, \gamma)$ is estimated by $\widehat{h}(\mathbb{z}, \gamma_n) = p^{-1}\mathrm{tr}\{(\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1}\mathcal{R}\mathcal{R}^T\}$ and

$$\widehat{\Xi}(\ell, \widehat{h}, \gamma_n) := \frac{-1}{2\pi i \widehat{\Delta}^{1/2}(f_\ell, \gamma_n)} \oint_{\mathcal{C}} f_\ell(\mathbb{z}) \widehat{h}(\mathbb{z}, \gamma_n) d\mathbb{z}$$

is a consistent estimator of $\Xi(f, h, \gamma)$. As an example of this scenario, assume that the p components of μ_j admit a natural ordering such that the dependence between their coordinates is a function of the difference between their indexes. Then we may set $\mathcal{R}\mathcal{R}^T$ to be a Toeplitz matrix (stationary auto-covariance structure).

(2) Only the spectral mass distribution of $\mathcal{R}\mathcal{R}^T$ in the form of $\tilde{\mathfrak{B}}_\infty$ described in (2.8) is specified.

The remainder of this section is devoted to dealing with the second scenario.

3.2. Polynomial alternatives

Even if $\tilde{\mathfrak{B}}_\infty$ is given, the estimation of $h(\mathbb{z}, \gamma)$ is still challenging since it involves the unknown limiting population spectral distribution L^Σ. In order to estimate $h(\mathbb{z}, \gamma)$, it is convenient to have it in a closed form. This is feasible if $\tilde{\mathfrak{B}}_\infty$ is a polynomial, which is true if $\mathcal{R}\mathcal{R}^T$ is a matrix polynomial in Σ_p. Since any smooth function can be approximated by polynomials, this formulation is quite flexible and practically beneficial. Assume therefore that

$$\mathcal{R}\mathcal{R}^T = \sum_{j=0}^{s} t_j \Sigma_p^j, \tag{3.1}$$

where $t_0, \ldots, t_s$ are pre-specified weights such that $\sum_{j=0}^{s} t_j \Sigma_p^j$ is nonnegative definite. Under the model,

$$h(\mathbb{z}, \gamma) = \lim_{p\to\infty} \frac{1}{p}\mathrm{tr}\Big[(\Theta^{-1}(\mathbb{z}, \gamma)\Sigma_p - \mathbb{z}I)^{-1} \sum_{j=0}^{s} t_j \Sigma_p^j\Big] = \sum_{j=0}^{s} t_j \rho_j(\mathbb{z}, \gamma),$$

where the functions $\rho_j(\mathbb{z}, \gamma)$ satisfy the recursive formula (see Ledoit and Péché, 2011)

$$\rho_0(\mathbb{z}, \gamma) = m(\mathbb{z}), \qquad \rho_{j+1}(\mathbb{z}, \gamma) = \Theta(\mathbb{z}, \gamma)\Big[\int x^j dL^\Sigma(x) + \mathbb{z}\rho_j(\mathbb{z}, \gamma)\Big].$$

For any $j \in \mathbb{N}$, $\int x^j dL^\Sigma(x)$, and consequently $\rho_j(\mathbb{z}, \gamma)$, can be estimated consistently (Bai, Chen and Yao, 2010, Lemma 1). Specifically, $p^{-1}\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_p)$ is a consistent estimator of $\int x dL^\Sigma(x)$.

In practice, we restrict to the case $s = 2$. There are several considerations that guided this choice of s as stated in Li et al. (2016). First, for $s = 2$, all quantities involved can be computed explicitly without requiring knowledge of higher-order moments of the observations. Also, the corresponding estimating equations for $h(\mathbb{z}, \gamma)$ are more stable as they do not involve higher-order spectral moments. Second, the choice of $s = 2$ yields a significant, yet nontrivial, concentration of the prior covariance of μ_j, $j = 1, \ldots, q$, (that is $\mathcal{R}\mathcal{R}^T$ up to a scaling factor) in the directions of the leading eigenvectors of Σ_p. Finally, the choice $s = 2$ allows for both convex and concave shapes of the spectral mass distribution $\tilde{\mathfrak{B}}_\infty$ since the latter becomes a quadratic function.

With $s = 2$, we estimate $\rho_0(\mathbb{z}, \gamma)$, $\rho_1(\mathbb{z}, \gamma)$, and $\rho_2(\mathbb{z}, \gamma)$ by

$$\begin{aligned} \widehat{\rho}_0(\mathbb{z}, \gamma_n) &= m_{n,p}(\mathbb{z}), \\ \widehat{\rho}_1(\mathbb{z}, \gamma_n) &= \widehat{\Theta}(\mathbb{z}, \gamma_n)[1 + \mathbb{z}m_{n,p}(\mathbb{z})], \\ \widehat{\rho}_2(\mathbb{z}, \gamma_n) &= \widehat{\Theta}(\mathbb{z}, \gamma_n)\big[p^{-1}\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_p) + \mathbb{z}\widehat{\rho}_1(\mathbb{z}, \gamma_n)\big] \end{aligned} \tag{3.2}$$

and $h(\mathbb{z}, \gamma)$ by

$$\widehat{h}(\mathbb{z}, \gamma_n) = \sum_{j=0}^{2} t_j \widehat{\rho}_j(\mathbb{z}, \gamma_n).$$

The algorithm for the data-driven shrinkage selection is stated next.

Algorithm 3.1 (Data-driven shrinkage selection)

1. *Specify prior weights* $\tilde{t} = (t_0, t_1, t_2)$*. The canonical choices are* $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$*;*

2. *Compute* $\widehat{h}(\mathbb{z},\gamma_n)=\sum_{j=0}^{2}t_j\widehat{\rho}_j(\mathbb{z},\gamma_n)$*;*
3. *For any* $\ell\in\mathcal{L}$*, numerically compute the integral*

$$\widehat{\Xi}(\ell,\widehat{h},\gamma_n)=\frac{-1}{2\pi i\widehat{\Delta}^{1/2}(f_\ell,\gamma_n)}\oint_{\mathcal{C}}f_\ell(\mathbb{z})\widehat{h}(\mathbb{z},\gamma_n)d\mathbb{z};$$

4. *Select* $\ell_{opt}(\tilde{t})=arg\max_{\ell\in\mathcal{L}}\widehat{\Xi}(\ell,\widehat{h},\gamma_n)$*.*

The behavior of the tests applied with the data-driven shrinkage selection is described in the following theorem.

Theorem 3.1 *Suppose **C1**–**C6** hold and* $\mathfrak{F}$ *satisfies conditions (i)–(iv). Then,*

(1) $\sup_{\ell\in\mathcal{L}}\sqrt{n}|\widehat{\Xi}(\ell,\widehat{h},\gamma_n)-\Xi(\ell,h,\gamma)|\xrightarrow{P}0$ *as* $n\to\infty$*.*

(2) Let ℓ^* *be any local maximizer of* $\Xi(\ell,h,\gamma)$ *in the interior of* $\mathcal{L}$*. Assume there exists a neighborhood of* ℓ^* *such that for all feasible points* $\ell\in\mathcal{L}$ *within the neighborhood, there exists a constant* $\mathcal{K}>0$ *such that*

$$\Xi(\ell,h,\gamma)-\Xi(\ell^*,h,\gamma)\leqslant-\mathcal{K}\|\ell-\ell^*\|_2^2. \tag{3.3}$$

Then, there exists a sequence $(\ell_n^*\colon n\in\mathbb{N})$ *of local maximizers of* $(\widehat{\Xi}(\ell,\widehat{h},\gamma_n)\colon n\in\mathbb{N})$ *satisfying*

$$n^{1/4}\|\ell_n^*-\ell^*\|_2=O_p(1)\qquad(n\to\infty). \tag{3.4}$$

Further, recalling notation in Section 2, under the null hypothesis,

$$\frac{\sqrt{n}}{\widehat{\Delta}^{1/2}(f_{\ell_n^*},\gamma_n)}\{\mathbf{M}(f_{\ell_n^*})-\widehat{\Omega}(f_{\ell_n^*},\gamma_n)I_q\}\Longrightarrow\mathbf{W}. \tag{3.5}$$

(3) Let ℓ^* *be any local maximizer of* $\Xi(\ell,h,\gamma)$ *on the boundary of* $\mathcal{L}$*. Assume there exists a neighborhood of* ℓ^* *such that for all feasible points* $\ell\in\mathcal{L}$ *within the neighborhood, there is a constant* $\mathcal{K}'>0$ *satisfying*

$$\Xi(\ell,h,\gamma)-\Xi(\ell^*,h,\gamma)\leqslant-\mathcal{K}'\|\ell-\ell^*\|_2. \tag{3.6}$$

Then, (3.4) and (3.5) still hold.

The two conditions (3.3) and (3.6) ensure that the parameter ℓ^* is locally identifiable in a neighborhood of ℓ^*. In general, the two conditions depend on the structure of L^{Σ}.

3.3. Combination of prior models

An extensive simulation analysis revealed that there is considerable variation in the shape of the power functions and the values of $\tilde{t}=(t_0,t_1,t_2)$, especially when the condition number of Σ_p is relatively large. In this subsection, we consider a convenient collection of priors that are representative of certain structural scenarios. A composite test, called

$\widehat{T}_{\max}$, is defined as the maximum of the standardized statistics $\widehat{T}(f_{\ell_i^*})$ where ℓ_i^* is obtained from Algorithm 3.1 under prior $\tilde{t}_i$, $i = 1, \ldots, m$. The following strategy is applicable to LR, LH and BNP. We therefore continue to use $\widehat{T}(f)$ to denote the general test statistic. In summary, we propose to test the hypothesis by rejecting for large values of the statistic

$$\widehat{T}_{\max} = \max_{\tilde{t} \in \widetilde{\Pi}} \widehat{T}(f_{\ell_i^*}),$$

where $\widetilde{\Pi} = \{\tilde{t}_1, \ldots, \tilde{t}_m\}$, $m \geqslant 1$, is a pre-specified finite class of weights. A simple but effective choice of $\widetilde{\Pi}$ consists of the three canonical weights $\tilde{t}_1 = (1,0,0)$, $\tilde{t}_2 = (0,1,0)$, $\tilde{t}_3 = (0,0,1)$.

Theorem 3.2 *Suppose **C1**–**C6** hold and $\mathfrak{F}$ satisfies condition (i)–(iv). For each $i = 1, \ldots, m$, assume that ℓ_{in}^* is a sequence of local maximizers of the empirical power function $\widehat{\Xi}(\ell, \widehat{h}, \gamma_n)$ under prior model with weight $\tilde{t}_i$ such that*

$$n^{1/4} \|\ell_{in}^* - \ell_i^*\|_2 = O_p(1).$$

(See (3.4)). Then, under the null hypothesis $H_0\colon BC = 0$,

$$\big(\widehat{T}(f_{\ell_{1n}^*}), \ldots, \widehat{T}(f_{\ell_{mn}^*})\big) \Longrightarrow \mathcal{N}\big(0, \Delta_*\big),$$

where Δ_* is an $m \times m$ matrix with diagonal entries 1 and (i,j)-th off-diagonal entry

$$\Delta^{-1/2}(f_{\ell_i^*}, \gamma) \Delta(f_{\ell_i^*}, f_{\ell_j^*}, \gamma) \Delta^{-1/2}(f_{\ell_j^*}, \gamma).$$

Theorem 3.2 shows that $\widehat{T}_{\max}$ has a non-degenerate limiting distribution under H_0. It is worth mentioning that LR, LH and BNP share the covariance matrix Δ_*. Theorem 3.2 can be used to determine the cut-off values of the test by deriving analytical formulas for the quantiles of the limiting distribution. Aiming to avoid complex calculations, a parametric bootstrap procedure is applied to approximate the cut-off values. Specifically, Δ_* is first estimated by $\widehat{\Delta}_*$, and then bootstrap replicates are generated by simulating from $\mathcal{N}(0, \widehat{\Delta}^*)$, thereby providing an approximation of the null distribution of $\widehat{T}_{\max}$. Replacing $\Delta(f_{\ell_i^*}, f_{\ell_j^*}, \gamma)$ with $\widehat{\Delta}(f_{\ell_i}, f_{\ell_j}, \gamma_n)$ yields the natural estimator.

Remark 3.1 *Observe that $\widehat{\Delta}_*$ defined above may not be nonnegative definite even though it is symmetric. If such a case occurs, the resulting estimator can be projected onto its closest non-negative definite matrix simply by setting the negative eigenvalues to zero. This covariance matrix estimator is denoted by $\widehat{\Delta}_*^+$ and it is used for generating the bootstraps samples.*

4. Ridge and higher-order regularizers

4.1. Ridge regularization

One of the most commonly used shrinkage procedures in statistics is ridge regularization, corresponding to choosing $f_\ell(x) = 1/(x - \ell)$, $\ell < 0$, so that $f_\ell(\widehat{\mathbf{\Sigma}}_p) = (\widehat{\mathbf{\Sigma}}_p - \ell I_p)^{-1}$. It is an effective way to shift $\widehat{\mathbf{\Sigma}}_p$ away from singularity by adding a ridge term $-\ell I_p$. In this subsection, we apply the results of Sections 2 and 3 using the ridge-shrinkage family

$$\mathfrak{F}_{ridge} := \{f_\ell(x) = (x - \ell)^{-1},\ \ell \in [\underline{\ell},\ \overline{\ell}]\}, \quad -\infty < \underline{\ell} < \overline{\ell} < 0.$$

In the literature, ridge-regularization was applied to high-dimensional one- and two-sample mean tests in Chen, Li and Zhong (2014) and Li et al. (2016). Hence, this subsection is a generalization of their methods to general linear hypotheses.

From the aspect of population covariance estimation, ridge-regularization can be viewed as an order-one estimation where Σ_p is estimated by a weighted average of $\widehat{\mathbf{\Sigma}}_p$ and I_p, namely $\alpha_0 I_p + \alpha_1 \widehat{\mathbf{\Sigma}}_p$. The estimator is equivalent to ridge-regularization with $\ell = -\alpha_0/\alpha_1$ for testing purposes. Within a restricted region of (α_1, α_2), the large eigenvalues of $\widehat{\mathbf{\Sigma}}_p$ are shrunk down and the small ones are lifted upward. It is a desired property since in high-dimensional settings, large sample eigenvalues are systematically biased upward and small sample eigenvalues downwards.

An important advantage of ridge regularization is that the test procedure is computationally efficient due to the fact that $\Omega(f_\ell, \gamma)$ and $\Delta(f_\ell, \gamma)$ admit closed forms as shown in Lemma 2.1. These quantities can be estimated by $\widehat{\Omega}_\ell(\gamma_n) = \widehat{\Theta}(\ell, \gamma_n) - 1$ and $\widehat{\Delta}_\ell(\gamma_n) = 2\widehat{\delta}(\ell, \ell, \gamma_n)$, respectively. A closed-form estimator $\widehat{\Xi}_\ell(\widehat{h}, \gamma_n)$ is then also available for $\Xi(\ell, h, \gamma)$. This leads to the following algorithm.

Algorithm 4.1 (Ridge-regularized test procedure)

1. *Specify prior weights* $\tilde{t} = (t_0, t_1, t_2)$*;*
2. *With* $m_{n,p}(\ell) = p^{-1}\mathrm{tr}(\widehat{\mathbf{\Sigma}}_p - \ell I_p)^{-1}$*, compute, for any* $\ell \in [\underline{\ell}, \overline{\ell}]$*,*

$$\begin{aligned}
\widehat{\Theta}(\ell, \gamma_n) &= \{1 - \gamma_n - \gamma_n \ell m_{n,p}(\ell)\}^{-1}, \\
\widehat{\Omega}_\ell(\gamma_n) &= \widehat{\Theta}(\ell, \gamma_n) - 1, \\
\widehat{\Delta}_\ell(\gamma_n) &= 2\gamma_n\{1 + \ell m_{n,p}(\ell)\}\widehat{\Theta}^3(\ell, \gamma_n) + 2\gamma_n \ell\{m_{n,p}(\ell) + \ell m'_{n,p}(\ell)\}\widehat{\Theta}^4(\ell, \gamma_n);
\end{aligned}$$

3. *For any* $\ell \in [\underline{\ell},\ \overline{\ell}]$*, compute* $\widehat{h}(\ell, \gamma_n) = \sum_{j=0}^{2} t_j \widehat{\rho}_j(\ell, \gamma_n)$ *as defined in (3.2) and*

$$\widehat{\Xi}_\ell(\widehat{h}, \gamma_n) = \frac{\widehat{h}(\ell, \gamma_n)}{\widehat{\Delta}_\ell^{1/2}(\gamma_n)};$$

4. *Select* $\ell^* = arg\max_{\ell \in [\underline{\ell},\ \overline{\ell}]} \widehat{\Xi}_\ell(\widehat{h}, \gamma_n)$*;*
5. *Use one of the standardized statistics*

$$\begin{aligned}
\widehat{T}^{\mathrm{LR}}(\ell^*) &:= \frac{\sqrt{n}\{1 + \widehat{\Omega}_{\ell*}(\gamma_n)\}}{q^{1/2}\widehat{\Delta}_{\ell*}^{1/2}(\gamma_n)}[T^{\mathrm{LR}}(\ell^*) - q\log\{1 + \widehat{\Omega}_{\ell*}(\gamma_n)\}], \\
\widehat{T}^{\mathrm{LH}}(\ell^*) &:= \frac{\sqrt{n}}{q^{1/2}\widehat{\Delta}_{\ell*}^{1/2}(\gamma_n)}[T^{\mathrm{LH}}(\ell^*) - q\widehat{\Omega}_{\ell*}(\gamma_n)], \\
\widehat{T}^{\mathrm{BNP}}(\ell^*) &:= \frac{\sqrt{n}\{1 + \widehat{\Omega}_{\ell*}(\gamma_n)\}^2}{q^{1/2}\widehat{\Delta}_{\ell*}^{1/2}(\gamma_n)}\Big[T^{\mathrm{BNP}}(\ell^*) - \frac{q\widehat{\Omega}_{\ell*}(\gamma_n)}{1 + \widehat{\Omega}_{\ell*}(\gamma_n)}\Big],
\end{aligned}$$

where

$$T^{\mathrm{LR}}(\ell^*) = \sum_{i=1}^{q} \log(1+\lambda_i), \qquad T^{\mathrm{LH}}(\ell^*) = \sum_{i=1}^{q} \lambda_i, \qquad T^{\mathrm{BNP}}(\ell^*) = \sum_{i=1}^{q} \frac{\lambda_i}{1+\lambda_i},$$

and $\lambda_1, \ldots, \lambda_q$ *are the eigenvalues of* $n^{-1} Q_n^T \mathbf{Y}^T (\widehat{\mathbf{\Sigma}}_p - \ell^* I_p)^{-1} \mathbf{Y} Q_n$. *Reject the null at asymptotic level* α *if the test statistic value exceeds* ξ_α.

Although in theory any negative ℓ^* is allowed in the test procedure, in practice, meaningful lower and upper bounds $\underline{\ell}$ and $\overline{\ell}$ are needed to ensure stability of the test statistics when $p \approx n$ or $p > n$ and also to carry out the search for optimal ℓ at a low computational cost. In our simulation settings we use $\overline{\ell} = -p^{-1}\mathrm{tr}(\widehat{\mathbf{\Sigma}}_p)/100$ and $\underline{\ell} = -20\lambda_{\max}(\widehat{\mathbf{\Sigma}}_p)$, which generally lead to quite robust performance.

The composite test procedure with ridge-regularization is summarized in Algoritm 4.2.

Algorithm 4.2 (Composite ridge-regularized test procedure)

1. *Select prior weights* $\widetilde{\Pi} = (\tilde{t}_1, \ldots, \tilde{t}_m)$. *The canonical choice is* $((1,0,0),(0,1,0),(0,0,1))$;
2. *For each* $\tilde{t}_j$ *in* $\widetilde{\Pi}$, *run Algorithm 4.1, get the standardized test statistic* $\widehat{T}(\ell_j^*)$ *and compute* $\widehat{T}_{\max} = \max_{1 \leqslant j \leqslant m} \widehat{T}(\ell_j^*)$;
3. *With the selected tuning parameters* $(\ell_1^*, \ell_2^*, \ell_3^*)$ *compute the matrix* $\widehat{\Delta}_*$ *whose diagonal elements are equal to one and whose* (i,j)*-th entry for* $i \neq j$ *is*

$$\widehat{\Delta}_{\ell_i^*}^{-1/2}(\gamma_n) \widehat{\Delta}_{\ell_i^*, \ell_j^*}(\gamma_n) \widehat{\Delta}_{\ell_j^*}^{-1/2}(\gamma_n),$$

where $\widehat{\Delta}_{\ell_i^*}(\gamma_n)$ *is defined in Step 2. of Algorithm 4.1 and*

$$\widehat{\Delta}_{\ell_i^*, \ell_j^*}(\gamma_n) = 2\widehat{\Theta}(\ell_i^*, \gamma_n)\widehat{\Theta}(\ell_j^*, \gamma_n)\Big[\frac{\ell_i^* \widehat{\Theta}(\ell_i^*, \gamma_n) - \ell_j^* \widehat{\Theta}(\ell_j^*, \gamma_n)}{\ell_i^* - \ell_j^*} - 1\Big];$$

4. *Project* $\widehat{\Delta}_*$ *to its closest non-negative definite matrix* $\widehat{\Delta}_*^+$ *by setting the negative eigenvalues to zero. Generate* $\varepsilon_1, \ldots, \varepsilon_G$ *with* $\varepsilon_b = \max_{1 \leqslant i \leqslant m} Z_i^{(b)}$ *with* $Z^{(b)} = [Z_i^{(b)}]_{i=1}^m \sim \mathcal{N}(0, \widehat{\Delta}_*^+)$.
5. *Compute the p-value as* $G^{-1} \sum_{b=1}^{G} \mathbb{1}\{\varepsilon_b > \widehat{T}_{\max}\}$.

4.2. Extension to higher-order regularizers

Through an extensive simulation study in a MANOVA setting, it is shown in Section 5 that the ridge-regularized tests compare favorably against a host of existing test procedures. This is consistent with the findings in Li et al. (2016) in the two-sample mean test framework. Ridge-shrinkage rescales $\widehat{\mathbf{H}}_p$ by $(\widehat{\mathbf{\Sigma}}_p - \ell I_p)^{-1}$ instead of $\widehat{\mathbf{\Sigma}}_p^{-1}$. Broader classes of scaling matrices have been studied extensively (see Ledoit and Wolf, 2012, for an overview). They can be set up in the form $f(\widehat{\mathbf{\Sigma}}_p)$. When $f(\cdot)$ is analytic, such scaling falls within the class of the proposed tests.

The flexibility provided by a larger class of scaling matrices can be useful to design test procedures for detecting a specific kind of alternative. The choice of the test procedure may for example be guided by questions such as *Which f leads to the best asymptotic power under a specific sequence of local alternatives, if H_0 is rejected based on large eigenvalues of $\mathbf{M}(f)$?* While a full characterization of this question is beyond the scope of this paper, a partial answer may be provided by restricting to functions f in the higher-order class

$$\mathfrak{F}_{\mathrm{high}} = \Big\{ f_\ell(x) = \Big[\sum_{j=0}^{\kappa} l_j x^j \Big]^{-1} : \ell = (l_0, \dots, l_\kappa)^T \in \mathcal{G} \Big\},$$

where $\mathcal{G}$ is such that f_ℓ is uniformly bounded and monotonically decreasing on $\mathcal{X}$, for any $\ell \in \mathcal{G}$. These higher-order shrinkage functions are weighted averages of ridge-type shrinkage functions. To see this, suppose the polynomial $\sum_{j=0}^{\kappa} l_j x^j$ has roots $r_1, \dots, r_{\kappa_0} \in \mathbb{C} \backslash \mathcal{X}$ with multiplicity $s_1, \dots, s_{\kappa_0} \in \mathbb{N}^+$. Via basic algebra, f_ℓ can be expressed as

$$f_\ell(x) = \Big[\sum_{j=0}^{\kappa} l_j x^j \Big]^{-1} = \sum_{j=1}^{\kappa_0} \sum_{i=1}^{s_{\kappa_0}} w_{ji} (x - r_j)^{-i}, \tag{4.1}$$

with some weights $w_{ji} \in \mathbb{C}$. If all roots are simple, f_ℓ is a weighted average of ridge-regularization with κ different parameters. Heuristically, it is expected that a higher order f_ℓ yields tests more robust against unfavorable selection of ridge shrinkage parameter.

The design of $\mathcal{G}$ is not easy when κ is large. Here, we select $\kappa = 3$, which is the minimum degree that allows f_ℓ^{-1} to be both locally convex and concave. In this case, the complexity of selecting the optimal regularizer is significantly higher than for ridge-regularization. Due to space limitations, we move the design of $\mathcal{G}$ and the test procedure when $\kappa = 3$ to Section S.1 of the Supplementary Material.

5. Simulations

In this section, the proposed tests are compared by means of a simulation study to two representative existing methods in the literature, Zhou, Guo and Zhang (2017) (ZGZ) and Cai and Xia (2014) (CX). We focus on one-way MANOVA, a set-up for which both competing methods are applicable. It is worth mentioning that CX requires a good estimator of the precision matrix Σ_p^{-1}, that is typically unavailable when both Σ_p and Σ_p^{-1} are dense. In the simulations, the true Σ_p^{-1} is utilized for CX, thus making it an oracle procedure. In the following, $\mathrm{LR}_{\mathrm{ridge}}$, $\mathrm{LH}_{\mathrm{ridge}}$, and $\mathrm{BNP}_{\mathrm{ridge}}$ denote the ridge-regularized tests presented in Algorithm 4.1. $\mathrm{LR}_{\mathrm{high}}$, $\mathrm{LH}_{\mathrm{high}}$, and $\mathrm{BNP}_{\mathrm{high}}$ denote the tests with higher-order shrinkage introduced in Section 4.2 with $\kappa = 3$. $\mathrm{LR}_{\mathrm{comp}}$, $\mathrm{LH}_{\mathrm{comp}}$ and $\mathrm{BNP}_{\mathrm{comp}}$ denote the composite tests of Algorithm 4.2 with the canonical choice of $\widetilde{\Pi} = ((1,0,0),(0,1,0),(0,0,1))$.

5.1. Settings

The observation matrix $\mathbf{Y}$ was generated as in (1.1) with normally distributed $\mathbf{Z}$. Specifically, we selected $k = 3$ or 5, and $N = 300$. For $k = 3$, the three groups had $75, 90$ and 135 observations, respectively. For $k = 5$, the design was balanced with each group containing 60 observations. The dimension p was $150, 600, 3000$, so that $\gamma_n = p/n \approx 0.5, 2$ and 10. The columns of B were the k group mean vectors. Accordingly, the columns of X were the group index indicators of observation subjects. We selected C to be the successive contrast matrix of order $q = k - 1$. This is a standard one-way MANOVA setting.

Under the null, B is the zero matrix. Under the alternative, for each setting of the parameters and each replicate, B is generated using one of the following models.

(i) *Dense alternative*: The entries of B are i.i.d. $\mathcal{N}(0, c^2)$ with $c = O(n^{-1/4}p^{-1/2})$ used to tune signal strength to a non-trivial level.
(ii) *Sparse altenative*: $B = c\mathcal{R}\boldsymbol{\mathcal{V}}$ with $c = O(n^{-1/4}p^{-1/2})$, where $\mathcal{R}$ is a diagonal $p \times p$ matrix with 10% randomly and uniformly selected diagonal entries being $\sqrt{10}$ and the remaining 90% being equal to 0, and $\boldsymbol{\mathcal{V}}$ is a $p \times p$ matrix with i.i.d. standard normal entries.

The following four models for the covariance matrix $\Sigma = \Sigma_p$ were considered. All models were further scaled so that $\mathrm{tr}(\Sigma_p) = p$.

(i) *Identity matrix* (ID): $\Sigma = I_p$.
(ii) *Dense case* Σ_{den}: Here $\Sigma = P\Sigma_{(1)}P^T$ with a unitary matrix P randomly generated from the Haar measure and resampled for each different setting, and a diagonal matrix $\Sigma_{(1)}$ whose eigenvalues are given by $\lambda_j = (0.1 + j)^6 + 0.05p^6$, $j = 1, \dots, p$. The eigenvalues of Σ decay slowly, so that no dominating leading eigenvalue exists.
(iii) *Toeplitz case* Σ_{toep}: Here Σ is a Teoplitz matrix with the (i, j)-th element equal to $0.5^{|i-j|}$. It is a setting where Σ^{-1} is sparse but Σ is dense.
(iv) *Discrete case* Σ_{dis}: Here $\Sigma = P\Sigma_{(2)}P^T$ with P generated in the same way as in (ii), and $\Sigma_{(2)}$ is a diagonal matrix with 40% eigenvalues 1, 40% eigenvalues 3 and 20% eigenvalues 10.

All tests were conducted at significance level $\alpha = 0.05$. Empirical sizes for the various tests are shown in Tables 5.1 and 5.2. Empirical power curves versus expected signal strength $n^{1/4}p^{1/2}c$ are reported in Figures 5.1–5.3. To better compare the power of each test, curves are displayed after size adjustment where the tests utilize the size-adjusted cut-off values based on the actual null distribution computed by simulations. Counterparts of Figures 5.1–5.3 that utilize asymptotic (approximate) cut-off values are reported in Section S.9. The difference between the two types is limited. LR, LH and BNP criteria behave similarly across simulation settings, as indicated by Theorem 2.4. Therefore, only one of them is displayed in each figure for ease of visualization. More figures can be found in Section S.8 of the Supplementary Material. Note that, in some of the settings, several of the power curves nearly overlap, creating an occlusion effect. Then, power curves corresponding to the composite tests are plotted as the top layer.

		$\Sigma = I_p$						$\Sigma = \Sigma_{den}$					
		$k=3$			$k=5$			$k=3$			$k=5$		
$n=300$, p	$=$	150	600	3000	150	600	3000	150	600	3000	150	600	3000
	$\tilde{t}_1$	5.4	5.2	5.1	5.2	5.1	5.1	4.9	4.4	4.7	4.4	3.3	4.2
$\mathrm{LR}_{\mathrm{ridge}}$	$\tilde{t}_2$	5.4	5.2	5.1	5.2	5.1	5.1	4.9	5.2	4.9	4.4	4.9	4.7
	$\tilde{t}_3$	5.3	5.2	5.1	5.2	5.1	5.1	5.8	5.9	5.1	5.3	5.2	4.9
	$\tilde{t}_1$	5.4	5.2	5.1	5.3	5.1	5.2	6.2	7.2	5.7	6.2	7.7	6.0
$\mathrm{LH}_{\mathrm{ridge}}$	$\tilde{t}_2$	5.4	5.2	5.1	5.3	5.1	5.2	6.2	5.9	5.2	6.2	5.9	5.1
	$\tilde{t}_3$	5.3	5.2	5.1	5.3	5.1	5.2	5.8	5.9	5.2	5.4	5.2	5.0
	$\tilde{t}_1$	5.3	5.2	5.0	5.2	5.0	5.0	4.0	2.5	3.7	2.9	1.3	3.1
$\mathrm{BNP}_{\mathrm{ridge}}$	$\tilde{t}_2$	5.4	5.2	5.0	5.2	5.0	5.0	4.0	4.7	4.6	2.9	3.9	4.4
	$\tilde{t}_3$	5.3	5.2	5.0	5.2	5.0	5.0	5.8	5.8	5.0	5.3	5.1	4.7
	$\tilde{t}_1$	6.5	6.3	5.3	6.5	5.3	5.5	6.0	5.8	5.1	6.5	5.9	4.5
$\mathrm{LR}_{\mathrm{high}}$	$\tilde{t}_2$	6.5	6.3	5.3	6.5	5.3	5.5	8.3	6.8	5.5	8.4	7.2	5.2
	$\tilde{t}_3$	6.6	6.3	5.3	6.6	5.3	5.5	6.7	6.7	5.5	6.4	7.1	5.2
	$\tilde{t}_1$	6.7	6.4	5.4	6.8	5.5	5.7	6.1	5.9	5.7	6.7	6.2	5.5
$\mathrm{LH}_{\mathrm{high}}$	$\tilde{t}_2$	6.7	6.4	5.4	6.8	5.4	5.7	8.3	6.8	5.6	8.5	7.3	5.5
	$\tilde{t}_3$	6.7	6.4	5.4	6.8	5.4	5.7	6.7	6.7	5.6	6.5	7.2	5.5
	$\tilde{t}_1$	6.2	6.3	5.2	6.1	5.3	5.2	5.9	5.7	4.6	6.4	5.5	3.7
$\mathrm{BNP}_{\mathrm{high}}$	$\tilde{t}_2$	6.3	6.3	5.2	6.1	5.2	5.2	8.3	6.7	5.3	8.3	7.0	4.9
	$\tilde{t}_3$	6.3	6.3	5.1	6.1	5.2	5.2	6.6	6.6	5.3	6.4	6.9	4.9
$\mathrm{LR}_{\mathrm{comp}}$		5.1	5.1	5.0	5.4	5.3	5.0	6.0	5.1	5.5	5.6	5.0	5.1
$\mathrm{LH}_{\mathrm{comp}}$		5.1	5.1	5.1	5.5	5.3	5.1	6.7	5.8	5.9	6.9	6.2	5.7
$\mathrm{BNP}_{\mathrm{comp}}$		5.1	5.0	5.0	5.4	5.2	5.0	5.4	4.5	5.1	4.7	4.4	4.6
ZGZ		5.6	5.7	5.2	5.6	4.8	5.2	5.9	5.5	5.4	5.4	5.4	5.3
CX (Oracle)		5.6	6.3	7.0	7.3	6.9	8.6	5.8	5.9	6.8	6.0	7.2	9.0

Table 5.1. Empirical sizes at level 5%. $\Sigma = \mathrm{ID}$ and Σ_{den}; $\tilde{t}_1 = (1,0,0)$, $\tilde{t}_2 = (0,1,0)$, $\tilde{t}_3 = (0,0,1)$.

		$\Sigma = \Sigma_{dis}$						$\Sigma = \Sigma_{toep}$					
		$k=3$			$k=5$			$k=3$			$k=5$		
$n=300$, p	$=$	150	600	3000	150	600	3000	150	600	3000	150	600	3000
	$\tilde{t}_1$	4.8	5.0	4.6	4.7	4.5	5.0	5.4	4.4	4.8	4.5	4.6	4.6
$\mathrm{LR}_{\mathrm{ridge}}$	$\tilde{t}_2$	5.1	5.2	4.9	5.2	4.6	5.1	5.4	4.9	4.9	4.9	4.8	5.0
	$\tilde{t}_3$	5.6	5.5	5.1	5.7	5.3	5.3	5.8	5.2	5.0	5.7	5.4	5.1
	$\tilde{t}_1$	5.8	6.0	5.2	6.6	6.3	5.6	6.4	5.3	5.2	6.2	6.3	5.3
$\mathrm{LH}_{\mathrm{ridge}}$	$\tilde{t}_2$	5.7	5.7	5.1	6.3	5.6	5.5	5.9	5.3	5.0	5.8	5.6	5.3
	$\tilde{t}_3$	5.6	5.5	5.2	5.8	5.3	5.4	5.8	5.3	5.1	5.7	5.4	5.2
	$\tilde{t}_1$	3.9	4.1	4.3	3.1	3.1	4.1	4.4	3.7	4.4	3.2	3.4	3.9
$\mathrm{BNP}_{\mathrm{ridge}}$	$\tilde{t}_2$	4.6	4.8	4.8	4.1	4.0	4.9	4.9	4.4	4.8	4.1	4.3	4.7
	$\tilde{t}_3$	5.5	5.5	5.0	5.7	5.2	5.1	5.8	5.2	5.0	5.6	5.4	5.1
	$\tilde{t}_1$	6.3	6.4	4.8	5.9	7.0	5.5	7.1	7.0	5.3	7.5	6.9	5.2
$\mathrm{LR}_{\mathrm{high}}$	$\tilde{t}_2$	7.9	6.5	4.8	8.3	7.1	5.5	7.6	7.2	5.3	7.8	7.0	5.2
	$\tilde{t}_3$	6.1	5.6	4.8	6.4	6.1	5.5	6.7	6.5	5.3	6.6	6.4	5.2
	$\tilde{t}_1$	6.6	6.5	5.0	6.2	7.2	5.7	7.2	7.2	5.5	7.7	7.0	5.5
$\mathrm{LH}_{\mathrm{high}}$	$\tilde{t}_2$	8.0	6.6	5.0	8.5	7.2	5.7	7.8	7.2	5.5	8.0	7.1	5.5
	$\tilde{t}_3$	6.2	5.6	5.0	6.5	6.2	5.7	6.7	6.5	5.5	6.7	6.5	5.5
	$\tilde{t}_1$	6.1	6.3	4.7	5.6	6.8	5.3	7.1	7.0	5.2	7.2	6.8	5.1
$\mathrm{BNP}_{\mathrm{high}}$	$\tilde{t}_2$	7.9	6.4	4.7	8.2	7.0	5.3	7.5	7.1	5.2	7.7	7.0	5.1
	$\tilde{t}_3$	6.1	5.5	4.7	6.4	6.0	5.3	6.6	6.4	5.2	6.5	6.3	5.1
$\mathrm{LR}_{\mathrm{comp}}$		6.2	5.2	5.0	5.2	5.3	5.5	5.9	5.0	5.1	5.5	4.9	4.9
$\mathrm{LH}_{\mathrm{comp}}$		7.0	5.9	5.3	6.5	6.4	6.0	6.6	5.6	5.3	6.6	5.7	5.3
$\mathrm{BNP}_{\mathrm{comp}}$		5.5	4.6	4.8	4.4	4.6	5.0	5.4	4.6	4.9	4.8	4.4	4.6
ZGZ		5.5	4.7	4.6	5.7	5.1	5.3	6.0	5.5	5.0	5.9	5.6	5.0
CX (Oracle)		5.3	5.9	6.6	6.8	7.2	8.6	5.3	6.2	6.8	6.8	7.2	8.4

Table 5.2. Empirical sizes at level 5%. $\Sigma = \Sigma_{dis}$ and Σ_{toep}; $\tilde{t}_1 = (1,0,0)$, $\tilde{t}_2 = (0,1,0)$, $\tilde{t}_3 = (0,0,1)$.

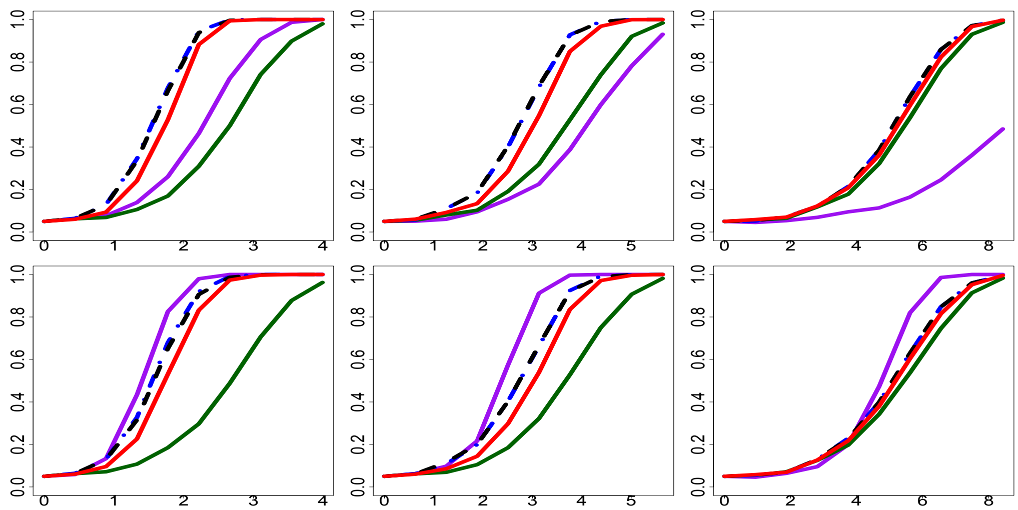

Figure 5.1: Size-adjusted power with $\Sigma = \Sigma_{den}$, $k = 5$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. BNP$_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); BNP$_{\mathrm{ridge}}$ (black, dashed) and BNP$_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (1, 0, 0)$.

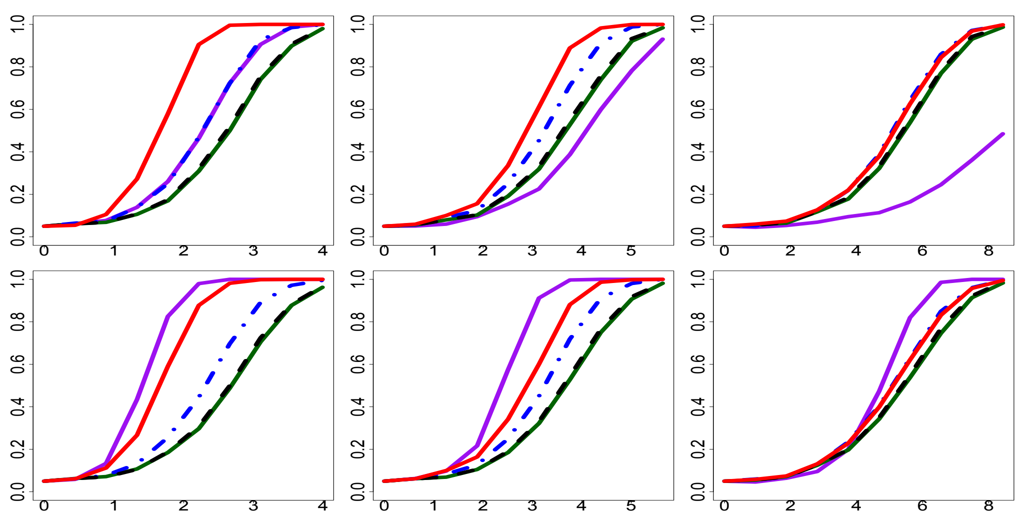

Figure 5.2: Size-adjusted power with $\Sigma = \Sigma_{den}$, $k = 5$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. LH$_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); LH$_{\mathrm{ridge}}$ (black, dashed) and LH$_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$.

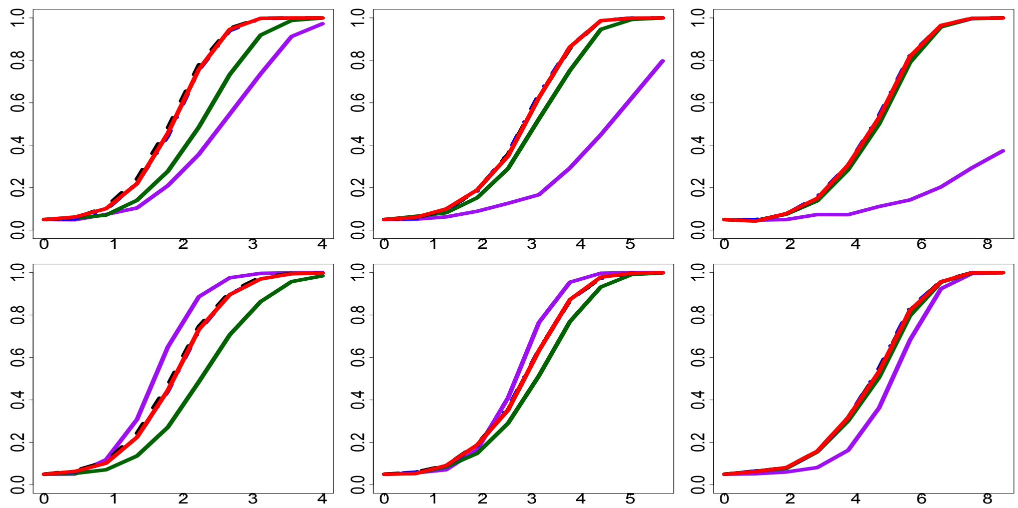

Figure 5.3: Size-adjusted power with $\Sigma = \Sigma_{toep}$, $k = 3$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LR}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LR}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LR}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$.

5.2. Summary of simulation results

Tables 5.1 and 5.2 show the empirical sizes of the proposed tests are mostly controlled under 7.5%. The slight oversize is caused by the fact that $\mathbf{M}(f)$ behaves like a quadratic form, therefore the finite sample distribution is skewed. LR and BNP tests are more conservative than LH tests because the former two calibrate the statistics by transforming eigenvalues of $\mathbf{M}(f)$. Ridge-regularized tests are slightly more conservative under higher-order shrinkage.

Note that in both simulation settings, B consists of independent entries. Therefore, $\tilde{t}_1 = (1, 0, 0)$ is considered as a correctly specified prior, while $\tilde{t}_2 = (0, 1, 0)$ and $\tilde{t}_3 = (0, 0, 1)$ are considered as moderately and severely misspecified, respectively. The composite tests combine $\tilde{t}_1$, $\tilde{t}_2$ and $\tilde{t}_3$, and are therefore considered as consistently capturing the correct prior. We shall treat the composite tests as a baseline to study the effect of prior misspecification, by comparing them to tests using a single $\tilde{t}$.

For each simulation configuration considered in this study, the proposed procedures are as powerful as the procedure with the best performance, except for the cases when B is sparse, p is small, and priors are severely misspecified in the proposed tests; see Figure S.8.6 in the Supplementary Material. We highlight the following observations based on the simulation results.

(1) The composite tests are slightly less efficient than $\mathrm{BNP}_{\mathrm{ridge}}$ and $\mathrm{BNP}_{\mathrm{high}}$ when the correct prior $\tilde{t}_1$ is used, as in Figure 5.1. However, as in Figure 5.2, when the prior is

severely misspecified, the composite test is significantly more powerful. It suggests that the composite tests are robust against prior misspecification, although losing some efficiency against tests with correctly specified priors.

(2) Although ridge-shrinkage and higher-order shrinkage behave similarly under the correct prior, the latter outperforms the former when the prior is misspecified; see Figure 5.2. This provides evidence for the robustness of high-order shrinkage against unfavorable ridge shrinkage parameter selection.

(3) ZGZ is a special case of the proposed test family with $f(x) = 1$ for all x, which amounts to replacing $\widehat{\mathbf{\Sigma}}_p$ with I_p. When $\Sigma_p = I_p$, ZGZ appears to be the reasonable option at least intuitively. Note, both $\mathfrak{F}_{\text{ridge}}$ and $\mathfrak{F}_{\text{high}}$ contain functions close to $f(x) = 1$. Figures for $\Sigma_p = I_p$ displayed in Section S.8 of the Supplementary Material show that the proposed tests perform as well as ZGZ in that case. It may be viewed as evidence of the effectiveness of the data-driven shrinkage selection strategy detailed in Section 3.

(4) Comparing to ZGZ, when the eigenvalues of Σ_p are disperse, the proposed tests are significantly more powerful when $p = 150$ and 600, but behave similarly as ZGZ when $p = 3000$. On the other hand, as in Figure 5.2, the ridge-regularized test with a severely misspecified prior $\tilde{t}_3$, is close to ZGZ.

(5) CX is a test specifically designed for sparse alternatives. The procedure shows its advantage in favorable settings, especially when $p = 150$. Simulation results suggest that the proposed tests are still comparable to CX even under sparse BC and Σ_p^{-1}, as long as the prior in use is not severely misspecified. When p is large, the proposed tests are significantly better when $\Sigma_p = I_p$. Evidence may be found in Figures S.8.10, S.8.11 and S.8.12 of the Supplementary Material.

6. Discussion

In this paper, we addressed the problem of testing general hypotheses in a high-dimensional setting by proposing a family of rotation-invariant tests that generalizes well-studied tests in the literature through utilization of a shrunken version of the empirical error covariance matrix. The shrinkage function is an analytic function on the support of the limiting spectrum of the empirical error covariance matrix. The asymptotic null distribution was built under finite fourth-moment assumption of the observations and a regime where the dimension of the observations (response) is proportional to the sample sizes, while the dimension of the regressors remains fixed. This class encompasses the MANOVA problem with a finite number of populations, and multivariate regression involving a finite number of predictors. We studied the asymptotic power of the proposed tests under a Bayesian framework involving a flexible class of local alternatives that determines the structure of the parameter of interest. We proposed a data-driven procedure for selection of the shrinkage function that relies on maximizing the asymptotic power of the test under specific classes of local alternatives. We also extended the procedure to propose a composite test that combines the optimally chosen tests associated with a finite collection of distinct local alternatives. Finally, we illustrated the test procedures by focusing on specific shrinkage functions including the ridge-regularization and a higher-order generalization. Simulation studies were conducted to show both the ridge and high-order regularizers have good power under various settings of population covariance and alternatives.

