PHASE TRANSITION OF THE IMAGE COVARIANCE MATRIX AND INFERENCE FOR *2D*-PRINCIPAL COMPONENTS

Shuhang Liu, Li Yang and Dandan Jiang*

School of Mathematics and Statistics

Xi'an Jiaotong University

Supplementary Material

Section S1 presents Central Limit Theorems (CLTs) for linear spectral statistics (LSS) involving matrix pairs under non-i.i.d. cases (Lemmas 1–2). Section S1.2 derives the cross-covariance formula (S1.4), and Section S1.3 establishes the simplified covariance structure under the shared-component dominance condition. Section S2 provides detailed calculations for Scenarios 1–3 of Example 1. For $H(x) = \frac{1}{3}\delta_{\{3\sigma^2\}} + \frac{1}{3}\delta_{\{2\sigma^2\}} + \frac{1}{3}\delta_{\{\sigma^2\}}$ and $f_0(x) = x$, Section S3 derives the limiting mean and variance of the empirical spectral process. Section S4 of the main manuscript contains the proofs of Lemma 2 and Theorems 2–3. Section S5 provides Algorithms 1–2 for estimating the shrinkage coefficient γ_r and scaling parameter ς_r^2 (Corollary 1), with the proof of Theorem 3 deferred to Section S6. Section S7 reports additional numerical results: Scenarios 1–3 for Models 1–2 (Section S7.1) and Model 3 (Section S7.2). Section S8 details supplementary facial database experiments: empirical support for multi-population heterogeneity (Section S8.1) and further FERET dataset analysis (Section S8.2).

S1 CLT for Linear Combinations of LSS under Non-i.i.d. Cases

S1.1 Auxiliary Lemmas

For Case 1, consider sample covariance matrices $\mathbf{S}_j^{(r)}$ and $\mathbf{S}_i^{(u)}$ for any $r \neq u \in [1:\tau]$, $j \in [1:h_r]$, and $i \in [1:h_u]$. Utilizing the contour integral representations in Jiang (2023), Lemma 1 establishes the Gaussian convergence for linear combinations of their empirical processes $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j$ and $\{\boldsymbol{\mathcal{Q}}_n^{(u)}(g)\}_i$. We first outline the required RMT regularity assumptions.

Assumption 1. For each $r \in [1:\tau]$ and $j \in [1:h_r]$, the entries $\{x_{jkt}^{(r)} : k \in [1:p], t \in [1:n]\}$ of $\mathbf{X}_j^{(r)}$ are i.i.d. with $\mathbb{E}(x_{j11}^{(r)}) = 0$, $\mathbb{E}|x_{j11}^{(r)}|^2 = 1$, and $\sup_n \mathbb{E}|x_{j11}^{(r)}|^4 < \infty$. For complex-valued variables, assume $\mathbb{E}\{(x_{j11}^{(r)})^2\} = 0$.

Assumption 2. For each $r \in [1:\tau]$, all $\ell, m \in [1:h_r]$, and all $t \in [1:n]$, assume $\mathbb{E}\boldsymbol{x}_{0t}^{(r)} = \mathbb{E}\boldsymbol{\varepsilon}_{\ell t}^{(r)} = \mathbf{0}$, $\mathbb{E}\boldsymbol{x}_{0t}^{(r)}(\boldsymbol{x}_{0t}^{(r)})^\top = \rho_r \mathbf{I}_p$, $\mathbb{E}\boldsymbol{\varepsilon}_{\ell t}^{(r)}(\boldsymbol{\varepsilon}_{mt}^{(r)})^\top = (1-\rho_r)\mathbf{I}_p \mathbf{1}_{\{\ell=m\}}$, $\mathbb{E}\boldsymbol{x}_{0t}^{(r)}(\boldsymbol{\varepsilon}_{\ell t}^{(r)})^\top = \mathbf{0}$, where $\rho_r \in [0,1]$, and $\{(x_{0kt}^{(r)}, \varepsilon_{1kt}^{(r)}, \ldots, \varepsilon_{h_r kt}^{(r)}) : k \in [1:p],\ t \in [1:n]\}$ are independent across the pairs (k,t).

Assumption 3. For each $r \in [1:\tau]$, $\mathbf{T}_r \coloneqq (\boldsymbol{\Sigma}^{(r)})^{1/2}$ is deterministic with $\sup_p \|\mathbf{T}_r\| < \infty$, and the ESD $H_n^{(r)}$ of $\boldsymbol{\Sigma}^{(r)}$ converges weakly to a probability distribution $H^{(r)}$.

Lemma 1. *Suppose Assumptions 1–3 hold and $p/n = c_n \to c \in (0, \infty)$. Consider Case 1 for the analytic functions $f, g \in \mathcal{A}$. Then the random variables $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j$ and $\{\boldsymbol{\mathcal{Q}}_n^{(u)}(g)\}_i$ both converge weakly to the Gaussian distribution with means $\mu_f^{(r)}, \mu_g^{(u)}$ and covariance functions $\nu_f^{(r)}, \nu_g^{(u)}$, respectively. By the additivity of Gaussian variables,*

$$\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j + \{\boldsymbol{\mathcal{Q}}_n^{(u)}(g)\}_i \Rightarrow \mathcal{N}(\mu_{f,g}^{(r,u)}, \nu_{f,g}^{(r,u)}),$$

where the symbol $\Rightarrow$ denotes convergence in distribution,

$$\mu_{f,g}^{(r,u)} := \mu_f^{(r)} + \mu_g^{(u)}, \quad \nu_{f,g}^{(r,u)} := \nu_f^{(r)} + \nu_g^{(u)},$$

and the mean and covariance functions can be calculated by following:

$$\begin{aligned}\mu_f^{(r)} =& -\frac{q}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{c\int \underline{s}^r(z)^3 x^2\{1+x\underline{s}^r(z)\}^{-3}\mathrm{d}H^{(r)}(x)}{[1-c\int \underline{s}^r(z)^2x^2\{1+x\underline{s}^r(z)\}^{-2}\mathrm{d}H^{(r)}(x)]^2}(\mathrm{d}z) \quad \text{(S1.1)}\\ &-\frac{\beta^{(r)}c}{2\pi i}\oint_{\mathcal{C}} f(z)\frac{\underline{s}^r(z)^3\int x\{1+x\underline{s}^r(z)\}^{-1}\mathrm{d}H^{(r)}(x)\int\{1+x\underline{s}^r(z)\}^{-2}\mathrm{d}H^{(r)}(x)}{1-c\int \underline{s}^r(z)^2x^2\{1+x\underline{s}^r(z)\}^{-2}dH^{(r)}(x)}(\mathrm{d}z),\end{aligned}$$

$$\begin{aligned}\nu_f^{(r)} =& -\frac{q+1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}\frac{f(z_1)f(z_2)}{\{\underline{s}^r(z_1)-\underline{s}^r(z_2)\}^2}\mathrm{d}\underline{s}^r(z_1)\mathrm{d}\underline{s}^r(z_2) \quad \text{(S1.2)}\\ &-\frac{\beta^{(r)}c}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\int\frac{x\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^r(z_1)\}^2}\frac{x\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^r(z_2)\}^2}\mathrm{d}\underline{s}^r(z_1)\mathrm{d}\underline{s}^r(z_2)\end{aligned}$$

and $\mu_f^{(u)}, \nu_f^{(u)}$ are the analogues of $\mu_f^{(r)}, \nu_f^{(r)}$. Here $\beta^{(r)} = \mathbb{E}|x_{jki}^{(r)}|^4 - q - 2$, where $q = 1$ for the real case and $q = 0$ for the complex case. Let $s^r(z)$ be the stieltjes transform of $F^{c,H^{(r)}}$, and $\underline{s}^r \equiv s_{\underline{F}^{c,H^{(r)}}}$ be the stieltjes transform of $\underline{F}^{c,H^{(r)}} \equiv (1-c)I_{[0,\infty)} + cF^{c,H^{(r)}}$.

We now focus on the theoretical results pertaining to Case 2. For the rth row-population $\mathbf{\Sigma}^{(r)}$, consider two sample covariance matrices $\mathbf{S}_j^{(r)}$ and $\mathbf{S}_i^{(r)}$ for any $i \neq j \in [1 : h_r]$. To facilitate the covariance calculations between $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j$ and $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_i$ under this within-population dependence, we impose Assumption 4, analogous to condition (5) in Bai et al. (2007), and establish Lemma 2.

Assumption 4. For each $r \in [1 : \tau]$, assume that $\max_{1\leq k\leq p} \sup_z \left|\mathbf{e}_k^\top \{\mathbf{I}_p + \underline{s}^{(r)}(z)\mathbf{\Sigma}^{(r)}\}^{-1}\mathbf{e}_k - \int \{1 + \underline{s}^{(r)}(z)x\}^{-1} \,\mathrm{d}H_n^{(r)}(x)\right| \to 0$, where $\mathbf{e}_k$ denotes the kth standard basis vector for $k \in [1 : p]$.

Lemma 2. *Suppose Assumptions 1–4 hold and $p/n = c_n \to c \in (0, \infty)$. Consider Case 2 and, based on the above Lemma 1, $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j$ and $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_i$ both converge weakly to the same Gaussian distribution with mean $\mu_f^{(r)}$ and covariance function $\nu_f^{(r)}$. By the additivity of Gaussian variables,*

$$\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j + \{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_i \Rightarrow \mathcal{N}(\mu_{f_{ji}}^{(r)}, \nu_{f_{ji}}^{(r)}),$$

where

$$\mu_{f_{ji}}^{(r)} := 2\mu_f^{(r)}, \quad \nu_{f_{ji}}^{(r)} := 2\nu_f^{(r)} + 2\nu_{\mathrm{Cov}\, f_{ji}}^{(r)}.$$

Here, $\mu_f^{(r)}$ and $\nu_f^{(r)}$ are the same as those defined in (S1.1) *and* (S1.2). *Moreover, the calculations of the covariance $\nu_{\mathrm{Cov}\, f_{ji}}^{(r)}$ fluctuate based on the* *specific correlation structure considered, where*

$$\begin{aligned}
\nu^{(r)}_{\mathrm{Cov}\, f_{ji}} &= \mathrm{Cov}\left(\{\boldsymbol{\mathcal{Q}}^{(r)}_n(f)\}_j, \{\boldsymbol{\mathcal{Q}}^{(r)}_n(f)\}_i\right) \qquad \text{(S1.3)}\\
&= -\frac{1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\partial_{z_1}\partial_{z_2}\,\mathrm{Cov}\left(\Xi^{(r)}_{n,j}(z_1), \Xi^{(r)}_{n,i}(z_2)\right)\mathrm{d}z_1\mathrm{d}z_2,
\end{aligned}$$

where $\Xi^{(r)}_{n,j}(z_1) = \mathrm{tr}(\mathbf{S}^{(r)}_j - z_1\mathbf{I})^{-1} - ps^r(z_1)$*, and* $\Xi^{(r)}_{n,i}(z_2)$ *is defined analogously. By leveraging the correlation exhibited between sample matrices* $\mathbf{X}^{(r)}_i$ *and* $\mathbf{X}^{(r)}_j$ *from the identical population, the covariance can be calculated by*

$$\nu^{(r)}_{\mathrm{Cov}\, f_{ji}} = \nu^{(r)}_{f,\rho} + o(1), \qquad \text{(S1.4)}$$

where the cross-covariance in Case 2 is

$$\begin{aligned}
\nu^{(r)}_{f,\rho} = -\frac{q+1}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\frac{\partial^2}{\partial z_1\partial z_2}\Big[&-\log\Big\{1 - \rho_r^2 c\,\underline{s}^{(r)}(z_1)\underline{s}^{(r)}(z_2)\\
&\times \int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}\{1+x\underline{s}^{(r)}(z_2)\}}\Big\}\Big]\,\mathrm{d}z_1\mathrm{d}z_2 \qquad \text{(S1.5)}\\
-\frac{\beta^{(r)}_\rho c}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\int\frac{x\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}^2}&\int\frac{x\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_2)\}^2}\,\mathrm{d}\underline{s}^{(r)}(z_1)\,\mathrm{d}\underline{s}^{(r)}(z_2).
\end{aligned}$$

Here $\beta^{(r)}_\rho = \mathbb{E}\{(x^{(r)}_{jkt})^2(x^{(r)}_{ikt})^2\} - 1 - (q+1)\rho_r^2$ *for* $t \in [1:n]$ *and* $k \in [1:p]$*, where* $q = 1$ *for the real case and* $q = 0$ *for the complex case. Moreover, the detailed proof of* (S1.4) *is provided in Section S1.2. The explicit expression of the limiting kernel obtained from the first logarithmic term in* $\nu^{(r)}_{f,\rho}$ *is given in* (S1.66)*.*

Remark 1. Even without Assumption 4, explicit formulas can be derived for special cases. Under diagonal loading, the first terms in (S1.2) and (S1.5) remain unchanged, while for finite $\beta^{(r)}$ and $\beta_\rho^{(r)}$, the contour integrals in the second terms are replaced by

$$\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2} f(z_1)f(z_2)\int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}^2\{1+x\underline{s}^{(r)}(z_2)\}^2}\,\mathrm{d}\underline{s}^{(r)}(z_1)\,\mathrm{d}\underline{s}^{(r)}(z_2).$$

Remark 2. When reducing to the special case of shared-component dominance, the cross-covariance term admits a simplified asymptotic representation: $\nu_{\mathrm{Cov}\,fji}^{(r)} = \nu_f^{(r)} + o(1)$. A derivation of this result is provided in Section S1.3.

S1.2 Proof of the Cross-covariance Formula (S1.4)

Fix the rth row-population $\mathbf{\Sigma}^{(r)} = \mathbf{T}_r\mathbf{T}_r^\top$ with $\mathbf{T}_r := (\mathbf{\Sigma}^{(r)})^{1/2}$, and take two distinct row indices $i \neq j \in [1 : h_r]$. Let

$$\mathbf{X}_j^{(r)} = (\boldsymbol{x}_{j1}^{(r)}, \ldots, \boldsymbol{x}_{jn}^{(r)}), \qquad \mathbf{X}_i^{(r)} = (\boldsymbol{x}_{i1}^{(r)}, \ldots, \boldsymbol{x}_{in}^{(r)}),$$

where $\boldsymbol{x}_{jt}^{(r)}, \boldsymbol{x}_{it}^{(r)} \in \mathbb{R}^p$ for all $t \in [1 : n]$. In Case 2, the two row-sample matrices share the same marginal distribution but may be dependent within the same row-population. Specifically, by Assumption 2, we have

$$\mathbb{E}\{\boldsymbol{x}_{jt}^{(r)}(\boldsymbol{x}_{jt}^{(r)})^*\} = \mathbb{E}\{\boldsymbol{x}_{it}^{(r)}(\boldsymbol{x}_{it}^{(r)})^*\} = \mathbf{I}_p, \qquad \mathbb{E}\{\boldsymbol{x}_{jt}^{(r)}(\boldsymbol{x}_{it}^{(r)})^*\} = \rho_r\mathbf{I}_p.$$

where $*$ denotes transpose in the real case and conjugate transpose in the complex case. Thus the within-population dependence enters the proof through the cross-covariance coefficient $\rho_r \in [0, 1]$. Let $q = 1$ in the real case and $q = 0$ in the complex case. For the complex case, we further assume $\mathbb{E}\{(x_{jkt}^{(r)})^2\} = 0$, $\mathbb{E}\{x_{jkt}^{(r)} x_{ikt}^{(r)}\} = 0$. The diagonal cross fourth-cumulant is defined by

$$\beta_\rho^{(r)} = \begin{cases} \mathbb{E}\{(x_{jkt}^{(r)})^2 (x_{ikt}^{(r)})^2\} - 1 - 2\rho_r^2, & \text{in the real case,} \\ \mathbb{E}\{|x_{jkt}^{(r)}|^2 |x_{ikt}^{(r)}|^2\} - 1 - \rho_r^2, & \text{in the complex case.} \end{cases}$$

Here $\beta_\rho^{(r)}$ is assumed to be common across all $k \in [1:p]$ and $t \in [1:n]$ within the rth row-population.

This dependence enters the proof through the mixed covariance of the corresponding quadratic forms. The proof proceeds with the uncentered covariance representations

$$\mathbf{S}_j^{(r)} = \frac{1}{n} \mathbf{T}_r \mathbf{X}_j^{(r)} (\mathbf{X}_j^{(r)})^\top \mathbf{T}_r^\top, \quad \mathbf{S}_i^{(r)} = \frac{1}{n} \mathbf{T}_r \mathbf{X}_i^{(r)} (\mathbf{X}_i^{(r)})^\top \mathbf{T}_r^\top. \tag{S1.6}$$

For $z = u + \mathrm{i}v \in \mathbb{C}^+$ with $v > 0$, denote

$$\mathbf{D}_j^{(r)}(z) = \mathbf{S}_j^{(r)} - z\mathbf{I}_p, \quad \mathbf{D}_i^{(r)}(z) = \mathbf{S}_i^{(r)} - z\mathbf{I}_p. \tag{S1.7}$$

Since $\mathbf{S}_j^{(r)}$ and $\mathbf{S}_i^{(r)}$ are real symmetric, the spectral theorem gives

$$\left\| \{\mathbf{D}_j^{(r)}(z)\}^{-1} \right\| = \max_{1 \le k \le p} \frac{1}{|\lambda_k(\mathbf{S}_j^{(r)}) - z|} \le \frac{1}{|\Im z|}.$$

The same bound holds for $\{\mathbf{D}_i^{(r)}(z)\}^{-1}$. And the corresponding resolvent traces are

$$\Xi_{n,j}^{(r)}(z) = \mathrm{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1} - ps^{(r)}(z),\ \Xi_{n,i}^{(r)}(z) = \mathrm{tr}\{\mathbf{D}_i^{(r)}(z)\}^{-1} - ps^{(r)}(z). \tag{S1.8}$$

The proof follows the martingale decomposition of resolvent traces used by Bai and Silverstein (2010) for linear spectral statistics. It is organized into three steps: Step 1 establishes the mixed quadratic-form covariance identity, Step 2 reduces the resolvent-trace covariance to martingale differences, and Step 3 identifies the limiting covariance through deterministic equivalents and Cauchy's integral representation.

Step 1. Mixed quadratic-form covariance.

We derive the mixed quadratic-form identity under non-i.i.d. cases, serving as the cross-sample analogue of the quadratic-form covariance identity in Bai and Silverstein (2010, Equation (9.8.6)). Fix distinct indices $i, j \in [1 : h_r]$. For the real case, let $\mathbf{A} = (A_{ab})$ and $\mathbf{B} = (B_{cd})$ be deterministic matrices with uniformly bounded spectral norms. Substituting the sum expansions of the quadratic forms and traces directly into the covariance yields

$$\begin{aligned}
&\mathbb{E}\big[\{(\boldsymbol{x}_{jt}^{(r)})^\top \mathbf{A}\boldsymbol{x}_{jt}^{(r)} - \operatorname{tr}\mathbf{A}\}\{(\boldsymbol{x}_{it}^{(r)})^\top \mathbf{B}\boldsymbol{x}_{it}^{(r)} - \operatorname{tr}\mathbf{B}\}\big] \\
&\quad = \mathbb{E}\Big[\Big\{\sum_{a,b=1}^{p} A_{ab} x_{jat}^{(r)} x_{jbt}^{(r)} - \sum_{a=1}^{p} A_{aa}\Big\}\Big\{\sum_{c,d=1}^{p} B_{cd} x_{ict}^{(r)} x_{idt}^{(r)} - \sum_{c=1}^{p} B_{cc}\Big\}\Big] \\
&\quad = \sum_{a,b,c,d=1}^{p} A_{ab} B_{cd} \mathbb{E}\{x_{jat}^{(r)} x_{jbt}^{(r)} x_{ict}^{(r)} x_{idt}^{(r)}\} - \sum_{a,c=1}^{p} A_{aa} B_{cc}. \qquad \text{(S1.9)}
\end{aligned}$$

We then evaluate the mixed fourth moment based on the possible coincidences among a, b, c, d. By the independence of coordinate triples across k

and the zero-mean condition, non-vanishing terms require indices to appear in pairs or as a quadruplet:

(a) If $a = b$, $c = d$, and $a \neq c$, then $\mathbb{E}\{(x_{jat}^{(r)})^2\}\mathbb{E}\{(x_{ict}^{(r)})^2\} = 1$, contributing $\sum_{a\neq c} A_{aa}B_{cc}$.

(b) If $a = c$, $b = d$, and $a \neq b$, independence yields $\mathbb{E}\{x_{jat}^{(r)}x_{iat}^{(r)}\}\mathbb{E}\{x_{jbt}^{(r)}x_{ibt}^{(r)}\} = \rho_r^2$, contributing $\rho_r^2 \sum_{a\neq b} A_{ab}B_{ab}$.

(c) If $a = d$, $b = c$, and $a \neq b$, similarly, the moment evaluates to ρ_r^2, contributing $\rho_r^2 \sum_{a\neq b} A_{ab}B_{ba}$.

(d) For the fully diagonal case $a = b = c = d = k$, the moment is $1 + 2\rho_r^2 + \beta_\rho^{(r)}$, which contributes $\sum_{k=1}^p A_{kk}B_{kk}\{1 + 2\rho_r^2 + \beta_\rho^{(r)}\}$.

All other configurations vanish. Then, summing these non-vanishing contributions and subtracting the centering term $\sum_{a,c=1}^p A_{aa}B_{cc} = \sum_{a\neq c} A_{aa}B_{cc} + \sum_{k=1}^p A_{kk}B_{kk}$, the (S1.9) simplifies to

$$\rho_r^2 \operatorname{tr}(\mathbf{A}\mathbf{B}^\top) + \rho_r^2 \operatorname{tr}(\mathbf{A}\mathbf{B}) + \beta_\rho^{(r)} \sum_{k=1}^p A_{kk}B_{kk}.$$

In particular, if $\mathbf{A}$ and $\mathbf{B}$ are symmetric, we have $\operatorname{tr}(\mathbf{A}\mathbf{B}^\top) = \operatorname{tr}(\mathbf{A}\mathbf{B})$, which further reduces the identity to

$$2\rho_r^2 \operatorname{tr}(\mathbf{A}\mathbf{B}) + \beta_\rho^{(r)} \sum_{k=1}^p A_{kk}B_{kk}.$$

For the complex case, expanding the covariance leads to analogous coincidences among $\mathbb{E}\{\overline{x}_{jat}^{(r)}x_{jbt}^{(r)}\overline{x}_{ict}^{(r)}x_{idt}^{(r)}\}$:

(a) The configuration where $a = b$, $c = d$, and $a \neq c$ yields 1, which is exactly cancelled by the centering term.

(b) The case $a = d$, $b = c$, and $a \neq b$ yields $\mathbb{E}\{\overline{x}_{jat}^{(r)} x_{iat}^{(r)}\}\mathbb{E}\{x_{jbt}^{(r)} \overline{x}_{ibt}^{(r)}\} = \rho_r^2$, contributing $\rho_r^2 \operatorname{tr}(\mathbf{AB})$.

(c) The case $a = c$, $b = d$, and $a \neq b$ involves pseudo-covariances (e.g., $\mathbb{E}\{\overline{x}_{jat}^{(r)} \overline{x}_{iat}^{(r)}\}$) that vanish by the complex condition.

(d) The fully diagonal case $a = b = c = d = k$ yields $1 + \rho_r^2 + \beta_\rho^{(r)}$, where the 1 is again cancelled by the centering term.

Thus, the complex covariance is $\rho_r^2 \operatorname{tr}(\mathbf{AB}) + \beta_\rho^{(r)} \sum_{k=1}^p A_{kk} B_{kk}$.

Consequently, in the present application where the real case involves deterministic symmetric matrices and the complex case satisfies the vanishing pseudo-covariance condition, the above two identities unify as

$$\begin{aligned}
&\mathbb{E}\Big[\big\{(\boldsymbol{x}_{jt}^{(r)})^* \mathbf{A} \boldsymbol{x}_{jt}^{(r)} - \operatorname{tr} \mathbf{A}\big\}\big\{(\boldsymbol{x}_{it}^{(r)})^* \mathbf{B} \boldsymbol{x}_{it}^{(r)} - \operatorname{tr} \mathbf{B}\big\}\Big] \\
&\qquad = (q+1)\rho_r^2 \operatorname{tr}(\mathbf{AB}) + \beta_\rho^{(r)} \sum_{k=1}^{p} A_{kk} B_{kk}.
\end{aligned} \tag{S1.10}$$

where $q = 1$ in the real case and $q = 0$ in the complex case.

Step 2. Martingale decomposition of resolvent traces.

Following the notation in the proof of Bai and Silverstein (2010, Theorem 9.10), define

$$\boldsymbol{r}_t^{(r,j)} = n^{-1/2} \mathbf{T}_r \boldsymbol{x}_{jt}^{(r)}, \quad \boldsymbol{r}_t^{(r,i)} = n^{-1/2} \mathbf{T}_r \boldsymbol{x}_{it}^{(r)}. \tag{S1.11}$$

Then

$$\mathbf{S}_j^{(r)} = \sum_{t=1}^{n} \boldsymbol{r}_t^{(r,j)} (\boldsymbol{r}_t^{(r,j)})^\top, \quad \mathbf{S}_i^{(r)} = \sum_{t=1}^{n} \boldsymbol{r}_t^{(r,i)} (\boldsymbol{r}_t^{(r,i)})^\top.$$

Let

$$\mathbf{D}_{j,t}^{(r)}(z) = \mathbf{D}_j^{(r)}(z) - \boldsymbol{r}_t^{(r,j)} (\boldsymbol{r}_t^{(r,j)})^\top, \quad \mathbf{D}_{i,t}^{(r)}(z) = \mathbf{D}_i^{(r)}(z) - \boldsymbol{r}_t^{(r,i)} (\boldsymbol{r}_t^{(r,i)})^\top. \tag{S1.12}$$

Throughout the proof, K denotes a positive constant independent of n, t, z, whose value may vary from line to line. Since the contours are bounded away from the real axis, there exists $v_0 > 0$ such that $|\Im z| \geq v_0$ for all $z \in \mathcal{C}$. Hence, by the spectral theorem,

$$\big\|\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\big\| + \big\|\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\big\| \leq K$$

uniformly in $t$and $z \in \mathcal{C}$. By the Sherman–Morrison rank-one perturbation formula, we obtain

$$\{\mathbf{D}_j^{(r)}(z)\}^{-1} = \{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} - \vartheta_t^{(r,j)}(z)\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\boldsymbol{r}_t^{(r,j)}(\boldsymbol{r}_t^{(r,j)})^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1},$$

where

$$\vartheta_t^{(r,j)}(z) = \big[1 + (\boldsymbol{r}_t^{(r,j)})^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\boldsymbol{r}_t^{(r,j)}\big]^{-1}.$$

Taking traces and differentiating with respect to z, we obtain

$$\operatorname{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1} - \operatorname{tr}\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} = \frac{\partial}{\partial z}\log \vartheta_t^{(r,j)}(z). \tag{S1.13}$$

The same identity holds with j replaced by i. Let

$$\mathcal{F}_t = \sigma\big\{(\boldsymbol{x}_{j\ell}^{(r)}, \boldsymbol{x}_{i\ell}^{(r)}) : 1 \le \ell \le t\big\}, \quad \mathbb{E}_t(\cdot) = \mathbb{E}(\cdot \mid \mathcal{F}_t). \tag{S1.14}$$

Since the column pairs are independent over t, and $\mathbf{D}_{j,t}^{(r)}(z)$ does not contain the tth column-pair, we have

$$(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} = 0.$$

Therefore, by (S1.13),

$$(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_{j}^{(r)}(z)\}^{-1} = \frac{\partial}{\partial z}(\mathbb{E}_t - \mathbb{E}_{t-1})\log\vartheta_t^{(r,j)}(z). \tag{S1.15}$$

Set

$$M_{n,j}^{(r),1}(z) = \operatorname{tr}\{\mathbf{D}_{j}^{(r)}(z)\}^{-1} - \mathbb{E}\operatorname{tr}\{\mathbf{D}_{j}^{(r)}(z)\}^{-1}.$$

The martingale decomposition gives

$$M_{n,j}^{(r),1}(z) = \sum_{t=1}^{n}(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_{j}^{(r)}(z)\}^{-1}.$$

Consequently, recall (S1.8). Since $ps^{(r)}(z)$ is deterministic, we have

$$\begin{aligned}
&\operatorname{Cov}\big(\Xi_{n,j}^{(r)}(z_1), \Xi_{n,i}^{(r)}(z_2)\big) = \operatorname{Cov}\big(\operatorname{tr}\{\mathbf{D}_{j}^{(r)}(z_1)\}^{-1}, \operatorname{tr}\{\mathbf{D}_{i}^{(r)}(z_2)\}^{-1}\big)\\
&= \mathbb{E}\Big[\Big\{\sum_{t=1}^{n}(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_{j}^{(r)}(z_1)\}^{-1}\Big\}\Big\{\sum_{s=1}^{n}(\mathbb{E}_s - \mathbb{E}_{s-1})\operatorname{tr}\{\mathbf{D}_{i}^{(r)}(z_2)\}^{-1}\Big\}\Big]\\
&= \sum_{t=1}^{n}\sum_{s=1}^{n}\mathbb{E}\Big[(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_{j}^{(r)}(z_1)\}^{-1}\,(\mathbb{E}_s - \mathbb{E}_{s-1})\operatorname{tr}\{\mathbf{D}_{i}^{(r)}(z_2)\}^{-1}\Big]\\
&= \sum_{t=1}^{n}\mathbb{E}\Big[(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_{j}^{(r)}(z_1)\}^{-1}\,(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_{i}^{(r)}(z_2)\}^{-1}\Big].
\end{aligned} \tag{S1.16}$$

Define

$$\tilde{b}_t^{(r,j)}(z) = \Big[1 + \frac{1}{n}\,\mathrm{tr}\,\big\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big\}\Big]^{-1} \tag{S1.17}$$

and

$$b_t^{(r,j)}(z) = \Big[1 + \frac{1}{n}\mathbb{E}\,\mathrm{tr}\,\big\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big\}\Big]^{-1}. \tag{S1.18}$$

Moreover, since $\mathbf{S}_{j,t}^{(r)}$ and $\boldsymbol{\Sigma}^{(r)}$ are nonnegative definite, for $z = u + \mathrm{i}v \in \mathbb{C}^+$, each of $\vartheta_t^{(r,j)}(z)$, $\tilde{b}_t^{(r,j)}(z)$, and $b_t^{(r,j)}(z)$ is bounded in absolute value by$|z|/v$. The same bounds hold with j replaced by i.

Since the two row samples have the same marginal distribution, recall Bai and Silverstein (2010, Equation (9.9.20)),

$$b_t^{(r,j)}(z) = b_t^{(r,i)}(z) = -z\underline{s}^{(r)}(z) + o(1)$$

uniformly for z on the contours. Define

$$\begin{aligned}\eta_t^{(r,j)}(z) &= (\boldsymbol{r}_t^{(r,j)})^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\boldsymbol{r}_t^{(r,j)} - \frac{1}{n}\,\mathrm{tr}\,\big\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big\} \\ &= \frac{1}{n}\Big[(\boldsymbol{x}_{jt}^{(r)})^\top\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{jt}^{(r)} - \mathrm{tr}\,\big\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big\}\Big].\end{aligned} \tag{S1.19}$$

Define $\eta_t^{(r,i)}(z)$ similarly. By the quadratic-form moment estimates used in the proof of Bai and Silverstein (2010, Theorem 9.10), uniformly for z on the contours,

$$\mathbb{E}|\eta_t^{(r,j)}(z)|^2 \le Kn^{-1}, \qquad \mathbb{E}|\eta_t^{(r,j)}(z)|^4 \le Kn^{-2}. \tag{S1.20}$$

Moreover, since

$$\frac{\partial}{\partial z}\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} = \{\mathbf{D}_{j,t}^{(r)}(z)\}^{-2},$$

the same argument gives

$$\mathbb{E}\Big|\frac{\partial}{\partial z}\eta_t^{(r,j)}(z)\Big|^2 \le Kn^{-1}, \qquad \mathbb{E}\Big|\frac{\partial}{\partial z}\eta_t^{(r,j)}(z)\Big|^4 \le Kn^{-2}. \tag{S1.21}$$

The same bounds hold with j replaced by i. By the definitions of $\tilde{b}_t^{(r,j)}(z)$ in (S1.17) and $\eta_t^{(r,j)}(z)$ in (S1.19),

$$\begin{aligned}
\vartheta_t^{(r,j)}(z) &= \big[1 + (\boldsymbol{r}_t^{(r,j)})^\top \{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\boldsymbol{r}_t^{(r,j)}\big]^{-1} \\
&= \big[1 + \frac{1}{n}\operatorname{tr}\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r + \eta_t^{(r,j)}(z)\big]^{-1} \\
&= \Big[\big\{1 + \frac{1}{n}\operatorname{tr}\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big\}\big\{1 + \tilde{b}_t^{(r,j)}(z)\eta_t^{(r,j)}(z)\big\}\Big]^{-1} \\
&= \tilde{b}_t^{(r,j)}(z)\big\{1 + \tilde{b}_t^{(r,j)}(z)\eta_t^{(r,j)}(z)\big\}^{-1}.
\end{aligned} \tag{S1.22}$$

Hence

$$\log\vartheta_t^{(r,j)}(z) = \log\tilde{b}_t^{(r,j)}(z) - \tilde{b}_t^{(r,j)}(z)\eta_t^{(r,j)}(z) + O\big(|\tilde{b}_t^{(r,j)}(z)\eta_t^{(r,j)}(z)|^2\big).$$

Since $\tilde{b}_t^{(r,j)}(z)$ and its derivative are uniformly bounded, (S1.20) and (S1.21) imply

$$\begin{aligned}
&\sum_{t=1}^n \mathbb{E}\Big|\frac{\partial}{\partial z}(\mathbb{E}_t - \mathbb{E}_{t-1})O\big(|\tilde{b}_t^{(r,j)}(z)\eta_t^{(r,j)}(z)|^2\big)\Big|^2 \\
&\quad \le K\sum_{t=1}^n \Big[\mathbb{E}|\eta_t^{(r,j)}(z)|^4 + \mathbb{E}\Big\{|\eta_t^{(r,j)}(z)|^2\Big|\frac{\partial}{\partial z}\eta_t^{(r,j)}(z)\Big|^2\Big\}\Big] = o(1).
\end{aligned} \tag{S1.23}$$

Also,

$$(\mathbb{E}_t - \mathbb{E}_{t-1})\log\tilde{b}_t^{(r,j)}(z) = 0,$$

because $\tilde{b}_t^{(r,j)}(z)$ is a function of the matrix obtained by removing the tth rank-one term and does not depend on the current column-pair. Therefore,

$$(\mathbb{E}_t - \mathbb{E}_{t-1})\,\mathrm{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1} = -\frac{\partial}{\partial z}(\mathbb{E}_t - \mathbb{E}_{t-1})\big[\tilde{b}_t^{(r,j)}(z)\eta_t^{(r,j)}(z)\big] + \xi_{nt,1}^{(r,j)}(z),$$

where, by (S1.23),

$$\sum_{t=1}^{n} \mathbb{E}|\xi_{nt,1}^{(r,j)}(z)|^2 = o(1). \tag{S1.24}$$

Next we replace $\tilde{b}_t^{(r,j)}(z)$ by $b_t^{(r,j)}(z)$. The estimates in the proof of Bai and Silverstein (2010, Theorem 9.10), in particular the estimates corresponding to Bai and Silverstein (2010, Equations (9.9.11)–(9.9.16)), give

$$\mathbb{E}\bigg|\frac{1}{n}\,\mathrm{tr}\left\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\right\} - \frac{1}{n}\mathbb{E}\,\mathrm{tr}\left\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\right\}\bigg|^2 \le Kn^{-1}, \tag{S1.25}$$

uniformly for z on the contours. Differentiating the trace and using the same argument yields

$$\mathbb{E}\bigg|\frac{\partial}{\partial z}\Big[\frac{1}{n}\,\mathrm{tr}\left\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\right\} - \frac{1}{n}\mathbb{E}\,\mathrm{tr}\left\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\right\}\Big]\bigg|^2 \le Kn^{-1}. \tag{S1.26}$$

Since the denominators in the definitions of $\tilde{b}_t^{(r,j)}(z)$ and $b_t^{(r,j)}(z)$ are bounded away from zero on the contours, (S1.25) and (S1.26) imply

$$\mathbb{E}|\tilde{b}_t^{(r,j)}(z) - b_t^{(r,j)}(z)|^2 \le Kn^{-1}, \qquad \mathbb{E}\bigg|\frac{\partial}{\partial z}\{\tilde{b}_t^{(r,j)}(z) - b_t^{(r,j)}(z)\}\bigg|^2 \le Kn^{-1}. \tag{S1.27}$$

The fourth-moment versions of the same estimates give

$$\mathbb{E}|\tilde{b}_t^{(r,j)}(z) - b_t^{(r,j)}(z)|^4 \le Kn^{-2}, \qquad \mathbb{E}\Big|\frac{\partial}{\partial z}\{\tilde{b}_t^{(r,j)}(z) - b_t^{(r,j)}(z)\}\Big|^4 \le Kn^{-2}. \tag{S1.28}$$

Combining (S1.20), (S1.21), (S1.27), and (S1.28), we have

$$\begin{aligned}
&\sum_{t=1}^{n} \mathbb{E}\Big|\frac{\partial}{\partial z}(\mathbb{E}_t - \mathbb{E}_{t-1})\big[\{\tilde{b}_t^{(r,j)}(z) - b_t^{(r,j)}(z)\}\eta_t^{(r,j)}(z)\big]\Big|^2 \\
&\quad \le K\sum_{t=1}^{n}\bigg[\Big\{\mathbb{E}|\tilde{b}_t^{(r,j)}(z) - b_t^{(r,j)}(z)|^4\Big\}^{1/2}\Big\{\mathbb{E}\Big|\frac{\partial}{\partial z}\eta_t^{(r,j)}(z)\Big|^4\Big\}^{1/2} \\
&\qquad\qquad + \Big\{\mathbb{E}\Big|\frac{\partial}{\partial z}\{\tilde{b}_t^{(r,j)}(z) - b_t^{(r,j)}(z)\}\Big|^4\Big\}^{1/2}\Big\{\mathbb{E}|\eta_t^{(r,j)}(z)|^4\Big\}^{1/2}\bigg] = o(1).
\end{aligned} \tag{S1.29}$$

It follows from (S1.24) and (S1.29) that

$$(\mathbb{E}_t - \mathbb{E}_{t-1})\,\mathrm{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1} = -\frac{\partial}{\partial z}\big[b_t^{(r,j)}(z)(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z)\big] + \xi_{nt}^{(r,j)}(z), \tag{S1.30}$$

where $\sum_{t=1}^{n}\mathbb{E}|\xi_{nt}^{(r,j)}(z)|^2 = o(1)$ uniformly for z on the contours. The same expansion and bound hold with j replaced by i.

