Eigenvalue distribution of a high-dimensional distance covariance matrix with application

Weiming Li, Qinwen Wang and Jianfeng Yao

Shanghai University of Finance and Economics,
Fudan University and The University of Hong Kong

Supplementary Material

This supplementary material contains some additional technical tools and the proofs of Theorem 1, Theorem 2, Theorem 3, Theorem 4 and Theorem 5 of the main paper. Throughout this supplementary material, $||\cdot||$ denotes the Euclidean norm for vectors, the spectral norm for matrices and the supremum norm for functions, respectively. $\mathbb{C}^+$ and $\mathbb{C}^-$ are referred as the upper and lower half complex plane (real axis excluded). K is used to denote some constant that can vary from place to place.

A Technical tools

Lemma 1. *[El Karoui (2010)] Consider the $n \times n$ kernel random matrix $\mathbf{M}$ with entries*

$$\mathbf{M}_{i,j} = f\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{p}\right).$$

Let us call $\boldsymbol{\psi}$ the vector with i-th entry $\boldsymbol{\psi}_i = \|\mathbf{x}_i\|_2^2/p - \tau/2$, where $\tau = 2\mathrm{tr}(\boldsymbol{\Sigma}_p)/p$. We assume that:
(a) $n \asymp p$, that is, n/p and p/n remain bounded as $p \to \infty$.
(b) $\boldsymbol{\Sigma}_p$ is a positive semi-definite $p \times p$ matrix, and $\|\boldsymbol{\Sigma}_p\| = \sigma_1(\boldsymbol{\Sigma}_p)$ remains bounded in p, that is, there exists $K > 0$, such that $\sigma_1(\boldsymbol{\Sigma}_p) \le K$, for all p.
(c) There exists $\ell \in \mathbb{R}$ such that $\lim_{p\to\infty} \mathrm{tr}(\boldsymbol{\Sigma}_p)/p = \ell$.
(d) $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{x}_i = \boldsymbol{\Sigma}_p^{1/2}\mathbf{w}_i$ for $i = 1, \ldots, n$.
(e) The entries of $\mathbf{w}_i$, a p-dimensional random vector, are i.i.d. Also, denoting by w_{ik} the kth entry of $\mathbf{w}_i$, we assume that $\mathbb{E}(w_{ik}) = 0$, $\mathbb{V}ar(w_{ik}) = 1$ and $\mathbb{E}(|w_{ik}|^{5+\varepsilon}) < \infty$ for some $\varepsilon > 0$.
(f) f is C^3 in a neighborhood of τ.
Then $\mathbf{M}$ can be approximated consistently in operator norm (and in proba-

bility) by the matrix $\widetilde{\mathbf{M}}$*, defined by*

$$\widetilde{\mathbf{M}} = f(\tau)\mathbf{1}\mathbf{1}' + f'(\tau)\left[\mathbf{1}\boldsymbol{\psi}' + \boldsymbol{\psi}\mathbf{1}' - 2\frac{\mathbf{X}'\mathbf{X}}{p}\right]$$
$$+ \frac{f''(\tau)}{2}\left[\mathbf{1}(\boldsymbol{\psi}\circ\boldsymbol{\psi})' + (\boldsymbol{\psi}\circ\boldsymbol{\psi})\mathbf{1}' + 2\boldsymbol{\psi}\boldsymbol{\psi}' + 4\frac{\mathrm{tr}(\boldsymbol{\Sigma}_p^2)}{p^2}\mathbf{1}\mathbf{1}'\right] + v_p\mathbf{I}_n,$$
$$v_p = f(0) + \tau f'(\tau) - f(\tau).$$

In other words,

$$||\mathbf{M} - \widetilde{\mathbf{M}}|| \to 0, \quad \textit{in probability.}$$

Lemma 2. *[Bai and Silverstein (2010)] Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $n\times n$ *Hermitian matrices. Then,*

$$||F^{\mathbf{A}} - F^{\mathbf{B}}|| \le \frac{1}{n}\mathrm{rank}(\mathbf{A}-\mathbf{B}) \quad \textit{and} \quad L^3(F^{\mathbf{A}}, F^{\mathbf{B}}) \le \frac{1}{n}\mathrm{tr}[(\mathbf{A}-\mathbf{B})(\mathbf{A}-\mathbf{B})^*],$$

where $L(F,G)$ *stands for the Lévy distance between the distribution functions* F *and* G*.*

Lemma 3. *Let* f*:* $\mathbb{R}^n \to \mathbb{R}$ *be any function of thrice differentiable in each argument. Let also* $\mathbf{x} = (x_1,\ldots,x_n)'$ *and* $\mathbf{y} = (y_1,\ldots,y_n)'$ *be two random vectors in* $\mathbb{R}^n$ *with i.i.d. elements, respectively, and set* $U = f(\mathbf{x})$ *and* $V = f(\mathbf{y})$*. If*

$$\gamma = \max\{\mathbb{E}|x_i|^3, \mathbb{E}|y_i|^3, 1 \le i \le n\} < \infty,$$

then for any thrice differentiable $g : \mathbb{R} \to \mathbb{R}$ *and any* $K > 0$*,*

$$|\mathbb{E}g(U) - \mathbb{E}g(V)| \le 2C_2(g)\gamma n\lambda_3(f),$$

where $\lambda_3(f) = \sup\left\{|\partial_i^k f(\mathbf{z})|^{3/k} : \mathbf{z} = (z_\ell), z_\ell \in \{x_\ell, y_\ell\}, 1 \le i \le n, 1 \le k \le 3\right\}$ *and* $C_2(g) = \frac{1}{6}\|g'\|_\infty + \frac{1}{2}\|g''\|_\infty + \frac{1}{6}\|g'''\|_\infty$*.*

This lemma follows directly from Corollary 1.2 in Chatterjee (2008) and its proof.