There are several future research directions that can be pursued. On the technical side, the analytic requirement of the shrinkage is still somewhat restrictive. One aim is to seek a generalization from analyticity to fourth-order continuous differentiability. A decision-theoretic selection of the shrinkage parameter that is optimal with respect to a broad class of local alternatives is an interesting theoretical challenge. Another challenge is to find suitable modifications to the tests that enable improvement of their power characteristics even when the dimension is an order of magnitude larger than the sample size, a setting that is outside the analytical framework adopted here.

References

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis* **2**. Wiley New York.

Bai, Z., Chen, J. and Yao, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Australian & New Zealand Journal of Statistics* **52** 423–437.

Bai, Z., Choi, K. P. and Fujikoshi, Y. (2017). Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica* **6** 311–329.

Bai, Z. and Silverstein, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *The Annals of Probability* **26** 316–345.

Bai, Z. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability* **32** 553–605.

Bai, Z. and Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.

Bai, Z., Jiang, D., Yao, J.-f. and Zheng, S. (2013). Testing linear hypotheses in high-dimensional regressions. *Statistics* **47** 1207–1223.

Billingsley, P. (1968). *Convergence of Probability Measures*. Probability and Mathematical Statistics. New York: Wiley.

Billingsley, P. (1995). *Probability and Measure*. Probability and Mathematical Statistics. New York: Wiley.

Biswas, M. and Ghosh, A. K. (2014). A nonparametric two-sample test applicable to high dimensional data. *Journal of Multivariate Analysis* **123** 160–171.

Cai, T. T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76** 349–372.

Cai, T. T. and Xia, Y. (2014). High-dimensional sparse MANOVA. *Journal of Multivariate Analysis* **131** 174–196.

Chang, J., Zheng, C., Zhou, W. and Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73** 1300-1310.

Chen, S. X., Li, J. and Zhong, P.-S. (2014). Two-sample tests for high dimensional means with thresholding and data transformation. *arXiv preprint arXiv:1410.2848*.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics* **38** 808–835.

Chen, L. S., Paul, D., Prentice, R. L. and Wang, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *Journal of the American Statistical Association* **106** 1345–1360.

El Karoui, N. and Kösters, H. (2011). Geometric sensitivity of random matrix results: consequences for shrinkage estimators of covariance and related statistical methods. *arXiv preprint arXiv:1105.1404*.

Fan, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proceedings of the National Academy of Sciences* **37** 760–766.

Fujikoshi, Y., Himeno, T. and Wakaki, H. (2004). Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size. *Journal of the Japan Statistical Society* **34** 19–26.

Guo, B. and Chen, S. X. (2016). Tests for high dimensional generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78** 1079–1102.

Hu, J., Bai, Z., Wang, C. and Wang, W. (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics* **69** 365–387.

Ledoit, O. and Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields* **151** 233–264.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* **40** 1024–1060.

Li, H., Aue, A., Paul, D., Peng, J. and Wang, P. (2016). An adaptable generalization of Hotelling's T^2 test in high dimension. *arXiv preprint arXiv:1609.08725*.

Liu, H., Aue, A. and Paul, D. (2015). On the Marčenko–Pastur law for linear time series. *The Annals of Statistics* **43** 675–712.

Lopes, M., Jacob, L. and Wainwright, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. In *Advances in Neural Information Processing Systems* 1206–1214.

Muirhead, R. J. (2009). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons.

Pan, G. and Zhou, W. (2011). Central limit theorem for Hotelling's T^2 statistic under large dimension. *The Annals of Applied Probability* **21** 1860–1910.

Paul, D. and Aue, A. (2014). Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference* **150** 1–29.

Silverstein, J. W. and Choi, S.-I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. *Journal of Multivariate Analysis* **54** 295–309.

Srivastava, M. S. and Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis* **97** 1927–1940.

Srivastava, R., Li, P. and Ruppert, D. (2016). RAPTT: An exact two-sample test in high dimensions using random projections. *Journal of Computational and Graphical Statistics* **25** 954–970.

Wang, L., Peng, B. and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association* **110** 1658–1669.

Yamada, T. and Himeno, T. (2015). Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. *Journal of Multivariate Analysis* **139** 7–27.

Yin, Y.-Q., Bai, Z. and Krishnaiah, P. R. (1988). On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields* **78** 509–521.

Zhou, B., Guo, J. and Zhang, J.-T. (2017). High-dimensional general linear hypothesis testing under heteroscedasticity. *Journal of Statistical Planning and Inference* **188** 36–54.

Appendix

This appendix contains the proofs of the main theorems. Additional proofs of supporting lemmas can be found in the Supplementary Material. The arguments used here build on those in Bai and Silverstein (2010) and Pan and Zhou (2011) but contain considerable innovations. First, the key assumption **C4** on the spectrum of Σ_p is different from its counterpart in Bai and Silverstein (2010). Therefore, additional results regarding the Stieltjes transforms $m_{n,p}(\mathbb{z})$ and $m(\mathbb{z})$ are needed here. We also need the uniform convergence of derivatives of $m_{n,p}(\mathbb{z})$ on a contour, which was not part of Bai and Silverstein (2010). These results are detailed in Section S.2 of the Supplementary Material. Second, Pan and Zhou (2011) only considered the case $\Sigma_p = I_p$ because of the invariance of Hotelling's T^2 statistic with respect to Σ_p. Important arguments in their paper, for example their Lemma 6, no longer hold under general covariance structures. Additionally, the calculation of the asymptotic variance and covariance of the quadratic form $n^{-1}Q_n^T\mathbf{Y}^T(\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}\mathbf{Y}Q_n$ is significantly more involved than for Hotelling's T^2 statistic. Third, our proof includes an important transformation of the quadratic form to deal with the complex correlation structure between $\mathbf{Y}Q_n$ and $\widehat{\mathbf{\Sigma}}_p$. We believe the trick can be used to generalize existing work in the literature, for example Bai, Choi and Fujikoshi (2017).

A.1. Proof of Theorem 2.1

Recall that

$$\begin{aligned} Q_n &= X^T(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1/2}, \\ \mathbf{M}(f) &= \frac{1}{n}Q_n^T\mathbf{Y}^T f(\widehat{\mathbf{\Sigma}}_p)\mathbf{Y}Q_n, \\ \widehat{\mathbf{\Sigma}}_p &= \frac{1}{n}\mathbf{Y}(I - X^T(XX^T)^{-1}X)\mathbf{Y}^T. \end{aligned}$$

Introduce the product $Q_n = U_n V_n$ with

$$U_n = X^T(XX^T)^{-1/2}\,, \tag{A.1}$$

$$V_n = (XX^T)^{-1/2}C[C^T(XX^T)^{-1}C]^{-1/2}\,. \tag{A.2}$$

This decomposition will aid the analysis of the correlation between $\mathbf{Y}Q_n$ and $\widehat{\mathbf{\Sigma}}_p$.

From now on, use $\Sigma_p^{T/2}$ to denote $(\Sigma_p^{1/2})^T$. Under the null hypothesis, the following representations hold:

$$\mathbf{M}(f) = \frac{1}{n}V_n^T U_n^T \mathbf{Z}^T \Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}U_n V_n,$$

$$\widehat{\mathbf{\Sigma}}_p = \frac{1}{n}\Sigma_p^{1/2}\mathbf{Z}(I - U_n U_n^T)\mathbf{Z}^T\Sigma_p^{T/2}.$$

Observe that the joint asymptotic normality of entries in $\sqrt{n}\mathbf{M}(f)$ is equivalent to the asymptotic normality of

$$n^{-1/2}\alpha^T V_n^T U_n^T \mathbf{Z}^T \Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}U_n V_n \eta$$

for arbitrary (but fixed) vectors α and $\eta \in \mathbb{R}^q$.

Recall that $\mathcal{X} = [0,\ \limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2]$. Let $\mathcal{C}$ be any contour enclosing $\mathcal{X}$ such that $f(\cdot)$ is analytic on its interior. With slight modifications, all arguments in the following hold for arbitrary such $\mathcal{C}$. For convenience, select $\mathcal{C}$ as rectangle with vertices $\underline{u} \pm iv_0$ and $\overline{u} \pm iv_0$, such that

$$v_0 > 0; \quad \overline{u} > \limsup \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2; \quad \underline{u} < 0.$$

Such a rectangle must exist.

By *Cauchy's integral formula*, if $\lambda_{\max}(\widehat{\mathbf{\Sigma}}_p) < \overline{u}$,

$$\begin{aligned} &n^{-1/2}\alpha^T V_n^T U_n^T \mathbf{Z}^T \Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}U_n V_n \eta \\ &\quad = \frac{-1}{2\pi i}\oint_{\mathcal{C}} f(\mathbb{z}) n^{-1/2}\alpha^T V_n^T U_n^T \mathbf{Z}^T \Sigma_p^{T/2}(\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1}\Sigma_p^{1/2}\mathbf{Z}U_n V_n \eta d\mathbb{z}. \end{aligned} \tag{A.3}$$

If $\lambda_{\max}(\widehat{\mathbf{\Sigma}}_p) \geqslant \overline{u}$, the above equality may not hold. However, if we can show that $\mathbb{P}(\lambda_{\max}(\widehat{\mathbf{\Sigma}}_p) \geqslant \overline{u})$ converges to 0, we can still acquire the weak limit of the left-hand side by deriving the weak limit of the right-hand side. Yin, Bai and Krishnaiah (1988, Theorem 3.1) implies that

$$\mathbb{P}(\lambda_{\max}(\widehat{\mathbf{\Sigma}}_p) \geqslant \overline{u}) \to 0. \tag{A.4}$$

Hence, it suffices to show the asymptotic normality of the process

$$\xi_n(\mathbb{z}, \alpha, \eta) = n^{-1/2}\alpha^T V_n^T U_n^T \mathbf{Z}^T \Sigma_p^{T/2}(\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1}\Sigma_p^{1/2}\mathbf{Z}U_n V_n \eta, \qquad \mathbb{z} \in \mathcal{C}.$$

Clearly, $\xi(\mathbb{z}, \alpha, \eta)$ is continuous with respect to $\mathbb{z}$. All asymptotic results are derived in the space of continuous functions on $\mathcal{C}$ with uniform topology. In the following, study the

$k \times k$ matrix process $U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widehat{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} \Sigma_p^{1/2} \mathbf{Z} U_n$, that is a component of $\xi(\mathbb{z}, \alpha, \eta)$. Note that a $k \times k$ complex-valued matrix can be viewed as a $2k^2$-variate real-valued random element. To unify dimensionality and avoid potential ambiguity, all objects are treated as random elements in the metric space $C(\mathcal{C}, \mathbb{R}^{2k^2})$ in the following. For dealing with matrices of dimension $q \times q$ (with $q < k$), note that a $q \times q$ matrix can be viewed as the q-th order leading submatrix of a $k \times k$ matrix whose other entries equal 0. Similarly, univariate functions may be viewed as 1×1 submatrices. Therefore, results in Chapter 2 of Billingsley (1968) apply with Euclidean distance replaced by Frobenius norm of a matrix, that is $\|A\|_F = (\sum_{i=1}^m \sum_{j=1}^r |a_{ij}|^2)^{1/2}$, where $A = [a_{ij}]_{ij}$.

We may proceed to prove the asymptotic normality of $\xi_n(\mathbb{z}, \alpha, \eta)$ on $\mathbb{z} \in \mathcal{C}$ directly. However, several technical challenges need to be addressed. First, in view of the spectral norm of $(\widehat{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1}$ being unbounded when $\mathbb{z}$ is close to the real axis and extreme eigenvalues of $\widehat{\mathbf{\Sigma}}_p$ exceed $\limsup \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2$, the tightness of the process $\xi_n(\mathbb{z}, \alpha, \eta)$ is unclear. Secondly, $\widehat{\mathbf{\Sigma}}_p$ is not a summation of independent terms, but contains $\mathbf{Z} U_n U_n^T \mathbf{Z}^T$, a component containing cross product terms between pairs of columns of $\mathbf{Z}$. These terms entangle the analysis of the correlation between $\widehat{\mathbf{\Sigma}}_p$ and each single column of $\mathbf{Z}$. For these technical reasons, we avoid directly working on $\xi_n(\mathbb{z}, \alpha, \eta)$ under **C1** on $\mathbb{z} \in \mathcal{C}$, but start with $n^{-1/2} U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} \Sigma_p^{1/2} \mathbf{Z} U_n$, a component of $\xi_n(\mathbb{z}, \alpha, \eta)$ with $\widehat{\mathbf{\Sigma}}_p$ replaced by an uncentered counterpart

$$\widetilde{\mathbf{\Sigma}}_p = \frac{1}{n} \Sigma_p^{1/2} \mathbf{Z} \mathbf{Z}^T \Sigma_p^{T/2}. \tag{A.5}$$

The relationship between $\widetilde{\mathbf{\Sigma}}_p$ and $\widehat{\mathbf{\Sigma}}_p$ is given by

$$\widehat{\mathbf{\Sigma}}_p = \widetilde{\mathbf{\Sigma}}_p - \frac{1}{n} \Sigma_p^{1/2} \mathbf{Z} U_n U_n^T \mathbf{Z}^T \Sigma_p^{T/2}. \tag{A.6}$$

Next, we modify the process and the distribution of $\mathbf{Z}$ as follows.

Process smoothing. Select a sequence of positive numbers ρ_n decaying to 0 with a rate such that for some $\omega \in (1, 2)$,

$$n\rho_n \downarrow 0, \quad \rho_n \geqslant n^{-\omega}.$$

Let $\mathcal{C}^+ = \mathcal{C} \cap \{u + iv \colon |v| \geqslant \rho_n\}$. Define

$$\begin{aligned}
\widetilde{\mathcal{Q}}_n(\mathbb{z}) &= n^{-1} U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} \Sigma_p^{1/2} \mathbf{Z} U_n, \qquad \text{if } \mathbb{z} \in \mathcal{C}^+, \\
\widetilde{\mathcal{Q}}_n(\mathbb{z}) &= \frac{\rho_n - v}{2\rho_n} \widetilde{\mathcal{Q}}_n(u + i\rho_n) + \frac{v + \rho_n}{2\rho_n} \widetilde{\mathcal{Q}}_n(u - i\rho_n), \qquad \text{if } \mathbb{z} \in \mathcal{C} \backslash \mathcal{C}^+.
\end{aligned}$$

To understand this definition better, note that if $\mathbb{z}$ is too close to the real axis, $\widetilde{\mathcal{Q}}_n(\mathbb{z})$ is modified to be the linear interpolation of its values at $u + i\rho_n$ and $u - i\rho_n$. Observe that V_n appearing in $\xi_n(\mathbb{z}, \alpha, \eta)$ was left out when defining $\widetilde{\mathcal{Q}}_n(\mathbb{z})$. This trick that helps transforming back to $\widehat{\mathbf{\Sigma}}_p$ from $\widetilde{\mathbf{\Sigma}}_p$; see (A.8). Note also that V_n is a sequence of deterministic matrices of fixed dimensions, having a limit under **C5** and **C6**. The reason to smooth the process is to guarantee a bound of order $O(\rho_n^{-1})$ on the spectral norm of $(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}$. It is crucial in the proof of tightness.

Variable truncation. **C1** will be temporarily replaced by the following truncated variable condition. Select a positive sequence ε_n such that

$$\varepsilon_n \to 0 \qquad \text{and} \qquad \varepsilon_n^{-4}\mathbb{E}[z_{11}^4 \mathbb{1}(|z_{11}| \geqslant \varepsilon_n n^{1/2})] \to 0.$$

The existence of such a sequence is shown in Yin, Bai and Krishnaiah (1988). We then truncate z_{ij} to be $z_{ij}\mathbb{1}(|z_{ij}| \leqslant \varepsilon_n n^{1/2})$. After that, we re-standardize the truncated variable to maintain zero mean and unit variance. Since we will mostly work on the truncated variables in the following sections, for notational simplicity, we shall use z_{ij} to denote the truncated random variables and $\breve{z}_{ij}$ to denote the original random variable satisfying **C1**. That is,

$$z_{ij} = \frac{\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| \leqslant \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| \leqslant \varepsilon_n n^{1/2})}{\{\mathbb{E}[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| \leqslant \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| \leqslant \varepsilon_n n^{1/2})]^2\}^{1/2}}.$$

For some constant $\mathcal{K}$, when n is sufficiently large,

$$|z_{ij}| \leqslant \mathcal{K}\varepsilon_n n^{1/2}, \quad \mathbb{E}[z_{ij}] = 0, \quad \mathbb{E}[z_{ij}^2] = 1, \quad \mathbb{E}[z_{ij}^4] < \infty. \tag{A.7}$$

The reason to truncate $\breve{z}_{ij}$ is to obtain a bound on the probability of extreme eigenvalues of $\widehat{\mathbf{\Sigma}}_p$ exceeding $\limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2$. A tail bound decaying fast enough is critical when proving tightness of the smoothed random processes on $\mathcal{C}$. Under the original condition **C1**, although (A.4) holds, such a tail bound is not available. After the truncation, the following lemma holds.

Lemma A.1 (Yin, Bai and Krishnaiah (1988); Bai and Silverstein (2004)) *Suppose the entries of* $\mathbf{Z}$ *satisfy* (A.7). *For any positive* ℓ *and any* $\mathfrak{D} \in (\limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2,\ \overline{u})$,

$$\mathbb{P}(\lambda_{\max}(\widetilde{\mathbf{\Sigma}}_p) \geqslant \mathfrak{D}) = o(n^{-\ell}).$$

It is argued later that the process smoothing and variable truncation steps do not change the weak limit of objects under consideration.

For arbitrary vectors a and $b \in \mathbb{R}^k$, define

$$G_n(\mathbb{z}, a, b) = a^T \widetilde{\mathcal{Q}}_n(\mathbb{z}) b.$$

Theorem A.1 *Suppose* $\mathbf{Z}$ *satisfies (A.7) and suppose* ***C2*–*C6*** *in Section 2 hold. Then,*

$$n^{1/2}\Big\{G_n(\mathbb{z}, a, b) - a^T b\, \frac{\Theta(\mathbb{z},\gamma) - 1}{\Theta(\mathbb{z},\gamma)}\Big\} \xrightarrow{D} \Psi^{(1)}(\mathbb{z}), \qquad \mathbb{z} \in \mathcal{C},$$

where $\xrightarrow{D}$ *denotes weak convergence in* $C(\mathcal{C}, \mathbb{R}^{2k^2})$, *and* $\Psi^{(1)}(\mathbb{z})$ *is a Gaussian process with zero mean and covariance function*

$$\Gamma^{(1)}(\mathbb{z}_1, \mathbb{z}_2) = \delta(\mathbb{z}_1, \mathbb{z}_2, \gamma)\Theta^{-2}(\mathbb{z}_1, \gamma)\Theta^{-2}(\mathbb{z}_2, \gamma)[\|a\|^2\|b\|^2 + (a^T b)^2].$$

See Section S.3 of the Supplementary Material for proof of the theorem.

Notice that $a^T b(\Theta(\mathbb{z},\gamma)-1)/\Theta(\mathbb{z},\gamma)$ is the pointwise asymptotic mean of $G_n(\mathbb{z},a,b)$. This expression suggests that in order to establish Theorem A.1 we might need to smooth $\Theta(\mathbb{z},\gamma)$ when the imaginary part of $\mathbb{z}$ in absolute value is smaller than ρ_n, in the same way as in the process smoothing strategy. Similar considerations apply to the treatment of the pointwise asymptotic covariance $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$ of $G_n(\mathbb{z},a,b)$. However, notice that $\Theta(\mathbb{z},\gamma)$ and $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$ are smooth functions of $\mathbb{z}$, $\mathbb{z}_1$ and $\mathbb{z}_2$ with bounded derivatives on $\mathcal{C}$. Therefore, when $\mathbb{z}\in\mathcal{C}\backslash\mathcal{C}^+$, $\Theta(\mathbb{z},\gamma)=\Theta(u\pm i\rho_n,\gamma)+O(\rho_n)$ and $\sqrt{n}\rho_n\to 0$. Similar results hold for $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$ when $\mathbb{z}_1$ and/or $\mathbb{z}_2$ are close to the real axis. We provide details in Section S.2 of the Supplementary Material, where the behavior of $\Theta(\mathbb{z},\gamma)$ and $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$ on $\mathcal{C}$ (correspondingly, $\mathcal{C}^2$) is discussed. Notably, $(\Theta(\mathbb{z},\gamma)-1)/\Theta(\mathbb{z},\gamma)$ is bounded away from 1 on $\mathcal{C}$. These results help in applying the delta-method for proving asymptotic normality.

The following result is an immediate consequence of Theorem A.1.

Lemma A.2 *Suppose* $\mathbf{Z}$ *satisfies (A.7) and suppose* ***C2***–***C6*** *in Section 2 hold. Then,*

$$n^{1/2}\Big\{\widetilde{\mathcal{Q}}_n(\mathbb{z})-\frac{\Theta(\mathbb{z},\gamma)-1}{\Theta(\mathbb{z},\gamma)}I_k\Big\}\xrightarrow{D}\Psi^{(2)}(\mathbb{z}),\qquad \mathbb{z}\in\mathcal{C},$$

where $\Psi^{(2)}(\mathbb{z})=[\Psi^{(2)}(\mathbb{z})]_{ij}$ *is a* $k\times k$ *symmetric Gaussian matrix process with zero mean and covariance, such that for* $i\leqslant j$, $i'\leqslant j'$,

$$\begin{aligned}
&\mathbb{E}[\Psi^{(2)}(\mathbb{z}_1)]_{ii}[\Psi^{(2)}(\mathbb{z}_2)]_{ii}=2\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)\Theta^{-2}(\mathbb{z}_1,\gamma)\Theta^{-2}(\mathbb{z}_2,\gamma)\ ,\\
&\mathbb{E}[\Psi^{(2)}(\mathbb{z}_1)]_{ij}[\Psi^{(2)}(\mathbb{z}_2)]_{ij}=\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)\Theta^{-2}(\mathbb{z}_1,\gamma)\Theta^{-2}(\mathbb{z}_2,\gamma),\qquad \text{if } i\neq j\ ,\\
&\mathbb{E}[\Psi^{(2)}(\mathbb{z}_1)]_{ij}[\Psi^{(2)}(\mathbb{z}_2)]_{i'j'}=0,\qquad \text{if } i\neq i' \text{ or } j\neq j'.
\end{aligned}$$

Next, transforming back to $\widehat{\mathbf{\Sigma}}_p$, define

$$\begin{aligned}
\widehat{\mathcal{Q}}_n(\mathbb{z})&=n^{-1}U_n^T\mathbf{Z}^T\Sigma_p^{T/2}(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}\Sigma_p^{1/2}\mathbf{Z}U_n,\qquad \mathbb{z}\in\mathcal{C}^+,\\
\widehat{\mathcal{Q}}_n(\mathbb{z})&=\frac{\rho_n-v}{2\rho_n}\widehat{\mathcal{Q}}_n(u+i\rho_n)+\frac{v+\rho_n}{2\rho_n}\widehat{\mathcal{Q}}_n(u-i\rho_n),\qquad \mathbb{z}\in\mathcal{C}\backslash\mathcal{C}^+.
\end{aligned}$$

Using the identity (A.5), and Lemma S.6 (*Woodbury matrix identity*) in the Supplementary Material, we get

$$\widehat{\mathcal{Q}}_n(\mathbb{z})=\widetilde{\mathcal{Q}}_n(\mathbb{z})[I_k-\widetilde{\mathcal{Q}}_n(\mathbb{z})]^{-1}.\tag{A.8}$$

Lemma A.3 now follows from Lemma A.2 and the delta method.

Lemma A.3 *Suppose* $\mathbf{Z}$ *satisfies (A.7) and suppose* ***C2***–***C6*** *in Section 2 hold. Then,*

$$n^{1/2}\{\widehat{\mathcal{Q}}_n(\mathbb{z})-\{\Theta(\mathbb{z},\gamma)-1\}I_k\}\xrightarrow{D}\Psi^{(3)}(\mathbb{z}),\qquad \mathbb{z}\in\mathcal{C},$$

where $\Psi^{(3)}(\mathbb{z})=[\Psi^{(3)}(\mathbb{z})]_{ij}$ *is a* $k\times k$ *symmetric Gaussian matrix process with zero mean and covariance, such that for* $i\leqslant j$, $i'\leqslant j'$,

$$\mathbb{E}[\Psi^{(3)}(\mathbb{z}_1)]_{ii}[\Psi^{(3)}(\mathbb{z}_2)]_{ii}=2\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma),$$

$$\mathbb{E}[\Psi^{(3)}(\mathbb{z}_1)]_{ij}[\Psi^{(3)}(\mathbb{z}_2)]_{ij} = \delta(\mathbb{z}_1,\mathbb{z}_2,\gamma), \quad \text{if } i \neq j,$$
$$\mathbb{E}[\Psi^{(3)}(\mathbb{z}_1)]_{ij}[\Psi^{(3)}(\mathbb{z}_2)]_{i'j'} = 0, \quad \text{if } i \neq i' \text{ or } j \neq j'.$$

The asymptotic normality of $\widehat{\mathcal{Q}}_n(\mathbb{z})$ follows since it is a smooth function of $\widetilde{\mathcal{Q}}_n(\mathbb{z})$. The calculation of the covariance kernel of $\Psi^{(3)}(\mathbb{z})$ is basic, though tedious, and hence details are omitted.

To smooth $\xi_n(\mathbb{z},\alpha,\eta)$ in the same way as $\widehat{\mathcal{Q}}_n(\mathbb{z})$, define

$$\widehat{\xi}_n(\mathbb{z},\alpha,\eta) = \xi_n(\mathbb{z},\alpha,\eta), \qquad \mathbb{z} \in \mathcal{C}^+,$$
$$\widehat{\xi}_n(\mathbb{z},\alpha,\eta) = \frac{\rho_n - v}{2\rho_n}\xi_n(u+i\rho_n,\alpha,\eta) + \frac{v+\rho_n}{2\rho_n}\xi_n(u-i\rho_n,\alpha,\eta), \qquad \mathbb{z} \in \mathcal{C}\backslash\mathcal{C}^+.$$

Note that $\widehat{\xi}_n(\mathbb{z},\alpha,\eta) = n^{1/2}\alpha^T V_n^T \widehat{\mathcal{Q}}_n(\mathbb{z}) V_n \eta$ and that V_n has orthonormal columns.

Lemma A.4 *Suppose that* $\mathbf{Z}$ *satisfies (A.7) and **C2**–**C6** hold. Then,*

$$\widehat{\xi}_n(\mathbb{z},\alpha,\eta) - n^{1/2}(\Theta(\mathbb{z},\gamma)-1)\alpha^T\eta \xrightarrow{D} \Psi^{(4)}(\mathbb{z}),$$

where $\Psi^{(4)}(\mathbb{z})$ *is a Gaussian process with zero mean and covariance function*

$$\Gamma^{(2)}(\mathbb{z}_1,\mathbb{z}_2) = \delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)[\|\alpha\|^2\|\eta\|^2 + (\alpha^T\eta)^2].$$

The following result is an immediate consequence of the foregoing:

$$\oint_{\mathcal{C}} \frac{f(\mathbb{z})\widehat{\xi}_n(\mathbb{z},\alpha,\eta)}{-2\pi i} d\mathbb{z} - n^{1/2}\Omega(f,\gamma)\alpha^T\eta \Longrightarrow \mathcal{N}(0, [\|\alpha\|^2\|\eta\|^2 + (\alpha^T\eta)^2]\Delta(f,\gamma)).$$

In Section S.6 of the Supplementary Material (see Lemma S.4 and (S.6.2) for details), we verify that, if we replace $\widehat{\xi}_n(\mathbb{z},\alpha,\eta)$ with $\xi_n(\mathbb{z},\alpha,\eta)$, and (A.7) with **C1**, the above result continues to hold.

Since $\sqrt{n}\alpha^T\mathbf{M}(f)\eta = (-2\pi i)^{-1}\oint_{\mathcal{C}} f(\mathbb{z})\xi_n(\mathbb{z},\alpha,\eta)d\mathbb{z}$ when $\lambda_{\max}(\widehat{\mathbf{\Sigma}}_p) < \overline{u}$ and (A.4) holds, the proof of Theorem 2.1 is complete.

A.2. Proof of Theorem 2.3

Define $T_n = C^T(n^{-1}XX^T)^{-1}C$. Then,

$$\begin{aligned}\sqrt{n}\mathbf{M}(f) =& \frac{1}{\sqrt{n}} Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}Q_n \\ &+ Q_n^T\mathbf{Z}^T\Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p)BCT_n^{-1/2} + \left(Q_n^T\mathbf{Z}^T\Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p)BCT_n^{-1/2}\right)^T \\ &+ \sqrt{n}T_n^{-1/2}C^TB^Tf(\widehat{\mathbf{\Sigma}}_p)BCT_n^{-1/2}.\end{aligned}$$

In view of Theorem 2.1 and Lemma 2.2 (whose proof is presented in Section S.5 of the Supplementary Material), we only need to show that under **C1**,

$$\left\| Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p) BC T_n^{-1/2} \right\|_2 \xrightarrow{P} 0, \tag{A.9}$$

$$\left\| \sqrt{n} T_n^{-1/2} C^T B^T f(\widehat{\mathbf{\Sigma}}_p) BC T_n^{-1/2} - \mathcal{H}(D, f) \right\|_2 \xrightarrow{P} 0. \tag{A.10}$$

For future use, we also show that the convergence is uniform over the class

$$\{B \in \mathbb{R}^{p\times k}, C \in \mathbb{R}^{k\times q} \colon \sqrt{n}\|BC\|_2^2 \leqslant \mathcal{K}\}, \qquad \text{for arbitrary } \mathcal{K} > 0.$$

Similar to the strategy in the proof of Theorem 2.1, we first consider z_{ij}'s satisfying (A.7). Observe when $\lambda_{\max}(\widetilde{\mathbf{\Sigma}}_p) < \mathfrak{D}$ for any $\mathfrak{D} \in (\limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2,\ \overline{u})$,

$$\begin{aligned}
&\|Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p) BC T_n^{-1/2}\|_2 \\
&= \Big\| \oint_{\mathcal{C}} \frac{-1}{2\pi i} f(\mathbb{z}) Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widehat{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC T_n^{-1/2} d\mathbb{z} \Big\|_2 \\
&\leqslant \Big\| \oint_{\mathcal{C}^+} \frac{-1}{2\pi i} f(\mathbb{z}) Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widehat{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC T_n^{-1/2} d\mathbb{z} \Big\|_2 + w_{1,n}\|BC\|_2,
\end{aligned}$$

where $w_{1,n} \to 0$ is a deterministic sequence. The last step is due to

$$\begin{aligned}
&\Big\| \int_{\mathcal{C}\backslash\mathcal{C}^+} \frac{-1}{2\pi i} f(\mathbb{z}) Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widehat{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC T_n^{-1/2} d\mathbb{z} \Big\|_2 \\
&\leqslant \mathcal{K} \rho_n \|T_n^{-1/2}\|_2 \ \|\mathbf{Z}^T \Sigma_p^{T/2}\|_2 \left[(\overline{u} - \mathfrak{D})^{-1} + |\underline{u}|^{-1} \right] \|BC\|_2 \\
&\leqslant \mathcal{K} \rho_n n^{1/2} \ \|T_n^{-1/2}\|_2 \ \mathfrak{D}^{1/2} \left[(\overline{u} - \mathfrak{D})^{-1} + |\underline{u}|^{-1} \right] \|BC\|_2,
\end{aligned}$$

where $\mathcal{K}$ is a universal constant.

Next, using Lemma S.6 of the Supplementary Material,

$$Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widehat{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC = V_n^T (I - \widetilde{\mathcal{Q}}_n(\mathbb{z}))^{-1} U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC.$$

In the following, $|d\mathbb{z}|$ refers to the differential form of the integral with respect to the length of a contour.

For any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\|Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p) BC T_n^{-1/2}\|_2 > \varepsilon \Big) \\
&\leqslant \mathbb{P}\Big(\Big\{ \|Q_n^T \mathbf{Z}^T \Sigma_p^{T/2} f(\widehat{\mathbf{\Sigma}}_p) BC T_n^{-1/2}\|_2 > \varepsilon \Big\} \bigcap \Big\{ \lambda_{\max}(\widetilde{\mathbf{\Sigma}}_p) < \mathfrak{D} \Big\} \Big) + \mathbb{P}\Big(\lambda_{\max}(\widetilde{\mathbf{\Sigma}}_p) \geqslant \mathfrak{D} \Big) \\
&\leqslant \mathbb{P}\Big(\Big\| \oint_{\mathcal{C}^+} \frac{-1}{2\pi i} f(\mathbb{z}) V_n^T (I - \widetilde{\mathcal{Q}}_n(\mathbb{z}))^{-1} U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC T_n^{-1/2} d\mathbb{z} \Big\|_2 \\
&\qquad > \varepsilon - w_{1,n}\|BC\|_2 \Big) + w_{2,n}
\end{aligned}$$

$$\begin{aligned}
&\leqslant \mathbb{P}\Big(\oint_{\mathcal{C}^+} \|(I-\widetilde{\mathcal{Q}}(\mathbb{z}))^{-1}\|_F^2 |d\mathbb{z}| \oint_{\mathcal{C}^+} \|U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} BCT_n^{-1/2}\|_F^2 |d\mathbb{z}| \\
&\qquad > \mathcal{K}_1 \{\varepsilon - w_{1,n}\|BC\|_2\}^2 \Big) + w_{2,n} \\
&\leqslant \mathbb{P}\Big(\oint_{\mathcal{C}^+} \|U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} BCT_n^{-1/2}\|_F^2 |d\mathbb{z}| > \frac{\mathcal{K}_1}{\mathcal{K}_2} \{\varepsilon - w_{1,n}\|BC\|_2\}^2 \Big) \\
&\qquad + \mathbb{P}\Big(\oint_{\mathcal{C}^+} \|(I-\widetilde{\mathcal{Q}}(\mathbb{z}))^{-1}\|_F^2 |d\mathbb{z}| > \mathcal{K}_2 \Big) + w_{2,n} \\
&\leqslant \frac{\sup_{\mathbb{z}\in\mathcal{C}^+} \mathbb{E}\|U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} BCT_n^{-1/2}\|_F^2}{\mathcal{K}_4 \{\varepsilon - w_{1,n}\|BC\|_2\}^2} + w_{3,n} + w_{2,n}.
\end{aligned}$$

Here, $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3, \mathcal{K}_4$ are appropriately large constants independent of BC, and

$$\begin{aligned}
w_{2,n} &= \mathbb{P}\Big(\lambda_{\max}(\widetilde{\mathbf{\Sigma}}_p) \geqslant \mathfrak{D} \Big) \to 0, \quad \text{due to Lemma A.1}, \\
w_{3,n} &= \mathbb{P}\Big(\oint_{\mathcal{C}^+} \|(I-\widetilde{\mathcal{Q}}(\mathbb{z}))^{-1}\|_F^2 |d\mathbb{z}| \geqslant \mathcal{K}_2 \Big) \to 0.
\end{aligned} \tag{A.11}$$

Note that (A.11) follows because $(I-\widetilde{\mathcal{Q}}(\mathbb{z}))^{-1} \xrightarrow{P} \Theta(\mathbb{z},\gamma)I_p$ (Lemma A.2) and that $\Theta(\mathbb{z},\gamma)$ is bounded on $\mathcal{C}$ (shown in Bai and Silverstein (2004)).

We claim, with proof presented in Section S.4 of the Supplementary Material, that

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \mathbb{E}\Big[\|U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} BCT_n^{-1/2}\|_F^2 \Big] \leqslant \mathcal{K}\|BC\|_2^2, \tag{A.12}$$

for a sufficiently large $\mathcal{K}$.

Now (A.9) follows when z_{ij}'s satisfy (A.7), since (2.3) implies $\|BC\|_2^2 \to 0$, as $n\to\infty$. In Section S.6 of the Supplementary Material, we will show that the difference between the left-hand side of (A.9) with and without variable truncation converges to 0 in probability. The convergence is also uniform on $BC \in \{\sqrt{n}\|BC\|_2^2 \leqslant \mathcal{K}\}$. It completes the proof of (A.9).

As for (A.10), again using Lemma S.6 of the Supplementary Material,

$$\begin{aligned}
C^T B^T (\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} BC = {} & C^T B^T (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} BC \\
& + \frac{1}{n} C^T B^T (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} \Sigma_p^{1/2} \mathbf{Z} U_n (I-\widetilde{\mathcal{Q}}(\mathbb{z}))^{-1} U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} BC.
\end{aligned}$$

Using analogous arguments, for some deterministic sequence $w_{4,n}\to 0$ and constants $\mathcal{K}_5, \mathcal{K}_6$,

$$\begin{aligned}
&\mathbb{P}\Big(\|\sqrt{n} T_n^{-1/2} C^T B^T f(\widehat{\mathbf{\Sigma}}_p) BCT_n^{-1/2} - \mathcal{H}(D,f)\|_2 > \varepsilon \Big) \\
&\leqslant \mathbb{P}\Big(\Big\| \oint_{\mathcal{C}^+} \frac{-1}{2\pi i} f(\mathbb{z}) \sqrt{n} T_n^{-1/2} C^T B^T (\widehat{\mathbf{\Sigma}}_p - \mathbb{z}I)^{-1} BCT_n^{-1/2} d\mathbb{z} - \mathcal{H}(D,f) \Big\|_2 \\
&\qquad > \varepsilon - w_{4,n}\|BC\|_2^2 \Big) + w_{2,n}
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \mathbb{P}\Big(\oint_{\mathcal{C}^+} \Big\| \sqrt{n} C^T B^T (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC - D(\mathbb{z},\gamma) \Big\|_2 |d\mathbb{z}| > \frac{\mathcal{K}_5}{2} \{\varepsilon - w_{4,n} \|BC\|_2^2\} \Big) \\
&\quad + \mathbb{P}\Big(\frac{1}{\sqrt{n}} \oint_{\mathcal{C}^+} \|(I - \widetilde{\mathcal{Q}}(\mathbb{z}))^{-1}\|_2 \|U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC\|_2^2 |d\mathbb{z}| > \frac{\mathcal{K}_5}{2} \{\varepsilon - \\
&w_{4,n} \|BC\|_2^2\} \Big) + w_{2,n} \\
&\leqslant \mathbb{P}\Big(\oint_{\mathcal{C}^+} \Big\| \sqrt{n} C^T B^T (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC - D(\mathbb{z},\gamma) \Big\|_2 |d\mathbb{z}| > \frac{\mathcal{K}_5}{2} \{\varepsilon - w_{4,n} \|BC\|_2^2\} \Big) \\
&\quad + \mathbb{P}\Big(\frac{1}{\sqrt{n}} \oint_{\mathcal{C}^+} \|U_n^T \mathbf{Z}^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC\|_2^2 |d\mathbb{z}| > \frac{\mathcal{K}_5}{2\mathcal{K}_6} \{\varepsilon - w_{4,n} \|BC\|_2^2\} \Big) \\
&\quad + \mathbb{P}\Big(\sup_{\mathbb{z} \in \mathcal{C}^+} \|(I - \widetilde{\mathcal{Q}}(\mathbb{z}))^{-1}\|_2 \geqslant \mathcal{K}_6 \Big) + w_{2,n}.
\end{aligned}
$$

The same arguments as those in proving (A.11) imply that, for sufficiently large $\mathcal{K}_6$,

$$
\mathbb{P}\Big(\sup_{\mathbb{z} \in \mathcal{C}^+} \|(I - \widetilde{\mathcal{Q}}(\mathbb{z}))^{-1}\|_2 \geqslant \mathcal{K}_6 \Big) \to 0.
$$

We only need to show

$$
\sup_{\mathbb{z} \in \mathcal{C}^+} \mathbb{E}\Big[\Big\| \sqrt{n} C^T B^T (\widetilde{\mathbf{\Sigma}}_p - \mathbb{z} I)^{-1} BC - D(\mathbb{z},\gamma) \Big\|_F^2 \Big] \leqslant \mathcal{K} \|BC\|_2^4, \tag{A.13}
$$

The proof is given in Section S.4 of the Supplementary Material. That (A.10) holds under **C1** (without variable truncation) will be addressed in Section S.6 of the Supplementary Material. The results indicate the convergence in (A.10) is uniform on $BC \in \{\sqrt{n}\|BC\|_2^2 \leqslant \mathcal{K}\}$.

A.3. Proof of Theorem 2.4

Throughout this section, $\mathbb{P}_*$ is the prior probability measure of μ, $\mathbb{P}_{BC}$ the probability measure of the observations conditional on BC, and $\mathbb{P}_{\mathbf{Z}}$ the probability measure of $\mathbf{Z}$.

Under the probabilistic local alternative model (2.6),

$$
\sqrt{n}\|BC\|_2^2 = O_p(1). \tag{A.14}
$$

Using Lemma S.13 and Lemma S.17 of the Supplementary Material, we can prove

$$
\sqrt{n} C^T B^T [\Theta^{-1}(\mathbb{z},\gamma)\Sigma_p - \mathbb{z} I]^{-1} BC - D_{\mathbb{P}_*}(\mathbb{z},\gamma) \xrightarrow{\mathbb{P}_*} 0, \qquad \text{pointwise for } \mathbb{z} \in \mathcal{C}, \tag{A.15}
$$

where $D_{\mathbb{P}_*}(\mathbb{z},\gamma) = h(\mathbb{z},\gamma)\mathcal{S}\mathcal{S}^T$.

In what follows, we only consider the (regularized version of) the LR criterion and show the convergence of Υ^{LR}. The convergence of Υ^{LH} and Υ^{BNP} can be proved using analogous arguments. To verify the theorem, it suffices to show that for any $\epsilon > 0$ and any $\zeta > 0$, there exists a sufficiently large N_0, such that when $n > N_0$,

$$
\mathbb{P}_*\Big(\Big| \Upsilon^{\mathrm{LR}}(BC, f) - \Phi\Big(-\xi_\alpha + \frac{\mathrm{tr}(\mathcal{H}(D_{\mathbb{P}_*}, f))}{q^{1/2}\Delta^{1/2}(f,\gamma)} \Big) \Big| > \epsilon \Big) < \zeta.
$$

Define $g = \{1 + \Omega(f,\gamma)\}\Delta^{-1/2}(f,\gamma)$ and its empirical counterpart $g_n = \{1 + \widehat{\Omega}(f,\gamma_n)\}$ $\widehat{\Delta}^{-1/2}(f,\gamma_n)$. Lemma 2.2 implies that $g_n \to g$ in probability. Recall

$$\mathcal{H}(D,f) = T^{-1/2}[(-2\pi i)^{-1}\oint_{\mathcal{C}} f(\mathbb{z})D(\mathbb{z},\gamma)d\mathbb{z}]T^{-1/2}.$$

Write

$$\begin{aligned}\mathbf{M}(f) =& \frac{1}{n}Q_n^T\mathbf{Z}^T\Sigma_p^{T/2}f(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}Q_n + \frac{g}{g_n\sqrt{n}}\mathcal{H}(D_{\mathbb{P}_*},f) + \frac{1}{g_n\sqrt{n}}\sigma_n(BC)\\ &+ \frac{1}{g_n\sqrt{n}}\sum_{i=1}^4 \eta^{(i)}(BC,\mathbf{Z}),\end{aligned}$$

where with notation $\mathcal{W}(f,\Sigma_p,\gamma) = (-2\pi i)^{-1}\oint_{\mathcal{C}} f(\mathbb{z})(\Theta(\mathbb{z},\gamma)\Sigma_p - \mathbb{z}I)^{-1}d\mathbb{z}$,

$$\begin{aligned}\sigma_n(BC) &= g\Big[\sqrt{n}T_n^{-1/2}C^TB^T\mathcal{W}(f,\Sigma_p,\gamma)BCT_n^{-1/2} - \mathcal{H}(D_{\mathbb{P}_*},f)\Big],\\ \eta^{(1)}(BC,\mathbf{Z}) &= [g_n - g]\sqrt{n}T_n^{-1/2}C^TB^T\mathcal{W}(f,\Sigma_p,\gamma)BCT_n^{-1/2},\\ \eta^{(2)}(BC,\mathbf{Z}) &= g_n\sqrt{n}T_n^{-1/2}C^TB^T\Big[f(\widehat{\mathbf{\Sigma}}_p) - \mathcal{W}(f,\Sigma_p,\gamma)\Big]BCT_n^{-1/2},\\ \eta^{(3)}(BC,\mathbf{Z}) &= g_nQ_n^T\mathbf{Z}^T\Sigma_p^{T/2}f(\widehat{\mathbf{\Sigma}}_p)BCT_n^{-1/2},\\ \eta^{(4)}(BC,\mathbf{Z}) &= g_nT_n^{-1/2}C^TB^Tf(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}Q_n.\end{aligned}$$

Therefore, by Lemma S.7, for $i = 1,2,\dots,q$,

$$\begin{aligned}&\Big|\lambda_i(\mathbf{M}(f)) - \lambda_i\Big(\frac{1}{n}Q_n^T\mathbf{Z}^T\Sigma_p^{T/2}f(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}Q_n + \frac{g}{g_n\sqrt{n}}\mathcal{H}(D_{\mathbb{P}_*},f)\Big)\Big|\\ &\qquad\leqslant \frac{1}{g_n\sqrt{n}}\|\sigma_n(BC)\|_2 + \frac{1}{g_n\sqrt{n}}\sum_{i=1}^4\|\eta^{(i)}(BC,\mathbf{Z})\|_2.\end{aligned}$$

Since the function $\log(1+x)$ is 1-Lipschitz when $x > 0$,

$$\sqrt{n}\frac{g_n}{q^{1/2}}\Big|T^{\mathrm{LR}}(f) - \sum_{i=1}^q \log\Big[1 + \lambda_i\Big(\frac{1}{n}Q_n^T\mathbf{Z}^T\Sigma_p^{T/2}f(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}Q_n\frac{g}{g_n\sqrt{n}}\mathcal{H}(D_{\mathbb{P}_*},f)\Big)\Big]\Big| \tag{A.16}$$

$$\leqslant q^{1/2}\|\sigma_n(BC)\|_2 + q^{1/2}\sum_{i=1}^4\|\eta^{(i)}(BC,\mathbf{Z})\|_2. \tag{A.17}$$

Define

$$\tilde{\Upsilon}(\xi) = \mathbb{P}_{\mathbf{Z}}\Big(\sqrt{n}\frac{g_n}{q^{1/2}}\sum_{i=1}^q \log\Big[1 + \lambda_i\Big(\frac{1}{n}Q_n^T\mathbf{Z}^T\Sigma_p^{T/2}f(\widehat{\mathbf{\Sigma}}_p)\Sigma_p^{1/2}\mathbf{Z}Q_n + \frac{g}{g_n\sqrt{n}}\mathcal{H}(D_{\mathbb{P}_*},f)\Big)\Big]$$

$$- \sqrt{n} g_n q^{1/2} \log(1 + \widehat{\Omega}(f, \gamma)) > \xi \Big)$$

and note that this quantity is independent of BC. By Theorem 2.1, Lemma 2.2 and an application of the delta method, for any fixed ξ,

$$\tilde{\Upsilon}(\xi) \to \Phi\left(-\xi + \frac{\operatorname{tr}(\mathcal{H}(D_{\mathbb{P}_*}, f))}{q^{1/2}\Delta^{1/2}(f, \gamma)}\right).$$

Hence, for any $\epsilon > 0$, we can find a sufficiently large N_1 such that when $n > N_1$,

$$\tilde{\Upsilon}(\xi_\alpha - 5q^{1/2}\epsilon) < \Phi\left(-\xi_\alpha + \frac{\operatorname{tr}(\mathcal{H}(D_{\mathbb{P}_*}, f))}{q^{1/2}\Delta^{1/2}(f, \gamma)} + 5q^{1/2}\epsilon\right) + \epsilon,$$

$$\tilde{\Upsilon}(\xi_\alpha + 5q^{1/2}\epsilon) > \Phi\left(-\xi_\alpha + \frac{\operatorname{tr}(\mathcal{H}(D_{\mathbb{P}_*}, f))}{q^{1/2}\Delta^{1/2}(f, \gamma)} - 5q^{1/2}\epsilon\right) - \epsilon.$$

Due to (A.14) and (A.15), there exists a constant $\mathcal{K}_\zeta$ and a sufficiently large N_2 such that when $n > N_2$, $\mathbb{P}_*(K^{(1)}) > 1 - \zeta$, where

$$K^{(1)} = \{BC\colon \sqrt{n}\|BC\|_2^2 \leqslant \mathcal{K}_\zeta\} \cap \{BC : |\sigma_n(BC)| \leqslant \epsilon\}.$$

Using the arguments in the proof of Theorem A.2, we have, for $i = 2, 3, 4$, uniformly on $K^{(1)}$,

$$\eta^{(i)}(BC, \mathbf{Z}) \xrightarrow{\mathbb{P}_{BC}} 0.$$

This convergence to zero is also valid for $\eta^{(1)}(BC, \mathbf{Z})$, since $g_n \xrightarrow{\mathbb{P}_{\mathbf{Z}}} g$ due to Lemma 2.2 and g_n is independent of BC. Therefore, we can find a sufficiently large N_3, such that when $n > N_3$, for any $BC \in K^{(1)}$, the event

$$K^{(2)}(BC) = \{\mathbf{Z}\colon \max_{i=1,\ldots,4} \|\eta^{(i)}(BC, \mathbf{Z})\|_2 \leqslant \epsilon\}$$

has measure at least $1-\epsilon$, that is, $\mathbb{P}_{BC}(K^{(2)}(BC)) > 1 - \epsilon$. Therefore, when $BC \in K^{(1)}$, and $n > \max(N_1, N_2, N_3)$, (A.17) implies

$$\begin{aligned}
\Upsilon^{\mathrm{LR}}(BC, f) &\leqslant \mathbb{P}_{BC}\Big(\Big\{\widehat{T}^{\mathrm{LR}}(f) > \xi_\alpha\Big\} \cap K^{(2)}(BC)\Big) + \mathbb{P}_{BC}(K^{(2)}(BC)^c) \\
&\leqslant \tilde{\Upsilon}(\xi_\alpha - 5q^{1/2}\epsilon) + \epsilon \\
&\leqslant \Phi\Big(- \xi_\alpha + \frac{\operatorname{tr}(\mathcal{H}(D_{\mathbb{P}_*}, f))}{q^{1/2}\Delta^{1/2}(f, \gamma)} + 5q^{1/2}\epsilon\Big) + 2\epsilon.
\end{aligned}$$

Conversely,

$$\begin{aligned}
\Upsilon^{\mathrm{LR}}(BC, f) &\geqslant \mathbb{P}_{BC}\Big(\Big\{\widehat{T}^{\mathrm{LR}}(f) > \xi_\alpha\Big\} \cap K^{(2)}(BC)\Big) \\
&\geqslant \tilde{\Upsilon}(\xi_\alpha + 5q^{1/2}\epsilon) \\
&\geqslant \Phi(-\xi_\alpha + \frac{\operatorname{tr}(\mathcal{H}(D_{\mathbb{P}_*}, f))}{q^{1/2}\Delta^{1/2}(f, \gamma)} - 5q^{1/2}\epsilon) - \epsilon.
\end{aligned}$$

This completes the proof, since $\mathbb{P}_*(K^{(1)}) > 1 - \zeta$.

