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

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

(1.1)

where (i) $\mathbf{B}$ is a $p \times k$ matrix of regression coefficients; (ii) $\mathbf{X}$ 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) $\Sigma_p$, a $p \times p$ nonnegative definite matrix, is the population covariance matrix of the errors, with $\Sigma_p^{1/2}$ a “square-root” of $\Sigma_p$ 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

$$H_0: \mathbf{BC} = 0 \quad \text{vs.} \quad H_\alpha: \mathbf{BC} \neq 0,$$

(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

$$\hat{\Sigma}_p = \frac{1}{n} Y(I - X^T(XX^T)^{-1}X)YT, \quad (1.3)$$

$$\hat{H}_p = \frac{1}{n} YXT(XX^T)^{-1}C[C^T(XX^T)^{-1}C]^{-1}C^T(XX^T)^{-1}XYT, \quad (1.4)$$

so that $\hat{\Sigma}_p$ is the residual covariance of the full model, an estimator of $\Sigma_p$, while $\hat{H}_p$ is the hypothesis sums of squares and cross products matrix, scaled by $n^{-1}$. In a one-way MANOVA set-up, $\hat{\Sigma}_p$ and $\hat{H}_p$ 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 $\hat{\Sigma}_p$ 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 $\Sigma_p$. One common feature is that all test statistics are linear functionals of the spectrum of $\hat{H}_p\hat{\Sigma}_p^{-1}$. Since this matrix is asymmetric, for convenience, a standard transformation is applied, giving the expressions of the invariant tests as follows. Define

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

$$M_0 = \frac{1}{n} Q_n^T \hat{\Sigma}_p^{-1} YQ_n.$$

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 $\hat{H}_p\hat{\Sigma}_p^{-1} = n^{-1} YQ_nQ_n^T Y^T \hat{\Sigma}_p^{-1}$ are the same as those of $M_0$. The test statistics for the LR, LH and BNP tests can be expressed as

$$T_{0LR} = \sum_{i=1}^{q} \log\{1 + \lambda_i(M_0)\},$$

$$T_{0LH} = \sum_{i=1}^{q} \lambda_i(M_0),$$

$$T_{0BNP} = \sum_{i=1}^{q} \lambda_i(M_0)/\{1 + \lambda_i(M_0)\}.$$  

The symbol $\lambda_i(\cdot)$ denotes the $i$-th largest eigenvalue of a symmetric matrix, further using the convention that $\lambda_{\text{max}}(\cdot)$ and $\lambda_{\text{min}}(\cdot)$ indicate the largest and smallest eigenvalue.
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 \( \hat{\Sigma}_p \) —, 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 \( \hat{\Sigma}_p \) 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. LR, LH and BNP tests can be generalized as linear spectral statistics of large-dimensional \( F \)-matrices, whose CLT is studied in Zheng (2012); Zheng, Bai and Yao (2017); Bodnar, Dette and Parolya (2017).

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 \( n \geq p \) 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 \( \hat{\Sigma}_p \) 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 \( \hat{\Sigma}_p \) 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). Other related works include Zhu and Bradic (2018).

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

\[
g(A) = R_A \text{diag}(g(\lambda_1(A)), \ldots, g(\lambda_p(A))) R_A^T,
\]

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 \( \bar{\Sigma}_p \), as elaborated in Section 2. The proposed statistics are functionals of eigenvalues of the regularized quadratic forms

\[
M(f) = \frac{1}{n} Q_n^T Y^T f(\bar{\Sigma}_p) Y Q_n.
\]

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

\[
T^{LR}(f) = \sum_{i=1}^{q} \log(1 + \lambda_i(M(f))),
\]

\[
T^{LH}(f) = \sum_{i=1}^{q} \lambda_i(M(f)),
\]

\[
T^{BNP}(f) = \sum_{i=1}^{q} \frac{\lambda_i(M(f))}{1 + \lambda_i(M(f))}.
\]

These test statistics are designed to capture possible departures from the null hypothesis, when \( \Sigma_p \) is replaced by \( f(\bar{\Sigma}_p) \), 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 \( M(f) \) has the same non-zero eigenvalues as \( f(\bar{\Sigma}_p)\bar{H}_p \). Thus, the proposed test family is a generalization of the classical statistics based on \( \Sigma_p^{-1} \bar{H}_p \). Importantly, \( M(f) \) — 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 \( \Sigma_p \) 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 \( \Sigma_p = I_p \), \( \Sigma_p \) 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. In the Appendix, a proof outline of the main theorem is presented, while technical details and proofs of other theorems 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

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

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

C1 (Moment conditions) $Z$ has i.i.d. entries $z_{ij}$ 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 $p, n \rightarrow \infty$ such that $\gamma_n = p/n \rightarrow \gamma \in (0, \infty)$ and $\sqrt{n} |\gamma_n - \gamma| \rightarrow 0$;

C3 (Boundedness of spectral norm) $\Sigma$ is non-negative definite; $\limsup_p \lambda_{\text{max}}(\Sigma_p) < \infty$;

C4 (Asymptotic stability of ESD) There exists a distribution $L^{\Sigma}$ with compact support in $[0, \infty)$, non-degenerate at zero, such that $\sqrt{n} D_W(F^{\Sigma}, L^{\Sigma}) \rightarrow 0$, as $n, p \rightarrow \infty$, where $D_W(F_1, F_2)$ denotes the Wasserstein distance between distributions $F_1$ and $F_2$, defined as

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

C5 (Asymptotically full rank) $X$ is of full rank and $n^{-1}XX^T$ converges to a positive definite $k \times k$ matrix. Moreover, $\limsup_{n \rightarrow \infty} \|X\|_{\text{max}} < \infty$;

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

Remark 2.1 The conditions are mild: C5 and C6 are commonly made in multivariate analysis for asymptotic results of regression models under non-Gaussianity. The
In the proposed inferential procedure. First, it imposes little constraint on the observations. Secondly, we always use $\gamma_n$ and $F_{\Sigma_0}$ to estimate $\gamma$ and $L_{\Sigma}$ in the proposed inferential procedure.

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

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

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

$$m(z) = \int \frac{dL^\Sigma_{\tau}(\tau)}{(1-\gamma - \gamma zm(z)) - z}.$$ \hspace{1cm} (2.1)

Equation (2.1) is often referred to as the Marchenko–Pastur equation. Moreover, pointwise almost surely for $z \in \mathbb{C}^+$, $m_{F_\Sigma}(z)$ converges to $m_{F_\Sigma}(z)$. The convergence holds even when $z \in \mathbb{R}_-$ (negative reals) with a smooth extension of $m_{F_\Sigma}$ 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 $m_{F_\Sigma}(z)$ as $m(z)$ and write $m_{F_{\Sigma_0}}(z)$ as $m_{n,p}(z)$. Note that

$$m_{n,p}(z) = p^{-1} \text{tr}(\hat{\Sigma}_p - zI_p)^{-1}$$

and define

$$\Theta(z, \gamma) = \{1 - \gamma - \gamma zm(z)\}^{-1}.$$ \hspace{1cm} (2.2)

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

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

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

$$p^{-1} \text{tr}[(\hat{\Sigma}_p - zI_p)^{-1}A] - p^{-1} \text{tr}[(\Theta^{-1}(z, \gamma) \Sigma_p - zI\}^{-1}A] \to 0, \quad \text{with probability 1.}$$

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 \( q \geq 1 \), denote by \( W = [w_{ij}]_{i,j=1}^q \) 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 \( 1 \leq i \leq j \leq q \). 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 \( C \) to be a closed contour enclosing \( \mathcal{X} \) such that \( f(\cdot) \) has a complex extension to the interior of \( C \). Further use \( C^2 \) to denote \( C \otimes C = \{(z_1, z_2) : z_1, z_2 \in C\} \).

**Theorem 2.1** Suppose C1–C6 hold. Under the null hypothesis \( H_0: BC = 0 \),

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

where \( \Longrightarrow \) denotes weak convergence and \( \Omega(f, \gamma) \) and \( \Delta(f, \gamma) \) are as follows.

\[
\Omega(f, \gamma) = \frac{-1}{2\pi i} \oint_C f(z)(\Theta(z, \gamma) - 1)dz.
\]

See (2.2) for the definition of \( \Theta(z, \gamma) \). For any two analytic functions \( f_1 \) and \( f_2 \),

\[
\Delta(f_1, f_2, \gamma) = \frac{2}{(2\pi i)^2} \iint_{C^2} f_1(z_1)f_2(z_2)\delta(z_1, z_2, \gamma)d\bar{z}_2dz_2,
\]

and \( \Delta(f, f, \gamma) \) is written as \( \Delta(f, \gamma) \) for simplicity. The kernel \( \delta(z_1, z_2, \gamma) \) is such that

\[
\delta(z_1, z_2, \gamma) = \Theta(z_1, \gamma)\Theta(z_2, \gamma)\left[\frac{z_1\Theta(z_1, \gamma) - z_2\Theta(z_2, \gamma)}{z_1 - z_2} - 1\right],
\]

\[
\delta(z, z, \gamma) = \lim_{z_2 \to z} \delta(z, z_2, \gamma) = \Theta^3(z, \gamma)\left[\frac{\partial \Theta(z, \gamma)}{\partial z} - 1\right]
\]

\[
= \gamma(1 + zm(z))\Theta(z, \gamma) + \gamma z(m(z) + zm'(z))\Theta^4(z, \gamma).
\]

The contour integral is taken counter-clockwise.

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

**Corollary 2.1** Under the conditions of Theorem 2.1, assume further that \( \Delta(f, \gamma) > 0 \). Let

\[
\hat{\lambda}_i = \frac{\sqrt{n}}{\Delta^{1/2}(f, \gamma)}\{\lambda_i(M(f)) - \Omega(f, \gamma)\}, \quad i = 1, \ldots, q.
\]

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

\[
\left(2^{q/2} \prod_{i=1}^q \Gamma(i/2)\right)^{-1} \prod_{i<j}(y_i - y_j)\exp\left(-\frac{1}{2} \sum_{i=1}^q y_i^2\right).
\]
Although without closed forms, \( \Omega(f, \gamma) \) and \( \Delta(f, \gamma) \) do not depend on the choice of \( C \) used to compute the contour integral. With the resolvent as kernel \( M(f) \) can be expressed as the integral of \( f(z)^{-1}Q_n^T Y (\bar{\Sigma}_p - zI_p)^{-1} Y Q_n \) on any contour \( C \), up to a scaling factor. The quadratic form \( n^{-1} Q_n^T Y (\bar{\Sigma}_p - zI_p)^{-1} Y Q_n \) is then shown to concentrate around \( [\Theta(z, \gamma) - 1]I_q \), which consequently serves as the integral kernel in \( \Omega(f, \gamma) \). The kernel \( \delta(z_1, z_2, \gamma) \) of \( \Delta(f, \gamma) \) is the limit of \( \mathbb{E}[n^{-1} \text{tr} (\bar{\Sigma}_p - z_1 I_p)^{-1} \Sigma_p (\bar{\Sigma}_p - z_2 I_p)^{-1} \Sigma_p] \).

Remark 2.2 Two sufficient conditions for \( \Delta(f, \gamma) > 0 \) are

1. \( f(x) > 0 \) for \( x \in \mathcal{X} \);
2. \( f(x) \geq 0 \) for \( x \in \mathcal{X} \), with \( f(x) \neq 0 \) for some \( x \in \mathcal{X} \), and \( \lim \inf \lambda_{\min}(\Sigma_p) > 0 \).

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 \( j, j_1, j_2 = 0, 1, 2, \ldots \),

\[
-\frac{1}{2\pi i} \int_C \frac{\partial^j f(x, \ell)}{\partial \ell^j} (\Theta(x, \gamma) - 1) dx = \frac{\partial^j (\Theta(\ell, \gamma) - 1)}{\partial \ell^j},
\]

\[
\frac{1}{(2\pi i)^2} \int_C \int_C \frac{\partial^{j_1} f(z_1, \ell_1)}{\partial \ell_1^{j_1}} \frac{\partial^{j_2} f(z_2, \ell_2)}{\partial \ell_2^{j_2}} \delta(z_1, z_2, \gamma) dz_1 dz_2 = \frac{\partial^{j_1+\ell_2} \delta(\ell_1, \ell_2, \gamma)}{\partial \ell_1^{j_1} \partial \ell_2^{j_2}}.
\]

The results continue to hold when \( \ell \in \mathbb{C} \setminus \mathcal{X} \).

Lemma 2.1 indicates that it is possible to have convenient and accurate estimators of the asymptotic mean and variance of \( M(f) \) 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 \( \ell_j \)'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 \( \hat{\Theta}(z, \gamma_n) \) and \( \hat{\delta}(z_1, z_2, \gamma_n) \) be the plug-in estimators of \( \Theta(z, \gamma) \) and \( \delta(z_1, z_2, \gamma) \), with \( (m_n(z), \gamma_n) \) estimated by \( (m_{n,f}(z), \gamma_n) \). For general \( f, f_1, f_2 \), we can estimate \( \Omega(f, \gamma) \) and \( \Delta(f_1, f_2, \gamma) \) by replacing \( \Theta(z, \gamma) \) and \( \delta(z_1, z_2, \gamma) \) with \( \hat{\Theta}(z, \gamma_n) \) and \( \hat{\delta}(z_1, z_2, \gamma_n) \). Denote the resulting estimators by \( \hat{\Omega}(f, \gamma_n) \) and \( \hat{\Delta}(f_1, f_2, \gamma_n) \). Then,

\[
\sqrt{n} | \hat{\Omega}(f, \gamma_n) - \Omega(f, \gamma) | \overset{P}{\rightarrow} 0,
\]

\[
\sqrt{n} | \hat{\Delta}(f_1, f_2, \gamma_n) - \Delta(f_1, f_2, \gamma) | \overset{P}{\rightarrow} 0,
\]

where \( \overset{P}{\rightarrow} \) indicates convergence in probability. Again, we write \( \hat{\Delta}(f, \gamma_n) \) as \( \hat{\Delta}(f, \gamma_n) \).

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

\[
\hat{\Omega}(f^{(j)}(x, \ell), \gamma_n) = \frac{\partial^j \hat{\Theta}(\ell, \gamma_n) - 1}{\partial \ell^j},
\]
In particular, for \(j, j_1, j_2 = 0\),
\[
\hat{\Omega}(f(x, \ell), \gamma_n) = \hat{\Theta}(\ell, \gamma_n) - 1,
\hat{\Delta}(f(x, \ell_1), f(x, \ell_2), \gamma_n) = 2\hat{\delta}(\ell_1, \ell_2, \gamma_n).
\]

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

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

**Theorem 2.2** Suppose \(C1-C6\) hold and \(\Delta(f, \gamma) > 0\). Under \(H_0: BC = 0\),
\[
\hat{T}^{LR}(f) := \frac{\sqrt{n}[1 + \hat{\Omega}(f, \gamma_n)]}{q^{1/2}\Delta^{1/2}(f, \gamma_n)} [T^{LR}(f) - q \log(1 + \hat{\Omega}(f, \gamma_n))] \Rightarrow \mathcal{N}(0, 1),
\]
\[
\hat{T}^{BH}(f) := \frac{\sqrt{n}}{q^{1/2}\Delta^{1/2}(f, \gamma_n)} \{T^{BH}(f) - q \hat{\Omega}(f, \gamma_n)\} \Rightarrow \mathcal{N}(0, 1),
\]
\[
\hat{T}^{BNP}(f) := \frac{\sqrt{n}[1 + \hat{\Omega}(f, \gamma_n)]^2}{q^{1/2}\Delta^{1/2}(f, \gamma_n)} \left\{T^{BNP}(f) - q \frac{\hat{\Omega}(f, \gamma_n)}{1 + \hat{\Omega}(f, \gamma_n)}\right\} \Rightarrow \mathcal{N}(0, 1).
\]

For any of the three tests, the null hypothesis is rejected at asymptotic level \(\alpha\), if \(\hat{T}(f) > \xi_\alpha\), where \(\xi_\alpha\) 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 \(\hat{T}^{LR}(f), \hat{T}^{BH}(f)\) and \(\hat{T}^{BNP}(f)\), the unifying notation \(\hat{T}(f)\) will be used to refer to each of the test statistics.

#### 2.3.1. Deterministic local alternatives

Consider a sequence of BC such that, on an open subset of \(\mathcal{C}\) containing \(\mathcal{X}\),
\[
\sqrt{n}C^T B^T \{\Theta^{-1}(z, \gamma)\Sigma_p - zI\}^{-1} BC \rightarrow D(z, \gamma) \quad \text{pointwise, as } n, p \rightarrow \infty.
\]

Observe that \(\mathbb{Y}Q_n = \sqrt{n}BC[C^T (n^{-1}XX^T)^{-1} C]^{-1/2} + \Sigma_p^{1/2}ZQ_n\) and define
\[
\mathcal{H}(D, f) = \int_{\mathcal{C}} f(z)D(z, \gamma)dz T^{-1/2}, \text{ where }
\]
(2.4)
\[ T = \lim_{n \to \infty} C^T (n^{-1}XX^T)^{-1}C. \] (2.5)

Note that \( T \) exists and is non-singular under \( C5 \) and \( C6 \). If further \( f(x) \geq 0 \) 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 \( n \to \infty \),

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

Denote the power functions of \( \hat{\mathcal{T}}(f) \) at asymptotic level \( \alpha \), conditional on \( BC \), by

\[ \Upsilon(BC, f) = \mathbb{P}(\hat{\mathcal{T}}(f) > \xi_\alpha \mid BC). \]

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

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

\[ \Upsilon(BC, f) \to \Phi(-\xi_\alpha + \frac{\text{tr}(\mathcal{H}(D, f))}{\sqrt{q^{1/2}\Delta^{1/2}(f, \gamma)}}), \]

where \( \Phi \) is the standard normal CDF.

**Remark 2.3** 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 \( [\mu_1, \ldots, \mu_q] \), it follows that

\[ \sqrt{n}C^T \mathcal{B}^{T} \{ \Theta^{-1}(z, \gamma) \Sigma_p - zI \}^{-1}BC = \sqrt{n} \left[ \mu^T \{ \Theta^{-1}(z, \gamma) \Sigma_p - zI_p \}^{-1} \mu \right]_{i,j=1}^q \]

**Remark 2.4** (a) Let \( \mathcal{E}_{m,p} \) denote the eigen-projection associated with \( \lambda_{m,p} = \lambda_m(\Sigma_p) \). Suppose that there exists a sequence (in \( p \)) of mappings \( \{ \mathcal{B}_{ij,p} \}_{i,j=1}^q \) from \( [0, \infty)^q \) to \( [0, \infty)^{2q} \), satisfying \( \mathcal{B}_{ij,p}(\lambda_{m,p}) = \sqrt{n} \mu^T \mathcal{E}_{m,p} \mu \), \( m = 1, \ldots, p \), and a mapping \( \{ \mathcal{B}_{ij,z} \}_{i,j=1}^q \) continuous on \( [0, \infty)^{2q} \) such that, as \( p \to \infty \) and for \( 1 \leq i, j \leq q \),

\[ \int |\mathcal{B}_{ij,p}(x) - \mathcal{B}_{ij,z}(x)|dF^{\Sigma_p}(x) \to 0. \]

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

\[ d_{ij}(z, \gamma) = \int \mathcal{B}_{ij,z}(x) dL^{\Sigma}(x) \] where

\[ d_{ij}(z, \gamma) = \int \mathcal{B}_{ij,z}(x) dL^{\Sigma}(x) \] where

\[ d_{ij}(z, \gamma) = \int \mathcal{B}_{ij,z}(x) dL^{\Sigma}(x) \]

(b) If \( \Sigma_p = I_p \), then (2.3) is satisfied if \( \sqrt{n} \mu^T \mu \to K_{ij} \), for some constants \( K_{ij} \), \( 1 \leq i, j \leq q \). In this case, \( D(z, \gamma) = (\Theta^{-1}(z, \gamma) - z)^{-1}[K_{ij}]_{i,j=1}^q \).
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

$$BC = n^{-1/4}p^{-1/2}\mathcal{R}\mathbf{V}\mathbf{S}^T,$$  \hspace{1cm} (2.6)

where $\mathbf{V}$ is a $p \times m$ stochastic matrix ($m \geq 1$ fixed) with independent elements $\nu_{ij}$ such that $E[\nu_{ij}] = 0$, $E[|\nu_{ij}|^2] = 1$ and $\max_{i,j} E[|\nu_{ij}|^4] \leq p^{c_2}$ for some $c_2 \in (0, 1)$; $\mathcal{R}$ is a $p \times p$ deterministic matrix and $\mathbf{S}$ is a fixed $q \times m$ matrix. Moreover, let $\|\mathcal{R}\|_2 \leq K_1 < \infty$ and suppose there is a nonrandom function $h(z, \gamma)$ such that, as $p \to \infty$, on an open subset of $\mathcal{C}$ containing $\mathcal{X}$,

$$p^{-1}\text{tr}\{(\Theta^{-1}(z, \gamma)\Sigma_p - zI)^{-1}\mathcal{R}\mathcal{R}^T\} \to h(z, \gamma) \quad \text{pointwise.}$$  \hspace{1cm} (2.7)

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

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 $\mathbf{S}$. Denote by $\mu_j$ the columns of $BC$ and by $V_j$ the columns of $\mathbf{V}$.

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

- (a) **Independent**: $\mu_j = n^{-1/4}p^{-1/2}\mathcal{R}V_j$;
- (b) **Longitudinal**: $\mu_j = n^{-1/4}p^{-1/2}\mathcal{R}(V_1 + V_2j + \cdots + V_mj^{m-1})$;
- (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, $\mu_j$ is the difference between the means of group $j$ and group $j + \text{Example 2.1}$
H. Li, A. Aue and D. Paul

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 \( \mu_j \) has a covariance matrix proportional to \( n^{-1/2}p^{-1}\mathbf{R}\mathbf{R}^T \). 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 2.4, is to postulate the existence of functions \( \tilde{B}_p \) satisfying
\[
\tilde{B}_p p \lambda_j,p q \text{tr} \mathbf{E}_{j,p} \mathbf{R}\mathbf{R}^T, j = 1, \ldots, p,
\]
and
\[
\hat{\tilde{B}}_p p \times \tilde{B}_8 p x q_{\tilde{E}} p p x q_{\Sigma} p x t 1 \gamma \text{tr} \mathbf{E}_{p } \mathbf{R}\mathbf{R}^T, z m \text{tr} p z qu z m \text{tr} p z.
\]
(2.8)

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

Another important feature of the probabilistic model is that it incorporates both dense and sparse alternatives through different specifications of the innovation variables \( \nu_{ij} \). 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 \( G_\eta \) 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, \( \mathbf{R} = \mathbf{I}_p \). More generally, the second specification above formulates a prior model for \( BC \) that is sparse in the coordinate system determined by \( \mathbf{R} \). In particular, if \( \mathbf{R}\mathbf{R}^T \) is a polynomial in \( \Sigma_p \) (see Section 3.2 for a discussion), \( BC \) can be seen as sparse in the spectral coordinates of \( \Sigma_p \).

**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
\[
Y(BC, f) \xrightarrow{L_1} \Phi \left( -\xi_\alpha + \frac{\text{tr}(SS^T T^{-1})}{q^{1/2} \Delta^{1/2}(f, \gamma)} \int \frac{-1}{2\pi i} f(z) h(z, \gamma) dz \right),
\]
(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 2.5 Even if the quantity $h_p(z, \gamma) = p^{-1} \text{tr}\{(\Theta^{-1}(z, \gamma)\Sigma_p - zI)^{-2}RR^T\}$ 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 $h(z, \gamma)$ is replaced by $h_p(z, \gamma)$.

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 $\Sigma_p$, besides $\Delta^{1/2}(f, \gamma)$, is only through the function $h(z, \nu)$, which incorporates the structure of the matrix $RR^T$. It is also noticeable that distributional characteristics of the variables $\nu_{ij}$ 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 $\mathcal{F}$ of shrinkage functions. The strategy is to maximize the local power function $\Upsilon(f, BC)$ 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 $\Sigma_p$ (and its interaction with $\mathcal{R}$) manifests itself through the function $h(z, \gamma)$. 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 $h(z, \gamma)$ 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

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

(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 $f_\ell$ with respect to $\ell$ are continuous on $\mathcal{L} \otimes \mathcal{Z}$;
(iii) The gradient $\nabla_\ell f_\ell$ and the Hessian $\nabla^2_\ell f_\ell$ of $f_\ell$ with respect to $\ell$ 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 \( h(z, \gamma) \) in (2.7) given, define

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

Theorem 2.4 suggests that \( \ell \) 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 \( S \). Since \( \Xi(\ell, h, \gamma) \) is continuous with respect to \( \ell \) under condition (i)–(iv), \( \ell_{opt} \) exists. Importantly, \( \Xi(\ell, h, \gamma) \) does not rely on \( S \). 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 \( h(z, \ell) \) is the limit of \( p^{-1} \text{tr}\{(\Theta^{-1}(z, \gamma) \Sigma_p - zI)^{-1} R R^T\} \). We next present two possible settings of \( R R^T \) under which \( h(z, \gamma) \) and consequently \( \Xi(\ell, h, \gamma) \) can be accurately estimated:

1. Suppose \( R R^T \) is specified. Then, \( h(z, \gamma) \) is estimated by \( \hat{h}(z, \gamma_n) = p^{-1} \text{tr}\{(\hat{\Sigma}_p - zI)^{-1} R R^T\} \) and

\[
\hat{\Xi}(\ell, \hat{h}, \gamma_n) := \frac{-1}{2\pi i \Delta^{1/2}(f_\ell, \gamma_n)} \oint_{\mathcal{C}} f_\ell(z) \hat{h}(z, \gamma_n) dz
\]

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

2. Only the spectral mass distribution of \( R R^T \) in the form of \( \mathfrak{B}_z \) 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 \( \mathfrak{B}_z \) is given, the estimation of \( h(z, \gamma) \) is still challenging since it involves the unknown limiting spectral distribution \( L^z \). In order to estimate \( h(z, \gamma) \), it is convenient to have it in a closed form. It is feasible if \( \mathfrak{B}_z \) is a polynomial, which is true if \( R R^T \) is a matrix polynomial in \( \Sigma_p \). Since any smooth function can be approximated by polynomials, this formulation is quite flexible and practically beneficial. Assume therefore that

\[
R R^T = \sum_{j=0}^{\gamma} t_j \Sigma_p^j,
\]

(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,

$$
\hat{h}(z, \gamma) = \lim_{p \to \infty} p^{-1} \mathrm{tr} \left[(\Theta^{-1}(z, \gamma) \Sigma_p - zI)^{-1} \sum_{j=0}^s t_j \Sigma_p^j \right] = \sum_{j=0}^s t_j \rho_j(z, \gamma),
$$

where the functions $\rho_j(z, \gamma)$ satisfy the recursive formula (see Ledoit and Pêché, 2011)

$$
\rho_0(z, \gamma) = m(z), \quad \rho_{j+1}(z, \gamma) = \Theta(z, \gamma) \left[ \int x^d dL^\Sigma(x) + z \rho_j(z, \gamma) \right].
$$

For any $j \in \mathbb{N}$, $\int x^d dL^\Sigma(x)$, and consequently $\rho_j(z, \gamma)$, can be estimated consistently (Bai, Chen and Yao, 2010, Lemma 1). Specifically, $p^{-1} \mathrm{tr}(\hat{\Sigma}_p)$ 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 $h(z, \gamma)$ 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 $\rho_j$, $j = 1, \ldots, q$, (that is $\mathcal{R} \mathcal{R}^T$ up to a scaling factor) in the directions of the leading eigenvectors of $\Sigma_p$.

Finally, the choice $s = 2$ allows for both convex and concave shapes of the spectral mass distribution $\Phi_x$ since the latter becomes a quadratic function.

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

$$
\hat{\rho}_0(z, \gamma_n) = m_{n,p}(z), \\
\hat{\rho}_1(z, \gamma_n) = \hat{\Theta}(z, \gamma_n)[1 + zm_{n,p}(z)], \\
\hat{\rho}_2(z, \gamma_n) = \hat{\Theta}(z, \gamma_n) \left[ p^{-1} \mathrm{tr}(\hat{\Sigma}_p) + z \hat{\rho}_1(z, \gamma_n) \right],
$$

$$
\hat{h}(z, \gamma_n) = \sum_{j=0}^2 t_j \hat{\rho}_j(z, \gamma_n). \quad (3.2)
$$

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

**Algorithm 3.1 (Data-driven shrinkage selection)**

1. Specify prior weights $\hat{\ell} = (t_0, t_1, t_2)$. The canonical choices are $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$.
2. Compute $\hat{h}(z, \gamma_n) = \sum_{j=0}^2 t_j \hat{\rho}_j(z, \gamma_n)$;
3. For any $\ell \in \mathcal{L}$, numerically compute the integral

$$
\hat{\mathcal{E}}(\ell, \hat{h}, \gamma_n) = \frac{-1}{2\pi i \Delta^{1/2}(f_\ell, \gamma_n)} \oint_{\mathcal{C}} f_\ell(z) \hat{h}(z, \gamma_n) dz;
$$

4. Select $\ell_{\text{opt}}(\hat{\ell}) = \arg\max_{\ell \in \mathcal{L}} \hat{\mathcal{E}}(\ell, \hat{h}, \gamma_n)$.

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 $\mathcal{F}$ satisfies conditions (i)-(iv). Then,

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

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

$$\Xi(\ell, h, \gamma) - \Xi(\ell^*, h, \gamma) \leq -K\|\ell - \ell^*\|^2.$$  \hfill (3.3)

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

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

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

$$\sqrt{n} \Im\{ \frac{\mathbf{M}(f_{\ell^*_n}) - \hat{\Omega}(f_{\ell^*_n}, \gamma_n)I_q}{\Delta^{1/2}(f_{\ell^*_n}, \gamma_n)} \} \to W.$$ \hfill (3.5)

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

$$\Xi(\ell, h, \gamma) - \Xi(\ell^*, h, \gamma) \leq -K'\|\ell - \ell^*\|^2.$$ \hfill (3.6)

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

The two conditions (3.3) and (3.6) ensure that the parameter $\ell^*$ is locally identifiable in a neighborhood of $\ell^*$. In general, the two conditions depend on the structure of $L^\Sigma$.

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 $\Sigma$ is relatively large. In this subsection, we consider a convenient collection of priors that are representative of certain structural scenarios. A composite test, called $\tilde{T}_{\text{max}}$, is defined as the maximum of the standardized statistics $\tilde{T}(f_{\ell^*_i})$ where $\ell^*_i$ 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 $\tilde{T}(f)$ to denote the general test statistic. In summary, we propose to test the hypothesis by rejecting for large values of the statistic

$$\tilde{T}_{\text{max}} = \max_{i \in \Pi} \tilde{T}(f_{\ell^*_i}),$$

where $\Pi = \{\tilde{t}_1, \ldots, \tilde{t}_m\}$, $m \geq 1$, is a pre-specified finite class of weights. A simple but effective choice of $\Pi$ consists of the three canonical weights $\tilde{t}_1 = (1, 0, 0)$, $\tilde{t}_2 = (0, 1, 0)$, $\tilde{t}_3 = (0, 0, 1)$. 
where \( \Delta \) (See (3.4)). Then, under the null hypothesis \( H_0 \) a parametric bootstrap procedure is applied to approximate the cut-off values. Specifically, for the quantiles of the limiting distribution. Aiming to avoid complex calculations, a theorem can be used to determine the cut-off values of the test by deriving analytical formulas of the test.

\( \hat{\Delta} \) is worth mentioning that LR, LH and BNP share the covariance matrix \( \Delta \) from \( \mu \). Suppose \( \text{Theorem 3.2} \)

| \( n = 300, p = \) | \( k = 3 \) | \( k = 5 \) |
|-----------------|---------|---------|
| \( \Sigma = I_p \) | \( \Sigma = \Sigma_{\text{den}} \) |
| LRridge | \( t_1 \) | 5.4 | 5.2 |
| LHridge | \( t_1 \) | 5.4 | 5.2 |
| BNPridge | \( t_2 \) | 5.4 | 5.2 |
| LRhigh | \( t_2 \) | 6.5 | 6.3 |
| LHhigh | \( t_2 \) | 6.7 | 6.4 |
| BNPhigh | \( t_2 \) | 6.3 | 6.3 |
| LRcomp | 5.1 | 5.1 | 5.0 |
| LHcomp | 5.1 | 5.1 | 5.5 |
| BNPcomp | 5.1 | 5.0 | 5.4 |
| ZGZ | 5.6 | 5.7 | 5.2 |
| CX (Oracle) | 5.6 | 6.3 | 7.0 |

Table 1. Empirical sizes at level 5%. \( \Sigma = I_p \) and \( \Sigma_{\text{den}} \); \( t_1 = (1, 0, 0, 1, 0) \), \( t_2 = (0, 1, 0) \), \( t_3 = (0, 0, 1) \).