B Proofs

At the beginning of this section, we first recall some notations for easy reading.

$$\mathbf{V}_x = \left(\frac{\|\mathbf{x}_k - \mathbf{x}_\ell\|}{\sqrt{p}}\right), \quad \mathbf{V}_y = \left(\frac{\|\mathbf{y}_k - \mathbf{y}_\ell\|}{\sqrt{q}}\right), \quad \mathbf{P}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n',$$

$$\gamma_x = \frac{1}{p}\mathrm{tr}\mathbf{\Sigma}_x, \quad \gamma_y = \frac{1}{q}\mathrm{tr}\mathbf{\Sigma}_y, \quad \kappa_x = \frac{1}{pn}\sum_{i=1}^n ||\mathbf{x}_i||^2, \quad \kappa_y = \frac{1}{qn}\sum_{i=1}^n ||\mathbf{y}_i||^2,$$

$$\mathbf{A}_n = \frac{1}{p}\mathbf{X}'\mathbf{X} + \gamma_x\mathbf{I}_n, \quad \mathbf{C}_n = \frac{1}{q}\mathbf{Y}'\mathbf{Y} + \gamma_y\mathbf{I}_n, \quad \mathbf{B}_n = \mathbf{A}_n^{\frac{1}{2}}\mathbf{C}_n\mathbf{A}_n^{\frac{1}{2}},$$

$$\mathbf{D}_x = \frac{1}{p}\mathbf{X}'\mathbf{X} + \kappa_x\mathbf{I}_n, \quad \mathbf{D}_y = \frac{1}{q}\mathbf{Y}'\mathbf{Y} + \kappa_y\mathbf{I}_n, \quad \mathbf{D}_z = \frac{1}{q}\mathbf{Z}'\mathbf{Z} + \kappa_z\mathbf{I}_n,$$

$$\mathbf{S}_{xy} = \mathbf{P}_n\mathbf{D}_x\mathbf{P}_n\mathbf{D}_y\mathbf{P}_n, \quad \mathbf{S}_{xz} = \mathbf{P}_n\mathbf{D}_x\mathbf{P}_n\mathbf{D}_z\mathbf{P}_n.$$

B.1 Proof of Theorem 1

The squared sample distance covariance $\mathcal{V}_n^2(\mathbf{x}, \mathbf{y})$ in (1.2) can be expressed as an inner product between the two matrices $\mathbf{P}_n\mathbf{V}_x\mathbf{P}_n$ and $\mathbf{P}_n\mathbf{V}_y\mathbf{P}_n$, that is,

$$\mathcal{V}_n^2(\mathbf{x}, \mathbf{y}) = \frac{\sqrt{pq}}{n^2}\mathrm{tr}\mathbf{P}_n\mathbf{V}_x\mathbf{P}_n\mathbf{V}_y\mathbf{P}_n.$$

Notice that the matrices $\mathbf{V}_x$ and $\mathbf{V}_y$ are exactly the Euclidean distance kernel matrices discussed in El Karoui (2010) with kernel function $f(x) = \sqrt{x}$. Applying their main theorem (see Lemma 1), the matrix

$$\mathbf{P}_n\mathbf{V}_x\mathbf{P}_n\mathbf{V}_y\mathbf{P}_n \tag{B.1}$$

can be approximated by a simplified random matrix $\mathbf{V}_n$ such that as (n, p, q) tend to infinity,

$$\|\mathbf{V}_n - \mathbf{P}_n\mathbf{V}_x\mathbf{P}_n\mathbf{V}_y\mathbf{P}_n\| \to 0 \tag{B.2}$$

in probability, where

$$\mathbf{V}_n \triangleq \frac{1}{2\sqrt{\gamma_x\gamma_y}}\mathbf{P}_n\left(\mathbf{A}_n + \frac{1}{8\gamma_x}\boldsymbol{\psi}_x\boldsymbol{\psi}_x'\right)\mathbf{P}_n\left(\mathbf{C}_n + \frac{1}{8\gamma_y}\boldsymbol{\psi}_y\boldsymbol{\psi}_y'\right)\mathbf{P}_n, \tag{B.3}$$

in which

$$\boldsymbol{\psi}_x = \frac{1}{p}\begin{pmatrix} \|\mathbf{x}_1\|^2 - \mathrm{tr}\mathbf{\Sigma}_x \\ \vdots \\ \|\mathbf{x}_n\|^2 - \mathrm{tr}\mathbf{\Sigma}_x \end{pmatrix} \quad \text{and} \quad \boldsymbol{\psi}_y = \frac{1}{q}\begin{pmatrix} \|\mathbf{y}_1\|^2 - \mathrm{tr}\mathbf{\Sigma}_y \\ \vdots \\ \|\mathbf{y}_n\|^2 - \mathrm{tr}\mathbf{\Sigma}_y \end{pmatrix}.$$

Then we replace the two traces γ_x and γ_y in $\mathbf{A}_n$ and $\mathbf{C}_n$ with their unbiased sample counterparts κ_x and κ_y, respectively, which does not affect the convergence in (B.2). Finally in (B.3), by removing the two rank-one matrices $(8\gamma_x)^{-1}\boldsymbol{\psi}_x\boldsymbol{\psi}_x^T$ and $(8\gamma_y)^{-1}\boldsymbol{\psi}_y\boldsymbol{\psi}_y^T$ (which have bounded spectral norm, almost surely), we get the conclusion of the theorem. The proof is thus complete.