Proof of Theorem 3.1

To show Result (1), first note that $\Delta(f_\ell, \gamma)$ is a continuous function of ℓ under Condition (ii) of $\mathfrak{F}$. It is also assumed that $\inf_{\ell\in\mathcal{L}} \Delta(f_\ell, \gamma) > 0$ in Condition (iv). Hence, $\Delta(f_\ell, \gamma)$ is bounded away from infinity and 0 on $\mathcal{L}$. Moreover, Conditions (i)–(iii) of $\mathfrak{F}$ imply

$$\sup_{\mathbb{z}\in\mathcal{C}} \sup_{\ell\in\mathcal{L}} |f_\ell(\mathbb{z})| < \infty.$$

In the proof of Lemma 2.2, presented in Section S.5 of the Supplementary Material, it is shown that

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \sqrt{n}|m_{n,p}(\mathbb{z}) - m(\mathbb{z})| \xrightarrow{P} 0,$$
$$\sup_{\mathbb{z}\in\mathcal{C}^+} \sqrt{n}|m'_{n,p}(\mathbb{z}) - m'(\mathbb{z})| \xrightarrow{P} 0.$$

It follows that

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \sqrt{n}|\widehat{h}(\mathbb{z}, \gamma_n) - h(\mathbb{z}, \gamma)| \xrightarrow{P} 0.$$

Hence,

$$\sup_{\ell\in\mathcal{L}} \sqrt{n}\Big| \oint_{\mathcal{C}^+} f_\ell(\mathbb{z})[\widehat{h}(\mathbb{z}, \gamma_n) - h(\mathbb{z}, \gamma)]d\mathbb{z}\Big| \xrightarrow{P} 0.$$

The boundedness of $\Theta(\mathbb{z}, \gamma)$ is deduced in Section S.2 of the Supplementary Material. It can be checked that $m_{n,p}(\mathbb{z})$ and $\widehat{\Theta}(\mathbb{z}, \gamma_n)$ are also bounded on $\mathcal{C}$ when $\lambda_{\max}(\widehat{\boldsymbol{\Sigma}}_p) < \mathfrak{D} < \overline{u}$. Since $\sqrt{n}\rho_n \to 0$, it follows that

$$\sup_{\ell\in\mathcal{L}} \sqrt{n}\Big| \oint_{\mathcal{C}\backslash\mathcal{C}^+} f_\ell(\mathbb{z})\widehat{h}(\mathbb{z}, \gamma_n)d\mathbb{z}\Big| \xrightarrow{P} 0,$$
$$\sup_{\ell\in\mathcal{L}} \sqrt{n}\Big| \oint_{\mathcal{C}\backslash\mathcal{C}^+} f_\ell(\mathbb{z})h(\mathbb{z}, \gamma_n)d\mathbb{z}\Big| \xrightarrow{P} 0.$$

The convergence of $\widehat{\Delta}(f_\ell, \gamma_n)$ is stated in Lemma 2.2. The proof of Lemma 2.2 reveals that the convergence of $\widehat{\Delta}(f_\ell, \gamma_n)$ follows from the uniform convergence of $\widehat{\delta}(\mathbb{z}_1, \mathbb{z}_2, \gamma_n)$ on $(\mathcal{C}^+)^2$. It implies that the convergence of $\widehat{\Delta}(f_\ell, \gamma_n)$ is uniform on $f_\ell \in \mathfrak{F}$, because

$$\sqrt{n}|\widehat{\Delta}(f_\ell, \gamma_n) - \Delta(f_\ell, \gamma)| \leqslant \mathcal{K} \sup_{\mathbb{z}\in\mathcal{C}} \sup_{\ell\in\mathcal{L}} |f_\ell(\mathbb{z})| \oint_{\mathcal{C}^2} \sqrt{n}|\widehat{\delta}(\mathbb{z}_1, \mathbb{z}_2, \gamma_n) - \delta(\mathbb{z}_1, \mathbb{z}_2, \gamma)||d\mathbb{z}_1||d\mathbb{z}_2|.$$

We therefore have

$$\sup_{\ell\in\mathcal{L}} \sqrt{n}|\widehat{\Delta}(f_\ell, \gamma_n) - \Delta(f_\ell, \gamma)| \xrightarrow{P} 0$$

and the proof of result (1) is complete.

As for (3.3), we only need to show for any $\varepsilon > 0$, there exists a constant $\mathcal{K}_\varepsilon > 0$ and an integer n_ε, such that for $t = \mathcal{K}_\varepsilon n^{-1/4}$ and any $n \geqslant n_\varepsilon$,

$$\mathbb{P}\Big(\widehat{\Xi}(\ell^*, \widehat{h}, \gamma_n) - \widehat{\Xi}(\ell^* + t\delta, \widehat{h}, \gamma_n) \geqslant 0 \text{ for all } \delta \text{ s.t. } \|\delta\|_2 = 1 \text{ and } \ell^* + t\delta \in \mathcal{L}\Big) \geqslant \varepsilon.$$

Under condition (3.3),

$$\begin{aligned}\widehat{\Xi}(\ell^*, \widehat{h}, \gamma_n) - \widehat{\Xi}(\ell^* + t\delta, \widehat{h}, \gamma_n) &= \Xi(\ell^*, h, \gamma) - \Xi(\ell^* + t\delta, h, \gamma) + o_p(n^{-1/2}) \\ &\geqslant \mathcal{K}t^2\|\delta\|_2^2 + o_p(n^{-1/2}).\end{aligned}$$

The remainder $o_p(n^{-1/2})$ is uniform for any t and δ. Therefore, if $t = O(n^{-1/4})$, $\widehat{\Xi}(\ell^*, \widehat{h}, \gamma_n) - \widehat{\Xi}(\ell^* + t\delta, \widehat{h}, \gamma_n)$ is positive with high probability for any δ with L_2-norm 1.

As for (3.5), write $\ell = (l_1, \dots, l_r)^T$ for any $\ell \in \mathcal{L}$. Similarly denote the local maximizer of $\widehat{\Xi}(\ell, \widehat{h}, \gamma_n)$ by $\ell_n^* = (l_{1n}^*, \dots, l_{rn}^*)^T$ and let $\ell^* = (l_1^*, l_2^*, \dots, l_r^*)^T$. Since the partial derivatives of f_{ℓ^*} are analytic under Condition (iii) of $\mathfrak{F}$, due to Theorem 2.1 and (S.5.1),

$$\begin{aligned} n^{1/4}\Big[\mathbf{M}\Big(\frac{\partial f_{\ell^*}}{\partial l_j^*}\Big) - \widehat{\Omega}\Big(\frac{\partial f_{\ell^*}}{\partial l_j^*}, \gamma_n\Big)I_q\Big] &= o_p(1), \quad j = 1, \dots, r. \\ \Big[\mathbf{M}\Big(\frac{\partial^2 f_{\ell^*}}{\partial l_j^* \partial l_{j'}^*}\Big) - \widehat{\Omega}\Big(\frac{\partial^2 f_{\ell^*}}{\partial l_j^* \partial l_{j'}^*}, \gamma_n\Big)I_q\Big] &= o_p(1), \quad j, j' = 1, \dots, r.\end{aligned}$$

Because the third-order derivatives are continuous functions on $\mathcal{L} \otimes \mathcal{Z}$, we can find a constant $\mathcal{K}_\nabla$ such that

$$\max_{1 \leqslant j, j', j'' \leqslant r} \sup_{\ell \in \mathcal{L}} \sup_{\mathbb{z} \in \mathcal{Z}} \Big|\frac{\partial^3 f_\ell(\mathbb{z})}{\partial l_j \partial l_{j'} \partial l_{j''}}\Big| \leqslant \mathcal{K}_\nabla.$$

Since for the constant function $f_0(x) = \mathcal{K}_\nabla$, $\mathbf{M}(f_0) = O_p(1)$,

$$\max_{1 \leqslant j, j', j'' \leqslant r} \sup_{\ell \in \mathcal{L}} \Big\|\mathbf{M}\Big(\frac{\partial^3 f_\ell}{\partial l_j \partial l_{j'} \partial l_{j''}}\Big)\Big\|_2 = O_p(1).$$

Similarly, we have

$$\max_{1 \leqslant j, j', j'' \leqslant r} \sup_{\ell \in \mathcal{L}} \Big|\widehat{\Omega}\Big(\frac{\partial^3 f_\ell}{\partial l_j \partial l_{j'} \partial l_{j''}}, \gamma_n\Big)\Big| = O_p(1).$$

A Taylor expansion shows that

$$\begin{aligned} &\Big\|\sqrt{n}\mathbf{M}(f_{\ell_n^*}) - \widehat{\Omega}(f_{\ell_n^*}, \gamma_n)I_q\Big\} - \sqrt{n}\Big\{\mathbf{M}(f_{\ell^*}) - \widehat{\Omega}(f_{\ell^*}, \gamma_n)I_q\Big\|_2 \\ \leqslant&\sqrt{n}\Big\|\mathbf{M}\Big((\ell_n^* - \ell^*)^T\nabla_\ell f_{\ell^*}\Big) - \widehat{\Omega}\Big((\ell_n^* - \ell^*)^T\nabla_\ell f_{\ell^*}, \gamma_n\Big)I_q\Big\|_2 \\ &+ \sqrt{n}\Big\|\mathbf{M}\Big((\ell_n^* - \ell^*)^T\nabla_\ell^2 f_{\ell^*}(\ell_n^* - \ell^*)\Big) - \widehat{\Omega}\Big((\ell_n^* - \ell^*)^T\nabla_\ell^2 f_{\ell^*}(\ell_n^* - \ell^*), \gamma_n\Big)I_q\Big\|_2 \\ &+ \sqrt{n}\|\ell_n^* - \ell^*\|_2^3 \cdot O_p(1) \\ =&o_p(1).\end{aligned}$$

Moreover, $\widehat{\Delta}(f_{\ell_n^*}, \gamma_n) \xrightarrow{P} \Delta(f_{\ell^*}, \gamma)$ and (3.5) follows.

When ℓ^* is on the boundary of $\mathcal{L}$, we can prove (3.4) and (3.5) along similar lines and details are consequently omitted.

Proof of Theorem 3.2

Following from the argument in the proof of Theorem 3.1,

$$\widehat{T}(f_{\ell_{in}^*}) = \widehat{T}(f_{\ell_i^*}) + o_p(1),$$

by *Slutsky's Theorem*, we only need to show

$$\Big(\widehat{T}(f_{\ell_1^*}), \ldots, \widehat{T}(f_{\ell_m^*})\Big) \Longrightarrow \mathcal{N}\Big(0, \Delta^*\Big).$$

Observe that the asymptotic normality of $\widehat{T}(f_{\ell_i^*})$ follows from that of the un-standardized statistic, $T(f_{\ell_i^*})$. It suffices to show that $\Big(T(f_{\ell_1^*}), \ldots, T(f_{\ell_m^*})\Big)$ is asymptotically normal.

If the (regularized version) of LH criterion is being adopted, that is,

$$T(f_{\ell_i^*}) = T^{\mathrm{LH}}(f_{\ell_i^*}) = \sum_{j=1}^{q} \lambda_j(\mathbf{M}(f_{\ell_i^*})),$$

the joint normality of $\Big(T(f_{\ell_1^*}), \ldots, T(f_{\ell_m^*})\Big)$ follows from Theorem 2.2 and the fact that, for any linear combination of $\Big(T(f_{\ell_1^*}), \ldots, T(f_{\ell_m^*})\Big)$, say with coefficients $a_1, \ldots, a_m$,

$$\sum_{i=1}^{m} a_i T^{\mathrm{LH}}(f_{\ell_i^*}) = T^{\mathrm{LH}}\Big(\sum_{i=1}^{m} a_i f_{\ell_i^*}\Big).$$

Since $\Delta(\sum_{i=1}^m a_i f_{\ell_i^*}, \gamma) = \sum_{i=1}^m \sum_{j=1}^m a_i a_j \Delta(f_{\ell_i^*}, f_{\ell_j^*}, \gamma)$, the asymptotic covariance kernel of

$$\Big(\widehat{T}(f_{\ell_1^*}), \ldots, \widehat{T}(f_{\ell_m^*})\Big)$$

can be verified to be Δ^* via elementary calculation.

If LR or BNP criterion with regularization is being adopted, due to delta-method,

$$T^{\mathrm{LR}}(f_{\ell_i^*}) = q\log(1 + \widehat{\Omega}(f_{\ell_i^*}, \gamma_n)) + \frac{T^{\mathrm{LH}}(f_{\ell_i^*}) - q\widehat{\Omega}(f_{\ell_i^*}, \gamma_n)}{1 + \widehat{\Omega}(f_{\ell_i^*}, \gamma_n)} + o_p(n^{-1/2}),$$

$$T^{\mathrm{BNP}}(f_{\ell_i^*}) = \frac{q\widehat{\Omega}(f_{\ell_i^*}, \gamma_n)}{1 + \widehat{\Omega}(f_{\ell_i^*}, \gamma_n)} + \frac{T^{\mathrm{LH}}(f_{\ell_i^*}) - q\widehat{\Omega}(f_{\ell_i^*}, \gamma_n)}{\{1 + \widehat{\Omega}(f_{\ell_i^*}, \gamma_n)\}^2} + o_p(n^{-1/2}).$$

It implies that any linear combination of $T^{\mathrm{LR}}(f_{\ell_i^*})$ or $T^{\mathrm{BNP}}(f_{\ell_i^*})$ can be expressed as a linear combination of $T^{\mathrm{LH}}(f_{\ell_i^*})$ with a negligible remainder. The proof is complete.

Supplementary Material

Supplementary Material includes additional simulation results and detailed proofs of the main theoretical results presented in this paper.

Supplementary Material for "High Dimensional General Linear Hypothesis Tests via Non-linear Spectral Shrinkage"

Haoran Li, Alexander Aue and Debashis Paul

University of California, Davis

S.1. Additional material to Section 4.2

Recall notation in Section 2–Section 3 of the manuscript and recall the higher order shrinkage family proposed in Section 4.2,

$$\mathfrak{F}_{\text{high}} = \Big\{ f_\ell(x) = \Big[\sum_{j=0}^{\kappa} l_j x^j \Big]^{-1}, \quad \ell = (l_0, \dots, l_\kappa)^T \in \mathcal{G} \Big\}.$$

In this section, we introduce the selection of $\mathcal{G}$ and the application of Section 2 and Section 3 with $f \in \mathfrak{F}_{\text{high}}$ when $\kappa = 3$.

If f_ℓ has three distinct roots r_1, r_2, r_3 and non-zero leading coefficient l_3, we have the following representation

$$f_\ell(x) = \Big[\sum_{j=0}^{3} l_j x^j \Big]^{-1} = \sum_{j=1}^{3} \omega_j (x - r_j)^{-1}$$

where $\omega_j = l_3^{-1} \prod_{i \neq j} (r_j - r_i)^{-1}$. With Lemma 2.1, we have the following closed forms of $\Omega(f_\ell, \gamma)$ and $\Delta(f_\ell, \gamma)$.

$$\begin{aligned} \Omega(f_\ell, \gamma) &= \sum_{j=1}^{3} \omega_j \{\Theta(r_j, \gamma) - 1\}; \\ \Delta(f_\ell, \gamma) &= 2 \sum_{j=1}^{3} \sum_{j'=1}^{3} \omega_j \omega_{j'} \delta(r_j, r_{j'}, \gamma) = 2 \sum_{j=1}^{3} \sum_{j'=1}^{3} \omega_j \overline{\omega}_{j'} \delta(r_j, \overline{r}_{j'}, \gamma). \end{aligned}$$

where $\overline{\omega}$ is the complex conjugate of ω and $\overline{r}$ is analogous.

The case that f_ℓ has a multiple root, say r_1, is the limit when r_2 and/or r_3 converges to r_1. As shown in (4.1), in that situation, the decomposition of f_ℓ involves $g_1(x) = (x - r_1)^{-2}$ or $g_2(x) = (x - r_1)^{-3}$. Although similar closed forms of $\Omega(f_\ell, \gamma)$ and $\Delta(f_\ell, \gamma)$ are also available, they involve first-order or second-order derivatives of $\Theta(\mathbb{z}, \gamma)$ and $\delta(\mathbb{z}, \mathbb{z}', \gamma)$ with respect to $\mathbb{z}$. The estimation of $\Omega(f_\ell, \gamma)$ and $\Delta(f_\ell, \gamma)$ will be less precise. However, the test procedure wouldn't benefit much from allowing the existence of a multiple root. Because the asymptotic power $\Upsilon(BC, f_\ell)$ under the local alternatives is smooth with respect to r_1, r_2, r_3. In the following, we shall simply restrict $\mathcal{G}$ to exclude the case where f_ℓ has a multiple root by forcing Restriction R3 shown below.

As stated in Section 4.2, it is reasonable to require f_ℓ being strictly positive and monotonically decreasing on $\mathcal{X}$ for any $\ell \in \mathcal{G}$. Moreover, since multiplying f_ℓ by a constant leads to an equivalent test procedure, we can fix $l_0 = f_\ell(0) = 1$.

In summary, we set $\mathcal{G}$ to be the set of coefficients $\ell = (1, l_1, l_2, l_3)$ such that

R1 (Compactness): $|l_i| \leqslant c_i$, $i = 1, 2$, and $0 < \underline{c_3} \leqslant |l_3| \leqslant \overline{c_3}$;

R2 (Monotonicity): $3l_3x^2 + 2l_2x + l_1 \geqslant 0$ for all $x \in [0, \bar{\lambda}]$;

R3 (Distinct roots): $\Big|18l_3l_2l_1 - 4l_2^3 + l_2^2l_1^2 - 4l_3l_1^3 - 27l_3^2\Big| \geqslant c_4$;

where $c_1, c_2, \underline{c_3}, \overline{c_3}, c_4$ are pre-specified positive constants and $\bar{\lambda}$ is a constant such that $\bar{\lambda} \geqslant \limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2$. In practice, we choose $\bar{\lambda} \geqslant \lceil \lambda_{\max}(\widehat{\mathbf{\Sigma}}_p) \rceil + 0.01p^{-1}\mathrm{tr}(\widehat{\mathbf{\Sigma}}_p)$.

Under R1 and R2, there exists a constant $\mathcal{K}$ such that

$$\inf_{\ell \in \mathcal{G}} \inf_{x \in [0, \bar{\lambda}]} \min_{j=1,2,3} |x - r_j| > \mathcal{K}.$$

Under R3, the roots r_1, r_2, r_3 are mutually exclusive and we can find a constant $\mathcal{K}'$ such that

$$\inf_{\ell \in \mathcal{G}} \min_{1 \leqslant j \neq j' \leqslant 3} |r_j - r_{j'}| > \mathcal{K}'.$$

The considerations lead to the following algorithm to determine the shrinkage function $f \in \mathfrak{F}_{\mathrm{high}}$ given $\mathcal{G}$ and the test procedure with the selected shrinkage. The practical suggestion of $\mathcal{G}$ is provided after.

Algorithm S.1.1 (Test procedure with higher order shrinkage) *Perform the following steps.*

1. *Specify prior weights* $\tilde{t} = (t_0, t_1, t_2)$. *Canonical choices are* $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$.
2. *For each* $\ell = (l_3, l_2, l_1, l_0) \in \mathcal{G}$, *calculate roots* r_1, r_2, r_3 *of* $\sum_{j=0}^3 l_j x^j = 0$, *and weights* $\omega_j = l_3^{-1} \prod_{i \neq j}(r_j - r_i)^{-1}$, $j = 1, 2, 3$.
3. *For* $\ell \in \mathcal{G}$, *compute the estimates*

$$\begin{aligned}
m_{n,p}(r_j) &= \frac{1}{p}\mathrm{tr}[(\widehat{\mathbf{\Sigma}}_p - r_jI_p)^{-1}],\ j = 1, 2, 3;\\
\widehat{\Theta}(r_j, \gamma_n) &= \{1 - \gamma_n - \gamma_n r_j m_{n,p}(r_j)\}^{-1},\ j = 1, 2, 3;\\
\widehat{\delta}(r_j, r_{j'}, \gamma_n) &= \widehat{\Theta}(r_j, \gamma_n)\widehat{\Theta}(r_{j'}, \gamma_n)\Big[\frac{r_j\widehat{\Theta}(r_j, \gamma_n) - r_{j'}\widehat{\Theta}(r_{j'}, \gamma_n)}{r_j - r_{j'}} - 1\Big],\ j \neq j';\\
\widehat{\delta}(r_j, r_j, \gamma_n) &= \gamma_n\{1 + r_j m_{n,p}(r_j)\}\widehat{\Theta}^3(r_j, \gamma)\\
&\qquad + \gamma_n r_j\{m_{n,p}(r_j) + r_j m'_{n,p}(r_j)\}\widehat{\Theta}^4(r_j, \gamma_n),\ j = 1, 2, 3;\\
\hat{\rho}_0(r_j, \gamma_n) &= m_{n,p}(r_j), j = 1, 2, 3;\\
\hat{\rho}_1(r_j, \gamma_n) &= \widehat{\Theta}(r_j, \gamma_n)[1 + r_j m_{n,p}(r_j)],\ j = 1, 2, 3;\\
\hat{\rho}_2(r_j, \gamma_n) &= \widehat{\Theta}(r_j, \gamma_n)[p^{-1}\mathrm{tr}(\widehat{\mathbf{\Sigma}}_p) + r_j\hat{\rho}_1(r_j, \gamma)],\ j = 1, 2, 3;
\end{aligned}$$

$$\widehat{\Omega}_\ell(\gamma_n) = \sum_{j=1}^{3} \omega_j\{\widehat{\Theta}(r_j,\gamma_n) - 1\};$$

$$\widehat{\Delta}_\ell(\gamma_n) = 2\sum_{j=1}^{3}\sum_{j'=1}^{3} \omega_j\overline{\omega}_{j'}\widehat{\delta}(r_j,\overline{r}_{j'},\gamma_n);$$

$$\widehat{\Xi}_\ell(\gamma_n) = \widehat{\Delta}_\ell^{-1/2}(\gamma_n)\sum_{j=1}^{3}\sum_{i=0}^{2} \omega_j t_i \hat{\rho}_i(r_j,\gamma_n).$$

4. *Select $\ell^* = arg\max_{\ell\in\mathcal{G}} \widehat{\Xi}_\ell(\gamma_n)$ through a grid search.*
5. *The same as Step 5 of Algorithm 4.1, use one of the following standardized statistics to reject the null at asymptotic level α, if $\widehat{T} > \xi_\alpha$.*

$$\widehat{T}^{\mathrm{LR}}(\ell^*) := \frac{\sqrt{n}\{1+\widehat{\Omega}_{\ell*}(\gamma_n)\}}{q^{1/2}\widehat{\Delta}_{\ell*}^{1/2}(\gamma_n)}[T^{\mathrm{LR}}(\ell^*) - q\log\{1+\widehat{\Omega}_{\ell*}(\gamma_n)\}];$$

$$\widehat{T}^{\mathrm{LH}}(\ell^*) := \frac{\sqrt{n}}{q^{1/2}\widehat{\Delta}_{\ell*}^{1/2}(\gamma_n)}\{T^{\mathrm{LH}}(\ell^*) - q\widehat{\Omega}_{\ell*}(\gamma_n)\};$$

$$\widehat{T}^{\mathrm{BNP}}(\ell^*) := \frac{\sqrt{n}\{1+\widehat{\Omega}_{\ell*}(\gamma_n)\}^2}{q^{1/2}\widehat{\Delta}_{\ell*}^{1/2}(\gamma_n)}\{T^{\mathrm{BNP}}(\ell^*) - \frac{q\widehat{\Omega}_{\ell*}(\gamma_n)}{1+\widehat{\Omega}_{\ell*}(\gamma_n)}\}.$$

where

$$T^{\mathrm{LR}}(\ell^*) = \sum_{i=1}^{q}\log(1+\lambda_i),\ T^{\mathrm{LH}}(\ell^*) = \sum_{i=1}^{q}\lambda_i,\ T^{\mathrm{BNP}}(\ell^*) = \sum_{i=1}^{q}\frac{\lambda_i}{1+\lambda_i},$$

and $\{\lambda_i\}_{i=1}^q$ are eigenvalues of $n^{-1}Q_n^T\mathbf{Y}^T f_{\ell*}(\widehat{\mathbf{\Sigma}}_p)\mathbf{Y}Q_n$.

The suggested grid in practice is as following. Denote $p^{-1}\mathrm{tr}(\widehat{\mathbf{\Sigma}}_p)$ as $\mathfrak{M}$. First, generate $l_3 \in \pm[10^{-3}\bar{\lambda}^{-2}\mathfrak{M}^{-1},\ 10^2\bar{\lambda}^{-2}\mathfrak{M}^{-1}]$. Secondly, it is efficient to focus on l_2's such that the inflection point $x = l_2/(-3l_3)$ of the cubic equation is around $[0,\bar{\lambda}]$. Hence, for any l_3, we select l_2 to be such that $l_2/(-3l_3) \in [-0.1\bar{\lambda},\ 1.1\bar{\lambda}]$. Thirdly, to avoid f_ℓ^{-1} being too steep, we select l_1 to be such that $l_3\mathfrak{M}^2 + l_2\mathfrak{M} + l_1 = [f_\ell^{-1}(\mathfrak{M}) - f_\ell^{-1}(0)]/\mathfrak{M} \in [0,2]$. Valid grid points are those satisfying R2 and R3. c_4 can be arbitrarily small. We suggest $c_4 = 10^{-6}\bar{\lambda}^{-2}$.

S.2. About $m(\mathbb{z})$, $\Theta(\mathbb{z},\gamma)$ and $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$

Recall that $m(\mathbb{z})$ is the Stieltjes transform of F^∞, that is the limit almost surely of $F^{\widehat{\mathbf{\Sigma}}_p}$ at any point of continuity of F^∞. $m(\mathbb{z})$ is the unique solution in $\mathbb{C}^+$ of (2.1)

$$m(\mathbb{z}) = \int\frac{dL^\Sigma(\tau)}{\tau(1-\gamma-\gamma\mathbb{z}m(\mathbb{z})) - \mathbb{z}},$$

where L^{Σ} is the limit of F^{Σ_p}.

We provide another formulation of the Marčenko-Pastur equation (2.1) that is more convenient under some situations. Define

$$\underline{F}^{\infty}(\tau) = (1-\gamma)\mathbb{1}_{[0,\infty)}(\tau) + \gamma F^{\infty}(\tau).$$

$\underline{F}^{\infty}$ is a valid c.d.f. for any γ, that is a mixture of F^{∞} and a point mass at 0 (if $\gamma < 1$). It is actually the limit almost surely of the ESD of $n^{-1}\mathbf{Z}^T\Sigma_p\mathbf{Z}$. Denote the Stieltjes transform of $\underline{F}^{\infty}$ to be $\underline{m}(\mathbb{z})$. Then, there is a 1-1 mapping between $m(\mathbb{z})$ and $\underline{m}(\mathbb{z})$ as

$$\underline{m}(\mathbb{z}) = \frac{\gamma - 1}{\mathbb{z}} + \gamma m(\mathbb{z}).$$

The Marčenko-Pastur equation has an equivalent formulation as

$$\underline{m}(\mathbb{z}) = \Big[-\mathbb{z} + \gamma \int \frac{\tau dL^{\Sigma}(\tau)}{1+\tau\underline{m}(\mathbb{z})}\Big]^{-1}.$$

With the help of $\underline{m}(\mathbb{z})$, we now study the domain of $\Theta(\mathbb{z},\gamma)$ on any contour $\mathcal{C}$ enclosing $\mathcal{X} = [0, \limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2]$. Observe that $\Theta(\mathbb{z},\gamma) = (\mathbb{z}\underline{m}(\mathbb{z}))^{-1}$. It is claimed in Bai and Silverstein (2004) that $\inf_{\mathbb{z}\in S}|\underline{m}(\mathbb{z})| > 0$, for any bounded subset S of $\mathbb{C}$. Therefore,

$$\sup_{\mathbb{z}\in\mathcal{C}} |\Theta(\mathbb{z},\gamma)| < \infty.$$

Because the support of F^{∞}, $\mathrm{sp}(F^{\infty}) \subset \mathcal{X}$,

$$\sup_{\mathbb{z}\in\mathcal{C}} |m'(\mathbb{z})| = \sup_{\mathbb{z}\in\mathcal{C}} \Big| \int (\tau-\mathbb{z})^{-2} dF^{\infty}(\tau)\Big| < \infty,$$
$$\sup_{\mathbb{z}\in\mathcal{C}} |m''(\mathbb{z})| = \sup_{\mathbb{z}\in\mathcal{C}} \Big| \int 2(\tau-\mathbb{z})^{-3} dF^{\infty}(\tau)\Big| < \infty.$$

Hence,

$$\sup_{\mathbb{z}\in\mathcal{C}} \Big|\frac{\partial}{\partial \mathbb{z}}\Theta(\mathbb{z},\gamma)\Big| < \infty,$$
$$\sup_{\mathbb{z}_1,\mathbb{z}_2\in\mathcal{C}} \Big|\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)\Big| < \infty,$$
$$\sup_{\mathbb{z}_1,\mathbb{z}_2\in\mathcal{C}} \Big|\frac{\partial}{\partial \mathbb{z}_1}\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)\Big| < \infty.$$

Moreover,

$$\inf_{\mathbb{z}\in\mathcal{C}} \Big|\frac{\Theta(\mathbb{z},\gamma)-1}{\Theta(\mathbb{z},\gamma)} - 1\Big| = \inf_{\mathbb{z}\in\mathcal{C}} \Big|\Theta^{-1}(\mathbb{z},\gamma)\Big| = \inf_{\mathbb{z}\in\mathcal{C}} \Big|\mathbb{z}\underline{m}(\mathbb{z})\Big| > 0.$$

About convergence of F^{Σ_p} to L^{Σ} It is claimed in Section 2.1 that, pointwise almost surely on $\mathbb{z} \in \mathbb{C}^+$, $m_{n,p}(\mathbb{z})$ converges to $m(\mathbb{z})$. In view that the convergence of F^{Σ_p} to

L^{Σ} can be arbitrarily slow, the convergence rate of $m_{n,p}(\mathbb{z})$ to $m(\mathbb{z})$ is also arbitrarily slow. Consequently, the convergence rate of $\widehat{\Theta}(\mathbb{z},\gamma_n)$ and $\widehat{\delta}(\mathbb{z}_1,\mathbb{z}_2,\gamma_n)$ can be slower than $n^{-1/2}$, while a faster rate than $n^{-1/2}$ is required in this paper.

To solve this problem, it is typical (see for example Bai and Silverstein (2010)), to replace $m(\mathbb{z})$ with a deterministic sequence $\{m_p^0(\mathbb{z}),\ p=1,2,\dots\}$, that is the unique solution of

$$m_p^0(\mathbb{z}) = \int \frac{dF^{\Sigma_p}(\tau)}{\tau(1-\gamma_n-\gamma_n\mathbb{z}m_p^0(\mathbb{z}))-\mathbb{z}}.$$

Notice that the last equation is the Marčenko-Pastur equation with the population spectral distribution F^{Σ_p} replacing the limiting spectral distribution. The convergence rate of $m_{n,p}(\mathbb{z})-m_p^0(\mathbb{z})$ is shown to be $O(n^{-1})$ in Bai and Silverstein (2004). The result does not rely on the convergence rate of F^{Σ_p} to L^{Σ}, since $m_p^0(\mathbb{z})$ is free of L^{Σ}.

To emphasize readability and succinctness of the paper, we adopt another solution, that is to impose a convergence rate on F^{Σ_p} to L^{Σ} and on γ_n to γ, as shown in **C2** and **C4**. Later, we will frequently refer to existing results in literature that are established using $\{m_p^0(\mathbb{z}),\ p=1,2,\dots\}$. It is necessary to study the difference between $m_p^0(\mathbb{z})$ and $m(\mathbb{z})$ under **C2** and **C4**.

Similarly to $\underline{m}(\mathbb{z})$, define

$$\underline{m}_p^0(\mathbb{z}) = \frac{\gamma_n-1}{\mathbb{z}} + \gamma_n m_p^0(\mathbb{z}).$$

The following formulation also holds

$$\underline{m}_p^0(\mathbb{z}) = \Big[-\mathbb{z}+\gamma_n\int\frac{\tau dF^{\Sigma_p}(\tau)}{1+\tau\underline{m}_p^0(\mathbb{z})}\Big]^{-1}.$$

It is claimed in Bai and Silverstein (2004) (see (4.2)) that when F^{Σ_p} converges to L^{Σ} at any continuity point of L^{Σ}, for any $\mathcal{C}$ bounded away from the support of F^{∞}, as $n\to\infty$,

$$\sup_{\mathbb{z}\in\mathcal{C}}|\underline{m}_p^0(\mathbb{z})-\underline{m}(\mathbb{z})|\to 0.$$

Therefore,

$$\sup_{\mathbb{z}\in\mathcal{C}}|m_p^0(\mathbb{z})-m(\mathbb{z})|\to 0.$$

The result still holds under our assumption in **C4** that $\sqrt{n}D_W(F^{\Sigma_p},L^{\Sigma})\to 0$, because the weak convergence of F^{Σ_p} is implied. Next, we show the convergence rate is faster than $n^{-1/2}$.

Lemma S.1 *Suppose **C2** and **C4** hold. For any contour $\mathcal{C}$ such that $\mathcal{X}$ is in the interior of $\mathcal{C}$,*

$$\sup_{\mathbb{z}\in\mathcal{C}}\sqrt{p}|m_p^0(\mathbb{z})-m(\mathbb{z})|\longrightarrow 0.$$

We only need to show

$$\sup_{z\in\mathcal{C}} \sqrt{p}|\underline{m}_p^0(z) - \underline{m}(z)| \longrightarrow 0.$$

Without loss of generality, suppose the contour $\mathcal{C}$ intersects with the real axis at two points, $\underline{u}$ and $\overline{u}$, with $\underline{u} < 0$ and $\overline{u} > \limsup_p \lambda(\Sigma_p)(1+\sqrt{\gamma})^2$.

Note that it is enough to show the convergence on $\mathcal{C}\backslash\{\underline{u}, \overline{u}\}$, since $\underline{m}_p^0(z)$ and $\underline{m}(z)$ are smooth on $\mathcal{C}$ (Silverstein and Choi, 1995).

Denote the support of $L(\Sigma)$ to be $\mathrm{sp}(L^\Sigma)$. We first show that

$$\inf_{z\in\mathcal{C}} \inf_{\tau\in\mathrm{sp}(L^\Sigma)} |1+\tau\underline{m}(z)| > 0. \tag{S.2.1}$$

When $z = \underline{u}$ or $\overline{u}$, it follows from Silverstein and Choi (1995, Theorem 4.1) that $-\underline{m}(z)^{-1} \in \mathrm{sp}(L^\Sigma)^c$. Therefore, $1+\tau\underline{m}(\underline{u}) \neq 0$ and $1+\tau\underline{m}(\overline{u}) \neq 0$ for any $\tau \in \mathrm{sp}(L^\Sigma)$. Following from continuity of $\underline{m}(z)$ and the fact that $\mathrm{sp}(L^\Sigma)$ is compact, we can find a neighborhood of $\underline{u}$ and a neighborhood of $\overline{u}$ such that there exists a $\epsilon > 0$, for any z in the two neighborhoods,

$$\inf_{\tau\in\mathrm{sp}(L^\Sigma)} |1+\tau\underline{m}(z)| > \epsilon.$$

When z is outside of the neighborhoods, so away from the real axis, Bai and Silverstein (1998, Lemma 2.11) indicates that

$$|1+\tau\underline{m}(z)| \leqslant \max\left(\frac{4\limsup_p \lambda_{\max}(\Sigma_p)}{\Im(z)}, 2\right),$$

where $\Im(z)$ is the imaginary part of z. It completes the proof of (S.2.1).

Moreover, since $\sup_{z\in\mathcal{C}} |m_p^0(z) - m(z)| \longrightarrow 0$, for all sufficiently large p,

$$\inf_{z\in\mathcal{C}} \inf_{\tau\in\mathrm{sp}(L^\Sigma)} |1+\tau\underline{m}_p^0(z)| > \epsilon/2.$$

Following from (S.2.1), uniformly on $\mathcal{C}$,

$$w_{n1} := \Big[\gamma_n \int \frac{\tau \mathbb{1}_{\mathrm{sp}(L^\Sigma)}(\tau) dF^{\Sigma_p}(\tau)}{1+\tau\underline{m}(z)} - \gamma \int \frac{\tau dL^\Sigma(\tau)}{1+\tau\underline{m}(z)}\Big] = o(n^{-1/2}).$$

Define

$$\tilde{m}_p(z) = \Big[-z + \gamma_n \int \frac{\tau \mathbb{1}_{\mathrm{sp}(L^\Sigma)}(\tau) dF^{\Sigma_p}(\tau)}{1+\tau\underline{m}(z)}\Big]^{-1}.$$

We have, uniformly on $\mathcal{C}$,

$$\tilde{m}_p(z) - \underline{m}(z) = \underline{m}(z)\Big[\underline{m}^{-1}(z) + w_{n1}\Big]^{-1} w_{n1} = o(n^{-1/2}),$$

since both $\underline{m}(z)$ and $\underline{m}^{-1}(z)$ are bounded on $\mathcal{C}$.

Now consider $\underline{m}_p^0(z) - \tilde{m}_p(z)$. The target is to show $\underline{m}_p^0(z) - \tilde{m}_p(z) = [\underline{m}_p^0(z) - \underline{m}(z)]R_p + o(n^{-1/2})$ for some R_p such that $\sup_{z\in\mathcal{C}\backslash\{\underline{u},\overline{u}\}} |R_p| < 1$ for sufficiently large p.

For simplicity, we shall write $\mathbb{1}_{\mathrm{sp}(L^{\Sigma})}$ as $\mathbb{1}_{\mathrm{sp}}$. First, observe that the imaginary part of $\underline{m}_p^0(\mathbb{z})$, denoted to be $v_p^0(\mathbb{z})$, is

$$v_p^0(\mathbb{z}) = \frac{\Im(\mathbb{z}) + \gamma_n \int \frac{\tau^2 v_p^0(\mathbb{z}) dF^{\Sigma_p}(\tau)}{|1+\tau \underline{m}_p^0(\mathbb{z})|^2}}{\left| -\mathbb{z} + \gamma_n \int \frac{\tau dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}_p^0(\mathbb{z})} \right|^2}. \tag{S.2.2}$$

The imaginary part of $\tilde{m}_p(\mathbb{z})$, denoted to be $\tilde{v}_p(\mathbb{z})$, is

$$\tilde{v}_p(\mathbb{z}) = \frac{\Im(\mathbb{z}) + \gamma_n \int \frac{\tau^2 \mathbb{1}_{\mathrm{sp}}(\tau) v(\mathbb{z}) dF^{\Sigma_p}(\tau)}{|1+\tau \underline{m}(\mathbb{z})|^2}}{\left| -\mathbb{z} + \gamma_n \int \frac{\tau \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}(\mathbb{z})} \right|^2} \tag{S.2.3}$$

where $v(\mathbb{z})$ is the imaginary part of $\underline{m}(\mathbb{z})$.

$$\begin{aligned}
\underline{m}_p^0(\mathbb{z}) - \tilde{m}_p(\mathbb{z}) &= \frac{[\underline{m}_p^0(\mathbb{z}) - \underline{m}(\mathbb{z})]\gamma_n \int \frac{\tau^2 \mathbb{1}_{\mathrm{sp}}(\tau)}{[1+\tau \underline{m}_p^0(\mathbb{z})][1+\tau \underline{m}(\mathbb{z})]} dF^{\Sigma_p}(\tau)}{\left[-\mathbb{z} + \gamma_n \int \frac{\tau \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}(\mathbb{z})} \right]\left[-\mathbb{z} + \gamma_n \int \frac{\tau dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}_p^0(\mathbb{z})} \right]} \\
&\quad - \frac{\gamma_n \int \frac{\tau}{[1+\tau \underline{m}_p^0(\mathbb{z})]} (1 - \mathbb{1}_{\mathrm{sp}}(\tau)) dF^{\Sigma_p}(\tau)}{\left[-\mathbb{z} + \gamma_n \int \frac{\tau \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}(\mathbb{z})} \right]\left[-\mathbb{z} + \gamma_n \int \frac{\tau dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}_p^0(\mathbb{z})} \right]} \\
&= [\underline{m}_p^0(\mathbb{z}) - \underline{m}(\mathbb{z})] R_p + w_{n2}, \text{ say.}
\end{aligned}$$

$\left| \int (1 - \mathbb{1}_{\mathrm{sp}}(\tau)) dF^{\Sigma_p}(\tau) \right| = \left| \int (1 - \mathbb{1}_{\mathrm{sp}}(\tau)) d[F^{\Sigma_p}(\tau) - L^{\Sigma}(\tau)] \right| = o(n^{-1/2})$. It follows that uniformly on $\mathcal{C}$, $\sqrt{n}|w_{n2}| \to 0$.

Next, we verify that when p is sufficiently large, $|R_p| < 1$. By Hölder's inequality, when $\Im(\mathbb{z}) \neq 0$,

$$\begin{aligned}
|R_p|^2 &= \left| \frac{\gamma_n \int \frac{\tau^2 \mathbb{1}_{\mathrm{sp}}(\tau)}{[1+\tau \underline{m}_p^0(\mathbb{z})][1+\tau \underline{m}(\mathbb{z})]} dF^{\Sigma_p}(\tau)}{\left[-\mathbb{z} + \gamma_n \int \frac{\tau \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}(\mathbb{z})} \right]\left[-\mathbb{z} + \gamma_n \int \frac{\tau dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}_p^0(\mathbb{z})} \right]} \right|^2 \\
&\leqslant \frac{\gamma_n \int \frac{\tau^2 \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}}{|1+\tau \underline{m}(\mathbb{z})|^2}}{\left| -\mathbb{z} + \gamma_n \int \frac{\tau \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}(\mathbb{z})} \right|^2} \frac{\gamma_n \int \frac{\tau^2 dF^{\Sigma_p}}{|1+\tau \underline{m}_p^0(\mathbb{z})|^2}}{\left| -\mathbb{z} + \gamma_n \int \frac{\tau dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}_p^0(\mathbb{z})} \right|^2} \\
&= \frac{\gamma_n \int \frac{\tau^2 \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}}{|1+\tau \underline{m}(\mathbb{z})|^2} \tilde{v}_p(\mathbb{z})}{\Im(\mathbb{z}) + \gamma_n \int \frac{\tau^2 \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}(\tau)}{|1+\tau \underline{m}(\mathbb{z})|^2} v(\mathbb{z})} \cdot \frac{\gamma_n \int \frac{\tau^2 dF^{\Sigma_p}}{|1+\tau \underline{m}_p^0(\mathbb{z})|^2} v_p^0(\mathbb{z})}{\Im(\mathbb{z}) + \gamma_n \int \frac{\tau^2 dF^{\Sigma_p}(\tau)}{|1+\tau \underline{m}_p^0(\mathbb{z})|^2} v_p^0(\mathbb{z})} \\
&= r_{n1} \cdot r_{n2}, \text{say.}
\end{aligned}$$

Because $\underline{m}_p^0(\mathbb{z})$ is the Stieltjes transform of a c.d.f supported within $\mathcal{X}$, we have that

$$0 < \inf_{\mathbb{z} \in \mathcal{C} \setminus \{\underline{u}, \overline{u}\}} v_p^0(\mathbb{z}) / \Im(\mathbb{z}) \leqslant \sup_{\mathbb{z} \in \mathcal{C} \setminus \{\underline{u}, \overline{u}\}} v_p^0(\mathbb{z}) / \Im(\mathbb{z}) < \infty.$$

Therefore, $r_{n2} < K < 1$ for all $\mathbb{z} \in \mathcal{C} \backslash \{\underline{u}, \overline{u}\}$.

We only need to show $r_{n1} \leqslant 1$. Because $\underline{m}(\mathbb{z})$ is the Stieltjes transform of $\underline{F}^{\infty}$, again,

$$0 < \inf_{\mathbb{z} \in \mathcal{C} \backslash \{\underline{u}, \overline{u}\}} v(\mathbb{z})/\Im(\mathbb{z}) \leqslant \sup_{\mathbb{z} \in \mathcal{C} \backslash \{\underline{u}, \overline{u}\}} v(\mathbb{z})/\Im(\mathbb{z}) < \infty.$$

Observe that

$$v(\mathbb{z}) = \frac{\Im(\mathbb{z}) + \gamma \int \frac{\tau^2 v(\mathbb{z}) dL^{\Sigma}(\tau)}{|1+\tau \underline{m}(\mathbb{z})|^2}}{\left| -\mathbb{z} + \gamma \int \frac{\tau dL^{\Sigma}(\tau)}{1+\tau \underline{m}(\mathbb{z})} \right|^2}$$

Comparing with (S.2.3), we have

$$\begin{aligned} \tilde{v}_p(\mathbb{z})/v(\mathbb{z}) = & \frac{\Im(\mathbb{z})/v(\mathbb{z}) + \gamma_n \int \frac{\tau^2 \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}(\tau)}{|1+\tau \underline{m}(\mathbb{z})|^2}}{\Im(\mathbb{z})/v(\mathbb{z}) + \gamma \int \frac{\tau^2 dL^{\Sigma}(\tau)}{|1+\tau \underline{m}(\mathbb{z})|^2}} \frac{\left| -\mathbb{z} + \gamma \int \frac{\tau dL^{\Sigma}(\tau)}{1+\tau \underline{m}(\mathbb{z})} \right|^2}{\left| -\mathbb{z} + \gamma_n \int \frac{\tau \mathbb{1}_{\mathrm{sp}}(\tau) dF^{\Sigma_p}(\tau)}{1+\tau \underline{m}(\mathbb{z})} \right|^2} \\ & \longrightarrow 1, \text{ uniformly for } \mathbb{z} \in \mathcal{C} \backslash \{\underline{u}, \overline{u}\}. \end{aligned}$$

Therefore, for sufficiently large n, $r_{n1} \leqslant 1$ for all $\mathbb{z} \in \mathcal{C} \backslash \{\underline{u}, \overline{u}\}$.