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

\[ n^{1/4} \| \ell^*_i - \ell^*_i \|_2 = O_p(1). \]

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

\[ (\hat{T}(f_{\ell^*_i}), \ldots, \hat{T}(f_{\ell^*_m})) \Rightarrow \mathcal{N}(0, \Delta^*_a), \]

where \( \Delta^*_a \) 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 \( \hat{T}_{\text{max}} \) has a non-degenerate limiting distribution under \( H_0 \). It is worth mentioning that LR, LH and BNP share the covariance matrix \( \Delta^*_a \). 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, \( \Delta^*_a \) is first estimated by \( \hat{\Delta}^*_a \), and then bootstrap replicates are generated by simulating from \( \mathcal{N}(0, \Delta^*_a) \), thereby providing an approximation of the null distribution of \( \hat{T}_{\text{max}} \). Replacing \( \Delta(f_{\ell^*_i}, f_{\ell^*_j}, \gamma) \) with \( \hat{\Delta}(f_{\ell^*_i}, f_{\ell^*_j}, \gamma_n) \) yields the natural estimator.
Remark 3.1 Observe that \( \hat{\Delta}_a \) 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 \( \hat{\Delta}^+_a \) 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 \( f_{\ell}(\hat{\Sigma}_p) = (\hat{\Sigma}_p - \ell I_p)^{-1} \). It is an effective way to shift \( \hat{\Sigma}_p \) 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 \( \mathcal{F}_{\text{ridge}} \):

\[
\mathcal{F}_{\text{ridge}} := \{ f_{\ell}(x) = (x - \ell)^{-1}, \ell \in [\ell, \bar{\ell}] \}, \quad -\infty < \ell < \bar{\ell} < 0.
\]

In the literature, ridge-regularization was applied to high-dimensional one- and two-sample mean tests in Chen et al. (2011) 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 \( \Sigma_p \) is estimated by a weighted average of \( \hat{\Sigma}_p \) and \( I_p \), namely \( \alpha_0 I_p + \alpha_1 \hat{\Sigma}_p \). The estimator is equivalent to ridge-regularization with \( \ell = -\alpha_0/\alpha_1 \) for testing purposes. Within a restricted region of \( (\alpha_1, \alpha_2) \), the large eigenvalues of \( \hat{\Sigma}_p \) 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 \( \Omega_{\ell}(\gamma_n) = \hat{\Theta}(\ell, \gamma_n) - 1 \) and \( \hat{\Delta}_{\ell}(\gamma_n) = 2\hat{\delta}(\ell, \ell, \gamma_n) \), respectively. A closed-form estimator \( \hat{\Xi}_{\ell}(h, \gamma_n) \) 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 \( \hat{\ell} = (t_0, t_1, t_2) \);
2. With \( m_{n,p}(\ell) = p^{-1} tr(\hat{\Sigma}_p - \ell I_p)^{-1} \), compute, for any \( \ell \in [\ell, \bar{\ell}] \),
   \[
   \hat{\Theta}(\ell, \gamma_n) = \{ 1 - \gamma_n - \gamma_n \ell m_{n,p}(\ell) \}^{-1},
   \hat{\Omega}_{\ell}(\gamma_n) = \hat{\Theta}(\ell, \gamma_n) - 1,
   \hat{\Delta}_{\ell}(\gamma_n) = 2\gamma_n \{ 1 + \ell m_{n,p}(\ell) \} \hat{\Theta}^3(\ell, \gamma_n) + 2\gamma_n \ell \{ m_{n,p}(\ell) + \ell m'_{n,p}(\ell) \} \hat{\Theta}^4(\ell, \gamma_n);
   \]
3. For any $\ell \in [\ell, \overline{\ell}]$, compute $\hat{h}(\ell, \gamma_n) = \sum_{j=0}^{2} t_j \hat{p}_j(\ell, \gamma_n)$ as defined in (3.2) and
\[
\hat{\xi}_\ell(\hat{h}, \gamma_n) = \frac{\hat{h}(\ell, \gamma_n)}{\Delta^{1/2}_\ell(\gamma_n)};
\]
4. Select $\ell^* = \arg\max_{\ell \in [\ell, \overline{\ell}]} \hat{\xi}_\ell(\hat{h}, \gamma_n)$;
5. Use one of the standardized statistics
\[
\hat{T}^{LR}(\ell^*) := \frac{\sqrt{n}(1 + \hat{\Omega}_{\ell^*}(\gamma_n))}{q^{1/2} \Delta^{1/2}_{\ell^*}(\gamma_n)} \left[ T^{LR}(\ell^*) - q \log\{1 + \hat{\Omega}_{\ell^*}(\gamma_n)\} \right],
\]
\[
\hat{T}^{LH}(\ell^*) := \frac{\sqrt{n}}{q^{1/2} \Delta^{1/2}_{\ell^*}(\gamma_n)} \left[ T^{LH}(\ell^*) - q \hat{\Omega}_{\ell^*}(\gamma_n) \right],
\]
\[
\hat{T}^{BNP}(\ell^*) := \frac{\sqrt{n}(1 + \hat{\Omega}_{\ell^*}(\gamma_n))^2}{q^{1/2} \Delta^{1/2}_{\ell^*}(\gamma_n)} \left[ T^{BNP}(\ell^*) - \frac{q \hat{\Omega}_{\ell^*}(\gamma_n)}{1 + \hat{\Omega}_{\ell^*}(\gamma_n)} \right],
\]
where
\[
T^{LR}(\ell^*) = \sum_{i=1}^{q} \log(1 + \lambda_i), \quad T^{LH}(\ell^*) = \sum_{i=1}^{q} \lambda_i, \quad T^{BNP}(\ell^*) = \sum_{i=1}^{q} \frac{\lambda_i}{1 + \lambda_i},
\]
and $\lambda_1, \ldots, \lambda_q$ are the eigenvalues of $n^{-1} Q_n^T Y^T (\hat{\Sigma}_p - \ell^* I_p)^{-1} Y Q_n$. Reject the null at asymptotic level $\alpha$ if the test statistic value exceeds $\xi_{\alpha}$.

Although in theory any negative $\ell^*$ is allowed in the test procedure, in practice, meaningful lower and upper bounds $\ell$ 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 $\ell$ at a low computational cost. In our simulation settings we use $\overline{\ell} = -p^{-1} \mathrm{tr}(\hat{\Sigma}_p)/100$ and $\ell = -20 \lambda_{\max}(\hat{\Sigma}_p)$, which generally lead to quite robust performance.

The composite test procedure with ridge-regularization is summarized below.

**Algorithm 4.2 (Composite ridge-regularized test procedure)**

1. Select prior weights $\tilde{\Pi} = (\tilde{t}_1, \ldots, \tilde{t}_m)$. The canonical choice is $((1, 0, 0), (0, 1, 0), (0, 0, 1))$;
2. For each $\tilde{t}_j$ in $\tilde{\Pi}$, run Algorithm 4.1, get the standardized test statistic $\hat{T}(\ell^*_j)$ and compute $\hat{T}_{\max} = \max_{1 \leq j \leq m} \hat{T}(\ell^*_j)$;
3. With the selected tuning parameters ($\ell^*_1, \ell^*_2, \ell^*_3$) compute the matrix $\hat{\Delta}_{\ell}$ whose diagonal elements are equal to one and whose $(i, j)$-th entry for $i \neq j$ is
\[
\hat{\Delta}^{-1/2}_{\ell^*_i}(\gamma_n) \hat{\Delta}_{\ell^*_i, \ell^*_j}(\gamma_n) \hat{\Delta}^{-1/2}_{\ell^*_j}(\gamma_n),
\]
where $\hat{\Delta}_{\ell^*_i}(\gamma_n)$ is defined in Step 2 of Algorithm 4.1 and
\[
\hat{\Delta}_{\ell^*_i, \ell^*_j}(\gamma_n) = 2 \hat{\Theta}(\ell^*_i, \gamma_n) \hat{\Theta}(\ell^*_j, \gamma_n) \left[ \frac{\ell^*_i \hat{\Theta}(\ell^*_i, \gamma_n) - \ell^*_j \hat{\Theta}(\ell^*_j, \gamma_n)}{\ell^*_i - \ell^*_j} - 1 \right].
\]
4. Project $\hat{\Delta}_b$ to its closest non-negative definite matrix $\hat{\Delta}_b^+$ by setting the negative eigenvalues to zero. Generate $\varepsilon_1, \ldots, \varepsilon_G$ with $\varepsilon_b = \max_{1 \leq i \leq m} Z_i^{(b)}$ with $Z_i^{(b)} = [Z_i^{(b)}]_{i=1}^m \sim \mathcal{N}(0, \hat{\Delta}_b^+)$. 

5. Compute the $p$-value as $G^{-1} \sum_{b=1}^G 1(\varepsilon_b > \hat{T}_{\max})$.

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 $\hat{\Delta}_b$ by $\ell I$ instead of $\hat{\Sigma}_p^{-1}$. 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 $f(\hat{\Sigma}_p)$. 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

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

where $\mathcal{G}$ is such that $f_\ell$ 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 $r_1, \ldots, r_\kappa \in \mathbb{C} \setminus \mathcal{X}$ with multiplicity $s_1, \ldots, s_\kappa \in \mathbb{N}^+$. Via basic algebra, $f_\ell$ can be expressed as

$$f_\ell(x) = \left( \sum_{j=0}^{\kappa} l_j x^j \right)^{-1} = \sum_{i=1}^{s_\kappa} w_{ji} (x - r_j)^{-i}, \quad (4.1)$$

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

The design of $\mathcal{G}$ is not easy when $\kappa$ is large. Here, we select $\kappa = 3$, which is the minimum degree that allows $f_\ell^{-1}$ 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 $\Sigma_p^{-1}$, that is typically unavailable when both $\Sigma_p$ and $\Sigma_p^{-1}$ are dense. In the simulations, the true $\Sigma_p^{-1}$ is utilized for CX, thus making it an oracle procedure. In the following, LR$_{\text{ridge}}$, LH$_{\text{ridge}}$, and BNP$_{\text{ridge}}$ denote the ridge-regularized tests presented in Algorithm 4.1. LR$_{\text{high}}$, LH$_{\text{high}}$, and BNP$_{\text{high}}$ denote the tests with higher-order shrinkage introduced in Section 4.2 with $\kappa = 3$. LR$_{\text{comp}}$, LH$_{\text{comp}}$ and BNP$_{\text{comp}}$ denote the composite ridge-regularized tests of Algorithm 4.2 with the canonical choice of $\hat{\Pi} = ((1, 0, 0), (0, 1, 0), (0, 0, 1))$.

5.1. Settings

The observation matrix $Y$ was generated as in (1.1) with normally distributed $Z$. 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.

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

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 alternative: $B = cR\mathbf{V}$ with $c = O(n^{-1/4}p^{-1/2})$, where $R$ 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 $\mathbf{V}$ is a $p \times p$ matrix with i.i.d. standard normal entries.

| $n = 300$, $p = 150$ | $k = 3$ | $\Sigma = \Sigma_{dis}$ | $\Sigma = \Sigma_{toep}$ |
|-----------------------|--------|-----------------|-----------------|
| 600 3000 150 | 150 600 3000 | 150 600 3000 | 150 600 3000 | 150 600 3000 |
| $\text{LR}_{\text{ridge}}$ | $t_1$ 4.8 | 5.0 | 4.6 | 4.7 | 4.5 | 5.0 | 5.4 | 4.4 | 4.8 | 4.5 | 4.6 | 4.6 |
| | $t_2$ 5.1 | 5.2 | 4.9 | 5.2 | 4.6 | 5.1 | 5.4 | 4.9 | 4.9 | 4.9 | 4.8 | 5.0 |
| | $t_3$ 5.6 | 5.5 | 5.1 | 5.7 | 5.3 | 5.3 | 5.8 | 5.2 | 5.0 | 5.7 | 5.4 | 5.1 |
| | $t_1$ 5.8 | 6.0 | 5.2 | 6.6 | 6.3 | 5.6 | 6.4 | 5.3 | 5.2 | 6.2 | 6.3 | 5.3 |
| | $t_2$ 5.7 | 5.7 | 5.1 | 6.3 | 5.6 | 5.5 | 5.9 | 5.3 | 5.0 | 5.8 | 5.6 | 5.3 |
| | $t_3$ 5.6 | 5.5 | 5.2 | 5.8 | 5.3 | 5.4 | 5.8 | 5.3 | 5.1 | 5.7 | 5.4 | 5.2 |
| $\text{LH}_{\text{ridge}}$ | $t_1$ 3.9 | 4.1 | 4.3 | 3.1 | 3.1 | 4.1 | 4.4 | 3.7 | 4.4 | 3.2 | 3.4 | 3.9 |
| | $t_2$ 4.6 | 4.8 | 4.8 | 4.1 | 4.0 | 4.9 | 4.9 | 4.4 | 4.8 | 4.1 | 4.3 | 4.7 |
| | $t_3$ 5.5 | 5.5 | 5.0 | 5.7 | 5.2 | 5.1 | 5.8 | 5.2 | 5.0 | 5.6 | 5.4 | 5.1 |
| $\text{BNP}_{\text{ridge}}$ | $t_1$ 6.3 | 6.4 | 4.8 | 5.9 | 7.0 | 5.5 | 7.1 | 7.0 | 5.3 | 7.5 | 6.9 | 5.2 |
| | $t_2$ 7.9 | 6.5 | 4.8 | 8.3 | 7.1 | 5.5 | 7.6 | 7.2 | 5.3 | 7.8 | 7.0 | 5.2 |
| | $t_3$ 6.1 | 5.6 | 4.8 | 6.4 | 6.1 | 5.5 | 6.7 | 6.5 | 5.3 | 6.6 | 6.4 | 5.2 |
| | $t_1$ 6.6 | 6.5 | 5.0 | 6.2 | 7.2 | 5.7 | 7.2 | 7.2 | 5.5 | 7.7 | 7.0 | 5.5 |
| | $t_2$ 8.0 | 6.6 | 5.0 | 8.5 | 7.2 | 5.7 | 7.8 | 7.2 | 5.5 | 8.0 | 7.1 | 5.5 |
| | $t_3$ 6.2 | 5.6 | 5.0 | 6.5 | 6.2 | 5.7 | 6.7 | 6.5 | 5.5 | 6.7 | 6.5 | 5.5 |
| $\text{BNP}_{\text{high}}$ | $t_1$ 6.1 | 6.3 | 4.7 | 5.6 | 6.8 | 5.3 | 7.1 | 7.0 | 5.2 | 7.2 | 6.8 | 5.1 |
| | $t_2$ 7.9 | 6.4 | 4.7 | 8.2 | 7.0 | 5.3 | 7.5 | 7.1 | 5.2 | 7.7 | 7.0 | 5.1 |
| | $t_3$ 6.1 | 5.5 | 4.7 | 6.4 | 6.0 | 5.3 | 6.6 | 6.4 | 5.2 | 6.5 | 6.3 | 5.1 |
| $\text{LR}_{\text{comp}}$ | 6.2 | 5.2 | 5.0 | 5.2 | 5.3 | 5.5 | 5.9 | 5.0 | 5.1 | 5.5 | 4.9 | 4.9 |
| $\text{LH}_{\text{comp}}$ | 7.0 | 5.9 | 5.3 | 6.5 | 6.4 | 6.0 | 6.6 | 5.6 | 5.3 | 6.6 | 5.7 | 5.3 |
| $\text{BNP}_{\text{comp}}$ | 5.5 | 4.6 | 4.8 | 4.4 | 4.6 | 5.0 | 5.4 | 4.6 | 4.9 | 4.8 | 4.4 | 4.6 |
| $\text{ZGZ}$ | 5.5 | 4.7 | 4.6 | 5.1 | 5.1 | 5.3 | 6.0 | 5.5 | 5.0 | 5.9 | 5.6 | 5.0 |
| $\text{CX}$ (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 2. Empirical sizes at level $5\%$. $\Sigma = \Sigma_{dis}$ and $\Sigma_{toep}$: $t_1 = (1, 0, 0)$, $t_2 = (0, 1, 0)$, $t_3 = (0, 0, 1)$.

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

(i) Identity matrix (ID): $\Sigma = I_p$.

(ii) Dense case $\Sigma_{den}$: 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 $\lambda_j = (0.1 + j)^0 + 0.05p^0$, $j = 1, \ldots, p$.

The eigenvalues of $\Sigma$ decay slowly, so that no dominating leading eigenvalue exists.

(iii) Toeplitz case $\Sigma_{toep}$: Here $\Sigma$ is a Toeplitz matrix with the $(i, j)$-th element equal to $0.5^{i-j}$. It is a setting where $\Sigma^{-1}$ is sparse but $\Sigma$ is dense.

(iv) Discrete case $\Sigma_{dis}$: 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 1 and 2. Empirical power curves versus expected signal strength
High-dimensional general linear hypothesis via spectral shrinkage

$n^{1/4}p^{1/2}c$ are reported in Figures 1–3. 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 1–3 that utilize asymptotic (approximate) cut-off values are reported in Section S.12 of the Supplementary Material. 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.11 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.

5.2. Summary of simulation results

Tables 1 and 2 show the empirical sizes of the proposed tests are mostly controlled under 7.5%. The slight oversize is caused by the fact that $M(f)$ 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 $M(f)$. 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.11.6 in the Supplementary Material. We highlight the following observations based on the simulation results.

![Figure 2: Size-adjusted power with $\Sigma = \Sigma_{d\text{en}}$, $k = 5$. Rows (top to bottom): $B = $ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. LH$_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); LH$_{\text{ridge}}$ (black, dashed) and LH$_{\text{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$.](image-url)
The composite tests are slightly less efficient than BNP_{ridge} and BNP_{high} when the correct prior $\tilde{t}_1$ is used, as in Figure 1. However, as in Figure 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.

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

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 2. This provides evidence for the robustness of high-order shrinkage against unfavorable ridge shrinkage parameter selection.

ZGZ is a special case of the proposed test family with $f(x) = 1$ for all $x$, which amounts to replacing $\tilde{\Sigma}_p$ with $I_p$. When $\Sigma_p = I_p$, ZGZ appears to be the reasonable option at least intuitively. Note, both $\tilde{\mathcal{F}}_{\text{ridge}}$ and $\tilde{\mathcal{F}}_{\text{high}}$ contain functions close to $f(x) = 1$. Figures for $\Sigma_p = I_p$ displayed in Section S.11 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.

Comparing to ZGZ, when the eigenvalues of $\Sigma_p$ 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 2, the ridge-regularized test with a severely misspecified prior $\tilde{t}_3$, is close to ZGZ.

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 $\Sigma_p^{-1}$, 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.11.10, S.11.11 and S.11.12 of the Supplementary Material.
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.

Bodnar, T., Dette, H. and Parolya, N. (2017). Testing for independence of large dimensional vectors. arXiv preprint arXiv:1708.03964.

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 Kôsters, H. (2011). Geometric sensitivity of random matrix re-
sults: 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.

Zheng, S. (2012). Central limit theorems for linear spectral statistics of large dimensional
Appendix: Proof of Theorem 2.1

This appendix contains a proof outline of Theorem 2.1. Additional proofs of supporting lemmas and other theorems can be found in the Supplementary Material.

Recall $Q_n = X^T (XX^T)^{-1} C [C^T (XX^T)^{-1} C]^{-1/2}$. Introduce $Q_n = U_n V_n$ with

$$U_n = X^T (XX^T)^{-1/2} \quad \text{and} \quad V_n = (XX^T)^{-1/2} C [C^T (XX^T)^{-1} C]^{-1/2}.$$

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

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

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

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

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

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

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

Recall that $X = [0, \limsup_p \lambda_{\text{max}}(\Sigma_p)(1 + \sqrt{q})^2]$. Let $C$ be any contour enclosing $X$ such that $f(\cdot)$ is analytic on its interior. With slight modifications, all arguments in the following hold for arbitrary such $C$. For convenience, select $C$ as rectangle with vertices $u \pm i\nu_0$ and $\overline{u} \pm i\nu_0$, such that $\nu_0 > 0$; $\overline{u} > \limsup \lambda_{\text{max}}(\Sigma_p)(1 + \sqrt{q})^2$; $u < 0$. Such a rectangle must exist.

By Cauchy’s integral formula, if $\lambda_{\text{max}}(\hat{\Sigma}_p) < \overline{u}$,

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

$$= -\frac{1}{2\pi i} \oint_C f(z) n^{-1/2} \alpha^T V_n^T U_n^T Z^T \Sigma_p^{T/2} (\hat{\Sigma}_p - zI)^{-1} \Sigma_p^{1/2} Z U_n V_n \eta dz. \quad (A.1)$$

If $\lambda_{\text{max}}(\hat{\Sigma}_p) \geq \overline{u}$, the above equality may not hold. However, if we can show that $\mathbb{P}(\lambda_{\text{max}}(\hat{\Sigma}_p) \geq \overline{u})$ 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
\[ \mathbb{P}(\lambda_{\max}(\hat{\Sigma}_p) \geq \|w\|) \to 0. \]  
(A.2)

Hence, it suffices to show the asymptotic normality of the process
\[ \xi_n(z, \alpha, \eta) = n^{-1/2} \alpha^T V_n^T U_n^T Z^T \Sigma_p^{T/2} (\hat{\Sigma}_p - zI)^{-1} \Sigma_p^{1/2} ZU_n V_n \eta, \quad z \in \mathbb{C}. \]

Clearly, \( \xi(z, \alpha, \eta) \) is continuous with respect to \( z \). All asymptotic results are derived in the space of continuous functions on \( \mathbb{C} \) with uniform topology. Results in Chapter 2 of Billingsley (1968) apply with Euclidean distance replaced by Frobenius norm of a matrix, that is \( \|A\|_{F} = (\sum_{i=1}^{n} \sum_{j=1}^{r} |a_{ij}|^2)^{1/2} \), where \( A = [a_{ij}]_{ij} \).

We may proceed to prove the asymptotic normality of \( \xi_n(z, \alpha, \eta) \) on \( z \in \mathbb{C} \) directly. However, several technical challenges need to be addressed. First, in view of the spectral nature of \( (\hat{\Sigma}_p - zI)^{-1} \) being unbounded when \( z \) is close to the real axis and extreme eigenvalues of \( \hat{\Sigma}_p \) exceed \( \lambda_{\max}(\Sigma_p)(1 + \sqrt{2})^2 \), the tightness of the process \( \xi_n(z, \alpha, \eta) \) is unclear. Secondly, \( \hat{\Sigma}_p \) is not a summation of independent terms, but contains \( ZU_n U_n^T Z^T \), a component containing cross product terms between pairs of columns of \( Z \). These terms entangle the analysis of the correlation between \( \hat{\Sigma}_p \) and each single column of \( Z \). For these technical reasons, we avoid directly working on \( \xi_n(z, \alpha, \eta) \) under \( C1 \) on \( z \in \mathbb{C} \), but start with \( n^{-1/2} U_n^T Z^T \Sigma_p^{T/2} (\hat{\Sigma}_p - zI)^{-1} \Sigma_p^{1/2} ZU_n \), a component of \( \xi_n(z, \alpha, \eta) \) with \( \hat{\Sigma}_p \) replaced by an uncentered counterpart
\[ \tilde{\Sigma}_p = \frac{1}{n} \Sigma_p^{1/2} ZZ^T \Sigma_p^{T/2}. \]  
(A.3)

The relationship between \( \tilde{\Sigma}_p \) and \( \hat{\Sigma}_p \) is given by \( \hat{\Sigma}_p = \tilde{\Sigma}_p - \frac{1}{n} \Sigma_p^{1/2} ZU_n U_n^T Z^T \Sigma_p^{T/2} \). Next, we modify the process and the distribution of \( Z \) as follows.

Process smoothing. Select a sequence of \( \rho_n > 0 \) such that for some \( \omega \in (1, 2) \)
\[ n \rho_n \downarrow 0, \quad \rho_n \geq n^{-\omega}. \]

Let \( C^+ = \mathbb{C} \cap \{u + iv : |v| \geq \rho_n\} \). Define
\[ \tilde{\Omega}_n(z) = n^{-1} U_n^T Z^T \Sigma_p^{T/2} (\hat{\Sigma}_p - zI)^{-1} \Sigma_p^{1/2} ZU_n, \quad \text{if } z \in \mathbb{C}^+, \]
\[ \tilde{\Omega}_n(z) = \frac{\rho_n - v}{2 \rho_n} \tilde{\Omega}_n(u + i \rho_n) + \frac{v + \rho_n}{2 \rho_n} \tilde{\Omega}_n(u - i \rho_n), \quad \text{if } z \in \mathbb{C} \setminus C^+. \]

To understand this definition better, note that if \( z \) is too close to the real axis, \( \tilde{\Omega}_n(z) \) 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 \( \xi_n(z, \alpha, \eta) \) was left out when defining \( \tilde{\Omega}_n(z) \). This trick helps transforming back to \( \hat{\Sigma}_p \) from \( \tilde{\Sigma}_p \); see (A.5). 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 \( (\hat{\Sigma}_p - zI_p)^{-1} \). 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 \( \varepsilon_n \) such that
\[
\varepsilon_n \to 0 \quad \text{and} \quad \varepsilon_n^{-4} \mathbb{E} [z_{11}^4 \mathbb{1}(|z_{11}| \geq \varepsilon_n n^{1/2})] \to 0.
\]
The existence of \( \varepsilon_n \) is shown in Yin, Bai and Krishnaiah (1988). We then truncate \( z_{ij} \)
to be \( z_{ij} \mathbb{1}(|z_{ij}| \leq \varepsilon_n n^{1/2}) \). The truncated variable is then standardized 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 \( \hat{z}_{ij} \) to denote the original random variable satisfying C1. That is,
\[
z_{ij} = \frac{\hat{z}_{ij} \mathbb{1}(\hat{z}_{ij} \leq \varepsilon_n n^{1/2}) - \mathbb{E}[\hat{z}_{ij} \mathbb{1}(\hat{z}_{ij} \leq \varepsilon_n n^{1/2})]}{\{\mathbb{E}[\hat{z}_{ij} \mathbb{1}(\hat{z}_{ij} \leq \varepsilon_n n^{1/2}) - \mathbb{E}[\hat{z}_{ij} \mathbb{1}(\hat{z}_{ij} \leq \varepsilon_n n^{1/2})]^{2}\}^{1/2}}.
\]

For some constant \( K \), when \( n \) is sufficiently large,
\[
|z_{ij}| \leq K \varepsilon_n n^{1/2}, \quad \mathbb{E}[z_{ij}] = 0, \quad \mathbb{E}[z_{ij}^2] = 1, \quad \mathbb{E}[z_{ij}^4] < \infty. \tag{A.4}
\]

The reason to truncate \( \hat{z}_{ij} \) is to obtain a bound on the probability of extreme eigenvalues of \( \hat{\Sigma}_p \) exceeding \( \limsup_p \lambda_{\max}(\Sigma_p)(1 + \sqrt{\gamma})^2 \), which is crucial to be able to prove tightness of the smoothed random processes on \( C \). Under the original condition C1, although (A.2) holds, such a tail bound is not available. After the truncation, the following lemma shown in Yin, Bai and Krishnaiah (1988); Bai and Silverstein (2004) holds.

**Lemma A.1** Suppose the entries of \( Z \) satisfy (A.4). For any positive \( \ell \) and any \( D \in (\limsup_p \lambda_{\max}(\Sigma_p)(1 + \sqrt{\gamma})^2, \infty) \),
\[
P(\lambda_{\max}(\hat{\Sigma}_p) \geq D) = o(n^{-\ell}).
\]

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

**Theorem A.1** For arbitrary vectors \( a \) and \( b \in \mathbb{R}^k \), define \( G_n(z, a, b) = a^T \hat{Q}_n(z)b \). Suppose \( Z \) satisfies (A.4) and suppose C2–C6 in Section 2 hold. Then,
\[
n^{1/2}\{G_n(z, a, b) - a^T b (1 - \Theta^{-1}(z, \gamma))\} \xrightarrow{D} \Psi^{(1)}(z), \quad z \in C,
\]
where \( \xrightarrow{D} \) denotes weak convergence in \( C(C, \mathbb{R}^2) \), and \( \Psi^{(1)}(z) \) is a Gaussian process with zero mean and covariance function
\[
\Gamma^{(1)}(z_1, z_2) = \delta(z_1, z_2, \gamma)\Theta^{-2}(z_1, \gamma)\Theta^{-2}(z_2, \gamma)[|a|^2||b|^2 + (a^T b)^2].
\]
See Section S.3 of the Supplementary Material for proof of the theorem.

Next, transforming back to \( \hat{\Sigma}_p \), define
\[
\hat{Q}_n(z) = n^{-1}U_n^T Z^T \Sigma_p^{-1/2} (\hat{\Sigma}_p - zI)^{-1} \Sigma_p^{1/2} ZU_n, \quad z \in \mathbb{C}^+,
\]
Lemma A.2 Suppose $Z$ satisfies (A.4) and suppose C2–C6 in Section 2 hold. Then,

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

where $\xrightarrow{D}$ denotes weak convergence in $C(\mathbb{C}, \mathbb{R}^{2k^2})$, and $\Psi^{(2)}(z) = [\Psi^{(2)}(z)]_{ij}$ is a $k \times k$ symmetric Gaussian matrix process with zero mean and covariance

$$\mathbb{E}[\Psi^{(2)}(z_1)]_{ij}[\Psi^{(2)}(z_2)]_{ij} = 2\delta(z_1, z_2, \gamma),$$
$$\mathbb{E}[\Psi^{(2)}(z_1)]_{ij}[\Psi^{(2)}(z_2)]_{ij} = \delta(z_1, z_2, \gamma), \quad \text{if } i \neq j,$$
$$\mathbb{E}[\Psi^{(2)}(z_1)]_{ij}[\Psi^{(2)}(z_2)]_{i'j'} = 0, \quad \text{if } i \neq i' \text{ or } j \neq j'.$$

Define a smoothed version of $\xi_n(z, \alpha, \eta)$ as

$$\hat{\xi}_n(z, \alpha, \eta) = \xi_n(z, \alpha, \eta), \quad z \in \mathbb{C}^+,$$
$$\hat{\xi}_n(z, \alpha, \eta) = \frac{p_n - u}{2p_n} \xi_n(u + i\rho_n, \alpha, \eta) + \frac{v + p_n}{2p_n} \xi_n(u - i\rho_n, \alpha, \eta), \quad z \in \mathbb{C}^+. $$

Lemma A.3 Suppose that $Z$ satisfies (A.4) and C2–C6 hold. Then,

$$\hat{\xi}_n(z, \alpha, \eta) - n^{1/2}(\Theta(z, \gamma) - 1)\alpha^T \eta \xrightarrow{D} \Psi^{(3)}(z),$$

where $\xrightarrow{D}$ denotes weak convergence in $C(\mathbb{C}, \mathbb{R}^{2})$, and $\Psi^{(3)}(z)$ is a Gaussian process with zero mean and covariance function $\Gamma^{(2)}(z_1, z_2) = \delta(z_1, z_2, \gamma)\left[[\alpha]\|\eta\|^2 + (\alpha^T \eta)^2\right]$. The following result is an immediate consequence of the foregoing:

$$\int_{\mathbb{C}} \frac{f(z)\hat{\xi}_n(z, \alpha, \eta)}{-2\pi i} dz - n^{1/2}\Omega(f, \gamma)\alpha^T \eta = \mathcal{N}(0, [[\alpha]\|\eta\|^2 + (\alpha^T \eta)^2]\Delta(f, \gamma)). \quad (A.6)$$

To complete the proof of Theorem 2.1, we further need to show (A.6) still holds if (A.4) is extended to C1 and $\hat{\xi}_n(z, \alpha, \eta)$ is replaced by $\xi_n(z, \alpha, \eta)$. We present the extension of (A.6) in Section S.9 of the Supplementary Material.

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

Outline

The structure of the Supplementary Material is as follows.

• In Section S.1, we present the methodology of applying the higher-order shrinkage family $\mathcal{F}_{\text{high}}$ when $\kappa = 3$ introduced in Section 4.2.

• In Section S.2, we study the properties of $m(z)$, $\Theta(z, \gamma)$, and $\delta(z_1, z_2, \gamma)$ on a contour $C$.

• In Section S.3, we present the proof of Theorem A.1. See the beginning of Section S.3 for the structure of subsections.

• In Section S.4, we present the proof of Theorem 2.3 under the truncated variable condition (A.4). The proof is extended from (A.4) to $C_1$ in Section S.9.

• In Section S.5, we present the proof of Theorem 2.4.

• In Section S.6, we present the proof of Theorem 3.1.

• In Section S.7, we present the proof of Theorem 3.2.

• In Section S.8, we present the proof of Remark 2.2, Lemma 2.1 and Lemma 2.2.

• In Section S.9, we discuss the truncation of random variables and the smoothing modification of $\xi_n(z, \alpha, \eta)$ near the real line, as introduced in Appendix. The results in this section show (A.6) still holds if we replace $\xi_n(z, \alpha, \eta)$ with $\xi_n(z, \alpha, \eta)$. Also, the results indicate $M(f)$ has the same weak limit when the observation errors satisfy (A.4) and when they satisfy $C_1$. It completes the proof of Theorem 2.1. Moreover, Section S.9 extends Section S.4 from (A.4) to $C_1$.

• Section S.10 contains technical lemmas.

• In Section S.11, we present additional simulation studies.

• In Section S.12, we present counterparts of Figure 1–Figure 3 in the manuscript but with asymptotic (approximate) cut-off values.
Notation list  For the benefit of readers, we collect the definition of commonly used notation through the Supplementary Material to the following list.