B.2 Proof of Theorem 2

Recall the approximation from Theorem 1,

$$\mathcal{V}_n^2(\mathbf{x},\mathbf{y}) = \frac{1}{2n^2}\sqrt{\frac{pq}{\gamma_x\gamma_y}}\mathrm{tr}\mathbf{S}_{xy} + o_p(1)$$

and notice that

$$\begin{aligned}\frac{1}{n}\mathrm{tr}(\mathbf{S}_{xy}) &= \frac{1}{npq}\mathrm{tr}(\mathbf{P}_n\mathbf{X}'\mathbf{X}\mathbf{P}_n\mathbf{Y}'\mathbf{Y}\mathbf{P}_n) + \frac{\kappa_y}{np}\mathrm{tr}(\mathbf{P}_n\mathbf{X}'\mathbf{X}) + \frac{\kappa_x}{nq}\mathrm{tr}(\mathbf{P}_n\mathbf{Y}'\mathbf{Y})\\ &\quad + \frac{n-1}{n}\kappa_x\kappa_y\\ &= \frac{1}{npq}\mathrm{tr}(\mathbf{X}'\mathbf{X}\mathbf{Y}'\mathbf{Y}) + 3\gamma_x\gamma_y + o_{a.s}(1).\end{aligned}$$

Moreover, from Equation (21) in Li and Yao (2018) and the independence between $\mathbf{X}$ and $\mathbf{Y}$,

$$\frac{1}{npq}\mathrm{tr}(\mathbf{X}'\mathbf{X}\mathbf{Y}'\mathbf{Y}) = \frac{1}{p}\mathrm{tr}(\boldsymbol{\Sigma}_x)\frac{1}{q}\mathrm{tr}(\boldsymbol{\Sigma}_y) + o_{a.s}(1).$$

Collecting the above results yields

$$\mathcal{V}_n^2(\mathbf{x},\mathbf{y}) = 2\sqrt{c_{n1}c_{n2}\gamma_x\gamma_y} + o_p(1).$$

On the other hand, applying Lemma 1, we have

$$\begin{aligned}\frac{1}{n}S_{2,n} =& \frac{1}{2n}\sqrt{\frac{pq}{\gamma_x\gamma_y}}\left(\frac{1}{n^2}\mathbf{1}'\mathbf{D}_x\mathbf{1} - 2\gamma_x\right)\left(\frac{1}{n^2}\mathbf{1}'\mathbf{D}_y\mathbf{1} - 2\gamma_y\right) + o_p(1)\\ =& 2\sqrt{c_{n1}c_{n2}\gamma_x\gamma_y} + o_p(1).\end{aligned}$$

Therefore, the statistic $T_n = n\mathcal{V}_n^2(\mathbf{x},\mathbf{y})/S_{2,n}$ converges to 1 in probability. The proof is complete.

B.3 Proof of Theorem 3

The strategy of the proof is as follows. First, we prove the theorem under Gaussian assumption. By virtue of rotation invariance property of Gaussian vectors, we may treat the two population covariance matrices $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ as diagonal ones, which can simplify the proof dramatically. Second, applying Lindeberg's replacement trick provided in Chatterjee (2008), we will remove the Gaussian assumption and show that the theorem still holds true for general distributions if the atoms (w_{ij}) have finite fourth moment, as stated in our Assumption (b).

Gaussian case: First, we have

$$|\kappa_x - \gamma_x| \xrightarrow{a.s.} 0 \quad \text{and} \quad |\kappa_y - \gamma_y| \xrightarrow{a.s.} 0, \tag{B.4}$$

as (n, p, q) tend to ∞. ¿From Lemma 2 and (B.4), we get

$$L^3(F^{\mathbf{S}_{xy}}, F^{\mathbf{B}_n}) \xrightarrow{a.s.} 0.$$

Hence, the matrices $\mathbf{S}_{xy}$ and $\mathbf{B}_n$ share the same limiting spectral distribution and thus we only focus on the convergence of $F^{\mathbf{B}_n}$. We first derive its limit conditioning on the sequence $(\mathbf{A}_n)$. Then the result holds unconditionally if the limit is independent of $(\mathbf{A}_n)$. Following standard strategies from random matrix theory, letting $s_{\mathbf{B}_n}(z)$ be the Stieltjes transform of $F^{\mathbf{B}_n}$, the convergence of $F^{\mathbf{B}_n}$ can be established through three steps:

Step 1: For any fixed $z \in \mathbb{C}^+$, $s_{\mathbf{B}_n}(z) - \mathbb{E}s_{\mathbf{B}_n}(z) \to 0$, almost surely.

Step 2: For any fixed $z \in \mathbb{C}^+$, $\mathbb{E}s_{\mathbf{B}_n}(z) \to s(z)$ with $s(z)$ satisfies the equations in (3).

Step 3: The uniqueness of the solution $s(z)$ to (3.1) on the set (3.2).

Step 1. Almost sure convergence of $s_{\mathbf{B}_n}(z) - \mathbb{E}s_{\mathbf{B}_n}(z)$.