Together, we proved

$$\underline{m}_p^0(\mathbb{z}) - \underline{m}(\mathbb{z}) = [\underline{m}_p^0(\mathbb{z}) - \underline{m}(\mathbb{z})] R_p + o(n^{-1/2}),$$

and $|R_n| < 1$ for all sufficiently large p. The uniform convergence of $\sqrt{n} |\underline{m}_p^0(\mathbb{z}) - \underline{m}(\mathbb{z})|$ follows.

Next, we show the derivative of $m_p^0(\mathbb{z})$ also converges to the derivative of $m(\mathbb{z})$.

Lemma S.2 *Suppose* ***C2*** *and* ***C4*** *hold. For any contour $\mathcal{C}$ such that $\mathcal{X}$ is in the interior of $\mathcal{C}$,*

$$\sup_{\mathbb{z} \in \mathcal{C}} \sqrt{p} \left| \frac{d}{d\mathbb{z}} m_p^0(\mathbb{z}) - \frac{d}{d\mathbb{z}} m(\mathbb{z}) \right| \longrightarrow 0.$$

We follow the arguments in the proof of Lemma S.1.

$$\underline{m}_p^0(\mathbb{z}) - \underline{m}(\mathbb{z}) = [\underline{m}_p^0(\mathbb{z}) - \underline{m}(\mathbb{z})] R_p + w_{n2} + \underline{m}(\mathbb{z}) [\underline{m}^{-1}(\mathbb{z}) + w_{n1}]^{-1} w_{n1}.$$

Take differentiation on both sides,

$$\begin{aligned} \frac{d}{d\mathbb{z}} \underline{m}_p^0(\mathbb{z}) - \frac{d}{d\mathbb{z}} \underline{m}(\mathbb{z}) = & [\frac{d}{d\mathbb{z}} \underline{m}_p^0(\mathbb{z}) - \frac{d}{d\mathbb{z}} \underline{m}(\mathbb{z})] R_p + [\underline{m}_p^0(\mathbb{z}) - \underline{m}(\mathbb{z})] \frac{d}{d\mathbb{z}} R_p \\ & + \frac{d}{d\mathbb{z}} w_{n2} + \frac{d}{d\mathbb{z}} \underline{m}(\mathbb{z}) [\underline{m}^{-1}(\mathbb{z}) + w_{n1}]^{-1} w_{n1} \\ & + \underline{m}(\mathbb{z}) \frac{d}{d\mathbb{z}} [\underline{m}^{-1}(\mathbb{z}) + w_{n1}]^{-1} w_{n1} + \underline{m}(\mathbb{z}) [\underline{m}^{-1}(\mathbb{z}) + w_{n1}]^{-1} \frac{d}{d\mathbb{z}} w_{n1} \end{aligned}$$

It is straightforward to verify, using arguments in the proof of Lemma S.1, that

$$\sup_{\mathbb{z} \in \mathcal{C}} |\frac{d}{d\mathbb{z}} R_p| < \infty,$$

$$\sup_{z\in\mathcal{C}}|\frac{d}{dz}[\underline{m}^{-1}(z)+w_{n1}]^{-1}|<\infty,$$
$$\sup_{z\in\mathcal{C}}\sqrt{n}\frac{d}{dz}w_{n1}\to 0,$$
$$\sup_{z\in\mathcal{C}}\sqrt{n}\frac{d}{dz}w_{n2}\to 0.$$

Together with the fact that $\underline{m}(z), \underline{m}^{-1}(z), d\underline{m}(z)/dz, d\underline{m}^{-1}(z)/dz$ are bounded,

$$\frac{d}{dz}\underline{m}_p^0(z)-\frac{d}{dz}\underline{m}(z)=[\frac{d}{dz}\underline{m}_p^0(z)-\frac{d}{dz}\underline{m}(z)]R_p+o(n^{-1/2}).$$

The uniform convergence of $\sqrt{n}|\frac{d}{dz}\underline{m}_p^0(z)-\frac{d}{dz}\underline{m}(z)|$ follows.

S.3. Proof of Theorem A.1

Recall notation in Section 2 and Appendix A.1 of the paper. Define

$$\Theta_n(z)=1+\gamma_n\frac{1}{p}\mathbb{E}\mathrm{tr}\{(\tilde{\boldsymbol{\Sigma}}_p-zI)^{-1}\Sigma_p\},$$
$$G_n^{(1)}(z,a,b)=n^{1/2}\{G_n(z,a,b)-\mathbb{E}G_n(z,a,b)\},$$
$$G_n^{(2)}(z,a,b)=n^{1/2}\{\mathbb{E}G_n(z,a,b)-a^Tb\frac{\Theta_n(z,\gamma)-1}{\Theta_n(z,\gamma)}\},$$
$$G_n^{(3)}(z,a,b)=n^{1/2}a^Tb\{\frac{\Theta_n(z)-1}{\Theta_n(z)}-\frac{\Theta(z,\gamma)-1}{\Theta(z,\gamma)}\}.$$

The rest of the section is organized as follows. In Subsection S.3.1, we show the finite dimensional convergence of $G_n^{(1)}(z,a,b)$. In Subsection S.3.2, we show the tightness of $G_n^{(1)}(z,a,b)$. In Subsection S.3.3, we show convergence of $G_n^{(2)}(z,a,b)$. In Subsection, S.3.4, we show convergence of $G_n^{(3)}(z,a,b)$. It completes the proof of Theorem A.1.

Notation We collect notation to the following list. Let

1. $\boldsymbol{u}_a=(u_{a1},\ldots,u_{aN})^T=n^{-1/2}U_na$;
2. $\boldsymbol{u}_b=(u_{b1},\ldots,u_{bN})^T=n^{-1/2}U_nb$;
3. $\bar{z}$ is the complex conjugate of z;
4. $\mathbf{z}_j$ is the jth column of $\mathbf{Z}$;
5. $\mathbf{Z}_j$ is $\mathbf{Z}$ with $\mathbf{z}_j$ replaced with the 0 vector;
6. $\mathbf{Z}_{ij}$ is $\mathbf{Z}$ with both $\mathbf{z}_i$ and $\mathbf{z}_j$ replaced with the 0 vectors;
7. $\mathbf{A}(z)=\Sigma_p^{T/2}(\tilde{\boldsymbol{\Sigma}}_p-zI)^{-1}\Sigma_p^{1/2}=\Sigma_p^{T/2}(\frac{1}{n}\Sigma_p^{1/2}\mathbf{Z}\mathbf{Z}^T\Sigma_p^{T/2}-zI)^{-1}\Sigma_p^{1/2}$;
8. $\mathbf{A}_j(z)=\Sigma_p^{T/2}(\frac{1}{n}\Sigma_p^{1/2}\mathbf{Z}_j\mathbf{Z}_j^T\Sigma_p^{T/2}-zI)^{-1}\Sigma_p^{1/2}$;
9. $\mathbf{A}_{ij}(z)=\Sigma_p^{T/2}(\frac{1}{n}\Sigma_p^{1/2}\mathbf{Z}_{ij}\mathbf{Z}_{ij}^T\Sigma_p^{T/2}-zI)^{-1}\Sigma_p^{1/2}$;
10. $\sigma(\mathbf{z}_1,\ldots,\mathbf{z}_N)$ is the σ-algebra generated by $\mathbf{z}_1,\ldots,\mathbf{z}_N$;

11. $\mathbb{E}_j[\cdot] = \mathbb{E}[\cdot \mid \sigma(\mathbf{z}_1, \ldots, \mathbf{z}_j)], j = 0, \ldots, N,$ with the convention $\mathbb{E}_0[\cdot] = \mathbb{E}[\cdot]$;
12. $\beta_j(\mathbb{z}) = \Big(1 + n^{-1}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j\Big)^{-1}$;
13. $\beta_j^{\mathrm{tr}}(\mathbb{z}) = \Big(1 + n^{-1}\mathrm{tr}\mathbf{A}_j(\mathbb{z})\Big)^{-1}$;
14. $\beta^{\mathbb{E}}(\mathbb{z}) = \Big(1 + \mathbb{E}n^{-1}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\Big)^{-1} = \Theta_{n-1}^{-1}(\mathbb{z})$;
15. $\theta_j(\mathbb{z}) = \frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j - \frac{1}{n}\mathrm{tr}\mathbf{A}_j(\mathbb{z})$;
16. $\varrho_j(\mathbb{z}) = \frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a\boldsymbol{u}_b^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j - \frac{1}{n}\boldsymbol{u}_a^T\mathbf{Z}_j^T\mathbf{A}_j^2(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b$.
17. e_i is the canonical vector with 1 on the ith coordinate.

The following identities holds

$$\begin{aligned}
\beta_j(\mathbb{z}) &= \beta_j^{\mathrm{tr}}(\mathbb{z}) - \beta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\theta_j(\mathbb{z}), \\
\beta_j(\mathbb{z}) &= \beta^{\mathbb{E}}(\mathbb{z}) - \beta_j(\mathbb{z})\beta^{\mathbb{E}}(\mathbb{z})(\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(\mathbb{z})).
\end{aligned} \tag{S.3.1}$$

In the following, $o_{L_1}(1)$ means a random variable converging to 0 in L_1-norm. Similarly, $o_{L_2}(1)$ means a random variable converging to 0 in L_2-norm. We shall use $\|\cdot\|_1$ to denote the entrywise matrix 1-norm, which is the sum of all matrix entries in absolute value. $\|\cdot\|_{\max}$ means the matrix max "norm", which is the maximum of all matrix entries in absolute value.

Under Condition **C4**, $\|U_n\|_{\max} = O(n^{-1/2})$. Therefore, for any fixed a and b,

$$\|\boldsymbol{u}_a\|_{\max} = O(n^{-1}), \quad \|\boldsymbol{u}_b\|_{\max} = O(n^{-1}).$$

S.3.1. Finite dimensional convergence of $G_n^{(1)}(\mathbb{z}, a, b)$

In this subsection, we show that

$$\sum_{i=1}^{r}\omega_i G_n^{(1)}(\mathbb{z}_i, a, b)$$

converges to a Gaussian random variable, where r is any positive integer, $\omega_1, \ldots, \omega_r$ and $\mathbb{z}_1, \ldots, \mathbb{z}_r$ are any complex numbers. In view of the smoothing step, $\mathbb{z}_1, \ldots, \mathbb{z}_r$ are required to have nonzero imaginary part. Without loss of generality, assume n is sufficiently large so that ρ_n is smaller than the imaginary part of $\mathbb{z}_1, \ldots, \mathbb{z}_r$.

S.3.1.1. Construction of martingale difference sequences

To lighten notation, the arguments a, b and γ may be dropped from some expressions whenever there is no scope of ambiguity. We represent $G_n^{(1)}(\mathbb{z})$ as the sum of a martingale difference sequence,

$$G_n^{(1)}(\mathbb{z}) = n^{1/2}\sum_{j=1}^{N}\{\mathbb{E}_j[G_n(\mathbb{z})] - \mathbb{E}_{j-1}[G_n(\mathbb{z})]\}$$

$$= n^{1/2} \sum_{j=1}^{N} \{\mathbb{E}_j[\boldsymbol{u}_a^T \mathbf{Z}^T \mathbf{A}(z) \mathbf{Z} \boldsymbol{u}_b - \boldsymbol{u}_a^T \mathbf{Z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b] - \mathbb{E}_{j-1}[\boldsymbol{u}_a^T \mathbf{Z}^T \mathbf{A}(z) \mathbf{Z} \boldsymbol{u}_b - \boldsymbol{u}_a^T \mathbf{Z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b]\}$$
$$= n^{1/2} \sum_{j=1}^{N} (\mathbb{E}_j - \mathbb{E}_{j-1})[d_1(z) + d_2(z) + d_3(z)],$$

where

$$d_1(z) = \boldsymbol{u}_a^T (\mathbf{Z} - \mathbf{Z}_j)^T \mathbf{A}(z) \mathbf{Z} \boldsymbol{u}_b,$$
$$d_2(z) = \boldsymbol{u}_a^T \mathbf{Z}_j^T (\mathbf{A}(z) - \mathbf{A}_j(z)) \mathbf{Z} \boldsymbol{u}_b,$$
$$d_3(z) = \boldsymbol{u}_a^T \mathbf{Z}_j^T \mathbf{A}_j(z) (\mathbf{Z} - \mathbf{Z}_j) \boldsymbol{u}_b.$$

$$d_1(z) = u_{aj} \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b + u_{aj} u_{bj} \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{z}_j - \frac{1}{n} u_{aj} \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{z}_j \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b \beta_j(z) - \frac{1}{n} u_{aj} u_{bj} (\mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{z}_j)^2 \beta_j(z)$$
$$= d_1^{(1)}(z) + d_1^{(2)}(z) + d_1^{(3)}(z) + d_1^{(4)}(z), \text{say.}$$

$d_1^{(1)}(z)$ is such that

$$n^{1/2} \sum_{j=1}^{N} (\mathbb{E}_j - \mathbb{E}_{j-1}) d_1^{(1)}(z) = n^{1/2} \sum_{j=1}^{N} \mathbb{E}_j[u_{aj} \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b].$$

By Lemma S.8, Lemma S.10 and LemmaS.13,

$$\mathbb{E}\Big| n^{1/2} \sum_{j=1}^{N} (\mathbb{E}_j - \mathbb{E}_{j-1}) d_1^{(2)}(z) \Big|^2$$
$$\leqslant n \sum_{j=1}^{N} |u_{aj} u_{bj}|^2 \mathbb{E}\Big| \mathbf{z}_j^T [\mathbb{E}_{j-1} \mathbf{A}_j(z)] \mathbf{z}_j - \mathrm{tr} \mathbb{E}_{j-1} \mathbf{A}_j(z) \Big|^2$$
$$\leqslant \mathcal{K} n \sum_{j=1}^{N} |u_{aj} u_{bj}|^2 n \mathbb{E} \|\mathbb{E}_{j-1} \mathbf{A}_j(z)\|^2 = o(1).$$

Due to (S.3.1), $d_1^{(3)}(z)$ is such that, we have

$$d_1^{(3)}(z) = \frac{1}{n} u_{aj} \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{z}_j \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b \beta_j(z) \beta_j^{\mathrm{tr}}(z) \theta_j(z) - u_{aj} \theta_j(z) \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b \beta_j^{\mathrm{tr}}(z) + u_{aj} (\beta_j^{\mathrm{tr}}(z) - 1) \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b.$$

By Lemma S.8, Lemma S.10 and Lemma S.18 , the first two terms above are such that

$$\mathbb{E} \frac{1}{n} \Big| \sum_{j=1}^{N} (\mathbb{E}_j - \mathbb{E}_{j-1}) u_{aj} \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{z}_j \mathbf{z}_j^T \mathbf{A}_j(z) \mathbf{Z}_j \boldsymbol{u}_b \beta_j(z) \beta_j^{\mathrm{tr}}(z) \theta_j(z) \Big|^2 = o(1),$$

$$\mathbb{E}\Big|n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})u_{aj}\theta_j(\mathbb{z})\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j^{\mathrm{tr}}(\mathbb{z})\Big|^2=o(1).$$

It leads to

$$n^{\frac{1}{2}}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})d_1^{(3)}(\mathbb{z})=\sum_{j=1}^{N}n^{\frac{1}{2}}\mathbb{E}_j[(\beta_j^{\mathrm{tr}}(\mathbb{z})-1)u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b]+o_p(1).$$

$$\begin{aligned}
&-(\mathbb{E}_j-\mathbb{E}_{j-1})d_1^{(4)}(\mathbb{z})\\
&=(\mathbb{E}_j-\mathbb{E}_{j-1})\frac{1}{n}u_{aj}u_{bj}(\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j)^2\beta_j^{\mathrm{tr}}(\mathbb{z})\\
&\quad-(\mathbb{E}_j-\mathbb{E}_{j-1})\frac{1}{n}u_{aj}u_{bj}(\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j)^2\beta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\theta_j(\mathbb{z})\\
&=(\mathbb{E}_j-\mathbb{E}_{j-1})\Big[\{nu_{aj}u_{bj}\theta_j^2(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\}+2\{u_{aj}u_{bj}\theta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\mathrm{tr}\mathbf{A}_j(\mathbb{z})\}\\
&\quad-\frac{1}{n}u_{aj}u_{bj}(\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j)^2\beta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\theta_j(\mathbb{z})\Big].
\end{aligned}$$

Using Lemma S.8, Lemma S.10 and Lemma S.13,

$$\begin{aligned}
&\mathbb{E}\Big|n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})nu_{aj}u_{bj}\theta_j^2(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\Big|^2\\
&\quad\leqslant\mathcal{K}n^3\sum_{j=1}^{N}|u_{aj}u_{bj}|^2\mathbb{E}|\theta_j^2(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})|^2=o(1).\\
&\mathbb{E}\Big|n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})u_{aj}u_{bj}\theta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\mathrm{tr}\mathbf{A}_j(\mathbb{z})\Big|^2\\
&\quad\leqslant\mathcal{K}n\sum_{j=1}^{N}|u_{aj}u_{bj}|^2\mathbb{E}|\theta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\mathrm{tr}\mathbf{A}_j(\mathbb{z})|^2=o(1).
\end{aligned}$$

Similarly, by Lemma S.18, we can show

$$\mathbb{E}\Big|n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})n^{-1}u_{aj}u_{bj}(\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j)^2\beta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\theta_j(\mathbb{z})\Big|^2=o(1).$$

Thus,

$$\mathbb{E}\Big|n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})d_1^{(4)}(\mathbb{z})\Big|^2=o(1).$$

All together

$$n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})d_1(\mathbb{z})=n^{1/2}\sum_{j=1}^{N}\mathbb{E}_j[u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j^{\mathrm{tr}}(\mathbb{z})]+o_p(1).$$

Further, we want to replace $\beta^{\mathrm{tr}}(\mathbb{z})$ with $\beta^{\mathbb{E}}(\mathbb{z})$. By (S.3.1), Lemma S.12, and Lemma S.16,

$$\begin{aligned}&\mathbb{E}|(\beta_j^{\mathrm{tr}}(\mathbb{z})-\beta^{\mathbb{E}}(\mathbb{z}))\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b|^2\\&\quad=\mathbb{E}|\beta_j^{\mathrm{tr}}(\mathbb{z})\beta^{\mathbb{E}}(\mathbb{z})\frac{1}{n}[\mathrm{tr}\mathbf{A}_j(\mathbb{z})-\mathbb{E}\mathrm{tr}\mathbf{A}_j(\mathbb{z})]\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b|^2=o(1).\end{aligned}$$

Therefore,

$$n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})d_1(\mathbb{z})=\beta^{\mathbb{E}}(\mathbb{z})\sum_{j=1}^{N}\mathbb{E}_j[n^{1/2}u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b]+o_p(1).$$

Secondly, using Lemma S.6,

$$\begin{aligned}d_2(\mathbb{z})=&-\frac{1}{n}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a\beta_j(\mathbb{z})\\&-\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\boldsymbol{u}_a^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j\beta_j(\mathbb{z})\\=&-\frac{1}{n}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a\beta_j(\mathbb{z})\\&-\varrho_j(\mathbb{z})\beta_j(\mathbb{z})-\frac{1}{n}\boldsymbol{u}_a^T\mathbf{Z}_j^T\mathbf{A}_j^2(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z})\\=&d_2^{(1)}(\mathbb{z})+d_2^{(2)}(\mathbb{z})+d_2^{(3)}(\mathbb{z}),\ \text{say.}\end{aligned}$$

Along very similar lines to those we use to simplify $d_1(\mathbb{z})$, together with

$$\mathbb{E}|\varrho_j(\mathbb{z})|^2\leqslant\mathcal{K}\mathbb{E}\frac{1}{n^2}\left|\boldsymbol{u}_a^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{A}_j(\overline{\mathbb{z}})\mathbf{Z}_j\boldsymbol{u}_a\boldsymbol{u}_b^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{A}_j(\overline{\mathbb{z}})\mathbf{Z}_j\boldsymbol{u}_b\right|=O(n^{-2}),$$

which is due to Lemma S.9 and Lemma S.15, we can show

$$\begin{aligned}&n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})d_2^{(1)}(\mathbb{z})\\&=\sum_{j=1}^{N}\mathbb{E}_j[(\beta_j^{\mathrm{tr}}(\mathbb{z})-1)n^{1/2}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a]+o_p(1)\\&=(\beta^{\mathbb{E}}(\mathbb{z})-1)\sum_{j=1}^{N}\mathbb{E}_j[n^{1/2}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a]+o_p(1).\\&n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})d_2^{(2)}(\mathbb{z})=-\beta^{\mathbb{E}}(\mathbb{z})n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})\varrho_j(\mathbb{z})\end{aligned}$$

$$
\begin{aligned}
&- n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})\varrho_j(\mathbb{z})\{\beta_j^{\mathrm{tr}}(\mathbb{z})-\beta^{\mathbb{E}}(\mathbb{z})\}\\
&+ n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})\varrho_j(\mathbb{z})\beta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\theta_j(\mathbb{z})\\
=& -\beta^{\mathbb{E}}(\mathbb{z})n^{1/2}\sum_{j=1}^{N}\mathbb{E}_j\varrho_j(\mathbb{z})+o_p(1),\\
& n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})d_2^{(3)}(\mathbb{z})\\
=& \sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})n^{-1/2}\boldsymbol{u}_a^T\mathbf{Z}_j^T\mathbf{A}_j^2(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z})\beta_j^{\mathrm{tr}}(\mathbb{z})\theta_j(\mathbb{z})=o_p(1).
\end{aligned}
$$

All together, we have

$$
\begin{aligned}
n^{1/2}\sum_{j=1}^{N}(\mathbb{E}_j-\mathbb{E}_{j-1})(d_2(\mathbb{z})) =&(\beta^{\mathbb{E}}(\mathbb{z})-1)\sum_{j=1}^{N}\mathbb{E}_j[n^{1/2}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a]\\
&-\beta^{\mathbb{E}}(\mathbb{z})n^{1/2}\sum_{j=1}^{N}\mathbb{E}_j\varrho_j(\mathbb{z}).
\end{aligned}
$$

Combining with $d_3(\mathbb{z})$, we have proved, thus far,

$$
G_n^{(1)}(\mathbb{z})=\sum_{j=1}^{N}\beta^{\mathbb{E}}(\mathbb{z})\mathcal{H}_n(\mathbb{z},j)+o_p(1)
$$

where

$$
\begin{aligned}
\mathcal{H}_n(\mathbb{z},j)=&\,\mathbb{E}_j[n^{1/2}u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b]+\mathbb{E}_j[n^{1/2}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a]\\
&-\mathbb{E}_jn^{1/2}\varrho_j(\mathbb{z})=\mathcal{H}_{[1]}(\mathbb{z},j)+\mathcal{H}_{[2]}(\mathbb{z},j)+\mathcal{H}_{[3]}(\mathbb{z},j),\ \text{say}.
\end{aligned}
$$

In summary, it suffices to find the weak limit of $\sum_{i=1}^{r}\sum_{j=1}^{N}\omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_n(\mathbb{z}_i,j)$, that is

$$
\sum_{i=1}^{r}\sum_{j=1}^{N}\omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)[\mathcal{H}_{[1]}(\mathbb{z}_i,j)+\mathcal{H}_{[2]}(\mathbb{z}_i,j)+\mathcal{H}_{[3]}(\mathbb{z}_i,j)].
$$

S.3.1.2. martingale central limit theorem

We use the following theorem to show finite dimensional convergence of

$$
\sum_{i=1}^{r}\sum_{j=1}^{N}\omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_n(\mathbb{z}_i,j).
$$

Theorem S.1 *(Theorem 35.12 of Billingsley (1995)). Suppose $Y_1, \dots, Y_n$ is a martinagle difference sequence with respsect to the increasing σ-field $\mathscr{F}_1, \dots, \mathscr{F}_n$, with finite second moments. If as $n \to 0$,*

(i) $\sum_{j=1}^{n} \mathbb{E}(Y_j^2 \mid \mathscr{F}_{j-1}) \to \sigma^2$ *in probability where σ^2 is a positive constant,*

(ii) $\sum_{j=1}^{n} \mathbb{E}\{Y_j^2 \mathbb{1}(|Y_j| \geqslant \epsilon)\} \to 0$ *for each $\epsilon > 0$,*

then,

$$\sum_{j=1}^{n} Y_j \to \mathcal{N}(0, \sigma^2), \qquad \textit{in distribution.}$$

It is easy to check that $\mathcal{H}_n(\mathbb{z}, j)$ has finite second moments. We show Condition (ii) first.

As for $\mathcal{H}_{[3]}(\mathbb{z}, j)$, because of insufficient finite moments of z_{ij}, we need to analyze $\varrho_j(\mathbb{z})$ carefully. Write

$$\begin{aligned}
\varrho_j(\mathbb{z}) &= \frac{1}{n} \sum_{i \neq \ell}^{p} e_i^T \mathbf{A}_j(\mathbb{z}) \mathbf{Z}_j \boldsymbol{u}_a \boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) e_\ell z_{ij} z_{\ell j} \\
&+ \frac{1}{n} \sum_{i=1}^{p} e_i^T \mathbf{A}_j(\mathbb{z}) \mathbf{Z}_j \boldsymbol{u}_a \boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) e_i \Big[z_{ij}^2 \mathbb{1}(|z_{ij}| \leqslant \log n) - \mathbb{E} z_{ij}^2 \mathbb{1}(|z_{ij}| \leqslant \log n) \Big] \\
&+ \frac{1}{n} \sum_{i=1}^{p} e_i^T \mathbf{A}_j(\mathbb{z}) \mathbf{Z}_j \boldsymbol{u}_a \boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) e_i \Big[z_{ij}^2 \mathbb{1}(|z_{ij}| > \log n) - \mathbb{E} z_{ij}^2 \mathbb{1}(|z_{ij}| > \log n) \Big] \\
&= \varrho_j^{(1)}(\mathbb{z}) + \varrho_j^{(2)}(\mathbb{z}) + \varrho_j^{(3)}(\mathbb{z}), \text{ say.}
\end{aligned}$$

By Lemma 5 of Pan and Zhou (2011),

$$\begin{aligned}
\mathbb{E}|\varrho_j^{(1)}(\mathbb{z})|^4 &\leqslant \mathcal{K} n^{-4} \mathbb{E}\Big| \boldsymbol{u}_a^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) \mathbf{A}_j(\overline{\mathbb{z}}) \mathbf{Z}_j \boldsymbol{u}_a \boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) \mathbf{A}_j(\overline{\mathbb{z}}) \mathbf{Z}_j \boldsymbol{u}_b \Big|^2 \\
&= O(n^{-4}).
\end{aligned}$$

By Lemma S.15,

$$\begin{aligned}
\mathbb{E}|\varrho_j^{(2)}(\mathbb{z})|^4 &\leqslant \mathcal{K} n^{-4} (\log n)^8 \mathbb{E}\Big(\sum_{i=1}^{p} |e_i^T \mathbf{A}_j(\mathbb{z}) \mathbf{Z}_j \boldsymbol{u}_a \boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) e_i| \Big)^4 \\
&\leqslant \mathcal{K} n^{-4} (\log n)^8 \mathbb{E}\Big| \boldsymbol{u}_a^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) \mathbf{A}_j(\overline{\mathbb{z}}) \mathbf{Z}_j \boldsymbol{u}_a \boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) \mathbf{A}_j(\overline{\mathbb{z}}) \mathbf{Z}_j \boldsymbol{u}_b \Big|^2 \\
&= O(n^{-4} (\log n)^8).
\end{aligned}$$

Note, conditional on $\mathbf{Z}_j$, all summation terms in $\varrho_j^{(3)}$ are mutually independent.

$$\mathbb{E}|\varrho_j^{(3)}(\mathbb{z})|^2 \leqslant \mathcal{K} n^{-2} \mathbb{E}\Big| \sum_{i=1}^{p} |e_i^T \mathbf{A}_j(\mathbb{z}) \mathbf{Z}_j \boldsymbol{u}_a \boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}) e_i| \Big|^2 \mathbb{E} z_{ij}^4 \mathbb{1}(|z_{ij}| > \log n)$$

$$= o(n^{-2}).$$

The last step is due to $\mathbb{E}z_{ij}^4\mathbb{1}(|z_{ij}| > \log n) \to 0$, as $n \to \infty$.

As for $\mathcal{H}_{[1]}(\mathbb{z}, j)$ and $\mathcal{H}_{[2]}(\mathbb{z}, j)$, Lemma S.16 says that

$$\mathbb{E}\Big|n^{1/2}u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\Big|^4 = O(n^{-2}),$$
$$\mathbb{E}\Big|n^{1/2}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a\Big|^4 = O(n^{-2}).$$

To verify condition (ii), for any positive ϵ,

$$\begin{aligned}
&\sum_{j=1}^N \mathbb{E}\Big[\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_n(\mathbb{z}_i, j)\Big|^2\mathbb{1}\big(\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_n(\mathbb{z}_i, j)\Big| \geqslant \epsilon\big)\Big]\\
&\leqslant 25\sum_{j=1}^N \mathbb{E}\Big[\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_{[1]}(\mathbb{z}_i, j)\Big|^2\mathbb{1}\big(\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_{[1]}(\mathbb{z}_i, j)\Big| \geqslant \epsilon/5\big)\Big]\\
&+ 25\sum_{j=1}^N \mathbb{E}\Big[\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_{[2]}(\mathbb{z}_i, j)\Big|^2\mathbb{1}\big(\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_{[2]}(\mathbb{z}_i, j)\Big| \geqslant \epsilon/5\big)\Big]\\
&+ 25\sum_{j=1}^N \mathbb{E}\Big[\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)n^{1/2}\varrho_j^{(1)}(\mathbb{z})\Big|^2\mathbb{1}\big(\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)n^{1/2}\varrho_j^{(1)}(\mathbb{z})\Big| \geqslant \epsilon/5\big)\Big]\\
&+ 25\sum_{j=1}^N \mathbb{E}\Big[\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)n^{1/2}\varrho_j^{(2)}(\mathbb{z})\Big|^2\mathbb{1}\big(\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)n^{1/2}\varrho_j^{(2)}(\mathbb{z})\Big| \geqslant \epsilon/5\big)\Big]\\
&+ 25\sum_{j=1}^N \mathbb{E}\Big[\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)n^{1/2}\varrho_j^{(3)}(\mathbb{z})\Big|^2\mathbb{1}\big(\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)n^{1/2}\varrho_j^{(3)}(\mathbb{z})\Big| \geqslant \epsilon/5\big)\Big]\\
&\leqslant \frac{\mathcal{K}}{\epsilon^2}\sum_{j=1}^N \mathbb{E}\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_{[1]}(\mathbb{z}_i, j)\Big|^4 + \frac{\mathcal{K}}{\epsilon^2}\sum_{j=1}^N \mathbb{E}\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_{[2]}(\mathbb{z}_i, j)\Big|^4\\
&+ \frac{\mathcal{K}}{\epsilon^2}\sum_{j=1}^N \mathbb{E}\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)n^{1/2}\varrho_j^{(1)}(\mathbb{z})\Big|^4 + \frac{\mathcal{K}}{\epsilon^2}\sum_{j=1}^N \mathbb{E}\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)n^{1/2}\varrho_j^{(2)}(\mathbb{z})\Big|^4\\
&+ \mathcal{K}n\sum_{j=1}^N \mathbb{E}\Big|\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\varrho_j^{(3)}(\mathbb{z})\Big|^2 \longrightarrow 0.
\end{aligned}$$

To verify Condition (i) of Theorem S.1, we next need to find the limit in probability of

$$\sum_{j=1}^N \mathbb{E}_{j-1}\big(\sum_{i=1}^r \omega_i\beta^{\mathbb{E}}(\mathbb{z}_i)\mathcal{H}_n(\mathbb{z}_i, j)\big)^2$$

$$= \sum_{i=1}^{r}\sum_{i'=1}^{r}\sum_{j=1}^{N} \mathbb{E}_{j-1}\omega_i\omega_{i'}\beta^{\mathbb{E}}(\mathbb{z}_i)\beta^{\mathbb{E}}(\mathbb{z}_{i'})\mathcal{H}_n(\mathbb{z}_i, j)\mathcal{H}_n(\mathbb{z}_{i'}, j).$$

In next four sections, for arbitrary $\mathbb{z}_1$ and $\mathbb{z}_2$ with nonzero imaginary part, we derive the limit in probability of

$$\sum_{j=1}^{N} \mathbb{E}_{j-1}\mathcal{H}_{[1]}(\mathbb{z}_1, j)\mathcal{H}_{[1]}(\mathbb{z}_2, j) \tag{S.3.2}$$

$$\sum_{j=1}^{N} \mathbb{E}_{j-1}\mathcal{H}_{[1]}(\mathbb{z}_1, j)\mathcal{H}_{[2]}(\mathbb{z}_2, j) \tag{S.3.3}$$

$$\sum_{j=1}^{N} \mathbb{E}_{j-1}\mathcal{H}_{[1]}(\mathbb{z}_1, j)\mathcal{H}_{[3]}(\mathbb{z}_2, j) \tag{S.3.4}$$

$$\sum_{j=1}^{N} \mathbb{E}_{j-1}\mathcal{H}_{[3]}(\mathbb{z}_1, j)\mathcal{H}_{[3]}(\mathbb{z}_2, j). \tag{S.3.5}$$

Note, $\mathcal{H}_{[2]}(\mathbb{z}_1, j)\mathcal{H}_{[2]}(\mathbb{z}_2, j)$ and $\mathcal{H}_{[2]}(\mathbb{z}_1, j)\mathcal{H}_{[3]}(\mathbb{z}_2, j)$ are just $\mathcal{H}_{[1]}(\mathbb{z}_1, j)\mathcal{H}_{[1]}(\mathbb{z}_2, j)$ and $\mathcal{H}_{[1]}(\mathbb{z}_1, j)\mathcal{H}_{[3]}(\mathbb{z}_2, j)$ respectively with a and b exchanged.

S.3.1.3. The limit of (S.3.2)

This subsection shows that, as $n \to \infty$,

$$\begin{aligned}&\sum_{j=2}^{N} \mathbb{E}_{j-1}\mathcal{H}_{[1]}(\mathbb{z}_1, j)\mathcal{H}_{[1]}(\mathbb{z}_2, j) \\ &= n\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\sum_{j=2}^{N} u_{aj}^2 \mathrm{tr}\left[\mathbb{E}_j \mathbf{A}_j(\mathbb{z}_1)\mathbb{E}_j\mathbf{A}_j(\mathbb{z}_2)\right]\sum_{i=1}^{j-1} u_{bi}^2 + o_p(1).\end{aligned} \tag{S.3.6}$$

Note that clearly when $j = 1$, $\mathbb{E}\mathcal{H}_{[1]}(\mathbb{z}_1, 1)\mathcal{H}_{[1]}(\mathbb{z}_2, 1) = 0$.

We introduce $\underline{\mathbf{Z}}_j = [\mathbf{z}_1, \ldots, \mathbf{z}_{j-1}, 0, \underline{\mathbf{z}}_{j+1}, \ldots, \underline{\mathbf{z}}_N]$ for $j = 2, 3, \ldots, N$, where $\underline{\mathbf{z}}_{j+1}, \ldots, \underline{\mathbf{z}}_N$ are i.i.d. copies of $\mathbf{z}_1$ and independent with $\mathbf{z}_1, \ldots, \mathbf{z}_{j-1}$. What's more, introduce $\underline{\mathbf{A}}_j(\mathbb{z})$ like $\mathbf{A}_j$, but $\underline{\mathbf{A}}_j(\mathbb{z})$ is now defined on $\underline{\mathbf{Z}}_j$ instead of $\mathbf{Z}_j$. Note, conditional on $\mathbf{z}_1, \ldots, \mathbf{z}_{j-1}$, $(\mathbf{Z}_j, \mathbf{A}_j)$ is independent with $(\underline{\mathbf{Z}}_j, \underline{\mathbf{A}}_j)$. Therefore, for $j \geqslant 2$,

$$\begin{aligned}&\mathbb{E}_{j-1}\Big[\mathbb{E}_j[\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_1)\mathbf{Z}_j\boldsymbol{u}_b]\mathbb{E}_j[\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_b]\Big] \\ =&\mathbb{E}_{j-1}[\boldsymbol{u}_b^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_j\boldsymbol{u}_b] \\ =&\sum_{i=1}^{j-1} u_{bi}^2\mathbb{E}_{j-1}\mathbf{z}_i^T\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\mathbf{z}_i + \sum_{i=1}^{j-1} u_{bi}\mathbb{E}_{j-1}\mathbf{z}_i^T\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b \\ &+ \sum_{i=j+1}^{N} u_{bi}\mathbb{E}_{j-1}\mathbf{z}_i^T\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_j\boldsymbol{u}_b = \mathcal{J}_j^{(1)} + \mathcal{J}_j^{(2)} + \mathcal{J}_j^{(3)}, \text{ say.}\end{aligned}$$

where $\underline{\mathbf{Z}}_{ij}, i < j$ is defined to be $[\mathbf{z}_1, \ldots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \ldots, \mathbf{z}_{j-1}, 0, \underline{\mathbf{z}}_{j+1}, \ldots, \underline{\mathbf{z}}_N]$.
(S.3.6) will follow, if we can show,

$$\sup_{1\leqslant j\leqslant N} \mathbb{E}\Big|\mathcal{J}_j^{(1)} - \beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\Big(\sum_{i=1}^{j-1} u_{bi}^2\Big)\mathrm{tr}\{[\mathbb{E}_j[\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)]]\}\Big| = o(1),$$
$$\sup_{1\leqslant j\leqslant N} \mathbb{E}|\mathcal{J}_j^{(2)}| = o(1),$$
$$\sup_{1\leqslant j\leqslant N} \mathbb{E}|\mathcal{J}_j^{(3)}| = o(1).$$

Similar to $\beta_j(\mathbb{z})$, define

$$\beta_{ij}(\mathbb{z}) = \frac{1}{1 + n^{-1}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z})\mathbf{z}_i},$$
$$\underline{\beta}_{ij}(\mathbb{z}) = \frac{1}{1 + n^{-1}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z})\mathbf{z}_i}$$

where $\underline{\mathbf{A}}_{ij}(\mathbb{z})$ is defined in the same way as $\mathbf{A}_{ij}(\mathbb{z})$, but with $\underline{\mathbf{Z}}_{ij}$ instead of $\mathbf{Z}_{ij}$.
As for $\mathcal{J}_j^{(1)}$,

$$\begin{aligned}
&\mathbb{E}\Big|\mathcal{J}_j^{(1)} - \beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\sum_{i=1}^{j-1} u_{bi}^2\mathrm{tr}\left[\mathbb{E}_j[\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)]\right]\Big| \\
&= \mathbb{E}\Big|\mathbb{E}_{j-1}\sum_{i=1}^{j-1} u_{bi}^2\beta_{ij}(\mathbb{z}_1)\underline{\beta}_{ij}(\mathbb{z}_2)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i - \\
&\qquad \sum_{i=1}^{j-1} u_{bi}^2\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathrm{tr}\left[\mathbb{E}_j[\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)]\right]\Big| \\
&\leqslant n^2\|\boldsymbol{u}_b\|_{\max}^2\mathbb{E}\Big|\beta_{ij}(\mathbb{z}_1)\underline{\beta}_{ij}(\mathbb{z}_2)\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i - \\
&\qquad \beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\frac{1}{n}\mathrm{tr}\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\}\Big| \\
&\leqslant n^2\|\boldsymbol{u}_b\|_{\max}^2\mathbb{E}\Big|(\beta_{ij}(\mathbb{z}_1) - \beta^{\mathbb{E}}(\mathbb{z}_1))\underline{\beta}_{ij}(\mathbb{z}_2)\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i\Big| \\
&+ n^2\|\boldsymbol{u}_b\|_{\max}^2\mathbb{E}\Big|\beta^{\mathbb{E}}(\mathbb{z}_1)(\underline{\beta}_{ij}(\mathbb{z}_2) - \beta^{\mathbb{E}}(\mathbb{z}_2))\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i\Big| \\
&+ n^2\|\boldsymbol{u}_b\|_{\max}^2|\beta^{\mathbb{E}}(\mathbb{z}_1)||\beta^{\mathbb{E}}(\mathbb{z}_2)|\mathbb{E}\Big|\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i - \frac{1}{n}\mathrm{tr}\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_1)\}\Big| \\
&= o(1).
\end{aligned}$$

For the last line, we need to the following arguments that are direct consequences of Lemma S.10, Lemma S.11, Lemma S.12 and Lemma S.13.

$$|\beta^{\mathbb{E}}(\mathbb{z})| < \infty,$$

$$
\begin{aligned}
&\mathbb{E}|\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z})\underline{\mathbf{A}}_{ij}(\mathbb{z})\mathbf{z}_i|^2 = O(1),\\
&\mathbb{E}|\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z})\underline{\mathbf{A}}_{ij}(\mathbb{z})\mathbf{z}_i - \frac{1}{n}\mathrm{tr}\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_1)\}|^2 = o(1),\\
&\mathbb{E}|\beta_{ij}(z) - \beta^{\mathbb{E}}(\mathbb{z})|^2 \leqslant \mathbb{E}\Big|\beta_{ij}(z)\beta^{\mathbb{E}}(\mathbb{z})(\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}\mathbf{z}_i - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_{ij})\Big|^2\\
&\qquad + \mathbb{E}\Big|\beta_{ij}(z)\beta^{\mathbb{E}}(\mathbb{z})(\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_{ij} - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j)\Big|^2 = o(1).
\end{aligned}
$$

As for $\mathcal{J}_j^{(2)}$, due to

$$
\begin{aligned}
&\mathbb{E}\Big|(\beta_{ij}(\mathbb{z}_1) - \beta^{\mathbb{E}}(\mathbb{z}_1))\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big| = o(1),\\
&\mathbb{E}\Big|\frac{1}{n}(\beta_{ij}(\mathbb{z}_1) - \beta^{\mathbb{E}}(\mathbb{z}_1))\mathrm{tr}\{\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big| = o(1),\\
&\mathbb{E}\Big|\frac{1}{n}\beta_{ij}(\mathbb{z}_1)\beta_{ij}(\mathbb{z}_2)[\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i - \mathrm{tr}\{\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\}]\cdot\\
&\qquad \mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big| = o(1),
\end{aligned}
$$

which are consequences of Lemma S.10, Lemma S.12, Lemma S.13 and Lemma S.16,

$$
\begin{aligned}
\mathcal{J}_j^{(2)} &= \mathbb{E}_{j-1}\sum_{i=1}^{j-1} u_{bi}\beta_{ij}(\mathbb{z}_1)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&= \mathbb{E}_{j-1}\sum_{i=1}^{j-1} u_{bi}\beta_{ij}(\mathbb{z}_1)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&\quad - \mathbb{E}_{j-1}\sum_{i=1}^{j-1}\frac{1}{n}u_{bi}\beta_{ij}(\mathbb{z}_1)\underline{\beta}_{ij}(\mathbb{z}_2)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&= \sum_{i=1}^{j-1} u_{bi}\beta^{\mathbb{E}}(\mathbb{z}_1)\mathbb{E}_{j-1}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&- \sum_{i=1}^{j-1}\frac{1}{n}u_{bi}\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathbb{E}_{j-1}\mathrm{tr}\{\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b + o_{L_1}(1)
\end{aligned}
$$

The residual term above is uniform over j.

We claim that the first two terms are also negligible. To see this, we first show

$$
\sup_j \mathbb{E}\Big|\sum_{i=1}^{j-1} u_{bi}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big|^2 = o(1), \tag{S.3.7}
$$

$$
\sup_j \mathbb{E}\Big|\sum_{i=1}^{j-1} u_{bi}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big|^2 = o(1). \tag{S.3.8}
$$

The proofs of the two are very similar. Therefore, we shall only present proof of (S.3.7).

Like $\mathbf{Z}_{ij}$, we define $\mathbf{Z}_{ii'j}$ to be $\mathbf{Z}$ with $\mathbf{z}_i, \mathbf{z}_{i'}, \mathbf{z}_j$ replaced by 0. Similarly, define $\underline{\mathbf{Z}}_{ii'j}$, $i < j, i' < j$ by replacing the i-th and i'-th column of $\underline{\mathbf{Z}}_j$ by the 0 vectors. Further, define $\mathbf{A}_{ii'j}$ and $\underline{\mathbf{A}}_{ii'j}$ to be the counterparts of $\mathbf{A}_{ij}$ and $\underline{\mathbf{A}}_{ij}$ with $\mathbf{Z}_{ij}$ replaced by $\mathbf{Z}_{ii'j}$, or with $\underline{\mathbf{Z}}_{ij}$ replaced by $\underline{\mathbf{Z}}_{ii'j}$. Accordingly, define

$$\begin{aligned}
\beta_{ij}^{\mathrm{tr}}(\mathbb{z}) &= \frac{1}{1 + n^{-1}\mathrm{tr}\mathbf{A}_{ij}(\mathbb{z})},\\
\theta_{ij}(\mathbb{z}) &= \frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z})\mathbf{z}_i - \frac{1}{n}\mathrm{tr}\mathbf{A}_{ij}(\mathbb{z}),\\
\beta_{ii'j}(\mathbb{z}) &= \frac{1}{1 + n^{-1}\mathbf{z}_{i'}^T\mathbf{A}_{ii'j}(\mathbb{z})\mathbf{z}_{i'}},\\
\underline{\beta}_{ii'j}(\mathbb{z}) &= \frac{1}{1 + n^{-1}\mathbf{z}_{i'}^T\underline{\mathbf{A}}_{ii'j}(\mathbb{z})\mathbf{z}_{i'}}.
\end{aligned}$$

By Lemma S.16, for squared-terms in the expansion of (S.3.7),

$$\begin{aligned}
&\sum_{i=1}^{j-1} u_{bi}^2\mathbb{E}\Big|\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big|^2\\
&\leqslant \sum_{i=1}^{N} u_{bi}^2\mathbb{E}\Big|\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big|^2 = O(n^{-1}).
\end{aligned}$$

For crossed-terms with $i \neq i'$, due to Lemma S.10, Lemma S.16, and Lemma S.17,

$$\begin{aligned}
&\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&\leqslant \mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ii'j}\boldsymbol{u}_b\mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&\quad + |u_{bi'}|\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&\quad + \frac{1}{n}\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\underline{\beta}_{ii'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ii'j}\boldsymbol{u}_b\cdot\\
&\quad\ \mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&\quad + \frac{1}{n}|u_{bi'}|\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\underline{\beta}_{ii'j}(\mathbb{z}_2)\mathbf{z}_{i'}\cdot\\
&\quad\ \mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&\quad + \frac{1}{n}\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\beta_{ii'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ii'j}\boldsymbol{u}_b\cdot\\
&\quad\ \mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&\quad + \frac{1}{n}|u_{bi'}|\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\beta_{ii'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\mathbf{z}_{i'}\cdot\\
&\quad\ \mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&\quad + \frac{1}{n^2}\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\beta_{ii'j}(\mathbb{z}_1)\cdot
\end{aligned}$$

$$
\begin{aligned}
&\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\underline{\beta}_{ii'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ii'j}\boldsymbol{u}_b\mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&+\frac{1}{n^2}|u_{bi'}|\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\mathbf{A}_{ii'j}(\mathbb{z}_1)\beta_{ii'j}(\mathbb{z}_1)\cdot\\
&\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\underline{\mathbf{A}}_{ii'j}(\mathbb{z}_2)\underline{\beta}_{ii'j}(\mathbb{z}_2)\mathbf{z}_{i'}\mathbf{z}_{i'}^T\mathbf{A}_{i'j}(\mathbb{z}_1)\underline{\mathbf{A}}_{i'j}(\mathbb{z}_2)\underline{\mathbf{Z}}_{i'j}\boldsymbol{u}_b|\\
&=o(1).
\end{aligned}
$$

Therefore, (S.3.7) follows. We conclude that the first term in the expansion of $\mathcal{J}_j$ is $o_{L_1}(1)$ uniformly for all j.

The second term in the expansion of $\mathcal{J}_j$ is such that,

$$
\begin{aligned}
&\mathbb{E}\Big|\Big[\mathrm{tr}\{\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\}-\mathrm{tr}\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\}\Big]\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big|^2\\
&\leqslant\mathcal{K}\mathbb{E}\Big|\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big|^2=O(1),
\end{aligned}
$$

due to an inequality similar to Lemma S.11.

Therefore,

$$
\begin{aligned}
&\sum_{i=1}^{j-1}\frac{1}{n}u_{bi}\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathbb{E}_{j-1}\mathrm{tr}\{\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&=\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathbb{E}_{j-1}\Big[\frac{1}{n}\mathrm{tr}\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\}\sum_{i=1}^{j-1}u_{bi}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big]+o_{L_1}(1)
\end{aligned}
$$

By (S.3.7), (S.3.8) and Lemma S.11,

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{n}\mathrm{tr}\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\}\sum_{i=1}^{j-1}u_{bi}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big]^2\\
&\leqslant\mathcal{K}\mathbb{E}\Big[\sum_{i=1}^{j-1}u_{bi}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\Big]^2=o(1).
\end{aligned}
$$

It implies that the second term in the expansion of $\mathcal{J}_j$ is also $o_{L_1}(1)$ uniformly for all j.