1. Recall $U_n = X^T(X X^T)^{-1/2}$ as defined in Appendix;
2. Recall $\tilde{\Sigma}_p = \frac{1}{n}\sum_{p}^{1/2}ZZ^T\zeta_p^{T/2}$ as defined in Appendix;
3. $u_a = (u_{a1}, \ldots, u_{aN})^T = n^{-1/2}U_n a$;
4. $u_b = (u_{b1}, \ldots, u_{bN})^T = n^{-1/2}U_n b$;
5. $\bar{z}$ is the complex conjugate of $z$;
6. $z_j$ is the $j$th conjugate of $Z$;
7. $Z_j$ is $Z$ with $z_j$ replaced with the 0 vector;
8. $Z_{ij}$ is $Z$ with both $z_i$ and $z_j$ replaced with the 0 vectors;
9. $A(z) = \sum_{i}^{T/2}(\tilde{\Sigma}_p - zI)^{-1}\sum_{i}^{1/2} = \sum_{i}^{T/2}(\frac{1}{n}\sum_{p}^{1/2}ZZ^T\zeta_p^{T/2} - zI)^{-1}\sum_{i}^{1/2}$;
10. $A_j(z) = \sum_{i}^{T/2}(\frac{1}{n}\sum_{p}^{1/2}Z_j^T\zeta_p^{T/2} - zI)^{-1}\sum_{i}^{1/2}$;
11. $A_{ij}(z) = \sum_{i}^{T/2}(\frac{1}{n}\sum_{p}^{1/2}Z_{ij}^T\zeta_p^{T/2} - zI)^{-1}\sum_{i}^{1/2}$;
12. $\sigma(z_1, \ldots, Z_N)$ is the $\sigma$-algebra generated by $z_1, \ldots, Z_N$;
13. $E_j[\cdot] = E[\cdot | \sigma(z_1, \ldots, z_j)], j = 0, \ldots, N$, with the convention $E_0[\cdot] = E[\cdot]$;
14. $\beta_j(z) = \left(1 + n^{-1}Z_j^T A_j(z)z_j \right)^{-1}$;
15. $\beta_j^{tr}(z) = \left(1 + n^{-1}tr A_j(z) \right)^{-1}$;
16. $\beta^Z(z) = \left(1 + E_n^{-1}tr A_1(z) \right)^{-1} = \Theta_n^{-1}(z)$;
17. $\theta_j(z) = \frac{1}{n}z_j^T A_j(z)z_j - \frac{1}{n}tr A_j(z)$;
18. $\phi_j(z) = \frac{1}{n}z_j^T A_j(z)Z_j u_a u_a^T Z_j^T A_j(z)z_j - \frac{1}{n}u_a^T Z_j^T A_j^2(z) Z_j u_b$;
19. $e_i$ is the canonical vector with 1 on the $i$th coordinate;
20. $\alpha_{L_1}(1)$ means a random variable converging to 0 in $L_1$-norm;
21. $\alpha_{L_2}(1)$ means a random variable converging to 0 in $L_2$-norm;
22. $\| \cdot \|_1$ is the entrywise matrix 1-norm, that is the sum of all matrix entries in absolute value.
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,

\[ \mathfrak{S}_{\text{high}} = \left\{ f_\ell(x) = \left[ \sum_{j=0}^{\infty} l_j x^j \right]^{-1}, \ \ell = (l_0, \ldots, l_\kappa)^T \in \mathcal{G} \right\}. \]

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

If \( f_\ell \) has three distinct roots, say \( r_1, r_2, r_3 \), and non-zero leading coefficient \( l_3 \), we have the following representation

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

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

\[ \Omega(f_\ell, \gamma) = \sum_{j=1}^{3} \omega_j \{\Theta(r_j, \gamma) - 1\}, \]

\[ \Delta(f_\ell, \gamma) = 2 \sum_{j=1}^{3} \sum_{j' = 1}^{3} \omega_j \omega_{j'} \delta(r_j, r_{j'}, \gamma) = 2 \sum_{j=1}^{3} \sum_{j' = 1}^{3} \omega_j \omega_{j'} \delta(r_j, r_{j'}, \gamma), \]

where \( \overline{\omega} \) is the complex conjugate of \( \omega \) and \( \overline{\gamma} \) is analogous.

The case that \( f_\ell \) 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 \( f_\ell \) 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 \( \Theta(x, \gamma) \) and \( \delta(x, x', \gamma) \) with respect to \( x \). 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 \( f_\ell \) has a multiple root by forcing Restriction R3 shown below.

As stated in Section 4.2, it is reasonable to require \( f_\ell \) being strictly positive and monotonically decreasing on \( \mathcal{X} \) for any \( \ell \in \mathcal{G} \). Moreover, since multiplying \( f_\ell \) 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 \( \ell = (1, l_1, l_2, l_3) \) satisfying the following restrictions.

R1 (Compactness): \( |l_i| \leq c_i, \ i = 1, 2, \) and \( 0 < \underline{c}_3 \leq |l_3| \leq \overline{c}_3 \).

R2 (Monotonicity): \( 3l_3 x^2 + 2l_2 x + l_1 \geq 0 \) for all \( x \in [0, \overline{x}]. \)
R3 (Distinct roots): \[18l_3l_2l_1 - 4l_2^3 + l_2^2l_1^2 - 4l_3l_1^3 - 27l_1^2 \geq c_4.\]

Here, \(c_1, c_2, c_3, c_4\) are pre-specified positive constants and \(\lambda\) is a constant such that 
\[\lambda \geq \lim_{p \to \infty} \lambda_{\max}(\Sigma_p)(1 + \sqrt{q})^2.\] In practice, we choose \(\lambda \geq [\lambda_{\max}(\Sigma_p)] + 0.01p^{-1}\text{tr}(\Sigma_p).\)

Under R1 and R2, there exists a constant \(K\) such that 
\[\inf_{l \in \mathcal{G}} \inf_{x \in [0, N]} \min_{j=1,2,3} |x - r_j| > K.\]

Under R3, the roots \(r_1, r_2, r_3\) are mutually exclusive and we can find a constant \(K'\) such that 
\[\inf_{l \in \mathcal{G}} \min_{1 \leq j \neq j' \leq 3} |r_j - r_{j'}| > K'.\]

The considerations lead to the following algorithm to determine the shrinkage function 
\(f \in \mathcal{G}_{\text{high}}\) given \(\mathcal{G}\) and to 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** \(\ell = (t_0, t_1, t_2)\). The 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 \(\omega_j = l_0^{-1} \prod_{j \neq j}(r_j - r_1)^{-1}\), \(j = 1, 2, 3\).

3. **For** \(\ell \in \mathcal{G}\), compute the estimates 
\[m_{n,p}(r_j) = \frac{1}{p} \text{tr}[(\hat{\Sigma}_p - r_j I_p)^{-1}], \quad j = 1, 2, 3;\]
\[\hat{\Theta}(r_j, \gamma_n) = (1 - \gamma_n - \gamma_n r_j m_{n,p}(r_j))^{-1}, \quad j = 1, 2, 3;\]
\[\hat{\delta}(r_j, r_j', \gamma_n) = \hat{\Theta}(r_j, \gamma_n) \hat{\Theta}(r_j', \gamma_n) \left[\frac{r_j \hat{\Theta}(r_j, \gamma_n) - r_j' \hat{\Theta}(r_j', \gamma_n)}{r_j - r_j'} - 1\right], \quad j \neq j';\]
\[\hat{\delta}(r_j, r_j, \gamma_n) = \gamma_n \{1 + r_j m_{n,p}(r_j)\} \hat{\Theta}^3(r_j, \gamma)\]
\[+ \gamma_n r_j \{m_{n,p}(r_j) + r_j m_{n,p}(r_j)\} \hat{\Theta}^4(r_j, \gamma_n), \quad j = 1, 2, 3;\]
\[\hat{\rho}_0(r_j, \gamma_n) = m_{n,p}(r_j), \quad j = 1, 2, 3;\]
\[\hat{\rho}_1(r_j, \gamma_n) = \hat{\Theta}(r_j, \gamma_n) [1 + r_j m_{n,p}(r_j)], \quad j = 1, 2, 3;\]
\[\hat{\rho}_2(r_j, \gamma_n) = \hat{\Theta}(r_j, \gamma_n) [p^{-1}\text{tr}(\hat{\Sigma}_p) + r_j \hat{\rho}_1(r_j, \gamma_n)], \quad j = 1, 2, 3;\]
\[\hat{\Omega}(\gamma_n) = \sum_{j=1}^3 \omega_j \{\hat{\Theta}(r_j, \gamma_n) - 1\};\]
Secondly, it is efficient to focus on \( \frac{3}{2} \sum_{j=1}^{3} \sum_{j'=1}^{3} \omega_j \varphi_{j'}(r_j, r_{j'}, \gamma_n) \);

\[
\hat{\Delta}_e(\gamma_n) = 2 \sum_{j=1}^{3} \sum_{j'=1}^{3} \omega_j \varphi_{j'}(r_j, r_{j'}, \gamma_n);
\]

\[
\hat{\Xi}_e(\gamma_n) = \hat{\Delta}_e^{-1/2}(\gamma_n) \sum_{j=1}^{3} \sum_{i=0}^{2} \omega_j t_i \hat{\alpha}(r_j, \gamma_n).
\]

4. Select \( \ell^* = \arg \max_{\ell \in \mathcal{G}} \hat{\Xi}_e(\gamma_n) \) 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 \( \alpha \), if \( T > \xi_{\alpha} \).

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

\[
\hat{T}^{BH}(\ell^*) := \frac{\sqrt{n} \{(1 + \hat{\Omega}_{\ell^*}(\gamma_n))\} \{T^{BH}(\ell^*) - q \hat{\Omega}_{\ell^*}(\gamma_n)\}}{q^{1/2} \hat{\Delta}_{\ell^*}^{-1/2}(\gamma_n)};
\]

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

Here,

\[
T^{LR}(\ell^*) = \sum_{i=1}^{q} \log(1 + \lambda_i), \quad T^{BH}(\ell^*) = \sum_{i=1}^{q} \lambda_i, \quad T^{BNP}(\ell^*) = \sum_{i=1}^{q} \frac{\lambda_i}{1 + \lambda_i},
\]

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

The suggested grid in practice is as follows. Denote \( p^{-1} \text{tr}(\hat{\mathbf{S}}_p) \) as \( \mathfrak{M} \). First, generate

\[
l_3 \in [10^{-3} \bar{\lambda}^{-2} \mathfrak{M}^{-1}, \ 10^2 \bar{\lambda}^{-2} \mathfrak{M}^{-1}].
\]

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 \( f_{\ell}^{-1} \) being too steep, we select \( l_1 \) to be such that

\[
l_1 \mathfrak{M}^2 + l_2 \mathfrak{M} + l_3 = \frac{f_{\ell}^{-1}(\mathfrak{M}) - f_{\ell}^{-1}(0)}{\mathfrak{M}} \in [0, 2].
\]

Valid grid points are those satisfying R2 and R3. \( c_4 \) can be arbitrarily small, while our suggestion is \( c_4 = 10^{-6} \bar{\lambda}^{-2} \).
S.2. About $m(z)$, $\Theta(z, \gamma)$ and $\delta(z_1, z_2, \gamma)$

Recall that $m(z)$ is the Stieltjes transform of $F^{\infty}$, that is the limit almost surely of $F^{\Sigma_p}$ at any point of continuity of $F^{\infty}$. On the other hand, $m(z)$ is the unique solution in $C^+$ of the Marčenko-Pastur equation (2.1), that is

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

where $L^{\Sigma}$ is the limit of $F^{\Sigma_p}$.

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

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

$F^{\infty}$ is a valid c.d.f. for any $\gamma$. It is actually the limit almost surely of the ESD of $n^{-1}Z^T \Sigma p Z$. Denote the Stieltjes transform of $F^{\infty}_p$ to be $m(z)$. Then, there is a 1-1 mapping between $m(z)$ and $\tilde{m}(z)$ as

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

The Marčenko-Pastur equation (2.1) has an equivalent formulation as

$$\tilde{m}(z) = \left[ - z + \gamma \int \frac{\tau dL^{\Sigma}(\tau)}{1 + \tau \tilde{m}(z)} \right]^{-1}.$$

With the help of $m(z)$, we now study the domain of $\Theta(z, \gamma)$ on any contour $C$ enclosing $\mathcal{X} = [0, \lim \sup_p \lambda_{\max}(\Sigma_p)(1 + \sqrt{\gamma})^2]$. Observe that $\Theta(z, \gamma) = (z m(z))^{-1}$. It is claimed in Bai and Silverstein (2004) that

$$\inf_{z \in \mathcal{X}} |m(z)| > 0,$$

for any bounded subset $S$ of $C$. Therefore,

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

Denote the support of $F^{\infty}$ to be $sp(F^{\infty})$. Clearly, $sp(F^{\infty}) \subset \mathcal{X}$. For any selection of the contour $C$,

$$\sup_{z \in C} |m'(z)| = \sup_{z \in C} \left| \int (\tau - z)^{-2} dF^{\infty}(\tau) \right| < \infty,$$

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

Hence,

$$\sup_{z \in C} \left| \frac{\partial}{\partial z} \Theta(z, \gamma) \right| < \infty,$$
About convergence of $F_{Σ_p}$ to $L^Σ$. It is claimed in Section 2.1 that, pointwise almost surely on $z \in \mathbb{C}^+$, $m_{n,p}(z)$ converges to $m(z)$. In view that the convergence of $F_{Σ_p}$ to $L^Σ$ can be arbitrarily slow, the convergence rate of $m_{n,p}(z)$ to $m(z)$ is also arbitrarily slow. Consequently, the convergence rate of $\hat{Θ}(z, γ_n)$ and $δ(z_1, z_2, γ_n)$ 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 to replace $m_{0,p}(z)$ with a deterministic sequence $t_{m_{0,p}}(z) = \int \frac{dF_{Σ_p}(τ)}{τ(1 - γ_n - γ_n zm_{0,p}(z)) - z}$.

See for example Bai and Silverstein (2010). 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 $m_{n,p}(z) - m_{0,p}(z)$ 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 $m_{0,p}(z)$ 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 $C_2$ and $C_4$. Later, we will frequently refer to existing results in literature that are established using $t_{m_{0,p}}(z)$, $p = 1, 2, \ldots$. It is necessary to study the difference between $m_{0,p}(z)$ and $m(z)$ under $C_2$ and $C_4$.

Similarly, define

$$m_{0,p}(z) = \frac{γ_n - 1}{z} + γ_n m_{0,p}(z).$$

The following formulation also holds

$$m_{0,p}(z) = -z + γ_n \int \frac{dF_{Σ_p}(τ)}{1 + τ m_{0,p}(z)} \tau^{-1}.$$

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 $C$ bounded away from the support of $F_{Σ}$, as $n \to \infty$,

$$\sup_{z \in C} |m_{0,p}(z) - m(z)| \to 0.$$
Therefore,
\[ \sup_{z \in \mathcal{C}} |m^0_p(z) - m(z)| \rightarrow 0. \]

The result still holds under our assumption \( \sqrt{n}D_W(F^{X_p}, L^2) \rightarrow 0 \) in \( \mathcal{C}^4 \), because the weak convergence of \( F^{X_p} \) is implied. Next, we show the convergence rate is faster than \( n^{-1/2} \).

**Lemma S.1** Suppose \( \mathcal{C}^2 \) and \( \mathcal{C}^4 \) hold. For any contour \( \mathcal{C} \) such that \( X \) is in the interior of \( \mathcal{C} \),
\[ \sup_{z \in \mathcal{C}} \sqrt{p}|m^0_p(z) - m(z)| \rightarrow 0. \]

We only need to show
\[ \sup_{z \in \mathcal{C}} \sqrt{p}|m^0_p(z) - m(z)| \rightarrow 0. \]

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

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

Denote the support of \( L(\Sigma) \) to be \( \text{sp}(L^2) \). We first show that
\[ \inf_{z \in \mathcal{C}} \inf_{\tau \in \text{sp}(L^2)} |1 + \tau m(z)| > 0. \]  
(S.2.1)

When \( z = \bar{u} \) or \( \bar{\pi} \), it follows from Silverstein and Choi (1995, Theorem 4.1) that
\[ -m(z)^{-1} \in \text{sp}(L^2)^c. \]

Therefore, \( 1 + \tau m(\bar{u}) \neq 0 \) and \( 1 + \tau m(\bar{\pi}) \neq 0 \) for any \( \tau \in \text{sp}(L^2) \). Following from continuity of \( m(z) \) and the fact that \( \text{sp}(L^2) \) is compact, we can find a neighborhood of \( \bar{u} \) and a neighborhood of \( \bar{\pi} \) such that there exists a \( \epsilon > 0 \), for any \( z \) in the two neighborhoods,
\[ \inf_{\tau \in \text{sp}(L^2)} |1 + \tau m(z)| > \epsilon. \]

When \( z \) is outside of the neighborhoods, so away from the real axis, Bai and Silverstein (1998, Lemma 2.11) indicates that
\[ |1 + \tau m(z)| \leq \max\left(\frac{4 \lim sup_p \lambda(\Sigma_p)(1 + \sqrt{\gamma})^2}{\Im(z)}, 2\right), \]

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

Moreover, since \( \sup_{z \in \mathcal{C}} |m^0_p(z) - m(z)| \rightarrow 0 \), for all sufficiently large \( p \),
\[ \inf_{z \in \mathcal{C}} \inf_{\tau \in \text{sp}(L^2)} |1 + \tau m^0_p(z)| > \epsilon/2. \]

Following from (S.2.1), uniformly on \( \mathcal{C} \),
\[ w_{n1} := \left[ \gamma_n \int \frac{T_{\Sigma}(\tau)}{1 + \tau m(z)} dF^{X_p}(\tau) - \gamma \int \frac{\tau dL^2(\tau)}{1 + \tau m(z)} \right] = o(n^{-1/2}). \]
Define
\[
\tilde{m}_p(z) = \left[ -z + \gamma_n \int \frac{\tau z_{sp}(\tau)}{1 + \tau m(z)} dF^{\Sigma_p}(\tau) \right]^{-1}.
\]
We have, uniformly on \( C \),
\[
\tilde{m}_p(z) - m(z) = m(z) \left[ m^{-1}(z) + w_{n1} \right]^{-1} w_{n1} = o(n^{-1/2}),
\]
since both \( m(z) \) and \( m^{-1}(z) \) are bounded on \( C \).

Now consider \( \bar{m}_p^0(z) - \tilde{m}_p(z) \). The target is to show
\[
\bar{m}_p^0(z) - \tilde{m}_p(z) = [m_p^0(z) - m(z)]R_p + o(n^{-1/2})
\]
for some \( R_p \) such that \( \sup_{z \in C \setminus [y, \pi]} |R_p| < 1 \) for sufficiently large \( p \).

For simplicity, we shall write \( \mathbb{1}_{sp}(L^p) \) as \( \mathbb{1}_{sp} \). First, observe that the imaginary part of \( \bar{m}_p^0(z) \), denoted to be \( \Im_0(z) \), is
\[
\Im_0(z) = \mathbb{3}(z) + \gamma_n \int \frac{\tau z_{sp}^0(\tau) dF^{\Sigma_p}(\tau)}{1 + \tau \hat{m}(z)}
\]
\[-z + \gamma_n \int \frac{\tau dF^{\Sigma_p}(\tau)}{1 + \tau \hat{m}(z)} \frac{1}{2}.
\]
The imaginary part of \( \tilde{m}_p(z) \), denoted to be \( \Im_0(z) \), is
\[
\bar{m}_p^0(z) - \tilde{m}_p(z) = m(z) \left[ m^{-1}(z) + w_{n1} \right]^{-1} w_{n1} = o(n^{-1/2}),
\]
where \( v(z) \) is the imaginary part of \( m(z) \).

Next, we verify that when \( p \) is sufficiently large, \( |R_p| < 1 \). By Hölder’s inequality, when \( \mathbb{3}(z) \neq 0, \)
\[
|R_p|^2 = \left| \gamma_n \int \frac{\tau z_{sp}^0(\tau) dF^{\Sigma_p}(\tau)}{1 + \tau \hat{m}(z)} \right| \left[ -z + \gamma_n \int \frac{\tau dF^{\Sigma_p}(\tau)}{1 + \tau \hat{m}(z)} \right]^{-2}.
\]
Assume that $n \geq 10$

Suppose Lemma S.2

Comparing with (S.2.3), we have

$$r_n \leq C, \quad \text{where } C = C(n).$$

Therefore, $r_n < K < 1$ for all $z \in \mathcal{C} \setminus \{u, \overline{u}\}$.

We only need to show $r_n \leq 1$. Because $\overline{m}(z)$ is the Stieltjes transform of $F^\infty$, again,

$$0 < \inf_{z \in \mathcal{C} \setminus \{u, \overline{u}\}} \frac{v_0^0(z)}{3(z)} \leq \sup_{z \in \mathcal{C} \setminus \{u, \overline{u}\}} \frac{v_0^0(z)}{3(z)} < \infty.$$ 

Observe that

$$v(z) = \frac{3(z) + \gamma \int \frac{\tau^2 v(z) dL^\infty(z)}{1 + \tau^2 m(z)^2}}{\left| -z + \gamma \int \frac{\tau dL^\infty(z)}{1 + \tau m(z)} \right|^2}.$$ 

Comparing with (S.2.3), we have

$$v_p(z) / v(z) = \frac{3(z) \left| v(z) + \gamma \int \frac{\tau^2 v(z) dL^\infty(z)}{1 + \tau^2 m(z)^2} \right|^2}{3(z) \left| v(z) + \gamma \int \frac{\tau dL^\infty(z)}{1 + \tau^2 m(z)^2} \right|^2}.$$ 

Therefore, for sufficiently large $n$, $r_n \leq 1$ for all $z \in \mathcal{C} \setminus \{u, \overline{u}\}$.

Together, we proved

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

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

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

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

$$\sup_{z \in \mathcal{C}} \sqrt{p} \left| \frac{d}{dz} m_p^0(z) - \frac{d}{dz} m(z) \right| \longrightarrow 0.$$
We follow the arguments in the proof of Lemma S.1.
\[
m_0^0(z) - m(z) = [m_0^0(z) - m(z)]R_p + w_{n2} + m(z)[m^{-1}(z) + w_{n1}]^{-1}w_{n1}.
\]

Take differentiation on both sides,
\[
\frac{d}{dz}m_0^0(z) - \frac{d}{dz}m(z) = \left(\frac{d}{dz}m_0^0(z) - \frac{d}{dz}m(z)\right)R_p + \left[m_0^0(z) - m(z)\right]R_p
+ \frac{d}{dz}w_{n2} + \frac{d}{dz}m(z)[m^{-1}(z) + w_{n1}]^{-1}w_{n1}
+ \frac{m(z)}{d}m(z)[m^{-1}(z) + w_{n1}]^{-1}w_{n1} + m(z)[m^{-1}(z) + w_{n1}]^{-1} \frac{d}{dz}w_{n1}.
\]

It is straightforward to verify, using arguments in the proof of Lemma S.1, that
\[
\sup_{z \in \mathbb{C}} \left| \frac{d}{dz}R_p \right| < \infty,
\]
\[
\sup_{z \in \mathbb{C}} \left| \frac{d}{dz}[m^{-1}(z) + w_{n1}]^{-1} \right| < \infty,
\]
\[
\sup_{z \in \mathbb{C}} \sqrt{n} \frac{d}{dz}w_{n1} \to 0,
\]
\[
\sup_{z \in \mathbb{C}} \sqrt{n} \frac{d}{dz}w_{n2} \to 0.
\]

Together with the fact that \(m(z), m^{-1}(z), dm(z)/dz, dm^{-1}(z)/dz\) are bounded,
\[
\frac{d}{dz}m_0^0(z) - \frac{d}{dz}m(z) = \left(\frac{d}{dz}m_0^0(z) - \frac{d}{dz}m(z)\right)R_p + o(n^{-1/2}).
\]

The uniform convergence of \(\sqrt{n} \frac{d}{dz}m_0^0(z) - \frac{d}{dz}m(z)\) follows.

**S.3. Proof of Theorem A.1**

Recall notation in Section 2 and Appendix of the paper. Define
\[
\Theta_n(z) = 1 + \gamma_n \frac{1}{p} \text{Tr}((\hat{\Sigma}_p - zI)^{-1} \Sigma_p),
\]
\[
G_n^{(1)}(z, a, b) = n^{1/2} \{G_n(z, a, b) - \mathbb{E}G_n(z, a, b)\},
\]
\[
G_n^{(2)}(z, a, b) = n^{1/2} \{\mathbb{E}G_n(z, a, b) - a^T b \Theta_n(z, \gamma) - 1 \},
\]
\[
G_n^{(3)}(z, a, b) = \frac{1}{2} a^T b \left[ \Theta_n(z) - 1 \Theta_n(z, \gamma) - 1 \right].
\]

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.
S.3.1. Finite dimensional convergence of $G^{(1)}_n(z, a, b)$

In this subsection, we show that

$$\sum_{i=1}^{r} \omega_i G^{(1)}_n(\tilde{z}_i, a, b)$$

converges to a Gaussian random variable, where $r$ is any positive integer, $\omega_1, \ldots, \omega_r$ and $z_1, \ldots, z_r$ are any complex numbers. In view of the smoothing modification as in Appendix, $z_1, \ldots, z_r$ are required to have nonzero imaginary part. Without loss of generality, assume $n$ is sufficiently large so that $\rho_n$ is smaller than the imaginary part of $z_1, \ldots, z_r$.

Before we proceed, recall the notation list. The following identities holds

$$\begin{align*}
\beta_j(x) &= \beta^T_j(x) - \beta_j(x) \beta^T_j(x) \theta_j(x), \\
\beta_j(x) &= \beta^R_j(x) - \beta_j(x) \beta^G_j(x) \left( \frac{1}{n} z_j^T A_j(z) z_j - \frac{1}{n} \text{tr} A_j(z) \right).
\end{align*}$$

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

$$\|u_a\|_{\text{max}} = O(n^{-1}), \quad \|u_b\|_{\text{max}} = O(n^{-1}).$$

S.3.1.1. Construction of martingale difference sequences

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

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

$$= n^{1/2} \sum_{j=1}^{N} \{ \mathbb{E}_j[u_a^T Z^T A(x) Z u_b - u_a^T Z_j^T A_j(x) Z_j u_b] \\
- \mathbb{E}_{j-1}[u_a^T Z^T A(x) Z u_b - u_a^T Z_j^T A_j(x) Z_j u_b] \}$$

$$= n^{1/2} \sum_{j=1}^{N} (\mathbb{E}_j - \mathbb{E}_{j-1})[d_1(z) + d_2(z) + d_3(z)],$$

where

$$\begin{align*}
d_1(z) &= u_a^T (Z - Z_j)^T A(x) Z u_b, \\
d_2(z) &= u_a^T Z_j^T (A(x) - A_j(x)) Z u_b, \\
d_3(z) &= u_a^T Z_j^T A_j(x) (Z - Z_j) u_b.
\end{align*}$$
\[ d_1(z) = u_{a_j}z_j^T A_j(z)Z_j u_b + u_{a_j}u_{b_j}z_j^T A_j(z)z_j \]
\[ - \frac{1}{n} u_{a_j}z_j^T A_j(z)z_j z_j^T A_j(z)Z_j u_b \beta_j(z) - \frac{1}{n} u_{a_j}u_{b_j} (z_j^T A_j(z)z_j)^2 \beta_j(z) \]
\[ = d_1^{(1)}(z) + d_1^{(2)}(z) + d_1^{(3)}(z) + d_1^{(4)}(z), \text{ say.} \]

\( d_1^{(1)}(z) \) is such that
\[ n^{1/2} \sum_{j=1}^N (E_j - E_{j-1})d_1^{(1)}(z) = n^{1/2} \sum_{j=1}^N E_j [u_{a_j}z_j^T A_j(z)Z_j u_b]. \]

By Lemma S.8, Lemma S.10, and Lemma S.13,
\[ E \left| n^{1/2} \sum_{j=1}^N (E_j - E_{j-1})d_1^{(2)}(z) \right|^2 \leq n \sum_{j=1}^N |u_{a_j}u_{b_j}|^2 E \left| z_j^T [E_{j-1} A_j(z)] z_j - \text{tr} E_{j-1} A_j(z) \right|^2 \]
\[ \leq \kappa n \sum_{j=1}^N |u_{a_j}u_{b_j}|^2 n E \left| E_{j-1} A_j(z) \right|^2 = o(1). \]

Due to (S.3.1), \( d_1^{(3)}(z) \) is such that
\[ d_1^{(3)}(z) = \frac{1}{n} u_{a_j} z_j^T A_j(z) z_j z_j^T A_j(z) Z_j u_b \beta_j(z) \beta_j^T(z) \theta_j(z) \]
\[ - u_{a_j} \theta_j(z) z_j^T A_j(z) Z_j u_b \beta_j^T(z) + u_{a_j} (\beta_j^T(z) - 1) z_j^T A_j(z) Z_j u_b. \]

By Lemma S.8, Lemma S.10, and Lemma S.18, the first two terms above are such that
\[ E \left| n^{1/2} \sum_{j=1}^N (E_j - E_{j-1}) u_{a_j} z_j^T A_j(z) z_j z_j^T A_j(z) Z_j u_b \beta_j(z) \beta_j^T(z) \theta_j(z) \right|^2 = o(1), \]
\[ E \left| n^{1/2} \sum_{j=1}^N (E_j - E_{j-1}) u_{a_j} \theta_j(z) z_j^T A_j(z) Z_j u_b \beta_j^T(z) \right|^2 = o(1). \]

It leads to
\[ n^{1/2} \sum_{j=1}^N (E_j - E_{j-1}) d_1^{(3)}(z) = \sum_{j=1}^N n^{1/2} E_j [(\beta_j^T(z) - 1) u_{a_j} z_j^T A_j(z) Z_j u_b] + o_p(1). \]
\[ - (E_j - E_{j-1}) d_1^{(4)}(z) \]
\[ = (E_j - E_{j-1}) \frac{1}{n} u_{a_j} u_{b_j} (z_j^T A_j(z) z_j^2) \beta_j^T(z) \]
Further, we want to replace $\beta_j^t(z)$ with $\beta^E_j(z)$. By (S.3.1), Lemma S.12, and Lemma S.16,

$$
\mathbb{E}[|\beta_j^t(z) - \beta^E_j(z)|^2] \\
= \mathbb{E}[|\beta_j^t(z)|^2] - \mathbb{E}[|\beta^E_j(z)|^2] - \mathbb{E}[\text{tr}A_j(z)]^2 + \mathbb{E}[\text{tr}A_j(z)]^2 \\
= o(1).
$$

Thus, we have

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

All together

$$
n^{1/2} \sum_{j=1}^N (\mathbb{E}_j - \mathbb{E}_{j-1})d_1(z) = n^{1/2} \sum_{j=1}^N \mathbb{E}_j \left| u_{a_j} z_j^T A_j(z) z_j \right|^2 + o_p(1).
$$

Further, we want to replace $\beta_j^t(z)$ with $\beta^E_j(z)$. By (S.3.1), Lemma S.12, and Lemma S.16,

$$
\mathbb{E}[|\beta_j^t(z) - \beta^E_j(z)|^2] \\
= \mathbb{E}[|\beta_j^t(z)|^2] - \mathbb{E}[|\beta^E_j(z)|^2] - \mathbb{E}[\text{tr}A_j(z)]^2 + \mathbb{E}[\text{tr}A_j(z)]^2 \\
= o(1).
$$

Therefore,

$$
n^{1/2} \sum_{j=1}^N (\mathbb{E}_j - \mathbb{E}_{j-1})d_1(z) = \beta^E(z) \sum_{j=1}^N \mathbb{E}_j \left| n^{1/2} u_{a_j} z_j^T A_j(z) z_j \right|^2 + o_p(1).
$$
Secondly, using Lemma S.6,

\[
d_2(z) = -\frac{1}{n}u_{b_j}z_j^T A_j(z)z_j^T u_n \beta_j(z)
- \frac{1}{n}z_j^T A_j(z)Z_j u_n u_n^T Z_j^T A_j(z)z_j \beta_j(z)
- \frac{1}{n}u_{b_j}z_j^T A_j(z)z_j^T A_j(z)Z_j u_n \beta_j(z)
- \rho_j(z) \beta_j(z) - \frac{1}{n}u_{b_j}z_j^T A_j^2(z)Z_j u_n \beta_j(z)
= d_2^{(1)}(z) + d_2^{(2)}(z) + d_2^{(3)}(z), \text{ say.}
\]

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

\[
\mathbb{E}\left|\rho_j(z)\right|^2 \leq \mathcal{K}\mathbb{E}\frac{1}{n^2} \left| u_{b_j}^T Z_j^T A_j(z)A_j(z)Z_j u_n u_n^T Z_j^T A_j(z)A_j(z)Z_j u_n \right| = O(n^{-2}),
\]

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

\[
n^{1/2} \sum_{j=1}^N (E_j - E_{j-1})d_2^{(1)}(z)
= \sum_{j=1}^N \mathbb{E}_j[(\beta_j^{(t)}(z) - 1)n^{1/2}u_{b_j}z_j^T A_j(z)Z_j u_n] + o_p(1)
= (\beta(z) - 1) \sum_{j=1}^N \mathbb{E}_j[n^{1/2}u_{b_j}z_j^T A_j(z)Z_j u_n] + o_p(1).
\]

\[
n^{1/2} \sum_{j=1}^N (E_j - E_{j-1})d_2^{(2)}(z) = -\beta(z)n^{1/2} \sum_{j=1}^N (E_j - E_{j-1})\rho_j(z)
- n^{1/2} \sum_{j=1}^N (E_j - E_{j-1})\rho_j(z)\{\beta_j^{(t)}(z) - \beta(z)\}
+ n^{1/2} \sum_{j=1}^N (E_j - E_{j-1})\rho_j(z)\beta_j(z)\beta_j^{(t)}(z)\theta_j(z)
= -\beta(z)n^{1/2} \sum_{j=1}^N \mathbb{E}_j\rho_j(z) + o_p(1).
\]

\[
n^{1/2} \sum_{j=1}^N (E_j - E_{j-1})d_2^{(3)}(z)
= \sum_{j=1}^N (E_j - E_{j-1})n^{-1/2}u_{b_j}^T Z_j^T A_j^2(z)Z_j u_n \beta_j(z)\beta_j^{(t)}(z)\theta_j(z) = o_p(1).
\]
All together, we have
\[
n^{1/2} \sum_{j=1}^{N} (E_j - E_{j-1}) (d_{2}(z)) = (\beta^{E}(z) - 1) \sum_{j=1}^{N} E_j [n^{1/2} u_{0j} z_j^T A_j(z) Z_j u_0] \\
- \beta^{E}(z)n^{1/2} \sum_{j=1}^{N} E_j \varrho_j(z).
\]
Combining with \(d_3(z)\), we have proved, thus far,
\[
G^{(1)}_n(z) = \sum_{j=1}^{N} \beta^{E}(z) \mathcal{H}_n(z, j) + o_p(1),
\]
where
\[
\mathcal{H}_n(z, j) = E_j [n^{1/2} u_{0j} z_j^T A_j(z) Z_j u_0] + E_j [n^{1/2} u_{0j} z_j^T A_j(z) Z_j u_0] \\
- E_j n^{1/2} \varrho_j(z) = \mathcal{H}_{11}(z, j) + \mathcal{H}_{23}(z, j) + \mathcal{H}_{33}(z, j), \text{ say.}
\]
In summary, it suffices to find the weak limit of \(\sum_{i=1}^{r} \sum_{j=1}^{N} \omega_i \beta^{E}(z_i) \mathcal{H}_n(z_i, j)\), that is
\[
\sum_{i=1}^{r} \sum_{j=1}^{N} \omega_i \beta^{E}(z_i) [\mathcal{H}_{11}(z_i, j) + \mathcal{H}_{23}(z_i, j) + \mathcal{H}_{33}(z_i, j)].
\]

### S.3.1.2. Martingale central limit theorem
We use the following theorem to show finite dimensional convergence of
\[
\sum_{i=1}^{r} \sum_{j=1}^{N} \omega_i \beta^{E}(z_i) \mathcal{H}_n(z_i, j).
\]

**Theorem S.1** (Theorem 35.12 of Billingsley (1995)). Suppose \(Y_1, \ldots, Y_n\) is a martingale difference sequence with respect to the increasing \(\sigma\)-field \(\mathcal{F}_1, \ldots, \mathcal{F}_n\), with finite second moments. If as \(n \to \infty\),

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

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

then,
\[
\sum_{j=1}^{n} Y_j \to \mathcal{N}(0, \sigma^2), \text{ in distribution.}
\]
It is easy to check that $H_a(z, j)$ has finite second moments. We show Condition (ii) first.