We assume $\mathbf{\Sigma}_y$ is diagonal, having the form

$$\mathbf{\Sigma}_y = \text{Diag}(\tau_1, \ldots, \tau_q).$$

By this and notations

$$\mathbf{r}_k = \frac{1}{\sqrt{q}} \mathbf{A}_n^{1/2} (w_{p+k,1}, \ldots, w_{p+k,n})', \quad k = 1, \ldots, q,$$

the matrix $\mathbf{B}_n$ can be expressed as

$$\mathbf{B}_n = \gamma_y \mathbf{A}_n + \sum_{k=1}^{q} \tau_k \mathbf{r}_k \mathbf{r}_k'. \tag{B.5}$$

It's "leave-one-out" version is denoted by $\mathbf{B}_{k,n} = \mathbf{B}_n - \tau_k \mathbf{r}_k \mathbf{r}_k'$, $k = 1, \ldots, q$. Let $\mathbb{E}_0(\cdot)$ be expectation and $\mathbb{E}_k(\cdot)$ be conditional expectation given $\mathbf{r}_1, \ldots, \mathbf{r}_k$. From the martingale decomposition and the identity

$$\mathbf{r}_k'(\mathbf{B}_n - z\mathbf{I}_n)^{-1} = \frac{\mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}}{1 + \tau_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k}, \tag{B.6}$$

we have

$$\begin{aligned} s_{\mathbf{B}_n}(z) - \mathbb{E}s_{\mathbf{B}_n}(z) =& \frac{1}{n}\sum_{k=1}^{q}(\mathbb{E}_k - \mathbb{E}_{k-1})\left[\operatorname{tr}(\mathbf{B}_n - z\mathbf{I}_n)^{-1} - \operatorname{tr}(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\right] \\ =& -\frac{1}{n}\sum_{k=1}^{q}(\mathbb{E}_k - \mathbb{E}_{k-1})\frac{\tau_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-2}\mathbf{r}_k}{1 + \tau_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k}. \end{aligned} \tag{B.7}$$

Similar to the arguments on pages 435-436 of Bai and Zhou (2008), the summands in (B.7) form a bounded martingale difference sequence, and hence $s_{\mathbf{B}_n}(z) - \mathbb{E}s_{\mathbf{B}_n}(z) \to 0$, almost surely.

Step 2. Convergence of $\mathbb{E}s_{\mathbf{B}_n}(z)$.

Let $s_{\mathbf{A}_n}(z)$ be the Stieltjes transform of $F^{\mathbf{A}_n}$. From Silverstein (1995), $s_{\mathbf{A}_n}(z)$ converges almost surely to $s_{\mathbf{A}}(z)$, which satisfies

$$z = -\frac{1}{s_{\mathbf{A}}(z)} + \int t + \frac{t}{1 + tc_1^{-1}s_{\mathbf{A}}(z)} dH_x(t). \tag{B.8}$$

Define two functions $w_n(z)$ and $m_n(z)$ as

$$w_n(z) = \frac{1}{n}\mathbb{E}\operatorname{tr}(\mathbf{B}_n - zI_n)^{-1}\mathbf{A}_n \quad \text{and} \quad m_n(z) = \gamma_y + \frac{1}{q}\sum_{k=1}^{q}\frac{\tau_k}{1 + \tau_k c_{n2}^{-1} w_n(z)}. \tag{B.9}$$

We first show that

$$m_n^{-1}(z) s_{\mathbf{A}_n}\left[z m_n^{-1}(z)\right] - \mathbb{E}s_{\mathbf{B}_n}(z) \to 0, \quad n \to \infty. \tag{B.10}$$

In fact, applying the identity (B.6), we have

$$\begin{aligned}
&\frac{1}{n}\mathrm{tr}\left[m_n(z)\mathbf{A}_n - z\mathbf{I}_n\right]^{-1} - \frac{1}{n}\mathrm{tr}(\mathbf{B}_n - z\mathbf{I}_n)^{-1}\\
=&\frac{1}{n}\mathrm{tr}\left[m_n(z)\mathbf{A}_n - z\mathbf{I}_n\right]^{-1}\left(\sum_{k=1}^{q}\tau_k\mathbf{r}_k\mathbf{r}_k' - (m_n(z)-\gamma_y)\mathbf{A}_n\right)(\mathbf{B}_n - z\mathbf{I}_n)^{-1}\\
=&\frac{1}{n}\sum_{k=1}^{q}\frac{\tau_k\mathbf{r}_k'(\mathbf{B}_{k,n} - zI_n)^{-1}\left[(m_n(z)\mathbf{A}_n - z\mathbf{I}_n\right]^{-1}\mathbf{r}_k}{1+\tau_k\mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k}\\
&-\frac{m_n(z)-\gamma_y}{n}\mathrm{tr}\left[m_n(z)\mathbf{A}_n - z\mathbf{I}_n\right]^{-1}\mathbf{A}_n(\mathbf{B}_n - z\mathbf{I}_n)^{-1}\\
=&\frac{1}{n}\sum_{k=1}^{q}\frac{\tau_k d_k}{1+\tau_k c_{n2}^{-1}w_n(z)},
\end{aligned}$$

where

$$\begin{aligned}
d_k =&\frac{1+\tau_k c_{n2}^{-1}w_n(z)}{1+\tau_k\mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k}\mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\left[m_n(z)\mathbf{A}_n - z\mathbf{I}_n\right]^{-1}\mathbf{r}_k\\
&-\frac{1}{q}\mathrm{tr}\left[m_n(z)\mathbf{A}_n - z\mathbf{I}_n\right]^{-1}\mathbf{A}_n(\mathbf{B}_n - z\mathbf{I}_n)^{-1}.
\end{aligned}$$

Following similar arguments on pages 85-87 of Bai and Silverstein (2010), one may obtain

$$\max_k \mathbb{E}(d_k) \to 0.$$

This result together with the fact

$$\inf_n |1+\tau_k c_{n2}^{-1}w_n(z)| \geq \inf_n \tau_k c_{n2}^{-1}|\Im(w_n(z))| > 0$$

imply the convergence in (B.10).