As for $\mathcal{J}_j^{(3)}$, due to

$$
\begin{aligned}
&\mathbb{E}|\theta_{ij}(\mathbb{z}_1)|^2=O(n^{-1}),\\
&\mathbb{E}|\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_j\boldsymbol{u}_b|^2=O(1),
\end{aligned}
$$

we get

$$
\begin{aligned}
\mathcal{J}_j^{(3)}&=\mathbb{E}_{j-1}\sum_{i=j+1}^{N}u_{bi}\mathbf{z}_i^T\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_j\boldsymbol{u}_b\\
&=\mathbb{E}_{j-1}\sum_{i=j+1}^{N}u_{bi}\beta_{ij}(\mathbb{z}_1)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_j\boldsymbol{u}_b
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}_{j-1} \sum_{i=j+1}^{N} u_{bi} \beta_{ij}^{\mathrm{tr}} \mathbf{z}_i^T \mathbf{A}_{ij}(\mathbb{z}_1) \underline{\mathbf{A}}_j(\mathbb{z}_2) \underline{\mathbf{Z}}_j \boldsymbol{u}_b \\
&- \mathbb{E}_{j-1} \sum_{i=j+1}^{N} u_{bi} \beta_{ij}(\mathbb{z}_1) \beta_{ij}^{\mathrm{tr}}(\mathbb{z}_1) \theta_{ij}(\mathbb{z}_1) \mathbf{z}_i^T \mathbf{A}_{ij}(\mathbb{z}_1) \underline{\mathbf{A}}_j(\mathbb{z}_2) \underline{\mathbf{Z}}_j \boldsymbol{u}_b \\
&= -\mathbb{E}_{j-1} \sum_{i=j+1}^{N} u_{bi} \beta_{ij}(\mathbb{z}_1) \beta_{ij}^{\mathrm{tr}}(\mathbb{z}_1) \theta_{ij}(\mathbb{z}_1) \mathbf{z}_i^T \mathbf{A}_{ij}(\mathbb{z}_1) \underline{\mathbf{A}}_j(\mathbb{z}_2) \underline{\mathbf{Z}}_j \boldsymbol{u}_b \\
&= o_{L_1}(1).
\end{aligned}
$$

(S.3.6) has been proved.

S.3.1.4. The limit of (S.3.3)

This subsection shows, as $n \to \infty$,

$$
\begin{aligned}
&\sum_{j=1}^{N} \mathbb{E}_{j-1} \mathcal{H}_{[1]}(\mathbb{z}_1, j) \mathcal{H}_{[2]}(\mathbb{z}_2, j) \\
&= n \beta^{\mathbb{E}}(\mathbb{z}_1) \beta^{\mathbb{E}}(\mathbb{z}_2) \sum_{i=1}^{N-1} u_{ai} u_{bi} \sum_{j=i+1}^{N} u_{aj} u_{bj} \mathrm{tr}\left[\mathbb{E}_j \mathbf{A}_j(\mathbb{z}_1) \mathbb{E}_j \mathbf{A}_j(\mathbb{z}_2)\right] \\
&\quad + o_p(1).
\end{aligned} \tag{S.3.9}
$$

Following notation defined in Section S.3.1.3,

$$
\begin{aligned}
&\mathbb{E}_{j-1}\left[\mathbb{E}_j[\mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_1) \mathbf{Z}_j \boldsymbol{u}_a] \mathbb{E}_j[\mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_b]\right] \\
&\quad = \mathbb{E}_{j-1}[\boldsymbol{u}_a^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}_1) \underline{\mathbf{A}}_j(\mathbb{z}_2) \underline{\mathbf{Z}}_j \boldsymbol{u}_b] \\
&\quad = \sum_{i=1}^{j-1} u_{ai} u_{bi} \mathbb{E}_{j-1} \mathbf{z}_i^T \mathbf{A}_j(\mathbb{z}_1) \underline{\mathbf{A}}_j(\mathbb{z}_2) \mathbf{z}_i + \sum_{i=1}^{j-1} u_{ai} \mathbb{E}_{j-1} \mathbf{z}_i^T \mathbf{A}_j(\mathbb{z}_1) \underline{\mathbf{A}}_j(\mathbb{z}_2) \underline{\mathbf{Z}}_{ij} \boldsymbol{u}_b \\
&\qquad + \sum_{i=j+1}^{N} u_{ai} \mathbb{E}_{j-1} \mathbf{z}_i^T \mathbf{A}_j(\mathbb{z}_1) \underline{\mathbf{A}}_j(\mathbb{z}_2) \underline{\mathbf{Z}}_j \boldsymbol{u}_b = \mathcal{J}_j^{(4)} + \mathcal{J}_j^{(5)} + \mathcal{J}_j^{(6)}, \text{ say.}
\end{aligned}
$$

For future use, we instead show the following results in L_2-norm.

$$
\begin{aligned}
&\sup_j \mathbb{E}\Big|\mathcal{J}_j^{(4)} - \beta^{\mathbb{E}}(\mathbb{z}_1) \beta^{\mathbb{E}}(\mathbb{z}_2) \sum_{i=1}^{j-1} u_{ai} u_{bi} \mathrm{tr} \mathbb{E}_j[\mathbf{A}_j(\mathbb{z}_1) \underline{\mathbf{A}}_j(\mathbb{z}_2)]\Big|^2 = o(1), \\
&\sup_j |\mathcal{J}_j^{(5)}|^2 = o(1), \\
&\sup_j |\mathcal{J}_j^{(6)}|^2 = o(1).
\end{aligned}
$$

As for $\mathcal{J}_j^{(4)}$, due to similar arguments to those for $\mathcal{J}_j^{(1)}$, we need to show

$$\mathbb{E}\Big|\sum_{i=1}^{j-1} u_{ai}u_{bi}\Big[\beta_{ij}(\mathbb{z}_1)\underline{\beta}_{ij}(\mathbb{z}_2)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i \\ - \beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathrm{tr}\big[\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\}\big]\Big]\Big|^2 = o(1).$$

It suffices to show

$$\mathbb{E}\frac{1}{n^2}\Big|\beta_{ij}(\mathbb{z}_1)\underline{\beta}_{ij}(\mathbb{z}_2)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i- \\ \beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathrm{tr}\big[\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\}\big]\Big|^2 = o(1).$$

It can be done using Cauchy-Schwarz inequality and the following results.

$$\begin{aligned}
&\mathbb{E}|\beta_{ij}(\mathbb{z}_1) - \beta^{\mathbb{E}}(\mathbb{z}_1)|^\ell = o(1), \quad \ell \geqslant 2,\\
&\mathbb{E}|\underline{\beta}_{ij}(\mathbb{z}_2) - \beta^{\mathbb{E}}(\mathbb{z}_2)|^\ell = o(1), \quad \ell \geqslant 2,\\
&\mathbb{E}|\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i - \frac{1}{n}\mathrm{tr}[\{\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\}]|^\ell = o(1), \quad \ell \geqslant 2.
\end{aligned}$$

Now consider $\mathcal{J}_j^{(5)}$,

$$\begin{aligned}
\mathcal{J}_j^{(5)} &= \mathbb{E}_{j-1}\sum_{i=1}^{j-1} u_{ai}\beta_{ij}(\mathbb{z}_1)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&= \mathbb{E}_{j-1}\sum_{i=1}^{j-1} u_{ai}\beta_{ij}(\mathbb{z}_1)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&- \mathbb{E}_{j-1}\sum_{i=1}^{j-1}\frac{1}{n}u_{ai}\beta_{ij}(\mathbb{z}_1)\underline{\beta}_{ij}(\mathbb{z}_2)\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&= \sum_{i=1}^{j-1} u_{ai}\beta^{\mathbb{E}}(\mathbb{z}_1)\mathbb{E}_{j-1}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b\\
&- \sum_{i=1}^{j-1}\frac{1}{n}u_{ai}\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathbb{E}_{j-1}\mathrm{tr}\{\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\}\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b + o_{L_2}(1).
\end{aligned}$$

The last step is due to

$$\begin{aligned}
&\mathbb{E}|\frac{1}{n}\mathbf{z}_i^T\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\mathbf{z}_i - \frac{1}{n}\mathrm{tr}\{\mathbf{A}_{ij}(\mathbb{z}_1)\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\}|^4 = o(1),\\
&\mathbb{E}|\mathbf{z}_i^T\underline{\mathbf{A}}_{ij}(\mathbb{z}_2)\underline{\mathbf{Z}}_{ij}\boldsymbol{u}_b|^4 = O(1).
\end{aligned}$$

Together with (S.3.7) and (S.3.8),

$$\sup_j \mathbb{E}|\mathcal{J}_j^{(5)}|^2 = o(1).$$

The proof of $\mathcal{J}_j^{(6)} = o_{L_2}(1)$ is very similar to that of $\mathcal{J}_j^{(3)}$. We omit details.

S.3.1.5. The limit of (S.3.4)

In this section, we will show

$$\sum_{j=1}^{N} \mathbb{E}_{j-1}\mathcal{H}_{[1]}(\mathbb{z}_1, j)\mathcal{H}_{[3]}(\mathbb{z}_2, j) \longrightarrow 0, \quad \text{in probability.} \tag{S.3.10}$$

Lemma S.19 indicates

$$\begin{aligned}
&\mathbb{E}_{j-1}\left[\mathbb{E}_j[\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_1)\mathbf{Z}_j\boldsymbol{\alpha}_b]\mathbb{E}_j(\varrho_j(\mathbb{z}_2))\right] \\
&= \frac{\mathbb{E}z_{11}^3}{n}\sum_{i=1}^{p}[\mathbb{E}_j h_i(\mathbb{z}_2, a, j)h_i(\mathbb{z}_2, b, j)][\mathbb{E}_j h_i(\mathbb{z}_1, b, j)]
\end{aligned}$$

where $h_i(\mathbb{z}, a, j)$ and $h_i(\mathbb{z}, b, j)$ are respectively the ith element of $\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a$ and $\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b$. Therefore,

$$(S.3.4) = -\mathbb{E}z_{11}^3\sum_{j=1}^{N} u_{aj}\sum_{i=1}^{p}[\mathbb{E}_j h_i(\mathbb{z}_2, a, j)h_i(\mathbb{z}_2, b, j)][\mathbb{E}_j h_i(\mathbb{z}_1, b, j)].$$

It is sufficient to show

$$\sup_{1\leqslant j\leqslant N}\mathbb{E}\sum_{i=1}^{p}\left|[\mathbb{E}_j h_i(\mathbb{z}_2, a, j)h_i(\mathbb{z}_2, b, j)][\mathbb{E}_j h_i(\mathbb{z}_1, b, j)]\right| \to 0.$$

We first show $h_i(\mathbb{z}, a, j)$ concentrates around its mean. Specifically, we have

$$\sup_{i\leqslant p; j\leqslant N}\mathbb{E}|h_i(\mathbb{z}, a, j) - \mathbb{E}h_i(\mathbb{z}, a, j)|^2 = O(n^{-1}), \tag{S.3.11}$$

$$\sup_{i\leqslant p; j\leqslant N}\mathbb{E}|h_i(\mathbb{z}, a, j) - \mathbb{E}h_i(\mathbb{z}, a, j)|^4 = o(n^{-1}). \tag{S.3.12}$$

To show (S.3.11),

$$\begin{aligned}
&h_i(\mathbb{z}, a, j) - \mathbb{E}h_i(\mathbb{z}, a, j) \\
&= \sum_{\ell=1,\ell\neq j}^{N}[\mathbb{E}_\ell - \mathbb{E}_{\ell-1}]h_i(\mathbb{z}, a, j) \\
&= \sum_{\ell=1,\ell\neq j}^{N}[\mathbb{E}_\ell - \mathbb{E}_{\ell-1}]e_i^T[\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_a - \mathbf{A}_{\ell j}(\mathbb{z})\mathbf{Z}_{\ell j}\boldsymbol{u}_a] \\
&= \sum_{\ell=1,\ell\neq j}^{N}[\mathbb{E}_\ell - \mathbb{E}_{\ell-1}]\{u_{a\ell}e_i^T\mathbf{A}_{\ell j}(\mathbb{z})\mathbf{z}_\ell - \frac{1}{n}e_i^T\mathbf{A}_{\ell j}(\mathbb{z})\mathbf{z}_\ell\mathbf{z}_\ell^T\mathbf{A}_{\ell j}(\mathbb{z})\mathbf{Z}_{\ell j}\boldsymbol{u}_a\beta_{\ell j}(\mathbb{z})
\end{aligned}$$

$$- \frac{u_{a\ell}}{n} e_i^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell \mathbf{z}_\ell^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell \beta_{\ell j}(\mathbb{z})\}$$

By Lemma S.14 and Lemma S.10, for all i, j, ℓ,

$$\mathbb{E}|e_i^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell|^4 = \mathbb{E}|\mathbf{z}_\ell^T \mathbf{A}_{\ell j}(\mathbb{z}) e_i e_i^T \mathbf{A}_{\ell j}(\overline{\mathbb{z}}) \mathbf{z}_\ell|^2 \leqslant \mathcal{K}\mathbb{E}\|\mathbf{A}_{\ell j}(\mathbb{z})\|^4 = O(1).$$

$$\mathbb{E}|\frac{1}{n} e_i^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell|^8 = n^{-8}\mathbb{E}|\mathbf{z}_\ell^T \mathbf{A}_{\ell j}(\mathbb{z}) e_i e_i^T \mathbf{A}_{\ell j}(\overline{\mathbb{z}}) \mathbf{z}_\ell|^4$$
$$\leqslant \mathcal{K}\mathbb{E}\|\mathbf{A}_{\ell j}(\mathbb{z})\|^8 [\varepsilon_n^4 n^{-6} + n^{-8}] = o(n^{-6}).$$

Together with Lemma S.13 and Lemma S.16,

$$\mathbb{E}|\frac{1}{n} e_i^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell \mathbf{z}_\ell^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{Z}_{\ell j} \boldsymbol{u}_a|^2 = O(n^{-2}),$$
$$\mathbb{E}|\frac{1}{n} e_i^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell \mathbf{z}_\ell^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{Z}_{\ell j} \boldsymbol{u}_a|^4 = o(n^{-2}),$$
$$\mathbb{E}|\frac{u_{a\ell}}{n} e_i^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell \mathbf{z}_\ell^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell|^2 \leqslant \mathcal{K}\|\boldsymbol{u}_a\|_{\max}^2 = O(n^{-2}).$$
$$\mathbb{E}|\frac{u_{a\ell}}{n} e_i^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell \mathbf{z}_\ell^T \mathbf{A}_{\ell j}(\mathbb{z}) \mathbf{z}_\ell|^4 = o(n^{-3}).$$

Using Lemma S.8, (S.3.11) and (S.3.12) holds.

Back to our goal, (S.3.11) and (S.3.12) lead to

$$\mathbb{E}\sum_{i=1}^p \Big|[\mathbb{E}_j h_i(\mathbb{z}_2, a, j) h_i(\mathbb{z}_2, b, j)][\mathbb{E}_j h_i(\mathbb{z}_1, b, j) - \mathbb{E}h_i(\mathbb{z}_1, b, j)]\Big|$$
$$\leqslant \mathbb{E}\sum_{i=1}^p \Big|[\mathbb{E}_j h_i(\mathbb{z}_2, a, j) \mathbb{E}h_i(\mathbb{z}_2, b, j)][\mathbb{E}_j h_i(\mathbb{z}_1, b, j) - \mathbb{E}h_i(\mathbb{z}_1, b, j)]\Big|$$
$$+ \mathbb{E}\sum_{i=1}^p \Big|[\mathbb{E}_j h_i(\mathbb{z}_2, a, j)\{h_i(\mathbb{z}_2, b, j) - \mathbb{E}h_i(\mathbb{z}_2, b, j)\}]\cdot$$
$$[\mathbb{E}_j h_i(\mathbb{z}_1, b, j) - \mathbb{E}h_i(\mathbb{z}_1, b, j)]\Big|$$

The second term above is such that

$$\Big(\mathbb{E}\sum_{i=1}^p \Big|[\mathbb{E}_j h_i(\mathbb{z}_2, a, j)\{h_i(\mathbb{z}_2, b, j) - \mathbb{E}h_i(\mathbb{z}_2, b, j)\}]\cdot$$
$$[\mathbb{E}_j h_i(\mathbb{z}_1, b, j) - \mathbb{E}h_i(\mathbb{z}_1, b, j)]\Big|\Big)^2$$
$$\leqslant \sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_2, a, j)|^2 \sum_{i=1}^p \mathbb{E}\Big|[h_i(\mathbb{z}_2, b, j) - \mathbb{E}h_i(\mathbb{z}_2, b, j)]\cdot$$
$$[\mathbb{E}_j h_i(\mathbb{z}_1, b, j) - \mathbb{E}h_i(\mathbb{z}_1, b, j)]\Big|^2$$
$$\leqslant \sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_2, a, j)|^2 \sum_{i=1}^p \Big(\mathbb{E}|[h_i(\mathbb{z}_2, b, j) - \mathbb{E}h_i(\mathbb{z}_2, b, j)]|^4\Big)^{1/2}.$$

$$\Big(\mathbb{E}|[h_i(\mathbb{z}_1,b,j)-\mathbb{E}h_i(\mathbb{z}_1,b,j)]|^4\Big)^{1/2}=o(1).$$

where the last step is due to (S.3.12) and

$$\sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_2,a,j)|^2=\mathbb{E}\boldsymbol{u}_a^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{A}_j(\overline{\mathbb{z}}_2)\mathbf{Z}_j\boldsymbol{u}_a=O(1). \tag{S.3.13}$$

The first term is such that

$$\begin{aligned}&\Big(\mathbb{E}\sum_{i=1}^p\Big|[\mathbb{E}_jh_i(\mathbb{z}_2,a,j)\mathbb{E}h_i(\mathbb{z}_2,b,j)][\mathbb{E}_jh_i(\mathbb{z}_1,b,j)-\mathbb{E}h_i(\mathbb{z}_1,b,j)]\Big|\Big)^2\\&\leqslant\sum_{i=1}^p\mathbb{E}|h_i(\mathbb{z}_2,a,j)|^2\sum_{i=1}^p|\mathbb{E}h_i(\mathbb{z}_2,b,j)|^2\mathbb{E}|h_i(\mathbb{z}_1,b,j)-\mathbb{E}h_i(\mathbb{z}_1,b,j)|^2\\&=o(1),\end{aligned}$$

The last line is due to

$$\sum_{i=1}^p|\mathbb{E}h_i(\mathbb{z}_2,b,j)|^2\mathbb{E}|h_i(\mathbb{z}_1,b,j)-\mathbb{E}h_i(\mathbb{z}_1,b,j)|^2\leqslant\mathcal{K}\frac{1}{n}\sum_{i=1}^p\mathbb{E}|h_i(\mathbb{z}_2,b,j)|^2=o(1),$$

which is a consequence of (S.3.11).

Therefore,

$$\begin{aligned}&\sup_{1\leqslant j\leqslant N}\mathbb{E}\sum_{i=1}^p\Big|[\mathbb{E}_jh_i(\mathbb{z}_2,a,j)h_i(\mathbb{z}_2,b,j)][\mathbb{E}_jh_i(\mathbb{z}_1,b,j)]\Big|\\&=\sup_{1\leqslant j\leqslant N}\mathbb{E}\sum_{i=1}^p\Big|[\mathbb{E}_jh_i(\mathbb{z}_2,a,j)h_i(\mathbb{z}_2,b,j)]\mathbb{E}h_i(\mathbb{z}_1,b,j)\Big|+o(1)\end{aligned}$$

$$\begin{aligned}&\sup_{1\leqslant j\leqslant N}\sum_{i=1}^p|\mathbb{E}h_i(\mathbb{z}_1,b,j)|\mathbb{E}\Big|h_i(\mathbb{z}_2,a,j)h_i(\mathbb{z}_2,b,j)\Big|\\&\leqslant\sup_{j\leqslant N;i\leqslant p}|\mathbb{E}h_i(\mathbb{z}_1,b,j)|\sup_{1\leqslant j\leqslant N}\sum_{i=1}^p\mathbb{E}\Big|h_i(\mathbb{z}_2,a,j)h_i(\mathbb{z}_2,b,j)\Big|\\&\leqslant\sup_{j\leqslant N;i\leqslant p}|\mathbb{E}h_i(\mathbb{z}_1,b,j)|\sup_{1\leqslant j\leqslant N}\Big(\mathbb{E}\boldsymbol{u}_a^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{A}_j(\overline{\mathbb{z}}_2)\mathbf{Z}_j\boldsymbol{u}_a\Big)^{1/2}\cdot\\&\qquad\Big(\mathbb{E}\boldsymbol{u}_b^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{A}_j(\overline{\mathbb{z}}_2)\mathbf{Z}_j\boldsymbol{u}_b\Big)^{1/2}.\end{aligned}$$

It implies that we only need to show

$$\sup_{j\leqslant N;i\leqslant p}|\mathbb{E}h_i(\mathbb{z}_1,b,j)|\to 0. \tag{S.3.14}$$

For all j,

$$\begin{aligned}\mathbb{E}h_i(\mathbb{z}, b, j) &= \sum_{\ell \neq j}^{N} u_{b\ell} \mathbb{E} e_i^T \mathbf{A}_{12}(\mathbb{z}) \mathbf{z}_1 \beta_{12}(\mathbb{z}) \\ &= \sum_{\ell \neq j}^{N} u_{b\ell} \mathbb{E} e_i^T \mathbf{A}_{12}(\mathbb{z}) \mathbf{z}_1 \beta_{12}(\mathbb{z}) \{1 + \frac{1}{n} \mathbb{E}\mathrm{tr}\mathbf{A}_{12}(\mathbb{z})\}^{-1} (\frac{1}{n} \mathbf{z}_1^T \mathbf{A}_{12} \mathbf{z}_1 - \frac{1}{n} \mathbb{E}\mathrm{tr}\mathbf{A}_{12}(\mathbb{z}))\end{aligned}$$

By Lemma S.13 and Lemma S.14,

$$\begin{aligned}&\mathbb{E}|e_i^T \mathbf{A}_{12}(\mathbb{z}) \mathbf{z}_1 \beta_{12}(\mathbb{z}) \{1 + \frac{1}{n} \mathbb{E}\mathrm{tr}\mathbf{A}_{12}(\mathbb{z})\}^{-1} (\frac{1}{n} \mathbf{z}_1^T \mathbf{A}_{12} \mathbf{z}_1 - \frac{1}{n} \mathbb{E}\mathrm{tr}\mathbf{A}_{12}(\mathbb{z}))| \\ &\leqslant \mathcal{K} \mathbb{E}|e_i^T \mathbf{A}_{12}(\mathbb{z}) \mathbf{z}_1 (\frac{1}{n} \mathbf{z}_1^T \mathbf{A}_{12} \mathbf{z}_1 - \frac{1}{n} \mathbb{E}\mathrm{tr}\mathbf{A}_{12}(\mathbb{z}))| = o(1).\end{aligned}$$

Consequently $\sup_j |\mathbb{E}h_i(\mathbb{z}, b, j)| \to 0$. We later prove a stronger result in (S.3.31). We have shown

$$\sup_{1 \leqslant j \leqslant N} \mathbb{E} \sum_{i=1}^{p} \Big| [\mathbb{E}_j h_i(\mathbb{z}_2, a, j) h_i(\mathbb{z}_2, b, j)][\mathbb{E}_j h_i(\mathbb{z}_1, b, j)] \Big| \to 0.$$

Thus,

$$\sum_{j=1}^{N} \mathbb{E}_{j-1} \mathcal{H}_{[1]}(\mathbb{z}_1, j) \mathcal{H}_{[3]}(\mathbb{z}_2, j) \longrightarrow 0, \quad \text{in probability.}$$

S.3.1.6. The limit of (S.3.5)

In this subsection, we show, as $n \to \infty$,

$$\begin{aligned}&\sum_{j=1}^{N} \mathbb{E}_{j-1} \mathcal{H}_{[3]}(\mathbb{z}_1, j) \mathcal{H}_{[3]}(\mathbb{z}_2, j) \\ &= (\beta^{\mathbb{E}}(\mathbb{z}_1) \beta^{\mathbb{E}}(\mathbb{z}_2))^2 \frac{1}{n} \sum_{j=1}^{N} [\mathrm{tr} \mathbb{E}_j \mathbf{A}_j(\mathbb{z}_1) \mathbb{E}_j \mathbf{A}_j(\mathbb{z}_2)]^2 \cdot \\ &\quad [(\sum_{i=1}^{j-1} u_{ai} u_{bi})^2 + \sum_{i=1}^{j-1} u_{ai}^2 \sum_{i=1}^{j-1} u_{bi}^2] + o_p(1).\end{aligned} \tag{S.3.15}$$

By Lemma S.20,

$$\begin{aligned}&\sum_{j=1}^{N} \mathbb{E}_{j-1} \mathcal{H}_{[3]}(\mathbb{z}_1, j) \mathcal{H}_{[3]}(\mathbb{z}_2, j) \\ &= (\mathbb{E} z_{11}^4 - 3) \frac{1}{n} \sum_{j=1}^{N} \sum_{i=1}^{p} [\mathbb{E}_j h_i(\mathbb{z}_1, a, j) h_i(\mathbb{z}_1, b, j)][\mathbb{E}_j h_i(\mathbb{z}_2, a, j) h_i(\mathbb{z}_2, b, j)]\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{j=1}^{N}\mathbb{E}_j[\sum_{i=1}^{p}h_i(\mathbb{z}_1,a,j)\underline{h}_i(\mathbb{z}_2,b,j)\sum_{i=1}^{p}h_i(\mathbb{z}_1,a,j)\underline{h}_i(\mathbb{z}_2,b,j)]\\
&+\frac{1}{n}\sum_{j=1}^{N}\mathbb{E}_j[\sum_{i=1}^{p}h_i(\mathbb{z}_1,b,j)\underline{h}_i(\mathbb{z}_2,b,j)\sum_{i=1}^{p}h_i(\mathbb{z}_1,a,j)\underline{h}_i(\mathbb{z}_2,a,j)]\\
&=\mathcal{J}_j^{(7)}+\mathcal{J}_j^{(8)}+\mathcal{J}_j^{(9)}.
\end{aligned}
$$

where similar to $h_i(\mathbb{z},a,j)$ and $h_i(\mathbb{z},b,j)$, $\underline{h}_i(\mathbb{z},a,j)$ and $\underline{h}_i(\mathbb{z},b,j)$ are respectively the ith elements of $\underline{\mathbf{A}}_j(\mathbb{z})\underline{\mathbf{Z}}_j\boldsymbol{u}_a$ and $\underline{\mathbf{A}}_j(\mathbb{z})\underline{\mathbf{Z}}_j\boldsymbol{u}_b$.

Consider $\mathcal{J}_j^{(7)}$ first. The target is to show

$$
\frac{1}{n}\sum_{j=1}^{N}\sum_{i=1}^{p}[\mathbb{E}_jh_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)][\mathbb{E}_jh_i(\mathbb{z}_2,a,j)h_i(\mathbb{z}_2,b,j)]=o_p(1). \tag{S.3.16}
$$

Split it into four terms,

$$
\begin{aligned}
&\sum_{i=1}^{p}[\mathbb{E}_jh_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)][\mathbb{E}_jh_i(\mathbb{z}_2,a,j)h_i(\mathbb{z}_2,b,j)]\\
&=\sum_{i=1}^{p}[\mathbb{E}_jh_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)][\mathbb{E}h_i(\mathbb{z}_2,a,j)][\mathbb{E}h_i(\mathbb{z}_2,b,j)]\\
&+\sum_{i=1}^{p}[\mathbb{E}_jh_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)][\mathbb{E}h_i(\mathbb{z}_2,a,j)][\mathbb{E}_jh_i(\mathbb{z}_2,b,j)-\mathbb{E}h_i(\mathbb{z}_2,b,j)]\\
&+\sum_{i=1}^{p}[\mathbb{E}_jh_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)][\mathbb{E}_jh_i(\mathbb{z}_2,a,j)-\mathbb{E}h_i(\mathbb{z}_2,a,j)][\mathbb{E}h_i(\mathbb{z}_2,b,j)]\\
&+\sum_{i=1}^{p}[\mathbb{E}_jh_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)]\cdot\\
&\qquad\mathbb{E}_j\{[h_i(\mathbb{z}_2,a,j)-\mathbb{E}h_i(\mathbb{z}_2,a,j)][h_i(\mathbb{z}_2,b,j)-\mathbb{E}h_i(\mathbb{z}_2,b,j)]\}\\
&=d_4^{(1)}+d_4^{(2)}+d_4^{(3)}+d_4^{(4)},\ \text{say.}
\end{aligned}
$$

Combining (S.3.13) and (S.3.14),

$$
\begin{aligned}
\mathbb{E}|d_4^{(1)}|&\leqslant\sup_i(|\mathbb{E}h_i(\mathbb{z}_2,a,j)|)\sup_i(|\mathbb{E}h_i(\mathbb{z}_2,b,j)|)\sum_{i=1}^{p}\mathbb{E}|h_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)|\\
&\leqslant\sup_i(|\mathbb{E}h_i(\mathbb{z}_2,a,j)|)\sup_i(|\mathbb{E}h_i(\mathbb{z}_2,b,j)|)\cdot\\
&\qquad\Big(\sum_{i=1}^{p}\mathbb{E}|h_i(\mathbb{z}_1,a,j)|^2\sum_{i=1}^{p}\mathbb{E}|h_i(\mathbb{z}_1,b,j)|^2\Big)^{1/2}\longrightarrow 0.
\end{aligned}
$$

Next, using (S.3.11) and (S.3.12),

$$\mathbb{E}|d_4^{(2)}| \leqslant \sup_i(|\mathbb{E}h_i(\mathbb{z}_2,a,j)|)\Big[\sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)|^2\Big]^{1/2}\cdot \Big[\sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_2,b,j)-\mathbb{E}h_i(\mathbb{z}_2,b,j)|^2\Big]^{1/2} \longrightarrow 0.$$

Similarly,

$$\mathbb{E}|d_4^{(3)}| \leqslant \sup_i(|\mathbb{E}h_i(\mathbb{z}_2,b,j)|)\Big[\sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)|^2\Big]^{1/2} \Big[\sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_2,a,j)-\mathbb{E}h_i(\mathbb{z}_2,a,j)|^2\Big]^{1/2} \longrightarrow 0.$$

$$\mathbb{E}|d_4^{(4)}| \leqslant (\sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_1,a,j)h_i(\mathbb{z}_1,b,j)|^2)^{1/2}\Big[\sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_2,a,j)-\mathbb{E}h_i(\mathbb{z}_2,a,j)|^4\cdot \sum_{i=1}^p \mathbb{E}|h_i(\mathbb{z}_2,b,j)-\mathbb{E}h_i(\mathbb{z}_2,b,j)|^4\Big]^{1/4} \longrightarrow 0.$$

Consider $\mathcal{J}_j^{(8)}$. The target is to show

$$\begin{aligned}&\mathbb{E}_j\sum_{i=1}^p h_i(\mathbb{z}_1,b,j)\underline{h}_i(\mathbb{z}_2,a,j)\sum_{i=1}^p \underline{h}_i(\mathbb{z}_2,b,j)h_i(\mathbb{z}_1,a,j)\\ &=\mathbb{E}_j\sum_{i=1}^p h_i(\mathbb{z}_1,b,j)\underline{h}_i(\mathbb{z}_2,a,j)\mathbb{E}_j\sum_{i=1}^p \underline{h}_i(\mathbb{z}_2,b,j)h_i(\mathbb{z}_1,a,j)+o_{L_1}(1).\end{aligned} \tag{S.3.17}$$

First we substitute $h_i(\mathbb{z}_1,a,j)$ with $\mathbb{E}_jh_i(\mathbb{z}_1,a,j)$ in $\mathcal{J}_j^{(8)}$ and show the resulting difference is small. That is to show

$$\begin{aligned}&\mathbb{E}_j\sum_{i=1}^p h_i(\mathbb{z}_1,b,j)\underline{h}_i(\mathbb{z}_2,a,j)\sum_{i=1}^p \underline{h}_i(\mathbb{z}_2,b,j)h_i(\mathbb{z}_1,a,j)\\ &=\mathbb{E}_j\sum_{i=1}^p h_i(\mathbb{z}_1,b,j)\underline{h}_i(\mathbb{z}_2,a,j)\sum_{i=1}^p \underline{h}_i(\mathbb{z}_2,b,j)\mathbb{E}_jh_i(\mathbb{z}_1,a,j)+o_{L_1}(1).\end{aligned} \tag{S.3.18}$$

It suffices to prove

$$\mathbb{E}\Big|\sum_{i=1}^p h_i(\mathbb{z}_1,b,j)\underline{h}_i(\mathbb{z}_2,a,j)\sum_{i=1}^p \underline{h}_i(\mathbb{z}_2,b,j)[h_i(\mathbb{z}_1,a,j)-\mathbb{E}_jh_i(\mathbb{z}_1,a,j)]\Big| = o(1).$$

Due to very similar arguments to (S.3.13),

$$
\begin{aligned}
&\mathbb{E}\Big|\sum_{i=1}^{p} h_i(\mathbb{z}_1,b,j)\underline{h}_i(\mathbb{z}_2,a,j)\Big|^2\\
&\leqslant \Big(\mathbb{E}\Big(\sum_{i=1}^{p}|h_i(\mathbb{z}_1,b,j)|^2\Big)^2\mathbb{E}\Big(\sum_{i=1}^{p}|h_i(\mathbb{z}_2,a,j)|^2\Big)^2\Big)^{1/2}\\
&=\Big(\mathbb{E}|\boldsymbol{u}_b^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z}_1)\mathbf{A}_j(\overline{\mathbb{z}}_1)\mathbf{Z}_j\boldsymbol{u}_b|^2\mathbb{E}|\boldsymbol{u}_a^T\mathbf{Z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{A}_j(\overline{\mathbb{z}}_2)\mathbf{Z}_j\boldsymbol{u}_a|^2\Big)^{1/2}=O(1).
\end{aligned}
$$

It suffices to show

$$
\mathbb{E}\Big|\sum_{i=1}^{p}\underline{h}_i(\mathbb{z}_2,b,j)[h_i(\mathbb{z}_1,a,j)-\mathbb{E}_j h_i(\mathbb{z}_1,a,j)]\Big|^2=o(1).
$$

Fixing j, we consider the σ-algebra generated by

$$
\{\mathbf{z}_1,\ldots,\mathbf{z}_{j-1},\mathbf{z}_{j+1},\ldots,\mathbf{z}_n\}\cup\{\underline{\mathbf{z}}_{j+1},\ldots\underline{\mathbf{z}}_N\}
$$

Define a filtration (in l) as

$$
\sigma(\mathbf{z}_1,\ldots,\mathbf{z}_{j-1},\underline{\mathbf{z}}_{j+1},\ldots\underline{\mathbf{z}}_N,\mathbf{z}_l,\ldots,\mathbf{z}_N),\quad l=j+1,\ldots,N.
$$

Rewrite $\sum_{i=1}^{p}\underline{h}_i(\mathbb{z}_2,b,j)[h_i(\mathbb{z}_1,a,j)-\mathbb{E}_j h_i(\mathbb{z}_1,a,j)]$ as a sum of martingale difference sequence with respect to the filtration.

Denote $\mathbb{E}_{j,l}$ to be the conditional expectation with respect to

$$
\sigma(\mathbf{z}_1,\ldots,\mathbf{z}_{j-1},\underline{\mathbf{z}}_{j+1},\ldots\underline{\mathbf{z}}_N,\mathbf{z}_l,\ldots,\mathbf{z}_N)
$$

for $l=j+1,\ldots,N$.

$$
\begin{aligned}
&\sum_{i=1}^{p}\underline{h}_i(\mathbb{z}_2,b,j)[h_i(\mathbb{z}_1,a,j)-\mathbb{E}_j h_i(\mathbb{z}_1,a,j)]\\
&=\sum_{l=j+1}^{N}(\mathbb{E}_{j,l}-\mathbb{E}_{j,l-1})\boldsymbol{u}_b^T\underline{\mathbf{Z}}_j^T\underline{\mathbf{A}}_j(\mathbb{z}_2)[\mathbf{A}_j(\mathbb{z}_1)\mathbf{Z}_j\boldsymbol{u}_a-\mathbf{A}_{ij}(\mathbb{z}_1)\mathbf{Z}_{ij}\boldsymbol{u}_a]\\
&=\sum_{l=j+1}^{N}(\mathbb{E}_{j,l}-\mathbb{E}_{j,l-1})u_{al}\boldsymbol{u}_b^T\underline{\mathbf{Z}}_j^T\underline{\mathbf{A}}_j(\mathbb{z}_2)\mathbf{A}_{lj}(\mathbb{z}_1)\mathbf{z}_l\beta_{lj}(\mathbb{z}_1)\\
&\quad-\frac{1}{n}\sum_{l=j+1}^{N}(\mathbb{E}_{j,l}-\mathbb{E}_{j,l-1})\boldsymbol{u}_b^T\underline{\mathbf{Z}}_j^T\underline{\mathbf{A}}_j(\mathbb{z}_2)\mathbf{A}_{lj}(\mathbb{z}_1)\mathbf{z}_l\mathbf{z}_l^T\mathbf{A}_{lj}(\mathbb{z}_1)\mathbf{Z}_{lj}\boldsymbol{u}_a\beta_{lj}(\mathbb{z}_1)\\
&=d_5^{(1)}+d_5^{(2)},\quad\text{say.}
\end{aligned}
$$

By Lemma S.8, Lemma S.10 and Lemma S.16,

$$\mathbb{E}|d_5^{(1)}|^2 \leqslant \mathcal{K} \sum_{l=j+1}^{N} |u_{al}|^2 \mathbb{E}|\boldsymbol{u}_b^T \underline{\mathbf{Z}}_j^T \underline{\mathbf{A}}_j(\mathbb{z}_2)\mathbf{A}_{lj}(\mathbb{z}_1)\mathbf{z}_l \beta_{lj}(\mathbb{z}_1)|^2 = O(n^{-1}).$$

$$\mathbb{E}|d_5^{(2)}|^2 \leqslant \mathcal{K}\frac{1}{n^2} \sum_{l=j+1}^{N} \mathbb{E}|\boldsymbol{u}_b^T \underline{\mathbf{Z}}_j^T \underline{\mathbf{A}}_j(\mathbb{z}_2)\mathbf{A}_{lj}(\mathbb{z}_1)\mathbf{z}_l\mathbf{z}_l^T \mathbf{A}_{ij}(\mathbb{z}_1)\mathbf{Z}_{lj}\boldsymbol{u}_a \beta_{lj}(\mathbb{z}_1)|^2$$

$$\leqslant \mathcal{K}\frac{1}{n^2} \sum_{l=j+1}^{N} \{\mathbb{E}|\boldsymbol{u}_b^T \underline{\mathbf{Z}}_j^T \underline{\mathbf{A}}_j(\mathbb{z}_2)\mathbf{A}_{lj}(\mathbb{z}_1)\mathbf{z}_l|^2 \mathbb{E}|\mathbf{z}_l^T \mathbf{A}_{lj}(\mathbb{z}_1)\mathbf{Z}_{lj}\boldsymbol{u}_a|^2\}^{1/2} = O(n^{-1}).$$

The proof of (S.3.18) is done.

We next continue to substitute $\underline{h}_i(\mathbb{z}_1, b, j)$ with $\mathbb{E}_j \underline{h}_i(\mathbb{z}_1, b, j)$ in

$$\mathbb{E}_j \sum_{i=1}^{p} h_i(\mathbb{z}_1, b, j)\underline{h}_i(\mathbb{z}_2, a, j) \sum_{i=1}^{p} \underline{h}_i(\mathbb{z}_2, b, j)\mathbb{E}_j h_i(\mathbb{z}_1, a, j)$$

and show the following line is $o(1)$.

$$\mathbb{E}\Big|\mathbb{E}_j \sum_{i=1}^{p} h_i(\mathbb{z}_1, b, j)\underline{h}_i(\mathbb{z}_2, a, j) \sum_{i=1}^{p} [\underline{h}_i(\mathbb{z}_2, b, j) - \mathbb{E}_j \underline{h}_i(\mathbb{z}_2, b, j)]\mathbb{E}_j h_i(\mathbb{z}_1, a, j)\Big|.$$

It can be done along very similar lines to the proof of (S.3.18). Therefore, we omit details.

Finally, we proved

$$\mathbb{E}_j \sum_{i=1}^{p} h_i(\mathbb{z}_1, b, j)\underline{h}_i(\mathbb{z}_2, a, j)\mathbb{E}_j \sum_{i=1}^{p} \underline{h}_i(\mathbb{z}_2, b, j)h_i(\mathbb{z}_1, a, j)$$
$$= \Big[\mathbb{E}_{j-1}\boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_j \boldsymbol{u}_a\Big]\Big[\mathbb{E}_{j-1}\boldsymbol{u}_b^T \underline{\mathbf{Z}}_j^T \underline{\mathbf{A}}_j(\mathbb{z}_2)\mathbf{A}_j(\mathbb{z}_1)\mathbf{Z}_j \boldsymbol{u}_a\Big] + o_{L_1}(1).$$

In Section S.3.1.4, we proved L_2-convergence of $\mathbb{E}_{j-1}\boldsymbol{u}_b^T \mathbf{Z}_j^T \mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)\underline{\mathbf{Z}}_j \boldsymbol{u}_a$. Therefore, it is straightforward that

$$\mathbb{E}_j \sum_{i=1}^{p} h_i(\mathbb{z}_1, b, j)\underline{h}_i(\mathbb{z}_2, a, j)\mathbb{E}_j \sum_{i=1}^{p} \underline{h}_i(\mathbb{z}_2, b, j)h_i(\mathbb{z}_1, a, j)$$
$$= \Big(\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathrm{tr}\mathbb{E}_j[\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)]\Big)^2 \Big(\sum_{i=1}^{j-1} u_{ai}u_{bi}\Big)^2 + o_{L_1}(1).$$

Consider $\mathcal{J}_j^{(9)}$. Repeat the arguments for $\mathcal{J}_j^{(8)}$ with limited modifications,

$$\mathbb{E}_j \sum_{i=1}^{p} h_i(\mathbb{z}_1, b, j)\underline{h}_i(\mathbb{z}_2, b, j)\mathbb{E}_j \sum_{i=1}^{p} \underline{h}_i(\mathbb{z}_2, a, j)h_i(\mathbb{z}_1, a, j)$$
$$= \Big(\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\mathrm{tr}\mathbb{E}_j[\mathbf{A}_j(\mathbb{z}_1)\underline{\mathbf{A}}_j(\mathbb{z}_2)]\Big)^2 (\sum_{i=1}^{j-1} u_{ai}u_{ai})(\sum_{i=1}^{j-1} u_{bi}u_{bi}) + o_{L_1}(1).$$

The proof of (S.3.15) is complete.

S.3.1.7. Calculation of asymptotic variance

Summarizing Section S.3.1.3 – Section S.3.1.6,

$$
\begin{aligned}
&\sum_{k=1}^{3}\sum_{k'=1}^{3}\sum_{j=2}^{N}\mathbb{E}_{j-1}\mathcal{H}_{[k]}(\mathbb{z}_1,j)\mathcal{H}_{[k']}(\mathbb{z}_2,j) = \\
&n\beta^{\mathbb{E}}(\mathbb{z}_1)\beta^{\mathbb{E}}(\mathbb{z}_2)\sum_{j=2}^{N}\mathrm{tr}[\mathbb{E}_j\mathbf{A}_j(\mathbb{z}_1)\mathbb{E}_j\mathbf{A}_j(\mathbb{z}_2)]\sum_{i=1}^{j-1}(u_{ai}^2u_{bj}^2+u_{aj}^2u_{bi}^2+2u_{ai}u_{bj}u_{aj}u_{bi}) \\
&+\frac{1}{n}\beta^{\mathbb{E}}(\mathbb{z}_1)^2\beta^{\mathbb{E}}(\mathbb{z}_2)^2\sum_{j=2}^{N}\{\mathrm{tr}[\mathbb{E}_j\mathbf{A}_j(\mathbb{z}_1)\mathbb{E}_j\mathbf{A}_j(\mathbb{z}_2)]\}^2\cdot \\
&\qquad\Big[(\sum_{i=1}^{j-1}u_{ai}u_{bi})^2+\sum_{i=1}^{j-1}u_{ai}^2\sum_{i=1}^{j-1}u_{bi}^2\Big]+o_p(1).
\end{aligned}
$$

We next try to find the limit of the right-hand side.

Recall the definition of $\underline{m}_p^0(\mathbb{z})$ in Section S.2. It is proved in Bai and Silverstein (2004, (2.18)) that

$$
\begin{aligned}
&\mathrm{tr}[\{\mathbb{E}_j\mathbf{A}_j(\mathbb{z}_1)\}\mathbf{A}_j(\mathbb{z}_2)]\Big\{1-\frac{j-1}{N^2}\underline{m}_p^0(\mathbb{z}_1)\underline{m}_p^0(\mathbb{z})\mathrm{tr}\Big[(I+\underline{m}_p^0(\mathbb{z}_2)\Sigma_p)^{-1}\Sigma_p\ \cdot \\
&(I+\underline{m}_p^0(\mathbb{z}_1)\Sigma_p)^{-1}\Sigma_p\Big]\Big\} = \frac{1}{\mathbb{z}_1\mathbb{z}_2}\mathrm{tr}[(I+\underline{m}_p^0(\mathbb{z}_2)\Sigma_p)^{-1}\Sigma_p(I+\underline{m}_p^0(\mathbb{z}_1)\Sigma_p)^{-1}\Sigma_p] \\
&+o_{L_1}(1).
\end{aligned}
$$

It is worth mentioning that in Bai and Silverstein (2004), the definition of the sample covariance matrix is $\frac{1}{N}\Sigma_p^{1/2}\mathbf{Z}\mathbf{Z}^T\Sigma_p^{T/2}$, while in this paper $\widetilde{\mathbf{\Sigma}}_p=\frac{1}{n}\Sigma_p^{1/2}\mathbf{Z}\mathbf{Z}^T\Sigma_p^{T/2}$. We shall not distinguish the difference because

$$
\Big(\frac{1}{N}\Sigma_p^{1/2}\mathbf{Z}\mathbf{Z}^T\Sigma_p^{T/2}-\mathbb{z}I\Big)^{-1}=\frac{N}{n}\Big(\widetilde{\mathbf{\Sigma}}_p-\frac{n}{N}\mathbb{z}I\Big)^{-1},
$$

and $\underline{m}_p^0(\mathbb{z})$ is continuous in $\mathbb{z}$.