As for $H_2(z, j)$, because of insufficient finite moments of $z_{ij}$, we need to analyze $\varrho_j(z)$ carefully. Write

$$
\varrho_j(z) = \frac{1}{n} \sum_{i \neq l} e_i^T A_j(z) z_{ij} u_a u_b^T Z_j^T A_j(z) e_l z_{lj} \\
+ \frac{1}{n} \sum_{i = 1}^p e_i^T A_j(z) Z_j u_a u_b^T Z_j^T A_j(z) e_i \left[ z_{ij}^2 \mathbb{1}(|z_{ij}| \leq \log n) - \mathbb{E} z_{ij}^2 \mathbb{1}(|z_{ij}| \leq \log n) \right] \\
+ \frac{1}{n} \sum_{i = 1}^p e_i^T A_j(z) Z_j u_a u_b^T Z_j^T A_j(z) e_i \left[ z_{ij}^2 \mathbb{1}(|z_{ij}| > \log n) - \mathbb{E} z_{ij}^2 \mathbb{1}(|z_{ij}| > \log n) \right] \\
= \varrho_j^{(1)}(z) + \varrho_j^{(2)}(z) + \varrho_j^{(3)}(z), \text{ say.}
$$

By Lemma 5 of Pan and Zhou (2011),

$$
\mathbb{E}|\varrho_j^{(1)}(z)|^4 \leq \mathcal{K} n^{-4} \mathbb{E} \left| u_a^T Z_j^T A_j(z) A_j(z) Z_j u_a u_b^T Z_j^T A_j(z) Z_j u_b \right|^2 \\
= O(n^{-4}).
$$

By Lemma S.15,

$$
\mathbb{E}|\varrho_j^{(2)}(z)|^4 \leq \mathcal{K} n^{-4} (\log n)^8 \mathbb{E} \left( \sum_{i = 1}^p |e_i^T A_j(z) Z_j u_a u_b^T Z_j^T A_j(z) e_i| \right)^4 \\
\leq \mathcal{K} n^{-4} (\log n)^8 \mathbb{E} \left| u_a^T Z_j^T A_j(z) A_j(z) Z_j u_a u_b^T Z_j^T A_j(z) A_j(z) Z_j u_b \right|^2 \\
= O(n^{-4} (\log n)^8).
$$

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

$$
\mathbb{E}|\varrho_j^{(3)}(z)|^2 \leq \mathcal{K} n^{-2} \mathbb{E} \left| \sum_{i = 1}^p |e_i^T A_j(z) Z_j u_a u_b^T Z_j^T A_j(z) e_i| \right|^2 \mathbb{E} z_{ij}^4 \mathbb{1}(|z_{ij}| > \log n) \\
= o(n^{-2}).
$$

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 $H_1(z, j)$ and $H_2(z, j)$, Lemma S.16 indicates that

$$
\mathbb{E} |n^{1/2} u_a z_j^T A_j(z) Z_j u_b|^4 = O(n^{-2}), \\
\mathbb{E} |n^{1/2} u_b z_j^T A_j(z) Z_j u_a|^4 = O(n^{-2}).
$$
To verify condition (ii), for any positive $\epsilon$,

$$
\sum_{j=1}^{N} \mathbb{E}\left[\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) \mathcal{H}_n(z_i, j)\right|^2 \mathbbm{1}\left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) \mathcal{H}_n(z_i, j)\right| \geq \epsilon\right)\right]
\leq 25 \sum_{j=1}^{N} \mathbb{E}\left[\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) \mathcal{H}_{[1]}(z_i, j)\right|^2 \mathbbm{1}\left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) \mathcal{H}_{[1]}(z_i, j)\right| \geq \epsilon/5\right)\right]
+ 25 \sum_{j=1}^{N} \mathbb{E}\left[\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) n^{1/2} \eta_j^{(1)}(z)\right|^2 \mathbbm{1}\left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) n^{1/2} \eta_j^{(1)}(z)\right| \geq \epsilon/5\right)\right]
+ 25 \sum_{j=1}^{N} \mathbb{E}\left[\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) n^{1/2} \eta_j^{(2)}(z)\right|^2 \mathbbm{1}\left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) n^{1/2} \eta_j^{(2)}(z)\right| \geq \epsilon/5\right)\right]
+ 25 \sum_{j=1}^{N} \mathbb{E}\left[\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) n^{1/2} \eta_j^{(3)}(z)\right|^2 \mathbbm{1}\left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) n^{1/2} \eta_j^{(3)}(z)\right| \geq \epsilon/5\right)\right]
\leq \frac{K}{\epsilon^2} \sum_{j=1}^{N} \mathbb{E} \left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) \mathcal{H}_{[1]}(z_i, j)\right|^4\right) + \frac{K}{\epsilon^2} \sum_{j=1}^{N} \mathbb{E} \left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) \mathcal{H}_{[2]}(z_i, j)\right|^4\right)
+ \frac{K}{\epsilon^2} \sum_{j=1}^{N} \mathbb{E} \left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) n^{1/2} \eta_j^{(1)}(z)\right|^4\right) + \frac{K}{\epsilon^2} \sum_{j=1}^{N} \mathbb{E} \left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) n^{1/2} \eta_j^{(2)}(z)\right|^4\right)
+ Kn \sum_{j=1}^{N} \mathbb{E} \left(\left|\sum_{i=1}^{r} \omega_i \beta^E(z_i) \eta_j^{(3)}(z)\right|^2\right) \longrightarrow 0.
$$

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

$$
\sum_{j=1}^{N} \mathbb{E}_{j-1}\left(\sum_{i=1}^{r} \omega_i \beta^E(z_i) \mathcal{H}_n(z_i, j)\right)^2
= \sum_{i=1}^{r} \sum_{j=1}^{N} \mathbb{E}_{j-1} \omega_i \omega_i \beta^E(z_i) \beta^E(z_i) \mathcal{H}_n(z_i, j) \mathcal{H}_n(z_i, j).
$$

In next four sections, for arbitrary $z_1$ and $z_2$ with nonzero imaginary part, we derive the limit in probability of

$$
\sum_{j=1}^{N} \mathbb{E}_{j-1} \mathcal{H}_{[1]}(z_1, j) \mathcal{H}_{[1]}(z_2, j),
$$

(S.3.2)
Note, $\mathcal{H}_2([z_1, j)\mathcal{H}_2([z_2, j)$ and $\mathcal{H}_2([z_1, j)\mathcal{H}_3([z_2, j)$ are just $\mathcal{H}_1([z_1, j)\mathcal{H}_1([z_2, j)$ and $\mathcal{H}_1([z_1, j)\mathcal{H}_3([z_2, j)$ respectively with $a$ and $b$ exchanged.

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

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

$$
\sum_{j=2}^{N} \mathbb{E}_{j-1} \mathcal{H}_1([z_1, j)\mathcal{H}_1([z_2, j) = n \beta^2(z_1) \beta^2(z_2) \sum_{j=2}^{N} u_{ij}^2 \text{tr} \left[ \mathbb{E}_j A_j(z_1) \mathbb{E}_j A_j(z_2) \right] \sum_{i=1}^{j-1} u_{bi}^2 + o_p(1).
$$

Note that clearly when $j = 1$, $\mathbb{E} \mathcal{H}_1([z_1, 1)\mathcal{H}_1([z_2, 1) = 0$.

We introduce

$$
Z_j = [z_1, \ldots, z_{j-1}, 0, z_{j+1}, \ldots, z_N]
$$

for $j = 2, 3, \ldots, N$, where $Z_{j+1}, \ldots, Z_N$ are i.i.d. copies of $Z_1$ and independent with $z_1, \ldots, z_{j-1}$. What’s more, introduce $A_j(z)$ like $A_j$, but $A_j(z)$ is now defined on $Z_j$ instead of $Z_j$. Note, conditional on $z_1, \ldots, z_{j-1}$, $(Z_j, A_j)$ is independent with $(Z_i, A_i)$. Therefore, for $j \geq 2$,

$$
\begin{align*}
\mathbb{E}_{j-1} \left[ \mathbb{E}_j [z_i^T A_j(z_1) Z_j u_b] \mathbb{E}_j [z_i^T A_j(z_2) Z_j u_b] \right] & = \mathbb{E}_{j-1} \left[ u_i^T Z_j^T A_j(z_1) A_j(z_2) Z_j u_b \right] \\
& = \sum_{i=1}^{j-1} u_{bi} \mathbb{E}_{j-1} [z_i^T A_j(z_1) A_j(z_2) Z_j u_b] + \sum_{i=1}^{j-1} u_{bi} \mathbb{E}_{j-1} [z_i^T A_j(z_1) A_j(z_2) Z_j u_b] \\
& \quad + \sum_{i=j+1}^{N} u_{bi} \mathbb{E}_{j-1} [z_i^T A_j(z_1) A_j(z_2) Z_j u_b] = \mathcal{J}_j^{(1)} + \mathcal{J}_j^{(2)} + \mathcal{J}_j^{(3)}, \text{ say.}
\end{align*}
$$

Here, $Z_{ij}, i < j$ is defined to be $[z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_{j-1}, 0, z_{j+1}, \ldots, z_N]$.

(S.3.6) will follow, if we can show,

$$
\sup_{1 \leq j \leq N} \mathbb{E} \left[ \mathcal{J}_j^{(1)} - \beta^2(z_1) \beta^2(z_2) \left( \sum_{i=1}^{j-1} u_{bi}^2 \right) \text{tr} \left[ \mathbb{E}_j [A_j(z_1) A_j(z_2)] \right] \right] = o(1),
$$
\[ \sup_{1 \leq j \leq N} \mathbb{E}[\mathcal{J}_j^{(2)}] = o(1), \]
\[ \sup_{1 \leq j \leq N} \mathbb{E}[\mathcal{J}_j^{(3)}] = o(1). \]

Similar to \( \beta_j(z) \), define
\[ \beta_{ij}(z) = \frac{1}{1 + n^{-1}z_i^T A_{ij}(z)z_i}, \]
\[ \beta_{ij}(z) = \frac{1}{1 + n^{-1}z_i^T A_{ij}(z)z_i}, \]
where \( A_{ij}(z) \) is defined in the same way as \( A_{ij}(z) \), but with \( Z_{ij} \) instead of \( Z_{ij} \).

As for \( \mathcal{J}_j^{(1)} \),
\[ \mathbb{E}[\mathcal{J}_j^{(1)}] = \beta^E(z_1)\beta^E(z_2) \sum_{i=1}^{j-1} u_{bi}^2 \text{tr} \left[ \mathbb{E}[A_j(z_1)A_j(z_2)] \right] \]
\[ = \mathbb{E} \left[ \sum_{i=1}^{j-1} u_{bi}^2 \beta_{ij}(z_1)\beta_{ij}(z_2)z_i^T A_{ij}(z_1)A_{ij}(z_2)z_i - \right. \]
\[ \left. \sum_{i=1}^{j-1} u_{bi}^2 \beta^E(z_1)\beta^E(z_2) \text{tr} \left[ \mathbb{E}[A_j(z_1)A_j(z_2)] \right] \right] \]
\[ \leq n^2\|u_b\|_\infty^2 \mathbb{E} \left| \beta_{ij}(z_1)\beta_{ij}(z_2) - \frac{1}{n} z_i^T A_{ij}(z_1)A_{ij}(z_2)z_i - \right. \]
\[ \left. \beta^E(z_1)\beta^E(z_2) \frac{1}{n} \text{tr} \{ A_j(z_1)A_j(z_2) \} \right| \]
\[ \leq n^2\|u_b\|_\infty^2 \mathbb{E} \left| (\beta_{ij}(z_1) - \beta^E(z_1))\beta_{ij}(z_2) - \beta^E(z_2) \right| \frac{1}{n} z_i^T A_{ij}(z_1)A_{ij}(z_2)z_i \]
\[ + n^2\|u_b\|_\infty^2 \mathbb{E} \left| \beta^E(z_1)(\beta_{ij}(z_2) - \beta^E(z_2)) \right| \frac{1}{n} z_i^T A_{ij}(z_1)A_{ij}(z_2)z_i \]
\[ + n^2\|u_b\|_\infty^2 \mathbb{E} \left| \beta^E(z_1) - \beta^E(z_2) \right| \left| \frac{1}{n} z_i^T A_{ij}(z_1)A_{ij}(z_2)z_i - \frac{1}{n} \text{tr} \{ A_j(z_1)A_j(z_2) \} \right| \]
\[ = o(1). \]

For the last line, we need the following arguments that are direct consequences of Lemma S.10, Lemma S.11, Lemma S.12, and Lemma S.13.
\[ |\beta^E(z)| < \infty, \]
\[ \mathbb{E} \left| \frac{1}{n} z_i^T A_{ij}(z)A_{ij}(z)z_i \right|^2 = O(1), \]
\[ \mathbb{E} \left| \frac{1}{n} z_i^T A_{ij}(z)A_{ij}(z)z_i - \frac{1}{n} \text{tr} \{ A_j(z_1)A_j(z_1) \} \right|^2 = o(1), \]
\[ \mathbb{E} |\beta_{ij}(z) - \beta^E(z)|^2 \leq \mathbb{E} \left| \beta_{ij}(z) - \beta^E(z) \right| \left| \frac{1}{n} z_i^T A_{ij}(z)z_i - \frac{1}{n} \text{tr} A_{ij} \right|^2. \]
The proofs of the two are very similar. Therefore, we shall only present proof of (S.3.7).

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

\[
\mathbb{E}
\begin{bmatrix}
(\beta_j(z_1) - \beta_i^E(z_1)) A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_b
\end{bmatrix} = o(1),
\]

\[
\mathbb{E}
\begin{bmatrix}
\frac{1}{n} (\beta_j(z_1) - \beta_i^E(z_1)) \text{tr}(A_{ij}(z_1) A_{ij}(z_2)) Z_{ij} A_{ij}(z_2) Z_{ij} u_b
\end{bmatrix} = o(1),
\]

\[
\mathbb{E}
\begin{bmatrix}
\frac{1}{n} \beta_j(z_1) \beta_j(z_2) [Z_{i}^T A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_b - \text{tr}(A_{ij}(z_1) A_{ij}(z_2))]
\end{bmatrix} = o(1),
\]

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

\[
\mathcal{J}_j^{(2)} = \mathbb{E}_{j-1} \sum_{i=1}^{j-1} u_{bi} \beta_{ij}(z_1) Z_{ij} A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_b
\]

\[
= \mathbb{E}_{j-1} \sum_{i=1}^{j-1} u_{bi} \beta_{ij}(z_1) Z_{ij} A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_b
\]

\[
- \mathbb{E}_{j-1} \sum_{i=1}^{j-1} \frac{1}{n} u_{bi} \beta_{ij}(z_1) \beta_{ij}(z_2) Z_{ij} A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_b
\]

\[
= \sum_{i=1}^{j-1} u_{bi} \beta_{ij}^E(z_1) E_{j-1} Z_{ij} A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_b
\]

\[
- \sum_{i=1}^{j-1} \frac{1}{n} u_{bi} \beta_{ij}^E(z_1) \beta_{ij}^E(z_2) E_{j-1} \text{tr}(A_{ij}(z_1) A_{ij}(z_2)) Z_{ij} A_{ij}(z_2) Z_{ij} u_b \]

\[+ o_L(1).
\]

The residual term above is uniform over \( j \).

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

\[
\sup_j \mathbb{E} \left| \sum_{i=1}^{j-1} u_{bi} Z_{ij} A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_b \right|^2 = o(1), \quad \text{(S.3.7)}
\]

\[
\sup_j \mathbb{E} \left| \sum_{i=1}^{j-1} u_{bi} Z_{ij} A_{ij}(z_2) Z_{ij} u_b \right|^2 = o(1). \quad \text{(S.3.8)}
\]

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

Like \( Z_{ij} \), we define \( Z_i^{ij'} \) to be \( Z \) with \( z_i, z_{i'}, z_j \) replaced by 0. Similarly, define \( Z_i^{ii'j'} \), \( i < j, i' < j \) by replacing the \( i \)-th and \( i' \)-th column of \( Z_j \) by the 0 vectors. Further, define \( A_{ii'j} \) and \( A_{ii'j} \) to be the counterparts of \( A_{ij} \) and \( A_{ij} \) with \( Z_{ij} \) replaced by \( Z_{ii'j} \), or with
$Z_{ij}$ replaced by $Z_{i'j'}$. Accordingly, define

$$\beta_{i'j'}(z) = \frac{1}{1 + n^{-1} \text{tr} A_{i'j'}(z)},$$

$$\eta_{i'j'}(z) = \frac{1}{n} z_i^T A_{i'j'}(z) z_{i'} - \frac{1}{n} \text{tr} A_{i'j'}(z),$$

$$\beta_{i'j'}(z) = \frac{1}{1 + n^{-1} z_i^T A_{i'j'}(z) z_{i'}},$$

$$\beta_{i'j'}(z) = \frac{1}{1 + n^{-1} z_i^T A_{i'j'}(z) z_{i'}}.$$

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

$$\sum_{i=1}^{j-1} u_{n,i}^2 \mathbb{E} \left| z_i^T A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_0 \right|^2 \leq \sum_{i=1}^{N} u_{n,i}^2 \mathbb{E} \left| z_i^T A_{ij}(z_1) A_{ij}(z_2) Z_{ij} u_0 \right|^2 = O(n^{-1}).$$

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

$$\mathbb{E} \left| z_i^T A_{ij}(z_1) A_{i'j'}(z_2) Z_{ij} u_0 \right|^2 \leq \mathbb{E} \left| z_i^T A_{i'j'}(z_1) A_{i'j'}(z_2) Z_{i'j'} u_0 \right|^2 + |u_{n,i'}|^2 \mathbb{E} \left| z_i^T A_{i'j'}(z_1) A_{i'j'}(z_2) Z_{i'j'} u_0 \right|^2 + \mathbb{E} \left| z_i^T A_{i'j'}(z_1) A_{i'j'}(z_2) Z_{i'j'} u_0 \right|^2$$

$$\mathbb{E} \left| z_i^T A_{i'j'}(z_1) A_{i'j'}(z_2) Z_{i'j'} u_0 \right|^2$$

$$\mathbb{E} \left| z_i^T A_{i'j'}(z_1) A_{i'j'}(z_2) Z_{i'j'} u_0 \right|^2$$

$$\mathbb{E} \left| z_i^T A_{i'j'}(z_1) A_{i'j'}(z_2) Z_{i'j'} u_0 \right|^2$$
we get
\[
\begin{align*}
\mathbf{A}_{ij}(x_1)\mathbf{z}_i \mathbf{z}_j^T \mathbf{A}_{ij}(x_2) \beta_{ij}(x_2) \mathbf{z}_i \mathbf{z}_j^T \mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b
\end{align*}
\]

\[= o(1).\]

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

The second term in the expansion of \(J_j\) is such that,
\[
\begin{align*}
\mathbb{E}\left[ \left| \text{tr}\{\mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\} - \text{tr}\{\mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\} \right| \mathbf{z}_i^T \mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b \right]^2 \\
\leq K\mathbb{E}\left[ \mathbf{z}_i^T \mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b \right]^2 = O(1),
\end{align*}
\]
due to an inequality similar to Lemma S.11.

Therefore,
\[
\begin{align*}
j-1 \\
\sum_{i=1}^{j-1} \frac{1}{n} u_{bi} \beta_{ij}(x_1) \beta_{ij}(x_2) & \mathbb{E}_{j-1} \text{tr}\{\mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\} \mathbf{z}_i^T \mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b \\
= \beta_{ij}(x_1) \beta_{ij}(x_2) & \mathbb{E}_{j-1} \left[ \frac{1}{n} \text{tr}\{\mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\} \sum_{i=1}^{j-1} u_{bi} \mathbf{z}_i^T \mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b \right] + o_{L_1}(1).
\end{align*}
\]

By (S.3.7), (S.3.8), and Lemma S.11,
\[
\begin{align*}
\mathbb{E}\left[ \frac{1}{n} \text{tr}\{\mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\} \sum_{i=1}^{j-1} u_{bi} \mathbf{z}_i^T \mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b \right]^2 \\
\leq K\mathbb{E}\left[ \sum_{i=1}^{j-1} u_{bi} \mathbf{z}_i^T \mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b \right]^2 = o(1).
\end{align*}
\]

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

As for \(J_j^{(3)}\), due to
\[
\begin{align*}
\mathbb{E} |\theta_{ij}(x_1)|^2 &= O(n^{-1}), \\
\mathbb{E} |\mathbf{z}_i^T \mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b|^2 &= O(1),
\end{align*}
\]
we get
\[
\begin{align*}
J_j^{(3)} = \mathbb{E}_{j-1} \sum_{i=j+1}^N u_{bi} \mathbf{z}_i^T \mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b \\
= \mathbb{E}_{j-1} \sum_{i=j+1}^N u_{bi} \beta_{ij}(x_1) \mathbf{z}_i^T \mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b \\
= \mathbb{E}_{j-1} \sum_{i=j+1}^N u_{bi} \beta_{ij}^T \mathbf{z}_i^T \mathbf{A}_{ij}(x_1)\mathbf{A}_{ij}(x_2)\mathbf{Z}_{ij} \mathbf{u}_b
\end{align*}
\]
As for $J$, this subsection shows, as $n \to \infty$. The limit of (S.3.3) has been proved.

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

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

$$
\sum_{j=1}^{N} E_{j-1} \mathcal{H}_{[1]}(z_j, J) \mathcal{H}_{[2]}(z_2, J)
\quad \quad,
$$

$$
= n \beta^R(z_1) \beta^R(z_2) \sum_{i=1}^{N-1} u_{ai} \sum_{j=i+1}^{N} u_{aj} u_{bj} \text{tr} \left[ E_j A_j(z_1) E_j A_j(z_2) \right] 
+ o_p(1).
$$

(S.3.6) has been proved.

Following notation defined in Section S.3.1.3,

$$
E_{j-1} \left[ E_j [z_j^T A_j(z_1) Z_j u_j] E_j [z_j^T A_j(z_2) Z_j u_j] \right]
$$

$$
= E_{j-1} [u_j^T Z_j^T A_j(z_1) A_j(z_2) Z_j u_j]
$$

$$
= \sum_{i=1}^{j-1} u_{ai} u_{bi} E_{j-1} [z_i^T A_j(z_1) A_j(z_2) Z_j u_b]
+ \sum_{i=1}^{j-1} u_{ai} E_{j-1} [z_i^T A_j(z_1) A_j(z_2) Z_j u_b]
+ \sum_{i=j+1}^{N} u_{ai} E_{j-1} [z_i^T A_j(z_1) A_j(z_2) Z_j u_b]
$$

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

$$
\sup_j \left| J_j^{(4)} - \beta^R(z_1) \beta^R(z_2) \sum_{i=1}^{j-1} u_{ai} u_{bi} \text{tr} E_j [A_j(z_1) A_j(z_2)] \right|^2 = o(1),
$$

$$
\sup_j |J_j^{(5)}|^2 = o(1),
$$

$$
\sup_j |J_j^{(6)}|^2 = o(1).
$$

As for $J_j^{(4)}$, due to similar arguments to those for $J_j^{(1)}$, we need to show

$$
\mathbb{E} \left[ \sum_{i=1}^{j-1} u_{ai} u_{bi} \beta_{ij}(z_1) \beta_{ij}(z_2) z_i^T A_{ij}(z_1) A_{ij}(z_2) z_i \right],
$$
It suffices to show
\[
\mathbb{E}\left(\frac{1}{n^2} \left| \beta_{ij}(z_1) - \beta_{ij}(z_2) \right|^2 \right) = o(1), \quad \ell \geq 2,
\]
\[
\mathbb{E}\left(\frac{1}{n} \left| \beta_{ij}(z_1) - \beta_{ij}(z_2) \right|^2 \right) = o(1), \quad \ell \geq 2,
\]
\[
\mathbb{E}\left(\frac{1}{n} \left| A_{ij}(z_1) - A_{ij}(z_2) \right|^2 \right) = o(1), \quad \ell \geq 2.
\]

Now consider \( J_j^{(5)} \),
\[
J_j^{(5)} = \mathbb{E}_{j-1} \sum_{i=1}^{j-1} u_{ai} \beta_{ij}(z_1) z_i^T A_{ij}(z_1) A_{ij}(z_2) z_i u_b
= \mathbb{E}_{j-1} \sum_{i=1}^{j-1} u_{ai} \beta_{ij}(z_1) z_i^T A_{ij}(z_1) A_{ij}(z_2) z_i z_i^T A_{ij}(z_2) z_i u_b
- \mathbb{E}_{j-1} \sum_{i=1}^{j-1} \frac{1}{n} u_{ai} \beta_{ij}(z_1) \beta_{ij}(z_2) z_i^T A_{ij}(z_1) A_{ij}(z_2) z_i z_i^T A_{ij}(z_2) z_i u_b
+ \sum_{i=1}^{j-1} \mathbb{E}_{j-1} \sum_{i=1}^{j-1} \frac{1}{n} u_{ai} \beta_{ij}(z_1) \beta_{ij}(z_2) \left( z_i^T A_{ij}(z_1) A_{ij}(z_2) z_i u_b \right)
- \sum_{i=1}^{j-1} \mathbb{E}_{j-1} \sum_{i=1}^{j-1} \frac{1}{n} u_{ai} \beta_{ij}(z_1) \beta_{ij}(z_2) \left( z_i^T A_{ij}(z_2) z_i u_b \right) + o_{L_2}(1).
\]

The last step is due to
\[
\mathbb{E}\left(\frac{1}{n} \left| z_i^T A_{ij}(z_1) A_{ij}(z_2) z_i - \frac{1}{n} \text{tr}\{A_{ij}(z_1) A_{ij}(z_2)\} \right|^4 \right) = o(1),
\]
\[
\mathbb{E}\left| z_i^T A_{ij}(z_2) z_i u_b \right|^4 = O(1).
\]

Together with (S.3.7) and (S.3.8),
\[
\sup_j \mathbb{E}\left| J_j^{(5)} \right|^2 = o(1).
\]

The proof of \( J_j^{(6)} = o_{L_2}(1) \) is very similar to that of \( J_j^{(3)} \). We omit details.
S.3.1.5. The limit of (S.3.4)

In this section, we will show

$$\sum_{j=1}^{N} E_{j-1} \mathcal{H}(x_1, j) \mathcal{H}(x_2, j) \rightarrow 0, \quad \text{in probability.} \quad (S.3.10)$$

Lemma S.19 indicates

$$E_{j-1} \left[ \mathbb{E}(z_{j}^T A_{j}(x_1) Z_j \alpha_j) \right] = \frac{E_{z_{1}}^{j}}{n} \sum_{i=1}^{p} [E_{j} h_i(x_2, a, j) h_i(x_2, b, j)] [E_{j} h_i(x_1, b, j)].$$

where $h_i(x_2, a, j)$ and $h_i(x_2, b, j)$ are respectively the $i$th element of $A_j(z) Z_j u_a$ and $A_j(z) Z_j u_b$. Therefore,

$$(S.3.4) = -\mathbb{E} \sum_{j=1}^{N} \sum_{i=1}^{p} [E_{j} h_i(x_2, a, j) h_i(x_2, b, j)] [E_{j} h_i(x_1, b, j)].$$

It is sufficient to show

$$\sup_{1 \leq i \leq N} \mathbb{E} \sum_{j=1}^{p} \left| [E_{j} h_i(x_2, a, j) h_i(x_2, b, j)] [E_{j} h_i(x_1, b, j)] \right| \rightarrow 0.$$ 

We first show $h_i(x_2, a, j)$ concentrates around its mean. Specifically, we have

$$\sup_{i \leq p; j \leq N} \mathbb{E} \left| h_i(x_2, a, j) - E h_i(x_1, a, j) \right|^2 = O(n^{-1}), \quad (S.3.11)$$

$$\sup_{i \leq p; j \leq N} \mathbb{E} \left| h_i(x_2, a, j) - E h_i(x_1, a, j) \right|^4 = o(n^{-1}). \quad (S.3.12)$$

To show (S.3.11),

$$h_i(x, a, j) - E h_i(x, a, j)$$

$$= \sum_{\ell=1, \ell \neq j}^{N} \left[ \mathbb{E}_{\ell} - \mathbb{E}_{\ell-1} \right] h_i(x, a, j)$$

$$= \sum_{\ell=1, \ell \neq j}^{N} \left[ \mathbb{E}_{\ell} - \mathbb{E}_{\ell-1} \right] e_{i}^T \left[ A_{j}(x) Z_j u_a - A_{\ell j}(x) Z_{\ell j} u_a \right]$$

$$= \sum_{\ell=1, \ell \neq j}^{N} \left[ \mathbb{E}_{\ell} - \mathbb{E}_{\ell-1} \right] \left\{ u_{a \ell} e_{i}^T A_{\ell j}(x) z_{\ell} - \frac{1}{n} e_{i}^T A_{\ell j}(x) z_{\ell} A_{\ell j}(x) Z_{\ell j} u_a \beta_{\ell j}(x) \right. \right.$$ 

$$\left. - \frac{u_{a \ell}}{n} e_{i}^T A_{\ell j}(x) z_{\ell} z_{\ell j}^T A_{\ell j}(x) z_{\ell j}(x) \right\}.$$
By Lemma S.14 and Lemma S.10, for all $i, j, \ell$,

$$\mathbb{E}|e_i^T A_{\ell j}(z)z_i|^4 = \mathbb{E}|z_i^T A_{\ell j}(z)e_i e_i^T A_{\ell j}(z)z_i|^2 \leq \mathcal{K} \mathbb{E}\|A_{\ell j}(z)\|^4 = O(1).$$

$$\mathbb{E}\frac{1}{n} e_i^T A_{\ell j}(z)z_i^4 = n^{-8} \mathbb{E}|z_i^T A_{\ell j}(z)e_i e_i^T A_{\ell j}(z)z_i|^4 \leq \mathcal{K} \mathbb{E}\|A_{\ell j}(z)\|^8 |z_i|^{4n^{-6} + n^{-8}} = o(n^{-6}).$$

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

$$\mathbb{E}\frac{1}{n} e_i^T A_{\ell j}(z)z_i z_i^T A_{\ell j}(z)z_i u_i|^2 = O(n^{-2}),$$

$$\mathbb{E}\frac{1}{n} e_i^T A_{\ell j}(z)z_i z_i^T A_{\ell j}(z)z_i u_i|^4 = o(n^{-2}),$$

$$\mathbb{E}\frac{u_a e_i^T A_{\ell j}(z)z_i z_i^T A_{\ell j}(z)z_i|^2}{n} \leq \mathcal{K} \|u_i\|^2_{\text{max}} = O(n^{-2}),$$

$$\mathbb{E}\frac{u_a e_i^T A_{\ell j}(z)z_i z_i^T A_{\ell j}(z)z_i|^4}{n} = o(n^{-3}).$$

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

$$\mathbb{E} \sum_{i=1}^p \left[ \left| \mathbb{E}_j h_i(x_2, a, j)h_i(x_2, b, j) \right| \left| \mathbb{E}_j h_i(x_1, b, j) - \mathbb{E} h_i(x_1, b, j) \right| \right]$$

$$\leq \mathbb{E} \sum_{i=1}^p \left[ \left| \mathbb{E}_j h_i(x_2, a, j)\mathbb{E} h_i(x_2, b, j) \right| \left| \mathbb{E}_j h_i(x_1, b, j) - \mathbb{E} h_i(x_1, b, j) \right| \right]$$

$$+ \mathbb{E} \sum_{i=1}^p \left[ \left| \mathbb{E}_j h_i(x_2, a, j) \{ h_i(x_2, b, j) - \mathbb{E} h_i(x_2, b, j) \} \right| \times$$

$$\left| \mathbb{E}_j h_i(x_1, b, j) - \mathbb{E} h_i(x_1, b, j) \right| \right].$$

The second term above is such that

$$\left( \mathbb{E} \sum_{i=1}^p \left[ \left| \mathbb{E}_j h_i(x_2, a, j) \{ h_i(x_2, b, j) - \mathbb{E} h_i(x_2, b, j) \} \right| \times$$

$$\left( \mathbb{E}_j h_i(x_1, b, j) - \mathbb{E} h_i(x_1, b, j) \right) \right] \right)^2$$

$$\leq \sum_{i=1}^p \mathbb{E}[h_i(x_2, a, j)]^2 \sum_{i=1}^p \mathbb{E}[h_i(x_2, b, j) - \mathbb{E} h_i(x_2, b, j)] \times$$

$$\left| \mathbb{E}_j h_i(x_1, b, j) - \mathbb{E} h_i(x_1, b, j) \right|^2$$

$$\leq \sum_{i=1}^p \mathbb{E}[h_i(x_2, a, j)]^2 \sum_{i=1}^p \left( \mathbb{E}[h_i(x_2, b, j) - \mathbb{E} h_i(x_2, b, j)]^4 \right)^{1/2} \times$$

$$\left( \mathbb{E}_j h_i(x_1, b, j) - \mathbb{E} h_i(x_1, b, j) \right)^{1/2}.$$
The last line is due to (S.3.12) and

\[
\sum_{i=1}^{p} \mathbb{E}|h_i(x_2, a, j)|^2 = \mathbb{E}u_a^T Z_j^T A_j(x_2) A_j(x_2) Z_j u_a = O(1). \tag{S.3.13}
\]

The first term is such that

\[
\left( \mathbb{E} \left[ \sum_{i=1}^{p} \left| \mathbb{E} h_i(x_2, a, j) \mathbb{E} h_i(x_2, b, j) \right| \mathbb{E} h_i(x_1, b, j) - \mathbb{E} h_i(x_1, b, j) \right]^2 \right)^{1/2} \leq k \frac{1}{n} \sum_{i=1}^{p} \mathbb{E}|h_i(x_2, b, j)|^2 = o(1),
\]

where the last step is due to (S.3.11). Therefore,

\[
\sup_{1 \leq j \leq N} \mathbb{E} \left| \sum_{i=1}^{p} \left| \mathbb{E} h_i(x_2, a, j) h_i(x_2, b, j) \right| \left| \mathbb{E} h_i(x_1, b, j) \right| \right| = \sup_{1 \leq j \leq N} \mathbb{E} \left| \sum_{i=1}^{p} \left| \mathbb{E} h_i(x_2, a, j) h_i(x_2, b, j) \right| \left| \mathbb{E} h_i(x_1, b, j) \right| + o(1),
\]

which is a consequence of (S.3.11).