Substituting (S1.30) into (S1.16), the terms involving $\xi_{nt}^{(r,j)}$ or $\xi_{nt}^{(r,i)}$ are negligible. Indeed,

$$\sum_{t=1}^{n}\mathbb{E}|\xi_{nt}^{(r,j)}(z_1)\xi_{nt}^{(r,i)}(z_2)| \le \Big(\sum_{t=1}^{n}\mathbb{E}|\xi_{nt}^{(r,j)}(z_1)|^2\Big)^{1/2}\Big(\sum_{t=1}^{n}\mathbb{E}|\xi_{nt}^{(r,i)}(z_2)|^2\Big)^{1/2} = o(1),$$

and similarly, using

$$\sum_{t=1}^{n}\mathbb{E}\Big|\frac{\partial}{\partial z}\big[b_t^{(r,j)}(z)(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z)\big]\Big|^2 = O(1),$$

the mixed terms are also $o(1)$. Hence

$$\begin{aligned}
&\mathrm{Cov}\left(\Xi_{n,j}^{(r)}(z_1), \Xi_{n,i}^{(r)}(z_2)\right) \qquad \text{(S1.31)}\\
&= \frac{\partial^2}{\partial z_1 \partial z_2} \sum_{t=1}^{n} \mathbb{E}\big[b_t^{(r,j)}(z_1) b_t^{(r,i)}(z_2) \big\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\big\}\big\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\big] + o(1).
\end{aligned}$$

Step 3. Identification of the limiting covariance.

We next evaluate the covariance term in (S1.31). By the definition of $\eta_t^{(r,j)}(z)$ in (S1.19), taking conditional expectation with respect to $\mathcal{F}_t$, and using that $\boldsymbol{x}_{jt}^{(r)}$ is $\mathcal{F}_t$-measurable, gives

$$\mathbb{E}_t \eta_t^{(r,j)}(z_1) = \frac{1}{n}\Big\{(\boldsymbol{x}_{jt}^{(r)})^\top \mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r\big] \boldsymbol{x}_{jt}^{(r)} - \mathrm{tr}\, \mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r\big]\Big\}.$$

Moreover, conditional on $\mathcal{F}_{t-1}$, the current vector $\boldsymbol{x}_{jt}^{(r)}$ is independent of $\mathbf{D}_{j,t}^{(r)}(z_1)$ and has covariance $\mathbf{I}_p$. Hence

$$\begin{aligned}
\mathbb{E}_{t-1}\eta_t^{(r,j)}(z_1) &= \frac{1}{n}\mathbb{E}_{t-1}\big[(\boldsymbol{x}_{jt}^{(r)})^\top \mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r \boldsymbol{x}_{jt}^{(r)}\big] - \frac{1}{n}\mathbb{E}_{t-1}\, \mathrm{tr}\left[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r\right] \\
&= \frac{1}{n}\mathbb{E}_{t-1}\, \mathrm{tr}\left[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r\right] - \frac{1}{n}\mathbb{E}_{t-1}\, \mathrm{tr}\left[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r\right] = 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1) \qquad \text{(S1.32)}\\
&\quad = \frac{1}{n}\Big\{(\boldsymbol{x}_{jt}^{(r)})^\top \mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r\big] \boldsymbol{x}_{jt}^{(r)} - \mathrm{tr}\, \mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r\big]\Big\}.
\end{aligned}$$

Similarly,

$$(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2) \qquad \text{(S1.33)}$$

$$= \frac{1}{n}\Big\{(\boldsymbol{x}_{it}^{(r)})^\top \mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\big]\boldsymbol{x}_{it}^{(r)} - \operatorname{tr}\mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\big]\Big\}.$$

Combining (S1.32) and (S1.33), we have

$$\begin{aligned}
&\mathbb{E}\big[\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\}\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\}\big] \qquad \text{(S1.34)}\\
&= \frac{1}{n^2}\mathbb{E}\Bigg[\Big\{(\boldsymbol{x}_{jt}^{(r)})^\top \mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\big]\boldsymbol{x}_{jt}^{(r)} - \operatorname{tr}\mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\big]\Big\}\\
&\qquad \times \Big\{(\boldsymbol{x}_{it}^{(r)})^\top \mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\big]\boldsymbol{x}_{it}^{(r)} - \operatorname{tr}\mathbb{E}_t\big[\mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\big]\Big\}\Bigg].
\end{aligned}$$

Recall (S1.14). Since the column pairs are independent over ℓ, and since $\mathbf{D}_{j,t}^{(r)}(z)$ and $\mathbf{D}_{i,t}^{(r)}(z)$ are obtained by deleting the tth rank-one terms, they do not involve the current column-pair $(\boldsymbol{x}_{jt}^{(r)}, \boldsymbol{x}_{it}^{(r)})$. Hence, conditional on $\mathcal{F}_{t-1}$, the future column-pairs are independent of the current column-pair, and

$$\mathbb{E}_t \mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r = \mathbb{E}_{t-1} \mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r.$$

The same identity holds with j replaced by i. Therefore, the matrices

$$\mathbf{A}_t(z_1) = \mathbb{E}_t \mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r, \quad \mathbf{B}_t(z_2) = \mathbb{E}_t \mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r \quad \text{(S1.35)}$$

are $\mathcal{F}_{t-1}$-measurable. Thus, after conditioning on $\mathcal{F}_{t-1}$, they may be treated as deterministic matrices in applying the mixed quadratic-form covariance formula (S1.10). For non-real z_1, z_2, these matrices are complex-valued. The identity extends from real symmetric matrices to complex deterministic matrices by bilinearity, after separating real and imaginary parts. Moreover, by the resolvent bounds on the contours and the boundedness of $\|\mathbf{\Sigma}^{(r)}\|$, $\|\mathbf{A}_t(z_1)\| + \|\mathbf{B}_t(z_2)\| \le K$ uniformly in t, z_1, z_2.

Consequently, applying (S1.10) conditionally on $\mathcal{F}_{t-1}$ gives

$$\begin{aligned}
&\mathbb{E}\Big[\big\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\big\}\big\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\Big] \\
&\quad = \frac{(q+1)\rho_r^2}{n^2}\mathbb{E}\,\mathrm{tr}\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\,\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\right] \qquad \text{(S1.36)} \\
&\qquad + \frac{\beta_\rho^{(r)}}{n^2}\mathbb{E}\sum_{k=1}^p\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\right]_{kk}\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\right]_{kk},
\end{aligned}$$

uniformly in t and uniformly for z_1, z_2 on the contours. Combining (S1.31) and (S1.36), we obtain

$$\begin{aligned}
&\mathrm{Cov}\left(\Xi_{n,j}^{(r)}(z_1), \Xi_{n,i}^{(r)}(z_2)\right) \\
&= \frac{\partial^2}{\partial z_1 \partial z_2}\sum_{t=1}^n \mathbb{E}\Big[b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\big\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\big\}\big\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\Big] + o(1) \\
&= \frac{\partial^2}{\partial z_1 \partial z_2}\sum_{t=1}^n b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\mathbb{E}\Big[\big\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\big\}\big\{(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\Big] + o(1) \\
&= \frac{\partial^2}{\partial z_1 \partial z_2}\sum_{t=1}^n b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\Big[\frac{(q+1)\rho_r^2}{n^2}\mathbb{E}\,\mathrm{tr}\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\,\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\right] \\
&\qquad\qquad + \frac{\beta_\rho^{(r)}}{n^2}\mathbb{E}\sum_{k=1}^p\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\right]_{kk}\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\right]_{kk}\Big] + o(1) \\
&= (q+1)\rho_r^2\mathcal{K}_{1,n}^{(r)}(z_1,z_2) + \beta_\rho^{(r)}\mathcal{K}_{2,n}^{(r)}(z_1,z_2) + o(1), \qquad \text{(S1.37)}
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{K}_{1,n}^{(r)}(z_1,z_2) &= \frac{\partial^2}{\partial z_1 \partial z_2}\frac{1}{n^2}\sum_{t=1}^n b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2) \qquad \text{(S1.38)} \\
&\qquad \times \mathbb{E}\,\mathrm{tr}\Big[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\,\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\Big]
\end{aligned}$$

and

$$\mathcal{K}_{2,n}^{(r)}(z_1, z_2) = \frac{\partial^2}{\partial z_1 \partial z_2} \frac{1}{n^2} \sum_{t=1}^{n} b_t^{(r,j)}(z_1) b_t^{(r,i)}(z_2) \tag{S1.39}$$
$$\times \mathbb{E} \sum_{k=1}^{p} \left[\mathbb{E}_t \mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r \right]_{kk} \left[\mathbb{E}_t \mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1} \mathbf{T}_r \right]_{kk}.$$

Write $\underline{s}_\ell^{(r)} = \underline{s}^{(r)}(z_\ell)$, for $\ell = 1, 2$. Let $c_n = p/n$ and $H_n^{(r)} = 1/p \sum_{k=1}^p \delta_{d_{rk}}$, where $d_{r1}, \ldots, d_{rp}$ are the eigenvalues of $\mathbf{\Sigma}^{(r)}$. Define

$$a_n^{(r)}(z_1, z_2) = c_n \, \underline{s}_1^{(r)} \underline{s}_2^{(r)} \int \frac{x^2 \, \mathrm{d}H_n^{(r)}(x)}{\{1 + x\underline{s}_1^{(r)}\}\{1 + x\underline{s}_2^{(r)}\}}. \tag{S1.40}$$

Then $a_n^{(r)}(z_1, z_2) \to a^{(r)}(z_1, z_2)$ uniformly on neighborhoods of the contours, where

$$a^{(r)}(z_1, z_2) = c \, \underline{s}_1^{(r)} \underline{s}_2^{(r)} \int \frac{x^2 \, \mathrm{d}H^{(r)}(x)}{\{1 + x\underline{s}_1^{(r)}\}\{1 + x\underline{s}_2^{(r)}\}}. \tag{S1.41}$$

For the first covariance component, define

$$L_{n,t}^{(r)}(z_1, z_2) = \frac{1}{n} b_t^{(r,j)}(z_1) b_t^{(r,i)}(z_2) \mathbb{E} \, \mathrm{tr} \left[\mathbb{E}_t \mathbf{T}_r^\top \{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1} \mathbf{T}_r \, \mathbb{E}_t \mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1} \mathbf{T}_r \right]. \tag{S1.42}$$

Then, by the definition of $\mathcal{K}_{1,n}^{(r)}(z_1, z_2)$ in (S1.38),

$$(q+1)\rho_r^2 \mathcal{K}_{1,n}^{(r)}(z_1, z_2) = (q+1)\rho_r^2 \frac{\partial^2}{\partial z_1 \partial z_2} \left\{ \frac{1}{n} \sum_{t=1}^{n} L_{n,t}^{(r)}(z_1, z_2) \right\}. \tag{S1.43}$$

For later use, put

$$\mathbf{\Phi}_{j,t}^{(r)}(z) = \left\{ z\mathbf{I}_p - \frac{n-1}{n} b_t^{(r,j)}(z) \mathbf{\Sigma}^{(r)} \right\}^{-1}, \quad \mathbf{\Phi}_{i,t}^{(r)}(z) = \left\{ z\mathbf{I}_p - \frac{n-1}{n} b_t^{(r,i)}(z) \mathbf{\Sigma}^{(r)} \right\}^{-1}.$$

For $a \in \{i, j\}$ and $\ell \neq t$, define

$$\mathbf{D}_{a,t\ell}^{(r)}(z) = \mathbf{D}_{a,t}^{(r)}(z) - \boldsymbol{r}_\ell^{(r,a)}(\boldsymbol{r}_\ell^{(r,a)})^\top$$

and

$$\vartheta_{t\ell}^{(r,a)}(z) = \big[1 + (\boldsymbol{r}_\ell^{(r,a)})^\top \{\mathbf{D}_{a,t\ell}^{(r)}(z)\}^{-1}\boldsymbol{r}_\ell^{(r,a)}\big]^{-1}.$$

Following (Bai and Silverstein, 2010, Formula (9.9.12)), for $a \in \{i, j\}$,

$$\{\mathbf{D}_{a,t}^{(r)}(z)\}^{-1} = -\boldsymbol{\Phi}_{a,t}^{(r)}(z) + b_t^{(r,a)}(z)\mathcal{A}_t^{(r,a)}(z) + \mathcal{B}_t^{(r,a)}(z) + \mathcal{C}_t^{(r,a)}(z),$$

where

$$\begin{aligned}
\mathcal{A}_t^{(r,a)}(z) &= \sum_{\ell\neq t} \boldsymbol{\Phi}_{a,t}^{(r)}(z)\big\{\boldsymbol{r}_\ell^{(r,a)}(\boldsymbol{r}_\ell^{(r,a)})^\top - \frac{1}{n}\boldsymbol{\Sigma}^{(r)}\big\}\{\mathbf{D}_{a,t\ell}^{(r)}(z)\}^{-1},\\
\mathcal{B}_t^{(r,a)}(z) &= \sum_{\ell\neq t} \big\{\vartheta_{t\ell}^{(r,a)}(z) - b_t^{(r,a)}(z)\big\}\boldsymbol{\Phi}_{a,t}^{(r)}(z)\boldsymbol{r}_\ell^{(r,a)}(\boldsymbol{r}_\ell^{(r,a)})^\top\{\mathbf{D}_{a,t\ell}^{(r)}(z)\}^{-1},\\
\mathcal{C}_t^{(r,a)}(z) &= \frac{1}{n}b_t^{(r,a)}(z)\boldsymbol{\Phi}_{a,t}^{(r)}(z)\boldsymbol{\Sigma}^{(r)}\sum_{\ell\neq t}\big[\{\mathbf{D}_{a,t\ell}^{(r)}(z)\}^{-1} - \{\mathbf{D}_{a,t}^{(r)}(z)\}^{-1}\big].
\end{aligned} \tag{S1.44}$$

Thus, we have

$$\begin{aligned}
&\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{a,t}^{(r)}(z)\}^{-1}\mathbf{T}_r \qquad \text{(S1.45)}\\
&\quad= b_t^{(r,a)}(z_1)\sum_{\ell<t}\mathbb{E}_t\mathbf{T}_r^\top\boldsymbol{\Phi}_{a,t}^{(r)}(z)\big\{\boldsymbol{r}_\ell^{(r,a)}(\boldsymbol{r}_\ell^{(r,a)})^\top - n^{-1}\boldsymbol{\Sigma}^{(r)}\big\}\{\mathbf{D}_{a,t\ell}^{(r)}(z)\}^{-1}\mathbf{T}_r\\
&\qquad - \mathbf{T}_r^\top\boldsymbol{\Phi}_{a,t}^{(r)}(z)\mathbf{T}_r + \mathbf{H}_t^{(r,a)}(z),
\end{aligned}$$

where the remainder $\mathbf{H}_t^{(r,a)}(z) := \mathbb{E}_t\mathbf{T}_r^\top\mathcal{B}_t^{(r,a)}(z)\mathbf{T}_r + \mathbb{E}_t\mathbf{T}_r^\top\mathcal{C}_t^{(r,a)}(z)\mathbf{T}_r$. Recall

(S1.45), we first consider the form in (S1.42),

$$\mathcal{R}_{0,1}^{(r)} = -\frac{1}{n} b_t^{(r,j)}(z_1)\{b_t^{(r,i)}(z_2)\}^2 \times \tag{S1.46}$$
$$\sum_{\ell<t} \mathbb{E}\,\mathrm{tr}\Big[\mathbf{T}_r^\top \mathbf{\Phi}_{j,t}^{(r)}(z_1)\mathbf{T}_r \times \mathbb{E}_t \mathbf{T}_r^\top \mathbf{\Phi}_{i,t}^{(r)}(z_2)\big\{\boldsymbol{r}_\ell^{(r,i)}(\boldsymbol{r}_\ell^{(r,i)})^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}\{\mathbf{D}_{i,t\ell}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\Big],$$

and

$$\mathcal{R}_{1,0}^{(r)} = -\frac{1}{n}\{b_t^{(r,j)}(z_1)\}^2 b_t^{(r,i)}(z_2) \times \tag{S1.47}$$
$$\sum_{\ell<t} \mathbb{E}\,\mathrm{tr}\Big[\mathbb{E}_t \mathbf{T}_r^\top \mathbf{\Phi}_{j,t}^{(r)}(z_1)\big\{\boldsymbol{r}_\ell^{(r,j)}(\boldsymbol{r}_\ell^{(r,j)})^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}\{\mathbf{D}_{j,t\ell}^{(r)}(z_1)\}^{-1}\mathbf{T}_r \times \mathbf{T}_r^\top \mathbf{\Phi}_{i,t}^{(r)}(z_2)\mathbf{T}_r\Big].$$

For example, in the first display, $\{\mathbf{D}_{i,t\ell}^{(r)}(z_2)\}^{-1}$ does not contain the ℓ-th rank-one term. Hence, conditional on the variables other than $\boldsymbol{x}_{i\ell}^{(r)}$, and $\mathbb{E}\big\{\boldsymbol{r}_\ell^{(r,i)}(\boldsymbol{r}_\ell^{(r,i)})^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\} = 0$. Therefore,

$$\mathcal{R}_{0,1}^{(r)} = 0, \quad \text{and} \quad \mathcal{R}_{1,0}^{(r)} = 0.$$

Then consider the form in (S1.42),

$$\mathcal{R}_{\neq}^{(r)} = \frac{1}{n}\{b_t^{(r,j)}(z_1) b_t^{(r,i)}(z_2)\}^2 \times \tag{S1.48}$$
$$\sum_{\substack{\ell,m<t\\ \ell\neq m}} \mathbb{E}\,\mathrm{tr}\Big[\mathbb{E}_t \mathbf{T}_r^\top \mathbf{\Phi}_{j,t}^{(r)}(z_1)\big\{\boldsymbol{r}_\ell^{(r,j)}(\boldsymbol{r}_\ell^{(r,j)})^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}\{\mathbf{D}_{j,t\ell}^{(r)}(z_1)\}^{-1}\mathbf{T}_r$$
$$\times\, \mathbb{E}_t \mathbf{T}_r^\top \mathbf{\Phi}_{i,t}^{(r)}(z_2)\big\{\boldsymbol{r}_m^{(r,i)}(\boldsymbol{r}_m^{(r,i)})^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}\{\mathbf{D}_{i,tm}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\Big].$$

For $\ell \neq m < t$, define

$$\mathbf{D}_{j,t\ell m}^{(r)}(z) = \mathbf{D}_{j,t}^{(r)}(z) - \boldsymbol{r}_\ell^{(r,j)}(\boldsymbol{r}_\ell^{(r,j)})^\top - \boldsymbol{r}_m^{(r,j)}(\boldsymbol{r}_m^{(r,j)})^\top,$$

and

$$\mathbf{D}_{i,tm\ell}^{(r)}(z) = \mathbf{D}_{i,t}^{(r)}(z) - \boldsymbol{r}_m^{(r,i)}(\boldsymbol{r}_m^{(r,i)})^\top - \boldsymbol{r}_\ell^{(r,i)}(\boldsymbol{r}_\ell^{(r,i)})^\top.$$

By the rank-one perturbation formula, for $\ell \neq m < t$,

$$\{\mathbf{D}_{j,t\ell}^{(r)}(z_1)\}^{-1} = \{\mathbf{D}_{j,t\ell m}^{(r)}(z_1)\}^{-1} - \vartheta_{t\ell m}^{(r,j)}(z_1)\{\mathbf{D}_{j,t\ell m}^{(r)}(z_1)\}^{-1}\boldsymbol{r}_m^{(r,j)}(\boldsymbol{r}_m^{(r,j)})^\top\{\mathbf{D}_{j,t\ell m}^{(r)}(z_1)\}^{-1},$$

where

$$\vartheta_{t\ell m}^{(r,j)}(z_1) = \big[1 + (\boldsymbol{r}_m^{(r,j)})^\top\{\mathbf{D}_{j,t\ell m}^{(r)}(z_1)\}^{-1}\boldsymbol{r}_m^{(r,j)}\big]^{-1}.$$

Similarly,

$$\{\mathbf{D}_{i,tm}^{(r)}(z_2)\}^{-1} = \{\mathbf{D}_{i,tm\ell}^{(r)}(z_2)\}^{-1} - \vartheta_{tm\ell}^{(r,i)}(z_2)\{\mathbf{D}_{i,tm\ell}^{(r)}(z_2)\}^{-1}\boldsymbol{r}_\ell^{(r,i)}(\boldsymbol{r}_\ell^{(r,i)})^\top\{\mathbf{D}_{i,tm\ell}^{(r)}(z_2)\}^{-1},$$

where

$$\vartheta_{tm\ell}^{(r,i)}(z_2) = \big[1 + (\boldsymbol{r}_\ell^{(r,i)})^\top\{\mathbf{D}_{i,tm\ell}^{(r)}(z_2)\}^{-1}\boldsymbol{r}_\ell^{(r,i)}\big]^{-1}.$$

Substituting these two identities into $\mathcal{R}_{\neq}^{(r)}$, we get

$$\mathcal{R}_{\neq}^{(r)} = \mathcal{R}_{\neq,0}^{(r)} + \mathcal{R}_{\neq,j}^{(r)} + \mathcal{R}_{\neq,i}^{(r)} + \mathcal{R}_{\neq,ji}^{(r)},$$

where

$$\begin{aligned}\mathcal{R}_{\neq,0}^{(r)} = \frac{1}{n}\{b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\}^2 \times \\ \sum_{\substack{\ell,m<t \\ \ell\neq m}} \mathbb{E}\operatorname{tr}\Big[\mathbb{E}_t\mathbf{T}_r^\top\boldsymbol{\Phi}_{j,t}^{(r)}(z_1)\big\{\boldsymbol{r}_\ell^{(r,j)}(\boldsymbol{r}_\ell^{(r,j)})^\top - n^{-1}\boldsymbol{\Sigma}^{(r)}\big\}\{\mathbf{D}_{j,t\ell m}^{(r)}(z_1)\}^{-1}\mathbf{T}_r \\ \times \mathbb{E}_t\mathbf{T}_r^\top\boldsymbol{\Phi}_{i,t}^{(r)}(z_2)\big\{\boldsymbol{r}_m^{(r,i)}(\boldsymbol{r}_m^{(r,i)})^\top - n^{-1}\boldsymbol{\Sigma}^{(r)}\big\}\{\mathbf{D}_{i,tm\ell}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\Big],\end{aligned} \tag{S1.49}$$

$$\mathcal{R}^{(r)}_{\neq,j} = -n^{-1}\{b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\}^2\times \tag{S1.50}$$
$$\sum_{\substack{\ell,m<t\\ \ell\neq m}} \mathbb{E}\,\mathrm{tr}\Big[\mathbb{E}_t\mathbf{T}_r^\top\mathbf{\Phi}^{(r)}_{j,t}(z_1)\big\{\boldsymbol{r}^{(r,j)}_\ell(\boldsymbol{r}^{(r,j)}_\ell)^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}$$
$$\times\,\vartheta^{(r,j)}_{t\ell m}(z_1)\{\mathbf{D}^{(r)}_{j,t\ell m}(z_1)\}^{-1}\boldsymbol{r}^{(r,j)}_m(\boldsymbol{r}^{(r,j)}_m)^\top\{\mathbf{D}^{(r)}_{j,t\ell m}(z_1)\}^{-1}\mathbf{T}_r$$
$$\times\,\mathbb{E}_t\mathbf{T}_r^\top\mathbf{\Phi}^{(r)}_{i,t}(z_2)\big\{\boldsymbol{r}^{(r,i)}_m(\boldsymbol{r}^{(r,i)}_m)^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}\{\mathbf{D}^{(r)}_{i,tm\ell}(z_2)\}^{-1}\mathbf{T}_r\Big],$$

$$\mathcal{R}^{(r)}_{\neq,i} = -n^{-1}\{b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\}^2\times \tag{S1.51}$$
$$\sum_{\substack{\ell,m<t\\ \ell\neq m}} \mathbb{E}\,\mathrm{tr}\Big[\mathbb{E}_t\mathbf{T}_r^\top\mathbf{\Phi}^{(r)}_{j,t}(z_1)\big\{\boldsymbol{r}^{(r,j)}_\ell(\boldsymbol{r}^{(r,j)}_\ell)^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}\{\mathbf{D}^{(r)}_{j,t\ell m}(z_1)\}^{-1}\mathbf{T}_r$$
$$\times\,\mathbb{E}_t\mathbf{T}_r^\top\mathbf{\Phi}^{(r)}_{i,t}(z_2)\big\{\boldsymbol{r}^{(r,i)}_m(\boldsymbol{r}^{(r,i)}_m)^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}$$
$$\times\,\vartheta^{(r,i)}_{tm\ell}(z_2)\{\mathbf{D}^{(r)}_{i,tm\ell}(z_2)\}^{-1}\boldsymbol{r}^{(r,i)}_\ell(\boldsymbol{r}^{(r,i)}_\ell)^\top\{\mathbf{D}^{(r)}_{i,tm\ell}(z_2)\}^{-1}\mathbf{T}_r\Big],$$

and

$$\mathcal{R}^{(r)}_{\neq,ji} = n^{-1}\{b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\}^2\times \tag{S1.52}$$
$$\sum_{\substack{\ell,m<t\\ \ell\neq m}} \mathbb{E}\,\mathrm{tr}\Big[\mathbb{E}_t\mathbf{T}_r^\top\mathbf{\Phi}^{(r)}_{j,t}(z_1)\big\{\boldsymbol{r}^{(r,j)}_\ell(\boldsymbol{r}^{(r,j)}_\ell)^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}$$
$$\times\,\vartheta^{(r,j)}_{t\ell m}(z_1)\{\mathbf{D}^{(r)}_{j,t\ell m}(z_1)\}^{-1}\boldsymbol{r}^{(r,j)}_m(\boldsymbol{r}^{(r,j)}_m)^\top\{\mathbf{D}^{(r)}_{j,t\ell m}(z_1)\}^{-1}\mathbf{T}_r$$
$$\times\,\mathbb{E}_t\mathbf{T}_r^\top\mathbf{\Phi}^{(r)}_{i,t}(z_2)\big\{\boldsymbol{r}^{(r,i)}_m(\boldsymbol{r}^{(r,i)}_m)^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big\}$$
$$\times\,\vartheta^{(r,i)}_{tm\ell}(z_2)\{\mathbf{D}^{(r)}_{i,tm\ell}(z_2)\}^{-1}\boldsymbol{r}^{(r,i)}_\ell(\boldsymbol{r}^{(r,i)}_\ell)^\top\{\mathbf{D}^{(r)}_{i,tm\ell}(z_2)\}^{-1}\mathbf{T}_r\Big].$$

After the three-deleted replacement, the remaining matrix factors in $\mathcal{R}^{(r)}_{\neq,0}$

are independent of the ℓ-th and m-th column-pairs. Since $\ell \neq m$ and

$$\mathbb{E}\{\boldsymbol{r}_s^{(r,a)}(\boldsymbol{r}_s^{(r,a)})^\top - n^{-1}\boldsymbol{\Sigma}^{(r)}\} = 0, \qquad a \in \{i,j\}, \text{ and } s \in \{m,\ell\}$$

we have $\mathcal{R}_{\neq,0}^{(r)} = 0$. Similarly, $\mathcal{R}_{\neq,j}^{(r)}$ contains a single centered ℓ-th factor independent of the remaining matrix factors, while $\mathcal{R}_{\neq,i}^{(r)}$ contains a single centered m-th factor. Therefore, $\mathcal{R}_{\neq,j}^{(r)} = \mathcal{R}_{\neq,i}^{(r)} = 0$.

It remains to control the term containing the two rank-one perturbation errors. For bounded matrix factors $\mathbf{A}_1, \ldots, \mathbf{A}_5$ independent of the ℓ-th and m-th column-pairs, the mixed quadratic-form identity gives, uniformly for $\ell \neq m < t$,

$$\begin{aligned}\Big|\mathbb{E}\,\mathrm{tr}\Big[&\mathbf{A}_1\{\boldsymbol{r}_\ell^{(r,j)}(\boldsymbol{r}_\ell^{(r,j)})^\top - n^{-1}\boldsymbol{\Sigma}^{(r)}\}\mathbf{A}_2\boldsymbol{r}_m^{(r,j)}(\boldsymbol{r}_m^{(r,j)})^\top\mathbf{A}_3\\ &\times\{\boldsymbol{r}_m^{(r,i)}(\boldsymbol{r}_m^{(r,i)})^\top - n^{-1}\boldsymbol{\Sigma}^{(r)}\}\mathbf{A}_4\boldsymbol{r}_\ell^{(r,i)}(\boldsymbol{r}_\ell^{(r,i)})^\top\mathbf{A}_5\Big]\Big| \leq Kn^{-2}.\end{aligned}$$

Since the coefficients $b_t^{(r,j)}(z_1)$, $b_t^{(r,i)}(z_2)$, $\vartheta_{t\ell m}^{(r,j)}(z_1)$, and $\vartheta_{tm\ell}^{(r,i)}(z_2)$ are uniformly bounded, this bound yields

$$|\mathcal{R}_{\neq,ji}^{(r)}| \leq \frac{K}{n}\sum_{\substack{\ell,m<t\\ \ell\neq m}} n^{-2} = O(n^{-1}) = o(1).$$

Therefore,

$$\mathcal{R}_{\neq}^{(r)} = \mathcal{R}_{\neq,0}^{(r)} + \mathcal{R}_{\neq,j}^{(r)} + \mathcal{R}_{\neq,i}^{(r)} + \mathcal{R}_{\neq,ji}^{(r)} = o(1).$$

Moreover, by the estimates corresponding to (Bai and Silverstein, 2010, Formulae (9.9.14)–(9.9.15)), for any matrix factor $\mathbf{M}_t(z')$ with uniformly bounded spectral norm,

$$\frac{1}{n}\mathbb{E}\Big|\,\mathrm{tr}\left[\mathbf{H}_t^{(r,a)}(z)\mathbf{M}_t(z')\right]\Big| = o(1), \qquad a \in \{i,j\}.$$

Therefore, all products containing one remainder factor $\mathbf{H}_t^{(r,j)}(z_1)$ or $\mathbf{H}_t^{(r,i)}(z_2)$ and one bounded factor from the other expansion are negligible.

Combining the preceding decomposition with the negligibility of the single centered terms, the different index centered products, and the remainder terms containing $\mathbf{H}_t^{(r,a)}$, we obtain

$$L_{n,t}^{(r)}(z_1,z_2) = \mathcal{I}_{n,t}^{(r)}(z_1,z_2) + \mathcal{J}_{n,t}^{(r)}(z_1,z_2) + o(1), \tag{S1.53}$$

uniformly in t, z_1, z_2, where

$$\mathcal{I}_{n,t}^{(r)}(z_1,z_2) = \frac{1}{n} b_t^{(r,j)}(z_1) b_t^{(r,i)}(z_2)\,\mathrm{tr}\left[\mathbf{T}_r^\top \mathbf{\Phi}_{j,t}^{(r)}(z_1)\mathbf{T}_r\,\mathbf{T}_r^\top \mathbf{\Phi}_{i,t}^{(r)}(z_2)\mathbf{T}_r\right], \tag{S1.54}$$

and the terms with the same index $\ell < t$ are collected in

$$\begin{aligned}\mathcal{J}_{n,t}^{(r)}(z_1,z_2) = \frac{1}{n}\{b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\}^2 \sum_{\ell<t} \mathbb{E}\,\mathrm{tr}\Big[\mathbb{E}_t\big[\mathbf{T}_r^\top \mathbf{\Phi}_{j,t}^{(r)}(z_1)\{\boldsymbol{r}_\ell^{(r,j)}(\boldsymbol{r}_\ell^{(r,j)})^\top - \frac{1}{n}\mathbf{\Sigma}^{(r)}\}\\ \times \{\mathbf{D}_{j,t\ell}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\big]\,\mathbb{E}_t\big[\mathbf{T}_r^\top \mathbf{\Phi}_{i,t}^{(r)}(z_2)\{\boldsymbol{r}_\ell^{(r,i)}(\boldsymbol{r}_\ell^{(r,i)})^\top - \frac{1}{n}\mathbf{\Sigma}^{(r)}\}\\ \times \{\mathbf{D}_{i,t\ell}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\big]\Big].\end{aligned} \tag{S1.55}$$

We first evaluate $\mathcal{I}_{n,t}^{(r)}(z_1,z_2)$. By the standard deterministic-equivalent estimate for the deleted resolvent traces, both $b_t^{(r,j)}(z)$ and $b_t^{(r,i)}(z)$ have the same deterministic equivalent:

$$b_t^{(r,j)}(z),\, b_t^{(r,i)}(z) = -z\underline{s}^{(r)}(z) + o(1)$$

uniformly on the contours. Hence, by the spectral decomposition of $\mathbf{\Sigma}^{(r)}$,

$$
\begin{aligned}
\mathcal{I}_{n,t}^{(r)}(z_1,z_2) &= \frac{1}{n} b_t^{(r,j)}(z_1) b_t^{(r,i)}(z_2) \sum_{k=1}^{p} \frac{d_{rk}^2}{\left\{z_1 - \frac{n-1}{n} b_t^{(r,j)}(z_1) d_{rk}\right\}\left\{z_2 - \frac{n-1}{n} b_t^{(r,i)}(z_2) d_{rk}\right\}} \\
&= c_n \underline{s}_1^{(r)} \underline{s}_2^{(r)} \int \frac{x^2 \,\mathrm{d}H_n^{(r)}(x)}{\{1 + x\underline{s}_1^{(r)}\}\{1 + x\underline{s}_2^{(r)}\}} + o(1) \\
&= a_n^{(r)}(z_1,z_2) + o(1), \qquad \text{(S1.56)}
\end{aligned}
$$

uniformly in t, z_1, z_2.