It is also proved in Bai and Silverstein (2004, (2.17)) that

$$
\frac{1}{1+N^{-1}\mathbb{E}\mathrm{tr}[\mathbf{A}_1(\mathbb{z})]}+\mathbb{z}\underline{m}_p^0(\mathbb{z})=O(n^{-1/2}). \tag{S.3.19}
$$

Note

$$
\begin{aligned}
&\frac{1}{n}\underline{m}_p^0(\mathbb{z}_1)\underline{m}_p^0(\mathbb{z}_2)\mathrm{tr}\Big[(I+\underline{m}_p^0(\mathbb{z}_2)\Sigma_p)^{-1}\Sigma_p(I+\underline{m}_p^0(\mathbb{z}_1)\Sigma_p)^{-1}\Sigma_p\Big] \\
&=\gamma_n\underline{m}_p^0(\mathbb{z}_1)\underline{m}_p^0(\mathbb{z}_2)\int\frac{\tau^2dF^{\Sigma_p}(\tau)}{[1+\tau\underline{m}_p^0(\mathbb{z}_1)][1+\tau\underline{m}_p^0(\mathbb{z}_2)]} \\
&=1+\frac{\underline{m}_p^0(\mathbb{z}_1)\underline{m}_p^0(\mathbb{z}_2)(\mathbb{z}_1-\mathbb{z}_2)}{\underline{m}_p^0(\mathbb{z}_2)-\underline{m}_p^0(\mathbb{z}_1)}.
\end{aligned}
$$

We get

$$\mathbb{E}\Big|\beta^{\mathbb{E}}(z_1)\beta^{\mathbb{E}}(z_2)\frac{1}{n}\mathrm{tr}[\mathbb{E}_j\mathbf{A}_j(z_1)\mathbb{E}_j\mathbf{A}_j(z_2)] - \frac{\mathcal{D}}{1-\frac{j-1}{N}\mathcal{D}}\Big| = o(1),$$

where

$$\mathcal{D} = 1 + \frac{\underline{m}_p^0(z_1)\underline{m}_p^0(z_2)(z_1-z_2)}{\underline{m}_p^0(z_2)-\underline{m}_p^0(z_1)}.$$

Define

$$\mathcal{P}_j = \frac{\mathcal{D}}{1-j\mathcal{D}/N}.$$

Therefore,

$$\begin{aligned}
&\sum_{i=1}^{3}\sum_{i'=1}^{3}\sum_{j=2}^{N}\mathbb{E}_{j-1}\mathcal{H}_{[i]}(z_1,j)\mathcal{H}_{[i']}(z_2,j)\\
=&n^2\sum_{j=2}^{N}\mathcal{P}_{j-1}\sum_{i=1}^{j-1}(u_{ai}^2u_{bj}^2+u_{aj}^2u_{bi}^2+2u_{ai}u_{bj}u_{aj}u_{bi})\\
&+n\sum_{j=2}^{N}\mathcal{P}_{j-1}^2\Big[(\sum_{i=1}^{j-1}u_{ai}u_{bi})^2+\sum_{i=1}^{j-1}u_{ai}^2\sum_{i=1}^{j-1}u_{bi}^2\Big]+o_p(1).\\
=&N^2\sum_{j=1}^{N}\mathcal{P}_{j}\sum_{i=1}^{j}(u_{ai}^2u_{bj}^2+u_{aj}^2u_{bi}^2+2u_{ai}u_{bj}u_{aj}u_{bi})\\
&+N\sum_{j=1}^{N}\mathcal{P}_{j}^2\Big[(\sum_{i=1}^{j}u_{ai}u_{bi})^2+\sum_{i=1}^{j}u_{ai}^2\sum_{i=1}^{j}u_{bi}^2\Big]+o_p(1).
\end{aligned}$$

The convergence of $\{\beta^{\mathbb{E}}(z_1)\beta^{\mathbb{E}}(z_2)\frac{1}{n}\mathrm{tr}[\mathbb{E}_j\mathbf{A}_j(z_1)\mathbb{E}_j\mathbf{A}_j(z_2)]\}^2$ follows from the fact that

$$\beta^{\mathbb{E}}(z_1)\beta^{\mathbb{E}}(z_2)\frac{1}{n}\mathrm{tr}[\mathbb{E}_j\mathbf{A}_j(z_1)\mathbb{E}_j\mathbf{A}_j(z_2)]$$

is bounded for any fixed z_1 and z_2 with non-zero imaginary part.

Define

$$\mathcal{O}_\ell = \sum_{i=1}^{\ell}(u_{ai}^2u_{b\ell}^2+u_{a\ell}^2u_{bi}^2+2u_{ai}u_{b\ell}u_{a\ell}u_{bi}).$$

The following result indicates $(\sum_{i=1}^{j}u_{ai}u_{bi})^2+\sum_{i=1}^{j}u_{ai}^2\sum_{i=1}^{j}u_{bi}^2$ is approximately the sum of $\mathcal{O}_\ell$, $\ell=1,\dots,j$.

$$\sum_{\ell=1}^{j}\mathcal{O}_\ell = \sum_{\ell=1}^{j}\sum_{i=1}^{\ell}(u_{ai}^2u_{b\ell}^2+u_{a\ell}^2u_{bi}^2+2u_{ai}u_{b\ell}u_{a\ell}u_{bi})$$

$$
\begin{aligned}
&= \sum_{i \leqslant \ell}^{j} (u_{ai}^2 u_{b\ell}^2 + u_{a\ell}^2 u_{bi}^2 + 2u_{ai}u_{b\ell}u_{a\ell}u_{bi}) \\
&= \sum_{i \geqslant \ell}^{j} (u_{ai}^2 u_{b\ell}^2 + u_{a\ell}^2 u_{bi}^2 + 2u_{ai}u_{b\ell}u_{a\ell}u_{bi}) \\
&= 1/2 \sum_{i=1}^{j} \sum_{\ell=1}^{j} (u_{ai}^2 u_{b\ell}^2 + u_{a\ell}^2 u_{bi}^2 + 2u_{ai}u_{b\ell}u_{a\ell}u_{bi}) + 2 \sum_{i=1}^{j} u_{ai}^2 u_{bi}^2 \\
&= \sum_{i=1}^{j} \sum_{\ell=1}^{j} (u_{ai}^2 u_{b\ell}^2 + u_{ai}u_{b\ell}u_{a\ell}u_{bi}) + O(n^{-3}) \\
&= (\sum_{i=1}^{j} u_{ai}u_{bi})^2 + \sum_{i=1}^{j} u_{ai}^2 \sum_{i=1}^{j} u_{bi}^2 + O(n^{-3}).
\end{aligned}
$$

Next, it can be verified that

$$
\mathcal{P}_j = \mathcal{P}_0 + \frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_j^2 + O(n^{-1}).
$$

To see this,

$$
\begin{aligned}
\frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_i^2 &\leqslant \frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_i \mathcal{P}_{i+1} = \sum_{i=1}^{j} (\mathcal{P}_{i+1} - \mathcal{P}_i) = -\mathcal{P}_0 + \mathcal{P}_j + O(n^{-1}), \\
\frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_i^2 &\geqslant \frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_{i-1} \mathcal{P}_i = \sum_{i=1}^{j} (\mathcal{P}_i - \mathcal{P}_{i-1}) = -\mathcal{P}_0 + \mathcal{P}_j.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\sum_{i=1}^{3} \sum_{i'=1}^{3} \sum_{j=2}^{N} \mathbb{E}_{j-1} \mathcal{H}_{[i]}(\mathbb{z}_1) \mathcal{H}_{[i']}(\mathbb{z}_2) \\
=& N^2 \sum_{j=1}^{N} \mathcal{P}_{j-1} \mathcal{O}_j + N \sum_{j=1}^{N} \mathcal{P}_j^2 \sum_{\ell=1}^{j} \mathcal{O}_\ell + o_p(1) \\
=& N \sum_{j=1}^{N} \mathcal{O}_j (\sum_{\ell=1}^{j} \mathcal{P}_\ell^2 + N\mathcal{P}_0) + N \sum_{j=1}^{N} \mathcal{P}_j^2 \sum_{\ell=1}^{j} \mathcal{O}_\ell + o_p(1) \\
=& N^2 \mathcal{P}_0 \sum_{j=1}^{N} \mathcal{O}_j + N \sum_{\ell \leqslant j}^{N} \mathcal{O}_j \mathcal{P}_\ell^2 + N \sum_{\ell \geqslant j}^{N} \mathcal{O}_j \mathcal{P}_\ell^2 + o_p(1) \\
=& N^2 \mathcal{P}_0 \sum_{j=1}^{N} \mathcal{O}_j + N \sum_{\ell=1}^{N} \sum_{j=1}^{N} \mathcal{O}_j \mathcal{P}_\ell^2 + o_p(1)
\end{aligned}
$$

$$
\begin{aligned}
&= N^2 \sum_{j=1}^{N} \mathcal{O}_j [\frac{1}{N} \sum_{j=1}^{N} \mathcal{P}_\ell^2 + \mathcal{P}_0] + o_p(1) \\
&= N^2 \mathcal{P}_N \sum_{j=1}^{N} \mathcal{O}_j + o_p(1) \\
&= N^2 \frac{\mathcal{D}}{1-\mathcal{D}} \Big[(\sum_{i=1}^{N} u_{ai} u_{bi})^2 + \sum_{i=1}^{N} u_{ai}^2 \sum_{i=1}^{N} u_{bi}^2 \Big] + o_p(1) \\
&= \frac{\mathcal{D}}{1-\mathcal{D}} [\|a\|^2 \|b\|^2 + (a^T b)^2] + o_p(1).
\end{aligned}
$$

Next, we express $\mathcal{D}/(1-\mathcal{D})$ in terms of $\delta(\mathbb{z}_1, \mathbb{z}_2, \gamma)$ and $\Theta(\mathbb{z}, \gamma)$. Using the results in Section S.2, we have

$$
\mathbb{z} \underline{m}_p^0(\mathbb{z}) = -\Theta(\mathbb{z}, \gamma) + o(1).
$$

It follows

$$
\frac{\mathcal{D}}{1-\mathcal{D}} = \delta(\mathbb{z}_1, \mathbb{z}_2, \gamma) \Theta^{-1}(\mathbb{z}_1, \gamma) \Theta^{-1}(\mathbb{z}_2, \gamma) + o(1).
$$

Thus,

$$
\begin{aligned}
&\sum_{k=1}^{3} \sum_{k'=1}^{3} \sum_{j=2}^{N} \beta^{\mathbb{E}}(\mathbb{z}_1) \beta^{\mathbb{E}}(\mathbb{z}_2) \mathbb{E}_{j-1} \mathcal{H}_{[k]}(\mathbb{z}_1, j) \mathcal{H}_{[k']}(\mathbb{z}_2, j) \xrightarrow{P} \\
&\qquad \Theta^{-2}(\mathbb{z}_1, \gamma) \Theta^{-2}(\mathbb{z}_2, \gamma) \delta(\mathbb{z}_1, \mathbb{z}_2, \gamma) [\|a\|^2 \|b\|^2 + (a^T b)^2].
\end{aligned}
$$

The proof of finite dimensional convergence of $G_n^{(1)}(\mathbb{z})$ is complete.

S.3.2. Tightness of $G_n^{(1)}(\mathbb{z}, a, b)$

In view of our smoothing strategy, to show the tightness, we first consider the case $\mathbb{z} = u + iv \in \mathcal{C}^+$, that is $|v| \geqslant \rho_n$.

Recall the definition

$$
G_n(\mathbb{z}, a, b) = \begin{cases} \boldsymbol{u}_a^T \mathbf{Z}^T \mathbf{A}(\mathbb{z}) \mathbf{Z} \boldsymbol{u}_b, & |v| \geqslant \rho_n \\ \frac{\rho_n - v}{2\rho_n} \boldsymbol{u}_a^T \mathbf{Z}^T \mathbf{A}(u + i\rho_n) \mathbf{Z} \boldsymbol{u}_b & \\ \quad + \frac{v + \rho_n}{2\rho_n} \boldsymbol{u}_a^T \mathbf{Z}^T \mathbf{A}(u - i\rho_n) \mathbf{Z} \boldsymbol{u}_b, & |v| < \rho_n. \end{cases}
$$

where $\mathbb{z} = u + iv$.

$$
G_n^{(1)}(\mathbb{z}, a, b) = n^{1/2} [G_n(\mathbb{z}, a, b) - \mathbb{E} G_n(\mathbb{z}, a, b)].
$$

We shall drop the arguments a and b in the rest of the section, since they are fixed.

We first show that $\mathbb{E}\|\mathbf{A}(\mathbb{z})\|^\ell$ is bounded on $\mathcal{C}^+$ for all $\ell \geqslant 1$. As in Lemma A.1, select an arbitrary constant $\mathfrak{D} \in (\limsup_{p\to\infty} \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2, \overline{u})$, and denote $\mathbb{G}$ to be the event $\{\lambda_{\max}(\widetilde{\mathbf{\Sigma}}_p) \geqslant \mathfrak{D}\}$. Lemma A.1 says, for any positive ℓ,

$$\mathbb{P}(\mathbb{G}) = o(n^{-\ell}).$$

Note, on $\mathbb{G}^c$, all eigenvalues of $\widetilde{\mathbf{\Sigma}}_p$ are bounded away from $\mathcal{C}^+$ with distance at least $\min\{\overline{u} - \mathfrak{D}, |\underline{u}|\}$. Thus, $\|(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}\|_2 < \mathcal{K}$ for some $\mathcal{K} > 0$. On $\mathbb{G}$, Lemma S.10 always holds. Therefore, there exists an universal constant $\mathcal{K}$ such that

$$\|\mathbf{A}(\mathbb{z})\| \leqslant \mathcal{K}[1 + |v|^{-1}\mathbb{1}(\mathbb{G})]\|\Sigma_p\| \leqslant \mathcal{K}[1 + \rho_n^{-1}\mathbb{1}(\mathbb{G})]\|\Sigma_p\|. \tag{S.3.20}$$

Thus, for any ℓ, there exists a constant $\mathcal{K}_\ell(\mathbf{A})$ such that

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \mathbb{E}\|\mathbf{A}(\mathbb{z})\|^\ell \leqslant \mathcal{K}_\ell(\mathbf{A}). \tag{S.3.21}$$

Similarly, we can show there exists constants $\mathcal{K}_\ell(\beta)$ and $\mathcal{K}(\beta^{\mathbb{E}})$ such that

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \mathbb{E}|\beta_j(\mathbb{z})|^\ell \leqslant \mathcal{K}_\ell(\beta), \tag{S.3.22}$$

$$\sup_{\mathbb{z}\in\mathcal{C}^+} |\beta^{\mathbb{E}}(\mathbb{z})| \leqslant \mathcal{K}(\beta^{\mathbb{E}}). \tag{S.3.23}$$

Another useful result is, on $\mathbb{G}^c$, a nonrandom bound holds as following.

$$|\beta_j(\mathbb{z})| = |1 - n^{-1}\mathbf{z}_j^T\mathbf{A}(\mathbb{z})\mathbf{z}_j| \leqslant 1 + \mathcal{K}\mathfrak{D}, \quad \text{on } \mathbb{G}^c. \tag{S.3.24}$$

(S.3.22), (S.3.23) and (S.3.24) are shown in Section 3 of Bai and Silverstein (2004).

Next, we show tightness of $G_n^{(1)}(\mathbb{z}, a, b)$. We use Theorem 12.3 of Billingsley (1968). The first condition of the theorem can be replaced by the tightness at any point in $[0, 1]$, as pointed out in Bai and Silverstein (2004). Therefore, we only need to show there exists a constant $\mathcal{K}$ such that

$$\mathbb{E}|G_n^{(1)}(\mathbb{z}_0)|^2 \leqslant \mathcal{K},$$

with $\mathbb{z}_0$ a complex number having non-zero imaginary part. In view of the construction and simplification of the martingale difference sequence in Subsection S.3.1, it is sufficient to show

$$\mathbb{E}|\sum_{j=1}^N \beta^{\mathbb{E}}(\mathbb{z}_0)\mathcal{H}_n(\mathbb{z}_0, j)|^2 \leqslant \mathcal{K}.$$

Due to (S.3.23), we only need to show

$$\sum_{j=1}^N \mathbb{E}|n^{1/2}u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_0)\mathbf{Z}_j\boldsymbol{u}_b|^2 \leqslant \mathcal{K},$$

$$\sum_{j=1}^N \mathbb{E}|n^{1/2}\varrho_j(\mathbb{z}_0)|^2 \leqslant \mathcal{K}.$$

We can prove the results using Lemma S.13, Lemma S.15 and Lemma S.16 with the nonrandom bound of $\|\mathbf{A}_j(\mathbb{z}_0)\|$ shown in Lemma S.10.

The second condition of Theorem 12.3 of Billingsley (1968) will be verified if we can show there exists a constant $\mathcal{K}$ such that for all sufficiently large n and $\mathbb{z}_1 \neq \mathbb{z}_2 \in \{u+iv \in \mathcal{C}$ and $|v| \geqslant \rho_n\}$,

$$\mathbb{E}\frac{|G_n^{(1)}(\mathbb{z}_1) - G_n^{(1)}(\mathbb{z}_2)|^2}{|\mathbb{z}_1 - \mathbb{z}_2|^2} \leqslant \mathcal{K}.$$

Define

$$\begin{aligned}
\tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2) &= \Sigma_p^{\frac{T}{2}}\Big[(\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^T\Sigma_p^{\frac{T}{2}} - \mathbb{z}_1 I)(\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^T\Sigma_p^{\frac{T}{2}} - \mathbb{z}_2 I)\Big]^{-1}\Sigma_p^{\frac{1}{2}},\\
\tilde{\mathbf{A}}_{j\boldsymbol{\cdot}}(\mathbb{z}_1, \mathbb{z}_2) &= \Sigma_p^{\frac{T}{2}}\Big[(\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}_j\mathbf{Z}_j^T\Sigma_p^{\frac{T}{2}} - \mathbb{z}_1 I)(\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^T\Sigma_p^{\frac{T}{2}} - \mathbb{z}_2 I)\Big]^{-1}\Sigma_p^{\frac{1}{2}},\\
\tilde{\mathbf{A}}_{\boldsymbol{\cdot} j}(\mathbb{z}_1, \mathbb{z}_2) &= \Sigma_p^{\frac{T}{2}}\Big[(\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}\mathbf{Z}^T\Sigma_p^{\frac{T}{2}} - \mathbb{z}_1 I)(\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}_j\mathbf{Z}_j^T\Sigma_p^{\frac{T}{2}} - \mathbb{z}_2 I)\Big]^{-1}\Sigma_p^{\frac{1}{2}},\\
\tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) &= \Sigma_p^{\frac{T}{2}}\Big[(\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}_j\mathbf{Z}_j^T\Sigma_p^{\frac{T}{2}} - \mathbb{z}_1 I)(\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}_j\mathbf{Z}_j^T\Sigma_p^{\frac{T}{2}} - \mathbb{z}_2 I)\Big]^{-1}\Sigma_p^{\frac{1}{2}}.
\end{aligned}$$

Along very similar lines to the proof of (S.3.21), we can show, for any $\ell \geqslant 1$, there exists a constant $\mathcal{K}_\ell(\tilde{\mathbf{A}})$

$$\begin{aligned}
\sup_{\mathbb{z}_1,\mathbb{z}_2\in\mathcal{C}^+} \mathbb{E}\|\tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2)\|^\ell &\leqslant \mathcal{K}_\ell(\tilde{\mathbf{A}}),\\
\sup_{\mathbb{z}_1,\mathbb{z}_2\in\mathcal{C}^+} \mathbb{E}\|\tilde{\mathbf{A}}_{j\boldsymbol{\cdot}}(\mathbb{z}_1, \mathbb{z}_2)\|^\ell &\leqslant \mathcal{K}_\ell(\tilde{\mathbf{A}}),\\
\sup_{\mathbb{z}_1,\mathbb{z}_2\in\mathcal{C}^+} \mathbb{E}\|\tilde{\mathbf{A}}_{\boldsymbol{\cdot} j}(\mathbb{z}_1, \mathbb{z}_2)\|^\ell &\leqslant \mathcal{K}_\ell(\tilde{\mathbf{A}}),\\
\sup_{\mathbb{z}_1,\mathbb{z}_2\in\mathcal{C}^+} \mathbb{E}\|\tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2)\|^\ell &\leqslant \mathcal{K}_\ell(\tilde{\mathbf{A}}).
\end{aligned} \tag{S.3.25}$$

Using the identity $\mathbf{A}(\mathbb{z}_1) - \mathbf{A}(\mathbb{z}_2) = (\mathbb{z}_1 - \mathbb{z}_2)\tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2)$.

$$\begin{aligned}
&\frac{G_n^{(1)}(\mathbb{z}_1) - G_n^{(1)}(\mathbb{z}_2)}{\mathbb{z}_1 - \mathbb{z}_2}\\
=&n^{1/2}\boldsymbol{u}_a^T\mathbf{Z}^T\tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2)\mathbf{Z}\boldsymbol{u}_b - n^{1/2}\mathbb{E}\boldsymbol{u}_a^T\mathbf{Z}^T\tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2)\mathbf{Z}\boldsymbol{u}_b\\
=&n^{1/2}\sum_{j=1}^N(\mathbb{E}_j - \mathbb{E}_{j-1})[\boldsymbol{u}_a^T\mathbf{Z}^T\tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2)\mathbf{Z}\boldsymbol{u}_b - \boldsymbol{u}_a^T\mathbf{Z}_j^T\tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_b]\\
=&n^{1/2}\sum_{j=1}^N(\mathbb{E}_j - \mathbb{E}_{j-1})[d_6 + d_7 + d_8 + d_9],
\end{aligned}$$

where

$$d_6 = \boldsymbol{u}_a^T\mathbf{Z}^T\tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2)\mathbf{Z}\boldsymbol{u}_b - \boldsymbol{u}_a^T\mathbf{Z}_j^T\tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2)\mathbf{Z}\boldsymbol{u}_b$$

$$d_7 = \boldsymbol{u}_a^T \mathbf{Z}_j^T \tilde{\mathbf{A}}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z} \boldsymbol{u}_b - \boldsymbol{u}_a^T \mathbf{Z}_j^T \tilde{\mathbf{A}}_{j\boldsymbol{\cdot}}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z} \boldsymbol{u}_b$$
$$d_8 = \boldsymbol{u}_a^T \mathbf{Z}_j^T \tilde{\mathbf{A}}_{j\boldsymbol{\cdot}}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z} \boldsymbol{u}_b - \boldsymbol{u}_a^T \mathbf{Z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z} \boldsymbol{u}_b$$
$$d_9 = \boldsymbol{u}_a^T \mathbf{Z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z} \boldsymbol{u}_b - \boldsymbol{u}_a^T \mathbf{Z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_b$$

Regarding d_6,

$$\begin{aligned} d_6 &= u_{aj} u_{bj} \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{z}_j \beta_j(\mathbb{z}_1) \beta_j(\mathbb{z}_2) \\ &\quad + u_{aj} \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_b \beta_j(\mathbb{z}_1) \\ &\quad - \frac{1}{n} u_{aj} \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{z}_j \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_b \beta_j(\mathbb{z}_1) \beta_j(\mathbb{z}_2). \end{aligned}$$

Combining Lemma S.13, Lemma S.16 and (S.3.25) we can show

$$\sup_j \sup_{\mathbb{z}_1, \mathbb{z}_2 \in \mathcal{C}^+} n^{-\ell} \mathbb{E} |\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{z}_j|^\ell < \mathcal{K}_\ell, \quad \text{for all } \ell \geqslant 2, \tag{S.3.26}$$
$$\sup_j \sup_{\mathbb{z}_1, \mathbb{z}_2 \in \mathcal{C}^+} n^{-\ell} \mathbb{E} |\mathbf{z}_j^T \mathbf{A}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{z}_j|^\ell < \mathcal{K}_\ell, \quad \text{for all } \ell \geqslant 2, \tag{S.3.27}$$
$$\sup_j \sup_{\mathbb{z}_1, \mathbb{z}_2 \in \mathcal{C}^+} \mathbb{E} |\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_b|^4 < \mathcal{K}, \tag{S.3.28}$$

for some constants $\mathcal{K}_\ell$ and $\mathcal{K}$.

Therefore, together with (S.3.22) and (S.3.24),

$$\sup_{\mathbb{z}_1, \mathbb{z}_2 \in \mathcal{C}^+} \mathbb{E} \Big| n^{1/2} \sum_{j=1}^N (\mathbb{E}_j - \mathbb{E}_{j-1}) d_6 \Big|^2 = O(1).$$

The other terms can be written as

$$\begin{aligned} d_7 &= -\frac{1}{n} \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_b \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_1) \mathbf{Z}_j \boldsymbol{u}_a \beta_j(\mathbb{z}_1) \\ &\quad - \frac{1}{n} u_{bj} \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{z}_j \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_1) \mathbf{Z}_j \boldsymbol{u}_a \beta_j(\mathbb{z}_1) \\ &\quad + \frac{1}{n^2} \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_1) \mathbf{Z}_j \boldsymbol{u}_a \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{z}_j \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_b \beta_j(\mathbb{z}_2) \beta_j(\mathbb{z}_1) \\ &\quad + \frac{1}{n^2} u_{bj} \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_1) \mathbf{Z}_j \boldsymbol{u}_a \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{z}_j \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_2) \mathbf{z}_j \beta_j(\mathbb{z}_2) \beta_j(\mathbb{z}_1) \\ d_8 &= -\frac{1}{n} \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_a \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_b \beta_j(\mathbb{z}_2) \\ &\quad - \frac{1}{n} u_{bj} \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_a \mathbf{z}_j^T \mathbf{A}_j(\mathbb{z}_2) \mathbf{z}_j \beta_j(\mathbb{z}_2) \\ d_9 &= u_{bj} \mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1, \mathbb{z}_2) \mathbf{Z}_j \boldsymbol{u}_a \end{aligned}$$

Next, we only present the proof of

$$\sup_{\mathbb{z}_1, \mathbb{z}_2 \in \mathcal{C}^+} \mathbb{E} \Big| n^{1/2} \sum_{j=1}^N (\mathbb{E}_j - \mathbb{E}_{j-1}) d_8 \Big|^2 = O(1).$$

The proof for d_7 and d_9 are very similar.

Due to (S.3.24), (S.3.26), (S.3.27) and (S.3.28),

$$\begin{aligned}
&\sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathbb{E}|\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z}_2)\mathbb{1}(\mathbb{G}^c)|^2\\
&\leqslant \sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathcal{K}\mathbb{E}|\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_b|^2\\
&\leqslant \sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathcal{K}\Big[\mathbb{E}|\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a|^4\mathbb{E}|\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_b|^4\Big]^{1/2}\\
&= O(1).\\
&\sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathbb{E}|\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z}_2)\mathbb{1}(\mathbb{G})|^2\\
&\leqslant \sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathbb{E}\|\mathbf{z}_j\|^4\|\tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\|^2\|\mathbf{Z}_j\boldsymbol{u}_a\|^2\|\mathbf{Z}_j\boldsymbol{u}_b\|^2\|\mathbf{A}_j(\mathbb{z}_2)\|^2\ \cdot\\
&\qquad (1+\|\mathbf{z}_j\|^2\|\mathbf{A}(\mathbb{z})\|)^2\mathbb{1}(\mathbb{G}) = o(1).
\end{aligned}$$

Combining with Lemma S.8, it follows that

$$\begin{aligned}
&\sup_{\mathbb{z}_1,\mathbb{z}_2\in\mathcal{C}^+} \mathbb{E}\Big|n^{1/2}\sum_{j=1}^N(\mathbb{E}_j-\mathbb{E}_{j-1})\frac{1}{n}\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z}_2)\Big|^2\\
&= O(1).
\end{aligned}$$

For the second term in d_8,

$$\begin{aligned}
&\sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathbb{E}|\frac{1}{n}\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{z}_j\beta_j(\mathbb{z}_2)\mathbb{1}(\mathbb{G}^c)|^2\\
&\leqslant \sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathcal{K}\mathbb{E}|\frac{1}{n}\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{z}_j|^2\\
&\leqslant \sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathcal{K}\Big[\mathbb{E}|\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a|^4\mathbb{E}|\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{z}_j|^4\Big]^{1/2}\\
&= O(1).\\
&\sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathbb{E}|\frac{1}{n}\mathbf{z}_j^T \tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\mathbf{Z}_j\boldsymbol{u}_a\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z}_2)\mathbf{z}_j\beta_j(\mathbb{z}_2)\mathbb{1}(\mathbb{G})|^2\\
&\leqslant \sup_j \sup_{\mathbb{z}_1,\mathbb{z}_2} \mathbb{E}n^{-2}\|\mathbf{z}_j\|^6\|\tilde{\mathbf{A}}_{jj}(\mathbb{z}_1,\mathbb{z}_2)\|^2\|\mathbf{Z}_j\boldsymbol{u}_a\|^2\|\mathbf{A}_j(\mathbb{z}_2)\|^2\ \cdot\\
&\qquad (1+\|\mathbf{z}_j\|^2\|\mathbf{A}(\mathbb{z})\|)^2\mathbb{1}(\mathbb{G}) = o(1).
\end{aligned}$$

When $|v| < \rho_n$, $G_n^{(1)}(\mathbb{z})$ is the connected line between $G_n^{(1)}(u+i\rho_n)$ and $G_n^{(1)}(u-i\rho_n)$. Set $\mathbb{z}_1 = u+i\rho_n$ and $\mathbb{z}_2 = u-i\rho_n$. All previous arguments in the subsection apply. Therefore, the slope of the connected line is bounded in expectation.

The proof of tightness of $G_n^{(1)}(\mathbb{z})$ is complete.

S.3.3. Convergence of $G_n^{(2)}$

Recall

$$
\begin{aligned}
&G_n^{(2)}(\mathbb{z}, a, b) = n^{1/2}\{\mathbb{E}G_n(\mathbb{z}, a, b) - a^T b\frac{\Theta_n(\mathbb{z},\gamma) - 1}{\Theta_n(\mathbb{z},\gamma)}\},\\
&\Theta_n(\mathbb{z}) = 1 + \gamma_n\frac{1}{p}\mathbb{E}\mathrm{tr}\mathbf{A}(\mathbb{z}).
\end{aligned}
$$

In this section, we show

$$
\sup_{\mathbb{z}\in\mathcal{C}^+} G_n^{(2)}(\mathbb{z}, a, b) \to 0.
$$

$$
\begin{aligned}
&n^{1/2}\mathbb{E}G_n(\mathbb{z}, a, b)\\
&= n^{1/2}\mathbb{E}\boldsymbol{u}_a^T\mathbf{Z}^T\mathbf{A}(\mathbb{z})\mathbf{Z}\boldsymbol{u}_b\\
&= n^{1/2}\sum_{j=1}^N \mathbb{E}u_{aj}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j\beta_j(\mathbb{z}) + n^{1/2}\sum_{j=1}^N \mathbb{E}u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z})\\
&= n^{1/2}\sum_{j=1}^N \mathbb{E}u_{aj}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j\beta^{\mathbb{E}}(\mathbb{z})\\
&\quad - n^{1/2}\sum_{j=1}^N \mathbb{E}u_{aj}u_{bj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j\beta_j(\mathbb{z})\beta^{\mathbb{E}}(\mathbb{z})\{\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(\mathbb{z})\}\\
&\quad - n^{1/2}\sum_{j=1}^N \mathbb{E}u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z})\beta^{\mathbb{E}}(\mathbb{z})\{\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(\mathbb{z})\}\\
&= n^{1/2}a^T b\frac{\gamma_n p^{-1}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})}{1 + \gamma_n p^{-1}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})}\\
&\quad + n^{1/2}a^T b\beta^{\mathbb{E}}(\mathbb{z})\mathbb{E}\beta_1(\mathbb{z})\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1 - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}\\
&\quad - n^{1/2}\sum_{j=1}^N \mathbb{E}u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b(\beta^{\mathbb{E}}(\mathbb{z}))^2\{\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(\mathbb{z})\}\\
&\quad + n^{1/2}\sum_{j=1}^N \mathbb{E}u_{aj}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z})(\beta^{\mathbb{E}}(\mathbb{z}))^2\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1 - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}^2\\
&= d_{10} + d_{11} + d_{12} + d_{13},\ \text{say.}
\end{aligned}
$$

First, due to (S.3.21), (S.3.22) and Lemma S.13,

$$
\sup_{\mathbb{z}\in\mathcal{C}^+}|\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z}) - \mathbb{E}\mathrm{tr}\mathbf{A}(\mathbb{z})| = \sup_{\mathbb{z}\in\mathcal{C}^+}|\mathbb{E}\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1^2(\mathbb{z})\mathbf{z}_1\beta_1(\mathbb{z})| = O(1).
$$

Together with (S.3.23), we get

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\Big(d_{10}-n^{1/2}a^Tb\frac{\Theta_n(\mathbb{z},\gamma)-1}{\Theta_n(\mathbb{z},\gamma)}\Big)\to 0.$$

Next we want to show $\sup_{\mathbb{z}\in\mathcal{C}^+}\Big(|d_{11}|+|d_{13}|\Big)\to 0$.

$$\begin{aligned}
d_{11}&=n^{1/2}a^Tb(\beta^{\mathbb{E}}(\mathbb{z}))^2\mathbb{E}\beta_1(\mathbb{z})\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}^2,\\
|d_{13}|&\leqslant n^{1/2}\|\boldsymbol{u}_a\|_{\max}\sum_{j=1}^N|\beta^{\mathbb{E}}(\mathbb{z})|^2\mathbb{E}\Big|\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z})\Big|\cdot\\
&\quad\Big|\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}\Big|^2.
\end{aligned}$$

Due to (S.3.21), (S.3.23), (S.3.28) and Lemma S.13,

$$\begin{aligned}
&\sup_{\mathbb{z}\in\mathcal{C}^+}\mathbb{E}|\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})|^4=o(n^{-1}),\\
&\sup_{\mathbb{z}\in\mathcal{C}^+}\mathbb{E}|\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\beta_j(\mathbb{z})|^2=O(1).
\end{aligned}\tag{S.3.29}$$

Thus,

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\Big(|d_{11}|+|d_{13}|\Big)\longrightarrow 0.$$

We next want to show

$$\sup_{\mathbb{z}\in\mathcal{C}^+}|d_{12}|\longrightarrow 0.$$

By Lemma S.20,

$$\begin{aligned}
&n^{1/2}\mathbb{E}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\{\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(\mathbb{z})\}\\
&=n^{-1/2}\mathbb{E}z_{11}^3\mathbb{E}\sum_{i=1}^p e_i^T\mathbf{A}_j(\mathbb{z})e_ie_i^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b.
\end{aligned}$$

We first show

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\sup_{1\leqslant i,m\leqslant p}\mathbb{E}\Big|e_i^T\mathbf{A}_j(\mathbb{z})e_m-\mathbb{E}e_i^T\mathbf{A}_j(\mathbb{z})e_m\Big|^2=O(n^{-1}),\tag{S.3.30}$$

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\sup_{1\leqslant i\leqslant m,1\leqslant j\leqslant N}\mathbb{E}\Big|e_i^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b-\mathbb{E}e_i^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b\Big|^2=O(n^{-1}).\tag{S.3.31}$$

For (S.3.30),

$$\mathbb{E}\Big|e_i^T\mathbf{A}_j(\mathbb{z})e_m-\mathbb{E}e_i^T\mathbf{A}_j(\mathbb{z})e_m\Big|^2$$

$$
\begin{aligned}
&= \mathbb{E}\Big|\sum_{\ell\neq j}^{N}(\mathbb{E}_\ell - \mathbb{E}_{\ell-1})[e_i^T\mathbf{A}_j(\mathbb{z})e_m - \mathbb{E}e_i^T\mathbf{A}_{j\ell}(\mathbb{z})e_m]\Big|^2 \\
&= \mathbb{E}\Big|\sum_{\ell\neq j}^{N}(\mathbb{E}_\ell - \mathbb{E}_{\ell-1})[\frac{1}{n}\mathbf{z}_\ell^T\mathbf{A}_{j\ell}(\mathbb{z})e_m e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell\beta_{\ell j}(\mathbb{z})]\Big|^2 \\
&\leqslant \mathcal{K}\frac{1}{n^2}\sum_{\ell\neq j}^{N}\mathbb{E}\Big|\mathbf{z}_\ell^T\mathbf{A}_{j\ell}(\mathbb{z})e_m e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell\beta_{\ell j}(\mathbb{z})\Big|^2.
\end{aligned}
$$

By Lemma S.14, there exists a $\mathcal{K}$,

$$
\mathbb{E}|\mathbf{z}_\ell^T\mathbf{A}_{j\ell}(\mathbb{z})e_m e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell|^2 \leqslant \mathcal{K}\mathbb{E}\|\mathbf{A}_{j\ell}(\mathbb{z})\|^4,
$$

Recall the definition of $\mathbb{G}^c$ in Subsection S.3.2. Using (S.3.21) and (S.3.24), on $\mathbb{G}^c$,

$$
\mathbb{E}\Big|\mathbf{z}_\ell^T\mathbf{A}_{j\ell}(\mathbb{z})e_m e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell\beta_{\ell j}(\mathbb{z})\Big|^2 \leqslant \mathbb{E}\mathcal{K}\|\mathbf{A}_{j\ell}(\mathbb{z})\|^4 = O(1). \tag{S.3.32}
$$

On $\mathbb{G}$,

$$
\begin{aligned}
&\mathbb{E}\Big|\mathbf{z}_\ell^T\mathbf{A}_{j\ell}(\mathbb{z})e_m e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell\beta_{\ell j}(\mathbb{z})\Big|^2 \\
&\qquad \leqslant \mathbb{E}\|\mathbf{z}_\ell\|^4\|\mathbf{A}_{j\ell}(\mathbb{z})\|^4(1+\|\mathbf{z}_\ell\|^2\|\mathbf{A}_j(\mathbb{z})\|^2)^2\mathbb{1}(\mathbb{G}) \to 0.
\end{aligned} \tag{S.3.33}
$$

It completes the proof of (S.3.30).

To show (S.3.31),

$$
\begin{aligned}
&\mathbb{E}|e_i^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b - \mathbb{E}e_i^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b|^2 \\
&= \mathbb{E}|\sum_{\ell\neq j}^{N}(\mathbb{E}_\ell - \mathbb{E}_{\ell-1})[e_i^T\mathbf{A}_j(\mathbb{z})\mathbf{Z}_j\boldsymbol{u}_b - e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{Z}_{j\ell}\boldsymbol{u}_b]|^2 \\
&\leqslant \mathcal{K}\sum_{\ell\neq j}^{N}\mathbb{E}\Big|u_{b\ell}e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell - \frac{1}{n}e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell\mathbf{z}_\ell^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{Z}_{j\ell}\boldsymbol{u}_b\beta_{\ell j}(\mathbb{z}) \\
&\qquad + u_{b\ell}(\beta_{\ell j}(\mathbb{z})-1)e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell\Big|^2.
\end{aligned}
$$

By Lemma S.14, (S.3.21),

$$
\sup_{\mathbb{z}\in\mathcal{C}^+}\sup_i\mathbb{E}|e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell|^4 \leqslant \mathcal{K}\sup_{\mathbb{z}\in\mathcal{C}^+}\mathbb{E}\|\mathbf{A}_{j\ell}(\mathbb{z})\|^4 = O(1).
$$

Together with Lemma S.14 and (S.3.22),

$$
\mathbb{E}|e_i^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{z}_\ell\mathbf{z}_\ell^T\mathbf{A}_{j\ell}(\mathbb{z})\mathbf{Z}_{j\ell}\boldsymbol{u}_b\beta_{\ell j}(\mathbb{z})|^2 = O(1).
$$

The proof of (S.3.31) is complete.

Following from (S.3.30) and (S.3.11),

$$\begin{aligned}
d_{12} &= -n^{1/2}(\beta^{\mathbb{E}}(z))^2\sum_{j=1}^{N}u_{aj}\mathbb{E}\mathbf{z}_j^T\mathbf{A}_j(z)\mathbf{Z}_j\boldsymbol{u}_b\{\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(z)\mathbf{z}_j-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(z)\}\\
&= -n^{-1/2}(\beta^{\mathbb{E}}(z))^2\sum_{j=1}^{N}u_{aj}\mathbb{E}z_{11}^3\mathbb{E}\sum_{i=1}^{p}e_i^T\mathbf{A}_j(z)e_ie_i^T\mathbf{A}_j(z)\mathbf{Z}_j\boldsymbol{u}_b\\
&= -n^{-1/2}(\beta^{\mathbb{E}}(z))^2\sum_{j=1}^{N}u_{aj}\mathbb{E}z_{11}^3\sum_{i=1}^{p}\mathbb{E}e_i^T\mathbf{A}_j(z)e_i\mathbb{E}e_i^T\mathbf{A}_j(z)\mathbf{Z}_j\boldsymbol{u}_b+o(1)\\
&= -n^{-1/2}(\beta^{\mathbb{E}}(z))^2\mathbb{E}z_{11}^3\sum_{j=1}^{N}u_{aj}\sum_{\ell\neq j}^{N}u_{b\ell}\sum_{i=1}^{p}\mathbb{E}e_i^T\mathbf{A}_1(z)e_i\mathbb{E}e_i^T\mathbf{A}_1(z)\mathbf{z}_2+o(1).
\end{aligned}$$

The residual term $o(1)$ is uniform on $\mathcal{C}^+$. Note

$$\begin{aligned}
&\mathbb{E}e_i^T\mathbf{A}_1(z)\mathbf{z}_2\\
&= \mathbb{E}e_i^T\mathbf{A}_{12}(z)\mathbf{z}_2\beta_{21}(z)\\
&= -\mathbb{E}e_i^T\mathbf{A}_{12}(z)\mathbf{z}_2\beta_{21}(z)\beta^{\mathbb{E}}(z)(\frac{1}{n}\mathbf{z}_2^T\mathbf{A}_{12}(z)\mathbf{z}_2-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(z))\\
&= -\mathbb{E}e_i^T\mathbf{A}_{12}(z)\mathbf{z}_2(\beta^{\mathbb{E}}(z))^2(\frac{1}{n}\mathbf{z}_2^T\mathbf{A}_{12}(z)\mathbf{z}_2-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(z))\\
&\quad+\mathbb{E}e_i^T\mathbf{A}_{12}(z)\mathbf{z}_2(\beta^{\mathbb{E}}(z))^2\beta_{21}(z)\{\frac{1}{n}\mathbf{z}_2^T\mathbf{A}_{12}(z)\mathbf{z}_2-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(z)\}^2\\
&= -\frac{1}{n}(\beta^{\mathbb{E}}(z))^2\mathbb{E}z_{11}^3\sum_{\ell=1}^{p}\mathbb{E}e_\ell^T\mathbf{A}_{12}(z)e_\ell e_\ell^T\mathbf{A}_{12}(z)e_i\\
&\quad+\mathbb{E}e_i^T\mathbf{A}_{12}(z)\mathbf{z}_2(\beta^{\mathbb{E}}(z))^2\beta_{21}(z)\{\frac{1}{n}\mathbf{z}_2^T\mathbf{A}_{12}(z)\mathbf{z}_2-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(z)\}^2.
\end{aligned}$$

By Lemma S.13, Lemma S.14, (S.3.21), (S.10),

$$\sup_{z\in\mathcal{C}^+;i}\mathbb{E}|e_i^T\mathbf{A}_{12}(z)\mathbf{z}_2(\beta^{\mathbb{E}}(z))^2\beta_{21}(z)\{\frac{1}{n}\mathbf{z}_2^T\mathbf{A}_{12}(z)\mathbf{z}_2-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(z)\}^2|=o(n^{-1/2}).$$

Therefore, again using (S.3.30),

$$\begin{aligned}
d_{12} &= n^{-3/2}(\beta^{\mathbb{E}}(z))^4(\mathbb{E}z_{11}^3)^2\sum_{j=1}^{N}u_{aj}\sum_{\ell\neq j}^{N}u_{b\ell}\sum_{i=1}^{p}\mathbb{E}e_i^T\mathbf{A}_j(z)e_i\cdot\\
&\quad\sum_{\ell=1}^{p}\mathbb{E}e_\ell^T\mathbf{A}_{12}(z)e_\ell e_\ell^T\mathbf{A}_{12}(z)e_i+o(1)\\
&= n^{-3/2}(\beta^{\mathbb{E}}(z))^4(\mathbb{E}z_{11}^3)^2\sum_{j=1}^{N}u_{aj}\sum_{\ell\neq j}^{N}u_{b\ell}\sum_{i=1}^{p}\sum_{\ell=1}^{p}\mathbb{E}e_i^T\mathbf{A}_j(z)e_i\mathbb{E}e_\ell^T\mathbf{A}_{12}(z)e_\ell\ \cdot
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}e_\ell^T\mathbf{A}_{12}(\mathbb{z})e_i + o(1)\\
&= n^{-5/2}(\beta^{\mathbb{E}}(\mathbb{z}))^4(\mathbb{E}z_{11}^3)^2 a^T U_n^T \mathbf{1}_n\mathbf{1}_n^T U_n b \ \cdot\\
&\quad \sum_{i=1}^p\sum_{\ell=1}^p \mathbb{E}e_i^T\mathbf{A}_1(\mathbb{z})e_i\mathbb{E}e_\ell^T\mathbf{A}_{12}(\mathbb{z})e_\ell\mathbb{E}e_\ell^T\mathbf{A}_{12}(\mathbb{z})e_i.
\end{aligned}$$

Note $a^T U_n^T\mathbf{1}_n = O(n^{1/2})$ because $\|U\|_{\max} = O(n^{-1/2})$. Meanwhile,

$$\sum_{i=1}^p\sum_{\ell=1}^p \mathbb{E}e_i^T\mathbf{A}_1(\mathbb{z})e_i\mathbb{E}e_\ell^T\mathbf{A}_{12}(\mathbb{z})e_\ell\mathbb{E}e_\ell^T\mathbf{A}_{12}(\mathbb{z})e_i = \alpha_1^T\mathbb{E}\mathbf{A}_{12}\alpha_2,$$

where α_1 is the diagonal of $\mathbb{E}\mathbf{A}_1(\mathbb{z})$ and α_2 is the diagonal of $\mathbb{E}\mathbf{A}_{12}(\mathbb{z})$.

$$\begin{aligned}
\sup_{\mathbb{z}\in\mathcal{C}^+}|\alpha_1^T\mathbb{E}\mathbf{A}_{12}\alpha_2| &\leqslant \sup_{\mathbb{z}\in\mathcal{C}^+}\|\mathbb{E}\mathbf{A}_{12}(\mathbb{z})\|_2\|\alpha_1\|_2\|\alpha_2\|_2\\
&\leqslant \sup_{\mathbb{z}\in\mathcal{C}^+} n\|\mathbb{E}\mathbf{A}_{12}(\mathbb{z})\|_2^2\|\mathbb{E}\mathbf{A}_1(\mathbb{z})\|_2\|\mathbb{E}\mathbf{A}_{12}(\mathbb{z})\|_2 = O(n).
\end{aligned}$$

It follows that $\sup_{\mathbb{z}\in\mathcal{C}^+}|d_{12}| = O(n^{-1/2})$.

S.3.4. Convergence of $G_n^{(3)}$

We next show

$$\sup_{\mathbb{z}\in\mathcal{C}^+} G_n^{(3)}(\mathbb{z},a,b)\to 0.$$

Using the idea of proof of Lemma 2 of Chen, Li and Zhong (2014), post-multiplying both sides of the identity $(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p) + \mathbb{z}I_p = \widetilde{\mathbf{\Sigma}}_p$ by $(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}$,

$$I_p + \mathbb{z}(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1} = \widetilde{\mathbf{\Sigma}}_p(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1} = \frac{1}{n}\sum_{i=1}^N \Sigma_p^{1/2}\mathbf{z}_i\mathbf{z}_i^T\Sigma_p^{T/2}(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}.$$

Taking trace and expectation on both sides,

$$\begin{aligned}
&\frac{n}{N}[1+\mathbb{z}\frac{1}{p}\mathbb{E}\mathrm{tr}(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}] = \frac{1}{p}\mathbb{E}\mathbf{z}_1^T\mathbf{A}(\mathbb{z})\mathbf{z}_1\\
&= \frac{\gamma_n^{-1}(\Theta_{n-1}(\mathbb{z})-1)}{\Theta_{n-1}(\mathbb{z})} - \frac{1}{p}\mathbb{E}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1\beta^{\mathbb{E}}(\mathbb{z})\beta_1(\mathbb{z})\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1 - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{(\Theta_{n-1}(\mathbb{z})-1)}{\Theta_{n-1}(\mathbb{z})} &= \frac{n}{N}\gamma_n + \frac{n}{N}\mathbb{z}\gamma_n\frac{1}{p}\mathbb{E}\mathrm{tr}(\widetilde{\mathbf{\Sigma}}_p - \mathbb{z}I_p)^{-1}\\
&\quad + \frac{1}{n}\mathbb{E}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1\beta^{\mathbb{E}}(\mathbb{z})\beta_1(\mathbb{z})\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1 - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}.
\end{aligned}$$

Note $(\Theta(\mathbb{z},\gamma)-1)\Theta^{-1}(\mathbb{z},\gamma)=\gamma+\gamma\mathbb{z}m(\mathbb{z})$. We only need to show

$$\sup_{\mathbb{z}\in\mathcal{C}^+} n^{1/2}\Big|\gamma_n+\gamma_n\mathbb{z}\frac{1}{p}\mathbb{E}\mathrm{tr}(\widetilde{\mathbf{\Sigma}}_p-\mathbb{z}I_p)^{-1}\}-(\gamma+\gamma\mathbb{z}m(\mathbb{z}))\Big|\longrightarrow 0, \tag{S.3.34}$$

$$\sup_{\mathbb{z}\in\mathcal{C}^+} n^{-\frac{1}{2}}\Big|\mathbb{E}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1\beta^{\mathbb{E}}(\mathbb{z})\beta_1(\mathbb{z})\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}\Big|\to 0. \tag{S.3.35}$$

(S.3.34) follows from the condition $n^{1/2}|\gamma_n-\gamma|\to 0$, Section 4 of Bai and Silverstein (2004) and Lemma S.1.