Therefore,

\[
\sup_{1 \leq j \leq N} \mathbb{E} \left| \sum_{i=1}^{p} \left| \mathbb{E} h_i(x_2, a, j) h_i(x_2, b, j) \right| \right| \leq \sup_{j \in N : j \leq p} \left| \mathbb{E} h_i(x_1, b, j) \right| \sup_{1 \leq j \leq N} \mathbb{E} \left| h_i(x_2, a, j) h_i(x_2, b, j) \right| \leq \sup_{j \in N : j \leq p} \left| \mathbb{E} h_i(x_1, b, j) \right| \sup_{1 \leq j \leq N} \left( \mathbb{E} u_a^T Z_j^T A_j(x_2) A_j(x_2) Z_j u_a \right)^{1/2} \times \left( \mathbb{E} u_a^T Z_j^T A_j(x_2) A_j(x_2) Z_j u_a \right)^{1/2}.
\]

It implies that we only need to show

\[
\sup_{j \in N : j \leq p} \left| \mathbb{E} h_i(x_1, b, j) \right| \to 0. \tag{S.3.14}
\]
For all $j$,

$$
\mathbb{E} h_i(z, b, j) = \sum_{\ell \neq j} u_{\ell b} \mathbb{E} e_i^T A_{12}(z) \beta_{12}(z)
$$

$$
= \sum_{\ell \neq j} u_{\ell b} \mathbb{E} e_i^T A_{12}(z) \beta_{12}(z) \{1 + \frac{1}{n} \text{etr} A_{12}(z) \}^{-1} \left( \frac{1}{n} z_i^T A_{12} z_1 - \frac{1}{n} \text{etr} A_{12}(z) \right).
$$

By Lemma S.13 and Lemma S.14,

$$
\mathbb{E} | e_i^T A_{12}(z) \beta_{12}(z) \{1 + \frac{1}{n} \text{etr} A_{12}(z) \}^{-1} \left( \frac{1}{n} z_i^T A_{12} z_1 - \frac{1}{n} \text{etr} A_{12}(z) \right) |
$$

$$
\leq \mathbb{K} | e_i^T A_{12}(z) \beta_{12}(z) \left( \frac{1}{n} z_i^T A_{12} z_1 - \frac{1}{n} \text{etr} A_{12}(z) \right) | = o(1).
$$

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

We have shown

$$
\sup_{1 \leq j \leq N} \mathbb{E} \sum_{i=1}^p \left[ | \mathbb{E} h_i(x_2, a, j) h_i(x_2, b, j) | | \mathbb{E} h_i(x_1, b, j) | \right] \rightarrow 0.
$$

Thus,

$$
\sum_{j=1}^N \mathbb{E} h_i(x_1, j) \mathcal{H}_{[3]}(x_2, j) \longrightarrow 0, \quad \text{in probability.}
$$

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

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

$$
\sum_{j=1}^N \mathbb{E} h_i(x_1, j) \mathcal{H}_{[3]}(x_2, j)
$$

$$
= \langle \beta^E(z_1) \beta^E(z_2) \rangle^2 \frac{1}{n} \sum_{j=1}^N \{ \text{tr} \mathbb{E} A_j(z_1) \mathbb{E} A_j(z_2) \}^2 \times (S.3.15)
$$

$$
\left( \sum_{i=1}^{j-1} u_{a_i u_{b_i}} \right)^2 + \sum_{i=1}^{j-1} u_{a_i}^2 \sum_{i=1}^{j-1} u_{b_i}^2 + o_p(1).
$$

By Lemma S.20,

$$
\sum_{j=1}^N \mathbb{E} h_i(x_1, j) \mathcal{H}_{[3]}(x_2, j)
$$

$$
= \langle \varepsilon_{11}^2 - 3 \rangle \frac{1}{n} \sum_{j=1}^N \sum_{i=1}^p \mathbb{E} h_i(x_1, a, j) h_i(x_1, b, j) \mathbb{E} h_i(x_2, a, j) h_i(x_2, b, j) \right].
$$
where similar to $h_i(x, a, j)$ and $h_i(z, b, j)$, $b_i(x, a, j)$ and $h_i(z, b, j)$ are respectively the $i$th elements of $\mathbf{A}_j(x)\mathbf{Z}_j\mathbf{u}_a$ and $\mathbf{A}_j(x)\mathbf{Z}_j\mathbf{u}_b$.

Consider $J_j^{(7)}$ first. The target is to show

$$\frac{1}{n} \sum_{j=1}^{N} \left[ \mathbb{E}_{x} \left[ h_i(z_1, a, j)h_i(z_1, b, j) \right] \right] = o_p(1). \quad (S.3.16)$$

Split it into four terms,

$$\sum_{i=1}^{p} \left[ \mathbb{E}_{x} \left[ h_i(z_1, a, j)h_i(z_1, b, j) \right] \right]$$

$$= \sum_{i=1}^{p} \left[ \mathbb{E}_{x} \left[ h_i(z_1, a, j)h_i(z_1, b, j) \right] \mathbb{E}_{x} \left[ h_i(z_2, a, j)h_i(z_2, b, j) \right] \right]$$

$$+ \sum_{i=1}^{p} \left[ \mathbb{E}_{x} \left[ h_i(z_1, a, j)h_i(z_1, b, j) \right] \mathbb{E}_{x} \left[ h_i(z_2, a, j) - h_i(z_2, b, j) \right] \right]$$

$$+ \sum_{i=1}^{p} \left[ \mathbb{E}_{x} \left[ h_i(z_1, a, j)h_i(z_1, b, j) \right] \mathbb{E}_{x} \left[ h_i(z_2, a, j) - h_i(z_2, b, j) \right] \right]$$

$$+ \sum_{i=1}^{p} \left[ \mathbb{E}_{x} \left[ h_i(z_1, a, j)h_i(z_1, b, j) \right] \right] \times$$

$$\mathbb{E}_{x} \left[ h_i(z_2, a, j) - h_i(z_2, b, j) \right] \right]$$

$$= d_{4}^{(1)} + d_{4}^{(2)} + d_{4}^{(3)} + d_{4}^{(4)}$$

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

$$\mathbb{E}[d_{4}^{(1)}] \leq \sup_{i} \left[ \mathbb{E}[h_i(z_2, a, j)] \mathbb{E}[h_i(z_2, b, j)] \right] \sum_{i=1}^{p} \mathbb{E}[h_i(z_1, a, j)h_i(z_1, b, j)]$$

$$\leq \sup_{i} \left[ \mathbb{E}[h_i(z_2, a, j)] \mathbb{E}[h_i(z_2, b, j)] \right] \times$$

$$\left( \sum_{i=1}^{p} \mathbb{E}[h_i(z_1, a, j)]^{2} \sum_{i=1}^{p} \mathbb{E}[h_i(z_1, b, j)]^{2} \right)^{1/2} \rightarrow 0.$$
It suffices to prove High-dimensional general linear hypothesis via spectral shrinkage

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

\[
\mathbb{E}[d_4^{(2)}] \leq \sup_i (\mathbb{E}[h_i(x_2, a, j)]) \left[ \sum_{i=1}^{p} \mathbb{E}[h_i(x_1, a, j)h_i(x_1, b, j)]^2 \right]^{1/2} \times \left[ \sum_{i=1}^{p} \mathbb{E}[h_i(x_2, b, j) - \mathbb{E}h_i(x_2, b, j)]^2 \right]^{1/2} \longrightarrow 0.
\]

Similarly,\n
\[
\mathbb{E}[d_4^{(3)}] \leq \sup_i (\mathbb{E}[h_i(x_2, b, j)]) \left[ \sum_{i=1}^{p} \mathbb{E}[h_i(x_1, a, j)h_i(x_1, b, j)]^2 \right]^{1/2} \times \left[ \sum_{i=1}^{p} \mathbb{E}[h_i(x_2, a, j) - \mathbb{E}h_i(x_2, a, j)]^2 \right]^{1/2} \longrightarrow 0,
\]

\[
\mathbb{E}[d_4^{(4)}] \leq \left( \sum_{i=1}^{p} \mathbb{E}[h_i(x_2, a, j)] \right)^{1/2} \left( \sum_{i=1}^{p} \mathbb{E}[h_i(x_2, a, j) - \mathbb{E}h_i(x_2, a, j)]^4 \right)^{1/4} \longrightarrow 0.
\]

Consider \( \mathcal{J}_j^{(8)} \). The target is to show\n
\[
\mathbb{E}_j \sum_{i=1}^{p} h_i(x_1, b, j)h_i(x_2, a, j) \sum_{i=1}^{p} h_i(x_2, b, j)h_i(x_1, a, j) = \mathbb{E}_j \sum_{i=1}^{p} h_i(x_2, b, j)h_i(x_1, a, j) + o_{L_4}(1).
\]

(S.3.17)

First we substitute \( h_i(x_1, a, j) \) with \( \mathbb{E}_j h_i(x_1, a, j) \) in \( \mathcal{J}_j^{(8)} \) and show the resulting difference is small. That is to show\n
\[
\mathbb{E}_j \sum_{i=1}^{p} h_i(x_1, b, j)h_i(x_2, a, j) \sum_{i=1}^{p} h_i(x_2, b, j)h_i(x_1, a, j) = \mathbb{E}_j \sum_{i=1}^{p} h_i(x_2, b, j)h_i(x_1, a, j) + o_{L_4}(1).
\]

(S.3.18)

It suffices to prove\n
\[
\mathbb{E} \left| \sum_{i=1}^{p} h_i(x_1, b, j)h_i(x_2, a, j) \sum_{i=1}^{p} h_i(x_2, b, j)h_i(x_1, a, j) - \mathbb{E}_j h_i(x_1, a, j) \right| = o(1).\n\]
Due to very similar arguments to (S.3.13),

\[
\mathbb{E}\left|\sum_{i=1}^{p} h_i(z_1, b, j) \overline{h_i}(z_2, a, j)\right|^2 \\
\leq \left(\mathbb{E}\left(\sum_{i=1}^{p} |h_i(z_1, b, j)|^2\right)^2 \mathbb{E}\left(\sum_{i=1}^{p} |h_i(z_2, a, j)|^2\right)^2\right)^{1/2} \\
= \left(\mathbb{E}\left|u_b^T Z_j^T A_j(z_1) A_j(z_1) Z_j u_a\right|^2 \mathbb{E}\left|u_a^T Z_j^T A_j(z_2) A_j(z_2) Z_j u_a\right|^2\right)^{1/2} = O(1).
\]

It suffices to show

\[
\mathbb{E}\left|\sum_{i=1}^{p} h_i(z_2, b, j) [h_i(z_1, a, j) - \mathbb{E}_j h_i(z_1, a, j)]\right|^2 = o(1).
\]

Fixing \( j \), we consider the \( \sigma \)-algebra generated by

\[
\{z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_N\} \cup \{z_{j+1}, \ldots, z_N\}.
\]

Define a filtration (in \( L \)) as

\[
\sigma(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_N, z_l, \ldots, z_N), \quad l = j + 1, \ldots, N.
\]

Rewrite \( \sum_{i=1}^{p} h_i(z_2, b, j) [h_i(z_1, a, j) - \mathbb{E}_j h_i(z_1, a, j)] \) 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

\[
\sigma(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_N, z_l, \ldots, z_N)
\]

for \( l = j + 1, \ldots, N \).

\[
\sum_{i=1}^{p} h_i(z_2, b, j) [h_i(z_1, a, j) - \mathbb{E}_j h_i(z_1, a, j)]
\]

\[
= \sum_{l=j+1}^{N} (\mathbb{E}_{j,l} - \mathbb{E}_{j,l-1}) u_b^T Z_j^T A_j(z_2) [A_j(z_1) Z_j u_a - A_{ij}(z_1) Z_{ij} u_a]
\]

\[
= \sum_{l=j+1}^{N} (\mathbb{E}_{j,l} - \mathbb{E}_{j,l-1}) u_a^T u_b^T Z_j^T A_j(z_2) A_{ij}(z_1) Z_{ij} \beta_{ij}(z_1)
\]

\[
- \frac{1}{n} \sum_{l=j+1}^{N} (\mathbb{E}_{j,l} - \mathbb{E}_{j,l-1}) u_b^T Z_j^T A_j(z_2) A_{ij}(z_1) Z_{ij} u_a \beta_{ij}(z_1)
\]

\[= d_5^{(1)} + d_5^{(2)}, \quad \text{say.} \]
By Lemma S.8, Lemma S.10, and Lemma S.16,
\[
E|d_5^{(1)}|^2 \leq \mathcal{K} \sum_{l=j+1}^{N} |u_{al}|^2 E|u_l^T Z_l^T A_j(z_2) A_{l_j}(z_1)z_l l_{lj}(z_1)|^2 = O(n^{-1}),
\]
\[
E|d_5^{(2)}|^2 \leq \mathcal{K} \frac{1}{n^2} \sum_{l=j+1}^{N} E|u_l^T Z_l^T A_j(z_2) A_{l_j}(z_1)z_l l_{lj}(z_1)|^2
\]
\[
\leq \mathcal{K} \frac{1}{n^2} \sum_{l=j+1}^{N} \{E|u_l^T Z_l^T A_j(z_2) A_{l_j}(z_1)z_l l_{lj}(z_1)|^2 \}^{1/2} = O(n^{-1}).
\]
The proof of (S.3.18) is complete.

Finally, we proved
\[
E \left[ \sum_{i=1}^{p} h_i(x_1, j) \right] = O(1).
\]
It can be done along very similar lines to the proof of (S.3.18). Therefore, we omit details.

In Section S.3.1.4, we proved \( L_2 \)-convergence of \( E_{j-1} u_b^T Z_j^T A_j(z_1)A_j(z_2) Z_j u_a \). Therefore, it is straightforward that
\[
E_j \sum_{i=1}^{p} h_i(x_1, j) \beta^2(z_1) \beta^2(z_2) \text{tr} E_j [A_j(z_1)A_j(z_2)] \left( \sum_{i=1}^{j-1} u_{ai} u_{bi} \right)^2 + O_{L_2}(1).
\]
Consider \( J_j^{(9)} \). Repeat the arguments for \( J_j^{(8)} \) with limited modifications,
\[
E_j \sum_{i=1}^{p} h_i(x_1, j) \beta^2(z_2) \text{tr} E_j [A_j(z_1)A_j(z_2)] \left( \sum_{i=1}^{j-1} u_{ai} u_{bi} \right)^2 + O_{L_2}(1).
\]
The proof of (S.3.15) is complete.
We next try to find the limit of the right-hand side.

Recall the definition of \( m_p^0(z) \) in Section S.2. It is proved in Bai and Silverstein (2004, (2.18)) that

\[
\sum_{k=1}^{3} \sum_{k'=1}^{3} \sum_{j=2}^{N} E_{i-1} \cdot H_{i}[k](z_1) \cdot H_{i}[k'](z_2) = n \beta^E(z_1) \beta^E(z_2) \sum_{j=2}^{N} \text{tr}[E_j A_j (z_1) E_j A_j (z_2)] \sum_{i=1}^{j-1} (u_{ai}^2 u_{bi}^2 + u_{ai}^2 u_{bi}^2 + 2u_{ai} u_{bj} u_{aj} u_{bi})
\]

\[
+ \frac{1}{n} \beta^E(z_1)^2 \beta^E(z_2)^2 \sum_{j=2}^{N} \{ \text{tr}[E_j A_j (z_1) E_j A_j (z_2)] \}^2 \times \left( \sum_{i=1}^{j-1} u_{ai} u_{bi} \right)^2 + \sum_{i=1}^{j-1} u_{ai}^2 \sum_{i=1}^{j-1} u_{bi}^2 \right] + o_p(1).
\]

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

Recall the definition of \( m_p^0(z) \) in Section S.2. It is proved in Bai and Silverstein (2004, (2.18)) that

\[
\text{tr}[E_j A_j (z_1) A_j (z_2)] \left( 1 - \frac{j-1}{N^2} m_p^0(z_1) m_p^0(z_2) \text{tr} \left[ (I + m_p^0(z_2) \Sigma_p)^{-1} \Sigma_p \times \right. \right.
\]

\[
(I + m_p^0(z_1) \Sigma_p)^{-1} \Sigma_p \right)] = \frac{1}{z_1 z_2} \text{tr} \left[ (I + m_p^0(z_2) \Sigma_p)^{-1} \Sigma_p (I + m_p^0(z_1) \Sigma_p)^{-1} \Sigma_p \right]
\]

\[
+ o_L(1).
\]

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

\[
\left( \frac{1}{N} \Sigma_p^{1/2} \Sigma_p^{1/2} - z I \right)^{-1} = \frac{N}{n} \left( \frac{1}{N} \Sigma_p - \frac{n}{N} z I \right)^{-1},
\]

and \( m_p^0(z) \) is continuous in \( z \).

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

\[
\frac{1}{I + N^{-1} \text{tr} [A_1(z)]} + zm_p^0(z) = O(n^{-1/2}). \quad (S.3.19)
\]

Note

\[
\frac{1}{n} m_p^0(z_1) m_p^0(z_2) \text{tr} \left[ (I + m_p^0(z_2) \Sigma_p)^{-1} \Sigma_p (I + m_p^0(z_1) \Sigma_p)^{-1} \Sigma_p \right]
\]

\[
= \gamma_n m_p^0(z_1) m_p^0(z_2) \frac{\tau^2 dF_{\tau_k}^*(\tau)}{[1 + \tau m_p^0(z_1)][1 + \tau m_p^0(z_2)]}
\]

\[
= 1 + \frac{m_p^0(z_1) m_p^0(z_2)(z_1 - z_2)}{m_p^0(z_2) - m_p^0(z_1)}.
\]
High-dimensional general linear hypothesis via spectral shrinkage

We get
\[ \mathbb{E} \left[ \beta^E(z_1) \beta^E(z_2) \right] \frac{1}{n} \text{tr} \left[ \mathbb{E}_j A_j (z_1) \mathbb{E}_j A_j (z_2) \right] - \frac{D}{1 - \frac{j}{N} D} = o(1), \]
where
\[ D = 1 + \frac{m_0^0(z_1) m_0^0(z_2) (z_1 - z_2)}{m_0^0(z_2) - m_0^0(z_1)}. \]
Define
\[ P_j = \frac{D}{1 - \frac{j}{N} D}. \]
Therefore,
\[ \sum_{i=1}^3 \sum_{j'=1}^3 \sum_{j=2}^N E_{j-1} \mathcal{H}_{[j]}(z_1, j) \mathcal{H}_{[j']}(z_2, j) \]
\[ = n^2 \sum_{j=2}^N P_{j-1} \sum_{i=1}^{j-1} (u_{ai}^2 u_{bi}^2 + u_{ai}^2 u_{bi}^2 + 2u_{ai} u_{bi} u_{aj} u_{bi}) \]
\[ + n \sum_{j=2}^N P_{j-1}^2 \left[ (\sum_{i=1}^{j-1} u_{ai} u_{bi})^2 + \sum_{i=1}^{j-1} u_{ai}^2 \sum_{i=1}^{j-1} u_{bi}^2 \right] + o_p(1) \]
\[ = N^2 \sum_{j=1}^N P_j \sum_{i=1}^j (u_{ai}^2 u_{bi}^2 + u_{ai}^2 u_{bi}^2 + 2u_{ai} u_{bi} u_{aj} u_{bi}) \]
\[ + N \sum_{j=1}^N P_j^2 \left[ (\sum_{i=1}^j u_{ai} u_{bi})^2 + \sum_{i=1}^j u_{ai}^2 \sum_{i=1}^j u_{bi}^2 \right] + o_p(1). \]

The convergence of \( \{ \beta^E(z_1) \beta^E(z_2) \} \frac{1}{n} \text{tr} \left[ \mathbb{E}_j A_j (z_1) \mathbb{E}_j A_j (z_2) \right] \) follows from the fact that
\[ \beta^E(z_1) \beta^E(z_2) \frac{1}{n} \text{tr} \left[ \mathbb{E}_j A_j (z_1) \mathbb{E}_j A_j (z_2) \right] \]
is bounded for any fixed \( z_1 \) and \( z_2 \) with non-zero imaginary part.

Define
\[ \mathcal{O}_\ell = \sum_{i=1}^\ell (u_{ai}^2 u_{bi}^2 + u_{ai}^2 u_{bi}^2 + 2u_{ai} u_{bi} u_{al} u_{bi}). \]

The following result indicates \( \sum_{i=1}^j u_{ai} u_{bi} \) is approximately the sum of \( \mathcal{O}_\ell, \ell = 1, \ldots, j. \)

\[ \sum_{\ell=1}^j \mathcal{O}_\ell = \sum_{\ell=1}^j \sum_{i=1}^\ell (u_{ai}^2 u_{bi}^2 + u_{ai}^2 u_{bi}^2 + 2u_{ai} u_{bi} u_{al} u_{bi}). \]
\[
\begin{align*}
&= \sum_{i \leq \ell} (u_{ai}^2 u_{bi}^2 + u_{al}^2 u_{bl}^2 + 2u_{ai}u_{bl} u_{al} u_{bi}) \\
&= \sum_{i > \ell} (u_{ai}^2 u_{bi}^2 + u_{al}^2 u_{bl}^2 + 2u_{ai}u_{bl} u_{al} u_{bi}) \\
&= \frac{1}{2} \sum_{i=1}^{j} \sum_{\ell=1}^{j} (u_{ai}^2 u_{bi}^2 + u_{al}^2 u_{bl}^2 + 2u_{ai}u_{bl} u_{al} u_{bi}) + 2 \sum_{i=1}^{j} u_{ai}^2 u_{bi} \\
&= \sum_{i=1}^{j} \sum_{\ell=1}^{j} (u_{ai}^2 u_{bi}^2 + u_{ai}u_{bl} u_{al} u_{bi}) + O(n^{-3}) \\
&= (\sum_{i=1}^{j} u_{ai}u_{bi})^2 + \sum_{i=1}^{j} u_{ai}^2 \sum_{i=1}^{j} u_{bi} + O(n^{-3}).
\end{align*}
\]

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,

\[
\frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_i^2 \leq \frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_i \mathcal{P}_{i+1} = \sum_{i=1}^{j} (\mathcal{P}_{i+1} - \mathcal{P}_i) = -\mathcal{P}_0 + \mathcal{P}_j + O(n^{-1}),
\]

\[
\frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_i^2 \geq \frac{1}{N} \sum_{i=1}^{j} \mathcal{P}_{i-1} \mathcal{P}_i = \sum_{i=1}^{j} (\mathcal{P}_i - \mathcal{P}_{i-1}) = -\mathcal{P}_0 + \mathcal{P}_j.
\]

It follows that

\[
\begin{align*}
&= \sum_{i=1}^{3} \sum_{i'=1}^{3} \sum_{j=2}^{N} E_{j-1} \mathcal{H}_{1}(z_1) \mathcal{H}_{1}(z_2) \\
&= N^2 \sum_{j=1}^{N} \mathcal{P}_{j-1} \mathcal{O}_j + N \sum_{j=1}^{N} \mathcal{P}_j^2 \sum_{\ell=1}^{j} \mathcal{O}_\ell + o_p(1) \\
&= N \sum_{j=1}^{N} \mathcal{O}_j (\sum_{\ell=1}^{j} \mathcal{P}_\ell^2 + N\mathcal{P}_0) + N \sum_{j=1}^{N} \mathcal{P}_j^2 \sum_{\ell=1}^{j} \mathcal{O}_\ell + o_p(1) \\
&= N^2 \mathcal{P}_0 \sum_{j=1}^{N} \mathcal{O}_j + N \sum_{\ell \leq j} \mathcal{O}_j \mathcal{P}_\ell^2 + N \sum_{\ell > j} \mathcal{O}_j \mathcal{P}_\ell^2 + o_p(1) \\
&= N^2 \mathcal{P}_0 \sum_{j=1}^{N} \mathcal{O}_j + N \sum_{\ell=1}^{N} \mathcal{O}_j \mathcal{P}_\ell^2 + o_p(1)
\end{align*}
\]
High-dimensional general linear hypothesis via spectral shrinkage

\[ N^2 \sum_{j=1}^{N} O_j \left[ \frac{1}{N} \sum_{j=1}^{N} P_j^2 + P_0 \right] + o_p(1) \]

\[ = N^2 P_N \sum_{j=1}^{N} O_j + o_p(1) \]

\[ = N^2 \frac{D}{1 - D} \left[ \left( \sum_{i=1}^{N} u_{ai} u_{bi} \right)^2 + \sum_{i=1}^{N} u_{ai}^2 \sum_{i=1}^{N} u_{bi}^2 \right] + o_p(1) \]

\[ = \frac{D}{1 - D} \left[ ||a||^2 ||b||^2 + (a^T b)^2 \right] + o_p(1). \]

Next, we express \( D_n \) in terms of \( \delta(z_1, z_2, \gamma) \) and \( \Theta(z, \gamma) \). Using the results in Section S.2, we have

\[ zm^0_n(z) = -\Theta(z, \gamma) + o(1). \]

It follows

\[ \frac{D}{1 - D} = \delta(z_1, z_2, \gamma) \Theta^{-1}(z_1, \gamma) \Theta^{-1}(z_2, \gamma) + o(1). \]

Thus,

\[ \sum_{k=1}^{3} \sum_{k'}^{3} \sum_{j=2}^{N} \beta_k^R(z_1) \beta_{k'}^R(z_2) E_{j-1} H_{k|j}(z_1, j) H_{k'|j}(z_2, j) \xrightarrow{P} \]

\[ \Theta^{-2}(z_1, \gamma) \Theta^{-2}(z_2, \gamma) \delta(z_1, z_2, \gamma) ||a||^2 ||b||^2 + (a^T b)^2. \]

The proof of finite dimensional convergence of \( G_n^{(1)}(z) \) is complete.

S.3.2. Tightness of \( G_n^{(1)}(z, a, b) \)

In view of our smoothing strategy, to show the tightness, we first consider the case \( z = u + iv \in C^+ \), that is \( |v| \geq \rho_n \).

Recall the definition

\[ G_n(z, a, b) = \begin{cases} 
    u_n^T Z^T A(z) Z u_b, & |v| \geq \rho_n, \\
    \frac{u_n^T Z^T A(u + i\rho_n) Z u_b}{2\rho_n} + \frac{v + \rho_n}{2\rho_n} u_n^T Z^T A(u - i\rho_n) Z u_b, & |v| < \rho_n,
\end{cases} \]

where \( z = u + iv \).

\[ G_n^{(1)}(z, a, b) = n^{1/2} [G_n(z, a, b) - E G_n(z, a, b)]. \]

We shall drop the arguments \( a \) and \( b \) in the rest of the section, since they are fixed.
We first show that $E_{\{A_pz_q}\{p_zq\}}$ is bounded on $C^+$ for all $\ell \geq 1$. As in Lemma A.1, select an arbitrary constant $D \in (\limsup_{p \to \infty} \lambda_{\max}(\Sigma_p)(1 + \sqrt{\tau}), \infty)$, and denote $G$ to be the event $\{\lambda_{\max}(\Sigma_p) \geq D\}$. Lemma A.1 says, for any positive $\ell$,

$$P(G) = o(n^{-\ell}).$$

Note, on $G^c$, all eigenvalues of $\Sigma_p$ are bounded away from $C^+$ with distance at least $\min\{|u| - D, |u|\}$. Thus, $\|\Sigma_p\| < \mathcal{K}$ for some $\mathcal{K} > 0$. On $G$, Lemma S.10 always holds. Therefore, there exists an universal constant $K$ such that $E_{\{A_pz_q\}} \leq K^\ell P_{\{A_pz_q\}}$. (S.3.21)

Similarly, we can show there exists constants $K_\ell(\beta)$ and $K(\beta^\mathbb{E})$ such that $E|\beta_j(z)|^\ell \leq K_\ell(\beta)$, (S.3.22) $E|\beta^\mathbb{E}(z)| \leq K(\beta^\mathbb{E})$. (S.3.23)

Another useful result is, on $G^c$, a nonrandom bound holds as

$$|\beta_j(z)| = |1 - n^{-1}z_j^T A(z)z_j| \leq 1 + K|\Sigma_p|, \quad \text{on } G^c.$$ (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 $G_n^{(1)}(a, b)$, $G_n^{(2)}(a, b)$. 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

$$E|G_n^{(1)}(z_0)|^2 \leq \mathcal{K},$$

with $z_0$ a complex number having non-zero imaginary part. In view of the construction and simplification of the martingale difference sequence in Section S.3.1, it is sufficient to show

$$E|\sum_{j=1}^N \beta_j^\mathbb{E}(z_0)\mathcal{H}(z_0, j)|^2 \leq \mathcal{K}.$$ (S.3.25)

Due to (S.3.23), we only need to show

$$\sum_{j=1}^N E|u_j z_j^T A_j(z_0)Z_j u_k|^2 \leq \mathcal{K},$$

$$\sum_{j=1}^N E|u_j\varphi_j(z_0)|^2 \leq \mathcal{K}.$$
We can prove the results using Lemma S.13, Lemma S.15, and Lemma S.16 with the nonrandom bound of \(|\mathbf{A}_j(z_0)|\) 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 \(z_1 \neq z_2 \in \{u + iv \in \mathcal{C} \text{ and } |v| \geq \rho_n\},\)

\[
\mathbb{E}\left[\frac{|G_n^{(1)}(z_1) - G_n^{(1)}(z_2)|^2}{|z_1 - z_2|^2}\right] \leq \mathcal{K}.
\]

Define

\[
\tilde{\mathbf{A}}(z_1, z_2) = \sum_p^{\frac{\ell}{2}} \left[ \left( -\frac{1}{n} \Sigma_p^\lambda \mathbf{ZZ}^T \Sigma_p^\lambda - z_1 I \right) \left( -\frac{1}{n} \Sigma_p^\lambda \mathbf{ZZ}^T \Sigma_p^\lambda - z_2 I \right) \right]^{-1} \Sigma_p^\lambda,
\]

\[
\tilde{\mathbf{A}}_j(z_1, z_2) = \sum_p^{\frac{\ell}{2}} \left[ \left( -\frac{1}{n} \Sigma_p^\lambda \mathbf{Z}_j \mathbf{Z}_j^T \Sigma_p^\lambda - z_1 I \right) \left( -\frac{1}{n} \Sigma_p^\lambda \mathbf{Z}_j \mathbf{Z}_j^T \Sigma_p^\lambda - z_2 I \right) \right]^{-1} \Sigma_p^\lambda,
\]

\[
\tilde{\mathbf{A}}_{jj}(z_1, z_2) = \sum_p^{\frac{\ell}{2}} \left[ \left( -\frac{1}{n} \Sigma_p^\lambda \mathbf{Z}_j \mathbf{Z}_j^T \Sigma_p^\lambda - z_1 I \right) \left( -\frac{1}{n} \Sigma_p^\lambda \mathbf{Z}_j \mathbf{Z}_j^T \Sigma_p^\lambda - z_2 I \right) \right]^{-1} \Sigma_p^\lambda.
\]

Along very similar lines to the proof of (S.3.21), we can show, for any \(\ell \geq 1\), there exists a constant \(\mathcal{K}_\ell(\mathbf{A})\)

\[
\sup_{z_1, z_2 \in \mathbb{C}^+} \mathbb{E}[\tilde{\mathbf{A}}(z_1, z_2)]^{\ell} \leq \mathcal{K}_\ell(\mathbf{A}),
\]

\[
\sup_{z_1, z_2 \in \mathbb{C}^+} \mathbb{E}[\tilde{\mathbf{A}}_j(z_1, z_2)]^{\ell} \leq \mathcal{K}_\ell(\mathbf{A}),
\]

\[
\sup_{z_1, z_2 \in \mathbb{C}^+} \mathbb{E}[\tilde{\mathbf{A}}_{jj}(z_1, z_2)]^{\ell} \leq \mathcal{K}_\ell(\mathbf{A}).
\]

Using the identity \(\mathbf{A}(z_1) - \mathbf{A}(z_2) = (z_1 - z_2)\tilde{\mathbf{A}}(z_1, z_2),\)

\[
\frac{G_n^{(1)}(z_1) - G_n^{(1)}(z_2)}{z_1 - z_2}
\]

\[= n^{1/2} u_a^T \mathbf{Z}^T \tilde{\mathbf{A}}(z_1, z_2) \mathbf{Z} u_b - n^{1/2} E u_a^T \mathbf{Z}^T \tilde{\mathbf{A}}(z_1, z_2) \mathbf{Z} u_b
\]

\[= n^{1/2} \sum_{j=1}^N (E_j - E_{j-1}) [u_a^T \mathbf{Z}^T \tilde{\mathbf{A}}(z_1, z_2) \mathbf{Z} u_b - u_a^T \mathbf{Z}^T \tilde{\mathbf{A}}(z_1, z_2) \mathbf{Z} u_b]
\]

\[= n^{1/2} \sum_{j=1}^N (E_j - E_{j-1}) [d_6 + d_7 + d_8 + d_9],
\]

where

\[d_6 = u_a^T \mathbf{Z}^T \tilde{\mathbf{A}}(z_1, z_2) \mathbf{Z} u_b - u_a^T \mathbf{Z}^T \tilde{\mathbf{A}}(z_1, z_2) \mathbf{Z} u_b,
\]
Next, we only present the proof of Combining Lemma S.13, Lemma S.16, and (S.3.25) we can show

\[ d_7 = u_n^T Z_j^T \tilde{A}(x_1, x_2) Z u_b - u_n^T Z_j^T \tilde{A}_j(x_1, x_2) Z u_b, \]

\[ d_8 = u_n^T Z_j^T \bar{A}_j(x_1, x_2) Z u_b - u_n^T Z_j^T \tilde{A}_j(x_1, x_2) Z u_b, \]

\[ d_9 = u_n^T Z_j^T \bar{A}_{jj}(x_1, x_2) Z u_b - u_n^T Z_j^T \bar{A}_j(x_1, x_2) Z u_b. \]

Regarding \( d_6 \),

\[ d_6 = u_{aj} z_j^T \bar{A}_{jj}(x_1, x_2) z_j \beta_j(x_1) \beta_j(x_2) \]

\[ + u_{aj} z_j^T \bar{A}_{jj}(x_1, x_2) Z_j u_b \beta_j(x_1) \]

\[ - \frac{1}{n} u_{aj} z_j^T \bar{A}_{jj}(x_1, x_2) Z_j z_j^T A_j (x_2) Z_j u_b \beta_j(x_1) \beta_j(x_2). \]

Combining Lemma S.13, Lemma S.16, and (S.3.25) we can show

\[ \sup_j \sup_{x_1, x_2 \in C^+} n^{-\ell} E[Z_j^T \bar{A}_{jj}(x_1, x_2) z_j \beta_j(x_1) \beta_j(x_2)] < \mathcal{K}_\ell, \quad \text{for all } \ell \geq 2, \quad (S.3.26) \]

\[ \sup_j \sup_{x_1, x_2 \in C^+} n^{-\ell} E[Z_j^T A_j (x_2) z_j] < \mathcal{K}_\ell, \quad \text{for all } \ell \geq 2, \quad (S.3.27) \]

\[ \sup_j \sup_{x_1, x_2 \in C^+} E[Z_j^T \bar{A}_{jj}(x_1, x_2) Z_j u_b] < \mathcal{K}, \quad (S.3.28) \]

for some constants \( \mathcal{K}_\ell \) and \( \mathcal{K} \).