We next find another link between $\mathbb{E}s_{\mathbf{B}_n}(z)$ and $w_n(z)$ by proving

$$1 + z\mathbb{E}s_{\mathbf{B}_n}(z) - \gamma_y w_n(z) - \frac{1}{n}\sum_{k=1}^{q}\frac{\tau_k w_n(z)}{c_{n2}+\tau_k w_n(z)} \to 0. \tag{B.11}$$

¿From the expression of $\mathbf{B}_n$ in (B.5) and the identity in (B.6), we have

$$\begin{aligned}\mathbf{I}_n + z(\mathbf{B}_n - z\mathbf{I})^{-1} =&\mathbf{B}_n(\mathbf{B}_n - z\mathbf{I}_n)^{-1}\\ =&\gamma_y \mathbf{A}_n(\mathbf{B}_n - z\mathbf{I})^{-1} + \sum_{k=1}^{q} \tau_k \mathbf{r}_k \mathbf{r}_k'(\mathbf{B}_n - z\mathbf{I}_n)^{-1}\\ =&\gamma_y \mathbf{A}_n(\mathbf{B}_n - z\mathbf{I})^{-1} + \sum_{k=1}^{q} \frac{\tau_k \mathbf{r}_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}}{1+\tau_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k}.\end{aligned} \tag{B.12}$$

Taking the trace on both sides of (B.12) and dividing by n, we get

$$\begin{aligned}1 + z\frac{1}{n}\mathrm{tr}(\mathbf{B}_n - z\mathbf{I}_n)^{-1} =&\gamma_y \frac{1}{n}\mathrm{tr}(\mathbf{B}_n - z\mathbf{I}_n)^{-1}\mathbf{A}_n + \frac{1}{n}\sum_{k=1}^{q} \frac{\tau_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k}{1+\tau_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k}\\ =&\gamma_y \frac{1}{n}\mathrm{tr}(\mathbf{B}_n - z\mathbf{I}_n)^{-1}\mathbf{A}_n + \frac{1}{n}\sum_{k=1}^{q} \frac{\tau_k c_{n2}^{-1} w_n(z)}{1+\tau_k c_{n2}^{-1} w_n(z)} + \varepsilon_n,\end{aligned}$$

where

$$\varepsilon_n = \frac{1}{n}\sum_{k=1}^{q} \frac{\tau_k [c_{n2}^{-1} w_n(z) - \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k]}{[1+\tau_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k][1+\tau_k c_{n2}^{-1} w_n(z)]}.$$

¿From the proof of (2.3) in Silverstein (1995), almost surely,

$$\inf_n \left|[1+\tau_k \mathbf{r}_k'(\mathbf{B}_{k,n} - z\mathbf{I}_n)^{-1}\mathbf{r}_k][1+\tau_k c_{n2}^{-1} w_n(z)]\right| > 0.$$

Moreover, following similar arguments on page 87 of Bai and Silverstein (2010), one may get

$$\frac{1}{n}\sum_{k=1}^{q} \mathbb{E}^{\frac{1}{2}} |c_{n2}^{-1} w_n(z) - \mathbf{r}_k'(\mathbf{B}_{k,n} - zI_n)^{-1}\mathbf{r}_k|^2 \to 0.$$

Therefore $\mathbb{E}(\varepsilon_n) \to 0$ and hence the convergence in (B.11) holds.

By considering a subsequence $\{n_k\}$ such that $w_{n_k}(z) \to w(z)$, from (B.8), (B.10) and (B.11), we have

$$\begin{aligned} m_{n_k}(z) &\to \int t + \frac{t}{1+tc_{n2}^{-1}w(z)} dH_y(t) \triangleq m(z),\\ s_{n_k}(z) &\to \frac{1}{m(z)} s_{\mathbf{A}}\left(\frac{z}{m(z)}\right),\\ zs_{n_k}(z) &\to -1 + w(z)\int t + \frac{t}{1+tc_2^{-1}w(z)} dH_y(t),\end{aligned}$$

as $k \to \infty$. These results demonstrate that $s_{n_k}(z)$ has a limit, say $s(z)$, which together with $(w(z), m(z),\ s_A(z))$ satisfy the following system of equations:

$$\begin{cases} s(z) = \dfrac{1}{m(z)} s_{\mathbf{A}}\left(\dfrac{z}{m(z)}\right), \\ zs(z) = -1 + w(z) \displaystyle\int t + \frac{t}{1 + tc_2^{-1} w(z)} dH_y(t), \\ z = -\dfrac{1}{s_{\mathbf{A}}(z)} + \displaystyle\int t + \frac{t}{1 + tc_1^{-1} s_{\mathbf{A}}(z)} dH_x(t), \\ m(z) = \displaystyle\int t + \frac{t}{1 + tc_2^{-1} w(z)} dH_y(t). \end{cases}$$

Cancelling the function $s_{\mathbf{A}}(z)$ from the above system yields an equivalent but simpler system of equations as shown in (3). Hence, the convergence of $s_n(z)$ is established if the system has a unique solution on the set (3.2).

Step 3. Uniqueness of the solution to (3).