As for (S.3.35),

$$\begin{aligned}
&\sup_{\mathbb{z}\in\mathcal{C}^+} n^{-1/2}\Big|\mathbb{E}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1\beta^{\mathbb{E}}(\mathbb{z})\beta_1(\mathbb{z})\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}\Big|\\
&=\sup_{\mathbb{z}\in\mathcal{C}^+} n^{1/2}\Big|\mathbb{E}\beta_1(\mathbb{z})\beta^{\mathbb{E}}(\mathbb{z})\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}\Big|\\
&=\sup_{\mathbb{z}\in\mathcal{C}^+} n^{1/2}\Big|\mathbb{E}\beta_1(\mathbb{z})(\beta^{\mathbb{E}}(\mathbb{z}))^2\{\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_1(\mathbb{z})\}^2\Big|=o(1).
\end{aligned}$$

S.4. Additional technical support of Theorem 2.3

In this section, we show (A.12) and (A.13), two key steps in the proof of Theorem 2.3 under the truncated random variable condition (A.7). Additional arguments that deal with the differnce between **C1** and (A.7) can be found in Section S.6.

Consider (A.12). It is sufficient to show

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\mathbb{E}|a^TU_n^T\mathbf{Z}^T(\widetilde{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}BCb|^2\leqslant\mathcal{K}\|BC\|_2^2,$$

for any vector (k-variate) a and (q-variate) b, such that $\|a\|_2\leqslant 1$ and $\|b\|_2\leqslant 1$.

Recall notation defined in Section S.3. Further, write $\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}_j\mathbf{Z}_j^T\Sigma_p^{\frac{T}{2}}$ as $\widetilde{\mathbf{\Sigma}}_p^{(j)}$ and $\frac{1}{n}\Sigma_p^{\frac{1}{2}}\mathbf{Z}_{jj'}\mathbf{Z}_{jj'}^T\Sigma_p^{\frac{T}{2}}$ as $\widetilde{\mathbf{\Sigma}}_p^{(jj')}$.

$$\begin{aligned}
&a^TU_n^T\mathbf{Z}^T(\widetilde{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}BCb\\
&=\sum_{j=1}^N\sqrt{n}u_{aj}\mathbf{z}_j^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(j)}-\mathbb{z}I)^{-1}BCb\beta_j(\mathbb{z})\\
&=\sum_{j=1}^N\sqrt{n}u_{aj}\mathbf{z}_j^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(j)}-\mathbb{z}I)^{-1}BCb\beta^{\mathbb{E}}(\mathbb{z})-\\
&\quad\sum_{j=1}^N\sqrt{n}u_{aj}\mathbf{z}_j^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(j)}-\mathbb{z}I)^{-1}BCb\beta_j(\mathbb{z})\beta^{\mathbb{E}}(\mathbb{z})(\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(\mathbb{z})\mathbf{z}_j-\frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(\mathbb{z}))\\
&=d_{13}+d_{14},\ \text{say}.
\end{aligned}$$

$$\mathbb{E}|d_{13}|^2 = n\sum_{j=1}^{N} u_{aj}^2|\beta^{\mathbb{E}}(z)|^2\mathbb{E}b^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p^{(1)} - \bar{z}I)^{-1}\Sigma_p(\widetilde{\mathbf{\Sigma}}_p^{(1)} - zI)^{-1}BCb+$$
$$n\sum_{j\neq j'} u_{aj}u_{aj'}|\beta^{\mathbb{E}}(z)|^2\mathbb{E}\mathbf{z}_1^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(1)} - zI)^{-1}BCbb^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p^{(2)} - \bar{z}I)^{-1}\Sigma_p^{\frac{1}{2}}\mathbf{z}_2$$
$$\leqslant \mathcal{K}|\beta^{\mathbb{E}}(z)|^2\|b\|_2^2\|BC\|_2^2\mathbb{E}\|(\widetilde{\mathbf{\Sigma}}_p^{(1)} - zI)^{-1}\|_2^2 + \mathcal{K}\mathbb{E}\frac{1}{n}\Big|\mathbf{z}_1^T\mathbf{A}_{12}(z)\mathbf{z}_2\Big|^2\cdot$$
$$\Big|\mathbf{z}_2^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(12)} - zI)^{-1}BCbb^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p^{(12)} - \bar{z}I)^{-1}\Sigma_p^{\frac{1}{2}}\mathbf{z}_1\beta_{12}(z)\beta_{21}(\bar{z})\Big|.$$

Using Lemma S.13, Lemma S.14, Lemma S.17,

$$\mathbb{E}|\mathbf{z}_1^T\mathbf{A}_{12}(z)\mathbf{z}_2|^4 \leqslant \mathcal{K}n^2\mathbb{E}\|\mathbf{A}_{12}(z)\|^4,$$
$$\mathbb{E}|\mathbf{z}_2^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(12)} - zI)^{-1}BCbb^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p^{(12)} - \bar{z}I)^{-1}\Sigma_p^{\frac{1}{2}}\mathbf{z}_1|^4$$
$$\leqslant \mathcal{K}\|BC\|_2^8\mathbb{E}\|(\widetilde{\mathbf{\Sigma}}_p^{(12)} - \bar{z}I)^{-1}\|_2^4\|(\widetilde{\mathbf{\Sigma}}_p^{(12)} - zI)^{-1}\|_2^4.$$

Together with (S.3.21), (S.3.22), and (S.3.23),

$$\sup_{z\in\mathcal{C}^+}\mathbb{E}|d_{13}|^2 \leqslant \mathcal{K}\|BC\|^2.$$

$$\mathbb{E}|d_{14}|^2 \leqslant \mathcal{K}|\beta^{\mathbb{E}}(z)|^2\mathbb{E}\Big|\mathbf{z}_1^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1}BCb\beta_j(z)\cdot$$
$$\sqrt{n}(\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(z)\mathbf{z}_j - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(z))\Big|^2.$$

Using Lemma S.13,

$$\mathbb{E}\Big|\sqrt{n}(\frac{1}{n}\mathbf{z}_j^T\mathbf{A}_j(z)\mathbf{z}_j - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{A}_j(z))\Big|^2 \leqslant \mathcal{K}\mathbb{E}\|\mathbf{A}_j(z)\|^2,$$
$$\mathbb{E}\Big|\mathbf{z}_1^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1}BCb\Big|^4 \leqslant \mathcal{K}\|BC\|_2^4\mathbb{E}\|(\widetilde{\mathbf{\Sigma}}_p - zI)^{-1}\|_2^4.$$

We have

$$\sup_{z\in\mathcal{C}^+}\mathbb{E}|d_{14}|^2 \leqslant \mathcal{K}\|BC\|^2.$$

(A.12) follows.

Consider (A.13). We first show

$$\sup_{z\in\mathcal{C}^+} n\mathbb{E}|a^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p - zI)^{-1}BCa - \mathbb{E}a^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p - zI)^{-1}BCa|^2 \leqslant \mathcal{K}\|BC\|_2^4,$$

for any vector (q-variate) a such that $\|a\|_2 \leqslant 1$.

We use Lemma S.8 to show the result. It is worth mentioning that when $z \in \mathbb{R}^-$, the result is shown in El Karoui and Kösters (2011). Note

$$\sqrt{n}a^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p - zI)^{-1}BCa - \mathbb{E}\sqrt{n}a^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p - zI)^{-1}BCa$$

$$
\begin{aligned}
&= \sum_{j=1}^{N} \mathbb{E}_j\Big[\sqrt{n}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} BCa\Big] \\
&\quad - \mathbb{E}_{j-1}\Big[\sqrt{n}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} BCa\Big].
\end{aligned}
$$

Note

$$
\begin{aligned}
&\sqrt{n}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} BCa \\
&= \frac{-1}{\sqrt{n}} a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} \Sigma_p^{1/2} \mathbf{z}_j \mathbf{z}_j^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} BCa\beta_j(z).
\end{aligned}
$$

Therefore, using Lemma S.8 (Burkholder's inequality) and Lemma S.14,

$$
\begin{aligned}
&n\mathbb{E}|a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} BCa - \mathbb{E}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} BCa|^2 \\
&\leqslant \mathcal{K}\mathbb{E}\Big|\mathbf{z}_j^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} BCaa^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} \Sigma_p^{1/2} \mathbf{z}_j \beta_j(z)\Big|^2.
\end{aligned}
$$

Observe that $\Sigma_p^{T/2}(\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} BCaa^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1}\Sigma_p^{1/2}$ is of rank one. Due to very similar lines to (S.3.32) and (S.3.33), we can get

$$
\mathbb{E}\Big|\mathbf{z}_j^T \Sigma_p^{T/2} (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} BCaa^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(j)} - zI)^{-1} \Sigma_p^{1/2} \mathbf{z}_j \beta_j(z)\Big|^2 \leqslant \mathcal{K}\|BC\|_2^4.
$$

Next, we show

$$
\sup_{z \in \mathcal{C}^+} n\Big|a^T C^T B^T \Big[\mathbb{E}(\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} - \{\Theta(z,\gamma)\Sigma_p - zI\}^{-1}\Big] BCa\Big|^2 \leqslant \mathcal{K}\|BC\|_2^4.
$$

$$
\begin{aligned}
&\sqrt{n}a^T C^T B^T \Big[\mathbb{E}(\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} - (\Theta(z,\gamma)\Sigma_p - zI)^{-1}\Big] BCa \\
&= \mathbb{E}\sqrt{n}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1}\Big[\Theta(z,\gamma)\Sigma_p - \widetilde{\mathbf{\Sigma}}_p\Big](\Theta(z,\gamma)\Sigma_p - zI)^{-1} BCa \\
&= \mathbb{E}\sqrt{n}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1}\Big[\Theta(z,\gamma)\Sigma_p - \frac{N}{n}\Sigma_p^{\frac{1}{2}} \mathbf{z}_1 \mathbf{z}_1^T \Sigma_p^{\frac{T}{2}}\Big] \cdot \\
&\qquad (\Theta(z,\gamma)\Sigma_p - zI)^{-1} BCa \\
&= -\frac{N}{n}\sqrt{n}\mathbb{E}\mathbf{z}_1^T \Sigma_p^{\frac{T}{2}} (\Theta(z,\gamma)\Sigma_p - zI)^{-1} BCaa^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(1)} - zI)^{-1} \Sigma_p^{\frac{1}{2}} \mathbf{z}_1 \beta_1(z) \\
&\quad + \sqrt{n}\Theta(z,\gamma)\mathbb{E}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} \Sigma_p (\Theta(z,\gamma)\Sigma_p - zI)^{-1} BCa \\
&= \sqrt{n}[\Theta(z,\gamma) - \frac{N}{n}\beta^{\mathbb{E}}(z)]\mathbb{E}a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p - zI)^{-1} \Sigma_p (\Theta(z,\gamma)\Sigma_p - zI)^{-1} BCa \\
&\quad - \frac{1}{\sqrt{n}}\Theta(z,\gamma)(z)\mathbb{E}\Big[a^T C^T B^T (\widetilde{\mathbf{\Sigma}}_p^{(1)} - zI)^{-1} \Sigma_p^{\frac{1}{2}} \mathbf{z}_1 \cdot \\
&\qquad \mathbf{z}_1^T \Sigma_p^{\frac{T}{2}} (\widetilde{\mathbf{\Sigma}}_p^{(1)} - zI)^{-1} \Sigma_p (\Theta(z,\gamma)\Sigma_p - zI)^{-1} BCa\beta_1(z)\Big]
\end{aligned}
$$

$$+\frac{N}{n}\sqrt{n}\mathbb{E}\mathbf{z}_1^T\Sigma_p^{\frac{T}{2}}(\Theta(\mathbb{z},\gamma)\Sigma_p-\mathbb{z}I)^{-1}BCaa^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p^{(1)}-\mathbb{z}I)^{-1}\cdot$$
$$\Sigma_p^{\frac{1}{2}}\mathbf{z}_1\beta_1(\mathbb{z})\beta^{\mathbb{E}}(\mathbb{z})(\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{A}_1(\mathbb{z})).$$

Due to (S.3.19) and (S.3.21), we can get

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\sqrt{n}|\Theta(\mathbb{z},\gamma)-\frac{N}{n}\beta^{\mathbb{E}}(\mathbb{z})|\to 0,$$
$$\sup_{\mathbb{z}\in\mathcal{C}^+}\mathbb{E}|a^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}\Sigma_p(\Theta(\mathbb{z},\gamma)\Sigma_p-\mathbb{z}I)^{-1}BCa|\leqslant\mathcal{K}\|BC\|_2^2.$$

Moreover, by Lemma S.13,

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\mathbb{E}|\mathbf{z}_1^T\Sigma_p^{\frac{T}{2}}(\widetilde{\mathbf{\Sigma}}_p^{(1)}-\mathbb{z}I)^{-1}\Sigma_p(\Theta(\mathbb{z},\gamma)\Sigma_p-\mathbb{z}I)^{-1}BCa\cdot$$
$$a^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p^{(1)}-\mathbb{z}I)^{-1}\Sigma_p^{\frac{1}{2}}\mathbf{z}_1|^2$$
$$\leqslant\mathcal{K}\mathbb{E}|a^TC^TB^T(\widetilde{\mathbf{\Sigma}}_p^{(1)}-\mathbb{z}I)^{-1}\cdot$$
$$\Sigma_p(\widetilde{\mathbf{\Sigma}}_p^{(1)}-\mathbb{z}I)^{-1}\Sigma_p(\Theta(\mathbb{z},\gamma)\Sigma_p-\mathbb{z}I)^{-1}BCa|^2\leqslant\|BC\|_2^4.$$

We next show

$$\mathbb{E}n\Big|[\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{A}_1(\mathbb{z})]\beta_1(\mathbb{z})\Big|^2=O(1).$$

Recall $\mathbb{G}$ is the event $\{\lambda_{\max}(\widetilde{\mathbf{\Sigma}}_p)\geqslant\mathfrak{D}\}$ for some $\mathfrak{D}\in(\limsup_p\lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2,\overline{u})$, defined in Subsection S.3.2. Recall (S.3.24). On $\mathbb{G}^c$, $\beta_1(\mathbb{z})$ is bounded and

$$\mathbb{E}n|\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{A}_1(\mathbb{z})|^2\leqslant\mathcal{K}\mathbb{E}\|\mathbf{A}_1(\mathbb{z})\|_2^2=O(1).$$

On $\mathbb{G}$,

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\mathbb{E}n\Big|[\frac{1}{n}\mathbf{z}_1^T\mathbf{A}_1(\mathbb{z})\mathbf{z}_1-\mathbb{E}\frac{1}{n}\mathrm{tr}\mathbf{A}_1(\mathbb{z})]\beta_1(\mathbb{z})\mathbb{1}(\mathbb{G})\Big|^2\to 0.$$

It completes the proof.

S.5. Proof of Remark 2.1, Lemma 2.1 and Lemma 2.2

S.5.1. Proof of Remark 2.1

The results in Subsection S.3.1.7 imply that $\delta(\mathbb{z}_1,\mathbb{z}_2,\gamma)$ is the limit in L_1-norm of $n^{-1}\mathrm{tr}[\mathbf{A}(\mathbb{z}_1)\mathbf{A}(\mathbb{z}_2)]$ pointwise on $(\mathcal{C}^+)^2$, when the random variables z_{ij}'s satisfy the truncated variable condition (A.7). Due to very similar lines to (S.3.21) and (S.3.25), we have

$$\sup_{\mathbb{z}_1,\mathbb{z}_2\in\mathcal{C}^+}\mathbb{E}\Big|n^{-1}\mathrm{tr}[\mathbf{A}(\mathbb{z}_1)\mathbf{A}(\mathbb{z}_2)]\Big|<\infty.$$

Besides, in Section S.2, it is claimed that $\delta(\mathbb{z}_1, \mathbb{z}_2, \gamma)$ and its partial derivatives are bounded on $\mathcal{C}$.

It follows from *Fubini's Theorem* and *Dominated Convergence Theorem* that

$$\frac{2}{(2\pi i)^2} \oiint_{(\mathcal{C}^+)^2} f(\mathbb{z}_1) f(\mathbb{z}_2) \frac{1}{n} \mathrm{tr}[\mathbf{A}(\mathbb{z}_1)\mathbf{A}(\mathbb{z}_2)] d\mathbb{z}_1 d\mathbb{z}_2 \xrightarrow{P} \Delta(f, \gamma).$$

On the other hand, when $\lambda_{\max}(\widetilde{\boldsymbol{\Sigma}}_p) < \mathfrak{D}$, for some $\mathfrak{D} \in (\limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2, \overline{u})$, $\|\mathbf{A}(\mathbb{z})\|$ is bounded. Therefore,

$$\oiint_{\mathcal{C}^2 \setminus (\mathcal{C}^+)^2} \left| f(\mathbb{z}_1) f(\mathbb{z}_2) \frac{1}{n} \mathrm{tr}[\mathbf{A}(\mathbb{z}_1)\mathbf{A}(\mathbb{z}_2)] \right| |d\mathbb{z}_1||d\mathbb{z}_2| \xrightarrow{P} 0.$$

Providing that $f(x) \geqslant 0$, for all sufficiently large n, with probability 1,

$$\frac{2}{(2\pi i)^2} \oiint_{\mathcal{C}^2} f(\mathbb{z}_1) f(\mathbb{z}_2) \frac{1}{n} \mathrm{tr}[\mathbf{A}(\mathbb{z}_1)\mathbf{A}(\mathbb{z}_2)] d\mathbb{z}_1 d\mathbb{z}_2 = \frac{1}{n} \mathrm{tr}[f(\widehat{\boldsymbol{\Sigma}}_p)\Sigma_p f(\widehat{\boldsymbol{\Sigma}}_p)\Sigma_p] \geqslant$$
$$\{\frac{1}{n}\mathrm{tr}[f(\widehat{\boldsymbol{\Sigma}}_p)\Sigma_p]\}^2 \geqslant \max\left\{\lambda^2_{\min}(\Sigma_p)[\frac{1}{n}\mathrm{tr}(f(\widehat{\boldsymbol{\Sigma}}_p))]^2,\ \lambda^2_{\min}(f(\widehat{\boldsymbol{\Sigma}}_p))[\frac{1}{n}\mathrm{tr}(\Sigma_p)]^2\right\}.$$

It is assumed that F^{Σ_p} converges to L^{Σ} in Wasserstein distance and the latter is non-degenerate at zero. If $f(x) > 0$ on the compact set $\mathcal{X}$, $\inf_{x \in \mathcal{X}} f(x) > 0$. Therefore, with high probability,

$$\lambda^2_{\min}(f(\widehat{\boldsymbol{\Sigma}}_p))[\frac{1}{n}\mathrm{tr}(\Sigma_p)]^2 > \mathcal{K} > 0.$$

It follows that $\Delta(f, \gamma) > 0$.

If $f(x)$ is only nonnegative but $\liminf_p \lambda_{\min}(\Sigma_p) > 0$, we only need to show that $n^{-1}\mathrm{tr}[f(\widehat{\boldsymbol{\Sigma}}_p)] > \mathcal{K} > 0$ with high probability for all sufficiently large n. By Bai and Silverstein (2004, Theorem 1.1), $\int f(\tau) dF^{\widehat{\boldsymbol{\Sigma}}_p}(\tau) \to \int f(\tau) dF^{\infty}(\tau)$ in probability. It follows that $\Delta(f, \gamma) > 0$ if $\int f(\tau) dF^{\infty}(\tau) > 0$.

The proof of Remark 2.1 is complete.

S.5.2. Proof of Lemma 2.1

Lemma 2.1 can be deduced by combining Lemma A.4 and Lemma S.3 shown in Section S.6.

S.5.3. Proof of Lemma 2.2

Observe that $\Theta(\mathbb{z}, \gamma)$ and $\Delta(\mathbb{z}, \gamma)$ are smooth functions of $m(\mathbb{z})$ and $m'(\mathbb{z})$. We further know that $\Theta(\mathbb{z}, \gamma)$ is bounded on $\mathcal{C}$ (see Section S.2). Also, it is easy to check that $\widehat{\Theta}(\mathbb{z}, \gamma_n)$

is bounded on $\mathcal{C}$ when $\lambda_{\max}(\widehat{\mathbf{\Sigma}}_p) < \mathfrak{D} < \overline{u}$. Consequently, it suffices to show the uniform convergence of $m_{n,p}(\mathbb{z})$ and $m'_{n,p}(\mathbb{z})$ on $\mathcal{C}^+$, that is,

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \sqrt{n}|m_{n,p}(\mathbb{z}) - m(\mathbb{z})| \xrightarrow{P} 0, \tag{S.5.1}$$

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \sqrt{n}|m'_{n,p}(\mathbb{z}) - m'(\mathbb{z})| \xrightarrow{P} 0. \tag{S.5.2}$$

Define $\bar{\mathbf{\Sigma}} = \frac{1}{n}\sum_{j=1}^n \Sigma_p^{1/2}\mathbf{z}_j\mathbf{z}_j^T\Sigma_p^{T/2}$ and

$$\tilde{m}(\mathbb{z}) = p^{-1}\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}} - \mathbb{z}I_p)^{-1}\Big].$$

Then, using the rank inequality, it is easy to check that

$$\sup_{\mathbb{z}\in\mathcal{C}^+} |\tilde{m}(\mathbb{z}) - m_{n,p}(\mathbb{z})| = o_p(n^{-1/2}),$$
$$\sup_{\mathbb{z}\in\mathcal{C}^+} \Big|\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z}) - \frac{d}{d\mathbb{z}}m_{n,p}(\mathbb{z})\Big| = o_p(n^{-1/2}).$$

The main reason is that

$$\Big\|F^{\widehat{\mathbf{\Sigma}}_p} - F^{\bar{\mathbf{\Sigma}}}\Big\|_\infty \leqslant \frac{1}{p}\mathrm{rank}(\widehat{\mathbf{\Sigma}}_p - \bar{\mathbf{\Sigma}}) \leqslant \frac{2k}{p}.$$

Together with Lemma S.1 and Lemma S.2, we only need to show

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \sqrt{n}|\tilde{m}(\mathbb{z}) - m_p^0(\mathbb{z})| \xrightarrow{P} 0,$$
$$\sup_{\mathbb{z}\in\mathcal{C}^+} \sqrt{n}\Big|\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z}) - \frac{d}{d\mathbb{z}}m_p^0(\mathbb{z})\Big| \xrightarrow{P} 0.$$

The convergence of $\tilde{m}(\mathbb{z})$ to $\underline{m}_p^0(\mathbb{z})$ is shown in Bai and Silverstein (2004) under (A.7). It indeed holds under **C1** (see Lemma S.5 in Section S.6). The proof of S.5.1 is complete.

As for the uniform convergence of $\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z})$, again, we first consider the truncated variables satisfying (A.7), with generalization to **C1** addressed by Lemma S.5 in Section S.6. We first show

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \sqrt{n}\Big|\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z}) - \mathbb{E}\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z})\Big| \xrightarrow{P} 0.$$

The convergence of finite-dimensional distributions is a direct consequence of Bai and Silverstein (2004). We only need to show its tightness. Using the same strategy as in Subsection S.3.2, it suffices to show that there exists a constant $\mathcal{K}$ such that for any $\mathbb{z}_1 \neq \mathbb{z}_2 \in \mathcal{C}^+$,

$$\mathbb{E}n\frac{|\{\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z}_1) - \mathbb{E}\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z}_1)\} - \{\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z}_2) - \mathbb{E}\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z}_2)\}|^2}{|\mathbb{z}_1 - \mathbb{z}_2|^2}$$

$$\leqslant 2n\mathbb{E}\Big|\frac{1}{p}\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}}-\mathbb{z}_1 I)^{-2}(\bar{\mathbf{\Sigma}}-\mathbb{z}_2 I)^{-1}\Big]-\mathbb{E}\frac{1}{p}\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}}-\mathbb{z}_1 I)^{-2}(\bar{\mathbf{\Sigma}}-\mathbb{z}_2 I)^{-1}\Big]\Big|^2$$
$$+ 2n\mathbb{E}\Big|\frac{1}{p}\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}}-\mathbb{z}_1 I)^{-1}(\bar{\mathbf{\Sigma}}-\mathbb{z}_2 I)^{-2}\Big]-\mathbb{E}\frac{1}{p}\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}}-\mathbb{z}_1 I)^{-1}(\bar{\mathbf{\Sigma}}-\mathbb{z}_2 I)^{-2}\Big]\Big|^2$$
$$\leqslant \mathcal{K}.$$

Define $\bar{\mathbf{\Sigma}}^{(j)}=\frac{1}{n}\sum_{i\neq j}^{n}\Sigma_p^{1/2}\mathbf{z}_i\mathbf{z}_i^T\Sigma_p^{T/2}$. We can write

$$\begin{aligned}&\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}}-\mathbb{z}_1 I)^{-2}(\bar{\mathbf{\Sigma}}-\mathbb{z}_2 I)^{-1}\Big]-\mathbb{E}\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}}-\mathbb{z}_1 I)^{-2}(\bar{\mathbf{\Sigma}}-\mathbb{z}_2 I)^{-1}\Big]\\ &=\sum_{j=1}^{n}(\mathbb{E}_j-\mathbb{E}_{j-1})\Big\{\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}}-\mathbb{z}_1 I)^{-2}(\bar{\mathbf{\Sigma}}-\mathbb{z}_2 I)^{-1}\Big]\\ &\qquad-\mathrm{tr}\Big[(\bar{\mathbf{\Sigma}}^{(j)}-\mathbb{z}_1 I)^{-2}(\bar{\mathbf{\Sigma}}^{(j)}-\mathbb{z}_2 I)^{-1}\Big]\Big\}.\end{aligned}$$

The rest of the proof is very similar to the work presented in Subsection S.3.2, also Section 3 of Bai and Silverstein (2004). We use Lemma S.8 to find the stochastic order of the sum of the martingale difference sequence. Similar to (S.3.21), we have $\sup_{\mathbb{z}\in\mathcal{C}^+}\mathbb{E}\|(\bar{\mathbf{\Sigma}}-\mathbb{z}I)^{-1}\|^{\ell}<\infty$ for any $\ell\in\mathbb{N}^+$. We omit details.

We next show

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\sqrt{n}\Big|\mathbb{E}\frac{d}{d\mathbb{z}}\tilde{m}(\mathbb{z})-\frac{d}{d\mathbb{z}}m_p^0(\mathbb{z})\Big|\to 0. \tag{S.5.3}$$

It suffices to show $\sup_{\mathbb{z}\in\mathcal{C}^+}\sqrt{n}|\mathbb{E}\frac{d}{d\mathbb{z}}\underline{\tilde{m}}(\mathbb{z})-\frac{d}{d\mathbb{z}}\underline{m}_p^0(\mathbb{z})|\to 0$, where $\underline{\tilde{m}}(\mathbb{z})=\frac{\gamma_n-1}{\mathbb{z}}+\gamma_n\tilde{m}(\mathbb{z})$.

Following notation in Section S.3.1 and Section S.4, it is shown in Bai and Silverstein (1998, (5.2)) that

$$\gamma_n\int\frac{dF^{\Sigma_p}(\tau)}{1+\tau\mathbb{E}\underline{\tilde{m}}(\mathbb{z})}+\mathbb{z}\gamma_n\mathbb{E}\tilde{m}(\mathbb{z})=A_n(\mathbb{z}),$$

where

$$\begin{aligned}A_n(\mathbb{z})=&\mathbb{E}\bar{\beta}_1(\mathbb{z})\Big[\frac{1}{n}\mathbf{z}_1^T\Sigma_p^T(\bar{\mathbf{\Sigma}}^{(1)}-\mathbb{z}I)^{-1}(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p+I)^{-1}\Sigma_p^{1/2}\mathbf{z}_1\\ &-\frac{1}{n}\mathbb{E}\mathrm{tr}[(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p+I)^{-1}\Sigma_p(\bar{\mathbf{\Sigma}}-\mathbb{z}I)^{-1}]\Big],\\ \bar{\beta}_1(\mathbb{z})=&\frac{1}{1+n^{-1}\mathbf{z}_1^T\Sigma_p^{T/2}(\bar{\mathbf{\Sigma}}^{(1)}-\mathbb{z}I)^{-1}\Sigma_p^{1/2}\mathbf{z}_1}.\end{aligned}$$

It is shown that $A_n(\mathbb{z})=o(n^{-1/2})$ uniformly on $\mathcal{C}^+$ in Bai and Silverstein (2004) (see (4.19)–(4.11)).

Next, we would like to first take differentiation on both sides with respect to $\mathbb{z}$, and then interchange differentiation and integration. Note that, similar to (S.3.20), (S.3.24), we have on $\mathbb{z}\in\mathcal{C}^+$,

$$\mathbb{E}\|(\bar{\mathbf{\Sigma}}-\mathbb{z}I)^{-2}\|^{\ell}\leqslant\mathbb{E}\mathcal{K}^{2\ell}[1+v^{-1}\mathbb{1}(\mathbb{G})]^{2\ell}<\infty,\ \text{for any } \ell\in\mathbb{N}^+,$$

$$\mathbb{E}|\frac{d}{d\mathbb{z}}\bar{\beta}_1(\mathbb{z})|^2 = \mathbb{E}|\frac{d}{d\mathbb{z}}(1 - n^{-1}\mathbf{z}_1^T\Sigma_p^{T/2}(\bar{\boldsymbol{\Sigma}} - \mathbb{z}I_p)^{-1}\Sigma_p^{1/2}\mathbf{z}_1)|^2$$
$$= \mathbb{E}|n^{-1}\mathbf{z}_1^T\Sigma_p^{T/2}(\bar{\boldsymbol{\Sigma}} - \mathbb{z}I)^{-2}\Sigma_p^{1/2}\mathbf{z}_1|^2 < \infty.$$

Moreover, Bai and Silverstein (2004, (4.3)) claims that

$$\sup_{\mathbb{z}\in\mathcal{C}^+} \|(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p + I)^{-1}\| < \infty.$$

Due to *Dominated Convergence Theorem*,

$$\begin{aligned}
&\frac{d}{d\mathbb{z}}A_n(\mathbb{z})\\
&= \mathbb{E}\frac{d}{d\mathbb{z}}\bar{\beta}_1(\mathbb{z})\Big[\frac{1}{n}\mathbf{z}_1^T\Sigma_p^T(\bar{\boldsymbol{\Sigma}}^{(1)} - \mathbb{z}I)^{-1}(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p + I)^{-1}\Sigma_p^{1/2}\mathbf{z}_1\\
&\qquad - \frac{1}{n}\mathbb{E}\mathrm{tr}\big[(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p + I)^{-1}\Sigma_p(\bar{\boldsymbol{\Sigma}} - \mathbb{z}I)^{-1}\big]\Big]\\
&\quad + \mathbb{E}\bar{\beta}_1(\mathbb{z})\Big[\frac{1}{n}\mathbf{z}_1^T\Sigma_p^T(\bar{\boldsymbol{\Sigma}}^{(1)} - \mathbb{z}I)^{-2}(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p + I)^{-1}\Sigma_p^{1/2}\mathbf{z}_1\\
&\qquad - \frac{1}{n}\mathbb{E}\mathrm{tr}\big[(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p + I)^{-1}\Sigma_p(\bar{\boldsymbol{\Sigma}} - \mathbb{z}I)^{-2}\big]\Big]\\
&\quad + \mathbb{E}\bar{\beta}_1(\mathbb{z})\Big[\frac{1}{n}\mathbf{z}_1^T\Sigma_p^T(\bar{\boldsymbol{\Sigma}}^{(1)} - \mathbb{z}I)^{-1}\Big\{\frac{d}{d\mathbb{z}}(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p + I)^{-1}\Big\}\Sigma_p^{1/2}\mathbf{z}_1\\
&\qquad - \frac{1}{n}\mathbb{E}\mathrm{tr}\Big[\Big\{\frac{d}{d\mathbb{z}}(\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\Sigma_p + I)^{-1}\Big\}\Sigma_p(\bar{\boldsymbol{\Sigma}} - \mathbb{z}I)^{-1}\Big]\Big]\\
&= d_{15} + d_{16} + d_{17},\ \text{say}.
\end{aligned}$$

Follow analogous lines as those in Subsection S.3.2, Subsection S.3.3 and Section S.4, we can show $\sup_{\mathbb{z}\in\mathcal{C}^+}\sqrt{n}[|d_{15}| + |d_{16}| + |d_{17}|] = o(1)$. Details are omitted.

Bai and Silverstein (2004, (4.12)) indicates

$$\begin{aligned}
&\mathbb{E}\underline{\tilde{m}}(\mathbb{z}) - \underline{m}_p^0(\mathbb{z})\\
&= -\underline{m}_p^0(\mathbb{z})A_n(\mathbb{z})\Big[1 - \gamma_n\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\underline{m}_p^0(\mathbb{z})\int\frac{\tau^2 dF^{\Sigma_p}(\tau)}{(1+\tau\mathbb{E}\underline{\tilde{m}}(\mathbb{z}))(1+\tau\underline{m}_p^0(\mathbb{z}))}\Big]^{-1}.
\end{aligned}$$

It is also claimed that the denominator on the right-hand side is bounded away from zero. Take differentiation on both sides,

$$\begin{aligned}
&\mathbb{E}\frac{d}{d\mathbb{z}}\underline{\tilde{m}}(\mathbb{z}) - \frac{d}{d\mathbb{z}}\underline{m}_p^0(\mathbb{z})\\
&= -\Big[\frac{d}{d\mathbb{z}}\underline{m}_p^0(\mathbb{z})\Big]A_n(\mathbb{z})\Big[1 - \gamma_n\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\underline{m}_p^0(\mathbb{z})\int\frac{\tau^2 dF^{\Sigma_p}(\tau)}{(1+\tau\mathbb{E}\underline{\tilde{m}}(\mathbb{z}))(1+\tau\underline{m}_p^0(\mathbb{z}))}\Big]^{-1}\\
&\quad - \underline{m}_p^0(\mathbb{z})\Big[\frac{d}{d\mathbb{z}}A_n(\mathbb{z})\Big]\Big[1 - \gamma_n\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\underline{m}_p^0(\mathbb{z})\int\frac{\tau^2 dF^{\Sigma_p}(\tau)}{(1+\tau\mathbb{E}\underline{\tilde{m}}(\mathbb{z}))(1+\tau\underline{m}_p^0(\mathbb{z}))}\Big]^{-1}
\end{aligned}$$

$$- \underline{m}_p^0(\mathbb{z})A_n(\mathbb{z})\frac{d}{d\mathbb{z}}\Big[1-\gamma_n\mathbb{E}\underline{\tilde{m}}(\mathbb{z})\underline{m}_p^0(\mathbb{z})\int\frac{\tau^2 dF^{\Sigma_p}(\tau)}{(1+\tau\mathbb{E}\underline{\tilde{m}}(\mathbb{z}))(1+\tau\underline{m}_p^0(\mathbb{z}))}\Big]^{-1}.$$

It follows that $\mathbb{E}\underline{\tilde{m}}(\mathbb{z}) - \underline{m}_p^0(\mathbb{z}) = o(n^{-1/2})$ uniformly on $\mathcal{C}^+$.

The proof of Lemma 2.2 is complete.

S.6. Truncation of random variables

In previous sections, we proved the asymptotics of various objects with the variable truncation step. To complete the technical support, we need to verify the truncation step does not change the weak limit of the objects.

Recall

$$z_{ij} = \frac{\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant \varepsilon_n n^{1/2})}{\{\mathbb{E}[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant \varepsilon_n n^{1/2})]^2\}^{1/2}}.$$

We define $\breve{\mathbf{Z}}$ and $\breve{\mathbf{\Sigma}}_p$ by copying the definition of $\mathbf{Z}$ and $\widehat{\mathbf{\Sigma}}_p$, but with $\mathbf{Z}$ replaced by $\breve{\mathbf{Z}}$. Similarly, $\breve{m}_{n,p}(\mathbb{z})$ and $\breve{m}'_{n,p}(\mathbb{z})$ are the counterparts of $m_{n,p}(\mathbb{z})$ and $m'_{n,p}(\mathbb{z})$.

Specifically, to verify that Theorem A.1 still holds without the variable truncation, we need to prove the following lemma.

Lemma S.3 *For any fixed α and η,*

$$\begin{aligned}\sup_{\mathbb{z}\in\mathcal{C}^+} n^{-1/2}\Big|&\alpha^T Q_n^T\mathbf{Z}^T\Sigma_p^{T/2}(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}\Sigma_p^{1/2}\mathbf{Z}Q_n\eta \\ &- \alpha^T Q_n^T\breve{\mathbf{Z}}^T\Sigma_p^{T/2}(\breve{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}\Sigma_p^{1/2}\breve{\mathbf{Z}}Q_n\eta\Big| \xrightarrow{P} 0.\end{aligned}$$

Secondly, to verify that Theorem 2.3 still holds without the variable truncation, we need to show (A.9) and (A.10) still hold without the variable truncation. It suffices to show the following lemma.

Lemma S.4

$$\begin{aligned}\sup_{\mathbb{z}\in\mathcal{C}^+}\Big\|&Q_n^T\breve{\mathbf{Z}}^T\Sigma_p^{T/2}(\breve{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}BCT_n^{-1/2} \\ &- Q_n^T\mathbf{Z}^T\Sigma_p^{T/2}(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}BCT_n^{-1/2}\Big\|_2 \xrightarrow{P} 0, \\ \sup_{\mathbb{z}\in\mathcal{C}^+}\sqrt{n}\Big\|&T_n^{-1/2}C^TB^T(\breve{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}BCT_n^{-1/2} \\ &- T_n^{-1/2}C^TB^T(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}BCT_n^{-1/2}\Big\|_2 \xrightarrow{P} 0.\end{aligned}$$

Lastly, to verify Lemma 2.2 under **C1**, we need to show the following lemma.

Lemma S.5

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\sqrt{n}|m_{n,p}(\mathbb{z})-\breve{m}_{n,p}(\mathbb{z})| \xrightarrow{P} 0,$$

$$\sup_{\mathbb{z}\in\mathcal{C}^+}\sqrt{n}|m'_{n,p}(\mathbb{z})-\breve{m}'_{n,p}(\mathbb{z})| \xrightarrow{P} 0.$$

In the following, we shall only prove Lemma S.3. The other two lemmas can be proved using analogous lines.

By Yin, Bai and Krishnaiah (1988),

$$\lambda_{\max}(\breve{\boldsymbol{\Sigma}}_p) \leqslant \limsup_p \lambda_{\max}(\Sigma_p)\lambda_{\max}(\frac{1}{n}\mathbf{Z}\mathbf{Z}^T) \xrightarrow{a.s.} \limsup_p \lambda_{\max}(\Sigma_p)(1+\sqrt{\gamma})^2.$$

Together with Lemma A.1, we only need to consider the case $\lambda_{\max}(\breve{\boldsymbol{\Sigma}}_p) < \mathfrak{D}$ and $\lambda_{\max}(\widehat{\boldsymbol{\Sigma}}_p) < \mathfrak{D}$ for a constant $\mathfrak{D}$ such that $(1+\sqrt{\gamma})^2 \limsup_{p\to\infty} \lambda_{\max}(\Sigma_p) < \mathfrak{D} < \overline{u}$. It follows, under this event,

$$\sup_{\mathbb{z}\in\mathcal{C}} \|(\widehat{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\|_2 \leqslant (\overline{u}-\mathfrak{D})^{-1} + |\underline{u}|^{-1},$$
$$\sup_{\mathbb{z}\in\mathcal{C}} \|(\breve{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\|_2 \leqslant (\overline{u}-\mathfrak{D})^{-1} + |\underline{u}|^{-1}.$$

$$\begin{aligned}
&n^{-1/2}\alpha^T Q_n^T \mathbf{Z}^T \Sigma_p^{T/2}(\widehat{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\Sigma_p^{1/2}\mathbf{Z}Q_n\eta \\
&\qquad - \alpha^T Q_n^T \breve{\mathbf{Z}}^T \Sigma_p^{T/2}(\breve{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\Sigma_p^{1/2}\breve{\mathbf{Z}}Q_n\eta \\
&= n^{-1/2}\alpha^T Q_n^T (\mathbf{Z}-\breve{\mathbf{Z}})^T \Sigma_p^{T/2}(\widehat{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\Sigma_p^{1/2}\mathbf{Z}Q_n\eta \\
&\quad + n^{-1/2}\alpha^T Q_n^T \breve{\mathbf{Z}}^T \Sigma_p^{T/2}[(\widehat{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1} - (\breve{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}]\Sigma_p^{1/2}\mathbf{Z}Q_n\eta \\
&\quad + n^{-1/2}\alpha^T Q_n^T \breve{\mathbf{Z}}^T \Sigma_p^{T/2}(\breve{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\Sigma_p^{1/2}(\mathbf{Z}-\breve{\mathbf{Z}})Q_n\eta \\
&= d_{18} + d_{19} + d_{20}, \text{ say.}
\end{aligned}$$

Therefore, we can find a constant $\mathcal{K}$ sufficiently large such that

$$|d_{18}| \leqslant n^{-1/2}\mathcal{K}\|\alpha^T Q_n^T(\mathbf{Z}-\breve{\mathbf{Z}})^T\|_2 \|\mathbf{Z}Q_n\eta\|_2.$$

By Lemma S.15, $n^{-1/2}\|\mathbf{Z}Q_n\eta\|_2 = O_p(1)$. We next show

$$\|(\mathbf{Z}-\breve{\mathbf{Z}})Q_n\alpha\|_2 \xrightarrow{P} 0.$$

Since

$$\begin{aligned}
&\mathbb{E}\alpha^T Q_n^T(\mathbf{Z}-\breve{\mathbf{Z}})^T(\mathbf{Z}-\breve{\mathbf{Z}})Q_n\alpha \\
&= \mathbb{E}\sum_{i=1}^N\sum_{j=1}^N [Q_n\alpha]_i[Q_n\alpha]_j \sum_{k=1}^p (z_{ik}-\breve{z}_{ik})(z_{jk}-\breve{z}_{jk}) = \|\alpha\|^2 p\mathbb{E}(z_{ij}-\breve{z}_{ij})^2. \\
&\breve{z}_{ij} - z_{ij} = \frac{\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| > \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| > \varepsilon_n n^{1/2})}{\{\mathbb{E}[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| \leqslant \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| \leqslant \varepsilon_n n^{1/2})]^2\}^{1/2}} \\
&\quad + \breve{z}_{ij}\Big[1 - \Big\{\mathbb{E}\Big[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| \leqslant \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| \leqslant \varepsilon_n n^{1/2})\Big]^2\Big\}^{-1/2}\Big].
\end{aligned}$$

We only need to show

$$\mathbb{E}p[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})-\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})]^2\longrightarrow 0,$$
$$p^{1/2}\Big[1-\Big\{\mathbb{E}\Big[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})-\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})\Big]^2\Big\}^{-1/2}\Big]\longrightarrow 0.$$

Note, ε_n is such that $\varepsilon_n\to 0$ and $\varepsilon_n^{-4}\mathbb{E}\breve{z}_{ij}^4\mathbb{1}(|\breve{z}_{ij}|\geqslant\varepsilon_n n^{1/2})\to 0$. For the first line above, we have

$$\begin{aligned}&\mathbb{E}p[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})-\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})]^2\leqslant 2\mathbb{E}p\breve{z}_{ij}^2\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})\\&\leqslant 2\gamma_n\varepsilon_n^{-2}\mathbb{E}\breve{z}_{ij}^4\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})\longrightarrow 0.\end{aligned}$$

For the second line, since $p^{1/2}(1-1/\sqrt{x})=p^{1/2}(x-1)/(\sqrt{x}+x)$, we only need to show $p\Big[1-\mathbb{E}[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})-\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})]^2\Big]\longrightarrow 0$.

$$\begin{aligned}&1-\mathbb{E}\Big[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})-\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})\Big]^2\\&=\mathbb{E}\breve{z}_{ij}^2-\mathbb{E}\breve{z}_{ij}^2\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})+\{\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})\}^2\\&=\mathbb{E}\breve{z}_{ij}^2\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})+[\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})]^2\\&\leqslant 2\mathbb{E}\breve{z}_{ij}^2\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})\leqslant 2\varepsilon_n^{-2}n^{-1}\mathbb{E}\breve{z}_{ij}^4\mathbb{1}(|\breve{z}_{ij}|>\varepsilon_n n^{1/2})=o(p^{-1}).\end{aligned}$$

Therefore, $|d_{18}|\xrightarrow{P}0$.

$$\begin{aligned}d_{19}&=n^{-1/2}\alpha^TQ_n^T\breve{\mathbf{Z}}^T\Sigma_p^{T/2}[(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}-(\breve{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}]\Sigma_p^{1/2}\mathbf{Z}Q_n\eta\\&=n^{-1/2}\alpha^TQ_n^T\breve{\mathbf{Z}}^T\Sigma_p^{T/2}(\widehat{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}[\breve{\mathbf{\Sigma}}_p-\widehat{\mathbf{\Sigma}}_p](\breve{\mathbf{\Sigma}}_p-\mathbb{z}I)^{-1}\Sigma_p^{1/2}\mathbf{Z}Q_n\eta.\end{aligned}$$

It is easy to check that Lemma S.15 still holds for $\breve{\mathbf{Z}}$ when $m=2$. We have

$$n^{-1}\mathbb{E}\alpha^TQ_n^T\breve{\mathbf{Z}}^T\breve{\mathbf{Z}}Q_n\alpha=O(1).$$

Thus, we only need to show,

$$n^{1/2}\|\breve{\mathbf{\Sigma}}_p-\widehat{\mathbf{\Sigma}}_p\|_2\leqslant n^{-1/2}\|\mathbf{Z}-\breve{\mathbf{Z}}\|_2(\|\mathbf{Z}\|_2+\|\breve{\mathbf{Z}}\|_2)\xrightarrow{P}0. \tag{S.6.1}$$

Since $n^{-1/2}\|\mathbf{Z}\|$ and $n^{-1/2}\|\breve{\mathbf{Z}}\|$ are $O_p(1)$ (Yin, Bai and Krishnaiah, 1988), we only need to show

$$\|\breve{\mathbf{Z}}-\mathbf{Z}\|_2\xrightarrow{P}0.$$

$$\begin{aligned}\breve{\mathbf{Z}}-\mathbf{Z}=&\Big[1-\Big\{\mathbb{E}\Big[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})-\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})\Big]^2\Big\}^{-1/2}\Big]\breve{\mathbf{Z}}\\&+\{\mathbb{E}[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})-\mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}|\leqslant\varepsilon_n n^{1/2})]^2\}^{-1/2}\mathfrak{F},\end{aligned}$$

where $\mathfrak{F}$ is a matrix with the (i,j)-th entry being $\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| > \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| > \varepsilon_n n^{1/2})$. It remains to show $\|\mathfrak{F}\|_2 \xrightarrow{P} 0$.

Note, $\mathfrak{F}$ has i.i.d. entries with mean 0 and variance

$$\mathbb{E}[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| > \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| > \varepsilon_n n^{1/2})]^2 = o(p^{-1}).$$

Due to the previous arguments. Again using results in Yin, Bai and Krishnaiah (1988), almost surely,

$$\left[\mathbb{E}[\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| > \varepsilon_n n^{1/2}) - \mathbb{E}\breve{z}_{ij}\mathbb{1}(|\breve{z}_{ij}| > \varepsilon_n n^{1/2})]^2\right]^{-1/2} n^{-1/2}\|\mathfrak{F}\|_2 = O_p(1).$$

Therefore, $\|\mathfrak{F}\|_2 \xrightarrow{P} 0$.

As for d_{20}, the argument for d_{18} works, since $n^{-1/2}\|\alpha^T Q_n^T \breve{\mathbf{Z}}^T\|_2 = O_p(1)$ also holds. It completes the proof of Lemma S.3.