Therefore, together with (S.3.22) and (S.3.24),

\[ \sup_{x_1, x_2 \in C^+} E\left[ n^{1/2} \sum_{j=1}^N (E_j - E_{j-1}) d_6 \right]^2 = O(1). \]

The other terms can be written as

\[ d_7 = -\frac{1}{n} z_j^T \tilde{A}_j(x_1, x_2) Z_j u_z z_j^T A_j (x_1) Z_j u_a \beta_j(x_1) \]

\[ - \frac{1}{n} u_{bj} z_j^T \bar{A}_j(x_1, x_2) z_j z_j^T A_j (x_1) Z_j u_a \beta_j(x_1) \]

\[ + \frac{1}{n^2} z_j^T A_j (x_1) Z_j u_z z_j^T \tilde{A}_j(x_1, x_2) z_j z_j^T A_j (x_2) Z_j u_b \beta_j(x_2) \beta_j(x_1) \]

\[ + \frac{1}{n^2} u_{bj} z_j^T A_j (x_1) Z_j u_z z_j^T \bar{A}_j(x_1, x_2) z_j z_j^T \bar{A}_j(x_2) z_j \beta_j(x_2) \beta_j(x_1); \]

\[ d_8 = -\frac{1}{n} z_j^T \tilde{A}_j(x_1, x_2) Z_j u_z z_j^T A_j (x_2) Z_j u_b \beta_j(x_2) \]

\[ - \frac{1}{n} u_{bj} z_j^T \bar{A}_j(x_1, x_2) Z_j u_z z_j^T A_j (x_2) Z_j \beta_j(x_2); \]

\[ d_9 = u_{bj} z_j^T \bar{A}_j(x_1, x_2) Z_j u_a. \]

Next, we only present the proof of

\[ \sup_{x_1, x_2 \in C^+} E\left[ n^{1/2} \sum_{j=1}^N (E_j - E_{j-1}) d_8 \right]^2 = O(1). \]
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),

\[
\sup_j \sup_{z_1, z_2} \mathbb{E}|z_j^T \tilde{A}_{jj}(z_1, z_2)Z_j u_a z_j^T A_j(x_2)Z_j u_b \beta_j(x_2)\mathbb{1}(G^c)|^2 \\
\leq \sup_j \sup_{z_1, z_2} \mathbb{E}|z_j^T \tilde{A}_{jj}(z_1, z_2)Z_j u_a z_j^T A_j(x_2)Z_j u_b|^2 \\
\leq \sup_j \mathbb{E}[|z_j^T \tilde{A}_{jj}(z_1, z_2)Z_j u_a|^4 |z_j^T A_j(x_2)Z_j u_b|^4]^{1/2} \\
= O(1).
\]

For the second term in $d_8$,

\[
\sup_j \sup_{z_1, z_2} \mathbb{E}\left|\frac{1}{n}z_j^T \tilde{A}_{jj}(z_1, z_2)Z_j u_a z_j^T A_j(x_2)Z_j u_b \beta_j(x_2)\mathbb{1}(G^c)\right|^2 \\
\leq \sup_j \sup_{z_1, z_2} \mathbb{E}\left|\frac{1}{n}z_j^T \tilde{A}_{jj}(z_1, z_2)Z_j u_a z_j^T A_j(x_2)z_j\right|^2 \\
\leq \sup_j \mathbb{E}[|z_j^T \tilde{A}_{jj}(z_1, z_2)Z_j u_a|^4 |\frac{1}{n}z_j^T A_j(x_2)z_j|^4]^{1/2} \\
= O(1).
\]

When $|v| < \rho_n$, $G_n^{(1)}(z)$ is the connected line between $G_n^{(1)}(u + i\rho_n)$ and $G_n^{(1)}(u - i\rho_n)$.

Set $z_1 = u + i\rho_n$ and $z_2 = u - i\rho_n$. All previous arguments in the subsection apply. Therefore, the slope of the connected line is bounded in expectation.

The proof of tightness of $G_n^{(1)}(z)$ is complete.
S.3.3. Convergence of $G_n^{(2)}$

Recall

$$G_n^{(2)}(z, a, b) = n^{1/2} \{ \text{E}G_n(z, a, b) - a^T b \frac{\Theta_n(z, \gamma)}{\Theta_n(z, \gamma)} \},$$

$$\Theta_n(z) = 1 + \gamma_n \frac{1}{p} \text{etr} A(z).$$

In this section, we show

$$\sup_{z \in \mathbb{C}^+} G_n^{(2)}(z, a, b) \to 0.$$ 

$$n^{1/2} \text{E}G_n(z, a, b)$$

$$= n^{1/2} \text{E} u_n^T Z^T A(x)^T X u_b$$

$$= n^{1/2} \sum_{j=1}^{N} E u_{\alpha_j} u_{\beta_j} z_j^T A_j(z) z_j \beta_j(x) + n^{1/2} \sum_{j=1}^{N} E u_{\alpha_j} z_j^T A_j(z) z_j u_{\beta_j} \beta_j(x)$$

$$= n^{1/2} \sum_{j=1}^{N} E u_{\alpha_j} u_{\beta_j} z_j^T A_j(z) z_j \beta^E(x)$$

$$- n^{1/2} \sum_{j=1}^{N} E u_{\alpha_j} u_{\beta_j} z_j^T A_j(z) z_j \beta_j(x) \beta^E(x) \left( \frac{1}{n} z_j^T A_j(z) z_j - \frac{1}{n} \text{etr} A_j(z) \right)$$

$$- n^{1/2} \sum_{j=1}^{N} E u_{\alpha_j} z_j^T A_j(z) z_j u_{\beta_j}(x) \beta^E(x) \left( \frac{1}{n} z_j^T A_j(z) z_j - \frac{1}{n} \text{etr} A_j(z) \right)$$

$$= n^{1/2} a^T b \frac{\gamma_n p^{-1} \text{etr} A_1(x)}{1 + \gamma_n p^{-1} \text{etr} A_1(x)}$$

$$+ n^{1/2} a^T b \beta^E(x) E \beta_1(x) \left( \frac{1}{n} z_1^T A_1(x) z_1 - \frac{1}{n} \text{etr} A_1(x) \right)$$

$$- n^{1/2} \sum_{j=1}^{N} E u_{\alpha_j} z_j^T A_j(z) z_j u_{\beta_j}(x) \beta^E(x) \left( \frac{1}{n} z_j^T A_j(z) z_j - \frac{1}{n} \text{etr} A_j(z) \right)$$

$$+ n^{1/2} \sum_{j=1}^{N} E u_{\alpha_j} z_j^T A_j(z) z_j u_{\beta_j}(x) (\beta^E(x))^2 \left( \frac{1}{n} z_j^T A_j(z) z_j - \frac{1}{n} \text{etr} A_j(z) \right)^2$$

$$= d_{10} + d_{11} + d_{12} + d_{13}, \text{ say.}$$

First, due to (S.3.21), (S.3.22), and Lemma S.13,

$$\sup_{z \in \mathbb{C}^+} |\text{etr} A_1(x) - \text{etr} A(x)| = \sup_{z \in \mathbb{C}^+} \left| \frac{1}{n} z_1^T A_1^2(z) z_1 \beta_1(z) \right| = O(1).$$
Together with (S.3.23), we get
\[ \sup_{x \in \mathbb{C}^+} \left( d_{10} - n^{1/2} a^T b \Theta_n(z, \gamma) - \frac{1}{\Theta_n(z, \gamma)} \right) \to 0. \]

Next we want to show \( \sup_{x \in \mathbb{C}^+} \left( |d_{11}| + |d_{13}| \right) \to 0. \)

\[ d_{11} = n^{1/2} a^T b(\beta^R(x))^2 \mathbb{E} \beta_1(z) \left\{ \frac{1}{n} z_j^T A_1(z) z_1 - \frac{1}{n} \text{tr} A_1(z) \right\}^2, \]
\[ |d_{13}| \leq n^{1/2} \|u_a\|_{\max} \sum_{j=1}^N |\beta^R(x)|^2 \mathbb{E} |z_j^T A_j(z) Z_j u_b \beta_j(z)| \times \left| \left\{ \frac{1}{n} z_j^T A_1(z) z_1 - \frac{1}{n} \text{tr} A_1(z) \right\}^2. \]

Due to (S.3.21), (S.3.23), (S.3.28), and Lemma S.13,
\[ \sup_{x \in \mathbb{C}^+} \mathbb{E} \left\{ \frac{1}{n} z_j^T A_1(z) z_1 - \frac{1}{n} \text{tr} A_1(z) \right\}^4 = o(n^{-1}), \]
\[ \sup_{x \in \mathbb{C}^+} \mathbb{E} |z_j^T A_j(z) Z_j u_b \beta_j(z)|^2 = O(1). \]
(S.3.29)

Thus,
\[ \sup_{x \in \mathbb{C}^+} \left( |d_{11}| + |d_{13}| \right) \to 0. \]

We next want to show \( \sup_{x \in \mathbb{C}^+} |d_{12}| \to 0. \)

By Lemma S.20,
\[ n^{1/2} \mathbb{E} z_j^T A_j(z) Z_j u_b \left\{ \frac{1}{n} z_j^T A_j(z) z_1 - \frac{1}{n} \text{tr} A_j(z) \right\} \]
\[ = n^{-1/2} \mathbb{E} z_j^3 \mathbb{E} \sum_{i=1}^p c_i^T A_j(z) c_i c_i^T A_j(z) Z_j u_b. \]

We first show
\[ \sup_{x \in \mathbb{C}^+} \sup_{1 \leq i, m \leq p} \mathbb{E} \left| c_i^T A_j(z) e_m - \mathbb{E} c_i^T A_j(z) e_m \right|^2 = O(n^{-1}), \]
(S.3.30)
\[ \sup_{x \in \mathbb{C}^+} \sup_{1 \leq i \leq m, 1 \leq j \leq N} \mathbb{E} \left| c_i^T A_j(z) Z_j u_b - \mathbb{E} c_i^T A_j(z) Z_j u_b \right|^2 = O(n^{-1}). \]
(S.3.31)

For (S.3.30),
\[ \mathbb{E} \left| c_i^T A_j(z) e_m - \mathbb{E} c_i^T A_j(z) e_m \right|^2 \]
The proof of (S.3.31) is complete.

By Lemma S.14, there exists a $K$,

$$
\mathbb{E}|z^T_\ell A_{j\ell}(z)e_m e^T_i A_{j\ell}(z)z_{\ell j\ell}(z)|^2 \leq K \mathbb{E}\|A_{j\ell}(z)\|^4.
$$

Recall the definition of $G^c$ in Section S.3.2. Using (S.3.21) and (S.3.24), on $G^c$,

$$
\mathbb{E}|z^T_\ell A_{j\ell}(z)e_m e^T_i A_{j\ell}(z)z_{\ell j\ell}(z)|^2 \leq \mathbb{E}\|A_{j\ell}(z)\|^4 = O(1). \quad \text{(S.3.32)}
$$

On $G$,

$$
\mathbb{E}|z^T_\ell A_{j\ell}(z)e_m e^T_i A_{j\ell}(z)z_{\ell j\ell}(z)|^2 \leq \mathbb{E}\|z_{\ell j\ell}(z)\|^2 \mathbb{E}\|A_{j\ell}(z)\|^2 1(G) \to 0. \quad \text{(S.3.33)}
$$

It completes the proof of (S.3.30).

To show (S.3.31),

$$
\mathbb{E}|e^T_i A_j(z)Z_j u_b - e^T_i A_j(z)Z_j u_b|^2 \\
= \mathbb{E}\left| \sum_{\ell \neq j} (E_{\ell} - E_{\ell-1}) [e^T_i A_j(z)Z_j u_b - e^T_j A_{j\ell}(z)Z_{j\ell} u_b] \right|^2 \\
\leq K \sum_{\ell \neq j} \mathbb{E}\left| u_{b\ell} e^T_i A_j(z)z_{\ell j\ell}(z) - \frac{1}{n} e^T_i A_{j\ell}(z)z_{\ell j\ell}(z)z^T_i A_{j\ell}(z)z_{\ell j\ell}(z) u_b \beta_{\ell j}(z) \right|^2 \\
+ u_{b\ell}(\beta_{\ell j}(z) - 1)e^T_i A_j(z)z_{\ell j\ell}(z).
$$

By Lemma S.14, (S.3.21),

$$
\sup_{z \in C^+} \mathbb{E}|e^T_i A_{j\ell}(z)z_{\ell j\ell}(z)|^4 \leq K \sup_{z \in C^+} \mathbb{E}\|A_{j\ell}(z)\|^4 = O(1).
$$

Together with Lemma S.14 and (S.3.22),

$$
\mathbb{E}|e^T_i A_{j\ell}(z)z_{\ell j\ell}(z)Z_{j\ell} u_b \beta_{\ell j}(z)|^2 = O(1).
$$

The proof of (S.3.31) is complete.
The residual term $o(1)$ is uniform on $C^+$. Note
\[
E_i^T A_1(z) z_2
= E_i^T A_{12}(z) z_2 \beta_{21}(z)
= -E_i^T A_{12}(z) z_2 \beta_{21}(z) \beta_2(z) \left( \frac{1}{n} z_2^T A_{12}(z) z_2 - \frac{1}{n} \text{Etr} A_1(z) \right)
= -E_i^T A_{12}(z) z_2 \beta_2(z) \beta_{21}(z) \left( \frac{1}{n} z_2^T A_{12}(z) z_2 - \frac{1}{n} \text{Etr} A_1(z) \right)
+ E_i^T A_{12}(z) z_2 \beta_2(z) \beta_{21}(z) \left( \frac{1}{n} z_2^T A_{12}(z) z_2 - \frac{1}{n} \text{Etr} A_1(z) \right)^2
= -\frac{1}{n} (\beta_2(z))^2 \beta_{21}(z) \left( \frac{1}{n} z_2^T A_{12}(z) z_2 - \frac{1}{n} \text{Etr} A_1(z) \right)^2
+ E_i^T A_{12}(z) z_2 \beta_2(z) \beta_{21}(z) \left( \frac{1}{n} z_2^T A_{12}(z) z_2 - \frac{1}{n} \text{Etr} A_1(z) \right)^2.
\]
By Lemma S.13, Lemma S.14, (S.3.21), and (S.10),
\[
\sup_{z \in C^+:i} |E_i^T A_{12}(z) z_2 \beta_2(z) \beta_{21}(z) \left( \frac{1}{n} z_2^T A_{12}(z) z_2 - \frac{1}{n} \text{Etr} A_1(z) \right)^2| = o(n^{-1/2}).
\]
Therefore, again using (S.3.30),
\[
d_{12} = n^{-3/2}(\beta_2(z))^4 (E \geq 1_1)^2 \sum_{j=1}^N u_{aj} \sum_{\ell \neq j} u_{bl} \sum_{i=1}^p E_i^T A_j(z) e_i \times
\sum_{\ell=1}^p E_{i\ell}^T A_{12}(z) e_{i\ell}^T A_{12}(z) e_{i\ell} + o(1)
= n^{-3/2}(\beta_2(z))^4 (E \geq 1_1)^2 \sum_{j=1}^N u_{aj} \sum_{\ell \neq j} u_{bl} \sum_{i=1}^p E_i^T A_j(z) e_i E_i^T A_{12}(z) e_{i \ell} \times
\]
We next show S.3.4. Convergence of $G^{(3)}$.

It follows that $\sup_{z \in \mathbb{C}^+} |d_{12}| = O(n^{-1/2}).$

\[ \text{Note } a^T U_n^T 1_n = O(n^{1/2}) \text{ because } \|U\|_{\text{max}} = O(n^{-1/2}). \text{ Meanwhile,} \]

\[ \sum_{i=1}^{P} \sum_{l=1}^{P} E e_i^T A_1(z) e_i E e_l^T A_12(z) e_l = \alpha_1^T E A_1 \alpha_2, \]

where $\alpha_1$ is the diagonal of $E A_1(z)$ and $\alpha_2$ is the diagonal of $E A_{12}(z)$.

\[ \sup_{z \in \mathbb{C}^+} |\alpha_1^T E A_{12} \alpha_2| \leq \sup_{z \in \mathbb{C}^+} \|E A_{12}(z)\|_2 \|\alpha_1\|_2 \|\alpha_2\|_2 \]

\[ \leq \sup_{z \in \mathbb{C}^+} n \|E A_{12}(z)\|_2^2 \|E A_1(z)\|_2 \|E A_{12}(z)\|_2 = O(n). \]

It follows that $\sup_{z \in \mathbb{C}^+} |d_{12}| = O(n^{-1/2}).$

\section*{S.3.4. Convergence of $G^{(3)}$}

We next show

\[ \sup_{z \in \mathbb{C}^+} G^{(3)}_n(z, a, b) \to 0. \]

Using the idea of proof of Lemma 2 of Chen et al. (2011), post-multiplying both sides of the identity $(\tilde{\Sigma}_p - z I_p) + z I_p = \tilde{\Sigma}_p$ by $(\tilde{\Sigma}_p - z I_p)^{-1},$

\[ I_p + z(\tilde{\Sigma}_p - z I_p)^{-1} = \tilde{\Sigma}_p(\tilde{\Sigma}_p - z I_p)^{-1} = \frac{1}{n} \sum_{i=1}^{N} \Sigma_p^{1/2} z_i z_i^T \Sigma_p^{1/2} (\tilde{\Sigma}_p - z I_p)^{-1}. \]

Taking trace and expectation on both sides,

\[ \frac{n}{N} \left[ 1 + z \frac{1}{p} \text{etr}(\tilde{\Sigma}_p - z I_p)^{-1} \right] = \frac{1}{p} E z_1^T A_1(z) z_1 \]

\[ = \frac{\gamma_{n-1}(z) - 1}{\Theta_{n-1}(z)} - \frac{1}{p} E z_1^T A_1(z) z_1 \beta(z) \beta(z) \left\{ \frac{1}{n} z_1^T A_1(z) z_1 - \frac{1}{n} \text{etr} A_1(z) \right\}. \]

Therefore,

\[ \frac{(\Theta_{n-1}(z) - 1)}{\Theta_{n-1}(z)} = \frac{n}{N} \gamma_n + \frac{n}{N} z \frac{1}{p} \text{etr}(\tilde{\Sigma}_p - z I_p)^{-1} \]

\[ + \frac{1}{n} E z_1^T A_1(z) z_1 \beta(z) \beta(z) \left\{ \frac{1}{n} z_1^T A_1(z) z_1 - \frac{1}{n} \text{etr} A_1(z) \right\}. \]
Note \((\Theta(z, \gamma) - 1)\Theta^{-1}(z, \gamma) = \gamma + \gamma zm(z)\). We only need to show

\[
\sup_{x \in \mathbb{C}^+} n^{1/2} \left| \gamma_n + \gamma_n x^{-1} \operatorname{Etr}(\tilde{\Sigma}_p - zI_p)^{-1} \right| - (\gamma + \gamma zm(z)) \to 0, \quad (S.3.34)
\]

\[
\sup_{x \in \mathbb{C}^+} n^{-1/2} \left| \mathbb{E} z_1^T A_1(z) x_1 \beta^2(z) \beta_1(z) \left\{ \frac{1}{n} z_1^T A_1(z) x_1 - \frac{1}{n} \operatorname{Etr} A_1(z) \right\} \right| \to 0. \quad (S.3.35)
\]

\((S.3.34)\) 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 \((S.3.35)\),

\[
\sup_{x \in \mathbb{C}^+} n^{-1/2} \left| \mathbb{E} z_1^T A_1(z) x_1 \beta^2(z) \beta_1(z) \left\{ \frac{1}{n} z_1^T A_1(z) x_1 - \frac{1}{n} \operatorname{Etr} A_1(z) \right\} \right|
\]

\[
= \sup_{x \in \mathbb{C}^+} n^{-1/2} \left| \mathbb{E} \beta_1(z) \beta^2(z) \left\{ \frac{1}{n} z_1^T A_1(z) x_1 - \frac{1}{n} \operatorname{Etr} A_1(z) \right\} \right|
\]

\[
= \sup_{x \in \mathbb{C}^+} n^{-1/2} \left| \mathbb{E} \beta_1(z) \beta^2(z) \left\{ \frac{1}{n} z_1^T A_1(z) x_1 - \frac{1}{n} \operatorname{Etr} A_1(z) \right\} \right| = o(1).
\]

Notice that \(a^T b(\Theta(z, \gamma) - 1)/\Theta(z, \gamma)\) is the pointwise asymptotic mean of \(G_n(z, a, b)\). This expression suggests that in order to establish Theorem A.1 we might need to smooth \(\Theta(z, \gamma)\) when the imaginary part of \(z\) in absolute value is smaller than \(\rho_n\), in the same way as in the process smoothing strategy. Similar considerations apply to the treatment of the pointwise asymptotic covariance \(\delta(z_1, z_2, \gamma)\) of \(G_n(z, a, b)\). However, notice that \(\Theta(z, \gamma)\) and \(\delta(z_1, z_2, \gamma)\) are smooth functions of \(z, z_1\) and \(z_2\) with bounded derivatives on \(\mathcal{C}\) (see Section S.2). Therefore, when \(z \in \mathcal{C} \setminus \mathbb{C}^+, \Theta(z, \gamma) = \Theta(u \pm i \rho_n, \gamma) + O(\rho_n)\) and \(\sqrt{n} \rho_n \to 0\). Similar results hold for \(\delta(z_1, z_2, \gamma)\) when \(z_1\) and/or \(z_2\) are close to the real axis.

### S.4. Proof of Theorem 2.3

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

\[
\sqrt{n} M(f) = \frac{1}{\sqrt{n}} Q_n^T Z^T \Sigma_p^{T/2} f(\tilde{\Sigma}_p) \Sigma_p^{1/2} Z Q_n
\]

\[
+ Q_n^T Z^T \Sigma_p^{T/2} f(\tilde{\Sigma}_p) B C T_n^{-1/2} + \left( Q_n^T Z^T \Sigma_p^{T/2} f(\tilde{\Sigma}_p) B C T_n^{-1/2} \right)^T
\]

\[
+ \sqrt{n} T_n^{-1/2} C^T B^T f(\tilde{\Sigma}_p) B C T_n^{-1/2}.
\]

In view of Theorem 2.1 and Lemma 2.2 (shown in Section S.8), we only need to show that under \(\textbf{C1}\),

\[
\left\| Q_n^T Z^T \Sigma_p^{T/2} f(\tilde{\Sigma}_p) B C T_n^{-1/2} \right\|_2 \overset{P}{\to} 0, \quad (S.4.1)
\]
\[ \sqrt{n}T_n^{-1/2}CT^T f(\hat{\Sigma}_p)BCT_n^{-1/2} - \mathcal{H}(D,f) \xrightarrow{P} 0. \quad (S.4.2) \]

For future references, we also show that the convergence is uniform over the class \[ \{B \in \mathbb{R}^{p \times k}, C \in \mathbb{R}^{k \times q} : \sqrt{n}\|BC\|_2^2 \leq K\}, \quad \text{for arbitrary } K > 0. \]

Similar to the strategy in the proof of Theorem 2.1, we first consider \( z_{ij} \)'s satisfying (A.4). Observe when \( \lambda_{\max}(\hat{\Sigma}_p) < D \) for any \( D \in (\limsup_p \lambda_{\max}(\Sigma_p)(1 + \sqrt{\gamma})^2, \pi) \),

\[
\|Q^T_nZ^T\Sigma_p^{1/2}f(\hat{\Sigma}_p)BCT_n^{-1/2}\|_2 \\
= \left\| \int_{c} \frac{-1}{2\pi i} f(z)Q^T_nZ^T\Sigma_p^{1/2}(\hat{\Sigma}_p - zI)^{-1}BCT_n^{-1/2}dz \right\|_2 \\
\leq \left\| \int_{c} \frac{-1}{2\pi i} f(z)Q^T_nZ^T\Sigma_p^{1/2}(\hat{\Sigma}_p - zI)^{-1}BCT_n^{-1/2}dz \right\|_2 + w_{1,n}\|BC\|_2,
\]

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

\[
\left\| \int_{c} \frac{-1}{2\pi i} f(z)Q^T_nZ^T\Sigma_p^{1/2}(\hat{\Sigma}_p - zI)^{-1}BCT_n^{-1/2}dz \right\|_2 \\
\leq K\rho_n |T_n^{-1/2}|_2 \|Z^T\Sigma_p^{1/2}\|_2 \left[(\pi - D)^{-1} + |u|^{-1}\right]\|BC\|_2 \\
\leq K\rho_n |T_n^{-1/2}|_2 \|D^{1/2}\left[(\pi - D)^{-1} + |u|^{-1}\right]\|BC\|_2,
\]

where \( K \) is a universal constant.

Next, using Lemma S.6,

\[ Q^T_nZ^T\Sigma_p^{1/2}(\hat{\Sigma}_p - zI)^{-1}BC = V_n^T(I - \hat{\Theta}_n(z))^{-1}U_n^TZ^T\Sigma_p^{1/2}(\hat{\Sigma}_p - zI)^{-1}BC. \]

In the following, \( |dz| \) refers to the differential form of the integral with respect to the length of a contour.

For any \( \varepsilon > 0 \),

\[
P\left(\|Q^T_nZ^T\Sigma_p^{1/2}f(\hat{\Sigma}_p)BCT_n^{-1/2}\|_2 > \varepsilon\right) \\
\leq P\left(\left\{\|Q^T_nZ^T\Sigma_p^{1/2}f(\hat{\Sigma}_p)BCT_n^{-1/2}\|_2 > \varepsilon\right\} \cap \{\lambda_{\max}(\hat{\Sigma}_p) < D\}\right) + P\left(\lambda_{\max}(\hat{\Sigma}_p) \geq D\right) \\
\leq P\left(\left\| \int_{c} \frac{-1}{2\pi i} f(z)V_n^T(I - \hat{\Theta}_n(z))^{-1}U_n^TZ^T\Sigma_p^{1/2}(\hat{\Sigma}_p - zI)^{-1}BCT_n^{-1/2}dz \right\|_2 \\
> \varepsilon - w_{1,n}\|BC\|_2\right) + w_{2,n} \\
\leq P\left(\int_{c} \|(I - \hat{\Theta}(z))^{-1}\|_2^p|dz| \int_{c} \|U_n^TZ^T\Sigma_p^{1/2}(\hat{\Sigma}_p - zI)^{-1}BCT_n^{-1/2}\|_2^2|dz| \\
> K_1(\varepsilon - w_{1,n}\|BC\|_2^2)\right) + w_{2,n}
\]
\[ p \text{High-dimensional general linear hypothesis via spectral shrinkage} 49 \]

Next, we prove

\[ \text{for any vector (} \text{and that } \Theta(\mathbf{z}, \gamma) \text{ is bounded on } C \text{(shown in Bai and Silverstein (2004)).} \]

Note that (S.4.3) follows because (I - \tilde{Q}(\mathbf{z}))^{-1} P \Theta(\mathbf{z}, \gamma) I_p (deduced from Lemma A.1) and that \Theta(\mathbf{z}, \gamma) is bounded on C (shown in Bai and Silverstein (2004)).

Here, \( K_1, K_2, K_3, K_4 \) are appropriately large constants independent of BC, and

\[ w_{2,n} = \mathbb{P} \left( \lambda_{\max}(\widetilde{\Sigma}_p) \geq \Theta \right) \to 0, \quad \text{due to Lemma A.1}, \]

\[ w_{3,n} = \mathbb{P} \left( \mathcal{F}_{\mathcal{C}}^+ \right) \to 0. \quad (S.4.3) \]

Note that (S.4.3) follows because (I - \tilde{Q}(\mathbf{z}))^{-1} P \Theta(\mathbf{z}, \gamma) I_p (deduced from Lemma A.1)

Next, we prove

\[ \sup_{z \in \mathcal{C}} \mathbb{E} \left[ \left\| U_n^T Z^T \Sigma_{p}^{1/2} (\widetilde{\Sigma}_p - \mathbf{z} I) - BCB^{-1/2} \mathbf{z} \right\|_F^2 \right] \leq K \| BC \|_2^2, \quad (S.4.4) \]

for a sufficiently large \( K \).

It is sufficient to show

\[ \sup_{z \in \mathcal{C}} \mathbb{E} \left[ \left\| a^T U_n^T Z^T (\widetilde{\Sigma}_p - \mathbf{z} I) - BCB^{-1/2} \mathbf{z} \right\|_F^2 \right] \leq K \| BC \|_2^2, \]

for any vector (\( k \)-variate) \( a \) and (\( q \)-variate) \( b \), such that \( \| a \|_2 \leq 1 \) and \( \| b \|_2 \leq 1 \).

Recall notation defined in Section S.3. Further, write \( \frac{1}{n} \Sigma_{p}^{1/2} Z_j Z_j^T \Sigma_{p}^{1/2} \) as \( \Sigma_p(j) \) and \( \frac{1}{n} \Sigma_{p}^{1/2} z_{jj'} Z_j^T z_{jj'} \Sigma_{p}^{1/2} \) as \( \Sigma_p(jj') \).

\[ a^T U_n^T Z^T (\widetilde{\Sigma}_p - \mathbf{z} I)^{-1} B B \]

\[ = \sum_{j=1}^{N} \sqrt{n} u_{aj} z_j^T \Sigma_{p}^{1/2} (\widetilde{\Sigma}_p(j) - \mathbf{z} I)^{-1} B B \]

\[ = \sum_{j=1}^{N} \sqrt{n} u_{aj} z_j^T \Sigma_{p}^{1/2} \left( \widetilde{\Sigma}_p(j) - \mathbf{z} I \right)^{-1} B B \]

\[ \times \left( \sum_{j=1}^{N} \sqrt{n} u_{aj} z_j^T \Sigma_{p}^{1/2} (\widetilde{\Sigma}_p(j) - \mathbf{z} I)^{-1} B B \right) \]

\[ = d_{13} + d_{14}, \text{ say.} \]

\[ \mathbb{E} [d_{13}]^2 = n \sum_{j=1}^{N} u_{aj}^2 |\beta_{j}^q(\mathbf{z})|^2 \mathbb{E} [b_i^T B^T (\widetilde{\Sigma}_p(j) - \mathbf{z} I)^{-1} \Sigma_p(jj') - \mathbf{z} I)^{-1} B B + \]
Using analogous arguments, for some deterministic sequence \(w\) uniform on \([0,1]\), without variable truncation converges to 0 in probability. The convergence is also shown in Section S.9, we will show that the difference between the left-hand side of (S.4.1) with \(z\) in (S.3.21), (S.3.22), and (S.3.23),

\[
\sup_{\|z\|_2 \leq 1} E|d_{11}|^2 \leq K\|BC\|^2.
\]

Using Lemma S.13, Lemma S.14, and Lemma S.17,

\[
E|z_j^T A_{12}(z)z_j|^4 \leq Kn^2E\|A_{12}(z)\|^4,
\]

Then (S.4.1) follows when \(z_j\)’s satisfy (A.4), since (2.3) implies \(\|BC\|^2 \to 0\), as \(n \to \infty\). In Section S.9, we will show that the difference between the left-hand side of (S.4.1) with and without variable truncation converges to 0 in probability. The convergence is also uniform on \(BC \in \{|\sqrt{n}\|BC\|_2 \leq K\}\). It completes the proof of (S.4.1).