Next we evaluate $\mathcal{J}_{n,t}^{(r)}(z_1,z_2)$. Fix $\ell < t$, and let

$$
\mathcal{F}_{t,\ell} = \sigma\{(\boldsymbol{x}_{jm}^{(r)}, \boldsymbol{x}_{im}^{(r)}) : 1 \le m \le t,\ m \ne \ell\}.
$$

After the ℓ-th rank-one term has been removed, the conditional expectations of the deleted resolvents are $\mathcal{F}_{t,\ell}$-measurable. Thus they may be treated as fixed matrices when conditioning on $\mathcal{F}_{t,\ell}$. Since

$$
\boldsymbol{r}_\ell^{(r,a)} (\boldsymbol{r}_\ell^{(r,a)})^\top - n^{-1}\mathbf{\Sigma}^{(r)} = \frac{1}{n} \mathbf{T}_r \big\{\boldsymbol{x}_{a\ell}^{(r)} (\boldsymbol{x}_{a\ell}^{(r)})^* - \mathbf{I}_p\big\} \mathbf{T}_r^\top, \qquad a \in \{i,j\},
$$

the conditional fourth-order contraction gives, uniformly in t, ℓ, z_1, z_2,

$$
\begin{aligned}
&\mathbb{E}\,\mathrm{tr}\Big[\mathbb{E}_t\big\{\mathbf{T}_r^\top \mathbf{\Phi}_{j,t}^{(r)}(z_1)\big(\boldsymbol{r}_\ell^{(r,j)}(\boldsymbol{r}_\ell^{(r,j)})^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big)\{\mathbf{D}_{j,t\ell}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\big\} \\
&\qquad \times \mathbb{E}_t\big\{\mathbf{T}_r^\top \mathbf{\Phi}_{i,t}^{(r)}(z_2)\big(\boldsymbol{r}_\ell^{(r,i)}(\boldsymbol{r}_\ell^{(r,i)})^\top - n^{-1}\mathbf{\Sigma}^{(r)}\big)\{\mathbf{D}_{i,t\ell}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\big\}\Big] \\
&= \frac{\rho_r^2}{n^2}\,\mathrm{tr}\left[\mathbf{\Phi}_{j,t}^{(r)}(z_1)\mathbf{\Sigma}^{(r)}\mathbf{\Phi}_{i,t}^{(r)}(z_2)\mathbf{\Sigma}^{(r)}\right] \qquad \text{(S1.57)} \\
&\qquad \times \mathbb{E}\,\mathrm{tr}\left[\mathbb{E}_t \mathbf{T}_r^\top\{\mathbf{D}_{j,t\ell}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\,\mathbb{E}_t \mathbf{T}_r^\top\{\mathbf{D}_{i,t\ell}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\right] + O(n^{-1}).
\end{aligned}
$$

Here the retained cross-pairing yields a product of two traces, each of order $O(p)$ by the resolvent bounds and the boundedness of $\mathbf{\Sigma}^{(r)}$ and $\mathbf{\Phi}_{a,t}^{(r)}(z)$.

Hence this term is $n^{-2}O(p^2) = O(1)$. The remaining contractions contain at most one such trace, or a diagonal fourth-cumulant sum bounded by $O(p)$ via the Cauchy–Schwarz inequality, and are therefore $n^{-2}O(p) = O(n^{-1})$ uniformly in t, ℓ, z_1, z_2. Thus we have $1/n \sum_{\ell<t} O(n^{-1}) = o(1)$.

By the rank-one perturbation formula and the uniform resolvent bounds, replacing $\mathbf{D}_{\cdot,t\ell}^{-1}$ by $\mathbf{D}_{\cdot,t}^{-1}$ inside the conditional normalized trace changes the expression by $o(1)$. More precisely,

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\operatorname{tr}\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t\ell}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\,\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{i,t\ell}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\right] \\
&\quad= \frac{1}{n}\mathbb{E}\operatorname{tr}\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\,\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\right] + o(1),
\end{aligned} \tag{S1.58}$$

uniformly in t, ℓ, z_1, z_2. Therefore, by the definition of $L_{n,t}^{(r)}(z_1, z_2)$ in (S1.42),

$$\begin{aligned}
&\mathbb{E}\operatorname{tr}\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t\ell}^{(r)}(z_1)\}^{-1}\mathbf{T}_r\,\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{i,t\ell}^{(r)}(z_2)\}^{-1}\mathbf{T}_r\right] \\
&\quad= \frac{n}{b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)}L_{n,t}^{(r)}(z_1, z_2) + o(n).
\end{aligned} \tag{S1.59}$$

The division by $b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)$ is valid since $b_t^{(r,j)}(z), b_t^{(r,i)}(z) = -z\underline{s}^{(r)}(z) + o(1)$ uniformly on the contours and the contours are chosen away from the singularities. Moreover, by the cyclic property of the trace and the same spectral decomposition as in (S1.56),

$$\frac{1}{n}b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\operatorname{tr}\left[\mathbf{\Phi}_{j,t}^{(r)}(z_1)\mathbf{\Sigma}^{(r)}\mathbf{\Phi}_{i,t}^{(r)}(z_2)\mathbf{\Sigma}^{(r)}\right] = a_n^{(r)}(z_1, z_2) + o(1). \tag{S1.60}$$

Combining (S1.57), (S1.59), and (S1.60), and multiplying by the outer factor n^{-1} in (S1.55), the contribution of a fixed $\ell < t$ to $\mathcal{J}_{n,t}^{(r)}(z_1, z_2)$ is

$$\frac{\rho_r^2}{n} a_n^{(r)}(z_1, z_2) L_{n,t}^{(r)}(z_1, z_2) + o(n^{-1}).$$

Summing over $\ell \in [1 : t-1]$, we obtain

$$\mathcal{J}_{n,t}^{(r)}(z_1, z_2) = \frac{t-1}{n} \rho_r^2 a_n^{(r)}(z_1, z_2) L_{n,t}^{(r)}(z_1, z_2) + o(1), \tag{S1.61}$$

uniformly in t, z_1, z_2.

Combining (S1.53), (S1.56), and (S1.61), we have

$$L_{n,t}^{(r)}(z_1, z_2) = a_n^{(r)}(z_1, z_2) + \frac{t-1}{n} \rho_r^2 a_n^{(r)}(z_1, z_2) L_{n,t}^{(r)}(z_1, z_2) + o(1), \tag{S1.62}$$

uniformly in t, z_1, z_2. The contours are selected so that

$$\inf_{0 \le u \le 1,\, z_1 \in \mathcal{C}_1,\, z_2 \in \mathcal{C}_2} \left| 1 - u \rho_r^2 a^{(r)}(z_1, z_2) \right| > 0.$$

Together with the uniform convergence of $a_n^{(r)}$, this implies that

$$1 - \frac{t-1}{n} \rho_r^2 a_n^{(r)}(z_1, z_2)$$

is bounded away from zero uniformly in t, z_1, z_2 for all sufficiently large n.

Hence

$$L_{n,t}^{(r)}(z_1, z_2) = \frac{a_n^{(r)}(z_1, z_2)}{1 - \frac{t-1}{n} \rho_r^2 a_n^{(r)}(z_1, z_2)} + o(1). \tag{S1.63}$$

Consequently,

$$(q+1)\rho_r^2 \frac{1}{n} \sum_{t=1}^{n} L_{n,t}^{(r)}(z_1, z_2) = (q+1)\rho_r^2 \frac{1}{n} \sum_{t=1}^{n} \frac{a_n^{(r)}(z_1, z_2)}{1 - \frac{t-1}{n} \rho_r^2 a_n^{(r)}(z_1, z_2)} + o(1)$$

$$\longrightarrow (q+1)\rho_r^2 \int_0^1 \frac{a^{(r)}(z_1,z_2)}{1-u\rho_r^2 a^{(r)}(z_1,z_2)}\,\mathrm{d}u$$

$$= -(q+1)\log\bigl\{1-\rho_r^2 a^{(r)}(z_1,z_2)\bigr\}. \qquad \text{(S1.64)}$$

The convergence above is uniform on neighborhoods of $\mathcal{C}_1 \times \mathcal{C}_2$, and the functions involved are analytic in z_1 and z_2. Therefore, by Cauchy's differentiation formula, the derivatives also converge uniformly. From (S1.43) and (S1.64), we obtain

$$(q+1)\rho_r^2 \mathcal{K}_{1,n}^{(r)}(z_1,z_2)$$
$$\longrightarrow (q+1)\frac{\partial^2}{\partial z_1 \partial z_2}\left[-\log\left\{1-\rho_r^2 c\,\underline{s}^{(r)}(z_1)\underline{s}^{(r)}(z_2)\int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}\{1+x\underline{s}^{(r)}(z_2)\}}\right\}\right].$$
$$\text{(S1.65)}$$

Equivalently, the limiting kernel in (S1.65) is

$$\frac{(q+1)\rho_r^2 c(\underline{s}^{(r)})'(z_1)(\underline{s}^{(r)})'(z_2)}{\bigl[1-\rho_r^2 c\,\underline{s}_1^{(r)}\underline{s}_2^{(r)}\int x^2\{1+x\underline{s}_1^{(r)}\}^{-1}\{1+x\underline{s}_2^{(r)}\}^{-1}\,\mathrm{d}H^{(r)}(x)\bigr]^2} \qquad \text{(S1.66)}$$
$$\times\left[\left\{\int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}_1^{(r)}\}^2\{1+x\underline{s}_2^{(r)}\}^2}\right\}\left\{1-\rho_r^2 c\,\underline{s}_1^{(r)}\underline{s}_2^{(r)}\int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}_1^{(r)}\}\{1+x\underline{s}_2^{(r)}\}}\right\}\right.$$
$$\left.+\rho_r^2 c\,\underline{s}_1^{(r)}\underline{s}_2^{(r)}\left\{\int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}_1^{(r)}\}^2\{1+x\underline{s}_2^{(r)}\}}\right\}\left\{\int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}_1^{(r)}\}\{1+x\underline{s}_2^{(r)}\}^2}\right\}\right].$$

For the fourth-cumulant term, we give the detailed diagonal calculation. Let $\underline{s}_n^{(r)}(z)$ denote the companion Stieltjes transform associated with $c_n = p/n$ and $H_n^{(r)}$, determined by

$$z = -\frac{1}{\underline{s}_n^{(r)}(z)} + c_n \int \frac{x}{1+x\underline{s}_n^{(r)}(z)}\,\mathrm{d}H_n^{(r)}(x). \qquad \text{(S1.67)}$$

Since $c_n \to c$, $H_n^{(r)} \Rightarrow H^{(r)}$, and $\sup_p \|\boldsymbol{\Sigma}^{(r)}\| < \infty$, we have

$$\underline{s}_n^{(r)}(z) - \underline{s}^{(r)}(z) \longrightarrow 0, \tag{S1.68}$$

uniformly for z on a fixed neighborhood of the contours. Recall (S1.39). For $k \in [1:p]$, define

$$\mathfrak{b}_{n,k}^{(r)}(z) = \big[\underline{s}_n^{(r)}(z)\mathbf{T}_r^\top\{\mathbf{I}_p + \underline{s}_n^{(r)}(z)\boldsymbol{\Sigma}^{(r)}\}^{-1}\mathbf{T}_r\big]_{kk},$$

$$\mathfrak{b}_{k}^{(r)}(z) = \big[\underline{s}^{(r)}(z)\mathbf{T}_r^\top\{\mathbf{I}_p + \underline{s}^{(r)}(z)\boldsymbol{\Sigma}^{(r)}\}^{-1}\mathbf{T}_r\big]_{kk},$$

$$\mathfrak{b}_{t,k}^{(r,j)}(z) = \big[b_t^{(r,j)}(z)\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big]_{kk}.$$

We first compare the conditional leave-one-out resolvent term with $\mathfrak{b}_{n,k}^{(r)}(z)$. Following the proof of Najim and Yao (2016, Proposition 5.4), we first replace the conditional expectation by the corresponding expectation, then use the rank-one resolvent identity to pass from the leave-one-out resolvent to the full resolvent, and finally apply Najim and Yao (2016, Equation (5.17)). For $k \in [1:p]$, consider the deterministic vector $\mathbf{T}_r\mathbf{e}_k$, where $\mathbf{e}_k$ denotes the kth standard basis vector. By Assumption 3,

$$\|\mathbf{T}_r\mathbf{e}_k\|^2 = \mathbf{e}_k^\top\boldsymbol{\Sigma}^{(r)}\mathbf{e}_k \le \|\boldsymbol{\Sigma}^{(r)}\|,$$

and hence these vectors have uniformly bounded norms. Since $\mathbb{E}_t$ is a conditional expectation, Jensen's inequality gives

$$\mathbb{E}\Big|\big[(\mathbb{E}_t - \mathbb{E})\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big]_{kk}\Big|^2$$

$$
\begin{aligned}
&= \mathbb{E}\Big|\mathbb{E}_t\big[\mathbf{T}_r^\top\big\{\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} - \mathbb{E}\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\big\}\mathbf{T}_r\big]_{kk}\Big|^2 \\
&\le \mathbb{E}\Big|\big[\mathbf{T}_r^\top\big\{\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} - \mathbb{E}\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\big\}\mathbf{T}_r\big]_{kk}\Big|^2. \qquad \text{(S1.69)}
\end{aligned}
$$

For each fixed j, the columns $\{\boldsymbol{x}_{jt}^{(r)} : t \in [1 : n]\}$ are independent, and their entries are i.i.d. with zero mean, unit variance, and finite fourth moment. Therefore, the argument of Najim and Yao (2016, Proposition 5.3), applied with $\boldsymbol{u}_n = \mathbf{T}_r\mathbf{e}_k$, also applies after deleting the tth column and gives

$$
\sup_{\substack{1\le t\le n\\1\le k\le p}} \mathbb{E}\Big|\big[\mathbf{T}_r^\top\big\{\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} - \mathbb{E}\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\big\}\mathbf{T}_r\big]_{kk}\Big|^2 = O(n^{-1}), \qquad \text{(S1.70)}
$$

uniformly for z on a fixed neighborhood of the contours. Let $\mathbf{D}_j^{(r)}(z)$ denote the corresponding matrix before deleting the tth rank-one term. By the rank-one resolvent identity

$$
\{\mathbf{D}_j^{(r)}(z)\}^{-1} - \{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} = -\vartheta_t^{(r,j)}(z)\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\boldsymbol{r}_t^{(r,j)}(\boldsymbol{r}_t^{(r,j)})^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}.
$$

Moreover,

$$
\begin{aligned}
&\mathbb{E}\left|(\boldsymbol{r}_t^{(r,j)})^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\mathbf{e}_k\right|^2 \\
&= \frac{1}{n}\mathbb{E}\left[\mathbf{e}_k^\top\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\boldsymbol{\Sigma}^{(r)}\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\mathbf{e}_k\right] = O(n^{-1}), \qquad \text{(S1.71)}
\end{aligned}
$$

and hence

$$
\sup_{\substack{1\le t\le n\\1\le k\le p}} \left|\big[\mathbf{T}_r^\top\big\{\mathbb{E}\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} - \mathbb{E}\{\mathbf{D}_j^{(r)}(z)\}^{-1}\big\}\mathbf{T}_r\big]_{kk}\right| = O(n^{-1}), \qquad \text{(S1.72)}
$$

uniformly for z on a fixed neighborhood of the contours. Under the identification $\mathbf{R}_n = \boldsymbol{\Sigma}^{(r)}$ and $\widetilde{t}_n(z) = \underline{s}_n^{(r)}(z)$, the deterministic matrix $\mathbf{T}_n(z)$ in Najim and Yao (2016) is $-1/z\{\mathbf{I}_p + \underline{s}_n^{(r)}(z)\boldsymbol{\Sigma}^{(r)}\}^{-1}$. Applying Najim and Yao (2016, Equation (5.17)) with $\boldsymbol{u} = \boldsymbol{v} = \mathbf{T}_r\mathbf{e}_k$, we obtain

$$\sup_{1\le k\le p}\left|\left[\mathbb{E}\mathbf{T}_r^\top\{\mathbf{D}_j^{(r)}(z)\}^{-1}\mathbf{T}_r + \frac{1}{z}\mathbf{T}_r^\top\{\mathbf{I}_p + \underline{s}_n^{(r)}(z)\boldsymbol{\Sigma}^{(r)}\}^{-1}\mathbf{T}_r\right]_{kk}\right| = O(n^{-1/2}), \tag{S1.73}$$

uniformly for z on a fixed neighborhood of the contours. Combining (S1.72) and (S1.73) gives

$$\sup_{\substack{1\le t\le n\\1\le k\le p}}\left|\left[\mathbb{E}\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r + \frac{1}{z}\mathbf{T}_r^\top\{\mathbf{I}_p + \underline{s}_n^{(r)}(z)\boldsymbol{\Sigma}^{(r)}\}^{-1}\mathbf{T}_r\right]_{kk}\right| = O(n^{-1/2}). \tag{S1.74}$$

It follows from (S1.69), (S1.70), and (S1.74) that

$$\sup_{\substack{1\le t\le n\\1\le k\le p}}\mathbb{E}\left|\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r + \frac{1}{z}\mathbf{T}_r^\top\{\mathbf{I}_p + \underline{s}_n^{(r)}(z)\boldsymbol{\Sigma}^{(r)}\}^{-1}\mathbf{T}_r\right]_{kk}\right|^2 = O(n^{-1}), \tag{S1.75}$$

uniformly for z on a fixed neighborhood of the contours. By adding and subtracting $-z\underline{s}_n^{(r)}(z)\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\right]_{kk}$, we have

$$\begin{aligned}
&\mathfrak{b}_{t,k}^{(r,j)}(z) - \mathfrak{b}_{n,k}^{(r)}(z)\\
&= \left\{b_t^{(r,j)}(z) + z\underline{s}_n^{(r)}(z)\right\}\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\right]_{kk} \qquad \text{(S1.76)}\\
&\quad - z\underline{s}_n^{(r)}(z)\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r + \frac{1}{z}\mathbf{T}_r^\top\{\mathbf{I}_p + \underline{s}_n^{(r)}(z)\boldsymbol{\Sigma}^{(r)}\}^{-1}\mathbf{T}_r\right]_{kk}.
\end{aligned}$$

The resolvent bounds and the boundedness of $\|\mathbf{\Sigma}^{(r)}\|$ imply

$$\sup_{\substack{1\le t\le n\\1\le k\le p}}\left|\left[\mathbb{E}_t\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\right]_{kk}\right|\le K.$$

Moreover, by the argument leading to Bai and Silverstein (2010, Equation (9.9.20)),

$$\sup_{1\le t\le n}\left|b_t^{(r,j)}(z)+z\underline{s}_n^{(r)}(z)\right|=o(1), \tag{S1.77}$$

uniformly for z on a fixed neighborhood of the contours. Hence, using (S1.75), we obtain

$$\begin{aligned}&\frac{1}{n^2}\sum_{t=1}^n\sum_{k=1}^p\mathbb{E}\left|\mathfrak{b}_{t,k}^{(r,j)}(z)-\mathfrak{b}_{n,k}^{(r)}(z)\right|^2\\&\le K\frac{p}{n}\sup_{1\le t\le n}\left|b_t^{(r,j)}(z)+z\underline{s}_n^{(r)}(z)\right|^2+K\frac{p}{n^2}=o(1).\end{aligned} \tag{S1.78}$$

Since $\mathbf{T}_r=(\mathbf{\Sigma}^{(r)})^{1/2}$,

$$\mathfrak{b}_{n,k}^{(r)}(z)=\left[\underline{s}_n^{(r)}(z)\mathbf{\Sigma}^{(r)}\{\mathbf{I}_p+\underline{s}_n^{(r)}(z)\mathbf{\Sigma}^{(r)}\}^{-1}\right]_{kk}=1-\mathbf{e}_k^\top\{\mathbf{I}_p+\underline{s}_n^{(r)}(z)\mathbf{\Sigma}^{(r)}\}^{-1}\mathbf{e}_k, \tag{S1.79}$$

and similarly,

$$\mathfrak{b}_k^{(r)}(z)=1-\mathbf{e}_k^\top\{\mathbf{I}_p+\underline{s}^{(r)}(z)\mathbf{\Sigma}^{(r)}\}^{-1}\mathbf{e}_k. \tag{S1.80}$$

On the fixed neighborhood under consideration, the two deterministic matrix inverses above are uniformly bounded. By the resolvent identity,

$$\max_{1\le k\le p}\left|\mathfrak{b}_{n,k}^{(r)}(z)-\mathfrak{b}_k^{(r)}(z)\right|\le\left\|\{\mathbf{I}_p+\underline{s}_n^{(r)}(z)\mathbf{\Sigma}^{(r)}\}^{-1}-\{\mathbf{I}_p+\underline{s}^{(r)}(z)\mathbf{\Sigma}^{(r)}\}^{-1}\right\|$$

$$\leq K \left| \underline{s}_n^{(r)}(z) - \underline{s}^{(r)}(z) \right| = o(1), \tag{S1.81}$$

uniformly for z on a fixed neighborhood of the contours, where the last equality follows from (S1.68). Consequently,

$$\frac{1}{n^2} \sum_{t=1}^{n} \sum_{k=1}^{p} \left| \mathfrak{b}_{n,k}^{(r)}(z) - \mathfrak{b}_k^{(r)}(z) \right|^2 \leq \frac{p}{n} \max_{1 \leq k \leq p} \left| \mathfrak{b}_{n,k}^{(r)}(z) - \mathfrak{b}_k^{(r)}(z) \right|^2 = o(1). \tag{S1.82}$$

Using $|x+y|^2 \leq 2|x|^2 + 2|y|^2$, and combining (S1.78) with (S1.82), we obtain

$$\frac{1}{n^2} \sum_{t=1}^{n} \sum_{k=1}^{p} \mathbb{E} \left| \mathfrak{b}_{t,k}^{(r,j)}(z) - \mathfrak{b}_k^{(r)}(z) \right|^2 = o(1). \tag{S1.83}$$

The same argument gives

$$\frac{1}{n^2} \sum_{t=1}^{n} \sum_{k=1}^{p} \mathbb{E} \left| \mathfrak{b}_{t,k}^{(r,i)}(z) - \mathfrak{b}_k^{(r)}(z) \right|^2 = o(1). \tag{S1.84}$$

For every $t \in [1:n]$ and $k \in [1:p]$,

$$\begin{aligned}
&\mathfrak{b}_{t,k}^{(r,j)}(z_1) \mathfrak{b}_{t,k}^{(r,i)}(z_2) - \mathfrak{b}_k^{(r)}(z_1) \mathfrak{b}_k^{(r)}(z_2) \\
&= \left\{ \mathfrak{b}_{t,k}^{(r,j)}(z_1) - \mathfrak{b}_k^{(r)}(z_1) \right\} \left\{ \mathfrak{b}_{t,k}^{(r,i)}(z_2) - \mathfrak{b}_k^{(r)}(z_2) \right\} \\
&\quad + \mathfrak{b}_k^{(r)}(z_1) \left\{ \mathfrak{b}_{t,k}^{(r,i)}(z_2) - \mathfrak{b}_k^{(r)}(z_2) \right\} + \mathfrak{b}_k^{(r)}(z_2) \left\{ \mathfrak{b}_{t,k}^{(r,j)}(z_1) - \mathfrak{b}_k^{(r)}(z_1) \right\}.
\end{aligned} \tag{S1.85}$$

By the Cauchy–Schwarz inequality and (S1.83)– (S1.84),

$$\begin{aligned}
&\frac{1}{n^2} \sum_{t=1}^{n} \sum_{k=1}^{p} \mathbb{E} \left| \mathfrak{b}_{t,k}^{(r,j)}(z_1) - \mathfrak{b}_k^{(r)}(z_1) \right| \left| \mathfrak{b}_{t,k}^{(r,i)}(z_2) - \mathfrak{b}_k^{(r)}(z_2) \right| \\
&\leq \left\{ \frac{1}{n^2} \sum_{t=1}^{n} \sum_{k=1}^{p} \mathbb{E} \left| \mathfrak{b}_{t,k}^{(r,j)}(z_1) - \mathfrak{b}_k^{(r)}(z_1) \right|^2 \right\}^{1/2} \times \left\{ \frac{1}{n^2} \sum_{t=1}^{n} \sum_{k=1}^{p} \mathbb{E} \left| \mathfrak{b}_{t,k}^{(r,i)}(z_2) - \mathfrak{b}_k^{(r)}(z_2) \right|^2 \right\}^{1/2}
\end{aligned}$$

$$= o(1). \tag{S1.86}$$

Moreover, the uniform resolvent bounds imply $\sup_{1\le k\le p}|\mathfrak{b}_k^{(r)}(z)| \le K$. Consequently,

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n}\sum_{k=1}^{p}|\mathfrak{b}_k^{(r)}(z_1)|\mathbb{E}\big|\mathfrak{b}_{t,k}^{(r,i)}(z_2)-\mathfrak{b}_k^{(r)}(z_2)\big| \\
&\le K\Big(\frac{p}{n}\Big)^{1/2}\Big\{\frac{1}{n^2}\sum_{t=1}^{n}\sum_{k=1}^{p}\mathbb{E}\big|\mathfrak{b}_{t,k}^{(r,i)}(z_2)-\mathfrak{b}_k^{(r)}(z_2)\big|^2\Big\}^{1/2} = o(1),
\end{aligned} \tag{S1.87}$$

and the same estimate holds after interchanging i and j. Therefore, (S1.85) yields

$$\frac{1}{n^2}\sum_{t=1}^{n}\sum_{k=1}^{p}\mathbb{E}\big[\mathfrak{b}_{t,k}^{(r,j)}(z_1)\mathfrak{b}_{t,k}^{(r,i)}(z_2)\big] = \frac{1}{n}\sum_{k=1}^{p}\mathfrak{b}_k^{(r)}(z_1)\mathfrak{b}_k^{(r)}(z_2)+o(1), \tag{S1.88}$$

uniformly for z_1 and z_2 on fixed neighborhoods of $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. By (S1.80), Assumption 4 implies

$$\max_{1\le k\le p}\Big|\mathfrak{b}_k^{(r)}(z)-\int\frac{\underline{s}^{(r)}(z)x}{1+\underline{s}^{(r)}(z)x}\,\mathrm{d}H_n^{(r)}(x)\Big| \longrightarrow 0, \tag{S1.89}$$

uniformly for z on a fixed neighborhood of the contours. Hence,

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{p}\mathfrak{b}_k^{(r)}(z_1)\mathfrak{b}_k^{(r)}(z_2) \\
&= c_n\Big[\int\frac{\underline{s}^{(r)}(z_1)x}{1+\underline{s}^{(r)}(z_1)x}\,\mathrm{d}H_n^{(r)}(x)\Big]\Big[\int\frac{\underline{s}^{(r)}(z_2)x}{1+\underline{s}^{(r)}(z_2)x}\,\mathrm{d}H_n^{(r)}(x)\Big]+o(1) \\
&\longrightarrow c\Big[\int\frac{\underline{s}^{(r)}(z_1)x}{1+\underline{s}^{(r)}(z_1)x}\,\mathrm{d}H^{(r)}(x)\Big]\Big[\int\frac{\underline{s}^{(r)}(z_2)x}{1+\underline{s}^{(r)}(z_2)x}\,\mathrm{d}H^{(r)}(x)\Big].
\end{aligned} \tag{S1.90}$$

Combining (S1.88) and (S1.90), we obtain

$$\frac{1}{n^2}\sum_{t=1}^{n}\sum_{k=1}^{p}\mathbb{E}\big[\mathfrak{b}_{t,k}^{(r,j)}(z_1)\mathfrak{b}_{t,k}^{(r,i)}(z_2)\big]$$

$$\longrightarrow c\left[\int \frac{\underline{s}^{(r)}(z_1)x}{1+\underline{s}^{(r)}(z_1)x}\,\mathrm{d}H^{(r)}(x)\right]\left[\int \frac{\underline{s}^{(r)}(z_2)x}{1+\underline{s}^{(r)}(z_2)x}\,\mathrm{d}H^{(r)}(x)\right], \quad \text{(S1.91)}$$

where the convergence is uniform on neighborhoods of $\mathcal{C}_1 \times \mathcal{C}_2$. Since the functions involved are analytic, Cauchy's differentiation formula yields

$$\mathcal{K}_{2,n}^{(r)}(z_1, z_2) \longrightarrow c(\underline{s}^{(r)})'(z_1)(\underline{s}^{(r)})'(z_2) \int \frac{x\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}^2} \times \int \frac{x\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_2)\}^2}. \quad \text{(S1.92)}$$

Even without Assumption 4, explicit formulas can be derived for special cases. When $\mathbf{\Sigma}^{(r)}$ is diagonal, its deterministic diagonal average can be evaluated directly in terms of $H_n^{(r)}$. Indeed,

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{p} \mathfrak{b}_k^{(r)}(z_1)\mathfrak{b}_k^{(r)}(z_2) \\
&= \frac{1}{n}\operatorname{tr}\left[\underline{s}^{(r)}(z_1)\mathbf{\Sigma}^{(r)}\{\mathbf{I}_p + \underline{s}^{(r)}(z_1)\mathbf{\Sigma}^{(r)}\}^{-1}\underline{s}^{(r)}(z_2)\mathbf{\Sigma}^{(r)}\{\mathbf{I}_p + \underline{s}^{(r)}(z_2)\mathbf{\Sigma}^{(r)}\}^{-1}\right] \\
&= c_n \int \frac{\underline{s}^{(r)}(z_1)\underline{s}^{(r)}(z_2)x^2}{\{1+x\underline{s}^{(r)}(z_1)\}\{1+x\underline{s}^{(r)}(z_2)\}}\,\mathrm{d}H_n^{(r)}(x) \\
&\longrightarrow c \int \frac{\underline{s}^{(r)}(z_1)\underline{s}^{(r)}(z_2)x^2}{\{1+x\underline{s}^{(r)}(z_1)\}\{1+x\underline{s}^{(r)}(z_2)\}}\,\mathrm{d}H^{(r)}(x), \qquad \text{(S1.93)}
\end{aligned}$$

uniformly for z_1 and z_2 on fixed neighborhoods of $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. Combining (S1.88) and (S1.93), and then applying Cauchy's differentiation formula, we obtain

$$\mathcal{K}_{2,n}^{(r)}(z_1, z_2) \longrightarrow c(\underline{s}^{(r)})'(z_1)(\underline{s}^{(r)})'(z_2) \times \int \frac{x^2\,\mathrm{d}H^{(r)}(x)}{\{1+x\underline{s}^{(r)}(z_1)\}^2\{1+x\underline{s}^{(r)}(z_2)\}^2}. \quad \text{(S1.94)}$$

It remains to verify the finite-dimensional convergence and tightness of the resolvent processes. As in Bai and Silverstein (2010, Theorem 9.10), the truncation, centralization and rescaling argument may be applied first. This reduction does not change the limiting distribution of the linear spectral statistics. After this reduction, the above moment bounds hold uniformly. Decompose $\Xi_{n,j}^{(r)}(z) = M_{n,j}^{(r),1}(z) + M_{n,j}^{(r),2}(z)$, where

$$M_{n,j}^{(r),1}(z) = \operatorname{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1} - \mathbb{E}\operatorname{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1}$$

and

$$M_{n,j}^{(r),2}(z) = \mathbb{E}\operatorname{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1} - ps^{(r)}(z).$$

For the random part,

$$M_{n,j}^{(r),1}(z) = \sum_{t=1}^{n}(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1}.$$

By (S1.30),

$$\begin{aligned} M_{n,j}^{(r),1}(z) &= \sum_{t=1}^{n}(\mathbb{E}_t - \mathbb{E}_{t-1})\operatorname{tr}\{\mathbf{D}_j^{(r)}(z)\}^{-1} \\ &= -\sum_{t=1}^{n}\frac{\partial}{\partial z}\left[b_t^{(r,j)}(z)(\mathbb{E}_t - \mathbb{E}_{t-1})\eta_t^{(r,j)}(z)\right] + \sum_{t=1}^{n}\xi_{nt}^{(r,j)}(z). \quad \text{(S1.95)} \end{aligned}$$

The error term is negligible in the covariance calculation, since

$$\sum_{t=1}^{n}\mathbb{E}|\xi_{nt}^{(r,j)}(z)|^2 = o(1),$$

uniformly for z on the contours.

For any fixed $z_1, \ldots, z_m$, the Cramér–Wold device and the martingale central limit theorem apply. The conditional covariance converges by

(S1.37)–(S1.92). The Lindeberg condition follows from

$$
\begin{aligned}
&\sum_{t=1}^{n}\mathbb{E}\Big|\frac{\partial}{\partial z}\big[b_t^{(r,j)}(z)(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,j)}(z)\big]\Big|^4\\
&\quad\le K\sum_{t=1}^{n}\mathbb{E}\Big[|\eta_t^{(r,j)}(z)|^4+\big|\frac{\partial}{\partial z}\eta_t^{(r,j)}(z)\big|^4\Big]=O(n^{-1}).
\end{aligned}
$$

Therefore the finite-dimensional distributions converge to those of a Gaussian process. For tightness, the resolvent identity gives

$$
\{\mathbf{D}_j^{(r)}(z_1)\}^{-1}-\{\mathbf{D}_j^{(r)}(z_2)\}^{-1}=(z_1-z_2)\{\mathbf{D}_j^{(r)}(z_1)\}^{-1}\{\mathbf{D}_j^{(r)}(z_2)\}^{-1}.
$$

Thus

$$
\begin{aligned}
&M_{n,j}^{(r),1}(z_1)-M_{n,j}^{(r),1}(z_2)\\
&\quad=(z_1-z_2)\big[\operatorname{tr}\{\mathbf{D}_j^{(r)}(z_1)\}^{-1}\{\mathbf{D}_j^{(r)}(z_2)\}^{-1}-\mathbb{E}\operatorname{tr}\{\mathbf{D}_j^{(r)}(z_1)\}^{-1}\{\mathbf{D}_j^{(r)}(z_2)\}^{-1}\big].
\end{aligned}
$$

Using the martingale moment estimate in the proof of Bai and Silverstein (2010, Theorem 9.10), together with the uniform resolvent bound on the contours, gives

$$
\mathbb{E}\big|M_{n,j}^{(r),1}(z_1)-M_{n,j}^{(r),1}(z_2)\big|^2\le K|z_1-z_2|^2
$$

and

$$
\mathbb{E}\big|M_{n,j}^{(r),1}(z_1)-M_{n,j}^{(r),1}(z_2)\big|^4\le K|z_1-z_2|^2.
$$

These estimates imply tightness of $M_{n,j}^{(r),1}(z)$ on the contour. The same argument applies to $M_{n,i}^{(r),1}(z)$. For the non-random term, the treatment in

Bai and Silverstein (2010, Theorem 9.10) gives

$$M_{n,j}^{(r),2}(z) \longrightarrow M^{(r),2}(z)$$

uniformly on the contours, where $M^{(r),2}(z)$ is deterministic and analytic. The same convergence holds for $M_{n,i}^{(r),2}(z)$. Since this term is non-random, it contributes to the limiting mean but not to the cross-covariance.

The preceding covariance calculation has been carried out for $z_1, z_2 \in \mathbb{C}^+$. For points on the remaining parts of the closed contours, the same argument applies. Indeed, if $\mathbf{S}$ denotes any of $\mathbf{S}_j^{(r)}, \mathbf{S}_i^{(r)}, \mathbf{S}_j^{(r)} - \boldsymbol{r}_t^{(r,j)}(\boldsymbol{r}_t^{(r,j)})^\top, \mathbf{S}_i^{(r)} - \boldsymbol{r}_t^{(r,i)}(\boldsymbol{r}_t^{(r,i)})^\top$, then $\mathbf{S}$ is real symmetric. Hence, for $z \in \mathbb{C}\backslash\mathbb{R}$, $\|(\mathbf{S} - z\mathbf{I}_p)^{-1}\| \leq 1/|\Im z|$. Therefore all rank-one perturbation identities, martingale decompositions and quadratic-form estimates used above remain valid with $|\Im z|$ in place of $\Im z$. Consequently, the covariance kernel obtained above extends to the full contours used in the Cauchy integral representation, in the same passage from the stieltjes transform process to the LSS as in Bai and Silverstein (2010).