The system of equations in (3) is equivalent to

$$\begin{cases} 1 + zs = wm, \\ m = \displaystyle\int t + \frac{t}{1 + tc_2^{-1} w} dH_y(t), \\ w = s \displaystyle\int t + \frac{t}{1 + tc_1^{-1}(1 + zs) w^{-1} s} dH_x(t). \end{cases} \tag{B.13}$$

Bringing $s = [wm - 1]/z$ into the third equation in (B.13), we have

$$w = \int \frac{t}{z}\Big(wm - 1\Big) + \frac{t\big(wm - 1\big)}{z + c_1^{-1} tm\big(wm - 1\big)} dH_x(t). \tag{B.14}$$

Now suppose the LSD $F \neq \delta_0$ and we have two solutions (s, w, m) and $(\tilde{s}, \tilde{w}, \tilde{m})$ to the system on the set (3.2) for a common $z \in \mathbb{C}^+$. Then, from (B.13) and (B.14), we can obtain

$$\begin{aligned} w - \tilde{w} = {} & (wm - \tilde{w}\tilde{m}) \\ & \times \int \left[\frac{t}{z} + \frac{tz}{\big(z + c_1^{-1} tm(wm - 1)\big)\big(z + c_1^{-1} t\tilde{m}(\tilde{w}\tilde{m} - 1)\big)} \right] dH_x(t) \\ & + (\tilde{m} - m) \int \frac{t^2 c_1^{-1} (wm - 1)(\tilde{w}\tilde{m} - 1)}{(z + c_1^{-1} tm(wm - 1))(z + c_1^{-1} t\tilde{m}(\tilde{w}\tilde{m} - 1))} dH_x(t), \end{aligned} \tag{B.15}$$

$$\tilde{m} - m = (w - \tilde{w}) \int \frac{t^2 c_2^{-1}}{\big(1 + tc_2^{-1} w\big)\big(1 + tc_2^{-1}\tilde{w}\big)} dH_y(t), \tag{B.16}$$

$$wm - \tilde{w}\tilde{m} = (w - \tilde{w}) \int \left(t + \frac{t}{\big(1 + tc_2^{-1} w\big)\big(1 + tc_2^{-1}\tilde{w}\big)} \right) dH_y(t). \tag{B.17}$$

Combining (B.15)-(B.17), if $w \neq \tilde{w}$, we have

$$B_1 B_2 + C_1 C_2 = 1, \tag{B.18}$$

where

$$B_1 = \int \frac{t}{z} + \frac{tz}{\big(z + c_1^{-1} tm(wm - 1)\big)\big(z + c_1^{-1} t\tilde{m}(\tilde{w}\tilde{m} - 1)\big)} dH_x(t),$$

$$B_2 = \int t + \frac{t}{\big(1 + tc_2^{-1} w\big)\big(1 + tc_2^{-1}\tilde{w}\big)} dH_y(t),$$

$$C_1 = \int \frac{t^2 c_1^{-1}(wm - 1)(\tilde{w}\tilde{m} - 1)}{(z + c_1^{-1} tm(wm - 1))(z + c_1^{-1} t\tilde{m}(\tilde{w}\tilde{m} - 1))} dH_x(t),$$

$$C_2 = \int \frac{t^2 c_2^{-1}}{\big(1 + tc_2^{-1} w\big)\big(1 + tc_2^{-1}\tilde{w}\big)} dH_y(t).$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|B_1 B_2|^2 \leq& \int \left|\frac{t}{z}\right| + \frac{|tz|}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \\
&\times \int \left|\frac{t}{z}\right| + \frac{|tz|}{|z + c_1^{-1} t\widetilde{m}(\widetilde{w}\widetilde{m} - 1)|^2} dH_x(t) \\
&\times \int t + \frac{t}{|1 + tc_2^{-1} w|^2} dH_y(t) \int t + \frac{t}{|1 + tc_2^{-1}\widetilde{w}|^2} dH_y(t) \\
=& \int \left|\frac{t}{z}\right| + \frac{|tz|}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \int t + \frac{t}{|1 + tc_2^{-1} w|^2} dH_y(t) \\
&\times \int \left|\frac{t}{z}\right| + \frac{|tz|}{|z + c_1^{-1} t\widetilde{m}(\widetilde{w}\widetilde{m} - 1)|^2} dH_x(t) \int t + \frac{t}{|1 + tc_2^{-1}\widetilde{w}|^2} dH_y(t) \\
:=& (\widetilde{B}_1 \widetilde{B}_2)^2,
\end{aligned}$$

$$\begin{aligned}
|C_1 C_2|^2 \leq& \int \frac{t^2 c_1^{-1}|wm - 1|^2}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \int \frac{t^2 c_1^{-1}|\widetilde{w}\widetilde{m} - 1|^2}{|z + c_1^{-1} t\widetilde{m}(\widetilde{w}\widetilde{m} - 1)|^2} dH_x(t) \\
&\times \int \frac{t^2 c_2^{-1}}{|1 + tc_2^{-1} w|^2} dH_y(t) \frac{t^2 c_2^{-1}}{|1 + tc_2^{-1}\widetilde{w}|^2} dH_y(t)
\end{aligned}$$

$$
\begin{aligned}
&= \int \frac{t^2 c_1^{-1}|wm-1|^2}{|z+c_1^{-1}tm(wm-1)|^2}dH_x(t)\int \frac{t^2 c_2^{-1}}{|1+tc_2^{-1}w|^2}dH_y(t)\\
&\quad \times \int \frac{t^2 c_1^{-1}|\widetilde{w}\widetilde{m}-1|^2}{|z+c_1^{-1}t\widetilde{m}(\widetilde{w}\widetilde{m}-1)|^2}dH_x(t)\int \frac{t^2 c_2^{-1}}{|1+tc_2^{-1}\widetilde{w}|^2}dH_y(t)\\
&:=(\widetilde{C}_1\widetilde{C}_2)^2.
\end{aligned}
$$