As for (S.4.2), again using Lemma S.6,

\[
C^T B^T(\hat{\Sigma}_p - zI)^{-1} BC = C^T B^T(\hat{\Sigma}_p - zI)^{-1} BC
\]

Using analogous arguments, for some deterministic sequence \(w_{4,n} \to 0\) and constants \(K_5, K_6,\)

\[
P\left(\sqrt{n}T_n^{-1/2}C^T B^T f(\hat{\Sigma}_p)BCT_n^{-1/2} - \mathcal{H}(D,f)\right) > \varepsilon\)
We only need to show
\[ \text{result is shown in El Karoui and Kösters (2011).} \]
Note for any vector \((q\text{-variate})\) \(a\) such that \(\|a\|_2 \leq 1\).
Moreover, note \(\text{we use Lemma S.8 to show the result. It is worth mentioning that when} z \in \mathbb{R}^n, \text{the}\)
result is shown in El Karoui and Kösters (2011). Note
\[
\begin{align*}
\sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa \\
\leq \sum_{j=1}^{N} E_j \left[ \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p^{(j)} - zI)^{-1} BCa \right] \\
- \sum_{j=1}^{N} E_j \left[ \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p^{(j)} - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa \right].
\end{align*}
\]
Moreover, note
\[
\sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p^{(j)} - zI)^{-1} BCa
\]
for any vector \((q\text{-variate})\) \(a\) such that \(\|a\|_2 \leq 1\).
We use Lemma S.8 to show the result. It is worth mentioning that when \(z \in \mathbb{R}^n, \text{the}\)
result is shown in El Karoui and Kösters (2011). Note
\[
\begin{align*}
\sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa \\
\leq \sum_{j=1}^{N} E_j \left[ \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p^{(j)} - zI)^{-1} BCa \right] \\
- \sum_{j=1}^{N} E_j \left[ \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p^{(j)} - zI)^{-1} BCa - \sqrt{n}a^T C^T B^T (\hat{\Sigma}_p - zI)^{-1} BCa \right].
\end{align*}
\]
Therefore, using Lemma S.8 (Burkholder’s inequality) and Lemma S.14,

\[ n\mathbb{E}|a^T C^T B^T (\hat{\Sigma}_p) - zI|^{-1}BCa (\hat{\Sigma}_p) - zI|^{-1}BCa|^2 \]

\[ \leq K \mathbb{E}|z_j^T \Sigma_p^{-1/2} (\hat{\Sigma}_p) - zI|^{-1}BCaa^T C^T B^T (\hat{\Sigma}_p) - zI|^{-1} \Sigma_p^{-1/2} z_j \beta_j(z)|^2. \]

Observe that \( \Sigma_p^{-1/2} (\hat{\Sigma}_p) - zI \) is of rank one. Due to very similar lines to (S.3.32) and (S.3.33), we can get

\[ \mathbb{E}|z_j^T \Sigma_p^{-1/2} (\hat{\Sigma}_p) - zI|^{-1}BCaa^T C^T B^T (\hat{\Sigma}_p) - zI|^{-1} \Sigma_p^{-1/2} z_j \beta_j(z)|^2 \leq K \|BC\|_2^2. \]

Next, we show

\[ \sup_{z \in C^*} \left| a^T C^T B^T \left[ \mathbb{E} (\hat{\Sigma}_p) - zI \right]^{-1} - \{ \Theta(z, \gamma) \Sigma_p - zI \}^{-1}BCa \right|^2 \leq K \|BC\|_2^4. \]

\[ \sqrt{n} a^T C^T B^T \left[ \mathbb{E} (\hat{\Sigma}_p) - zI \right]^{-1} - (\Theta(z, \gamma) \Sigma_p - zI)^{-1}BCa \]

\[ = \mathbb{E} \sqrt{n} a^T C^T B^T (\hat{\Sigma}_p) - zI)^{-1} \left( \Theta(z, \gamma) \Sigma_p - \hat{\Sigma}_p \right) \left( \Theta(z, \gamma) \Sigma_p - zI \right)^{-1}BCa \]

\[ = \mathbb{E} \sqrt{n} a^T C^T B^T (\hat{\Sigma}_p) - zI)^{-1} \left[ \Theta(z, \gamma) \Sigma_p - \frac{N}{n} \Sigma_p^{1/2} z_1 z_1^T \Sigma_p^{1/2} \right] \times \]

\[ \left( \Theta(z, \gamma) \Sigma_p - zI \right)^{-1}BCa \]

\[ = -\frac{N}{n} \sqrt{n} \mathbb{E} z_1^T \Sigma_p^{1/2} (\Theta(z, \gamma) \Sigma_p - zI)^{-1}BCaa^T C^T B^T (\hat{\Sigma}_p) - zI)^{-1} \Sigma_p^{1/2} z_1 \beta_1(z) \]

\[ + \sqrt{n} \mathbb{E} (\Theta(z, \gamma) \Sigma_p - zI)^{-1}BCa \]

\[ = \sqrt{n} |\Theta(z, \gamma) - \frac{N}{n} \beta^2(z)| \mathbb{E} a^T C^T B^T (\hat{\Sigma}_p) - zI)^{-1} \Sigma_p^{1/2} z_1 \beta_1(z) \]

\[ - \frac{1}{\sqrt{n}} \Theta(z, \gamma)(z) \mathbb{E} a^T C^T B^T (\hat{\Sigma}_p) - zI)^{-1} \Sigma_p^{1/2} z_1 \times \]

\[ z_1^T \Sigma_p^{1/2} (\hat{\Sigma}_p) - zI)^{-1} \Sigma_p (\Theta(z, \gamma) \Sigma_p - zI)^{-1}BCa \beta_1(z) \]

\[ + \frac{N}{n} \sqrt{n} \mathbb{E} z_1^T \Sigma_p^{1/2} (\Theta(z, \gamma) \Sigma_p - zI)^{-1}BCaa^T C^T B^T (\hat{\Sigma}_p) - zI)^{-1} \times \]

\[ \Sigma_p^{1/2} z_1 \beta_1(z)^2 \]

\[ \left( \frac{1}{n} z_1^T A_1(z) z_1 - \mathbb{E} \frac{1}{n} \text{tr} A_1(z) \right). \]

Due to (S.3.19) and (S.3.21), we can get

\[ \sup_{z \in C^*} \sqrt{n} |\Theta(z, \gamma) - \frac{N}{n} \beta^2(z)| \to 0, \]
Theorem 2.4

Let \( \mu \) denote the prior probability measure of \( \Phi \), \( \mathbb{P}_{BC} \) the probability measure of the observations conditional on \( BC \), and \( \mathbb{P}_Z \) the probability measure of \( Z \).

Under the probabilistic local alternative model (2.6),

\[
\sqrt{n} |BC|_2^2 = O_p(1). \tag{S.5.1}
\]

Using Lemma S.13 and Lemma S.17, we can prove

\[
\sqrt{n} C^T B^T [\Theta^{-1}(z, \gamma) \Sigma_p - zI]^{-1} BC - D_{\Phi*}(z, \gamma) \xrightarrow{\mathbb{P}_*} 0, \quad \text{pointwise for } z \in C, \tag{S.5.2}
\]

where \( D_{\Phi*}(z, \gamma) = h(z, \gamma) SS^T \).

In what follows, we only consider the (regularized version of) the LR criterion and show the convergence of \( \Upsilon^{LH} \). The convergence of \( \Upsilon^{LH} \) and \( \Upsilon^{BGP} \) 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 \( n > N_0 \),

\[
\mathbb{P}_n \left( \left| \Upsilon_{LR}^{(BC, f)} - \Phi \left( -\xi_0 + \frac{\text{tr}(\mathcal{H}(D_{\mathcal{X}^*}, f))}{q^{1/2}\Delta^{1/2}(f, \gamma)} \right) \right| > \epsilon \right) < \zeta.
\]

Define \( g = \{1 + \Omega(f, \gamma)\} \Delta^{-1/2}(f, \gamma) \) and its empirical counterpart \( g_n = \{1 + \hat{\Omega}(f, \gamma_n)\} \hat{\Delta}^{-1/2}(f, \gamma_n) \). Lemma 2.2 (proved in Section S.8) implies that \( g_n \to g \) in probability. Recall

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

Write

\[
M(f) = \frac{1}{n} Q_n^T Z_n^T \Sigma_p^{1/2} f(\hat{\Sigma}_p) \Sigma_p^{1/2} Z_n + \frac{g}{g_n \sqrt{n}} \mathcal{H}(D_{\mathcal{X}^*}, f) + \frac{1}{g_n \sqrt{n}} \sigma_n(BC)
\]

\[+ \frac{1}{g_n \sqrt{n}} \sum_{i=1}^{q} \eta^{(i)}(BC, Z),\]

where with notation \( W(f, \Sigma_p, \gamma) = (-2\pi i)^{-1} \int_{\mathbb{C}} f(z)(\Theta(z, \gamma)\Sigma_p - zI)^{-1}dz, \)

\[
\sigma_n(BC) = \left[ \frac{\sqrt{n}}{T_n^{-1/2} C^T B W(f, \Sigma_p, \gamma) B C T_n^{-1/2}} - \mathcal{H}(D_{\mathcal{X}^*}, f) \right],
\]

\[
\eta^{(1)}(BC, Z) = [g_n - g] \frac{\sqrt{n}}{T_n^{-1/2} C^T B W(f, \Sigma_p, \gamma) B C T_n^{-1/2}},
\]

\[
\eta^{(2)}(BC, Z) = g_n \frac{\sqrt{n}}{T_n^{-1/2} C^T B W(f, \Sigma_p, \gamma) B C T_n^{-1/2}} \left( f(\hat{\Sigma}_p) - W(f, \Sigma_p, \gamma) \right) B C T_n^{-1/2},
\]

\[
\eta^{(3)}(BC, Z) = g_n Q_n^T Z_n^T \Sigma_p^{1/2} f(\hat{\Sigma}_p) B C T_n^{-1/2},
\]

\[
\eta^{(4)}(BC, Z) = g_n Q_n^T Z_n^T \Sigma_p^{1/2} Z_n.
\]

Therefore, by Lemma S.7, for \( i = 1, 2, \ldots, q, \)

\[
\left| \lambda_i(M(f)) - \lambda_i \left( \frac{1}{n} Q_n^T Z_n^T \Sigma_p^{1/2} f(\hat{\Sigma}_p) \Sigma_p^{1/2} Z_n + \frac{g}{g_n \sqrt{n}} \mathcal{H}(D_{\mathcal{X}^*}, f) \right) \right| \leq \frac{1}{g_n \sqrt{n}} \| \sigma_n(BC) \|_2 + \frac{1}{g_n \sqrt{n}} \sum_{i=1}^{q} \| \eta^{(i)}(BC, Z) \|_2.
\]

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

\[
\sqrt{n} \frac{g_n}{q^{1/2}} |T_{LR}(f) - \sum_{i=1}^{q} \log \left[ 1 + \lambda_i \left( \frac{1}{n} Q_n^T Z_n^T \Sigma_p^{1/2} f(\hat{\Sigma}_p) \Sigma_p^{1/2} Z_n + \frac{g}{g_n \sqrt{n}} \mathcal{H}(D_{\mathcal{X}^*}, f) \right) \right]| \leq q^{1/2} \| \sigma_n(BC) \|_2 + q^{1/2} \sum_{i=1}^{q} \| \eta^{(i)}(BC, Z) \|_2.
\]

(S.5.3)
Define
\[
\tilde{\Upsilon}(\xi) = \mathbb{P}_Z \left( \sqrt{n} \frac{g_n}{q^{1/2}} \sum_{i=1}^{q} \log \left[ 1 + \lambda_i \left( \frac{1}{n} Q_n^T \Sigma_p^{1/2} f(\hat{\Sigma}_p) \Sigma_p^{1/2} Z Q_n + \frac{g_n}{g_n^2} \mathcal{H}(D_{\hat{p}_s}, f) \right) \right] - \sqrt{n} g_n q^{1/2} \log(1 + \hat{\Omega}(f, \gamma)) > \xi \right)
\]
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 \( \xi \),
\[
\tilde{\Upsilon}(\xi) \to \Phi \left( -\xi + \frac{\text{tr}(\mathcal{H}(D_{\hat{p}_s}, f))}{q^{1/2} \Delta^{1/2}(f, \gamma)} \right).
\]
Hence, for any \( \epsilon > 0 \), we can find a sufficiently large \( N_1 \) such that when \( n > N_1 \),
\[
\tilde{\Upsilon}(\xi_{\alpha} - 5q^{1/2} \epsilon) < \Phi \left( -\xi_{\alpha} + \frac{\text{tr}(\mathcal{H}(D_{\hat{p}_s}, f))}{q^{1/2} \Delta^{1/2}(f, \gamma)} + 5q^{1/2} \epsilon \right) + \epsilon,
\]
\[
\tilde{\Upsilon}(\xi_{\alpha} + 5q^{1/2} \epsilon) > \Phi \left( -\xi_{\alpha} + \frac{\text{tr}(\mathcal{H}(D_{\hat{p}_s}, f))}{q^{1/2} \Delta^{1/2}(f, \gamma)} - 5q^{1/2} \epsilon \right) - \epsilon.
\]

Due to (S.5.1) and (S.5.2), there exists a constant \( K_\xi \) and a sufficiently large \( N_2 \) such that when \( n > N_2 \), \( \mathbb{P}_X(K^{(1)}) > 1 - \zeta \), where
\[
K^{(1)} = \{ BC : \sqrt{n} \| BC \|_2 \leq K_\xi \} \cap \{ BC : |\sigma_n(BC)| \leq \epsilon \}.
\]
Using the arguments in Section S.4, we have, for \( i = 2, 3, 4 \), uniformly on \( K^{(1)} \),
\[
\eta^{(i)}(BC, Z) \overset{p_{BC}}{\to} 0.
\]
This convergence to zero is also valid for \( \eta^{(i)}(BC, Z) \), since \( g_n \overset{P}{\rightarrow} g \) 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
\[
K^{(2)}(BC) = \{ Z : \max_{i=1,\ldots,4} \| \eta^{(i)}(BC, Z) \|_2 \leq \epsilon \}
\]
has measure at least \( 1 - \epsilon \), that is, \( \mathbb{P}_{BC}(K^{(2)}(BC)) > 1 - \epsilon \). Therefore, when \( BC \in K^{(1)} \), and \( n > \max(N_1, N_2, N_3) \), (S.5.4) implies
\[
\Upsilon^{LR}(BC, f) \leq \mathbb{P}_{BC} \left( \left\{ \tilde{T}^{LR}(f) > \xi_{\alpha} \right\} \cap K^{(2)}(BC) \right) + \mathbb{P}_{BC}(K^{(2)}(BC)^c)
\]
\[
\leq \tilde{T}(\xi_{\alpha} - 5q^{1/2} \epsilon) + \epsilon
\]
\[
\leq \Phi \left( -\xi_{\alpha} + \frac{\text{tr}(\mathcal{H}(D_{\hat{p}_s}, f))}{q^{1/2} \Delta^{1/2}(f, \gamma)} + 5q^{1/2} \epsilon \right) + 2 \epsilon.
\]
Conversely,
\[
\Upsilon^{LR}(BC, f) \geq \mathbb{P}_{BC} \left( \left\{ \tilde{T}^{LR}(f) > \xi_{\alpha} \right\} \cap K^{(2)}(BC) \right)
\]
\[ \geq \bar{\Phi}(\xi + 5q^{1/2} \epsilon) \]
\[ \geq \Phi(-\xi + \frac{\text{tr}(H(Dq, f))}{q^{1/2} \Delta^{1/2}(f, \gamma)} - 5q^{1/2} \epsilon) - \epsilon. \]

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

**S.6. Proof of Theorem 3.1**

To show Result (1) of Theorem 3.1, first note that \( \Delta_{\ell, \gamma} \) is a continuous function of \( \ell \) under Condition (ii) of \( \mathcal{F} \). It is also assumed that \( \inf_{\ell \in \mathcal{L}} \Delta(f, \gamma) > 0 \) in Condition (iv). Hence, \( \Delta(f, \gamma) \) is bounded away from infinity and 0 on \( \mathcal{L} \). Moreover, Conditions (i)–(iii) of \( \mathcal{F} \) imply

\[ \sup_{z \in \mathcal{C}} \| f(z) \| < \infty. \]

In the proof of Lemma 2.2, presented in Section S.8, it is shown that

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

It follows that

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

Hence,

\[ \sup_{\ell \in \mathcal{L}} \sqrt{n} \left| \int_{\mathcal{C}^+} f(\zeta) [\hat{h}(\zeta, \gamma_n) - h(\zeta, \gamma)] d\zeta \right| \xrightarrow{P} 0. \]

The boundedness of \( \Theta(z, \gamma) \) is deduced in Section S.2. It can be checked that \( m_{n, p}(z) \) and \( \hat{\Theta}(z, \gamma_n) \) are also bounded on \( \mathcal{C} \) when \( \lambda_{\max}(\hat{\Sigma}_{np}) < \mathcal{D} < \bar{n} \). Since \( \sqrt{n} \rho \rightarrow 0 \), it follows that

\[ \sup_{\ell \in \mathcal{L}} \sqrt{n} \left| \int_{\mathcal{C}^+, \mathcal{C}^+} f_{\ell}(\zeta) \hat{h}(\zeta, \gamma_n) d\zeta \right| \xrightarrow{P} 0, \]
\[ \sup_{\ell \in \mathcal{L}} \sqrt{n} \left| \int_{\mathcal{C}^+, \mathcal{C}^+} f_{\ell}(\zeta) h(\zeta, \gamma_n) d\zeta \right| \xrightarrow{P} 0. \]

The convergence of \( \hat{\Delta}(f, \gamma_n) \) is stated in Lemma 2.2. The proof of Lemma 2.2 reveals that the convergence of \( \hat{\Delta}(f, \gamma_n) \) follows from the uniform convergence of \( \hat{\delta}(z_1, z_2, \gamma_n) \) on \( (\mathcal{C}^+)^2 \). It implies that the convergence of \( \hat{\Delta}(f, \gamma_n) \) is uniform on \( f_\ell \in \mathcal{F} \), because

\[ \sqrt{n} |\hat{\Delta}(f, \gamma_n) - \Delta(f, \gamma)| \leq K \sup_{z \in \mathcal{L}} \sup_{\ell \in \mathcal{L}} |f(\zeta)| \int_{\mathcal{C}^2} \sqrt{n} |\delta(z_1, z_2, \gamma_n) - \delta(z_1, z_2, \gamma)| dz_1 |dz_2|. \]
We therefore have
\[
\sup_{\ell \in \mathcal{L}} \sqrt{n} \big| \hat{\Delta}(f_\ell, \gamma_n) - \Delta(f_\ell, \gamma) \big| \xrightarrow{P} 0
\]
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 \( K_\varepsilon > 0 \) and an integer \( n_\varepsilon \), such that for \( t = K_\varepsilon n^{-1/4} \) and any \( n \geq n_\varepsilon \),
\[
P \left( \hat{\mathcal{E}}(\ell^*, \hat{h}, \gamma_n) - \hat{\mathcal{E}}(\ell^* + t\delta, \hat{h}, \gamma_n) \geq 0 \text{ for all } \delta \text{ s.t. } \|\delta\|_2 = 1 \text{ and } \ell^* + t\delta \in \mathcal{L} \right) \geq \varepsilon.
\]
Under condition (3.3),
\[
\hat{\mathcal{E}}(\ell^*, \hat{h}, \gamma_n) - \hat{\mathcal{E}}(\ell^* + t\delta, \hat{h}, \gamma_n) = \Xi(\ell^*, h, \gamma) - \Xi(\ell^* + t\delta, h, \gamma) + o_p(n^{-1/2})
\]
\[
\geq K\ell_2^2[\delta]_2^2 + o_p(n^{-1/2}).
\]
The remainder \( o_p(n^{-1/2}) \) is uniform for any \( t \) and \( \delta \). Therefore, if \( t = O(n^{-1/4}) \),
\[ \hat{\mathcal{E}}(\ell^*, \hat{h}, \gamma_n) - \hat{\mathcal{E}}(\ell^* + t\delta, \hat{h}, \gamma_n) \text{ is positive with high probability for any } \delta \text{ with } L_2\text{-norm } 1. \]

As for (3.5), write \( \ell = (l_1, \ldots, l_r)^T \) for any \( \ell \in \mathcal{L} \). Similarly denote the local maximizer of \( \hat{\mathcal{E}}(\ell, \hat{h}, \gamma_n) \) by \( \ell^*_n = (l^*_n, \ldots, l^*_n)^T \) and let \( \ell^* = (l^*_1, l^*_2, \ldots, l^*_r)^T \). Since the partial derivatives of \( f_{\ell^*} \) are analytic under Condition (iii) of \( \mathcal{F} \), due to Theorem 2.1 and (S.8.1),
\[
n^{1/4} \left[ M \left( \frac{\partial f_{\ell^*}}{\partial \ell_j^*_1} \right) - \hat{\Omega} \left( \frac{\partial f_{\ell^*}}{\partial \ell_j^*_1}, \gamma_n \right) I_q \right] = o_p(1), \quad j = 1, \ldots, r.
\]
\[
\left[ M \left( \frac{\partial^2 f_{\ell^*}}{\partial \ell_j^*_1 \partial \ell_{j'}^*_1} \right) - \hat{\Omega} \left( \frac{\partial^2 f_{\ell^*}}{\partial \ell_j^*_1 \partial \ell_{j'}^*_1}, \gamma_n \right) I_q \right] = o_p(1), \quad j, j' = 1, \ldots, r.
\]
Because the third-order derivatives are continuous functions on \( \mathcal{L} \otimes \mathcal{Z} \), we can find a constant \( K_\varphi \) such that
\[
\max_{1 \leq j, j', j'' \leq r} \sup_{\ell \in \mathcal{L}} \sup_{z \in \mathcal{Z}} \left| \frac{\partial^3 f_\ell(z)}{\partial \ell_j \partial \ell_{j'} \partial \ell_{j''}} \right| \leq K_\varphi.
\]
Since for the constant function \( f_0(x) = K_\varphi \), \( M(f_0) = O_p(1) \),
\[
\max_{1 \leq j, j' \leq r} \sup_{\ell \in \mathcal{L}} \left\| M \left( \frac{\partial^3 f_\ell}{\partial \ell_j \partial \ell_{j'} \partial \ell_{j''}} \right) \right\|_2 = O_p(1).
\]
Similarly, we have
\[
\max_{1 \leq j, j', j'' \leq r} \sup_{\ell \in \mathcal{L}} \left| \hat{\Omega} \left( \frac{\partial^3 f_\ell}{\partial \ell_j \partial \ell_{j'} \partial \ell_{j''}}, \gamma_n \right) \right| = O_p(1).
\]
An application of Taylor expansion shows that
\[
\left\| \sqrt{n} M(f_{\ell^*_n}) - \hat{\Omega}(f_{\ell^*_n}, \gamma_n) I_q \right\|_2 - \sqrt{n} \left\| M(f_{\ell^*}) - \hat{\Omega}(f_{\ell^*}, \gamma_n) I_q \right\|_2
\]
Moreover, \( \hat{\Delta}(f_{\ell_n^*}, \gamma_n) \xrightarrow{p} \Delta(f_{\ell_n^*}, \gamma) \) and (3.5) follows.

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

**S.7. Proof of Theorem 3.2**

Following from the lines in the proof of Theorem 3.1,

\[
\hat{T}(f_{\ell_n^*}) = \hat{T}(f_{\ell_n^*}) + o_p(1),
\]

by Slutsky’s Theorem, we only need to show

\[
\left( \hat{T}(f_{\ell_n^*}), \ldots, \hat{T}(f_{\ell_n^*}) \right) \implies \mathcal{N}(0, \Delta^*).
\]

Observe that the asymptotic normality of \( \hat{T}(f_{\ell_n^*}) \) follows from that of the un-standardized statistic, \( T(f_{\ell_n^*}) \). It suffices to show that \( \left( T(f_{\ell_n^*}), \ldots, T(f_{\ell_n^*}) \right) \) is asymptotically normal.

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

\[
T(f_{\ell_n^*}) = T^{LH}(f_{\ell_n^*}) = \sum_{j=1}^{q} \lambda_j(M(f_{\ell_n^*})),
\]

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

\[
\sum_{i=1}^{m} a_i T^{LH}(f_{\ell_n^*}) = T^{LH} \left( \sum_{i=1}^{m} a_i f_{\ell_n^*} \right).
\]

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

\[
\left( \hat{T}(f_{\ell_n^*}), \ldots, \hat{T}(f_{\ell_n^*}) \right)
\]

can be verified to be \( \Delta^* \) via elementary calculation.

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

\[
T^{LR}(f_{\ell_n^*}) = q \log(1 + \hat{\Omega}(f_{\ell_n^*}, \gamma_n)) + \frac{T^{LH}(f_{\ell_n^*}) - q\hat{\Omega}(f_{\ell_n^*}, \gamma_n)}{1 + \hat{\Omega}(f_{\ell_n^*}, \gamma_n)} + o_p(n^{-1/2}),
\]
\[ T_{\text{BNP}}(f_{t^*}) = \frac{q\widehat{\Omega}(f_{t^*}, \gamma_n)}{1 + \widehat{\Omega}(f_{t^*}, \gamma_n)} + \frac{T_{\text{LH}}(f_{t^*}) - q\widehat{\Omega}(f_{t^*}, \gamma_n)}{(1 + \widehat{\Omega}(f_{t^*}, \gamma_n))^2} + o_p(n^{-1/2}). \]

It implies that any linear combination of \( T_{\text{LH}}(f_{t^*}) \) or \( T_{\text{BNP}}(f_{t^*}) \) can be expressed as a linear combination of \( T_{\text{LH}}(f_{t^*}) \) with a negligible remainder. The proof is complete.

### S.8. Proof of Remark 2.2, Lemma 2.1 and Lemma 2.2

#### S.8.1. Proof of Remark 2.2

The results in Section S.3.1.7 imply that \( \delta(z_1, z_2, \gamma) \) is the limit in \( L_1 \)-norm of \( n^{-1} \text{tr}[A(z_1)A(z_2)] \) pointwise on \((C^+)^2\), when the random variables \( z_{ij} \)'s satisfy the truncated variable condition (A.4). Due to very similar lines to (S.3.21) and (S.3.25), we have

\[
\sup_{z_1, z_2 \in C^+} \mathbb{E} \left| n^{-1} \text{tr}[A(z_1)A(z_2)] \right| < \infty.
\]

Besides, in Section S.2, it is shown that \( \delta(z_1, z_2, \gamma) \) and its partial derivatives are bounded on \( C \).

It follows from *Fubini’s Theorem* and *Dominated Convergence Theorem* that

\[
\frac{2}{(2\pi)^2} \iint_{(C^+)^2} f(z_1) f(z_2) \frac{1}{n} \text{tr}[A(z_1)A(z_2)] dz_1 dz_2 \xrightarrow{p} \Delta(f, \gamma).
\]

On the other hand, when \( \lambda_{\max}(\hat{\Sigma}_p) < D \), for some \( D \in (\limsup_{\gamma} \lambda_{\max}(\Sigma_p)(1 + \sqrt{\gamma})^2, \infty) \), \( \|A(z)\| \) is bounded. Therefore,

\[
\iint_{C^2 \in (C^+)^2} \left| f(z_1) f(z_2) \frac{1}{n} \text{tr}[A(z_1)A(z_2)] \right| dz_1 dz_2 \xrightarrow{L^1} 0.
\]

Providing that \( f(x) \geq 0 \), for all sufficiently large \( n \), with high probability,

\[
\frac{2}{(2\pi)^2} \iint_{C^2} f(z_1) f(z_2) \frac{1}{n} \text{tr}[A(z_1)A(z_2)] dz_1 dz_2 = \frac{1}{n} \text{tr}[f(\hat{\Sigma}_p)\Sigma_p f(\hat{\Sigma}_p)\Sigma_p] \geq \max \left\{ \lambda^2_{\min}(\Sigma_p), \left[ \frac{1}{n} \text{tr}(f(\hat{\Sigma}_p)) \right]^2, \lambda^2_{\min}(f(\hat{\Sigma}_p))\left[ \frac{1}{n} \text{tr}(\Sigma_p) \right]^2 \right\}.
\]

It is assumed that \( F_{\Sigma_p} \) converges to \( L^2 \) 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,

\[
\lambda^2_{\min}(f(\hat{\Sigma}_p))\left[ \frac{1}{n} \text{tr}(\Sigma_p) \right]^2 > K > 0.
\]

It follows that \( \Delta(f, \gamma) > 0 \).
If \( f(x) \) is only nonnegative, \( f(x) > 0 \) for some \( x \in \mathcal{X} \), and \( \lim \inf_p \lambda_{\min}(\Sigma_p) > 0 \), we only need to show that \( n^{-1} \text{tr}[f(\Sigma_p)] > K > 0 \) with high probability for all sufficiently large \( n \). By Bai and Silverstein (2004, Theorem 1.1),

\[
\int f(\tau)dF_\Sigma^\tau(\tau) \overset{p}{\to} \int f(\tau)dF^\infty(\tau).
\]

It follows that \( \Delta(f, \gamma) > 0 \) if \( \int f(\tau)dF^\infty(\tau) > 0 \). Due to Identity Theorem, \( f(\cdot) \) can not be constantly 0 on any open subset of \( \mathcal{X} \). Otherwise, \( f(x) = 0 \) for any \( x \in \mathcal{X} \). Together with the continuity of \( F^\infty \), it follows that \( \int f(\tau)dF^\infty(\tau) > 0 \).

The proof of Remark 2.2 is complete.

**S.8.2. Proof of Lemma 2.1**

Lemma 2.1 can be deduced by combining Lemma A.3 and Lemma S.3 shown in Section S.9.

**S.8.3. Proof of Lemma 2.2**

Observe that \( \Theta(z, \gamma) \) and \( \Delta(z, \gamma) \) are smooth functions of \( m(z) \) and \( m'(z) \). We further know that \( \Theta(z, \gamma) \) is bounded on \( \mathcal{C} \) (see Section S.2). Also, it is straightforward to check that \( \hat{\Theta}(z, \gamma_n) \) is bounded on \( \mathcal{C} \) when \( \lambda_{\max}(\hat{\Sigma}_p) < \mathcal{D} < \pi \). Consequently, it suffices to show the uniform convergence of \( m_{n,p}(z) \) and \( m'_{n,p}(z) \) on \( \mathcal{C}^+ \), that is,

\[
\sup_{z \in \mathcal{C}^+} \sqrt{n}|m_{n,p}(z) - m(z)| \overset{p}{\to} 0, \quad (S.8.1)
\]

\[
\sup_{z \in \mathcal{C}^+} \sqrt{n}|m'_{n,p}(z) - m'(z)| \overset{p}{\to} 0. \quad (S.8.2)
\]

Define \( \hat{\Sigma}_p = \frac{1}{n} \sum_{j=1}^n \Sigma_p^{1/2} z_j z_j^T \Sigma_p^{T/2} \) and

\[
\hat{m}(z) = p^{-1} \text{tr}\left((\hat{\Sigma}_p - zI_p)^{-1}\right).
\]

Then, due to the rank inequality,

\[
\left\| F_\hat{\Sigma}_p - F_\Sigma \right\|_{\infty} \leq \frac{1}{p} \text{rank}(\hat{\Sigma}_p - \Sigma_p) \leq \frac{2k}{p}.
\]

Therefore,

\[
\sup_{z \in \mathcal{C}^+} |\hat{m}(z) - m_{n,p}(z)| = o_p(n^{-1/2}),
\]

\[
\sup_{z \in \mathcal{C}^+} \left| \frac{d}{dz} \hat{m}(z) - \frac{d}{dz} m_{n,p}(z) \right| = o_p(n^{-1/2}).
\]
Together with Lemma S.1 and Lemma S.2, we only need to show

\[
\sup_{z \in \mathbb{C}^+} \sqrt{n} |\dot{m}(z) - m^0_p(z)| \xrightarrow{P} 0,
\]

\[
\sup_{z \in \mathbb{C}^+} \sqrt{n} \left| \frac{d}{dz} \dot{m}(z) - \frac{d}{dz} m^0_p(z) \right| \xrightarrow{P} 0.
\]

The convergence of \( \dot{m}(z) \) to \( m^0_p(z) \) is shown in Bai and Silverstein (2004) under (A.4). It indeed holds under \( C_1 \) (see Lemma S.5 in Section S.9). The proof of S.8.1 is complete.

As for the uniform convergence of \( \frac{d}{dz} \dot{m}(z) \), again, we first consider the truncated variables satisfying (A.4), with generalization to \( C_1 \) addressed in Lemma S.5 in Section S.9. We first show

\[
\sup_{z \in \mathbb{C}^+} \sqrt{n} \left| \frac{d}{dz} \dot{m}(z) - \mathbb{E} \frac{d}{dz} \dot{m}(z) \right| \xrightarrow{P} 0.
\]

The convergence of finite-dimensional distributions of the process \( \sqrt{n} \left| \frac{d}{dz} \dot{m}(z) - \mathbb{E} \frac{d}{dz} \dot{m}(z) \right| \) is a direct consequence of Bai and Silverstein (2004). We only need to show its tightness. Using the same strategy as in Section S.3.2, it suffices to show that there exists a constant \( K \) such that for any \( z_1 \neq z_2 \in \mathbb{C}^+ \),

\[
\mathbb{E}_n \left[ \left( \frac{d}{dz} \dot{m}(z_1) - \mathbb{E} \frac{d}{dz} \dot{m}(z_1) \right) - \left( \frac{d}{dz} \dot{m}(z_2) - \mathbb{E} \frac{d}{dz} \dot{m}(z_2) \right) \right]^2
\]

\[
\leq 2n \mathbb{E} \frac{1}{p} \text{tr} \left[ (\Sigma_p - z_1 I)^{-2}(\Sigma_p - z_2 I)^{-1} \right] - \mathbb{E} \frac{1}{p} \text{tr} \left[ (\Sigma_p - z_1 I)^{-2}(\Sigma_p - z_2 I)^{-1} \right]^2
\]

\[
+ 2n \mathbb{E} \frac{1}{p} \text{tr} \left[ (\Sigma_p - z_1 I)^{-1}(\Sigma_p - z_2 I)^{-2} \right] - \mathbb{E} \frac{1}{p} \text{tr} \left[ (\Sigma_p - z_1 I)^{-1}(\Sigma_p - z_2 I)^{-2} \right]^2
\]

\[
\leq K.
\]

Define \( \Sigma^{(j)}_p = \frac{1}{n} \sum_{i \neq j}^{n} \Sigma^{1/2}_p z_i z_i^T \Sigma^T_p \). We can write

\[
\text{tr} \left[ (\Sigma_p - z_1 I)^{-2}(\Sigma_p - z_2 I)^{-1} \right] - \mathbb{E} \text{tr} \left[ (\Sigma_p - z_1 I)^{-2}(\Sigma_p - z_2 I)^{-1} \right]
\]

\[
= \sum_{j=1}^{n} \left( \mathbb{E}_j - \mathbb{E}_{j-1} \right) \left\{ \text{tr} \left[ (\Sigma_p - z_1 I)^{-2}(\Sigma_p - z_2 I)^{-1} \right] \right\}
\]

\[
- \text{tr} \left[ (\Sigma^{(j)}_p - z_1 I)^{-2}(\Sigma^{(j)}_p - z_2 I)^{-1} \right].
\]

The rest of the proof is very similar to the lines presented in Section S.3.2, also Section 3 of Bai and Silverstein (2004). Lemma S.8 is critical in the proof. We omit details.