Finally, by the Cauchy integral representation,

$$\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j = -\frac{1}{2\pi \mathrm{i}} \oint_{\mathcal{C}_1} f(z_1)\Xi_{n,j}^{(r)}(z_1)\, dz_1. \tag{S1.96}$$

The empirical process $\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_i$ is defined analogously, with $j, z_1, \mathcal{C}_1$ replaced by $i, z_2, \mathcal{C}_2$, respectively. Substituting (S1.37), (S1.65), and (S1.92) gives

$$\operatorname{Cov}\left(\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_j, \{\boldsymbol{\mathcal{Q}}_n^{(r)}(f)\}_i\right) \longrightarrow \nu_{f,\rho}^{(r)},$$

where

$$\begin{aligned}
\nu_{f,\rho}^{(r)} = -\frac{q+1}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} f(z_1) f(z_2) \frac{\partial^2}{\partial z_1 \partial z_2} \Big[-\log \big\{ 1 - \rho_r^2 c\, \underline{s}^{(r)}(z_1) \underline{s}^{(r)}(z_2) \\
\times \int \frac{x^2 \, \mathrm{d}H^{(r)}(x)}{\{1 + x\underline{s}^{(r)}(z_1)\}\{1 + x\underline{s}^{(r)}(z_2)\}} \big\} \Big] \, \mathrm{d}z_1 \mathrm{d}z_2 \\
- \frac{\beta_\rho^{(r)} c}{4\pi^2} \oint_{\mathcal{C}_1} \oint_{\mathcal{C}_2} f(z_1) f(z_2) \int \frac{x \, \mathrm{d}H^{(r)}(x)}{\{1 + x\underline{s}^{(r)}(z_1)\}^2} \int \frac{x \, \mathrm{d}H^{(r)}(x)}{\{1 + x\underline{s}^{(r)}(z_2)\}^2} \, \mathrm{d}\underline{s}^{(r)}(z_1) \, \mathrm{d}\underline{s}^{(r)}(z_2).
\end{aligned}$$

This proves the claimed cross-covariance formula.

S1.3 Proof of the Cross-covariance under Shared-component Dominance

We next show that the simplified covariance formula used in the main text is recovered under the shared-component dominance condition for two rows in the same row-population. Fix $r \in [1:\tau]$ and take two distinct row indices $i \neq j \in [1:h_r]$. we formulate the shared-component dominance condition through the smallness of the difference between two dependent row samples. Specifically, assume that, uniformly for $t \in [1:n]$,

$$\frac{1}{p}\mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^2 = o(n^{-1}),\ \mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^4 = O\Big[\big\{\mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^2\big\}^2\Big]. \tag{S1.97}$$

The first condition means that the row-specific discrepancy is negligible relative to the p-dimensional signal size, while the second condition controls its fourth moment. This is the sense in which the shared component dominates the difference between the two row samples. Then, define the perturbation between the two sample covariance matrices by

$$\Delta^{(r,ji)} := \mathbf{S}_j^{(r)} - \mathbf{S}_i^{(r)}. \tag{S1.98}$$

Since (S1.6), (S1.98) can be further deduced

$$\begin{aligned}
\Delta^{(r,ji)} &= \frac{1}{n}\mathbf{T}_r\{\mathbf{X}_j^{(r)}(\mathbf{X}_j^{(r)})^\top - \mathbf{X}_i^{(r)}(\mathbf{X}_i^{(r)})^\top\}\mathbf{T}_r^\top \\
&= \frac{1}{n}\mathbf{T}_r\Big[\mathbf{X}_i^{(r)}\{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}^\top + \{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}(\mathbf{X}_i^{(r)})^\top \\
&\qquad + \{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}\{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}^\top\Big]\mathbf{T}_r^\top.
\end{aligned} \tag{S1.99}$$

Throughout this proof, K denotes a positive constant independent of n, p, t, and z, whose value may change from line to line. By the uniform boundedness of $\|\mathbf{T}_r\|$, we have $\|\mathbf{T}_r\mathbf{A}\mathbf{T}_r^\top\|_F \leq \|\mathbf{T}_r\|^2\|\mathbf{A}\|_F \leq K\|\mathbf{A}\|_F$ for every conformable matrix $\mathbf{A}$. Together with $\|\mathbf{A}\mathbf{B}^\top\|_F \leq \|\mathbf{A}\|\,\|\mathbf{B}\|_F$ and Cauchy's inequality, this gives

$$\begin{aligned}
\mathbb{E}\big\|\Delta^{(r,ji)}\big\|_F^2 &= \mathbb{E}\Big\|\frac{1}{n}\mathbf{T}_r\big[\mathbf{X}_i^{(r)}\{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}^\top + \{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}(\mathbf{X}_i^{(r)})^\top \\
&\qquad + \{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}\{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}^\top\big]\mathbf{T}_r^\top\Big\|_F^2 \\
&\leq \frac{K}{n^2}\mathbb{E}\Big[\big\|\mathbf{X}_i^{(r)}\{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}^\top\big\|_F^2 + \big\|\{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}(\mathbf{X}_i^{(r)})^\top\big\|_F^2
\end{aligned}$$

$$
\begin{aligned}
&\quad + \big\|\{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}\{\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\}^\top\big\|_F^2\Big] \\
&\le \frac{K}{n^2}\mathbb{E}\big[\|\mathbf{X}_i^{(r)}\|^2\|\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\|_F^2 + \|\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\|_F^4\big] \quad \text{(S1.100)} \\
&\le \frac{K}{n^2}\big\{\mathbb{E}\|\mathbf{X}_i^{(r)}\|^4\big\}^{1/2}\big\{\mathbb{E}\|\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\|_F^4\big\}^{1/2} + \frac{K}{n^2}\mathbb{E}\|\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\|_F^4.
\end{aligned}
$$

Moreover, since the column pairs are independent over t,

$$
\mathbb{E}\|\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\|_F^2 = \sum_{t=1}^n \mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^2 = n \cdot o\Big(\frac{p}{n}\Big) = o(p) = o(n), \quad \text{(S1.101)}
$$

where $p/n \to c \in (0, \infty)$. Also, combining (S1.97) and (S1.101),

$$
\begin{aligned}
\mathbb{E}\|\mathbf{X}_j^{(r)} - \mathbf{X}_i^{(r)}\|_F^4 &= \mathbb{E}\big[\sum_{t=1}^n \|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^2\big]^2 \\
&= \sum_{t=1}^n \mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^4 + 2\sum_{1\le s<t\le n} \mathbb{E}\|\boldsymbol{x}_{js}^{(r)} - \boldsymbol{x}_{is}^{(r)}\|^2\mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^2 \\
&\le Kn\big[\sup_{1\le t\le n} \mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^2\big]^2 + Kn^2\big[\sup_{1\le t\le n} \mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^2\big]^2 \\
&= O\Big[n^2\big\{\sup_{1\le t\le n} \mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\|^2\big\}^2\Big] = o(n^2). \qquad \text{(S1.102)}
\end{aligned}
$$

Under the standard spectral-norm bound for the sample matrix, we have $\mathbb{E}\|\mathbf{X}_i^{(r)}\|^4 = O(n^2)$. Substituting (S1.102) into (S1.100), we get

$$
\mathbb{E}\big\|\Delta^{(r,ji)}\big\|_F^2 \le \frac{K}{n^2}\big\{O(n^2)\big\}^{1/2}\big\{o(n^2)\big\}^{1/2} + \frac{K}{n^2}o(n^2) = o(1). \qquad \text{(S1.103)}
$$

Therefore, we have $\big\|\Delta^{(r,ji)}\big\|_F = o_p(1)$ and $\big\|\Delta^{(r,ji)}\big\| = o_p(1)$. Recall (S1.11), for $t \in [1:n]$, define

$$
\Delta_t^{(r,ji)} \coloneqq \boldsymbol{r}_t^{(r,j)}(\boldsymbol{r}_t^{(r,j)})^\top - \boldsymbol{r}_t^{(r,i)}(\boldsymbol{r}_t^{(r,i)})^\top. \qquad \text{(S1.104)}
$$

Then, recall the definition of $D_{j,t}^{(r)}(z)$ in (S1.12),

$$\begin{aligned}\Delta^{(r,ji)} - \Delta_t^{(r,ji)} &= \mathbf{S}_j^{(r)} - \boldsymbol{r}_t^{(r,j)}(\boldsymbol{r}_t^{(r,j)})^\top - \big\{\mathbf{S}_i^{(r)} - \boldsymbol{r}_t^{(r,i)}(\boldsymbol{r}_t^{(r,i)})^\top\big\} \\ &= \mathbf{D}_{j,t}^{(r)}(z) - \mathbf{D}_{i,t}^{(r)}(z). \end{aligned} \tag{S1.105}$$

The right-hand side does not depend on z. Repeating the argument leading to (S1.103), but with the tth column removed, gives

$$\sup_{1\le t\le n} \mathbb{E}\big\|\Delta^{(r,ji)} - \Delta_t^{(r,ji)}\big\|_F^2 = o(1). \tag{S1.106}$$

We now compare the two rank-one perturbation formula. From (S1.105), $\mathbf{D}_{j,t}^{(r)}(z) = \mathbf{D}_{i,t}^{(r)}(z) + \Delta^{(r,ji)} - \Delta_t^{(r,ji)}$. Therefore, by the resolvent identity,

$$\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} = \{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1} - \{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}(\Delta^{(r,ji)} - \Delta_t^{(r,ji)})\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}. \tag{S1.107}$$

Multiplying (S1.107) by $\mathbf{T}_r^\top$ and $\mathbf{T}_r$, and then taking conditional expectation $\mathbb{E}_t(\cdot)$, defined in (S1.14), yields

$$\begin{aligned}&\mathbb{E}_t\big[\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big] \\ &\quad= \mathbb{E}_t\big[\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big] - \mathbb{E}_t\big[\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}(\Delta^{(r,ji)} - \Delta_t^{(r,ji)})\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big].\end{aligned} \tag{S1.108}$$

On the contours used in the Cauchy integral representation, the resolvents are uniformly bounded, that is, $\big\|\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\big\| \le K$, and $\big\|\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\big\| \le K$. Therefore, by Jensen's inequality and the boundedness of $\|\mathbf{T}_r\|$, combining (S1.106),

$$
\begin{aligned}
&\mathbb{E}\Big\|\mathbb{E}_t\big[\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}(\Delta^{(r,ji)}-\Delta_t^{(r,ji)})\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big]\Big\|_F \\
&\quad\le \mathbb{E}\mathbb{E}_t\Big\|\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}(\Delta^{(r,ji)}-\Delta_t^{(r,ji)})\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\Big\|_F \\
&\quad\le K\mathbb{E}\mathbb{E}_t\big\|\Delta^{(r,ji)}-\Delta_t^{(r,ji)}\big\|_F = K\mathbb{E}\big\|\Delta^{(r,ji)}-\Delta_t^{(r,ji)}\big\|_F \\
&\quad\le K\Big\{\mathbb{E}\big\|\Delta^{(r,ji)}-\Delta_t^{(r,ji)}\big\|_F^2\Big\}^{1/2} = o(1),
\end{aligned}
\tag{S1.109}
$$

uniformly in t and z on the contours. Following the leading covariance representation (S1.31) already obtained in Section S1.2,

$$
\begin{aligned}
&\mathrm{Cov}\left(\Xi_{n,j}^{(r)}(z_1), \Xi_{n,i}^{(r)}(z_2)\right) \\
&\quad= \frac{\partial^2}{\partial z_1 \partial z_2}\sum_{t=1}^n \mathbb{E}\Big[b_t^{(r,j)}(z_1)b_t^{(r,i)}(z_2)\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\big\}\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\Big] + o(1),
\end{aligned}
\tag{S1.110}
$$

where $\eta_t^{(r,j)}(z_1)$ and $\eta_t^{(r,i)}(z_2)$ are given by (S1.19), and the deterministic coefficients $b_t^{(r,j)}(z_1)$ and $b_t^{(r,i)}(z_2)$ are defined as in (S1.18).

We shall show that, under the shared-component dominance condition, the first martingale increment $(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)$ in (S1.110) may be replaced by the corresponding i-sample martingale increment. By adding and subtracting $n^{-1}(\boldsymbol{x}_{jt}^{(r)})^\top\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{jt}^{(r)}$, we have

$$
\begin{aligned}
&\eta_t^{(r,j)}(z) - \eta_t^{(r,i)}(z) \\
&\quad= \frac{1}{n}\Big[(\boldsymbol{x}_{jt}^{(r)})^\top\mathbf{T}_r^\top\big[\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}-\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\big]\mathbf{T}_r\boldsymbol{x}_{jt}^{(r)} \\
&\qquad\qquad - \mathrm{tr}\Big\{\mathbf{T}_r^\top\big[\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}-\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\big]\mathbf{T}_r\Big\}\Big]
\end{aligned}
\tag{S1.111}
$$

$$+\frac{1}{n}\Big[(\boldsymbol{x}_{jt}^{(r)})^\top \mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{jt}^{(r)} - (\boldsymbol{x}_{it}^{(r)})^\top \mathbf{T}_r^\top \{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{it}^{(r)}\Big].$$

We control the two terms on the right-hand side of (S1.111) separately. For the first term, set

$$\mathbf{M}_t(z) = \mathbf{T}_r^\top\big[\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} - \{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\big]\mathbf{T}_r \coloneqq (a_{kl})_{p\times p}.$$

Conditional on the rank-one perturbation formula, $\mathbf{M}_t(z)$ is fixed and is independent of the current vector $\boldsymbol{x}_{jt}^{(r)} \coloneqq (x_{jkt}^{(r)}) \in \mathbb{R}^p$. Recall that, for entries from the rth row-population,

$$\mathbb{E}\big[(x_{jkt}^{(r)}x_{jlt}^{(r)} - \delta_{kl})(x_{jat}^{(r)}x_{jbt}^{(r)} - \delta_{ab})\big] = \delta_{ka}\delta_{lb} + q\delta_{kb}\delta_{la} + \beta^{(r)}\mathbf{1}_{\{k=l=a=b\}},$$

where Kronecker delta function $\delta_{ab} = \mathbf{1}_{\{a=b\}}$, $q = 1$ in the real case and $q = 0$ in the complex case, and $\beta^{(r)} = \mathbb{E}\{(x_{jkt}^{(r)})^4\} - q - 2$ is common over k and t within the rth row-population. Therefore,

$$\begin{aligned}
&\mathbb{E}\Big[\big|(\boldsymbol{x}_{jt}^{(r)})^\top\mathbf{M}_t(z)\boldsymbol{x}_{jt}^{(r)} - \operatorname{tr}\mathbf{M}_t(z)\big|^2 \,\Big|\, \mathbf{M}_t(z)\Big] = \mathbb{E}\Big[\big|\sum_{k,l=1}^p a_{kl}(x_{jkt}^{(r)}x_{jlt}^{(r)} - \delta_{kl})\big|^2 \,\Big|\, \mathbf{M}_t(z)\Big]\\
&\quad= \sum_{k,l,a,b=1}^p a_{kl}\overline{a}_{ab}\,\mathbb{E}\big[(x_{jkt}^{(r)}x_{jlt}^{(r)} - \delta_{kl})(x_{jat}^{(r)}x_{jbt}^{(r)} - \delta_{ab})\big]\\
&\quad= \sum_{k,l=1}^p |a_{kl}|^2 + q\sum_{k,l=1}^p a_{kl}\overline{a}_{lk} + \beta^{(r)}\sum_{k=1}^p |a_{kk}|^2\\
&\quad\le \sum_{k,l=1}^p |a_{kl}|^2 + q\big|\sum_{k,l=1}^p a_{kl}\overline{a}_{lk}\big| + |\beta^{(r)}|\sum_{k=1}^p |a_{kk}|^2 \qquad\qquad (\text{S1.112})\\
&\quad\le \|\mathbf{M}_t(z)\|_F^2 + q\big(\sum_{k,l=1}^p |a_{kl}|^2\big)^{1/2}\big(\sum_{k,l=1}^p |a_{lk}|^2\big)^{1/2} + K\|\mathbf{M}_t(z)\|_F^2 \le K\|\mathbf{M}_t(z)\|_F^2.
\end{aligned}$$

where $\overline{a}_{lk}$ denotes the complex conjugate of a_{lk}. Hence, combining (S1.112) and (S1.106), the first term of (S1.111) can be controlled,

$$
\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\left[(\boldsymbol{x}_{jt}^{(r)})^\top \mathbf{M}_t(z)\boldsymbol{x}_{jt}^{(r)} - \operatorname{tr}\mathbf{M}_t(z)\right]\right|^2 \\
&\quad \le \frac{K}{n^2}\mathbb{E}\|\mathbf{M}_t(z)\|_F^2 \\
&\quad = \frac{K}{n^2}\mathbb{E}\left\|\mathbf{T}_r^\top\left\{\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1} - \{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\right\}\mathbf{T}_r\right\|_F^2 \\
&\quad = \frac{K}{n^2}\mathbb{E}\left\|\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}(\Delta^{(r,ji)} - \Delta_t^{(r,ji)})\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\right\|_F^2 \\
&\quad \le \frac{K}{n^2}\mathbb{E}\left[\|\mathbf{T}_r\|^4\left\|\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\right\|^2\left\|\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\right\|^2\left\|\Delta^{(r,ji)} - \Delta_t^{(r,ji)}\right\|_F^2\right] \\
&\quad \le \frac{K}{n^2}\mathbb{E}\left\|\Delta^{(r,ji)} - \Delta_t^{(r,ji)}\right\|_F^2 = o(n^{-2}). \qquad \text{(S1.113)}
\end{aligned}
$$

Second, we control the remaining quadratic-form difference. Using the identity $\mathbf{a}^\top\mathbf{A}\mathbf{a} - \mathbf{b}^\top\mathbf{A}\mathbf{b} = (\mathbf{a}-\mathbf{b})^\top\mathbf{A}(\mathbf{a}+\mathbf{b})$, with $\mathbf{a} = \boldsymbol{x}_{jt}^{(r)}$, $\mathbf{b} = \boldsymbol{x}_{it}^{(r)}$, and $\mathbf{A} = \mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r$, together with Cauchy's inequality, the shared-component dominance condition (S1.97), and the marginal moment bound $\mathbb{E}\|\boldsymbol{x}_{jt}^{(r)} + \boldsymbol{x}_{it}^{(r)}\|^4 = O(p^2)$, we obtain

$$
\begin{aligned}
&\mathbb{E}\left|\frac{1}{n}\left[(\boldsymbol{x}_{jt}^{(r)})^\top\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{jt}^{(r)} - (\boldsymbol{x}_{it}^{(r)})^\top\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{it}^{(r)}\right]\right|^2 \\
&\quad = \frac{1}{n^2}\mathbb{E}\left|(\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)})^\top\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r(\boldsymbol{x}_{jt}^{(r)} + \boldsymbol{x}_{it}^{(r)})\right|^2 \\
&\quad \le \frac{K}{n^2}\mathbb{E}\left[\left\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\right\|^2\left\|\boldsymbol{x}_{jt}^{(r)} + \boldsymbol{x}_{it}^{(r)}\right\|^2\right] \\
&\quad \le \frac{K}{n^2}\left\{\mathbb{E}\left\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\right\|^4\right\}^{1/2}\left\{\mathbb{E}\left\|\boldsymbol{x}_{jt}^{(r)} + \boldsymbol{x}_{it}^{(r)}\right\|^4\right\}^{1/2} \\
&\quad = \frac{K}{n^2}O\left[\mathbb{E}\left\|\boldsymbol{x}_{jt}^{(r)} - \boldsymbol{x}_{it}^{(r)}\right\|^2\right]O(p) = \frac{K}{n^2}\,o\left(\frac{p}{n}\right)O(p) = o(n^{-1}). \qquad \text{(S1.114)}
\end{aligned}
$$

Combining (S1.111), (S1.113), and (S1.114), we obtain

$$
\begin{aligned}
&\mathbb{E}\Big|(\mathbb{E}_t-\mathbb{E}_{t-1})\big\{\eta_t^{(r,j)}(z)-\eta_t^{(r,i)}(z)\big\}\Big|^2\\
&\quad\le K\mathbb{E}\big|\eta_t^{(r,j)}(z)-\eta_t^{(r,i)}(z)\big|^2\\
&\quad\le K\mathbb{E}\bigg|\frac{1}{n}\Big[(\boldsymbol{x}_{jt}^{(r)})^\top\mathbf{T}_r^\top\big[\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}-\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\big]\mathbf{T}_r\boldsymbol{x}_{jt}^{(r)}\\
&\qquad\qquad-\operatorname{tr}\Big\{\mathbf{T}_r^\top\big[\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}-\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\big]\mathbf{T}_r\Big\}\Big]\bigg|^2\\
&\qquad+K\mathbb{E}\bigg|\frac{1}{n}\Big[(\boldsymbol{x}_{jt}^{(r)})^\top\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{jt}^{(r)}-(\boldsymbol{x}_{it}^{(r)})^\top\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{it}^{(r)}\Big]\bigg|^2\\
&\quad=o(n^{-2})+o(n^{-1})=o(n^{-1}),
\end{aligned}
\tag{S1.115}
$$

uniformly in t and z on the contours. Next, we compare $b_t^{(r,j)}(z)$ and $b_t^{(r,i)}(z)$. Recall (S1.18) and the same definition holds for $b_t^{(r,i)}(z)$. Since the denominators are bounded away from zero on the contours, we have

$$
\begin{aligned}
&\big|b_t^{(r,j)}(z)-b_t^{(r,i)}(z)\big|\\
&=\Big|b_t^{(r,j)}(z)b_t^{(r,i)}(z)\frac{1}{n}\mathbb{E}\operatorname{tr}\Big[\mathbf{T}_r^\top\big[\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}-\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\big]\mathbf{T}_r\Big]\Big|\\
&\le\frac{K}{n}\mathbb{E}\Big|\operatorname{tr}\Big[\mathbf{T}_r^\top\big[\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}-\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\big]\mathbf{T}_r\Big]\Big|\\
&=\frac{K}{n}\mathbb{E}\Big|\operatorname{tr}\big[\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}(\Delta^{(r,ji)}-\Delta_t^{(r,ji)})\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big]\Big|\\
&=\frac{K}{n}\mathbb{E}\Big|\operatorname{tr}\big[\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}(\Delta^{(r,ji)}-\Delta_t^{(r,ji)})\big]\Big|\\
&\le\frac{K}{n}\mathbb{E}\Big[\big\|\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\big\|_F\big\|\Delta^{(r,ji)}-\Delta_t^{(r,ji)}\big\|_F\Big]\\
&\le K\frac{\sqrt{p}}{n}\Big\{\mathbb{E}\big\|\Delta^{(r,ji)}-\Delta_t^{(r,ji)}\big\|_F^2\Big\}^{1/2}=o(1),
\end{aligned}
\tag{S1.116}
$$

uniformly in t and z on the contours. Here the last inequality uses $|\operatorname{tr}(\mathbf{AB})|\le$

$\|\mathbf{A}\|_F\|\mathbf{B}\|_F$, and $\|\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\mathbf{T}_r^\top\{\mathbf{D}_{i,t}^{(r)}(z)\}^{-1}\|_F = O(\sqrt{p})$.

We now substitute these two replacements into (S1.110), and it remains to control the two replacement errors. Indeed, we have

$$
\begin{aligned}
\mathbb{E}\big|(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,j)}(z)\big|^2 &\le 2\mathbb{E}\big|\mathbb{E}_t\eta_t^{(r,j)}(z)\big|^2+2\mathbb{E}\big|\mathbb{E}_{t-1}\eta_t^{(r,j)}(z)\big|^2\le 4\mathbb{E}\big|\eta_t^{(r,j)}(z)\big|^2\\
&= 4\mathbb{E}\Big|\frac{1}{n}\Big[(\boldsymbol{x}_{jt}^{(r)})^\top\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\boldsymbol{x}_{jt}^{(r)}-\operatorname{tr}\big\{\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big\}\Big]\Big|^2\\
&\le \frac{K}{n^2}\mathbb{E}\big\|\mathbf{T}_r^\top\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\mathbf{T}_r\big\|_F^2\le\frac{K}{n^2}\mathbb{E}\Big[\|\mathbf{T}_r\|^4\big\|\{\mathbf{D}_{j,t}^{(r)}(z)\}^{-1}\big\|_F^2\Big]\\
&\le \frac{Kp}{n^2}=O(n^{-1}), \qquad\qquad (\text{S1.117})
\end{aligned}
$$

where the last equality follows from $p/n\to c\in(0,\infty)$. Since $b_t^{(r,j)}(z_1)$ and $b_t^{(r,i)}(z_1)$ are deterministic coefficients, (S1.116) gives

$$
\sup_{1\le t\le n}\Big|b_t^{(r,j)}(z_1)-b_t^{(r,i)}(z_1)\Big|=o(1).
$$

Using the boundedness of $b_t^{(r,i)}(z_2)$, Cauchy's inequality, and (S1.117) together with its analogue with j replaced by i, we obtain

$$
\begin{aligned}
&\Bigg|\sum_{t=1}^n\mathbb{E}\Big[\big\{b_t^{(r,j)}(z_1)-b_t^{(r,i)}(z_1)\big\}b_t^{(r,i)}(z_2)\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\big\}\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\Big]\Bigg|\\
&\quad\le K\sup_{1\le t\le n}\big|b_t^{(r,j)}(z_1)-b_t^{(r,i)}(z_1)\big|\sum_{t=1}^n\mathbb{E}\Big|\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\big\}\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\Big|\\
&\quad\le K\sup_{1\le t\le n}\big|b_t^{(r,j)}(z_1)-b_t^{(r,i)}(z_1)\big|\Big\{\sum_{t=1}^n\mathbb{E}\big|(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,j)}(z_1)\big|^2\Big\}^{1/2}\\
&\qquad\qquad\times\Big\{\sum_{t=1}^n\mathbb{E}\big|(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big|^2\Big\}^{1/2}\\
&\quad=o(1)\cdot O(1)\cdot O(1)=o(1). \qquad\qquad (\text{S1.118})
\end{aligned}
$$

Then, by the boundedness of $b_t^{(r,i)}(z_1)$ and $b_t^{(r,i)}(z_2)$, Cauchy's inequality, (S1.115), and the analogue of (S1.117) with j replaced by i, we have

$$\begin{aligned}
&\Big|\sum_{t=1}^{n}\mathbb{E}\Big[b_t^{(r,i)}(z_1)b_t^{(r,i)}(z_2)\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})[\eta_t^{(r,j)}(z_1)-\eta_t^{(r,i)}(z_1)]\big\}\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\Big]\Big| \\
&\quad\le K\Big\{\sum_{t=1}^{n}\mathbb{E}\big|(\mathbb{E}_t-\mathbb{E}_{t-1})[\eta_t^{(r,j)}(z_1)-\eta_t^{(r,i)}(z_1)]\big|^2\Big\}^{1/2} \\
&\qquad\quad\times\Big\{\sum_{t=1}^{n}\mathbb{E}\Big|(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\Big|^2\Big\}^{1/2} \\
&\quad= o(1)\cdot O(1)=o(1). \qquad\qquad (\text{S1.119})
\end{aligned}$$

Combining (S1.110), (S1.118), and (S1.119), we obtain

$$\begin{aligned}
&\operatorname{Cov}\big(\Xi_{n,j}^{(r)}(z_1),\Xi_{n,i}^{(r)}(z_2)\big) \\
&\quad=\frac{\partial^2}{\partial z_1\partial z_2}\sum_{t=1}^{n}\mathbb{E}\Big[b_t^{(r,i)}(z_1)b_t^{(r,i)}(z_2)\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,i)}(z_1)\big\}\big\{(\mathbb{E}_t-\mathbb{E}_{t-1})\eta_t^{(r,i)}(z_2)\big\}\Big]+o(1) \\
&\quad=\operatorname{Cov}\big(\Xi_{n,i}^{(r)}(z_1),\Xi_{n,i}^{(r)}(z_2)\big)+o(1), \qquad\qquad (\text{S1.120})
\end{aligned}$$

uniformly for z_1, z_2 on the contours. The same conclusion remains valid after differentiating with respect to z_1 and z_2, since the derivatives only replace resolvent factors by higher powers of resolvents, which are still uniformly bounded on the contours.

Finally, using the Cauchy integral representation in (S1.96) and its analogue for the ith row sample, and integrating (S1.120) over the two contours, we obtain

$$\mathrm{Cov}\left(\{\mathcal{Q}_n^{(r)}(f)\}_j, \{\mathcal{Q}_n^{(r)}(f)\}_i\right) = \mathrm{Var}\left(\{\mathcal{Q}_n^{(r)}(f)\}_i\right) + o(1).$$

Hence, under the shared-component dominance condition,

$$\nu^{(r)}_{\mathrm{Cov}\,fji} = \nu^{(r)}_f + o(1).$$

The proof is completed.

S2 Detailed Calculations of Example 1

Let $f_0(x) = x$. We define the first two moments of the population spectral distributions $H^{(r)}$ as

$$m_1^{(r)} = \int x \,\mathrm{d}H^{(r)}(x), \quad m_2^{(r)} = \int x^2 \,\mathrm{d}H^{(r)}(x), \text{ for } r = 1, 2, 3.$$

In the fourth-cumulant term, the coefficient multiplying $\beta^{(r)}$ or $\beta^{(r)}_\rho$ depends on the setting under consideration. We denote this coefficient by $\kappa_x^{(r)}$, where $\kappa_x^{(r)} = \{m_1^{(r)}\}^2$ under Assumption 4, and $\kappa_x^{(r)} = m_2^{(r)}$ in the diagonal case.

Recall the three population spectral distributions introduced in Example 1: $H^{(1)} = \delta_{\{1\}}$, $H^{(2)} = \delta_{\{2\}}/2 + \delta_{\{1\}}/2$, and $H^{(3)} = \delta_{\{3\}}/3 + \delta_{\{2\}}/3 + \delta_{\{1\}}/3$. Direct integration yields their respective moments as

$$\left(m_1^{(1)}, m_1^{(2)}, m_1^{(3)}\right) = (1, 3/2, 2), \qquad \left(m_2^{(1)}, m_2^{(2)}, m_2^{(3)}\right) = (1, 5/2, 14/3).$$

Consequently, $\kappa_x^{(1)} = 1$ in both settings. For the second and third populations, Assumption 4 gives $\kappa_x^{(2)} = 9/4$ and $\kappa_x^{(3)} = 4$, whereas in the diagonal case one obtains $\kappa_x^{(2)} = 5/2$ and $\kappa_x^{(3)} = 14/3$. Note that $\int x \, \mathrm{d}F^{c,H^{(r)}}(x) = m_1^{(r)}$, meaning the mean correction identically vanishes, i.e., $\mu_x^{(r)} = 0$ for all r. For the jth sample drawn from the rth population, we have

$$\{\boldsymbol{\mathcal{Q}}_n^{(r)}(f_0)\}_j = \sum_{i=1}^{p} \lambda_i(\mathbf{S}_j^{(r)}) - m_1^{(r)} p.$$

We first list the variance $\nu_x^{(r)}$ and the Case II cross-covariance $\nu_{x,\rho}^{(r)}$ for each population. The calculation for $H^{(1)}$ follows Example 1 in Jiang (2023), and that for $H^{(2)}$ follows Example 1 in Wang and Jiang (2025). The detailed calculation for $H^{(3)}$ is provided in Section S3.

$$H^{(1)}: \quad \nu_x^{(1)} = c\{(q+1) + \beta^{(1)}\}, \qquad \nu_{x,\rho}^{(1)} = c\{(q+1)\rho_1^2 + \beta_\rho^{(1)}\}.$$

$$H^{(2)}: \quad \nu_x^{(2)} = c\big\{(5/2)(q+1) + \beta^{(2)}\kappa_x^{(2)}\big\}, \quad \nu_{x,\rho}^{(2)} = c\big\{(5/2)(q+1)\rho_2^2 + \beta_\rho^{(2)}\kappa_x^{(2)}\big\}.$$

$$H^{(3)}: \quad \nu_x^{(3)} = c\big\{(14/3)(q+1) + \beta^{(3)}\kappa_x^{(3)}\big\}, \nu_{x,\rho}^{(3)} = c\big\{(14/3)(q+1)\rho_3^2 + \beta_\rho^{(3)}\kappa_x^{(3)}\big\}.$$

We now evaluate the limiting variance ν_x of $\boldsymbol{\mathcal{G}}_n(f_0)$ under the three row-population configurations in Example 1. By the independence of row samples from different row-populations in Case I, the only nonzero cross-covariance contribution comes from the two dependent row samples within the same row-population.

Scenario 1: $(h_1, h_2, h_3) = (2, 1, 1)$. The centering term is $b_x = p\{2m_1^{(1)} +$

$m_1^{(2)} + m_1^{(3)}\} = (11/2)p$. $\boldsymbol{\mathcal{G}}_n(f_0)$ admits the decomposition

$$\boldsymbol{\mathcal{G}}_n(f_0) = \sum_{j=1}^{2} \{\boldsymbol{\mathcal{Q}}_n^{(1)}(f_0)\}_j + \{\boldsymbol{\mathcal{Q}}_n^{(2)}(f_0)\}_1 + \{\boldsymbol{\mathcal{Q}}_n^{(3)}(f_0)\}_1.$$

Accounting for the covariance within the first population, we sum the components to obtain

$$\nu_x = 2\nu_x^{(1)} + \nu_x^{(2)} + \nu_x^{(3)} + 2\,\mathrm{Cov}\big(\{\boldsymbol{\mathcal{Q}}_n^{(1)}(f_0)\}_1, \{\boldsymbol{\mathcal{Q}}_n^{(1)}(f_0)\}_2\big).$$

Substituting the respective values yields the explicit variances under Assumption 4 and the diagonal case:

$$\nu_{x,\mathcal{S}_1}^{\mathrm{A4}} = c\Big\{(q+1)(55/6 + 2\rho_1^2) + 2\beta^{(1)} + (9/4)\beta^{(2)} + 4\beta^{(3)} + 2\beta_\rho^{(1)}\Big\},$$
$$\nu_{x,\mathcal{S}_1}^{\mathrm{diag}} = c\Big\{(q+1)(55/6 + 2\rho_1^2) + 2\beta^{(1)} + (5/2)\beta^{(2)} + (14/3)\beta^{(3)} + 2\beta_\rho^{(1)}\Big\}.$$

The same covariance decomposition gives the following two cases.

Scenario 2: $(h_1, h_2, h_3) = (1, 2, 1)$. The centering term is $b_x = p\{m_1^{(1)} + 2m_1^{(2)} + m_1^{(3)}\} = 6p$ and summing $\nu_x^{(1)} + 2\nu_x^{(2)} + \nu_x^{(3)} + 2\nu_{x,\rho}^{(2)}$ yields

$$\nu_{x,\mathcal{S}_2}^{\mathrm{A4}} = c\Big\{(q+1)(32/3 + 5\rho_2^2) + \beta^{(1)} + (9/2)\beta^{(2)} + 4\beta^{(3)} + (9/2)\beta_\rho^{(2)}\Big\},$$
$$\nu_{x,\mathcal{S}_2}^{\mathrm{diag}} = c\Big\{(q+1)(32/3 + 5\rho_2^2) + \beta^{(1)} + 5\beta^{(2)} + (14/3)\beta^{(3)} + 5\beta_\rho^{(2)}\Big\}.$$

Scenario 3: $(h_1, h_2, h_3) = (1, 1, 2)$. The centering term is $b_x = p\{m_1^{(1)} + m_1^{(2)} + 2m_1^{(3)}\} = (13/2)p$ and summing $\nu_x^{(1)} + \nu_x^{(2)} + 2\nu_x^{(3)} + 2\nu_{x,\rho}^{(3)}$ provides

$$\nu_{x,\mathcal{S}_3}^{\mathrm{A4}} = c\Big\{(q+1)\big(77/6 + (28/3)\rho_3^2\big) + \beta^{(1)} + (9/4)\beta^{(2)} + 8\beta^{(3)} + 8\beta_\rho^{(3)}\Big\},$$
$$\nu_{x,\mathcal{S}_3}^{\mathrm{diag}} = c\Big\{(q+1)\big(77/6 + (28/3)\rho_3^2\big) + \beta^{(1)} + (5/2)\beta^{(2)} + (28/3)\beta^{(3)} + (28/3)\beta_\rho^{(3)}\Big\}.$$

Consequently, the standardized statistic is

$$T_x = \nu_x^{-1/2}\Big[\sum_{i=1}^{p} h\lambda_i - b_x\Big] \Rightarrow \mathcal{N}(0,1),$$

where the mean correction is $\mu_x = 0$, and ν_x is taken from the expression corresponding to the specified row-population configuration and setting.