Then (B.18) implies

$$
\begin{aligned}
1 &= |B_1B_2+C_1C_2|\\
&\le \sqrt{(\widetilde{B}_1^2+\widetilde{C}_1^2)(\widetilde{B}_2^2+\widetilde{C}_2^2)}\\
&= \Bigg\{\int \left|\frac{t}{z}\right| + \frac{|tz|}{|z+c_1^{-1}tm(wm-1)|^2}dH_x(t)\int t+\frac{t}{|1+tc_2^{-1}w|^2}dH_y(t)\\
&\quad + \int \frac{t^2c_1^{-1}|wm-1|^2}{|z+c_1^{-1}tm(wm-1)|^2}dH_x(t)\int \frac{t^2c_2^{-1}}{|1+tc_2^{-1}w|^2}dH_y(t)\Bigg\}^{1/2}\\
&\quad \times \Bigg\{\int \left|\frac{t}{z}\right| + \frac{|tz|}{|z+c_1^{-1}t\widetilde{m}(\widetilde{w}\widetilde{m}-1)|^2}dH_x(t)\int t+\frac{t}{|1+tc_2^{-1}\widetilde{w}|^2}dH_y(t)\\
&\quad + \int \frac{t^2c_1^{-1}|\widetilde{w}\widetilde{m}-1|^2}{|z+c_1^{-1}t\widetilde{m}(\widetilde{w}\widetilde{m}-1)|^2}dH_x(t)\int \frac{t^2c_2^{-1}}{|1+tc_2^{-1}\widetilde{w}|^2}dH_y(t)\Bigg\}^{1/2}.
\end{aligned}
\tag{B.19}
$$

On the other hand, taking the imaginary part on both sides of the second equation in (B.13) and (B.14), we obtain

$$
\Im(\overline{m}) = \int \frac{t^2c_2^{-1}\Im(w)}{|1+tc_2^{-1}w|^2}dH_y(t), \tag{B.20}
$$

$$
\begin{aligned}
\Im(w) &= \Im(wm\overline{z}-\overline{z})\int \frac{t}{|z|^2}+\frac{t}{|z+c_1^{-1}tm\big(wm-1\big)|^2}dH_x(t)\\
&\quad + \Im(\overline{m})\int \frac{t^2c_1^{-1}|wm-1|^2}{|z+c_1^{-1}tm(wm-1)|^2}dH_x(t).
\end{aligned}
\tag{B.21}
$$

Further, if it holds

$$
\Im(wm\overline{z}-\overline{z}) > |z|\Im(w)\int t+\frac{t}{|1+tc_2^{-1}w|^2}dH_y(t), \tag{B.22}
$$

then for $w\in\mathbb{C}^+$, combining the above three equations (B.20), (B.21) and

(B.22) will lead to

$$1 > \int \frac{t}{|z|} + \frac{t|z|}{|z + c_1^{-1} tm\big(wm-1\big)|^2} dH_x(t) \int t + \frac{t}{|1+tc_2^{-1}w|^2} dH_y(t)$$
$$+ \int \frac{t^2 c_1^{-1}|wm-1|^2}{|z + c_1^{-1} tm(wm-1)|^2} dH_x(t) \int \frac{t^2 c_2^{-1}}{|1+tc_2^{-1}w|^2} dH_y(t). \quad \text{(B.23)}$$

Such inequality also holds true if we replace w and m by $\tilde{w}$ and $\tilde{m}$, that is,

$$1 > \int \frac{t}{|z|} + \frac{t|z|}{|z + c_1^{-1} t\tilde{m}\big(\tilde{w}\tilde{m}-1\big)|^2} dH_x(t) \int t + \frac{t}{|1+tc_2^{-1}\tilde{w}|^2} dH_y(t)$$
$$+ \int \frac{t^2 c_1^{-1}|\tilde{w}\tilde{m}-1|^2}{|z + c_1^{-1} t\tilde{m}(\tilde{w}\tilde{m}-1)|^2} dH_x(t) \int \frac{t^2 c_2^{-1}}{|1+tc_2^{-1}\tilde{w}|^2} dH_y(t). \quad \text{(B.24)}$$

Combining (B.23) and (B.24) will lead to a contradiction to (B.19), which means that we could only have one solution (s, w, m) satisfying the system of equations (3.1) on the set (3.2).

So it is sufficient to prove the assertion (B.22) on some open set of $\mathbb{C}^+$. In fact, using the first and second equations in (B.13), we have

$$\Im(wm\overline{z} - \overline{z}) = |z|^2 \Im(s),$$
$$\Im(zs) = \Im(wm) = \int t + \frac{t}{|1+tc_2^{-1}w|^2} dH_y(t) \Im(w).$$

Then assertion (B.22) is equivalent to

$$\Im(s) > \frac{1}{|z|}\Im(zs). \quad \text{(B.25)}$$

Actually, for any subsequence $\{n_k\}$ such that

$$s_{n_k}(z) = \frac{1}{n_k}\mathbb{E}\mathrm{tr}(\mathbf{B}_{n_k} - zI_n)^{-1}$$

converges, the empirical distribution $F^{\mathbf{B}_{n_k}}$ has a limit F (may depend on $\{n_k\}$), as $k \to \infty$, whose support is bounded upward by a constant, say K, which dose not depend on $\{n_k\}$. Moreover, the limit $s(z)$ of $s_{n_k}(z)$ is the Stieltjes transform of F, i.e.

$$s(z) = \int \frac{1}{x-z} dF(x).$$

This implies

$$\Im(s(z)) = \int \frac{1}{|x-z|^2} dF(x) \Im(z),$$
$$\Im(zs(z)) = \int \frac{x}{|x-z|^2} dF(x) \Im(z).$$

Therefore, (B.25) is true whenever $|z| > K$, which completes our proof.