We next show

\[
\sup_{z \in \mathbb{C}^+} \sqrt{n} \left| \mathbb{E} \frac{d}{dz} \dot{m}(z) - \frac{d}{dz} m^0_p(z) \right| \rightarrow 0. \tag{S.8.3}
\]

It suffices to show \( \sup_{z \in \mathbb{C}^+} \sqrt{n} \left| \mathbb{E} \frac{d}{dz} \dot{m}(z) - \frac{d}{dz} m^0_p(z) \right| \rightarrow 0 \), where \( \dot{m}(z) = \frac{\gamma_n - 1}{z} + \gamma_n \dot{m}(z) \).
Following notation in Section S.3.1 and Section S.4, it is shown in Bai and Silverstein (1998, (5.2)) that

$$
\gamma_n \int \frac{dF_{\Sigma_p}(\tau)}{1 + \tau E\hat{m}(z)} + z\gamma_n E\hat{m}(z) = A_n(z),
$$

where

$$
A_n(z) = E\hat{\beta}_1(z)\left[\frac{1}{n}z_1^T \Sigma_p^{(1)} - zI\right]^{-1}(E\hat{m}(z)\Sigma_p + I)^{-1}\Sigma_p^{1/2} z_1
- \frac{1}{n} E\text{tr}\left[(E\hat{m}(z)\Sigma_p + I)^{-1}\Sigma_p(z - zI)^{-1}\right],
$$

$$
\hat{\beta}_1(z) = \frac{1}{1 + n^{-1}z_1^T \Sigma_p^{T/2}(\Sigma_p - zI)^{-1}\Sigma_p^{1/2} z_1}.
$$

It is shown that $A_n(z) = o(n^{-1/2})$ uniformly on $C^+$ in Bai and Silverstein (2004) (see (4.19)–(4.11)).

Next, we first take differentiation on both sides with respect to $z$, and then interchange differentiation and integration. Note that, similar to (S.3.20), (S.3.24), we have on $z \in C^+$,

$$
E\|\Sigma_p - zI\|_{1+\ell} < \infty, \text{ for any } \ell \in \mathbb{N}^{+},
$$

$$
E\left|\frac{d}{dz} \hat{\beta}_1(z)\right| = E\left|\frac{d}{dz} \left(1 - n^{-1}z_1^T \Sigma_p^{T/2}(\Sigma_p - zI)^{-1}\Sigma_p^{1/2} z_1\right)\right|
= E\left|n^{-1}z_1^T \Sigma_p^{T/2}(\Sigma_p - zI)^{-2}\Sigma_p^{1/2} z_1\right| < \infty.
$$

Moreover, Bai and Silverstein (2004, (4.3)) claims that

$$
\sup_{z \in C^+} \|(E\hat{m}(z)\Sigma_p + I)^{-1}\| < \infty.
$$

Due to the Dominated Convergence Theorem,

$$
\frac{d}{dz} A_n(z)
= E\frac{d}{dz} \hat{\beta}_1(z)\left[\frac{1}{n}z_1^T \Sigma_p^{(1)} - zI\right]^{-1}(E\hat{m}(z)\Sigma_p + I)^{-1}\Sigma_p^{1/2} z_1
- \frac{1}{n} E\text{tr}\left[(E\hat{m}(z)\Sigma_p + I)^{-1}\Sigma_p(z - zI)^{-1}\right]
+ E\hat{\beta}_1(z)\left[\frac{1}{n}z_1^T \Sigma_p^{(1)} - zI\right]^{-2}(E\hat{m}(z)\Sigma_p + I)^{-1}\Sigma_p^{1/2} z_1
- \frac{1}{n} E\text{tr}\left[(E\hat{m}(z)\Sigma_p + I)^{-1}\Sigma_p - zI\right]^{-2}\]
+ E\hat{\beta}_1(z)\left[\frac{1}{n}z_1^T \Sigma_p^{(1)} - zI\right]^{-1}\left\{\frac{d}{dz} (E\hat{m}(z)\Sigma_p + I)^{-1}\right\}\Sigma_p^{1/2} z_1
- \frac{1}{n} E\text{tr}\left\{\frac{d}{dz} (E\hat{m}(z)\Sigma_p + I)^{-1}\right\}\Sigma_p(z - zI)^{-1}].
High-dimensional general linear hypothesis via spectral shrinkage

\[ = d_{15} + d_{16} + d_{17}, \text{ say.} \]

Follow analogous lines as those in Section S.3.2, Section S.3.3, and Section S.4, we can show \( \sup_{x \in \mathbb{C}^+} \sqrt{n}[|d_{15}| + |d_{16}| + |d_{17}|] = o(1) \). Details are omitted.

Bai and Silverstein (2004, (4.12)) indicates

\[
E \tilde{m}(z) - m_p^0(z) = -m_p^0(z)A_n(z) \left[ 1 - \gamma_n E \tilde{m}(z)m_p^0(z) \int \frac{\tau^2 dF_S^\tau(\tau)}{(1 + \tau E \tilde{m}(z))(1 + \tau m_p^0(z))} \right]^{-1}.
\]

It is also claimed that the denominator on the right-hand side is bounded away from zero. Take differentiation on both sides,

\[
E \frac{d}{dz} \tilde{m}(z) - \frac{d}{dz} m_p^0(z)
\]

\[
= - \left[ \frac{d}{dz} m_p^0(z) \right] A_n(z) \left[ 1 - \gamma_n E \tilde{m}(z)m_p^0(z) \int \frac{\tau^2 dF_S^\tau(\tau)}{(1 + \tau E \tilde{m}(z))(1 + \tau m_p^0(z))} \right]^{-1}
\]

\[
- m_p^0(z) \left[ \frac{d}{dz} A_n(z) \right] \left[ 1 - \gamma_n E \tilde{m}(z)m_p^0(z) \int \frac{\tau^2 dF_S^\tau(\tau)}{(1 + \tau E \tilde{m}(z))(1 + \tau m_p^0(z))} \right]^{-1}
\]

\[
- m_p^0(z)A_n(z) \frac{d}{dz} \left[ 1 - \gamma_n E \tilde{m}(z)m_p^0(z) \int \frac{\tau^2 dF_S^\tau(\tau)}{(1 + \tau E \tilde{m}(z))(1 + \tau m_p^0(z))} \right]^{-1}.
\]

It follows that \( E \tilde{m}(z) - m_p^0(z) = o(n^{-1/2}) \) uniformly on \( \mathbb{C}^+ \).

The proof of Lemma 2.2 is complete.

**S.9. Random variable truncation and process smoothing**

In this section, we present the proof of the following claims.

1. (A.6) still holds if the observation errors satisfy \( \mathbf{C1} \) instead of (A.4).
2. When the observation errors satisfy \( \mathbf{C1} \),

\[
\left| \int_C f(z) \xi_n(z, \alpha, \eta)dz - \int_C f(z) \xi_n(z, \alpha, \eta)dz \right| \xrightarrow{P} 0.
\]

3. (S.4.1) and (S.4.2) still hold if the observation errors satisfy \( \mathbf{C1} \) instead of (A.4).

Claim 1 and Claim 2, together with Appendix, complete the proof of Theorem 2.1. Claim 3 completes the proof of Theorem 2.3 by extending results in Section S.4 from (A.4) to \( \mathbf{C1} \).

Recall the convention made in Appendix that \( \tilde{z}_{ij} \)'s denotes observation errors satisfying \( \mathbf{C1} \) and \( \bar{z}_{ij} \) is the truncated random variable satisfying (A.4). The relation between \( \tilde{z}_{ij} \)
and $z_{ij}$ is such that

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

We further define $\bar{Z}$ and $\bar{\Sigma}_n$ by mimicking the definition of $Z$ and $\hat{\Sigma}_n$, but with $z_{ij}$ replaced by $\bar{z}_{ij}$. Similarly, $\bar{m}_{n,p}(z)$ and $\bar{m}'_{n,p}(z)$ are the counterparts of $m_{n,p}(z)$ and $m'_{n,p}(z)$.

To show Claim 1, we only need to prove the following lemma.

**Lemma S.3** For any fixed $\alpha$ and $\eta$,\[
\sup_{x \in \mathbb{C}^n} n^{-1/2} \left| \alpha^T Q_n^T Z^T \Sigma_p^{1/2} (\hat{\Sigma}_n - zI) - 1 \Sigma_p^{1/2} \hat{Z} Q_n \eta \right| - \alpha^T Q_n^T Z^T \Sigma_p^{1/2} (\hat{\Sigma}_n - zI) - 1 \Sigma_p^{1/2} \hat{Z} Q_n \eta \mathop\xrightarrow{P} 0.
\]

**Proof** By Yin, Bai and Krishnaiah (1988),\[
\lambda_{\max}(\hat{\Sigma}_n) \leq \limsup_{p \to \infty} \lambda_{\max}(\Sigma_p) \lambda_{\max}\left(\frac{1}{n} ZZ^T\right) \xrightarrow{a.s.} \limsup_{p \to \infty} \lambda_{\max}(\Sigma_p)(1 + \sqrt{\gamma})^2.
\]

Together with Lemma A.1, we only need to consider the case $\lambda_{\max}(\hat{\Sigma}_n) < D$ and $\lambda_{\max}(\hat{\Sigma}_n) < D$ for a constant $D$ such that\[
(1 + \sqrt{\gamma})^2 \limsup_{p \to \infty} \lambda_{\max}(\Sigma_p) < D < \pi.
\]

It follows that, under this event,\[
\sup_{x \in \mathbb{C}^n} \| (\hat{\Sigma}_n - zI)^{-1} \|_2 \leq (\pi - D)^{-1} + |\eta|^{-1},
\]
\[
\sup_{x \in \mathbb{C}^n} \| (\hat{\Sigma}_n - zI)^{-1} \|_2 \leq (\pi - D)^{-1} + |\eta|^{-1}.
\]

Observe\[
n^{-1/2} \alpha^T Q_n^T Z^T \Sigma_p^{1/2} (\hat{\Sigma}_n - zI)^{-1} \Sigma_p^{1/2} \hat{Z} Q_n \eta - \alpha^T Q_n^T Z^T \Sigma_p^{1/2} (\hat{\Sigma}_n - zI)^{-1} \Sigma_p^{1/2} Z Q_n \eta
\]
\[
= n^{-1/2} \alpha^T Q_n^T (Z - \bar{Z})^T \Sigma_p^{1/2} (\hat{\Sigma}_n - zI)^{-1} \Sigma_p^{1/2} \bar{Z} Q_n \eta
\]
\[
+ n^{-1/2} \alpha^T Q_n^T \bar{Z}^T \Sigma_p^{1/2} (\hat{\Sigma}_n - zI)^{-1} \Sigma_p^{1/2} Z Q_n \eta
\]
\[
+ n^{-1/2} \alpha^T Q_n^T \bar{Z}^T \Sigma_p^{1/2} (\hat{\Sigma}_n - zI)^{-1} \Sigma_p^{1/2} (Z - \bar{Z}) Q_n \eta
\]
\[
= d_{18} + d_{19} + d_{20}, \text{ say.}
\]

Therefore, we can find a constant $K$ sufficiently large such that\[
|d_{18}| \leq n^{-1/2} K \| \alpha^T Q_n^T (Z - \bar{Z})^T \|_2 \| Z Q_n \eta \|_2.
\]
For the second line, since
\[ \| (Z - \tilde{Z}) Q_n \alpha \|_2^2 \rightarrow P \cdot 0. \]

Since
\[
E_{\alpha}^T Q_n^T (Z - \tilde{Z})^T (Z - \tilde{Z}) Q_n \alpha
\]
\[= E \sum_{i=1}^{N} \sum_{j=1}^{N} [Q_n \alpha_i][Q_n \alpha_j] \sum_{k=1}^{P} (z_{ik} - \tilde{z}_{ik})(z_{jk} - \tilde{z}_{jk}) = \| \alpha \|^2 pE (z_{ij} - \tilde{z}_{ij})^2
\]
and
\[
\tilde{z}_{ij} - z_{ij} = E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2}) - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2})
\]
\[+ \tilde{z}_{ij} \left[ 1 - \left\{ E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2}) - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2}) \right\}^2 \right]^{-1/2},
\]
we only need to show
\[
E p[\tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2}) - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2})]^2 \longrightarrow 0,
\]
\[p^{1/2} \left[ 1 - \left\{ E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2}) - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2}) \right\}^2 \right]^{-1/2} \longrightarrow 0.
\]
Note, $\varepsilon_n$ is such that $\varepsilon_n \rightarrow 0$ and
\[
\varepsilon_n^{-4} E z_{ij} 1(|z_{ij}| \geq \varepsilon_n n^{1/2}) \rightarrow 0.
\]
For the first line above, we have
\[
E p[\tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2}) - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2})]^2 \leq 2 n^{4} E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2})
\]
\[\leq 2 \gamma_n \varepsilon_n^{-2} E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2}) \longrightarrow 0.
\]
For the second line, since
\[
(1 - 1/\sqrt{x}) = (x - 1)/(\sqrt{x} + x),
\]
we only need to show
\[
p \left[ 1 - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2}) - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2}) \right] \longrightarrow 0.
\]
Indeed,
\[
1 - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2}) - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2})
\]
\[= E \tilde{z}_{ij}^2 - E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2}) + (E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| \leq \varepsilon_n n^{1/2})]^2
\]
\[= E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2}) + [E \tilde{z}_{ij} 1(|\tilde{z}_{ij}| > \varepsilon_n n^{1/2})]^2.
\]
define completes the proof of Lemma S.3.

Therefore, $|d_{18}| \xrightarrow{P} 0$.

$$d_{19} = n^{-1/2} \alpha^T Q_n^T \tilde{Z}^T \Sigma_p^T/2 [ (\tilde{\Sigma}_p - zI)^{-1} - (\Sigma_p - zI)^{-1} ] \Sigma_p^{1/2} Z Q_n \eta$$

$$= n^{-1/2} \alpha^T Q_n^T \tilde{Z}^T \Sigma_p^T/2 [ (\tilde{\Sigma}_p - zI)^{-1} - (\Sigma_p - zI)^{-1} ] \Sigma_p^{1/2} Z Q_n \eta.$$ 

It is easy to check that Lemma S.15 still holds for $\tilde{Z}$ when $m = 2$. We have

$$n^{-1} \mathbb{E} \alpha^T Q_n^T \tilde{Z}^T \tilde{Z} Q_n \alpha = O(1).$$

Thus, we only need to show,

$$n^{1/2} |\tilde{\Sigma}_p - \Sigma_p|_2 \leq n^{-1/2} \|Z - \tilde{Z}\|_2 (\|Z\|_2 + \|\tilde{Z}\|_2) \xrightarrow{P} 0.$$ (S.9.1)

Since $n^{-1/2} \|Z\|_2$ and $n^{-1/2} \|\tilde{Z}\|_2$ are $O_p(1)$ (Yin, Bai and Krishnaiah, 1988), we only need to show

$$\|\tilde{Z} - Z\|_2 \xrightarrow{P} 0.$$

$$\tilde{Z} - Z = \left[ 1 - \left\{ \mathbb{E} [\tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| \leq \varepsilon_n n^{1/2}) - \mathbb{E} \tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| \leq \varepsilon_n n^{1/2})] \right\}^{-1/2} \right] Z$$

$$+ \left\{ \mathbb{E} [\tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| \leq \varepsilon_n n^{1/2}) - \mathbb{E} \tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| \leq \varepsilon_n n^{1/2})] \right\}^{-1/2} F,$$

where $F$ is a matrix with the $(i, j)$-th entry being $\tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| > \varepsilon_n n^{1/2}) - \mathbb{E} \tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| > \varepsilon_n n^{1/2})$. It remains to show $\|F\|_2 \xrightarrow{P} 0$.

Note, $F$ has i.i.d. entries with mean 0 and variance

$$\mathbb{E} [\tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| > \varepsilon_n n^{1/2}) - \mathbb{E} \tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| > \varepsilon_n n^{1/2})]^2 = o(p^{-1}).$$

Again using results in Yin, Bai and Krishnaiah (1988), almost surely,

$$\left[ \mathbb{E} [\tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| > \varepsilon_n n^{1/2}) - \mathbb{E} \tilde{\varepsilon}_{ij} 1(|\tilde{\varepsilon}_{ij}| > \varepsilon_n n^{1/2})]^2 \right]^{-1/2} n^{-1/2} \|F\|_2 = O_p(1).$$

Therefore, $\|F\|_2 \xrightarrow{P} 0$.

As for $d_{20}$, the argument for $d_{18}$ works, since $n^{-1/2} \|\alpha^T Q_n^T \tilde{Z}^T\|_2 = O_p(1)$ also holds. It completes the proof of Lemma S.3.

To show Claim 2, for notational simplicity, we will not change the notation but redefine $\xi_n(z, \alpha, \beta)$ and $\zeta_n(z, \alpha, \beta)$ with $Z$ replaced by $\tilde{Z}$. When $\lambda_{\max}(\Sigma_p) < \pi$,

$$\left| \int_C f(z) \tilde{\xi}_n(z, \alpha, \beta) dz - \int_C f(z) \xi_n(z, \alpha, \beta) dz \right|$$ (S.9.2)
\[ \leq \mathcal{K}_n \rho n^{-1/2} \| \tilde{Z} \tilde{Z}^T \|_2 \| \alpha \|_2 \| \eta \|_2 (|\pi - \lambda_{\text{max}}(\tilde{\Sigma}_p)|^{-1} + |\xi|^{-1}) \].

It is straightforward to see that the right-hand side above converges to 0 in probability since \( \rho_n \rightarrow 0 \). The proof of Claim 2 is complete.

The results so far indicate that under \( C_1 \),
\[ -\frac{1}{2\pi i} \oint_C f(z) \xi_n(z, \alpha, \eta) dz - n^{1/2} \Omega(f, \gamma) \alpha^T \eta \Longrightarrow N(0, \|\alpha\|_2^2 \|\eta\|_2^2 + (\alpha^T \eta)^2) \Delta(f, \gamma). \]

As stated in Appendix, when \( \lambda_{\text{max}}(\tilde{\Sigma}_p) < \pi \),
\[ n^{-1/2} \alpha^T V_n^T U_n^T \tilde{Z} \Sigma_p^{1/2} f(\tilde{\Sigma}_p) \Sigma_p^{1/2} \tilde{Z} U_n V_n \eta = -\frac{1}{2\pi i} \oint_C f(z)n^{-1/2} \alpha^T V_n^T U_n^T \tilde{Z} \Sigma_p^{1/2} (\tilde{\Sigma}_p - zI)\Sigma_p^{1/2} \tilde{Z} U_n V_n \eta dz. \]

Therefore, for arbitrary \( \alpha \) and \( \eta \),
\[ \sqrt{n} \alpha \mathbf{M}(f) \eta - n^{1/2} \Omega(f, \gamma) \alpha^T \eta \Longrightarrow N(0, \|\alpha\|_2^2 \|\eta\|_2^2 + (\alpha^T \eta)^2) \Delta(f, \gamma). \]

It completes the proof of Theorem 2.1.

As for Claim 3, it suffices to show the following lemma.

**Lemma S.4**
\[
\sup_{x \in \mathcal{C}^+} \left\| Q_n^T \tilde{Z} \Sigma_p^{1/2} (\tilde{\Sigma}_p - zI)^{-1} BCT_n^{-1/2} - Q_n^T \tilde{Z} \Sigma_p^{1/2} (\tilde{\Sigma}_p - zI)^{-1} BCT_n^{-1/2} \right\|_2 \xrightarrow{P} 0,
\]
\[
\sup_{x \in \mathcal{C}^+} \sqrt{n} \left\| T_n^{-1/2} C B^T (\tilde{\Sigma}_p - zI)^{-1} BCT_n^{-1/2} - T_n^{-1/2} C B^T (\tilde{\Sigma}_p - zI)^{-1} BCT_n^{-1/2} \right\|_2 \xrightarrow{P} 0.
\]

To complete the proof of Lemma 2.2 presented in Section S.8.3, we need the following lemma.

**Lemma S.5**
\[
\sup_{x \in \mathcal{C}^+} \sqrt{n} |m_{n,p}(x) - \bar{m}_{n,p}(x)| \xrightarrow{P} 0,
\]
\[
\sup_{x \in \mathcal{C}^+} \sqrt{n} |m'_{n,p}(x) - \bar{m}'_{n,p}(x)| \xrightarrow{P} 0.
\]

The two lemmas can be proved using analogous lines as the proof of Lemma S.3. We omit details.
S.10. Technical lemmas

We present a collection of lemmas when the random variables satisfy (A.4). That is, for some constant $K$, when $n$ is sufficiently large,

$$|z_{ij}| \leq K \varepsilon_n n^{1/2}, \quad E[z_{ij}] = 0, \quad E[z_{ij}^2] = 1, \quad E[z_{ij}^4] < \infty,$$

where $\varepsilon_n$ is a positive sequence such that

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

See Appendix for details. Proofs of these lemmas are omitted because similar results can be found in the literature. See for example (3.10)-(3.14) of Pan and Zhou (2011). Recall the notation list. The following matrices can be either real- or complex-valued.

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

$$s_{i+j+1}(A + C) \leq s_{i+1}(A) + s_{j+1}(C),$$

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 $\sigma$-field $\{\sigma_i\}$. Then for $m \geq 2$

$$E\left[\sum_i |Y_i|^m\right] \leq K_m E\left(\sum_i E(\|Y_i\|^2 | \sigma_{i-1})\right)^{m/2} + K_m E\left(\sum_i |Y_i|^m\right).$$

Lemma S.9 (Lemma 2.7 of Bai and Silverstein (1998)) Let $Y = (Y_1, \ldots, Y_p)^T$, where $Y_i$’s are i.i.d. real r.v.’s with mean 0 and variance 1. Let $B = (b_{ij})_{p \times p}$, a deterministic complex matrix. Then for any $m \geq 2$, we have

$$E|Y^TBY - \text{tr}B|^m \leq K_m (EY_i^4 \text{tr}B^*B^*)^{m/2} + K_mEY_i^{2m}\text{tr}(B^*B)^{m/2},$$

where $B^*$ denotes the complex conjugate transpose of $B$, and $K$ is a constant only depending on $m$.

Lemma S.10 ((3.4) of Bai and Silverstein (1998)) For any $z = u + iv$ with $v > 0$,

$$\|A(z)\|_2 \leq \frac{\|\Sigma_p\|}{v}, \quad |\beta_1(z)| \leq \frac{|z|}{v}, \quad |\beta_1^\text{tr}(z)| \leq \frac{|z|}{v}. $$
Lemma S.11 (Lemma 2.10 of Bai and Silverstein (1998)) For any matrix $D$ and $z = u + iv$ with $v > 0$,
\[
\left| \text{tr}(A(z)D - A_j(z)D) \right| \leq \frac{|D\Sigma_p|_2}{v}.
\]

Lemma S.12 For $m \geq 2$ and any fixed $z$ with non-zero imaginary part,
\[
n^{-m}E\left| \text{tr}A(z) - \text{E} \text{tr}A(z) \right|^{m} = O(n^{-m/2}).
\]

Lemma S.13 For a sequence of deterministic matrices $D$ such that $\|D\|_2 < \infty$, $m \geq 2$,
\[
n^{-m}E\left| z_1^T Dz_1 - \text{tr}(D) \right|^{m} \leq K_m n^{-m} \left\{ \left( \text{tr}(DD^*) \right)^{m/2} + \varepsilon_n^{2m-4} n^{m-2} \text{tr}(DD^*)^{m/2} \right\} \leq K_m \|D\|_2^m \varepsilon_n^{2m-4} n^{-1}
\]
for some constant $K_m$.

Lemma S.14 For sequences of deterministic matrices $D$ and $G$ such that $\|D\|_2 < \infty$ and $\|G\|_2 < \infty$, for $m \geq 2$,
\[
n^{-m}E\left| z_1^T Dc_i e_j^T Gz_1 \right|^{m} \leq K_m \|D\|_2^m \|G\|_2^m \varepsilon_n^{2m-4} n^{-2} + n^{-m} = O(n^{-2} \varepsilon_n^{2m-4}),
\]
for some constant $K_m$.

Lemma S.15 For a sequence of deterministic matrices $D$ such that $\|D\|_2 < \infty$ and a sequence of vector $u$ such that $\limsup_{n \to \infty} n \|u\|_{\text{max}} \leq K_{\text{max}} < \infty$,
\[
E\left| u^T Z^T DZ u \right|^{m} \leq K_m K_{\text{max}}^m \|D\|_2^m,
\]
for $m \geq 2$ and some constant $K_m > 0$.

Lemma S.16 For a sequence of deterministic matrices $D$ such that $\|D\|_2 < \infty$ and a sequence of vector $u$ such that $\limsup_{n \to \infty} n \|u\|_{\text{max}} \leq K_{\text{max}} < \infty$,
\[
E\left| z_1^T DZ_1 u \right|^{m} \leq K_m K_{\text{max}}^m \|D\|_2^m n^{m/2-2} \varepsilon_n^{m-4},
\]
for $m \geq 4$ and some constant $K_m > 0$.

Lemma S.17 For a sequence of deterministic matrices $D$ such that $\|D\|_2 < \infty$,
\[
E\left| z_1^T Dz_2 \right|^{m} \leq K_m \|D\|_2^m n^{m-2} \varepsilon_n^{m-4},
\]
for $m \geq 4$ and some constant $K_m > 0$.

Lemma S.18 For sequences of deterministic matrices $D_1, \ldots, D_m, G_1, \ldots, G_s, J$ such that $\|D_j\|_2 < \infty$, $\|G_j\|_2 < \infty$, and $\|J\|_2 < \infty$, and a sequence of vector $u$ such that $\limsup_{n \to \infty} n \|u\|_{\text{max}} \leq K_{\text{max}} < \infty$,
\[
E\left| \prod_{i=1}^{m} \frac{1}{n} z_i^T D_i z_i \prod_{j=1}^{s} \frac{1}{n} (z_i^T G_j z_1 - \text{tr}G_j) (z_i^T J z_1 u) \right|^{m}.
\]
\[ \leq K_{m,s,t} \prod_{i=1}^{m} \| D_i \|_2 \prod_{j=1}^{s} \| G_j \|_2 J^{x^t} J_{\max}^{-1/2} E_{\| \varepsilon \|_n^2} \varepsilon_n^{s-2}, \]

where \( m \geq 0, s \geq 1, 0 \leq t \leq 2 \) and some constant \( K_{m,s,t} > 0. \)

**Lemma S.19** For a vector \( r \) and deterministic matrices \( D = (d_{ij}) \) and \( G, \)

\[ E[(z_i^T D z_1 - \text{tr} D) z_i^T G] = E z_{i1}^2 \sum_{i=1}^{p} d_{ii} e_i^T G e_i, \]

where \( e_i \) is the canonical vector with the \( i \)th entry 1.

**Lemma S.20** For deterministic matrices \( D = (d_{ij}) \) and \( G = (g_{ij}), \)

\[ E[(z_i^T D z_1 - \text{tr} D)(z_i^T G z_1 - \text{tr} G)] = |E z_{i1}^2|^{2} \text{tr} D G^T + \text{tr} D G + (E z_{i1}^4 - |E z_{i1}^2|^{2} - 2) \sum_{i=1}^{p} d_{ii} g_{ii}. \]
S.11. Additional simulation studies

Figure S.11.1 – Figure S.11.12 display additional size-adjusted power curves.

Figure S.11.1: Size-adjusted power with $\Sigma = \Sigma_{\text{den.}}, \ k = 5$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. LH$_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); LH$_{\text{ridge}}$ (black, dashed) and LH$_{\text{high}}$ (blue, dotted-dashed) with $\tilde{t} = (1,0,0)$.

Figure S.11.2: Size-adjusted power with $\Sigma = \Sigma_{\text{den.}}, \ k = 3$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. BNP$_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); BNP$_{\text{ridge}}$ (black, dashed) and BNP$_{\text{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0,1,0)$. 
Figure S.11.3: Size-adjusted power with $\Sigma = \Sigma_{\text{den}}, k = 5$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\text{LR}_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\text{LR}_{\text{ridge}}$ (black, dashed) and $\text{LR}_{\text{high}}$ (blue, dotted-dashed) with $\tilde{\ell} = (0, 0, 1)$.

Figure S.11.4: Size-adjusted power with $\Sigma = \Sigma_{\text{toep}}, k = 3$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. $\text{LH}_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); $\text{LH}_{\text{ridge}}$ (black, dashed) and $\text{LH}_{\text{high}}$ (blue, dotted-dashed) with $\tilde{\ell} = (1, 0, 0)$. 
Figure S.11.5: Size-adjusted power with $\Sigma = \Sigma_{toep}$, $k = 5$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. BNP$_{comp}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); BNP$_{ridge}$ (black, dashed) and BNP$_{high}$ (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$.

Figure S.11.6: Size-adjusted power with $\Sigma = \Sigma_{toep}$, $k = 3$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. LR$_{comp}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); LR$_{ridge}$ (black, dashed) and LR$_{high}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$. 

High-dimensional general linear hypothesis via spectral shrinkage
Figure S.11.7: Size-adjusted power with $\Sigma = \Sigma_{\text{dis}}$, $k = 5$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. LH$_\text{comp}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); LH$_\text{ridge}$ (black, dashed) and LH$_\text{high}$ (blue, dotted-dashed) with $\tilde{t} = (1, 0, 0)$.

Figure S.11.8: Size-adjusted power with $\Sigma = \Sigma_{\text{dis}}$, $k = 3$. Rows (top to bottom): $B =$ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. BNP$_\text{comp}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); BNP$_\text{ridge}$ (black, dashed) and BNP$_\text{high}$ (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$. 
Figure S.11.9: Size-adjusted power with \( \Sigma = \Sigma_{\text{dis}}, \ k = 5 \). Rows (top to bottom): \( B = \) Dense and Sparse; Columns (left to right): \( p = 150, 600, 3000 \). \( \text{LR}_{\text{comp}} \) (red, solid); \( \text{ZGZ} \) (green, solid); oracle CX (purple, solid); \( \text{LR}_{\text{ridge}} \) (black, dashed) and \( \text{LR}_{\text{high}} \) (blue, dotted-dashed) with \( \tilde{t} = (0, 0, 1) \).

Figure S.11.10: Size-adjusted power with \( \Sigma = I_p, \ k = 3 \). Rows (top to bottom): \( B = \) Dense and Sparse; Columns (left to right): \( p = 150, 600, 3000 \). \( \text{LH}_{\text{comp}} \) (red, solid); \( \text{ZGZ} \) (green, solid); oracle CX (purple, solid); \( \text{LH}_{\text{ridge}} \) (black, dashed) and \( \text{LH}_{\text{high}} \) (blue, dotted-dashed) with \( \tilde{t} = (1, 0, 0) \).
Figure S.11.11: 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$_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); BNP$_{\text{ridge}}$ (black, dashed) and BNP$_{\text{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 1, 0)$. 

Figure S.11.12: Size-adjusted power with $\Sigma = I_p$, $k = 3$. Rows (top to bottom): $B = $ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. LR$_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); LR$_{\text{ridge}}$ (black, dashed) and LR$_{\text{high}}$ (blue, dotted-dashed) with $\tilde{t} = (0, 0, 1)$. 
S.12. Nominal power curves

Figure S.12.13–Figure S.12.15 are counterparts of Figure 1–Figure 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.12.13: Empirical power with $\Sigma = \Sigma_{\text{den}}, k = 5$. Rows (top to bottom): $B = \text{Dense and Sparse}$; Columns (left to right): $p = 150, 600, 3000$. BNP$_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); BNP$_{\text{ridge}}$ (black, dashed) and BNP$_{\text{high}}$ (blue, dotted-dashed) with $\bar{t} = (1, 0, 0)$.

Figure S.12.14: Empirical power with $\Sigma = \Sigma_{\text{den}}, k = 5$. Rows (top to bottom): $B = \text{Dense and Sparse}$; Columns (left to right): $p = 150, 600, 3000$. LH$_{\text{comp}}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); LH$_{\text{ridge}}$ (black, dashed) and LH$_{\text{high}}$ (blue, dotted-dashed) with $\bar{t} = (0, 0, 1)$. 
Figure S.12.15: Empirical power with $\Sigma = \Sigma_{toep}, k = 3$. Rows (top to bottom): $B = $ Dense and Sparse; Columns (left to right): $p = 150, 600, 3000$. LR$_{comp}$ (red, solid); ZGZ (green, solid); oracle CX (purple, solid); LR$_{ridge}$ (black, dashed) and LR$_{high}$ (blue, dotted-dashed) with $\ell = (0, 1, 0)$. 