S3 Calculations of the Population $H^{(3)}$ in Example 1

S3.1 Calculations of $\mu_x^{(3)}$

We primarily derive the computation procedure for $\mu_x^{(3)}$ and $\nu_x^{(3)}$. It is straightforward to ascertain that the mean $\tilde{\mu}_x$ when $H(x) = \delta_{\{3\sigma^2\}}/3 + \delta_{\{2\sigma^2\}}/3 + \delta_{\{\sigma^2\}}/3$ is identical to σ^2 multiplied by the mean $\mu_x^{(3)}$ when $H^{(3)}(x) = \delta_{\{3\}}/3 + \delta_{\{2\}}/3 + \delta_{\{1\}}/3$. Consequently, we initially concentrate on the computation of $\mu_x^{(3)}$.

Let $\underline{s}(z)$ be represented as $\underline{s}$ for the sake of simplicity. Refer to (3.1) as presented in Bai and Yao (2012a) as $H^{(3)} = \delta_{\{3\}}/3 + \delta_{\{2\}}/3 + \delta_{\{1\}}/3$ for $z \in \mathbb{C}^+$,

$$z = -\frac{1}{\underline{s}} + \int \frac{cx}{1 + x\underline{s}} \mathrm{d}H^{(3)}(x) = -\frac{1}{\underline{s}} + \frac{c}{3\{1+\underline{s}\}} + \frac{2c}{3\{1+2\underline{s}\}} + \frac{c}{\{1+3\underline{s}\}}.$$

The first part of $\mu_x^{(3)}$ is

$$
\begin{aligned}
\mu_{x_1}^{(3)} &= -\frac{q}{2\pi i}\oint \frac{cz\underline{s}^3\int x^2\{1+x\underline{s}\}^{-3}\mathrm{d}H(x)}{[1-c\underline{s}^2\int x^2\{1+x\underline{s}\}^{-2}\mathrm{d}H(x)]^2}\mathrm{d}z \\
&= -\frac{q}{2\pi i}\oint \Big[-\frac{1}{\underline{s}}+\frac{c}{3\{1+\underline{s}\}}+\frac{2c}{3\{1+2\underline{s}\}}+\frac{c}{\{1+3\underline{s}\}}\Big] \\
&\quad\times \frac{c\underline{s}^3\big[\{1+\underline{s}\}^{-3}+4\{1+2\underline{s}\}^{-3}+9\{1+3\underline{s}\}^{-3}\big]}{3-c\underline{s}^2\big[\{1+\underline{s}\}^{-2}+4\{1+2\underline{s}\}^{-2}+9\{1+3\underline{s}\}^{-2}\big]} \\
&\quad\times \Big[\frac{1}{\underline{s}^2}-\frac{c}{3\{1+\underline{s}\}^2}-\frac{4c}{3\{1+2\underline{s}\}^2}-\frac{3c}{\{1+3\underline{s}\}^2}\Big] \\
&= -\frac{q}{2\pi i}\oint \frac{c^2\underline{s}N_1[\underline{s}]N_2[\underline{s}]}{3\{1+\underline{s}\}^2\{1+2\underline{s}\}^2\{1+3\underline{s}\}^2 D[\underline{s}]}\mathrm{d}\underline{s},
\end{aligned}
$$

where

$$
\begin{aligned}
N_1[\underline{s}] =& \big[\{1+2\underline{s}\}\{1+3\underline{s}\}+2\{1+\underline{s}\}\{1+3\underline{s}\}+3\{1+\underline{s}\}\{1+2\underline{s}\}\big] \\
&-3\{1+\underline{s}\}\{1+2\underline{s}\}\{1+3\underline{s}\}, \\
N_2[\underline{s}] =& \{1+2\underline{s}\}^3\{1+3\underline{s}\}^3+4\{1+\underline{s}\}^3\{1+3\underline{s}\}^3+9\{1+\underline{s}\}^3\{1+2\underline{s}\}^3, \\
D[\underline{s}] =& 3\{1+\underline{s}\}^2\{1+2\underline{s}\}^2\{1+3\underline{s}\}^2 \\
&-c\underline{s}^2\big[\{1+2\underline{s}\}^2\{1+3\underline{s}\}^2+4\{1+\underline{s}\}^2\{1+3\underline{s}\}^2+9\{1+\underline{s}\}^2\{1+2\underline{s}\}^2\big].
\end{aligned}
$$

The contour of $\underline{s}$ encloses the points -1, $-1/2$, and $-1/3$ for $0 < c \leq 1$, but it encloses no poles when $c > 1$. Thus $\mu_{x_1}^{(3)}$ in both cases can be calculated as

$$
\mu_{x_1}^{(3)} = -\frac{q}{2\pi i}\times 0 = 0.
$$

Similarly, the second part of $\mu_x^{(3)}$ can be computed as

$$
\begin{aligned}
\mu_{x_2}^{(3)} &= -\frac{\beta c}{2\pi i}\oint \frac{z\underline{s}^3\int x\{1+x\underline{s}\}^{-1}\mathrm{d}H(x)\int\{1+x\underline{s}\}^{-2}\mathrm{d}H(x)}{1-c\underline{s}^2\int x^2\{1+x\underline{s}\}^{-2}\mathrm{d}H(x)}\mathrm{d}z \\
&= -\frac{\beta c}{2\pi i}\oint\Big[-\frac{1}{\underline{s}}+\frac{c}{3\{1+\underline{s}\}}+\frac{2c}{3\{1+2\underline{s}\}}+\frac{c}{\{1+3\underline{s}\}}\Big] \\
&\quad\times\frac{\underline{s}^3\big[\{1+\underline{s}\}^{-1}+2\{1+2\underline{s}\}^{-1}+3\{1+3\underline{s}\}^{-1}\big]}{9-3c\underline{s}^2\big[\{1+\underline{s}\}^{-2}+4\{1+2\underline{s}\}^{-2}+9\{1+3\underline{s}\}^{-2}\big]} \\
&\quad\times\Big[\frac{1}{\{1+\underline{s}\}^2}+\frac{1}{\{1+2\underline{s}\}^2}+\frac{1}{\{1+3\underline{s}\}^2}\Big] \\
&\quad\times\Big[\frac{1}{\underline{s}^2}-\frac{c}{3\{1+\underline{s}\}^2}-\frac{4c}{3\{1+2\underline{s}\}^2}-\frac{3c}{\{1+3\underline{s}\}^2}\Big] \\
&= -\frac{\beta c}{2\pi i}\times 2\pi i\times 0 \\
&= 0.
\end{aligned}
$$

Thus,

$$\tilde{\mu}_x = \mu_x^{(3)} = 0.$$

S3.2 Calculations of $\nu_x^{(3)}$

Now, consider the calculation of $\tilde{\nu}_x$. When $H(x) = \delta_{\{3\sigma^2\}}/3 + \delta_{\{2\sigma^2\}}/3 + \delta_{\{\sigma^2\}}/3$, for $z \in \mathbb{C}^+$, we have

$$z = -\frac{1}{\underline{s}} + \int \frac{cx}{1+x\underline{s}}\mathrm{d}H(x) = -\frac{1}{\underline{s}} + \frac{c\sigma^2}{3\{1+\sigma^2\underline{s}\}} + \frac{2c\sigma^2}{3\{1+2\sigma^2\underline{s}\}} + \frac{c\sigma^2}{1+3\sigma^2\underline{s}}.$$

The first part of $\tilde{\nu}_x$ is

$$\tilde{\nu}_{x_1} = -\frac{q+1}{4\pi^2}\oint\oint \frac{z_1 z_2}{\big\{\underline{s}(z_1-\underline{s}(z_2)\big\}^2}\mathrm{d}\underline{s}(z_1)\mathrm{d}\underline{s}(z_2).$$

Let $\underline{s}_i$ represent $\underline{s}(z_i)$ for $i = 1, 2$ for the sake of simplicity. For a fixed $\underline{s}_2$, the contour of $\underline{s}_1$ encircles $-1/\sigma^2$, $-1/(2\sigma^2)$, and $-1/(3\sigma^2)$, whereas it encircles 0 if $c > 1$. Thus, we have

$$
\begin{aligned}
&\oint \frac{z_1}{\{\underline{s}_1 - \underline{s}_2\}^2} \mathrm{d}\underline{s}_1 \\
&= \oint -\frac{1}{\underline{s}_1\{\underline{s}_1 - \underline{s}_2\}^2} + \frac{c\sigma^2}{3\{1+\sigma^2\underline{s}_1\}\{\underline{s}_1 - \underline{s}_2\}^2} + \frac{2c\sigma^2}{3\{1+2\sigma^2\underline{s}_1\}\{\underline{s}_1 - \underline{s}_2\}^2} \\
&\qquad + \frac{c\sigma^2}{\{1+3\sigma^2\underline{s}_1\}\{\underline{s}_1 - \underline{s}_2\}^2} \mathrm{d}\underline{s}_1 \\
&= \begin{cases} 2\pi i\Big[\dfrac{c\sigma^4}{3\{1+\sigma^2\underline{s}_2\}^2} + \dfrac{4c\sigma^4}{3\{1+2\sigma^2\underline{s}_2\}^2} + \dfrac{3c\sigma^4}{\{1+3\sigma^2\underline{s}_2\}^2}\Big], & \text{for } 0 < c \le 1, \\ -2\pi i \dfrac{1}{\underline{s}_2^2}, & \text{for } 1 < c. \end{cases}
\end{aligned}
$$

When $0 < c \le 1$, we have

$$
\begin{aligned}
\tilde{\nu}_{x_1} &= -\frac{(q+1)2\pi i}{4\pi^2} \oint \Big[\frac{c\sigma^4}{3\{1+\sigma^2\underline{s}_2\}^2} + \frac{4c\sigma^4}{3\{1+2\sigma^2\underline{s}_2\}^2} + \frac{3c\sigma^4}{\{1+3\sigma^2\underline{s}_2\}^2}\Big] \\
&\qquad\qquad \times \Big[-\frac{1}{\underline{s}_2} + \frac{c\sigma^2}{3\{1+\sigma^2\underline{s}_2\}} + \frac{2c\sigma^2}{3\{1+2\sigma^2\underline{s}_2\}} + \frac{c\sigma^2}{1+3\sigma^2\underline{s}_2}\Big] \mathrm{d}\underline{s}_2 \\
&= -\frac{(q+1)2\pi i}{4\pi^2} \times 2\pi i \times \Big(\frac{c\sigma^4}{3} + \frac{4c\sigma^4}{3} + 3c\sigma^4\Big) \\
&= \frac{14}{3}(q+1)c\sigma^4.
\end{aligned}
$$

On the other hand, analogous computations are performed for the scenario when $c > 1$ with the pole at 0. The value of $\tilde{\nu}_{x_1}$ can be determined as

$$
\begin{aligned}
\tilde{\nu}_{x_1} &= \frac{(q+1)2\pi i}{4\pi^2} \oint \frac{1}{\underline{s}_2^2}\Big[-\frac{1}{\underline{s}_2} + \frac{c\sigma^2}{3\{1+\sigma^2\underline{s}_2\}} + \frac{2c\sigma^2}{3\{1+2\sigma^2\underline{s}_2\}} + \frac{c\sigma^2}{1+3\sigma^2\underline{s}_2}\Big] \mathrm{d}\underline{s}_2 \\
&= \frac{(q+1)2\pi i}{4\pi^2} \times 2\pi i \times \Big(-\frac{14}{3}c\sigma^4\Big) \\
&= \frac{14}{3}(q+1)c\sigma^4.
\end{aligned}
$$

The results are consistent across all instances of c. Under Assumption 4, and the second part of $\tilde{\nu}_x$ is

$$\tilde{\nu}_{x_2,\mathrm{A4}} = -\frac{\beta c}{4\pi^2}\oint\oint z_1 z_2\Big[\frac{\sigma^2}{3\{1+\sigma^2\underline{s}_1\}^2}+\frac{2\sigma^2}{3\{1+2\sigma^2\underline{s}_1\}^2}+\frac{\sigma^2}{\{1+3\sigma^2\underline{s}_1\}^2}\Big]$$
$$\times\Big[\frac{\sigma^2}{3\{1+\sigma^2\underline{s}_2\}^2}+\frac{2\sigma^2}{3\{1+2\sigma^2\underline{s}_2\}^2}+\frac{\sigma^2}{\{1+3\sigma^2\underline{s}_2\}^2}\Big]\mathrm{d}\underline{s}_1\mathrm{d}\underline{s}_2.$$

The contour encompasses residues of $-1/\sigma^2$, $-1/(2\sigma^2)$, and $-1/(3\sigma^2)$ when $0 < c \leq 1$, whereas it encloses 0 in the alternative scenario where $c > 1$. Simple calculations for all scenarios reveal that the conclusions are identical, albeit with different poles,

$$\oint z_1\Big[\frac{\sigma^2}{3\{1+\sigma^2\underline{s}_1\}^2}+\frac{2\sigma^2}{3\{1+2\sigma^2\underline{s}_1\}^2}+\frac{\sigma^2}{\{1+3\sigma^2\underline{s}_1\}^2}\Big]\mathrm{d}\underline{s}_1$$
$$=\oint\Big[-\frac{1}{\underline{s}_1}+\frac{c\sigma^2}{3\{1+\sigma^2\underline{s}_1\}}+\frac{2c\sigma^2}{3\{1+2\sigma^2\underline{s}_1\}}+\frac{c\sigma^2}{1+3\sigma^2\underline{s}_1}\Big]$$
$$\times\Big[\frac{\sigma^2}{3\{1+\sigma^2\underline{s}_1\}^2}+\frac{2\sigma^2}{3\{1+2\sigma^2\underline{s}_1\}^2}+\frac{\sigma^2}{\{1+3\sigma^2\underline{s}_1\}^2}\Big]\mathrm{d}\underline{s}_1$$
$$= 2\pi i\times 2\sigma^2$$
$$= -4\pi i\sigma^2.$$

Therefore, $\tilde{\nu}_{x_2}$ can be calculated as

$$\tilde{\nu}_{x_2,\mathrm{A4}} = -\frac{\beta c}{4\pi^2}\times(4\pi i\sigma^2)^2 = 4\beta c\sigma^4.$$

In the diagonal case, the second part of $\tilde{\nu}_x$ is different. By Remark 1, the product of two marginal integrals in the fourth-cumulant term is replaced by the diagonal spectral average. Hence,

$$\tilde{\nu}_{x_2,\mathrm{diag}} = -\frac{\beta c}{4\pi^2}\oint\oint z_1 z_2\bigg[\frac{\sigma^4}{3\{1+\sigma^2\underline{s}_1\}^2\{1+\sigma^2\underline{s}_2\}^2} + \frac{4\sigma^4}{3\{1+2\sigma^2\underline{s}_1\}^2\{1+2\sigma^2\underline{s}_2\}^2} + \frac{3\sigma^4}{\{1+3\sigma^2\underline{s}_1\}^2\{1+3\sigma^2\underline{s}_2\}^2}\bigg]\mathrm{d}\underline{s}_1\mathrm{d}\underline{s}_2.$$

For $a = 1, 2, 3$, write

$$I_a = \oint \frac{z(\underline{s})}{\{1+a\sigma^2\underline{s}\}^2}\mathrm{d}\underline{s},$$

where

$$z(\underline{s}) = -\frac{1}{\underline{s}} + \frac{c\sigma^2}{3\{1+\sigma^2\underline{s}\}} + \frac{2c\sigma^2}{3\{1+2\sigma^2\underline{s}\}} + \frac{c\sigma^2}{1+3\sigma^2\underline{s}}.$$

When $0 < c \le 1$, the contour of $\underline{s}$ encloses $-1/\sigma^2$, $-1/(2\sigma^2)$, and $-1/(3\sigma^2)$.

Therefore, we have

$$I_1 = 2\pi i\Big\{1 - \frac{25c}{12} + \frac{4c}{3} + \frac{3c}{4}\Big\} = 2\pi i,$$
$$I_2 = 2\pi i\Big\{\frac{c}{3} + 1 - \frac{10c}{3} + 3c\Big\} = 2\pi i,$$
$$I_3 = 2\pi i\Big\{\frac{c}{12} + \frac{4c}{3} + 1 - \frac{17c}{12}\Big\} = 2\pi i.$$

Thus, when $0 < c \le 1$,

$$\tilde{\nu}_{x_2,\mathrm{diag}} = -\frac{\beta c}{4\pi^2}\Big\{\frac{\sigma^4}{3}I_1^2 + \frac{4\sigma^4}{3}I_2^2 + 3\sigma^4 I_3^2\Big\} = -\frac{\beta c}{4\pi^2}\Big\{\frac{\sigma^4}{3} + \frac{4\sigma^4}{3} + 3\sigma^4\Big\}(2\pi i)^2$$
$$= \frac{14}{3}\beta c\sigma^4.$$

When $c > 1$, the contour encloses the pole at 0. Since

$$\operatorname*{Res}_{\underline{s}=0}\frac{z(\underline{s})}{\{1+a\sigma^2\underline{s}\}^2} = -1, \qquad a = 1, 2, 3,$$

we have $I_a = -2\pi i$ for $a = 1, 2, 3$. Hence the same value is obtained:

$$\tilde{\nu}_{x_2,\mathrm{diag}} = -\frac{\beta c}{4\pi^2}\Big\{\frac{\sigma^4}{3} + \frac{4\sigma^4}{3} + 3\sigma^4\Big\}(-2\pi i)^2 = \frac{14}{3}\beta c\sigma^4.$$

Finally, combining this with the first part $\tilde{\nu}_{x_1} = (14/3)(q+1)c\sigma^4$, we obtain,

$$\tilde{\nu}_{x,\mathrm{A4}} = \Big\{\frac{14}{3}(q+1) + 4\beta\Big\}c\sigma^4, \quad \tilde{\nu}_{x,\mathrm{diag}} = \Big\{\frac{14}{3}(q+1) + \frac{14}{3}\beta)\Big\}c\sigma^4.$$

S4 Proofs of Lemma 2 and Theorems 2–3

Following the structure in (2.4), the spectral behavior of $\mathbf{G}$ is inherently determined by the heterogeneous covariance matrices $\{\mathbf{\Sigma}^{(r)}\}_{r=1}^{\tau}$, which are characterized by generalized spiked models (Bai and Yao, 2012b). Under the homogeneous non-spiked baseline, these covariance matrices share a common orthogonal eigenbasis $\mathbf{U} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_p)$ and take the form

$$\mathbf{\Sigma}^{(r)} = \mathbf{U}(\mathbf{D}_0 + \mathbf{\Lambda}_r)\mathbf{U}^\top = \mathbf{U}\,\mathrm{diag}(\sigma_1^{(r)}, \dots, \sigma_p^{(r)})\mathbf{U}^\top, \text{ for each } r \in [1:\tau].$$

Here, $\mathbf{D}_0 = \mathrm{diag}(d_1, \dots, d_p)$ is the non-spiked diagonal matrix, and $\mathbf{\Lambda}_r = \mathrm{diag}(\lambda_1^{(r)}, \dots, \lambda_{\kappa_r}^{(r)}, 0, \dots, 0)$ is the spiked diagonal matrix with its non-zero entries sorted in descending order, where κ_r denotes the number of the rth population spiked eigenvalues. Then, we define the joint support of the spiked components across all populations as $\mathcal{M} = \{1, \dots, M\}$, where $M \coloneqq \max_{1\le r\le\tau} \kappa_r$. Therefore, for $r \in [1:\tau]$, the rth row-population eigenvalue $\sigma_m^{(r)}$ is

$$\sigma_m^{(r)} = \begin{cases} d_m + \lambda_m^{(r)}, & m \in [1:M], \\ d_m, & m \in [M+1:p]. \end{cases} \tag{S4.121}$$

We concatenate the sample matrices from (2.5) along the sample dimension, and denote the resulting matrix by $\breve{\mathbf{W}} = \big(\mathbf{W}_1^{(1)}, \ldots, \mathbf{W}_{h_1}^{(1)}, \ldots, \mathbf{W}_1^{(\tau)}, \ldots, \mathbf{W}_{h_\tau}^{(\tau)}\big) \in \mathbb{R}^{p\times N}$ with $N = nh$. Accordingly, (2.5) admits the equivalent form $\mathbf{G} = N^{-1}\breve{\mathbf{W}}\breve{\mathbf{W}}^\top$. Due to the spectral invariance of $\mathbf{G}$ under orthogonal transformations, its extreme eigenvalues are identical to those of $N^{-1}\check{\mathbf{W}}\check{\mathbf{W}}^\top$, where $\check{\mathbf{W}} \coloneqq \mathbf{U}^\top\breve{\mathbf{W}}$. Define

$$\check{\mathbf{G}} \coloneqq \frac{1}{N}\check{\mathbf{W}}\check{\mathbf{W}}^\top = \mathbf{U}^\top\big(\frac{1}{N}\breve{\mathbf{W}}\breve{\mathbf{W}}^\top\big)\mathbf{U} = \mathbf{U}^\top\mathbf{G}\mathbf{U}.$$

Since $\mathbf{U}$ is orthogonal and is a permutation matrix, $\check{\mathbf{G}}$ and $\mathbf{G}$ are orthogonally similar. Hence

$$\lambda_j(\check{\mathbf{G}}) = \lambda_j(\mathbf{G}), \qquad \text{for any } j = [1:p].$$

Therefore, it is sufficient to study $\check{\mathbf{G}}$. Then, Assumption 2 is imposed on the rotated innovations $\mathbf{U}^\top\boldsymbol{x}_{jt}^{(r)}$ in this common eigenbasis. More precisely, we have

$$\mathbf{U}^\top\boldsymbol{w}_{jt}^{(r)} = \operatorname{diag}\big(\{\sigma_1^{(r)}\}^{1/2}, \ldots, \{\sigma_p^{(r)}\}^{1/2}\big)\mathbf{U}^\top\boldsymbol{x}_{jt}^{(r)}.$$

Consequently, the joint spiked components and the non-spiked components in $\check{\mathbf{W}}$ can be decoupled into separate submatrices, and we partition $\check{\mathbf{W}}$ as follows:

$$\check{\mathbf{W}} = \begin{pmatrix}\mathbf{Y}_1\\ \mathbf{Y}_0\end{pmatrix} \in \mathbb{R}^{p\times N}, \quad \mathbf{Y}_1 = \begin{pmatrix}\boldsymbol{y}_1\\ \vdots\\ \boldsymbol{y}_M\end{pmatrix} \in \mathbb{R}^{M\times N}, \quad \mathbf{Y}_0 = \begin{pmatrix}\boldsymbol{y}_{M+1}\\ \vdots\\ \boldsymbol{y}_p\end{pmatrix} \in \mathbb{R}^{p_0\times N}.$$

where the submatrix $\mathbf{Y}_1$ corresponds to the index set $\mathcal{M}$, and $\mathbf{Y}_0$ contains the remaining $p_0 = p - M$ noise components, whose corresponding sample covariance matrices will be examined separately in our subsequent analysis.

To tackle the within-population dependence structure specified in Assumption 2 we shall use the following notation based on the non-spiked components. For any $m \in [M+1 : p]$, let $\boldsymbol{y}_m$ denote the corresponding row of the non-spiked block $\mathbf{Y}_0$. Define

$$\boldsymbol{y}_m = (\boldsymbol{y}_m^{(1)}, \boldsymbol{y}_m^{(2)}, \ldots, \boldsymbol{y}_m^{(\tau)}) \in \mathbb{R}^{1\times N},$$

the entry $y_{jmt}^{(r)}$ of $\boldsymbol{y}_m^{(r)} \in \mathbb{R}^{1\times nh_r}$ is therefore

$$y_{jmt}^{(r)} = \{\sigma_m^{(r)}\}^{1/2} x_{jmt}^{(r)} = \{\sigma_m^{(r)}\}^{1/2}\big(x_{0mt}^{(r)} + \varepsilon_{jmt}^{(r)}\big), \quad j \in [1 : h_r],\, t \in [1 : n].$$

Then, for $i, j \in [1 : h_r]$, $s, t \in [1 : n]$, under Assumption 2, we obtain

$$\begin{aligned}\mathbb{E}\big\{y_{jmt}^{(r)} y_{ims}^{(r)}\big\} &= d_m \mathbb{E}\Big[\big(x_{0mt}^{(r)} + \varepsilon_{jmt}^{(r)}\big)\big(x_{0ms}^{(r)} + \varepsilon_{ims}^{(r)}\big)\Big] \\ &= d_m\big[(1-\rho_r)\mathbf{1}_{\{i=j\}} + \rho_r\big]\mathbf{1}_{\{t=s\}}.\end{aligned}$$

Hence, we have $\mathbb{E}\big\{(\boldsymbol{y}_m^{(r)})^\top \boldsymbol{y}_m^{(r)}\big\} = d_m \mathbf{R}_{h_r}(\rho_r) \otimes \mathbf{I}_n \in \mathbb{R}^{nh_r \times nh_r}$, where $\otimes$ denotes the Kronecker product, and $\mathbf{R}_{h_r}(\rho_r) = (1-\rho_r)\mathbf{I}_{h_r} + \rho_r \mathbf{1}_{h_r}\mathbf{1}_{h_r}^\top$. Thus, the covariance matrix of $\boldsymbol{y}_m^\top$ evaluates to $\mathbb{E}(\boldsymbol{y}_m^\top \boldsymbol{y}_m) = d_m \mathbf{B}_0$, where the matrix $\mathbf{B}_0$ characterizes the within-population sample dependence structure, given by

$$\mathbf{B}_0 \coloneqq \operatorname{diag}\big(\mathbf{R}_{h_1}(\rho_1) \otimes \mathbf{I}_n, \ldots, \mathbf{R}_{h_\tau}(\rho_\tau) \otimes \mathbf{I}_n\big) \in \mathbb{R}^{N\times N}. \qquad \text{(S4.122)}$$

Defining the diagonal matrix $\mathbf{A}_0 = \mathrm{diag}(d_{M+1}, \ldots, d_p)$, we can express the non-spiked block via the separable representation $\mathbf{Y}_0 = \mathbf{A}_0^{1/2}\mathbf{Z}_0\mathbf{B}_0^{1/2}$, where $\mathbf{Z}_0 = (z_{it})_{p_0 \times N}$ is a matrix of standardized innovations. We assume that there exists a deterministic sequence $N^{-1/2} \le \phi_N \le N^{-c_\phi}$ for some $c_\phi > 0$, such that

$$\max_{1 \le i \le p_0,\, 1 \le t \le N} |z_{it}| \le N^{1/2}\phi_N. \tag{S4.123}$$

Next, we consider the analogy with the spiked block $\mathbf{Y}_1$. Let $\boldsymbol{y}_m$ denote its corresponding row for any $m \in [1 : M]$. The covariance matrix of $\boldsymbol{y}_m^\top$ is then given by

$$\mathbb{E}(\boldsymbol{y}_m^\top \boldsymbol{y}_m) = \boldsymbol{\Gamma}_m \mathbf{B}_0 = \mathrm{diag}\left(\sigma_m^{(1)}\mathbf{R}_{h_1}(\rho_1) \otimes \mathbf{I}_n, \ldots, \sigma_m^{(\tau)}\mathbf{R}_{h_\tau}(\rho_\tau) \otimes \mathbf{I}_n\right),$$

where the diagonal matrix $\boldsymbol{\Gamma}_m$ incorporates the heterogeneous perturbation strengths across populations:

$$\boldsymbol{\Gamma}_m := \mathrm{diag}\left(\sigma_m^{(1)}\mathbf{I}_{nh_1}, \ldots, \sigma_m^{(\tau)}\mathbf{I}_{nh_\tau}\right) \in \mathbb{R}^{N \times N}. \tag{S4.124}$$

Thus, the multi-population heterogeneity affects the phase transition of $\mathbf{G}$ through the matrices $\boldsymbol{\Gamma}_m\mathbf{B}_0$ for $m \in [1 : M]$.

Since outlier eigenvalues are separated from the spectral bulk, we first characterize the rightmost edge of the non-spiked reference matrix. Define

$$\mathbf{Q}_0 = \frac{1}{N}\mathbf{Y}_0\mathbf{Y}_0^\top = \frac{1}{N}\mathbf{A}_0^{1/2}\mathbf{Z}_0\mathbf{B}_0\mathbf{Z}_0^\top\mathbf{A}_0^{1/2}. \tag{S4.125}$$

We also introduce its dual matrix

$$\widetilde{\mathbf{Q}}_0 = \frac{1}{N}\mathbf{Y}_0^\top \mathbf{Y}_0 = \frac{1}{N}\mathbf{B}_0^{1/2}\mathbf{Z}_0^\top \mathbf{A}_0\mathbf{Z}_0\mathbf{B}_0^{1/2}. \tag{S4.126}$$

The matrices $\mathbf{Q}_0$ and $\widetilde{\mathbf{Q}}_0$ have the same nonzero eigenvalues. Thus $\widetilde{\mathbf{Q}}_0$ is the dual separable sample covariance matrix associated with the non-spiked reference model. We denote the empirical spectral distributions (ESD) of $\mathbf{A}_0$ and $\mathbf{B}_0$ by $F_{p_0}^{\mathbf{A}_0}$ and $F_N^{\mathbf{B}_0}$, respectively, and set $c_N = p_0/N$. We assume that there exists a constant $c_0 \in (0,1)$ such that $c_0 \le c_N \le c_0^{-1}$ and

$$\max\{\|\mathbf{A}_0\|, \|\mathbf{B}_0\|\} \le c_0^{-1} \quad , F_{p_0}^{\mathbf{A}_0}([0,c_0]) \vee F_N^{\mathbf{B}_0}([0,c_0]) \le 1 - c_0. \tag{S4.127}$$

For any $z \in \mathbb{C}^+$, we define $s_{1,N}(z)$ and $s_{2,N}(z)$ be the unique solutions in $\mathbb{C}^+$ to the following system of self-consistent equations

$$\begin{cases} s_{1,N}(z) = -c_N \displaystyle\int \frac{t}{z\{1+ts_{2,N}(z)\}}\,\mathrm{d}F_{p_0}^{\mathbf{A}_0}(t), \\ s_{2,N}(z) = -\displaystyle\int \frac{t}{z\{1+ts_{1,N}(z)\}}\,\mathrm{d}F_N^{\mathbf{B}_0}(t). \end{cases} \tag{S4.128}$$

Define

$$s_{c_N}(z) = -\frac{1}{z}\int \frac{1}{1+ts_{2,N}(z)}\,\mathrm{d}F_{p_0}^{\mathbf{A}_0}(t), \qquad z \in \mathbb{C}^+. \tag{S4.129}$$

Then s_{c_N} is the stieltjes transform of a deterministic probability measure, denoted by F_{c_N}, supported on $[0,\infty)$. By Ding and Yang (2021, Lemma 2.5), its support on $(0,\infty)$ can be written as a finite union of intervals, i.e.

$$\operatorname{supp}(F_{c_N}) \cap (0,\infty) = \bigcup_{k=1}^{L_N} [e_{2k,N}, e_{2k-1,N}] \cap (0,\infty).$$

The endpoints $e_{k,N}$ are called the spectral edges. In particular, we focus on the rightmost edge

$$\lambda_{+,N} \coloneqq e_{1,N} = \sup\{\operatorname{supp}(F_{c_N}) \cap (0,\infty)\}. \tag{S4.130}$$

We assume that the rightmost edge $\lambda_{+,N}$ is strictly regular. Specifically, there exists a constant $c_0 \in (0,1)$ such that,

$$1 + s_{1,N}(\lambda_{+,N})\lambda_{\max}(\mathbf{B}_0) \geq c_0, \quad 1 + s_{2,N}(\lambda_{+,N})\lambda_{\max}(\mathbf{A}_0) \geq c_0. \tag{S4.131}$$

Under the preliminary assumption of the common eigenbasis $\mathbf{U}$ and the characterization of the bulk edge $\lambda_{+,N}$ of the non-spiked reference matrix $\mathbf{Q}_0$, the proof of Theorem 3 is organized as follows. Step 1 reduces the eigenvalue equation of $\mathbf{G}$, outside the spectrum of $\mathbf{Q}_0$, to an exact finite-dimensional determinant equation. Step 2 proves the uniform concentration of the spiked row quadratic forms around their trace counterparts. Step 3 replaces the random trace terms by deterministic equivalents and derives the threshold function in Theorem 2. Step 4 applies Rouché's theorem to count the zeros near each deterministic root and proves the almost sure convergence of the separated sample outliers. Step 5 proves that the remaining sample eigenvalues after the separated outliers stick to the bulk edge $\lambda_{+,N}$.

Step 1. Exact finite-dimensional determinant equation.

For $z \notin \operatorname{spec}(\mathbf{Q}_0) \cup \{0\}$, we partition

$$\check{\mathbf{G}} - z\mathbf{I}_p = \begin{pmatrix} N^{-1}\mathbf{Y}_1\mathbf{Y}_1^\top - z\mathbf{I}_M & N^{-1}\mathbf{Y}_1\mathbf{Y}_0^\top \\ N^{-1}\mathbf{Y}_0\mathbf{Y}_1^\top & \mathbf{Q}_0 - z\mathbf{I}_{p_0} \end{pmatrix}.$$

By the Schur complement formula,

$$\begin{aligned} \det(\check{\mathbf{G}} - z\mathbf{I}_p) &= \det(\mathbf{Q}_0 - z\mathbf{I}_{p_0}) \\ &\quad \times \det\left[N^{-1}\mathbf{Y}_1\mathbf{Y}_1^\top - z\mathbf{I}_M - N^{-2}\mathbf{Y}_1\mathbf{Y}_0^\top(\mathbf{Q}_0 - z\mathbf{I}_{p_0})^{-1}\mathbf{Y}_0\mathbf{Y}_1^\top\right]. \end{aligned} \tag{S4.132}$$

Recall (S4.126), define the resolvent $\mathbf{P}_0(z) \coloneqq (\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1}$. Observe that

$$N^{-1}\mathbf{Y}_0^\top(\mathbf{Q}_0 - z\mathbf{I}_{p_0})^{-1}\mathbf{Y}_0 = \widetilde{\mathbf{Q}}_0(\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1}. \tag{S4.133}$$

Indeed,

$$(\mathbf{Q}_0 - z\mathbf{I}_{p_0})\mathbf{Y}_0 = \left(N^{-1}\mathbf{Y}_0\mathbf{Y}_0^\top - z\mathbf{I}_{p_0}\right)\mathbf{Y}_0 = \mathbf{Y}_0\left(N^{-1}\mathbf{Y}_0^\top\mathbf{Y}_0 - z\mathbf{I}_N\right) = \mathbf{Y}_0(\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N).$$

Since $\mathbf{Q}_0$ and $\widetilde{\mathbf{Q}}_0$ share the same nonzero eigenvalues with $z \neq 0$, the invertibility of $\mathbf{Q}_0 - z\mathbf{I}_{p_0}$ implies the invertibility of $\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N$. Hence

$$(\mathbf{Q}_0 - z\mathbf{I}_{p_0})^{-1}\mathbf{Y}_0 = \mathbf{Y}_0(\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1}.$$

Multiplying from the left by $N^{-1}\mathbf{Y}_0^\top$ yields

$$N^{-1}\mathbf{Y}_0^\top(\mathbf{Q}_0 - z\mathbf{I}_{p_0})^{-1}\mathbf{Y}_0 = N^{-1}\mathbf{Y}_0^\top\mathbf{Y}_0(\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1} = \widetilde{\mathbf{Q}}_0(\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1},$$

which proves (S4.133). Moreover,

$$\widetilde{\mathbf{Q}}_0(\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1} = \mathbf{I}_N + z(\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1} = \mathbf{I}_N + z\mathbf{P}_0(z). \tag{S4.134}$$

Combining (S4.133) and (S4.134), we obtain

$$
\begin{aligned}
&N^{-1}\mathbf{Y}_1\mathbf{Y}_1^\top - z\mathbf{I}_M - N^{-2}\mathbf{Y}_1\mathbf{Y}_0^\top(\mathbf{Q}_0 - z\mathbf{I}_{p_0})^{-1}\mathbf{Y}_0\mathbf{Y}_1^\top \\
&\quad = N^{-1}\mathbf{Y}_1\mathbf{Y}_1^\top - z\mathbf{I}_M - N^{-1}\mathbf{Y}_1\{\mathbf{I}_N + z\mathbf{P}_0(z)\}\mathbf{Y}_1^\top \\
&\quad = -z\left\{\mathbf{I}_M + N^{-1}\mathbf{Y}_1\mathbf{P}_0(z)\mathbf{Y}_1^\top\right\}. \qquad \text{(S4.135)}
\end{aligned}
$$

Substituting this into (S4.132), we obtain

$$
\det(\check{\mathbf{G}} - z\mathbf{I}_p) = \det(\mathbf{Q}_0 - z\mathbf{I}_{p_0})(-z)^M D_N(z),
$$

where

$$
D_N(z) \coloneqq \det\{\mathbf{I}_M + N^{-1}\mathbf{Y}_1\mathbf{P}_0(z)\mathbf{Y}_1^\top\}. \qquad \text{(S4.136)}
$$

Since $\mathbf{G}$ and $\check{\mathbf{G}}$ share the same spectrum, the secular equation reduces to $D_N(z) = 0$, i.e.

$$
\det\{\mathbf{I}_M + N^{-1}\mathbf{Y}_1\mathbf{P}_0(z)\mathbf{Y}_1^\top\} = 0.
$$

Step 2. Quadratic-form concentration of the spiked rows.