Weak limits remain unchanged after process smoothing Lemma S.3 implies Lemma A.4 also holds when $\mathbf{Z}$ satisfies **C1**, Now, re-define $\xi_n(\mathbb{z}, \alpha, \beta)$ with $\mathbf{Z}$ replaced by $\breve{\mathbf{Z}}$. When $\lambda_{\max}(\breve{\boldsymbol{\Sigma}}_p) < \overline{u}$,

$$\begin{aligned}
&\Big|\oint_{\mathcal{C}} f(\mathbb{z})\widehat{\xi}_n(\mathbb{z}, \alpha, \eta)d\mathbb{z} - \oint_{\mathcal{C}} f(\mathbb{z})\xi_n(\mathbb{z}, \alpha, \eta)d\mathbb{z}\Big| \\
&\leqslant \mathcal{K}\rho_n n^{-1/2}\|\breve{\mathbf{Z}}\breve{\mathbf{Z}}^T\|_2\|\alpha\|_2\|\eta\|_2(|\overline{u} - \lambda_{\max}(\breve{\boldsymbol{\Sigma}}_p)|^{-1} + |\underline{u}^{-1}|).
\end{aligned} \tag{S.6.2}$$

The right-hand side above converges to 0 in probability since $\rho_n n^{1/2} \to 0$. Therefore, under **C1**,

$$\begin{aligned}
&\frac{-1}{2\pi i}\oint_{\mathcal{C}} f(\mathbb{z})\xi_n(\mathbb{z}, \alpha, \eta)d\mathbb{z} - n^{1/2}\Omega(f, \gamma)\alpha^T\eta \\
&\quad \xrightarrow{D} \mathcal{N}(0, [\|\alpha\|_2^2\|\eta\|_2^2 + (\alpha^T\eta)^2]\Delta(f, \gamma)).
\end{aligned}$$

When $\lambda_{\max}(\breve{\boldsymbol{\Sigma}}_p) < \overline{u}$,

$$\begin{aligned}
&n^{-1/2}\alpha^T V_n^T U_n^T \breve{\mathbf{Z}}^T \Sigma_p^{T/2} f(\breve{\boldsymbol{\Sigma}}_p)\Sigma_p^{1/2}\breve{\mathbf{Z}}U_n V_n\eta \\
=&\frac{-1}{2\pi i}\oint_{\mathcal{C}} f(\mathbb{z})n^{-1/2}\alpha^T V_n^T U_n^T \breve{\mathbf{Z}}^T \Sigma_p^{T/2}(\breve{\boldsymbol{\Sigma}}_p - \mathbb{z}I)^{-1}\Sigma_p^{1/2}\breve{\mathbf{Z}}U_n V_n\eta d\mathbb{z}.
\end{aligned}$$

Therefore, for arbitrary α and η,

$$\left[\sqrt{n}\alpha\mathbf{M}(f)\eta - n^{1/2}\Omega(f, \gamma)\alpha^T\eta\right] \Longrightarrow \mathcal{N}(0, [\|\alpha\|_2^2\|\eta\|_2^2 + (\alpha^T\eta)^2]\Delta(f, \gamma)).$$

S.7. Technical lemmas

There are a collection of lemmas built under (A.7). Proofs of these lemmas are omitted. Similar work exists literature. See for example (3.10)- (3.14) of Pan and Zhou (2011). Recall the notation list in Section S.3.

Lemma S.6 *(Woodbury formula) The following identity holds*

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

for matrices A, U, C, V of conformable sizes and assuming all inverse operations are well-defined.

Lemma S.7 *(Fan (1951)) Let A and C be two $p \times n$ complex matrices. Then, for any nonnegative integers i and j, we have*

$$s_{i+j+1}(A + C) \leqslant s_{i+1}(A) + s_{j+1}(C),$$

where $s_i(\cdot)$ is the i-th largest singular value of a matrix.

Lemma S.8 *(Burkholder) Let $\{Y_i\}$ be a complex martingale difference sequence with respect to the increasing σ-field $\{\sigma_i\}$. Then for $m \geqslant 2$*

$$\mathbb{E}\Big|\sum_i Y_i\Big|^m \leqslant \mathcal{K}_m \mathbb{E}(\sum_i \mathbb{E}(|Y_i|^2 \mid \sigma_{i-1}))^{m/2} + \mathcal{K}_m \mathbb{E}(\sum_i |Y_i|^m).$$

Lemma S.9 *(Lemma 2.7 of Bai and Silverstein (1998)) Let $\mathbf{Y} = (Y_1, \dots, Y_p)^T$, where Y_i's are i.i.d. real r.v.'s with mean 0 and variance 1. Let $\mathbf{B} = (b_{ij})_{p\times p}$, a deterministic complex matrix. Then for any $m \geqslant 2$, we have*

$$\mathbb{E}|\mathbf{Y}^T\mathbf{B}\mathbf{Y} - \mathrm{tr}\mathbf{B}|^m \leqslant \mathcal{K}_m(\mathbb{E}Y_1^4 \mathrm{tr}\mathbf{B}\mathbf{B}^*)^{m/2} + \mathcal{K}_m \mathbb{E}Y_1^{2m} \mathrm{tr}[(\mathbf{B}\mathbf{B}^*)^{m/2}],$$

where $\mathbf{B}^*$ denotes the complex conjugate transpose of $\mathbf{B}$, and $\mathcal{K}$ is a constant only depending on m.

Lemma S.10 *((3.4) of Bai and Silverstein (1998)) For any $\mathbb{z} = u + iv$ with $v > 0$,*

$$\|\mathbf{A}(\mathbb{z})\|_2 \leqslant \frac{\|\Sigma_p\|}{v}, \qquad |\beta_1(\mathbb{z})| \leqslant \frac{|\mathbb{z}|}{v}, \qquad |\beta_1^{\mathrm{tr}}(\mathbb{z})| \leqslant \frac{|\mathbb{z}|}{v}.$$

Lemma S.11 *(Lemma 2.10 of Bai and Silverstein (1998)) For any matrix $\mathbf{D}$ and $z = u + iv$ with $v > 0$,*

$$\Big|\mathrm{tr}\{\mathbf{A}(\mathbb{z})\mathbf{D} - \mathbf{A}_j(\mathbb{z})\mathbf{D}\}\Big| \leqslant \frac{\|\mathbf{D}\Sigma_p\|_2}{v}.$$

Lemma S.12 *For $m \geqslant 2$ and any fixed z with non-zero imaginary part,*

$$n^{-m}\mathbb{E}\Big|\mathrm{tr}\mathbf{A}(\mathbb{z}) - \mathbb{E}\mathrm{tr}\mathbf{A}(\mathbb{z})\Big|^m = O(n^{-m/2}).$$

Lemma S.13 *For a sequence of deterministic matrices $\mathbf{D}$ such that $\|\mathbf{D}\|_2 < \infty$, $m \geqslant 2$,*

$$\begin{aligned} &n^{-m}\mathbb{E}\Big|\mathbf{z}_1^T\mathbf{D}\mathbf{z}_1 - \mathrm{tr}(\mathbf{D})\Big|^m \\ &\leqslant \mathcal{K}_m n^{-m}\Big\{[\mathrm{tr}(\mathbf{D}\mathbf{D}^*)]^{m/2} + \varepsilon_n^{2m-4}n^{m-2}\mathrm{tr}[(\mathbf{D}\mathbf{D}^*)^{m/2}]\Big\} \leqslant \mathcal{K}_m\|\mathbf{D}\|_2^m\varepsilon_n^{2m-4}n^{-1} \end{aligned}$$

for some constant $\mathcal{K}_m$.

Lemma S.14 *For sequences of deterministic matrices* $\mathbf{D}$ *and* $\mathbf{G}$ *such that* $\|\mathbf{D}\|_2 < \infty$ *and* $\|\mathbf{G}\|_2 < \infty$, *for* $m \geqslant 2$,

$$n^{-m}\mathbb{E}\Big|\mathbf{z}_1^T\mathbf{D}e_ie_j^T\mathbf{G}\mathbf{z}_1\Big|^m \leqslant \mathcal{K}_m\|\mathbf{D}\|_2^m\|\mathbf{G}\|_2^m[\varepsilon_n^{2m-4}n^{-2} + n^{-m}] = O(n^{-2}\varepsilon_n^{2m-4}),$$

for some constant $\mathcal{K}_m$.

Lemma S.15 *For a sequence of deterministic matrices* $\mathbf{D}$ *such that* $\|\mathbf{D}\|_2 < \infty$ *and a sequence of vector* $\boldsymbol{u}$ *such that* $\limsup_{n\to\infty} n\|\boldsymbol{u}\|_{\max} \leqslant \mathcal{K}_{\max} < \infty$,

$$\mathbb{E}|\boldsymbol{u}^T\mathbf{Z}^T\mathbf{D}\mathbf{Z}\boldsymbol{u}|^m \leqslant \mathcal{K}_m\mathcal{K}_{\max}^{2m}\|\mathbf{D}\|_2^m,$$

for $m \geqslant 2$ *and some constant* $\mathcal{K}_m > 0$.

Lemma S.16 *For a sequence of deterministic matrices* $\mathbf{D}$ *such that* $\|\mathbf{D}\|_2 < \infty$ *and a sequence of vector* $\boldsymbol{u}$ *such that* $\limsup_n \to \infty n\|\boldsymbol{u}\|_{\max} \leqslant \mathcal{K}_{\max} < \infty$,

$$\mathbb{E}\Big|\mathbf{z}_1^T\mathbf{D}\mathbf{Z}_1\boldsymbol{u}\Big|^m \leqslant \mathcal{K}_m\mathcal{K}_{\max}^m\|\mathbf{D}\|_2^m n^{m/2-2}\varepsilon_n^{m-4},$$

for $m \geqslant 4$ *and some constant* $\mathcal{K}_m > 0$.

Lemma S.17 *For a sequence of deterministic matrices* $\mathbf{D}$ *such that* $\|\mathbf{D}\|_2 < \infty$,

$$\mathbb{E}\big|\mathbf{z}_1^T\mathbf{D}\mathbf{z}_2\big|^m \leqslant \mathcal{K}_m\|\mathbf{D}\|_2^m n^{m-2}\varepsilon_n^{m-4},$$

for $m \geqslant 4$ *and some constant* $\mathcal{K}_m > 0$.

Lemma S.18 *For sequences of deterministic matrices* $\mathbf{D}_1, \ldots, \mathbf{D}_m$, $\mathbf{G}_1, \ldots, \mathbf{G}_s$, *and* $\mathbf{J}$ *such that* $\|\mathbf{D}_j\|_2 < \infty$, $\|\mathbf{G}_j\|_2 < \infty$, *and* $\|\mathbf{J}\|_2 < \infty$, *and a sequence of vector* $\boldsymbol{u}$ *such that* $\limsup_{n\to\infty} n\|\boldsymbol{u}\|_{\max} \leqslant \mathcal{K}_{\max} < \infty$,

$$\begin{aligned}&\mathbb{E}\Big|\prod_{i=1}^m \frac{1}{n}\mathbf{z}_1^T\mathbf{D}_i\mathbf{z}_1 \prod_{j=1}^s \frac{1}{n}(\mathbf{z}_1^T\mathbf{G}_j\mathbf{z}_1 - \mathrm{tr}\mathbf{G}_j)(\mathbf{z}_1^T\mathbf{J}\mathbf{Z}_1\boldsymbol{u})^t\Big| \\ &\leqslant \mathcal{K}_{m,s,t}\prod_{i=1}^m \|\mathbf{D}_i\|_2 \prod_{j=1}^s \|\mathbf{G}_j\|\|\mathbf{J}\|_2^t\mathcal{K}_{\max}^t n^{-1/2}\varepsilon_n^{\max(s-2,0)},\end{aligned}$$

where $m \geqslant 0, s \geqslant 1, 0 \leqslant t \leqslant 2$ *and some constant* $\mathcal{K}_{m,s,t} > 0$.

Lemma S.19 *For a vector* $\mathbf{r}$ *and deterministic matrices* $\mathbf{D} = (d_{ij})$ *and* $\mathbf{G}$,

$$\mathbb{E}[(\mathbf{z}_1^T\mathbf{D}\mathbf{z}_1 - \mathrm{tr}\mathbf{D})\mathbf{z}_1^T\mathbf{G}\mathbf{r}] = \mathbb{E}z_{11}^3\sum_{i=1}^p d_{ii}e_i^T\mathbf{G}\mathbf{r},$$

where e_i *is the canonical vector with the* i*th entry* 1.

Lemma S.20 *For deterministic matrices* $\mathbf{D} = (d_{ij})$ *and* $\mathbf{G} = (g_{ij})$,

$$\begin{aligned}\mathbb{E}[(\mathbf{z}_1^T\mathbf{D}\mathbf{z}_1 - \mathrm{tr}\mathbf{D})(\mathbf{z}_1^T\mathbf{G}\mathbf{z}_1 - \mathrm{tr}\mathbf{G})] &= |\mathbb{E}z_{11}^2|^2\mathrm{tr}\mathbf{D}\mathbf{G}^T + \mathrm{tr}\mathbf{D}\mathbf{G} + (\mathbb{E}z_{11}^4 \\ &\quad - |\mathbb{E}z_{11}^2|^2 - 2)\sum_{i=1}^p d_{ii}g_{ii}.\end{aligned}$$

S.8. Additional simulation studies

Figure S.8.1 – Figure S.8.12 display additional size-adjusted power curves.

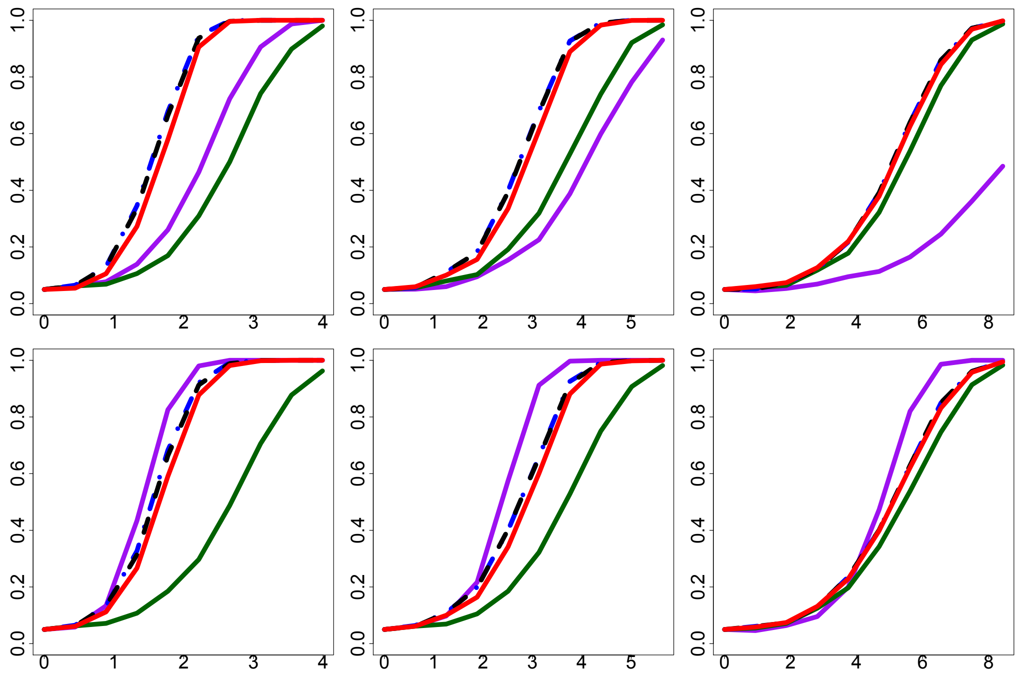

Figure S.8.1: Size-adjusted power with $\Sigma = \Sigma_{den}$, $k = 5$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LH}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LH}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LH}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (1, 0, 0)$.

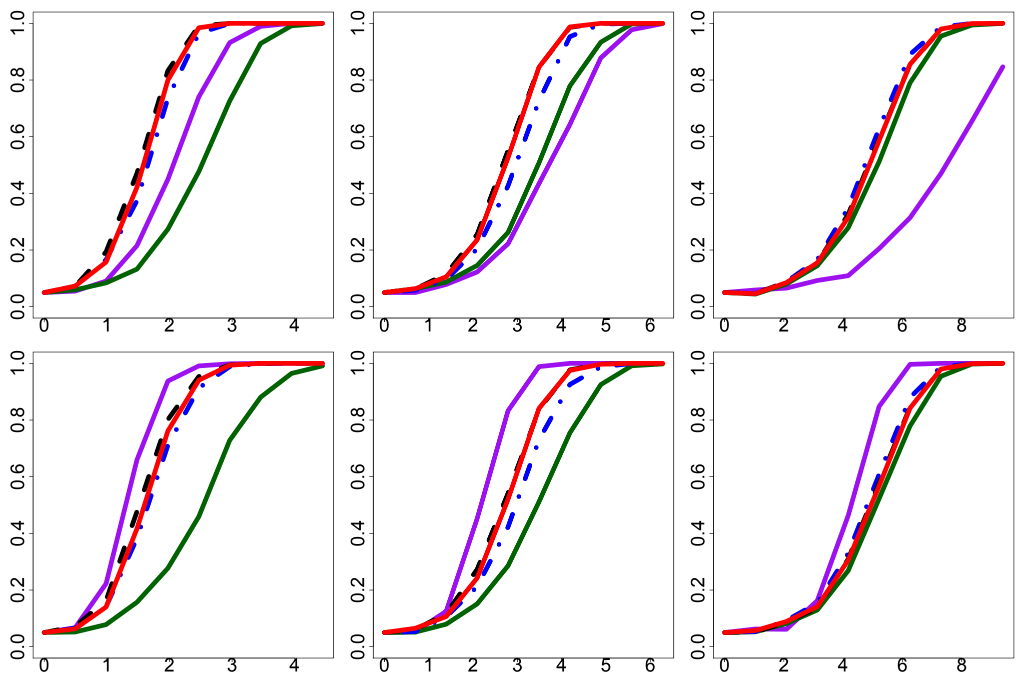

Figure S.8.2: Size-adjusted power with $\Sigma = \Sigma_{den}$, $k = 3$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. BNP_{comp} (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\text{BNP}_{\text{ridge}}$ (black, dashed) and BNP_{high} (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$.

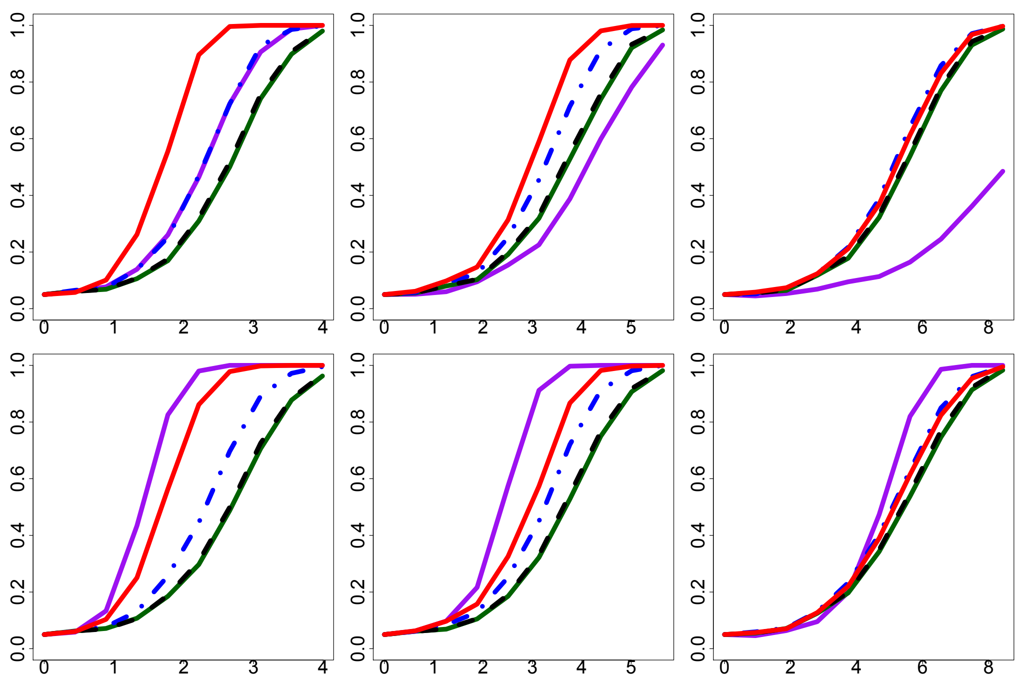

Figure S.8.3: Size-adjusted power with $\Sigma = \Sigma_{den}$, $k = 5$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LR}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LR}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LR}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$.

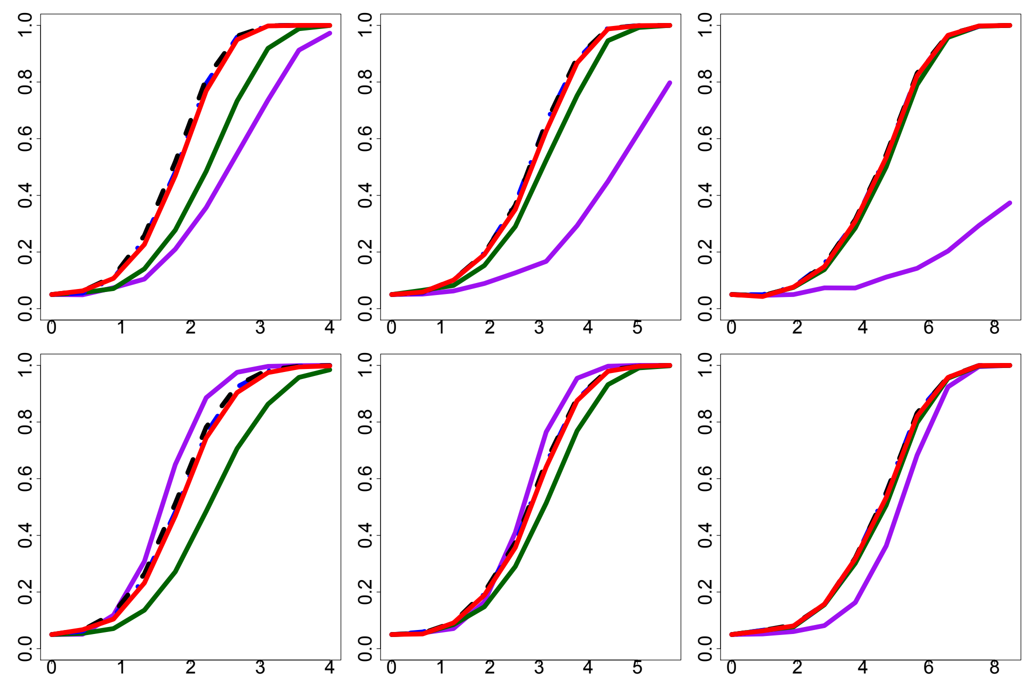

Figure S.8.4: Size-adjusted power with $\Sigma = \Sigma_{toep}$, $k = 3$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LH}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LH}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LH}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (1, 0, 0)$.

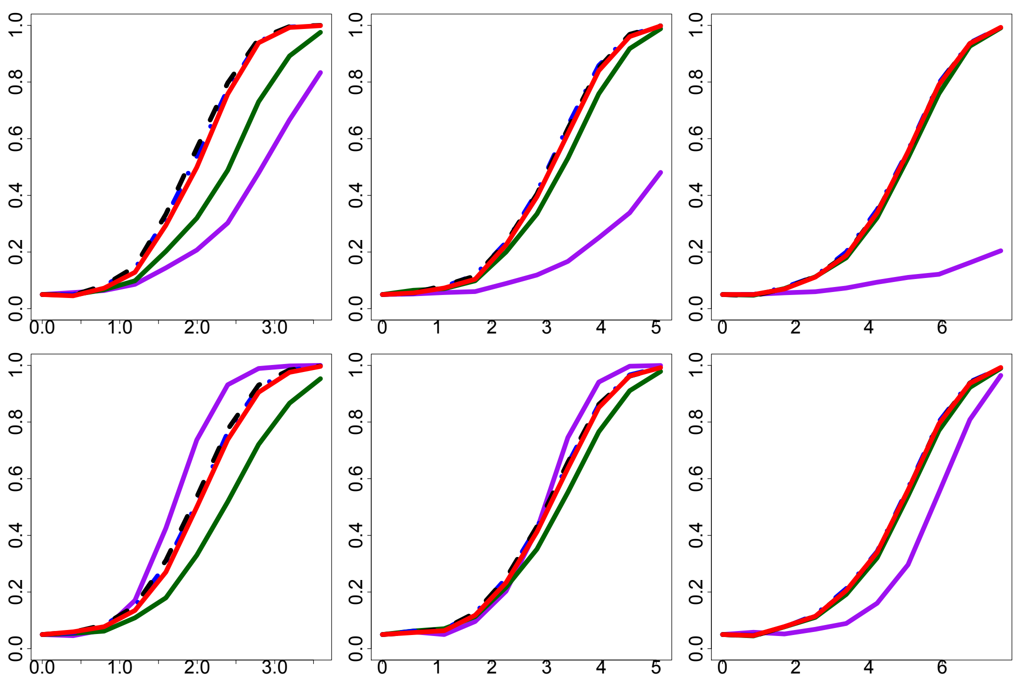

Figure S.8.5: Size-adjusted power with $\Sigma = \Sigma_{toep}$, $k = 5$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{BNP}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{BNP}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{BNP}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$.

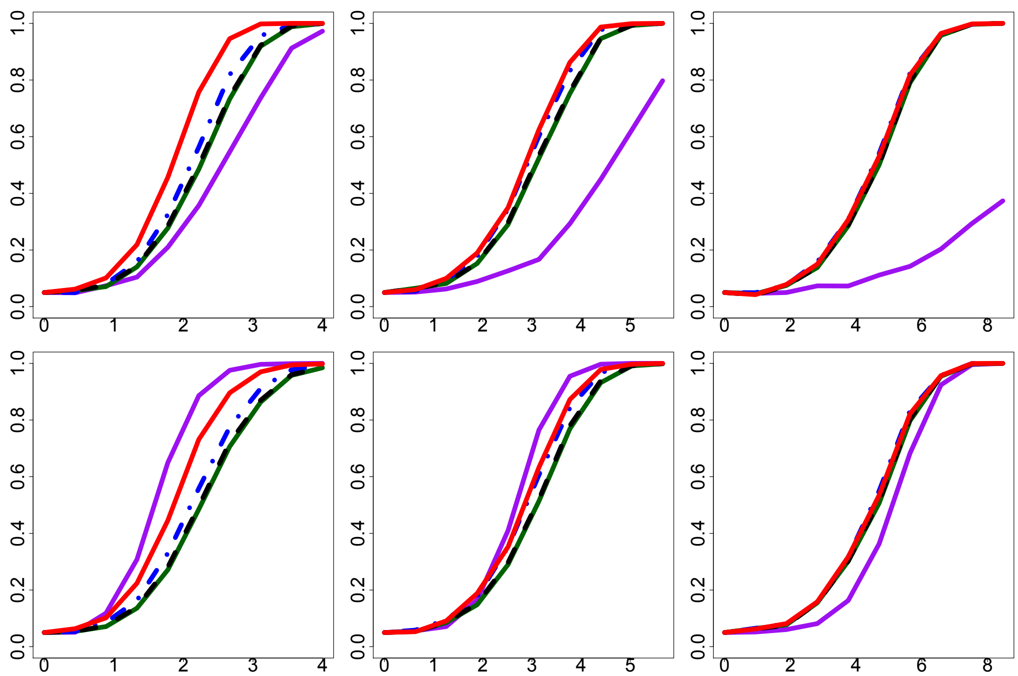

Figure S.8.6: Size-adjusted power with $\Sigma = \Sigma_{toep}$, $k = 3$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LR}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LR}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LR}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$.

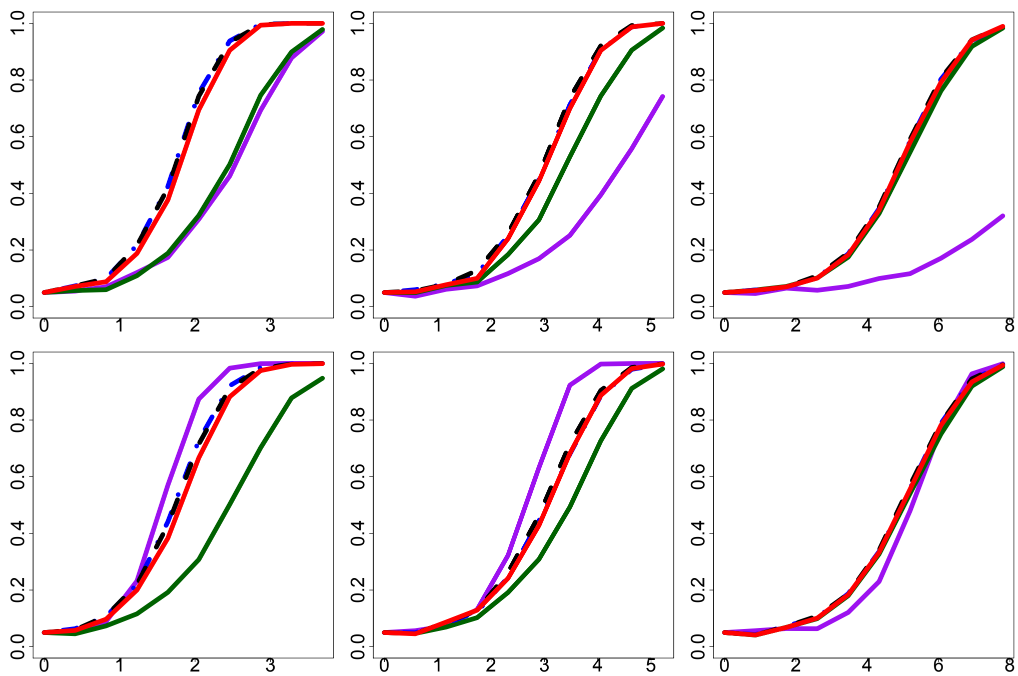

Figure S.8.7: Size-adjusted power with $\Sigma = \Sigma_{dis}$, $k = 5$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LH_{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LH_{ridge}}$ (black, dashed) and $\mathrm{LH_{high}}$ (blue, dotted-dashed) with $\tilde{t} = (1, 0, 0)$.

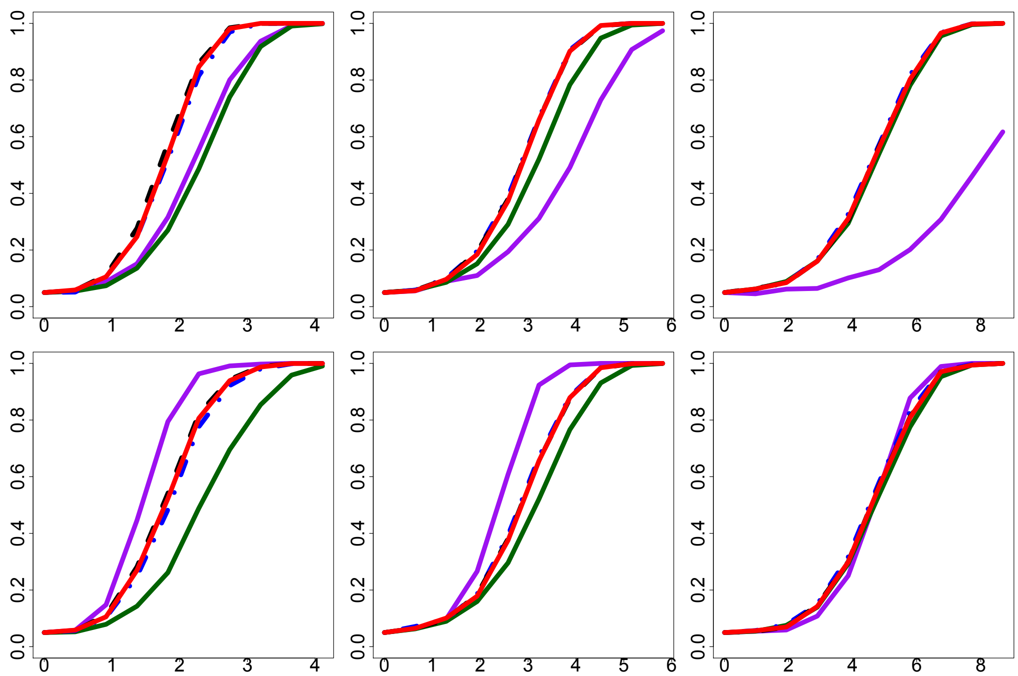

Figure S.8.8: Size-adjusted power with $\Sigma = \Sigma_{dis}$, $k = 3$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{BNP}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{BNP}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{BNP}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$.

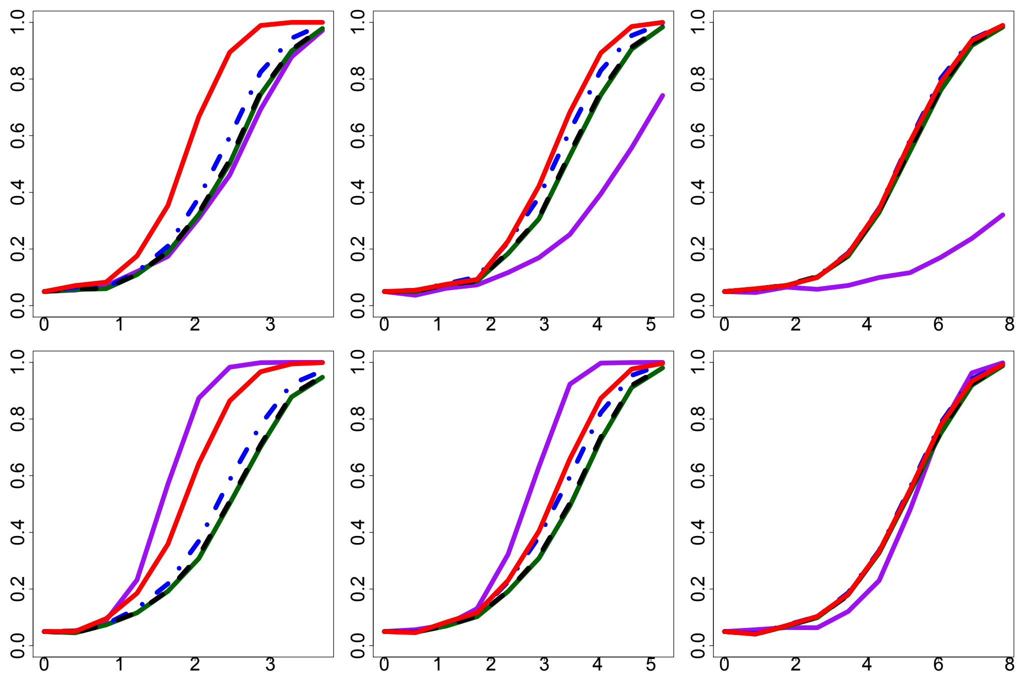

Figure S.8.9: Size-adjusted power with $\Sigma = \Sigma_{dis}$, $k = 5$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LR}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LR}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LR}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$.

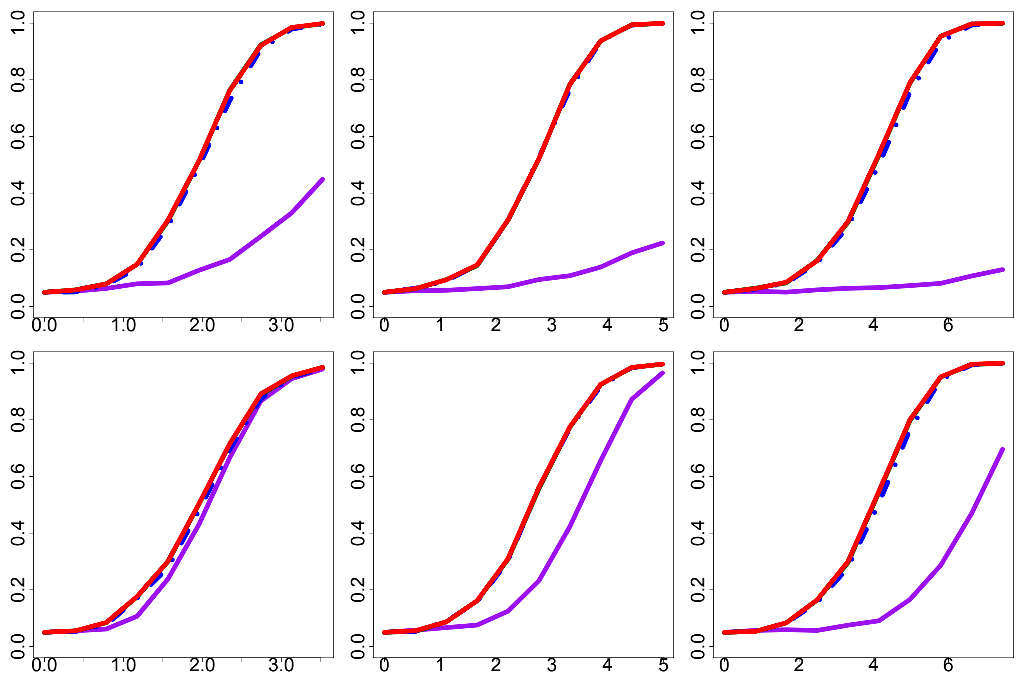

Figure S.8.10: Size-adjusted power with $\Sigma = I_p$, $k = 3$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LH}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LH}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LH}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (1, 0, 0)$.

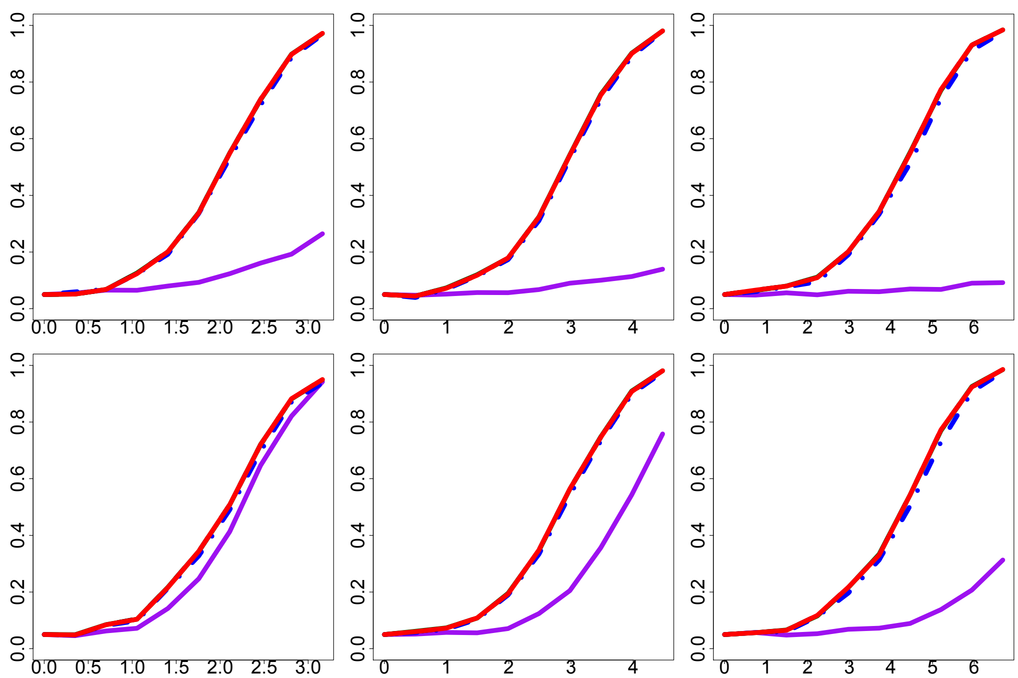

Figure S.8.11: Size-adjusted power with $\Sigma = I_p$, $k = 5$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. BNP_{comp} (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\text{BNP}_{\text{ridge}}$ (black, dashed) and BNP_{high} (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$.

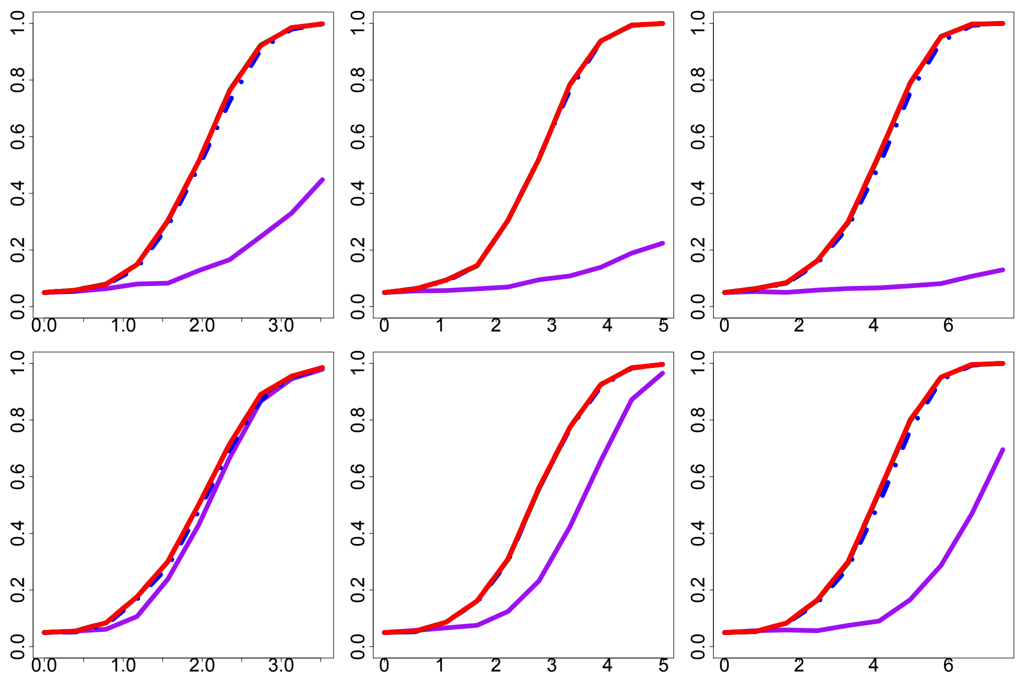

Figure S.8.12: Size-adjusted power with $\Sigma = I_p$, $k = 3$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LR}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LR}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LR}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$.

S.9. Nominal power curves

Figure S.9.1–Figure S.9.3 are counterparts of Figure 5.1–Figure 5.3 in the manuscript but with asymptotic (approximate) cut-off values. We identify that the difference between them and the size-adjusted power curves are negligible.

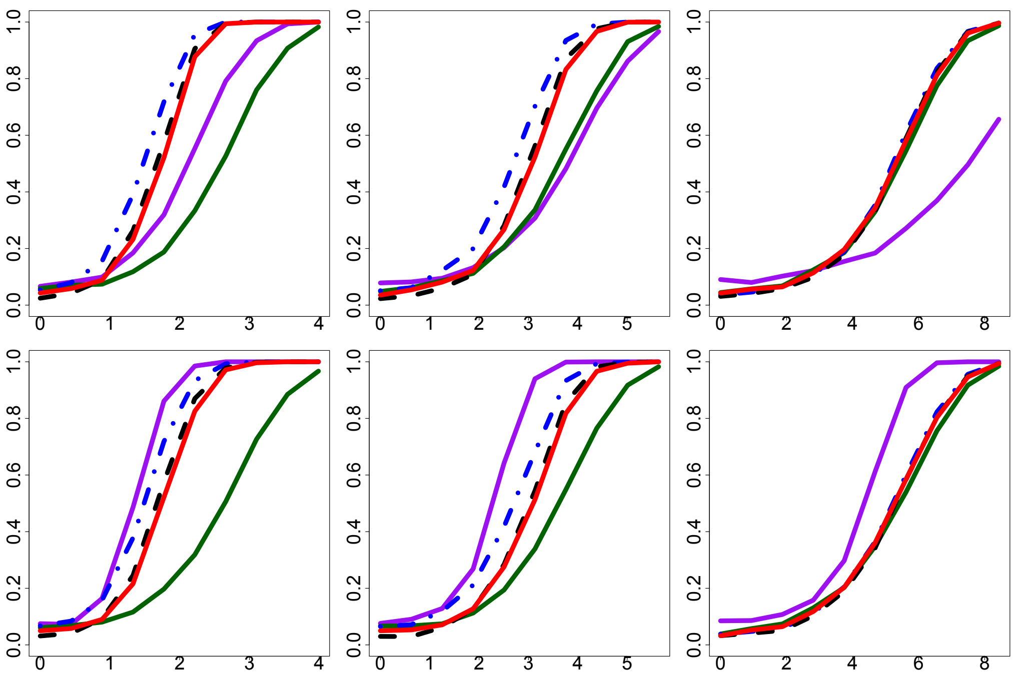

Figure S.9.1: Empirical power with $\Sigma = \Sigma_{den}$, $k = 5$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. BNP_{comp} (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\text{BNP}_{\text{ridge}}$ (black, dashed) and BNP_{high} (blue, dotted-dashed) with $\tilde{t} = (1, 0, 0)$.

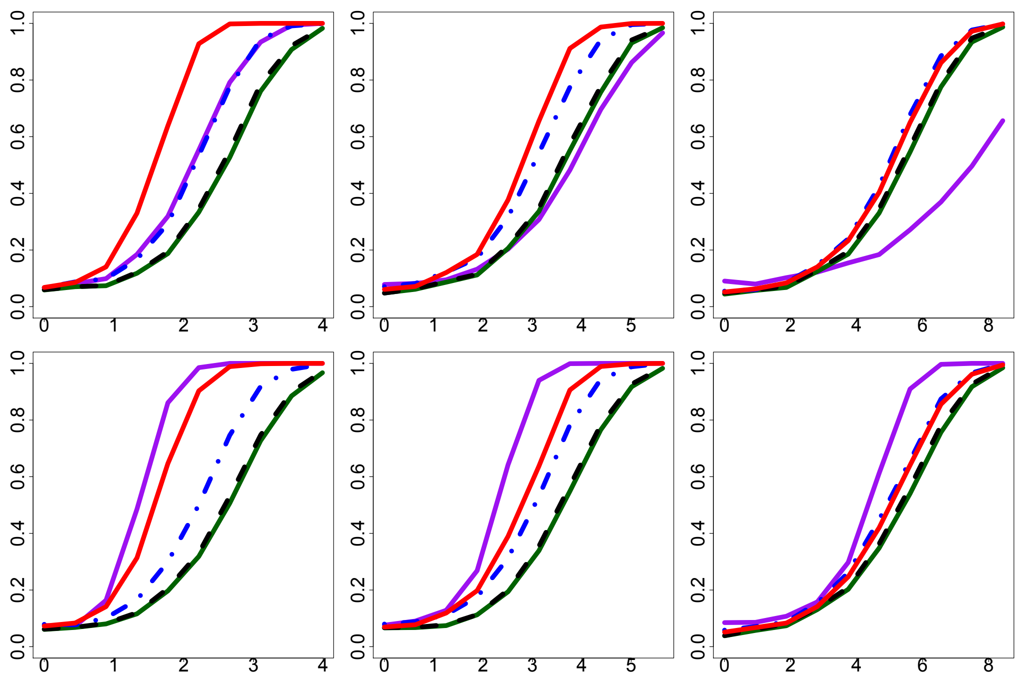

Figure S.9.2: Empirical power with $\Sigma = \Sigma_{den}$, $k = 5$. Rows (top to bottom): B = Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LH}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LH}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LH}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$.

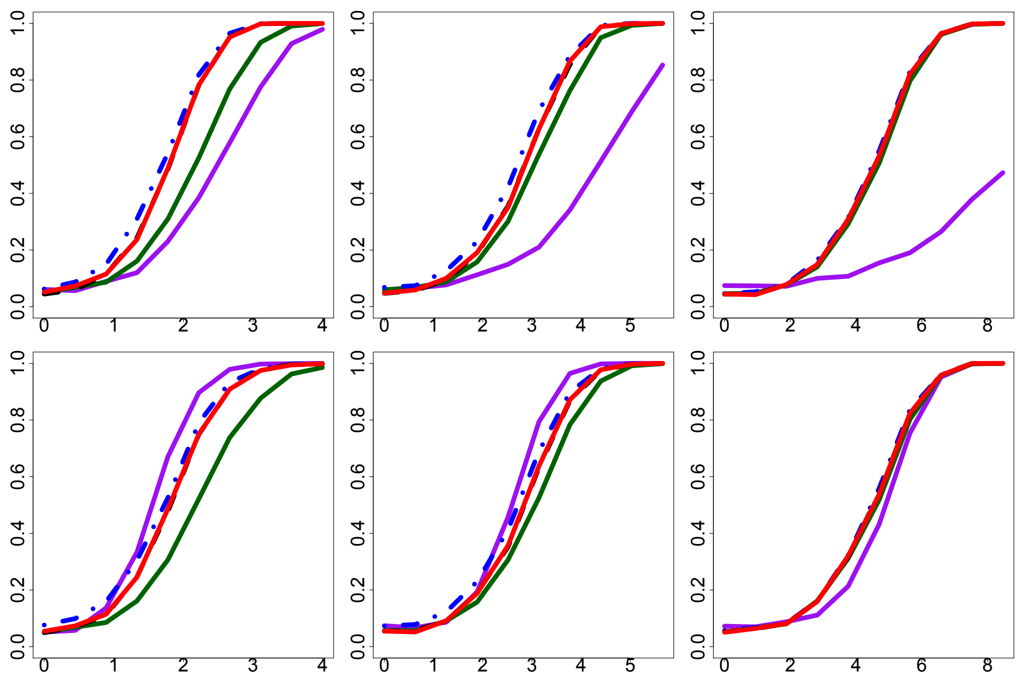

Figure S.9.3: Empirical power with $\Sigma = \Sigma_{toep}$, $k = 3$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\mathrm{LR}_{\mathrm{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\mathrm{LR}_{\mathrm{ridge}}$ (black, dashed) and $\mathrm{LR}_{\mathrm{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$.