Let $\mathcal{D}$ be a deterministic compact subset of $\mathbb{C} \setminus \{0\}$ separated from the non-spiked bulk support. That is, for some constant $c_{\mathcal{D}} > 0$,

$$
\operatorname{dist}\big(\mathcal{D}, \operatorname{supp}(F_{c_N})\big) \geq c_{\mathcal{D}}, \qquad \inf_{z\in\mathcal{D}} |z| \geq c_{\mathcal{D}}. \qquad \text{(S4.137)}
$$

Then, define

$$
\mathbf{B}_{mk} \coloneqq \mathbb{E}(\boldsymbol{y}_k^\top \boldsymbol{y}_m) = \mathbf{1}_{\{m=k\}}\boldsymbol{\Gamma}_m\mathbf{B}_0 \quad \text{for any } m, k \in \mathcal{M}. \qquad \text{(S4.138)}
$$

We first prove a high-moment bound for deterministic test matrices. Specifically, for any deterministic matrix $\mathbf{A} \in \mathbb{C}^{N\times N}$ satisfying $\|\mathbf{A}\| \le C$ and for any fixed integer $q \ge 1$, we show that

$$\mathbb{E}\left|N^{-1}\left\{\boldsymbol{y}_m \mathbf{A} \boldsymbol{y}_k^\top - \operatorname{tr}(\mathbf{B}_{mk}\mathbf{A})\right\}\right|^{2q} \le C_q\left(N^{-1/2} + \phi_N^2\right)^q. \tag{S4.139}$$

This bound suffices to establish almost sure convergence. By the independence of the spiked rows and $\mathbf{Y}_0$, the resolvent $\mathbf{P}_0(z) = (\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1}$ is conditionally deterministic given $\mathbf{Y}_0$. Substituting $\mathbf{A} = \mathbf{P}_0(z)$ into (S4.139), invoking the uniform boundedness of $\mathbf{P}_0(z)$ on $\mathcal{D}$, and applying a finite-net argument over $\mathcal{D}$, we shall obtain

$$\sup_{z\in\mathcal{D}}\left|N^{-1}\left\{\boldsymbol{y}_m \mathbf{P}_0(z) \boldsymbol{y}_k^\top - \operatorname{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z)\}\right\}\right| = o_{a.s.}(1), \qquad m,k\in\mathcal{M}. \tag{S4.140}$$

First, we add the following assumption.

Assumption 5. There exists a deterministic sequence $N^{-1/2} \le \phi_N \le N^{-c_\phi}$ for some $c_\phi > 0$, such that $\max_{1\le m\le M,\,1\le r\le\tau,\,1\le j\le h_r,\,1\le t\le n}|x_{jmt}^{(r)}| \le N^{1/2}\phi_N$.

For each $m \in [1:M]$, write

$$\boldsymbol{y}_m = (\boldsymbol{y}_m^{(1)}, \boldsymbol{y}_m^{(2)}, \ldots, \boldsymbol{y}_m^{(\tau)}) \in \mathbb{R}^{1\times N},$$

where $\boldsymbol{y}_m^{(r)} = (y_{1m1}^{(r)}, \ldots, y_{1mn}^{(r)}, \ldots, y_{h_r m1}^{(r)}, \ldots, y_{h_r mn}^{(r)}) \in \mathbb{R}^{1\times nh_r}$.

To use the independence with respect to the sample index t, we regroup the entries by a deterministic permutation. Specifically, for $t \in [1:n]$, define

$$\boldsymbol{\zeta}_{mt}^{(r)} = (y_{1mt}^{(r)}, \ldots, y_{h_r mt}^{(r)})^\top \in \mathbb{R}^{h_r}, \qquad r \in [1:\tau],$$

and

$$\boldsymbol{\zeta}_{mt} \coloneqq \big((\boldsymbol{\zeta}_{mt}^{(1)})^\top, \ldots, (\boldsymbol{\zeta}_{mt}^{(\tau)})^\top\big)^\top \in \mathbb{R}^h.$$

Thus we have

$$\begin{aligned}
&(\boldsymbol{\zeta}_{m1}^\top, \ldots, \boldsymbol{\zeta}_{mn}^\top) \\
&= \big(y_{1m1}^{(1)}, \ldots, y_{h_1 m1}^{(1)}, \ldots, y_{1m1}^{(\tau)}, \ldots, y_{h_\tau m1}^{(\tau)}, \ldots, y_{1mn}^{(1)}, \ldots, y_{h_1 mn}^{(1)}, \ldots, y_{1mn}^{(\tau)}, \ldots, y_{h_\tau mn}^{(\tau)}\big).
\end{aligned}$$

This vector is obtained from the original $\boldsymbol{y}_m$ by a deterministic permutation of its coordinates. Since the same permutation is applied to $\boldsymbol{y}_m$, $\boldsymbol{y}_k$, $\mathbf{A}$, and $\mathbf{B}_{mk}$, the quantities $\boldsymbol{y}_m \mathbf{A} \boldsymbol{y}_k^\top$ and $\mathrm{tr}\{\mathbf{B}_{mk}\mathbf{A}\}$ are unchanged. Hence, after this relabeling, we keep the same notation and write

$$\boldsymbol{y}_m = (\boldsymbol{\zeta}_{m1}^\top, \ldots, \boldsymbol{\zeta}_{mn}^\top) \in \mathbb{R}^{1\times N}.$$

Moreover, by the definition of $y_{jmt}^{(r)}$,

$$\boldsymbol{\zeta}_{mt}^{(r)} = \{\sigma_m^{(r)}\}^{1/2}\big(x_{0mt}^{(r)} + \varepsilon_{1mt}^{(r)}, \ldots, x_{0mt}^{(r)} + \varepsilon_{h_r mt}^{(r)}\big)^\top.$$

By construction, for $m, k \in \mathcal{M}$, in the original ordering, recall (S4.138)

$$\mathbf{B}_{mk} = \mathbb{E}(\boldsymbol{y}_k^\top \boldsymbol{y}_m) = \mathbf{1}_{\{m=k\}} \boldsymbol{\Gamma}_m \mathbf{B}_0.$$

After relabeling the coordinates by the sample index t, we use the same symbol $\mathbf{B}_{mk}$ for the corresponding permuted covariance matrix. Then

$$\mathbb{E}\big(\boldsymbol{\zeta}_{kt}\boldsymbol{\zeta}_{mt}^{\top}\big) = \mathbf{1}_{\{m=k\}}\,\mathrm{diag}\left(\sigma_m^{(1)}\mathbf{R}_{h_1}(\rho_1), \ldots, \sigma_m^{(\tau)}\mathbf{R}_{h_\tau}(\rho_\tau)\right).$$

We partition $\mathbf{A} \in \mathbb{C}^{N\times N}$ according to the same grouping:

$$\mathbf{A} = (\mathbf{A}_{\mu\nu})_{\mu,\nu=1}^{n} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1n} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{n1} & \mathbf{A}_{n2} & \cdots & \mathbf{A}_{nn} \end{pmatrix}, \qquad \mathbf{A}_{\mu\nu} \in \mathbb{C}^{h\times h}.$$

Therefore, we have

$$\begin{aligned}
&N^{-1}\left\{\boldsymbol{y}_m\mathbf{A}\boldsymbol{y}_k^{\top} - \mathrm{tr}(\mathbf{B}_{mk}\mathbf{A})\right\} \\
&= N^{-1}\Big[\sum_{\mu=1}^{n}\sum_{\nu=1}^{n}\boldsymbol{\zeta}_{m\mu}^{\top}\mathbf{A}_{\mu\nu}\boldsymbol{\zeta}_{k\nu} - \sum_{\mu=1}^{n}\mathrm{tr}\left\{\mathbb{E}(\boldsymbol{\zeta}_{k\mu}\boldsymbol{\zeta}_{m\mu}^{\top})\mathbf{A}_{\mu\mu}\right\}\Big] \\
&= \sum_{\mu=1}^{n}\sum_{\nu=1}^{n}N^{-1}\boldsymbol{\zeta}_{m\mu}^{\top}\mathbf{A}_{\mu\nu}\boldsymbol{\zeta}_{k\nu} - \sum_{\mu=1}^{n}N^{-1}\,\mathrm{tr}\left\{\mathbb{E}(\boldsymbol{\zeta}_{k\mu}\boldsymbol{\zeta}_{m\mu}^{\top})\mathbf{A}_{\mu\mu}\right\} \\
&= \sum_{\mu=1}^{n}\sum_{\nu=1}^{n}\left[(N^{-1/2}\boldsymbol{\zeta}_{m\mu})^{\top}\mathbf{A}_{\mu\nu}(N^{-1/2}\boldsymbol{\zeta}_{k\nu}) - \mathbf{1}_{\{\mu=\nu\}}\,\mathrm{tr}\left\{\mathbb{E}\{(N^{-1/2}\boldsymbol{\zeta}_{k\mu})(N^{-1/2}\boldsymbol{\zeta}_{m\mu})^{\top}\}\mathbf{A}_{\mu\mu}\right\}\right].
\end{aligned}$$

Under Assumptions 1–2, 5 and (S4.123), assuming $\max_{1\le m\le M,1\le r\le\tau}\sigma_m^{(r)} = O(1)$, we have uniformly over $m \in [1:M]$ and $t \in [1:n]$,

$$\|N^{-1/2}\boldsymbol{\zeta}_{mt}\| \le C\phi_N,\ \mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^2 \le CN^{-1},\ \mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^4 \le CN^{-2}. \tag{S4.141}$$

For $2 \le s \le 4$, Hölder's inequality yields

$$\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^s \le \left(\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^4\right)^{(s-2)/2}\left(\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^2\right)^{(4-s)/2} \le C_sN^{-s/2}.$$

For $s \geq 4$, we have

$$\begin{aligned}
\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^s &= \mathbb{E}\left[\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^{s-4}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^4\right] \\
&\leq (C\phi_N)^{s-4}\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^4 \\
&\leq C_s N^{-2}\phi_N^{s-4}.
\end{aligned}$$

Thus, for every fixed $s \geq 2$,

$$\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^s \leq C_s \begin{cases} N^{-s/2}, & 2 \leq s \leq 4, \\ N^{-2}\phi_N^{s-4}, & s \geq 4. \end{cases} \tag{S4.142}$$

We now prove the high-moment bound for the centered quadratic form. Fix an integer $q \geq 1$. Expanding the $2q$-th absolute moment gives

$$\begin{aligned}
&\mathbb{E}\Big|N^{-1}\{\boldsymbol{y}_m\mathbf{A}\boldsymbol{y}_k^\top - \mathrm{tr}(\mathbf{B}_{mk}\mathbf{A})\}\Big|^{2q} \\
&\quad= \sum_{\mu_1,\nu_1,\ldots,\mu_{2q},\nu_{2q}=1}^{n} \mathbb{E}\prod_{\ell=1}^{2q}\Big[(N^{-1/2}\boldsymbol{\zeta}_{m\mu_\ell})^\top\mathbf{A}_{\mu_\ell\nu_\ell}(N^{-1/2}\boldsymbol{\zeta}_{k\nu_\ell}) \\
&\qquad\qquad - \mathbf{1}_{\{\mu_\ell=\nu_\ell\}}\,\mathrm{tr}\left\{\mathbb{E}\{(N^{-1/2}\boldsymbol{\zeta}_{k\mu_\ell})(N^{-1/2}\boldsymbol{\zeta}_{m\mu_\ell})^\top\}\mathbf{A}_{\mu_\ell\mu_\ell}\right\}\Big],
\end{aligned} \tag{S4.143}$$

where, in the complex case, the same argument applies to the conjugated factors. For an index configuration $\pi = (\mu_1, \nu_1, \ldots, \mu_{2q}, \nu_{2q})$, define the index multiplicity

$$d_\pi(t) \coloneqq \#\{\ell : \mu_\ell = t\} + \#\{\ell : \nu_\ell = t\}, \qquad t \in [1:n],$$

along with the set of appearing indices and its cardinality:

$$V(\pi) \coloneqq \{t \in [1:n] : d_\pi(t) > 0\}, \qquad v(\pi) \coloneqq |V(\pi)|.$$

For notational simplicity, we denote the centered summand by

$$H_{\mu\nu}^{(mk)} := (N^{-1/2}\boldsymbol{\zeta}_{m\mu})^\top \mathbf{A}_{\mu\nu}(N^{-1/2}\boldsymbol{\zeta}_{k\nu}) - \mathbf{1}_{\{\mu=\nu\}} \operatorname{tr}\big[\mathbb{E}\{(N^{-1/2}\boldsymbol{\zeta}_{k\mu})(N^{-1/2}\boldsymbol{\zeta}_{m\mu})^\top\}\mathbf{A}_{\mu\mu}\big].$$

Then a typical term in the $2q$-th moment expansion is

$$\mathbb{E}\prod_{\ell=1}^{2q} H_{\mu_\ell\nu_\ell}^{(mk)}.$$

Suppose that $d_\pi(t) = 1$ for some index t, meaning t appears exactly once among the $4q$ indices $\mu_1, \nu_1, \ldots, \mu_{2q}, \nu_{2q}$. Consequently, there exists a unique ℓ_0 such that either $\mu_{\ell_0} = t$ (with $\nu_{\ell_0} \neq t$) or $\nu_{\ell_0} = t$ (with $\mu_{\ell_0} \neq t$).

If $\mu_{\ell_0} = t$ and $\nu_{\ell_0} \neq t$, we have $\mathbf{1}_{\{\mu_{\ell_0}=\nu_{\ell_0}\}} = 0$, which simplifies the factor to

$$H_{\mu_{\ell_0}\nu_{\ell_0}}^{(mk)} = (N^{-1/2}\boldsymbol{\zeta}_{mt})^\top \mathbf{A}_{t\nu_{\ell_0}}(N^{-1/2}\boldsymbol{\zeta}_{k\nu_{\ell_0}}).$$

Observe that all factors in $\prod_{\ell\neq\ell_0} H_{\mu_\ell\nu_\ell}^{(mk)}$ as well as the vector $\boldsymbol{\zeta}_{k\nu_{\ell_0}}$ involve only sample indices strictly different from t. They are therefore entirely independent of $\boldsymbol{\zeta}_{mt}$. Using the centeredness $\mathbb{E}\boldsymbol{\zeta}_{mt} = \mathbf{0}$, we naturally obtain

$$\begin{aligned}
\mathbb{E}\prod_{\ell=1}^{2q} H_{\mu_\ell\nu_\ell}^{(mk)} &= \mathbb{E}\Big[\prod_{\ell\neq\ell_0} H_{\mu_\ell\nu_\ell}^{(mk)}\, (N^{-1/2}\boldsymbol{\zeta}_{mt})^\top \mathbf{A}_{t\nu_{\ell_0}}(N^{-1/2}\boldsymbol{\zeta}_{k\nu_{\ell_0}})\Big] \\
&= \mathbb{E}\Big[\prod_{\ell\neq\ell_0} H_{\mu_\ell\nu_\ell}^{(mk)}\, \big\{\mathbb{E}(N^{-1/2}\boldsymbol{\zeta}_{mt})\big\}^\top \mathbf{A}_{t\nu_{\ell_0}}(N^{-1/2}\boldsymbol{\zeta}_{k\nu_{\ell_0}})\Big] \\
&= 0.
\end{aligned}$$

By symmetry, the case where $\nu_{\ell_0} = t$ and $\mu_{\ell_0} \neq t$ follows an identical rationale. Indeed, the factor becomes

$$H_{\mu_{\ell_0}\nu_{\ell_0}}^{(mk)} = (N^{-1/2}\boldsymbol{\zeta}_{m\mu_{\ell_0}})^\top \mathbf{A}_{\mu_{\ell_0}t}(N^{-1/2}\boldsymbol{\zeta}_{kt}),$$

where all remaining random vectors are independent of $\boldsymbol{\zeta}_{kt}$. Since $\mathbb{E}\boldsymbol{\zeta}_{kt} = \mathbf{0}$, the expectation similarly vanishes, i.e. $\mathbb{E}\prod_{\ell=1}^{2q} H_{\mu_\ell\nu_\ell}^{(mk)} = 0$.

Consequently, a non-vanishing term requires $d_\pi(t) \geq 2$ for all $t \in V(\pi)$. Furthermore, because each of the $2q$ factors supplies two index positions, the total degree satisfies $\sum_{t\in V(\pi)} d_\pi(t) = 4q$. Combining this identity with the constraint $d_\pi(t) \geq 2$, we deduce that

$$2v(\pi) \leq \sum_{t\in V(\pi)} d_\pi(t) = 4q, \tag{S4.144}$$

yielding the upper bound $v(\pi) \leq 2q$ for the number of distinct sample indices. Moreover, factors involving disjoint sets of sample indices are independent. Thus, if a maximal group of factors linked through common sample indices contains only one factor, then the corresponding expectation is zero because $H_{\mu\nu}^{(mk)}$ is centered. Therefore, for a contributing configuration, each such maximal group contains at least two factors. Let $\kappa(\pi)$ be the number of these maximal groups. Since there are $2q$ factors in total, we have $\kappa(\pi) \leq q$.

We first establish the bounds for the individual components of $H_{\mu\nu}^{(mk)}$. By the Cauchy–Schwarz inequality, the random bilinear part is bounded by

$$\left|(N^{-1/2}\boldsymbol{\zeta}_{m\mu})^\top \mathbf{A}_{\mu\nu}(N^{-1/2}\boldsymbol{\zeta}_{k\nu})\right| \leq \|\mathbf{A}_{\mu\nu}\|_F \|N^{-1/2}\boldsymbol{\zeta}_{m\mu}\| \|N^{-1/2}\boldsymbol{\zeta}_{k\nu}\|. \tag{S4.145}$$

For the deterministic trace part, the Cauchy–Schwarz inequality gives

$$\begin{aligned}
&\Big|\operatorname{tr}\big[\mathbb{E}\{(N^{-1/2}\boldsymbol{\zeta}_{k\mu})(N^{-1/2}\boldsymbol{\zeta}_{m\mu})^\top\}\mathbf{A}_{\mu\mu}\big]\Big| \le \Big\|\mathbb{E}\big\{(N^{-1/2}\boldsymbol{\zeta}_{k\mu})(N^{-1/2}\boldsymbol{\zeta}_{m\mu})^\top\big\}\Big\|_F\|\mathbf{A}_{\mu\mu}\|_F \\
&\quad\le \mathbb{E}\big\{\|N^{-1/2}\boldsymbol{\zeta}_{k\mu}\|\|N^{-1/2}\boldsymbol{\zeta}_{m\mu}\|\big\}\|\mathbf{A}_{\mu\mu}\|_F \le \big(\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{k\mu}\|^2\big)^{1/2}\big(\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{m\mu}\|^2\big)^{1/2}\|\mathbf{A}_{\mu\mu}\|_F \\
&\quad\le CN^{-1}\|\mathbf{A}_{\mu\mu}\|_F. \qquad \text{(S4.146)}
\end{aligned}$$

For each $t \in V(\pi)$, let $b_1, \ldots, b_{d_\pi(t)} \in \{m, k\}$ denote the labels of the block variables carrying the sample index t. For an integer $s \ge 0$, define $\omega(0) \coloneqq 1$, $\omega(1) \coloneqq N^{-1/2}$, and, for $s \ge 2$,

$$\omega(s) \coloneqq \begin{cases} N^{-s/2}, & 2 \le s \le 4, \\ N^{-2}\phi_N^{s-4}, & s \ge 5. \end{cases} \qquad \text{(S4.147)}$$

Then, by Hölder's inequality and (S4.141), we have $\mathbb{E}\prod_{i=1}^{d_\pi(t)}\|N^{-1/2}\boldsymbol{\zeta}_{b_i t}\| \le C_q\omega(d_\pi(t))$. Indeed, if $2 \le d_\pi(t) \le 4$, then

$$\mathbb{E}\prod_{i=1}^{d_\pi(t)}\|N^{-1/2}\boldsymbol{\zeta}_{b_i t}\| \le \prod_{i=1}^{d_\pi(t)}\big(\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{b_i t}\|^{d_\pi(t)}\big)^{1/d_\pi(t)} \le C_q N^{-d_\pi(t)/2} = C_q\omega(d_\pi(t)).$$

If $d_\pi(t) \ge 5$, then

$$\mathbb{E}\prod_{i=1}^{d_\pi(t)}\|N^{-1/2}\boldsymbol{\zeta}_{b_i t}\| \le \prod_{i=1}^{d_\pi(t)}\big(\mathbb{E}\|N^{-1/2}\boldsymbol{\zeta}_{b_i t}\|^{d_\pi(t)}\big)^{1/d_\pi(t)} \le C_q N^{-2}\phi_N^{d_\pi(t)-4} = C_q\omega(d_\pi(t)).$$

We now bound the product $\prod_{\ell=1}^{2q} H_{\mu_\ell\nu_\ell}^{(mk)}$ for a fixed contributing configuration. Expanding each $H_{\mu_\ell\nu_\ell}^{(mk)}$, every factor contributes either a random bilinear term or a deterministic trace term. For each $t \in V(\pi)$, let $e_\pi(t)$ be the number of random block variables with sample index t, and let $p_\pi(t)$ be the number of deterministic trace factors associated with the diagonal block t. Since each trace factor corresponds to the case $\mu = \nu = t$, it accounts for two occurrences of the index t. Hence

$$d_\pi(t) = e_\pi(t) + 2p_\pi(t).$$

The $p_\pi(t)$ trace factors contribute at most $CN^{-p_\pi(t)}$, while the $e_\pi(t)$ random block variables contribute at most $C_q\omega(e_\pi(t))$. By the definition of $\omega(d_\pi(t))$ and the lower bound $N^{-1/2} \leq \phi_N$, we have $N^{-p_\pi(t)}\omega(e_\pi(t)) \leq C\omega(d_\pi(t))$. Therefore,

$$\left|\mathbb{E}\prod_{\ell=1}^{2q} H_{\mu_\ell\nu_\ell}^{(mk)}\right| \leq C_q \prod_{\ell=1}^{2q} \|\mathbf{A}_{\mu_\ell\nu_\ell}\|_F \prod_{t\in V(\pi)} \omega(d_\pi(t)). \tag{S4.148}$$

where $\omega(\cdot)$ is defined in (S4.147).

We now sum over the contributing configurations. We partition the index configurations into equivalence classes based on their equality relations. Let $\mathfrak{p}$ represent a fixed contributing partition of the $4q$ index positions, and let $\mathcal{C}(\mathfrak{p})$ be the equivalence class of all configurations $\pi = (\mu_1, \nu_1, \ldots, \mu_{2q}, \nu_{2q})$ that induce this partition. For all $\pi \in \mathcal{C}(\mathfrak{p})$, the number $v(\pi)$ and the multiplicities $\{d_\pi(t) : t \in V(\pi)\}$ are fixed up to a relabeling of the sample indices. Then, put $a_{\mu\nu} := \|\mathbf{A}_{\mu\nu}\|_F$. Since $\|\mathbf{A}\| \leq C$ and h is fixed, we have

$$\max_{\mu,\nu} a_{\mu\nu} \leq C, \qquad \sum_{\mu,\nu=1}^{n} a_{\mu\nu}^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}^*) \leq N\|\mathbf{A}\|^2 \leq CN.$$

Let $\kappa(\pi)$ denote the number of maximal groups of factors linked through common sample indices. Since each contributing maximal group contains at least two factors, we have $\kappa(\pi) \leq q$. For a fixed pattern $\mathfrak{p}$, the summation over $\pi \in \mathcal{C}(\mathfrak{p})$ factorizes across these independent groups. Within each maximal group, successive applications of the Cauchy–Schwarz inequality, together with the bounds $\|\mathbf{A}\| \leq C$ and $\sum_{\mu,\nu} a_{\mu\nu}^2 \leq CN$, control the corresponding sub-sum by $C_q N$. Multiplying the contributions from all $\kappa(\pi)$ groups yields

$$\sum_{\pi \in \mathcal{C}(\mathfrak{p})} \prod_{\ell=1}^{2q} a_{\mu_\ell \nu_\ell} \leq C_q N^{\kappa(\pi)}. \tag{S4.149}$$

Next, define

$$r(\pi) := \sum_{t \in V(\pi):\, d_\pi(t) \geq 5} \{d_\pi(t) - 4\}.$$

The quantity $r(\pi)$ records the excess multiplicity beyond four among the sample indices with multiplicity at least five. By the definition of $\omega(d_\pi(t))$ in (S4.147) and combining (S4.144),

$$\begin{aligned}\prod_{t \in V(\pi)} \omega(d_\pi(t)) &= \prod_{t \in V(\pi):\, 2 \leq d_\pi(t) \leq 4} N^{-d_\pi(t)/2} \prod_{t \in V(\pi):\, d_\pi(t) \geq 5} N^{-2} \phi_N^{d_\pi(t)-4} \\ &= N^{-2q + r(\pi)/2} \phi_N^{r(\pi)}.\end{aligned} \tag{S4.150}$$

Combining (S4.149) and (S4.150), we obtain

$$\sum_{\pi \in \mathcal{C}(\mathfrak{p})} \prod_{\ell=1}^{2q} \|\mathbf{A}_{\mu_\ell \nu_\ell}\|_F \prod_{t \in V(\pi)} \omega(d_\pi(t)) \leq C_q N^{\kappa(\pi)} N^{-2q + r(\pi)/2} \phi_N^{r(\pi)} \tag{S4.151}$$

$$= C_q (N^{-1/2})^{4q-2\kappa(\pi)-r(\pi)} (\phi_N^2)^{r(\pi)/2}.$$

Since each contributing configuration satisfies $d_\pi(t) \geq 2$ for all $t \in V(\pi)$, the excess multiplicity satisfies

$$r(\pi) \leq \sum_{t \in V(\pi)} \{d_\pi(t) - 2\} = 4q - 2v(\pi).$$

Moreover, each maximal group of factors linked through common sample indices contains at least one appearing sample index, so $v(\pi) \geq \kappa(\pi)$. Since each contributing maximal group contains at least two factors and there are $2q$ factors in total, we also have $\kappa(\pi) \leq q$. Hence

$$4q - 2\kappa(\pi) - \frac{r(\pi)}{2} \geq 2q - \kappa(\pi) \geq q.$$

Therefore, since $N^{-1/2} + \phi_N^2 < 1$ for all sufficiently large N,

$$\begin{aligned} (N^{-1/2})^{4q-2\kappa(\pi)-r(\pi)} (\phi_N^2)^{r(\pi)/2} &\leq \left(N^{-1/2} + \phi_N^2\right)^{4q-2\kappa(\pi)-r(\pi)/2} \\ &\leq \left(N^{-1/2} + \phi_N^2\right)^q. \end{aligned} \tag{S4.152}$$

Since the number of index-equality patterns of $\mu_1, \nu_1, \ldots, \mu_{2q}, \nu_{2q}$ depends only on q, combining (S4.151) and (S4.152), summing over all contributing patterns yields

$$\mathbb{E}\left| N^{-1}\left\{\boldsymbol{y}_m \mathbf{A} \boldsymbol{y}_k^\top - \mathrm{tr}(\mathbf{B}_{mk}\mathbf{A})\right\}\right|^{2q} \quad \leq C_q\left(N^{-1/2} + \phi_N^2\right)^q. \tag{S4.153}$$

We now substitute $\mathbf{A} = \mathbf{P}_0(z)$. By the spectral confinement of the non-spiked part and the separation of $\mathcal{D}$ from the non-spiked bulk support, there exists a constant $C_0 > 0$ such that, with

$$\mathcal{E}_N \coloneqq \Big\{ \sup_{z\in\mathcal{D}} \|\mathbf{P}_0(z)\| \le C_0 \Big\},$$

we have $\sum_{N=1}^{\infty} \mathbb{P}(\mathcal{E}_N^c) < \infty$. Conditioning on $\mathbf{Y}_0$, the resolvent $\mathbf{P}_0(z)$ is deterministic with respect to the spiked rows. Hence, on the event $\mathcal{E}_N$, applying (S4.153) with $\mathbf{A} = \mathbf{P}_0(z)$ gives, for each fixed $z \in \mathcal{D}$,

$$\mathbb{E}\left[\left| N^{-1} \left\{ \boldsymbol{y}_m \mathbf{P}_0(z) \boldsymbol{y}_k^\top - \operatorname{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z)\} \right\} \right|^{2q} \middle| \mathbf{Y}_0 \right] \le C_q \left(N^{-1/2} + \phi_N^2 \right)^q.$$

Since $\phi_N \le N^{-c_\phi}$, we have

$$N^{-1/2} + \phi_N^2 \le 2N^{-\alpha}, \qquad \alpha \coloneqq \min\{1/2, 2c_\phi\} > 0.$$

Therefore, by Markov's inequality, for any fixed $\eta > 0$,

$$\mathbb{P}\left(\left| N^{-1} \left[\boldsymbol{y}_m \mathbf{P}_0(z) \boldsymbol{y}_k^\top - \operatorname{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z)\} \right] \right| > \eta,\ \mathcal{E}_N \right) \le C_{q,\eta} \left(N^{-1/2} + \phi_N^2 \right)^q \le C_{q,\eta} N^{-\alpha q}.$$

To obtain a uniform bound over $\mathcal{D}$, let $\mathcal{N}_\delta \subset \mathcal{D}$ be a finite δ-net with $|\mathcal{N}_\delta| \le C\delta^{-2}$. Take $\delta = N^{-L}$ for a fixed large constant $L > 0$. Then $|\mathcal{N}_\delta| \le CN^{2L}$, and hence

$$\mathbb{P}\left(\max_{z_0\in\mathcal{N}_\delta} \left| N^{-1} \left[\boldsymbol{y}_m \mathbf{P}_0(z_0) \boldsymbol{y}_k^\top - \operatorname{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z_0)\} \right] \right| > \eta,\ \mathcal{E}_N \right) \le C_{q,\eta} N^{2L-\alpha q}.$$

Choose q sufficiently large such that $\alpha q > 2L + 2$. Then

$$\sum_{N=1}^{\infty} \mathbb{P}\left(\max_{z_0\in\mathcal{N}_\delta} \left| N^{-1} \left[\boldsymbol{y}_m \mathbf{P}_0(z_0) \boldsymbol{y}_k^\top - \operatorname{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z_0)\} \right] \right| > \eta,\ \mathcal{E}_N \right) < \infty. \tag{S4.154}$$

It remains to control the discretization error. By the same high-moment argument, applying (S4.141) to $\sum_{t=1}^{n} \|N^{-1/2}\boldsymbol{\zeta}_{mt}\|^2 = N^{-1}\|\boldsymbol{y}_m\|^2$,

$$N^{-1}\|\boldsymbol{y}_m\|^2 = O_{a.s.}(1), \qquad N^{-1}\|\boldsymbol{y}_k\|^2 = O_{a.s.}(1),$$

uniformly over the finite set $\mathcal{M}$. Moreover,

$$N^{-1}\|\mathbf{B}_{mk}\|_* \leq N^{-1}\mathrm{rank}(\mathbf{B}_{mk})\|\mathbf{B}_{mk}\| \leq \|\mathbf{B}_{mk}\| \leq C,$$

where $\|\cdot\|_*$ denotes the nuclear norm. Thus, on an almost sure event and on $\mathcal{E}_N$, combining the resolvent identity

$$\mathbf{P}_0(z) - \mathbf{P}_0(z_0) = (z - z_0)\mathbf{P}_0(z)\mathbf{P}_0(z_0),$$

for any $z \in \mathcal{D}$, choose $z_0 \in \mathcal{N}_\delta$ such that $|z - z_0| \leq \delta = N^{-L}$,

$$\begin{aligned}
&\Big|N^{-1}\big[\boldsymbol{y}_m\mathbf{P}_0(z)\boldsymbol{y}_k^\top - \mathrm{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z)\}\big] - N^{-1}\big[\boldsymbol{y}_m\mathbf{P}_0(z_0)\boldsymbol{y}_k^\top - \mathrm{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z_0)\}\big]\Big| \\
&\quad \leq |z - z_0|N^{-1}\|\mathbf{P}_0(z)\|\|\mathbf{P}_0(z_0)\|\left(\|\boldsymbol{y}_m\|\|\boldsymbol{y}_k\| + \|\mathbf{B}_{mk}\|_*\right) \\
&\quad \leq C|z - z_0| \leq CN^{-L} = o(1). \qquad \text{(S4.155)}
\end{aligned}$$

Combining (S4.154), (S4.155), and $\sum_N \mathbb{P}(\mathcal{E}_N^c) < \infty$, the Borel–Cantelli lemma implies that, for every fixed $\eta > 0$,

$$\limsup_{N\to\infty} \sup_{z\in\mathcal{D}} \Big|N^{-1}\big[\boldsymbol{y}_m\mathbf{P}_0(z)\boldsymbol{y}_k^\top - \mathrm{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z)\}\big]\Big| \leq \eta \quad a.s.$$

Since $\eta > 0$ is arbitrary, taking a countable sequence of rational $\eta \downarrow 0$ gives, for fixed $m, k \in \mathcal{M}$,

$$\sup_{z\in\mathcal{D}} \Big|N^{-1}\big[\boldsymbol{y}_m\mathbf{P}_0(z)\boldsymbol{y}_k^\top - \mathrm{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z)\}\big]\Big| = o_{a.s.}(1).$$

Since $M = |\mathcal{M}|$ is fixed, taking a finite union over $m, k \in \mathcal{M}$ preserves the almost sure convergence. Hence

$$\max_{m,k\in\mathcal{M}} \sup_{z\in\mathcal{D}} \left| N^{-1}\big[\boldsymbol{y}_m \mathbf{P}_0(z) \boldsymbol{y}_k^\top - \operatorname{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z)\}\big] \right| = o_{a.s.}(1).$$

Step 3. Weighted dual-resolvent deterministic equivalent and the threshold function.

Recall that

$$\mathbf{P}_0(z) = (\widetilde{\mathbf{Q}}_0 - z\mathbf{I}_N)^{-1}.$$

A deterministic equivalent for $\mathbf{P}_0(z)$ is given by

$$\mathcal{P}_0(z) = \{-z(\mathbf{I}_N + s_{1,N}(z)\mathbf{B}_0)\}^{-1}, \tag{S4.156}$$

where $s_{1,N}(z)$ and $s_{2,N}(z)$ solve the self-consistent equations, defined in (S4.128). Let $\mathcal{D}$ be a deterministic compact subset of $\mathbb{C} \setminus \{0\}$ separated from the non-spiked bulk support, defined in (S4.137). By the separable representation of $\mathbf{Y}_0$, $\widetilde{\mathbf{Q}}_0$ is the dual sample covariance matrix in the separable covariance model with left covariance $\mathbf{A}_0$, right covariance $\mathbf{B}_0$, and aspect ratio $c_N = p_0/N$. Hence the anisotropic local law for separable covariance matrices (Ding and Yang, 2021, Theorem S.3.9) applies to $\widetilde{\mathbf{Q}}_0$. Therefore, for any deterministic unit vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^N$, the estimate

$$\left|\boldsymbol{u}^*\{\mathbf{P}_0(z) - \mathcal{P}_0(z)\}\boldsymbol{v}\right| \prec \phi_N + \Psi_N(z) \tag{S4.157}$$

holds uniformly for $z \in \mathcal{D}$, where ϕ_N is the bounded-support sequence in Assumption 5, and $\Psi_N(z)$ denotes the control parameter in the anisotropic local law. Here and below, for two sequences of nonnegative random variables X_N and Y_N, the notation $X_N \prec Y_N$ means stochastic domination (Ding and Yang, 2021, Definition 3.4), namely, for any $\varepsilon > 0$ and $D > 0$, there exists $N_0 = N_0(\varepsilon, D)$ such that

$$\mathbb{P}\left(X_N > N^{\varepsilon} Y_N\right) \leq N^{-D}, \qquad N \geq N_0. \tag{S4.158}$$

Since $\mathcal{D}$ is separated from the deterministic bulk support, the control parameter in the anisotropic local law is uniformly of polynomial order on $\mathcal{D}$. Together with $\phi_N \leq N^{-c_\phi}$, there exists a constant $c_* > 0$ such that

$$\epsilon_N := \sup_{z \in \mathcal{D}} \{\phi_N + \Psi_N(z)\} \leq N^{-c_*}$$

for all sufficiently large N. Fix $0 < \varepsilon < c_*/2$. Then $N^{\varepsilon}\epsilon_N = o(1)$. Moreover, for every $D > 0$,

$$\mathbb{P}\Big(\sup_{z \in \mathcal{D}} \big|\boldsymbol{u}^* \{\mathbf{P}_0(z) - \mathcal{P}_0(z)\} \boldsymbol{v}\big| > N^{\varepsilon} \epsilon_N\Big) \leq N^{-D} \tag{S4.159}$$

for deterministic unit vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^N$, uniformly in their choices.

For each $m \in [1 : M]$, recall the definitions of $\mathbf{B}_0$ in (S4.122) and $\boldsymbol{\Gamma}_m$ in (S4.124). Both matrices are block diagonal with respect to the same population partition. Hence they commute, and $\boldsymbol{\Gamma}_m \mathbf{B}_0$ is a deterministic symmetric nonnegative definite matrix. Moreover, and $\max_{1 \leq m \leq M, 1 \leq r \leq \tau} \sigma_m^{(r)} =$

$O(1)$,

$$\|\boldsymbol{\Gamma}_m\mathbf{B}_0\| \le C, \quad N^{-1}\|\boldsymbol{\Gamma}_m\mathbf{B}_0\|_* = N^{-1}\operatorname{tr}(\boldsymbol{\Gamma}_m\mathbf{B}_0) \le C, \quad m \in [1:M]. \tag{S4.160}$$

Denote the spectral decomposition of $\boldsymbol{\Gamma}_m\mathbf{B}_0$ by

$$\boldsymbol{\Gamma}_m\mathbf{B}_0 = \sum_{\ell=1}^{N} b_{m\ell}\boldsymbol{\xi}_{m\ell}\boldsymbol{\xi}_{m\ell}^*, \qquad b_{m\ell} \ge 0, \qquad \|\boldsymbol{\xi}_{m\ell}\| = 1. \tag{S4.161}$$

Since $M = |\mathcal{M}|$ is fixed, the deterministic family

$$\mathcal{V}_N = \{\boldsymbol{\xi}_{m\ell} : 1 \le m \le M,\ 1 \le \ell \le N\}$$

has cardinality at most $MN = O(N)$. Applying (S4.159) to all vectors in $\mathcal{V}_N$, and choosing the probability exponent $D > 3$, the union bound gives

$$\sum_{N=1}^{\infty} \mathbb{P}\Big(\max_{1\le m\le M}\max_{1\le \ell\le N}\sup_{z\in\mathcal{D}}\big|\boldsymbol{\xi}_{m\ell}^*\{\mathbf{P}_0(z) - \mathcal{P}_0(z)\}\boldsymbol{\xi}_{m\ell}\big| > N^{\varepsilon}\epsilon_N\Big) < \infty.$$

Therefore, by the Borel–Cantelli lemma,

$$\max_{1\le m\le M}\max_{1\le \ell\le N}\sup_{z\in\mathcal{D}}\big|\boldsymbol{\xi}_{m\ell}^*\{\mathbf{P}_0(z) - \mathcal{P}_0(z)\}\boldsymbol{\xi}_{m\ell}\big| = o_{a.s.}(1). \tag{S4.162}$$

Consequently, for each $m \in [1:M]$,

$$\begin{aligned}
&\sup_{z\in\mathcal{D}}\Big|N^{-1}\operatorname{tr}\big[\boldsymbol{\Gamma}_m\mathbf{B}_0\{\mathbf{P}_0(z) - \mathcal{P}_0(z)\}\big]\Big| \\
&\quad = \sup_{z\in\mathcal{D}}\Big|N^{-1}\sum_{\ell=1}^{N} b_{m\ell}\boldsymbol{\xi}_{m\ell}^*\{\mathbf{P}_0(z) - \mathcal{P}_0(z)\}\boldsymbol{\xi}_{m\ell}\Big| \\
&\quad \le N^{-1}\sum_{\ell=1}^{N} b_{m\ell}\max_{1\le \ell\le N}\sup_{z\in\mathcal{D}}\big|\boldsymbol{\xi}_{m\ell}^*\{\mathbf{P}_0(z) - \mathcal{P}_0(z)\}\boldsymbol{\xi}_{m\ell}\big|
\end{aligned}$$

$$\leq N^{-1}\|\boldsymbol{\Gamma}_m\mathbf{B}_0\|_* \max_{1\leq\ell\leq N}\sup_{z\in\mathcal{D}}\left|\boldsymbol{\xi}_{m\ell}^*\{\mathbf{P}_0(z)-\mathcal{P}_0(z)\}\boldsymbol{\xi}_{m\ell}\right| = o_{a.s.}(1),$$

where the last step follows from (S4.160) and (S4.162). Since M is fixed, this gives

$$\max_{1\leq m\leq M}\sup_{z\in\mathcal{D}}\left|N^{-1}\operatorname{tr}\{\boldsymbol{\Gamma}_m\mathbf{B}_0\mathbf{P}_0(z)\} - N^{-1}\operatorname{tr}\{\boldsymbol{\Gamma}_m\mathbf{B}_0\mathcal{P}_0(z)\}\right| = o_{a.s.}(1). \tag{S4.163}$$

Combining (S4.163) with Step 2, namely

$$\sup_{z\in\mathcal{D}}\left\|N^{-1}\mathbf{Y}_1\mathbf{P}_0(z)\mathbf{Y}_1^\top - \left[N^{-1}\operatorname{tr}\{\mathbf{B}_{mk}\mathbf{P}_0(z)\}\right]_{m,k\in\mathcal{M}}\right\| = o_{a.s.}(1), \tag{S4.164}$$

and using $\mathbf{B}_{mk} = \mathbf{1}_{\{m=k\}}\boldsymbol{\Gamma}_m\mathbf{B}_0$, we obtain

$$\sup_{z\in\mathcal{D}}\left\|N^{-1}\mathbf{Y}_1\mathbf{P}_0(z)\mathbf{Y}_1^\top - \operatorname{diag}\left[N^{-1}\operatorname{tr}\{\boldsymbol{\Gamma}_m\mathbf{B}_0\mathcal{P}_0(z)\}\right]_{1\leq m\leq M}\right\| = o_{a.s.}(1). \tag{S4.165}$$

Define, for any $z\in\mathcal{D}$,

$$\psi_{m,N}(z) \coloneqq -N^{-1}\operatorname{tr}\{\boldsymbol{\Gamma}_m\mathbf{B}_0\mathcal{P}_0(z)\}, \qquad m\in[1:M]. \tag{S4.166}$$

It follows from (S4.165) that, in terms of the matrix norm,

$$\sup_{z\in\mathcal{D}}\left\|N^{-1}\mathbf{Y}_1\mathbf{P}_0(z)\mathbf{Y}_1^\top + \operatorname{diag}\left(\psi_{m,N}(z)\right)_{m\in[1:M]}\right\| = o_{a.s.}(1). \tag{S4.167}$$

Therefore, for $D_N(z) = \det\left\{\mathbf{I}_M + N^{-1}\mathbf{Y}_1\mathbf{P}_0(z)\mathbf{Y}_1^\top\right\}$ defined in (S4.136), its deterministic approximation is given by

$$D_N^0(z) \coloneqq \det\left[\operatorname{diag}\left(1-\psi_{m,N}(z)\right)_{m\in[1:M]}\right] = \prod_{1\leq m\leq M}\{1-\psi_{m,N}(z)\}. \tag{S4.168}$$

Since the functions $\psi_{m,N}(z)$ are uniformly bounded on $\mathcal{D}$, the convergence in (S4.167) implies that $N^{-1}\mathbf{Y}_1\mathbf{P}_0(z)\mathbf{Y}_1^\top$ is also uniformly bounded on $\mathcal{D}$ almost surely. Because M is fixed, the determinant is locally Lipschitz on bounded subsets of $\mathbb{C}^{M\times M}$. Applying this continuity property to the bounded matrices yields

$$\sup_{z\in\mathcal{D}} |D_N(z) - D_N^0(z)| = o_{a.s.}(1). \tag{S4.169}$$

We now turn to the proof of the threshold criterion, which entails restricting the deterministic function to the real axis strictly beyond the bulk edge, namely $(\lambda_{+,N}, \infty)$. For $x > \lambda_{+,N}$ defined in (S4.130), the self-consistent equations admit real continuations outside the support of F_{c_N}. On the physical branch,

$$s_{1,N}(x) < 0, \qquad s_{2,N}(x) < 0.$$

Since $\mathbf{\Gamma}_m$ and $\mathbf{B}_0$ commute, they can be diagonalized simultaneously. Let $b_{01}, \dots, b_{0N}$ be the eigenvalues of $\mathbf{B}_0$, and let $\gamma_{m1}, \dots, \gamma_{mN}$ be the corresponding diagonal entries of $\mathbf{\Gamma}_m$ in the same eigenbasis. Then, for $i \in [1 : N]$, we have $b_{0i} \geq 0$ and $\gamma_{mi} \geq 0$. Moreover, by the regularity of the right edge in (S4.131) and since $x > \lambda_{+,N}$ lies in the rightmost component of the complement of the support, we have

$$1 + s_{1,N}(x) b_{0i} > 0, \qquad i \in [1 : N]. \tag{S4.170}$$

By the definition of $\mathcal{P}_0(x)$ in (S4.156),

$$\mathcal{P}_0(x) = -\frac{1}{x}(\mathbf{I}_N + s_{1,N}(x)\mathbf{B}_0)^{-1}.$$

Therefore, (S4.166) can be written as

$$\begin{aligned}\psi_{m,N}(x) &= -\frac{1}{N}\operatorname{tr}\Big[\boldsymbol{\Gamma}_m\mathbf{B}_0\big\{-\frac{1}{x}(\mathbf{I}_N + s_{1,N}(x)\mathbf{B}_0)^{-1}\big\}\Big] \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{\gamma_{mi}b_{0i}}{x\{1+s_{1,N}(x)b_{0i}\}}. \end{aligned} \tag{S4.171}$$

By (S4.170), every summand in (S4.171) is nonnegative. Since $m \in [1:M]$, the corresponding spiked component is nondegenerate, and thus $\boldsymbol{\Gamma}_m\mathbf{B}_0 \neq 0$. Hence at least one summand is positive, and

$$\psi_{m,N}(x) > 0, \qquad x > \lambda_{+,N}. \tag{S4.172}$$

The continuity of $\psi_{m,N}$ on $(\lambda_{+,N}, \infty)$ follows from the continuity of $s_{1,N}(x)$ and the positivity of the denominators in (S4.171).

It remains to prove strict monotonicity. Recall the self-consistent equations in (S4.128), on the real branch,

$$\begin{cases} s_{1,N}(x) = -\dfrac{c_N}{x}\displaystyle\int \frac{t}{1+ts_{2,N}(x)}\,\mathrm{d}F_{p_0}^{\mathbf{A}_0}(t), \\ s_{2,N}(x) = -\dfrac{1}{x}\displaystyle\int \frac{t}{1+ts_{1,N}(x)}\,\mathrm{d}F_{N}^{\mathbf{B}_0}(t). \end{cases} \tag{S4.173}$$

Define

$$\alpha_N(x) = c_N\int \frac{t^2}{\{1+ts_{2,N}(x)\}^2}\,\mathrm{d}F_{p_0}^{\mathbf{A}_0}(t),\ \ \beta_N(x) = \int \frac{t^2}{\{1+ts_{1,N}(x)\}^2}\,\mathrm{d}F_{N}^{\mathbf{B}_0}(t). \tag{S4.174}$$

Differentiating (S4.173) with respect to x gives

$$\begin{cases} s'_{1,N}(x) = -\dfrac{s_{1,N}(x)}{x} + \dfrac{\alpha_N(x)}{x} s'_{2,N}(x), \\ s'_{2,N}(x) = -\dfrac{s_{2,N}(x)}{x} + \dfrac{\beta_N(x)}{x} s'_{1,N}(x). \end{cases} \tag{S4.175}$$

Equivalently,

$$\begin{pmatrix} 1 & -\alpha_N(x)/x \\ -\beta_N(x)/x & 1 \end{pmatrix} \begin{pmatrix} s'_{1,N}(x) \\ s'_{2,N}(x) \end{pmatrix} = \begin{pmatrix} -s_{1,N}(x)/x \\ -s_{2,N}(x)/x \end{pmatrix}.$$

For $x > \lambda_{+,N}$, the real solution of the self-consistent equations is regular outside the support of F_{c_N}. In particular, the Jacobian of the self-consistent system is non-singular, or equivalently,

$$\Delta_N(x) \coloneqq 1 - \frac{\alpha_N(x)\beta_N(x)}{x^2} > 0, \qquad x > \lambda_{+,N}. \tag{S4.176}$$

Indeed, for the finite-N self-consistent system, zeros of $\Delta_N(x)$ characterize the spectral edges of F_{c_N}. Since $\lambda_{+,N}$ is the rightmost edge, no such zero occurs on $(\lambda_{+,N}, \infty)$. Solving (S4.175), we obtain

$$s'_{1,N}(x) = \frac{-s_{1,N}(x)/x - \alpha_N(x)s_{2,N}(x)/x^2}{\Delta_N(x)}. \tag{S4.177}$$

Since

$$s_{1,N}(x) < 0, \qquad s_{2,N}(x) < 0, \qquad \alpha_N(x) \geq 0, \qquad \Delta_N(x) > 0,$$

we have

$$s'_{1,N}(x) > 0, \qquad x > \lambda_{+,N}. \tag{S4.178}$$

Differentiating (S4.171), we get

$$\psi'_{m,N}(x) = -\frac{1}{N}\sum_{i=1}^{N} \frac{\gamma_{mi} b_{0i}\{1 + s_{1,N}(x)b_{0i} + x b_{0i} s'_{1,N}(x)\}}{x^2\{1 + s_{1,N}(x)b_{0i}\}^2}. \qquad \text{(S4.179)}$$

For each $i \in [1 : N]$, (S4.170) and (S4.178) imply $1+s_{1,N}(x)b_{0i}+xb_{0i}s'_{1,N}(x) > 0$. Together with $\gamma_{mi}b_{0i} \geq 0$, this shows that every nonzero summand in (S4.179) is strictly negative. Since $\mathbf{\Gamma}_m\mathbf{B}_0 \neq 0$, at least one summand is nonzero. Therefore,

$$\psi'_{m,N}(x) < 0, \qquad x > \lambda_{+,N}. \qquad \text{(S4.180)}$$

Thus $\psi_{m,N}$ is strictly decreasing on $(\lambda_{+,N}, \infty)$.

Finally, as $x \to \infty$, the self-consistent equations imply

$$s_{1,N}(x) = O(x^{-1}), \qquad s_{2,N}(x) = O(x^{-1}).$$

Hence, for all sufficiently large x,

$$1 + s_{1,N}(x)b_{0i} \geq \frac{1}{2}, \qquad i \in [1 : N].$$

Using (S4.171) and (S4.160), we obtain

$$0 < \psi_{m,N}(x) \leq \frac{2}{xN}\sum_{i=1}^{N} \gamma_{mi}b_{0i} = \frac{2}{xN}\,\mathrm{tr}(\mathbf{\Gamma}_m\mathbf{B}_0) \leq \frac{C}{x} \to 0.$$

Therefore, we obtain

$$\lim_{x\to\infty} \psi_{m,N}(x) = 0. \qquad \text{(S4.181)}$$

Since $\psi_{m,N}$ is positive, continuous and strictly decreasing on $(\lambda_{+,N}, \infty)$, together with (S4.181), the equation

$$\psi_{m,N}(x) = 1$$

has at most one solution on $(\lambda_{+,N}, \infty)$. If $\lim_{x \downarrow \lambda_{+,N}} \psi_{m,N}(x) > 1$, then the intermediate value theorem yields a unique solution $\theta_{m,N} > \lambda_{+,N}$ such that $\psi_{m,N}(\theta_{m,N}) = 1$. Conversely, if such a solution $\theta_{m,N}$ exists, then strict monotonicity gives

$$\lim_{x \downarrow \lambda_{+,N}} \psi_{m,N}(x) > \psi_{m,N}(\theta_{m,N}) = 1.$$

Therefore, the equation $\psi_{m,N}(x) = 1$ admits a unique solution $\theta_{m,N} > \lambda_{+,N}$ if and only if $\lim_{x \downarrow \lambda_{+,N}} \psi_{m,N}(x) > 1$. Accordingly, define the set of supercritical indices within $\mathcal{M}$ as

$$\mathcal{M}_+ = \Big\{ 1 \le m \le M : \lim_{x \downarrow \lambda_{+,N}} \psi_{m,N}(x) > 1 \Big\}, \tag{S4.182}$$

and define subcritical set

$$\mathcal{M}_0 = \mathcal{M} \setminus \mathcal{M}_+. \tag{S4.183}$$

This proves Theorem 2.

Step 4. Rouché localization of outlier eigenvalues.

Recall from (S4.169) and (S4.168) that

$$\sup_{z \in \mathcal{D}} |D_N(z) - D_N^0(z)| = o_{a.s.}(1), \qquad D_N^0(z) = \prod_{1 \le m \le M} \{1 - \psi_{m,N}(z)\}.$$

We utilize this deterministic approximation to localize the zeros of $D_N(z)$ within a compact set $\mathcal{D}$, separated from the non-spiked bulk support. Let $\theta_{1,N} > \cdots > \theta_{L,N} > \lambda_{+,N}$ be the distinct deterministic roots of the equations

$$\psi_{m,N}(x) = 1, \qquad m \in \mathcal{M}_+,$$

where $\mathcal{M}_+$ is defined in (S4.182). For each $\ell \in [1 : L]$, define

$$\mathcal{M}_\ell = \{m \in \mathcal{M}_+ : \theta_{m,N} = \theta_{\ell,N}\}, \quad q_\ell = |\mathcal{M}_\ell|, \quad \bigcup_{\ell=1}^{L} \mathcal{M}_\ell = \mathcal{M}_+.$$

By definition, $1 - \psi_{m,N}(\theta_{\ell,N}) = 0$ for $m \in \mathcal{M}_\ell$, and $1 - \psi_{m,N}(\theta_{\ell,N}) \neq 0$ for $m \notin \mathcal{M}_\ell$. Consequently, D_N^0 vanishes at $\theta_{\ell,N}$.

We next demonstrate the nondegeneracy of the corresponding factors, which establishes that the algebraic multiplicity of this zero is exactly q_ℓ. We assume that there exists a constant $c_0 > 0$ such that,

$$\min_{1\le\ell\le L}(\theta_{\ell,N} - \lambda_{+,N}) \ge c_0, \qquad \min_{\ell\neq k}|\theta_{\ell,N} - \theta_{k,N}| \ge c_0, \tag{S4.184}$$

for all sufficiently large N. Choose a fixed $\varepsilon \in (0, c_0/3)$, then,

$$0 < \varepsilon < \frac{1}{3}\min\Big\{\min_{1\le\ell\le L}(\theta_{\ell,N} - \lambda_{+,N}),\, \min_{\ell\neq k}|\theta_{\ell,N} - \theta_{k,N}|\Big\},$$

for all sufficiently large N. For each $\ell \in [1 : L]$, define the open disk and its boundary by

$$\mathbb{D}_{\ell,N}(\varepsilon) = \{z : |z - \theta_{\ell,N}| < \varepsilon\}, \qquad \Upsilon_{\ell,N}(\varepsilon) = \partial\mathbb{D}_{\ell,N}(\varepsilon).$$

By (S4.184) and the uniform boundedness of the roots $\theta_{\ell,N}$, the closed disks $\overline{\mathbb{D}_{\ell,N}(\varepsilon)}$ are mutually disjoint and contained in a fixed compact set that is uniformly separated from the non-spiked bulk support, for all sufficiently large N. Therefore, the uniform estimates established in Steps 2–3 apply on these contours.

For any $m \in \mathcal{M}_\ell$, $\psi_{m,N}(\theta_{\ell,N}) = 1$. Applying the derivative formula (S4.179) obtained in Step 3, in conjunction with (S4.170) and (S4.178), yields for $x > \lambda_{+,N}$,

$$|\psi'_{m,N}(x)| \geq \frac{1}{N}\sum_{i=1}^{N} \frac{\gamma_{mi} b_{0i}}{x^2\{1 + s_{1,N}(x) b_{0i}\}} = \frac{\psi_{m,N}(x)}{x}.$$

Evaluating this inequality at $x = \theta_{\ell,N}$ yields

$$|\psi'_{m,N}(\theta_{\ell,N})| \geq \frac{1}{\theta_{\ell,N}}.$$

Furthermore, the roots $\theta_{\ell,N}$ are uniformly bounded from above. Moreover, by (S4.160) and the definition of $\psi_{m,N}$, the roots $\theta_{\ell,N}$ are uniformly bounded above. Since $\psi_{m,N}(\theta_{\ell,N}) = 1$, for all $\ell \in [1 : L]$ and all sufficiently large N, there exists a constant $C > 0$ such that $\theta_{\ell,N} \leq C$. Thus, there exists a constant $c_1 > 0$, independent of m, ℓ, and N, such that

$$|\psi'_{m,N}(\theta_{\ell,N})| \geq c_1, \qquad m \in \mathcal{M}_\ell, \quad \ell \in [1 : L], \tag{S4.185}$$

for all sufficiently large N. Therefore, each factor $1 - \psi_{m,N}(z)$ with $m \in \mathcal{M}_\ell$ possesses a simple zero at $z = \theta_{\ell,N}$. Because exactly q_ℓ such factors vanish at $\theta_{\ell,N}$, the function D_N^0 has a zero at $\theta_{\ell,N}$ with algebraic multiplicity q_ℓ.

Moreover, by the analyticity of the factors $1 - \psi_{m,N}(z)$ on $\mathcal{D}$, the root separation assumption (S4.184), the nondegeneracy (S4.185), and the fact that M is fixed, D_N^0 has no zero on the boundaries $\Upsilon_{\ell,N}(\varepsilon)$. Hence, after possibly decreasing ε, there exists a constant $c_\varepsilon > 0$ such that, for all sufficiently large N,

$$\min_{1\leq \ell \leq L} \inf_{z\in\Upsilon_{\ell,N}(\varepsilon)} |D_N^0(z)| \geq c_\varepsilon. \tag{S4.186}$$

Combining (S4.169) and (S4.186), we obtain, almost surely for all sufficiently large N and any $\ell \in [1 : L]$,

$$\sup_{z\in\Upsilon_{\ell,N}(\varepsilon)} |D_N(z) - D_N^0(z)| < \inf_{z\in\Upsilon_{\ell,N}(\varepsilon)} |D_N^0(z)|.$$

By Rouché's theorem, D_N and D_N^0 share the same number of zeros in $\mathbb{D}_{\ell,N}(\varepsilon)$, counting multiplicities. Hence, almost surely for all sufficiently large N,

$$N_{\mathbb{D}_{\ell,N}(\varepsilon)}(D_N) = N_{\mathbb{D}_{\ell,N}(\varepsilon)}(D_N^0) = q_\ell, \qquad \ell \in [1 : L], \tag{S4.187}$$

where $N_{\mathbb{D}}(f)$ denotes the number of zeros of f in $\mathbb{D}$, counted with algebraic multiplicity. By the spectral norm bound for separable covariance matrices and the bounded spiked strengths, there exists a constant $R > 0$ such that, almost surely for all sufficiently large N, we have $\|\mathbf{G}\| \leq R$ and

$\lambda_{+,N}, \theta_{1,N}, \ldots, \theta_{L,N} \in [0, R]$. In addition, define the complement interval

$$\mathcal{J}_N(\varepsilon) = \Big([\lambda_{+,N} + \varepsilon, R] \setminus \bigcup_{\ell=1}^{L} (\theta_{\ell,N} - \varepsilon, \theta_{\ell,N} + \varepsilon)\Big), \tag{S4.188}$$

where $R > 0$ is chosen such that all eigenvalues under consideration are contained in $[0, R]$ almost surely for all sufficiently large N. By the root separation assumption and the established nondegeneracy, D_N^0 has no zeros on $\mathcal{J}_N(\varepsilon)$. Thus, there exists a constant $c'_\varepsilon > 0$ such that, for all sufficiently large N,

$$\inf_{x \in \mathcal{J}_N(\varepsilon)} |D_N^0(x)| \geq c'_\varepsilon.$$

Coupled with (S4.169), this implies that, almost surely for all sufficiently large N, D_N has no real zero in $\mathcal{J}_N(\varepsilon)$. Consequently, all real zeros of D_N in the outlier region $[\lambda_{+,N} + \varepsilon, R]$ are contained in the disjoint disks $\mathbb{D}_{\ell,N}(\varepsilon)$, for any $\ell \in [1 : L]$.

It remains to link these zeros with the eigenvalues of $\mathbf{G}$. According to the determinant equation derived in Step 1, for any $z \in \mathcal{D}$,

$$\det(\check{\mathbf{G}} - z\mathbf{I}_p) = \det(\mathbf{Q}_0 - z\mathbf{I}_{p_0})(-z)^M D_N(z).$$

Since $\mathcal{D}$ is separated from zero and from the non-spiked spectrum almost surely for all sufficiently large N, we have

$$\det(\mathbf{Q}_0 - z\mathbf{I}_{p_0}) \neq 0, \qquad (-z)^M \neq 0, \qquad \text{for any } z \in \mathcal{D}.$$

Thus, the zeros of D_N in $\mathcal{D}$ precisely coincide with the eigenvalues of $\check{\mathbf{G}}$ located in $\mathcal{D}$, counting multiplicities.

Since $\mathbf{G}$ and $\check{\mathbf{G}}$ share the same spectrum, the same conclusion holds for $\mathbf{G}$. Therefore, almost surely, for all sufficiently large N, the matrix $\mathbf{G}$ has exactly q_ℓ eigenvalues in $\mathbb{D}_{\ell,N}(\varepsilon)$, for each $\ell \in [1 : L]$. The preceding exclusion of real zeros outside the disks shows that there are no additional eigenvalues of $\mathbf{G}$ in $[\lambda_{+,N} + \varepsilon, R]$. Since $\theta_{1,N} > \cdots > \theta_{L,N} > \lambda_{+,N}$ and the disks are mutually disjoint, these eigenvalues are ordered according to the deterministic roots. Hence the eigenvalues in $\mathbb{D}_{\ell,N}(\varepsilon)$ correspond to the sample ranks

$$\mathcal{I}_\ell = \{\omega_{\ell-1} + 1, \dots, \omega_\ell\}, \qquad \omega_0 = 0, \qquad \omega_\ell = \sum_{t=1}^{\ell} q_t.$$

We conclude that, almost surely for all sufficiently large N,

$$\max_{i \in \mathcal{I}_\ell} |\lambda_i(\mathbf{G}) - \theta_{\ell,N}| \le \varepsilon, \qquad \ell \in [1 : L].$$

Taking a sequence of rational values $\varepsilon \downarrow 0$ yields the final result:

$$\max_{i \in \mathcal{I}_\ell} |\lambda_i(\mathbf{G}) - \theta_{\ell,N}| \xrightarrow{a.s.} 0, \qquad \ell \in [1 : L]. \tag{S4.189}$$

This completes the proof of the outlier component (i) of Theorem 3.

Step 5. Edge sticking of non-outlier eigenvalues.

It remains to prove that the eigenvalues of $\mathbf{G}$ which are not contained in the separated outlier clusters stick to the right edge $\lambda_{+,N}$.

We first control the right edge of the non-spiked reference matrix $\mathbf{Q}_0 = N^{-1}\mathbf{Y}_0\mathbf{Y}_0^\top$. Let F_{c_N} be the deterministic spectral distribution associated with $\mathbf{Q}_0$, whose rightmost edge is $\lambda_{+,N}$. For $j \geq 1$, define the classical location $\chi_{j,N}$, counted from the right edge, by

$$1 - F_{c_N}(\chi_{j,N}) = \frac{j}{p_0}. \tag{S4.190}$$

By the eigenvalue rigidity theorem for non-spiked separable covariance matrices (Ding and Yang, 2021, Theorem S.3.11), applied to the $\mathbf{Q}_0$, the following high-probability estimate holds: for every fixed integer ϖ, there exists a deterministic sequence $\delta_N \downarrow 0$ such that, for every $D > 0$,

$$\mathbb{P}\Big(\max_{1\leq j\leq\varpi} |\lambda_j(\mathbf{Q}_0) - \chi_{j,N}| > \delta_N \Big) \leq N^{-D} \tag{S4.191}$$

for all sufficiently large N.

We now extend this high-probability estimate to almost sure convergence. Given any $D > 2$, the summability $\sum_{N=1}^{\infty} N^{-D} < \infty$, combined with (S4.191) and the Borel–Cantelli lemma, ensures that the event $\max_{1\leq j\leq\varpi} |\lambda_j(\mathbf{Q}_0) - \chi_{j,N}| > \delta_N$ occurs only finitely many times. Consequently, $\max_{1\leq j\leq\varpi} |\lambda_j(\mathbf{Q}_0) - \chi_{j,N}| \leq \delta_N$ holds almost surely for all sufficiently large N. Since $\delta_N \downarrow 0$, we obtain

$$\max_{1\leq j\leq\varpi} |\lambda_j(\mathbf{Q}_0) - \chi_{j,N}| = o_{a.s.}(1). \tag{S4.192}$$

Denote by ρ_{c_N} the density of F_{c_N} on the rightmost component of its support.

By the square-root behavior of the deterministic density at the regular rightmost edge (Ding and Yang, 2021, Lemma S.3.4), there exist constants $C_1, C_2, \delta > 0$ such that

$$C_1\sqrt{\lambda_{+,N} - x} \leq \rho_{c_N}(x) \leq C_2\sqrt{\lambda_{+,N} - x}, \qquad x \in [\lambda_{+,N} - \delta, \lambda_{+,N}],$$

for all sufficiently large N. For fixed j, we have

$$\frac{j}{p_0} = \int_{\chi_{j,N}}^{\lambda_{+,N}} \rho_{c_N}(x)\,\mathrm{d}x \asymp (\lambda_{+,N} - \chi_{j,N})^{3/2}.$$

Therefore, there exists a constant $C > 0$ such that

$$\max_{1\leq j\leq \varpi} |\chi_{j,N} - \lambda_{+,N}| \leq C\Big(\frac{\varpi}{p_0}\Big)^{2/3} = o(1), \tag{S4.193}$$

where $p_0 \asymp N$. Combining (S4.192) and (S4.193) gives

$$\max_{1\leq j\leq \varpi} |\lambda_j(\mathbf{Q}_0) - \lambda_{+,N}| = o_{a.s.}(1). \tag{S4.194}$$

We next prove the upper bound for the non-outlier eigenvalues of $\mathbf{G}$. Let $\eta > 0$ be fixed. Choose the radius $\varepsilon > 0$ in Step 4 so that $0 < \varepsilon < \min\{\eta, c_0/3\}$. By Step 4, almost surely for all sufficiently large N, all real zeros of D_N in $[\lambda_{+,N} + \varepsilon, R]$ are contained in the disjoint disks $\mathbb{D}_{\ell,N}(\varepsilon)$ for any $\ell \in [1 : L]$, and the total number of eigenvalues of $\mathbf{G}$ in these disks is $K = \sum_{\ell=1}^{L} q_\ell$. Since

$$[\lambda_{+,N} + \eta, R] \subset [\lambda_{+,N} + \varepsilon, R],$$

there are no additional eigenvalues of $\mathbf{G}$ above $\lambda_{+,N} + \eta$ outside the K separated outlier eigenvalues. Therefore, almost surely for all sufficiently large N, there are at most K eigenvalues of $\mathbf{G}$ larger than $\lambda_{+,N} + \eta$. Hence

$$\lambda_{K+1}(\mathbf{G}) \leq \lambda_{+,N} + \eta \tag{S4.195}$$

almost surely for all sufficiently large N. It remains to obtain the lower bound. Since $\mathbf{Q}_0$ is a principal submatrix of $\check{\mathbf{G}}$, Cauchy's interlacing theorem gives

$$\lambda_j(\check{\mathbf{G}}) \geq \lambda_j(\mathbf{Q}_0), \qquad 1 \leq j \leq p_0.$$

Since $\mathbf{G}$ and $\check{\mathbf{G}}$ have the same eigenvalues, we have

$$\lambda_j(\mathbf{G}) \geq \lambda_j(\mathbf{Q}_0), \qquad 1 \leq j \leq p_0.$$

By (S4.194), let $\varrho_N = \max_{1\leq j\leq\varpi} |\lambda_j(\mathbf{Q}_0) - \lambda_{+,N}|$, we obtain $\varrho_N = o_{a.s.}(1)$. Therefore, for $j \in \{K+1, \ldots, \varpi\}$,

$$\lambda_j(\mathbf{G}) \geq \lambda_j(\mathbf{Q}_0) \geq \lambda_{+,N} - \varrho_N = \lambda_{+,N} - o_{a.s.}(1).$$

Together with (S4.195), we have, for $j \in \{K+1, \ldots, \varpi\}$,

$$\lambda_{+,N} - o_{a.s.}(1) \leq \lambda_j(\mathbf{G}) \leq \lambda_{+,N} + \eta$$

almost surely for all sufficiently large N. Since $\eta > 0$ is arbitrary, taking a sequence of rational values $\eta \downarrow 0$ gives

$$\max_{K+1\leq j\leq\varpi} |\lambda_j(\mathbf{G}) - \lambda_{+,N}| \xrightarrow{a.s.} 0. \tag{S4.196}$$

This proves the edge sticking part (ii) of Theorem 3.

S5 Preprocessing Strategies for G

In practice, the assumptions of the baseline model often fail to hold due to heterogeneous non-spiked components, i.e., $\mathbf{D}_r \neq \mathbf{D}_u$ for all $r \neq u \in [1:\tau]$. To address this more general setting, inspired by Ledoit and Wolf (2020), we adopt two preprocessing strategies depending on the value of c and formalized in Corollary 1. After preprocessing, $\widetilde{\mathbf{G}}$ conforms to the baseline structure, and the theoretical results remain consistent with Theorem 3.

Corollary 1. *Given the rth row-population, define the regularized sample covariance matrix $\widetilde{\mathbf{S}}_j^{(r)}$ by*

$$\widetilde{\mathbf{S}}_j^{(r)} = \begin{cases} (1-\gamma_r)\mathbf{S}_j^{(r)} + \gamma_r \mathbf{I}_p, & \textit{for } c < 1, \\ \varsigma_r^{-2}\mathbf{S}_j^{(r)} = \mathbf{U}_r \left[\dfrac{\operatorname{diag}\left(\lambda_1(\mathbf{S}_j^{(r)}), \lambda_2(\mathbf{S}_j^{(r)}), \ldots, \lambda_p(\mathbf{S}_j^{(r)})\right)}{\varsigma_r^2} \right] \mathbf{U}_r^\top, & \textit{for } c \geq 1, \end{cases} \tag{S5.197}$$

where shrinkage coefficient $0 < \gamma_r < 1$, scaling parameter $\varsigma_r^2 > 0$, and the symbol $\operatorname{diag}(\cdot)$ *denotes the diagonal matrix.*

Remark 3. When the baseline assumptions are violated and $\widetilde{\mathbf{G}}$ is regularized as in (S5.197), the mean and covariance terms $\left(\mu_{f_0}^{(r)}, \nu_{f_0}^{(r)}\right)$ in (2.13) should be adjusted accordingly:

$$\left(\tilde{\mu}_{f_0}^{(r)}, \tilde{\nu}_{f_0}^{(r)}\right) = \begin{cases} \left((1-\gamma_r)\mu_{f_0}^{(r)} + \gamma_r,\ (1-\gamma_r)^2\nu_{f_0}^{(r)}\right), & \text{for} \quad c < 1, \\ \left(\hat{\varsigma}_r^{-2}\mu_{f_0}^{(r)},\ \hat{\varsigma}_r^{-4}\nu_{f_0}^{(r)}\right), & \text{for} \quad c \geq 1. \end{cases}$$

The above regularization stabilizes the sample spectrum and improves the bulk fitting to the M-P law. Specifically, when $c < 1$, linear shrinkage contracts $\mathbf{S}_j^{(r)}$ toward $\mathbf{I}_p$, thereby damping estimation noise in the eigenvalues. When $c \geq 1$, global scaling calibrates the non-spiked bulk to follow the M-P law while preserving the relative magnitudes of the spiked ones. In finite samples, however, close spikes may cluster near the bulk edge and thus be practically indistinguishable from the noise spectrum. This section

---

Algorithm 1 Linear shrinkage for $c < 1$

Require: Sample covariance matrices $\{\mathbf{S}_j^{(r)}\}$ for $r = 1\ldots,\tau,\ j = 1\ldots,h_r$; number of folds K_f; grid step size d_γ.

Ensure: Shrinking covariance matrix $\widetilde{\mathbf{G}}$; shrinkage coefficient $\{\hat{\gamma}_r\}_{r=1}^{\tau}$.

$\Theta \leftarrow \{d_\gamma, 2d_\gamma, \ldots, 1-d_\gamma\}$
for $r = 1$ to τ **do**
  Partition samples $\{\mathbf{S}_j^{(r)}\}$ into K disjoint folds
  Select optimal parameter via cross-validation:

$$\hat{\gamma}_r \leftarrow \arg\max_{\gamma\in\Theta} \frac{1}{K_f}\sum_{k=1}^{K_f}\mathcal{L}_k(\gamma)$$

  for $j = 1$ to h_r **do**
    Shrinking covariance matrix: $\widetilde{\mathbf{S}}_j^{(r)} \leftarrow (1-\hat{\gamma}_r)\mathbf{S}_j^{(r)} + \hat{\gamma}_r\mathbf{I}_p$
  end for
end for
Calculate shrinking image covariance matrix: $\widetilde{\mathbf{G}} \leftarrow \frac{1}{h}\sum_{r=1}^{\tau}\sum_{j=1}^{h_r}\widetilde{\mathbf{S}}_j^{(r)}$
return $\widetilde{\mathbf{G}}$ and $\{\hat{\gamma}_r\}$

---

describes the computational procedures for the shrinkage coefficient γ_r and the scaling parameter ς_r^2 introduced in Corollary 1 for situations where the assumptions of the baseline model are violated. First, the shrinkage coefficient $\hat{\gamma}_r \in (0,1)$ is determined via K-fold cross-validation by maximizing the log-likelihood function

$$\log \mathcal{L}(\gamma_r) = -\frac{n}{2}\Big\{p\log(2\pi) + \log\det(\widetilde{\mathbf{S}}_j^{(r)})\Big\} - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{w}_i^{(r)})^\top(\widetilde{\mathbf{S}}_j^{(r)})^{-1}\boldsymbol{w}_i^{(r)}.$$

The complete estimation procedure is summarized in the Algorithm 1. Second, leveraging the robustness of the sample median against extreme eigen-

Algorithm 2 Global scaling for $c \geq 1$

Require: Sample covariance matrices $\{\mathbf{S}_j^{(r)}\}$ for $r = 1\ldots,\tau,\ j = 1\ldots,h_r$; truncation parameter d; M-P bulk edges $\lambda_\pm$; the M-P median $m(c)$.
Ensure: Scaled image covariance matrix $\widetilde{\mathbf{G}}$; scaling parameters $\{\hat{\varsigma}_r^2\}_{r=1}^{\tau}$.
Set empirical spiked threshold: $\lambda_{\text{thr}} \leftarrow d\lambda_+$
for $r = 1$ to τ **do**
  for $j = 1$ to h_r **do**
    Obtain eigenvalues $\lambda_{ji}^{(r)} \leftarrow \lambda_i(\mathbf{S}_j^{(r)})$, sorted in descending order;
    Identify empirical spiked indices: $\mathcal{I}_{\text{emp}}^{(r)} \leftarrow \{i \mid \lambda_{ji}^{(r)} > \lambda_{\text{thr}}\}$
    if bulk eigenvalues exist **then**
      Extract bulk eigenvalues: $\lambda_{\text{bulk}} \leftarrow \{\lambda_i \mid i \notin \mathcal{I}_{\text{emp}}^{(r)}\}$
      Estimate empirical median: $m_{\text{emp}} \leftarrow \text{median}(\lambda_{\text{bulk}})$
      Estimate scaling factor: $\hat{\varsigma}_r^2 \leftarrow m_{\text{emp}}/m(c)$
    end if
    Scaled covariance matrix: $\widetilde{\mathbf{S}}_j^{(r)} \leftarrow \mathbf{S}_j^{(r)}/\hat{\varsigma}_r^2$
  end for
end for
Calculate scaled image covariance matrix: $\widetilde{\mathbf{G}} \leftarrow 1/h\sum_{r=1}^{\tau}\sum_{j=1}^{h_r}\widetilde{\mathbf{S}}_j^{(r)}$
return $\widetilde{\mathbf{G}}, \{\hat{\varsigma}_r^2\}$

values, we estimate ς_r^2 by the ratio of the empirical median of the non-spiked eigenvalues to its theoretical counterpart under the Marčenko-Pastur (M-P) law:

$$\hat{\varsigma}_r^2 = \frac{\mathrm{median}\{\lambda_{bulk}^{(r)}\}}{m(c)},$$

where $m(c)$ denotes the theoretical median of the M-P law, defined as the solution to the equation

$$\int_{(1-\sqrt{c})^2}^{m(c)} \frac{1}{2\pi c x}\sqrt{(\lambda_+ - x)(x-\lambda_-)}\, dx = 0.5 - \max(0, 1-1/c),$$

with $\lambda_\pm$ being the lower and upper spectral edges of the M-P law, $\lambda_\pm = (1 \pm \sqrt{c})^2$.

To further robustify the estimator $\hat{\varsigma}_r^2$, we introduce a truncation parameter d to establish a threshold λ_{thr}. This strategy mitigates potential bias by excluding distant spikes while retaining those proximal to the bulk to stabilize the scale calibration. The complete procedure is formalized in Algorithm 2.

S6 Proof of Theorem 4

Recall the empirical process $\boldsymbol{\mathcal{G}}_n(f_0)$ associated with $\mathbf{G}$, defined in (2.7) under the linear function $f_0(x) = x$. For the first term in (4.20), we have

$$\sum_{i=1}^{p} \lambda_i(\mathbf{G}) = \frac{1}{h}\boldsymbol{\mathcal{G}}_n(f_0) - \frac{p}{h}\sum_{r=1}^{\tau} h_r \int x F^{c,H^{(r)}}(\mathrm{d}x). \tag{S6.198}$$

By Theorem 1, we have

$$\boldsymbol{\mathcal{G}}_n(f_0) = h\sum_{i=1}^{p}\lambda_i(\mathbf{G}) - p\sum_{r=1}^{\tau} h_r \int xF^{c,H^{(r)}}(\mathrm{d}x) \Rightarrow \mathcal{N}(\mu_x^{\mathbf{G}}, \nu_x^{\mathbf{G}}), \quad \text{(S6.199)}$$

where $\mu_x^{\mathbf{G}}$ and $\nu_x^{\mathbf{G}}$ are given by (2.13). By the first-order limits in Theorem 3,

$$\frac{\sum_{\ell=1}^{L}\sum_{i\in\mathcal{I}_\ell}\lambda_i(\mathbf{G})}{\sum_{\ell=1}^{L} q_\ell\theta_{\ell,N}} - 1 \longrightarrow 0 \qquad \text{a.s.} \quad \text{(S6.200)}$$

Finally, under the homogeneous non-spiked baseline, we combine the equations (S6.199) and (S6.200) and use Slutsky's theorem to obtain

$$T_{x,H} = \nu_{x,H}^{-\frac{1}{2}}\left[\sum_{i=1}^{p}\lambda_i(\mathbf{G}) - \sum_{\ell=1}^{L}\sum_{i\in\mathcal{I}_\ell}\lambda_i(\mathbf{G}) - b_{x,H} - \mu_{x,H}\right] \Rightarrow \mathcal{N}(0,1),$$

where

$$b_{x,H} = p\int x\,\mathrm{d}F^{c,H^{\mathbf{\Sigma}_0}}(x) - \sum_{\ell=1}^{L} q_\ell\theta_{\ell,N}. \quad \text{(S6.201)}$$

For the hypothesis (4.19) and under $\mathcal{H}_0$, the outlier equations

$$\psi_{m,N}(x) = 1, \qquad \text{for } x > \lambda_{+,N} \text{ with } m = [1:K_0]$$

yield K_0 deterministic outlier locations, denoted by $\theta_{1,N},\ldots,\theta_{K_0,N}$, where repeated roots are counted according to their occurrences. If some of these roots coincide, they are grouped into L_0 distinct values $\{\theta_\ell\}_{L_0}^{\ell=1}$, with corresponding multiplicities $\{q_\ell\}_{L_0}^{\ell=1}$, where $\sum_{\ell=1}^{L_0} q_\ell = K_0$. Hence the deterministic outlier contribution can be written as $\sum_{\ell=1}^{L_0} q_\ell\theta_{\ell,N}$. Since the present statistic uses $f_0(x) = x$, this grouped expression is equivalently

$$\sum_{\ell=1}^{L_0} q_\ell\theta_{\ell,N} = \sum_{m=1}^{K_0}\theta_{m,N}.$$

Therefore, explicit form for the centering term in (S6.201), and we have the test statistic

$$T_{x,H}(K_0) = \nu_{x,H}^{-\frac{1}{2}} \Big[\sum_{i=1}^{p} \lambda_i(\mathbf{G}) - \sum_{m=1}^{K_0} \lambda_m(\mathbf{G}) - b_{x,H}(K_0) - \mu_{x,H} \Big], \quad \text{(S6.202)}$$

where

$$b_{x,H}(K_0) = p \int x \, \mathrm{d}F^{c,H^{\mathbf{\Sigma}_0}}(x) - \sum_{m=1}^{K_0} \theta_{m,N}.$$

Proof is completed.

Remark 4. In implementation, the spiked strengths $\sigma_m^{(r)}$ are estimated by nonlinear shrinkage methods (Ledoit and Wolf, 2020). The dependence parameter ρ_r is generally unknown in practice, and we estimate it by normalizing the averaged pairwise trace covariance within the rth row-population:

$$\widehat{\rho}_r = \frac{\frac{2}{h_r(h_r-1)} \sum_{1 \le i < j \le h_r} \mathrm{tr}\{n^{-1} \mathbf{W}_j^{(r)} (\mathbf{W}_i^{(r)})^\top\}}{\frac{1}{h_r} \sum_{j=1}^{h_r} \mathrm{tr}\{n^{-1} \mathbf{W}_j^{(r)} (\mathbf{W}_j^{(r)})^\top\}}. \quad \text{(S6.203)}$$

In computation, we use $\widehat{\rho}_r^c = \min\{1, \max\{0, \widehat{\rho}_r\}\}$. The plug-in values are used in the threshold equations.

S7 Simulation Study

S7.1 Various Scenarios in Example 1

Table 1 reports the empirical sizes for Models 1 and 2 with $\rho = 0.9$ across Scenarios 1–3 in Example 1. Under the null hypothesis $K_0 = 0$, the empirical sizes of the proposed test statistic in (2.15) remain generally well-calibrated to the significance level $\alpha = 0.05$ across the experimental conditions, for different structural configurations (h_1, h_2, h_3) of the image covariance matrix $\mathbf{G}$. Complementing these results, Figure 1 displays the empirical histograms of the standardized test statistic for Gamma populations in Model 2 with $c = 1$. The empirical distributions closely approximate the standard normal distribution $\mathcal{N}(0, 1)$, as indicated by the overlaid red dashed curves. Collectively, these results show that the proposed test maintains accurate empirical sizes across different structural configurations of the image covariance matrix $\mathbf{G}$ in the non-i.i.d. setting, providing further numerical evidence for Theorem 1.

Table 1: Empirical size with $\rho = 0.9$ for different scenarios in Example 1.

			Gaussian population						**Gamma population**					
			Model 1			**Model 2**			**Model 1**			**Model 2**		
n	c	p	Scen. 1	Scen. 2	Scen. 3	Scen. 1	Scen. 2	Scen. 3	Scen. 1	Scen. 2	Scen. 3	Scen. 1	Scen. 2	Scen. 3
300	0.5	150	0.046	0.044	0.048	0.047	0.061	0.037	0.047	0.051	0.052	0.047	0.043	0.052
	1.0	300	0.062	0.053	0.050	0.057	0.051	0.059	0.057	0.051	0.051	0.051	0.044	0.050
	1.5	450	0.047	0.049	0.055	0.046	0.049	0.049	0.056	0.058	0.049	0.038	0.047	0.061
600	0.5	300	0.041	0.050	0.056	0.047	0.047	0.046	0.048	0.055	0.042	0.051	0.049	0.051
	1.0	600	0.045	0.055	0.050	0.045	0.039	0.055	0.062	0.050	0.052	0.050	0.039	0.050
	1.5	900	0.057	0.056	0.050	0.039	0.051	0.063	0.060	0.054	0.042	0.046	0.047	0.062
900	0.5	450	0.053	0.043	0.047	0.057	0.046	0.049	0.046	0.063	0.043	0.060	0.050	0.051
	1.0	900	0.039	0.055	0.044	0.046	0.053	0.039	0.051	0.047	0.049	0.057	0.050	0.050
	1.5	1350	0.059	0.056	0.054	0.047	0.046	0.059	0.043	0.047	0.060	0.052	0.051	0.038

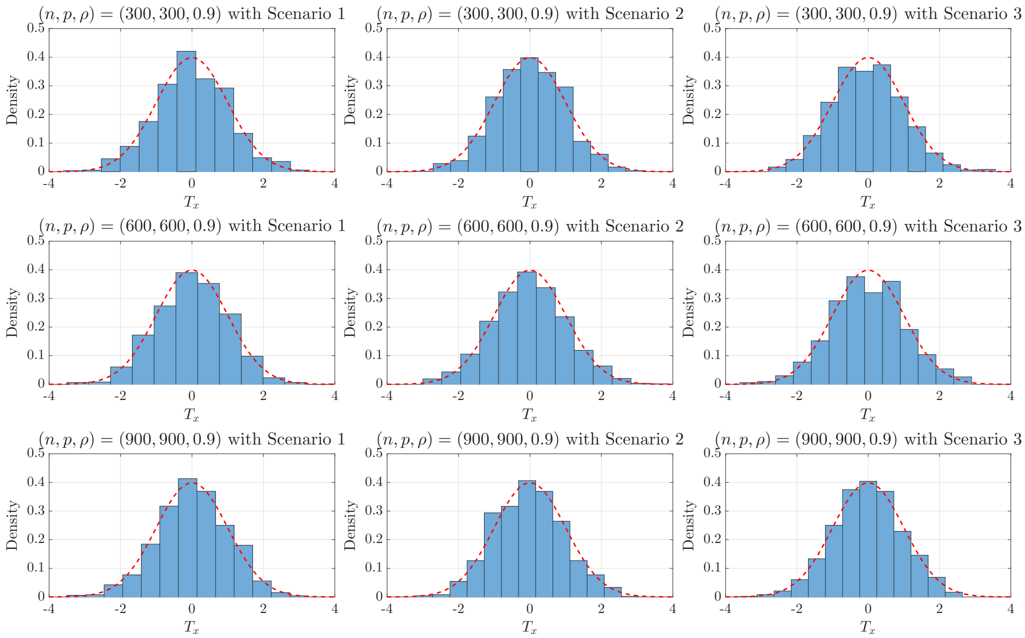

Figure 1: Empirical distributions of T_x in Model 2 under Gamma.

S7.2 Performance of the Test Statistics for PC Number Detection on Model 3

Table 2 summarizes the empirical sizes for the sequential testing procedure under Model 3, calculated at the significance level $\alpha = 0.05$ over 1,000 Monte Carlo replications. Across varying (n, p, c) configurations, we observe that the empirical size consistently reaches its minimum at the true $K_0 = 6$ (bold). Complementing this, the Kolmogorov-Smirnov (KS) distance serves as an auxiliary diagnostic, which similarly attains its minimum at this specific K_0, thereby corroborating the asymptotic validity of the test

Table 2: Empirical sizes for Model 3 under varying (n, c, p) and K_0

n	c	p	$K_0 = 1$	$K_0 = 2$	$K_0 = 3$	$K_0 = 4$	$K_0 = 5$	$K_0 = 6$	$K_0 = 7$
600	0.5	300	0.565 (0.337)	0.449 (0.273)	0.351 (0.214)	0.244 (0.170)	0.116 (0.080)	**0.050** (0.030)	0.927 (0.908)
600	1.0	600	0.435 (0.242)	0.357 (0.187)	0.221 (0.144)	0.141 (0.120)	0.076 (0.054)	**0.042** (0.035)	0.593 (0.737)
600	1.5	900	0.373 (0.214)	0.269 (0.148)	0.176 (0.105)	0.130 (0.075)	0.079 (0.028)	**0.067** (0.026)	0.411 (0.623)
1200	0.5	600	0.440 (0.251)	0.316 (0.192)	0.250 (0.141)	0.153 (0.111)	0.085 (0.040)	**0.046** (0.019)	0.925 (0.919)
1200	1.0	1200	0.316 (0.172)	0.224 (0.119)	0.151 (0.088)	0.095 (0.067)	0.073 (0.035)	**0.057** (0.031)	0.597 (0.751)
1200	1.5	1800	0.240 (0.146)	0.167 (0.120)	0.111 (0.083)	0.068 (0.046)	0.063 (0.036)	**0.054** (0.029)	0.367 (0.620)
1800	0.5	900	0.367 (0.217)	0.252 (0.161)	0.168 (0.128)	0.118 (0.088)	0.072 (0.045)	**0.042** (0.039)	0.920 (0.910)
1800	1.0	1800	0.254 (0.159)	0.177 (0.108)	0.131 (0.088)	0.086 (0.070)	0.068 (0.039)	**0.051** (0.020)	0.583 (0.739)
1800	1.5	2700	0.187 (0.117)	0.135 (0.078)	0.100 (0.061)	0.072 (0.049)	0.061 (0.038)	**0.056** (0.036)	0.363 (0.593)

Note: Empirical size (KS-distance in parentheses).

statistic.

Furthermore, Figure 2 illustrates that standard heuristics, such as the scree plot, the Kaiser criterion, and the 85% cumulative variance threshold, fail under strong dependence. As the dimension grows, the accumulation of bulk eigenvalues near the spectral edge blurs the separation between spikes and the bulk and shifts the apparent elbow to later components, causing visually based rules to erroneously absorb close spikes and overestimate K_0.

These results demonstrate that the test is well-calibrated, a finding fur-

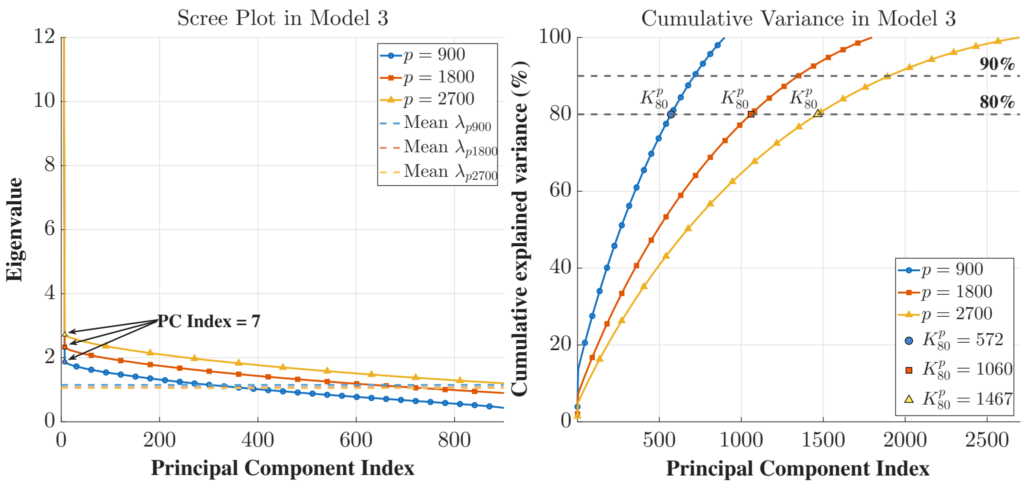

Figure 2: Scree plots and cumulative variance in Model 3.

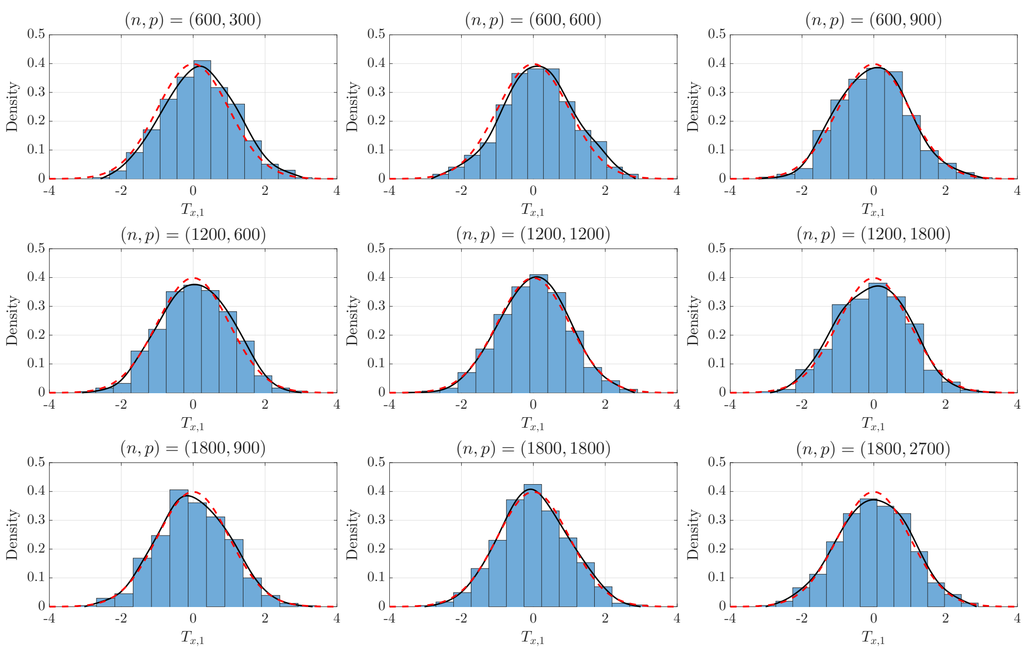

Figure 3: Empirical distributions of $T_{x,1}$ under the null in Model 3.

ther corroborated by the empirical density histograms in Figure 3. Specifically, under the null hypothesis $K = K_0$, the standardized test statistic closely follows the standard normal distribution $\mathcal{N}(0,1)$, aligning with the stable size properties reported in Table 2. Overall, these numerical results indicate that the true value K_0 corresponds to a stabilization point in the empirical size curve, thereby substantiating the accuracy and robustness of the proposed methodology in multi-population settings.

S8 Supplementary Experiments on the Facial Databases

S8.1 Empirical Support for Multi-population Heterogeneity

To empirically examine the heterogeneity of row covariance structures, for each $j \in [1:h]$, let $\boldsymbol{w}_{jt} \in \mathbb{R}^p$ denote the jth row vector of the t-th image sample matrix and $\bar{\boldsymbol{w}}_j = 1/n \sum_{i=1}^n \boldsymbol{w}_{ji}$. For any $i, j = [1:h]$ we compute the pairwise distance matrix $D = (d(i,j))_{1 \le i,j \le h}$, where the normalized covariance distance defined as

$$d(i,j) = \left\| \frac{\mathbf{S}_i}{\mathrm{tr}(\mathbf{S}_i)} - \frac{\mathbf{S}_j}{\mathrm{tr}(\mathbf{S}_j)} \right\|_F, \text{ with } \mathbf{S}_j = \frac{1}{n-1} \sum_{t=1}^n (\boldsymbol{w}_{jt} - \bar{\boldsymbol{w}}_j)^\top (\boldsymbol{w}_{jt} - \bar{\boldsymbol{w}}_j).$$

Here, $d(i,j)$ provides an empirical diagnostic, after trace normalization, for the heterogeneity among covariance structures associated with different image rows.

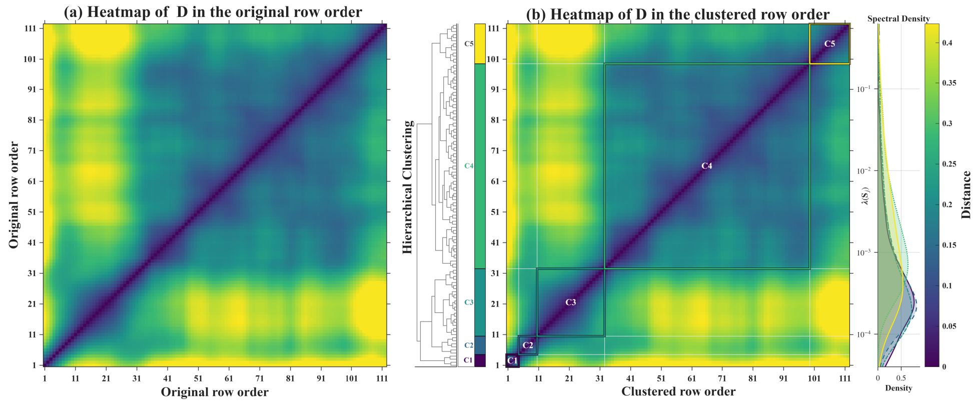

Figure 4: Heatmap of D in the original and clustered row order (ORL).

As an illustration using the ORL facial image database (400 images sized 92×112 pixels), we computed matrix D based on covariance discrepancies. As shown in Figure 4(a), the original-order heatmap exhibits clear block and bright band patterns rather than a homogeneous structure, indicating that covariance structures across image rows are spatially heterogeneous. Following hierarchical reordering (Figure 4(b)), the global pattern remains largely consistent, implying that contiguous rows naturally aggregate into distinct row-populations. The annotated diagonal blocks (C1)–(C5) confirm small intra-cluster distances alongside varying inter-cluster heterogeneity. Furthermore, discernible variations in the empirical spectral densities across these five clusters provide supplementary evidence that the empirically derived row-populations possess heterogeneous covariance structures. Applying this diagnostic to the FERET dataset as shown in Figure 5,

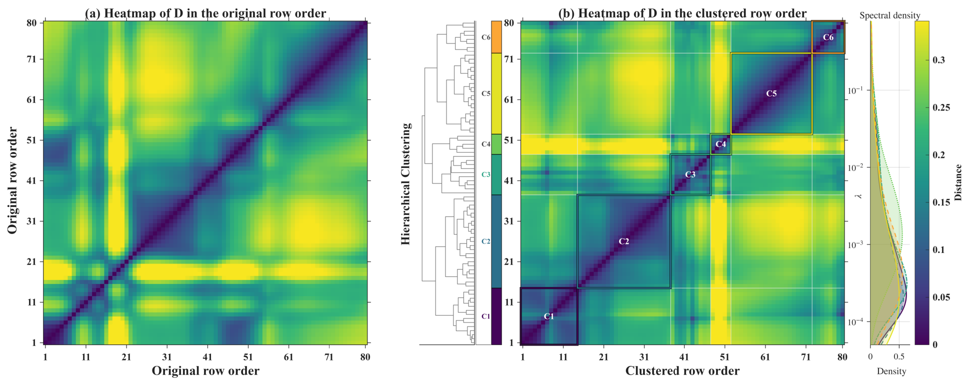

Figure 5: Heatmap of D in the original and clustered row order (FERET).

the clustered row order partially reverses the original spatial sequence, indicating that intrinsic covariance similarities do not strictly follow the physical top-to-bottom arrangement. Nonetheless, the reordered heatmap clearly delineates 6 distinct clusters (C1)–(C6) with distinct diagonal block structures. For the Extended Yale B dataset in Figure 6, hierarchical reordering partitions the fine-grained original heatmap into 11 smaller diagonal blocks (C1)–(C11), revealing a more granular row-population structure than in the ORL and FERET datasets.

Across all examined datasets, variations in the empirical spectral densities among these labeled clusters consistently validate the presence of heterogeneous row-population covariance structures as shown in Figure 7.

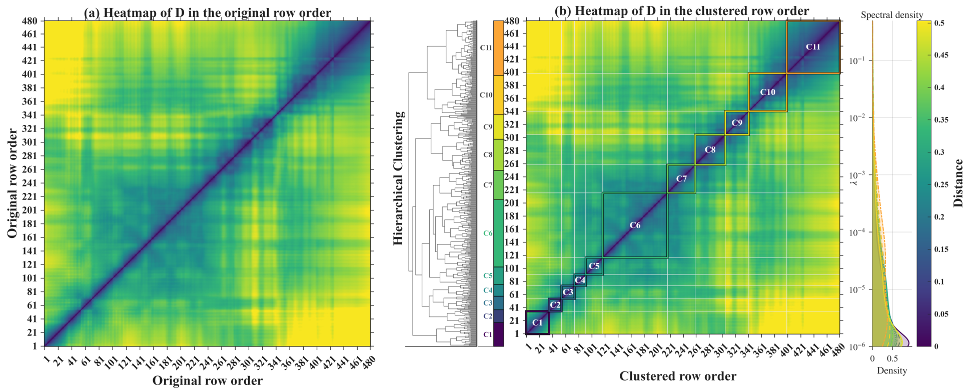

Figure 6: Heatmap of D in the original and clustered row order (Yale B).

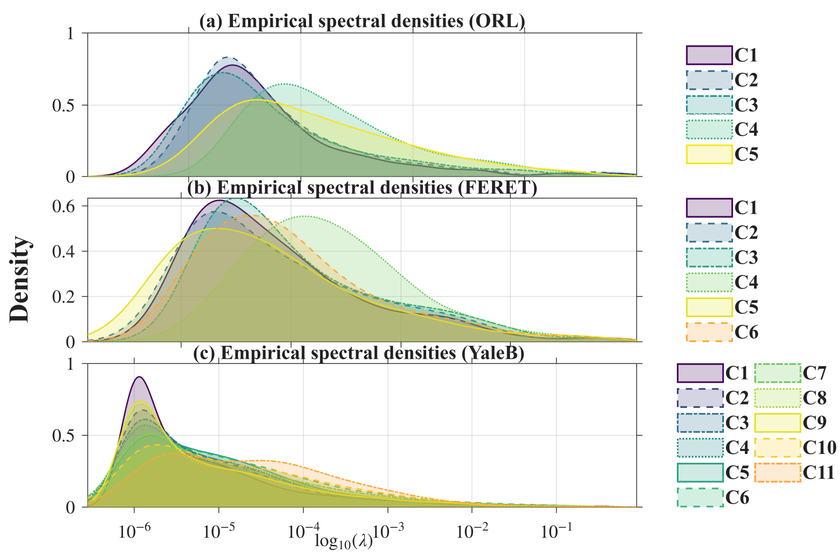

Figure 7: Empirical spectral densities of row-population covariance estimates.

S8.2 Experiment on the FERET Database

To evaluate performance under broader class diversity and varying facial expressions, we utilize a subset of the FERET database containing 1,400 images of 200 individuals. Unlike the ORL dataset, these images are resized to 80×80 pixels and feature minimal background clutter and significant variations in expression and lighting. In the experimental setup, the image covariance matrix $\mathbf{G}$ for the FERET database is constructed using a training set consisting of one image randomly chosen from the first five images of each subject, with the remainder used for testing.

Table 3: Empirical size and KS-distance for (4.24) in the FERET database.

	$K_0 = 1$	$K_0 = 2$	$K_0 = 3$	$K_0 = 4$	$K_0 = 5$
Empirical size	1.000	1.000	**0.045**	1.000	1.000
KS-distance metric	(1.000)	(1.000)	(0.025)	(0.898)	(0.936)

Table 3 reports the empirical sizes of rejecting the null hypothesis (4.24) together with the corresponding KS-distance metric, based on 300 replications at a significance level $\alpha = 0.05$ along the sequential testing procedure. Following Flowchart 1, the number of principal components K_0 for $\mathbf{G}$ is selected according to the elbow pattern of the empirical size curve, with the KS distance serving as an auxiliary diagnostic for the adequacy of the null approximation. Under this criterion, we obtain $K_0 = 3$ for the FERET database. Figure 8 further compares the test-set classification accuracies of the competing methods. As indicated by the red solid curve, our sequential testing procedure yields stable principal component selection and

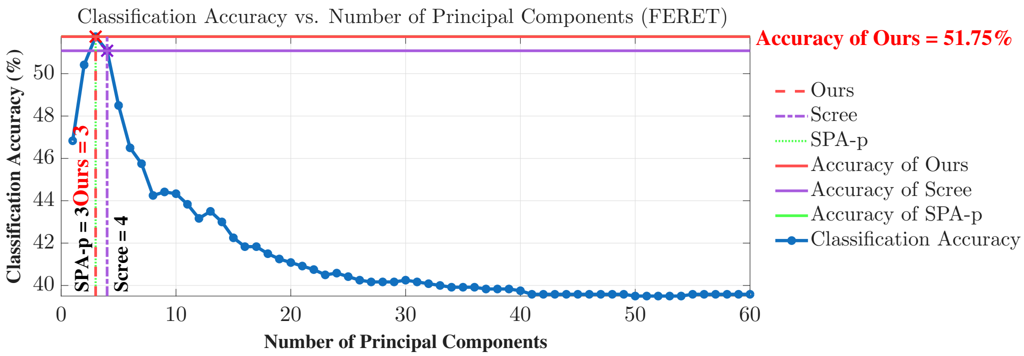

Figure 8: Test-set classification accuracy vs. number of PCs (FERET).

consistently competitive classification performance.

References

Bai, Z. and J. W. Silverstein (2010). *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. New York: Springer.

Bai, Z. and J. Yao (2012a). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis 106*, 167–177.

Bai, Z. and J. Yao (2012b). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis 106*, 167–177.

Bai, Z. D., B. Miao, and G. Pan (2007). On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability 35*(4), 1532–1572.

Ding, X. and F. Yang (2021). Spiked separable covariance matrices and principal components. *The Annals of Statistics 49*(2), 1113–1138.

Jiang, D. (2023). A universal test on spikes in a high-dimensional generalized spiked model and its applications. *Statistica Sinica 33*, 1749–1770.

Ledoit, O. and M. Wolf (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics 48*(5), 3043–3065.

Najim, J. and J. Yao (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *The Annals of Applied Probability 26*(3), 1837–1887.

Wang, R. and D. Jiang (2025). Inference on testing the number of spikes in a high-dimensional generalized spiked Fisher matrix. *Electronic Journal of Statistics 19*(1), 1286–1312.