Non-Gaussian case: since the two sets of samples $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$ are independent, we first fix the sequence of matrices $(\mathbf{A}_n)$ and show that, without the Gaussian assumption, the empirical spectral distribution $F^{\mathbf{S}_{xy}}$ will still converge weakly to the same spectral distribution F under Assumptions (a)-(c). Next, the same trick can be applied to $\{\mathbf{x}_i\}$, which will not be detailed here. Our strategy to remove the Gaussian assumption is based on Lemma 3, an extension of Lindeberg's argument for general smooth functions, see also Corollary 1.2 in Chatterjee (2008). As a special case, letting g be the identity function and f be the Stieltjes transform, the theorem will ensure that the order of the difference in expectation between the two Stieltjes transforms under the Gaussian distribution and a non-Gaussian one is $O(n^{-1/2})$ whenever the two distributions match the first two moments and have finite fourth moment. Hence, such difference can be negligible as $n \to \infty$, by which and the "Step 1" for Gaussian case the proof is done.

Recall that

$$\mathbf{B}_n = \mathbf{A}_n^{1/2} \Big(\frac{1}{q} \mathbf{Y}' \mathbf{Y} + \gamma_y \mathbf{I} \Big) \mathbf{A}_n^{1/2} = \mathbf{A}_n^{1/2} \Big(\frac{1}{q} \mathbf{W}' \mathbf{\Sigma}_y \mathbf{W} + \gamma_y I \Big) \mathbf{A}_n^{1/2},$$

where the table $\mathbf{W}$ consists i.i.d. standard Gaussian random variables and we vectorize it as a qn-dimensional random vector, denoted as $\mathbf{w} = (w_{ij})$. Therefore, the Stieltjes transform $s_n(z)$ of $F^{\mathbf{B}_n}$ can be viewed as a function of the random vector $\mathbf{w}$, defined as

$$U := f(\mathbf{w}) = \frac{1}{n} \mathrm{tr}(\mathbf{B}_n - z\mathbf{I})^{-1},$$

Similarly, we denote by

$$V := f(\tilde{\mathbf{w}})$$

the non-Gaussian counterpart of U, where $\tilde{\mathbf{w}} = (\tilde{w}_{ij})$ have the same first two moments as $\{w_{ij}\}$ and finite fourth moment. Let $\bar{\mathbf{w}} = (\bar{w}_{ij})$ be a mixture of

$\mathbf{w}$ and $\tilde{\mathbf{w}}$ by taking $\bar{w}_{ij} \in \{w_{ij}, \tilde{w}_{ij}\}$ for $i = p+1, \ldots, p+q$ and $j = 1, \ldots, n$, whose matrix form is denoted by $\overline{W}$. Applying Lemma 3, one gets

$$|\mathbb{E}(U) - \mathbb{E}(V)| \leq Kqn\lambda_3(f), \tag{B.26}$$

where

$$\lambda_3(f) = \sup\left\{ \left|\frac{\partial^k f(\bar{\mathbf{w}})}{\partial \bar{w}_{ij}^k}\right|^{3/k} : p+1 \leq i \leq p+q, 1 \leq j \leq n, 1 \leq k \leq 3, \bar{\mathbf{w}} \in \mathbb{R}^{qn} \right\}.$$

Hence, the remaining work is to find a bound for $\lambda_3(f)$, which can be achieved from bounding the first three derivatives of f with respect to $\bar{w}_{ij}$. To this end, following the same truncation, centralization and rescaling steps as in Bai and Silverstein (2010) (see Eq. (4.3.4)) and the "no eigenvalues" argument under finite fourth moment condition in Bai and Silverstein (1998), without loss of generality, we assume that the atoms $(\bar{w}_{ij})$ satisfy the following:

$$\mathbb{E}(\bar{w}_{ij}) = 0,\ \mathbb{V}ar(\bar{w}_{ij}) = 1,\ |\bar{w}_{ij}| \leq \sqrt{n},\ \mathbf{e}_i'\overline{\mathbf{W}}\,\overline{\mathbf{W}}'\mathbf{e}_i \leq Kn,$$

for all i and j, where the vector $\mathbf{e}_i$ is the ith canonical basis on $\mathbb{R}^q$. For convenience, we still use notations $(w_{ij}, \mathbf{w}, \mathbf{W})$ instead of $(\bar{w}_{ij}, \bar{\mathbf{w}}, \overline{\mathbf{W}})$ in what follows.

Let $\mathbf{G} = (\mathbf{B}_n - zI)^{-1}$, then the first three derivatives of $f(\mathbf{w})$ with respect to w_{ij} are the following:

$$\begin{aligned}
\frac{\partial f(\mathbf{w})}{\partial w_{ij}} &= \frac{1}{n}\mathrm{tr}\mathbf{G}' = -\frac{1}{n}\mathrm{tr}\mathbf{B}_n'\mathbf{G}^2,\\
\frac{\partial^2 f(\mathbf{w})}{\partial w_{ij}^2} &= -\frac{1}{n}\mathrm{tr}(\mathbf{B}_n''\mathbf{G}^2 + 2\mathbf{B}_n'\mathbf{G}\mathbf{G}') = -\frac{1}{n}\mathrm{tr}\mathbf{B}_n''\mathbf{G}^2 + \frac{2}{n}\mathrm{tr}\mathbf{B}_n'\mathbf{G}^2\mathbf{B}_n'\mathbf{G},\\
\frac{\partial^3 f(\mathbf{w})}{\partial w_{ij}^3} &= \frac{4}{n}\mathrm{tr}\mathbf{B}_n''\mathbf{G}^2\mathbf{B}_n' - \frac{6}{n}\mathrm{tr}\mathbf{B}_n'\mathbf{G}^2\mathbf{B}_n'\mathbf{G}\mathbf{B}_n'\mathbf{G} + \frac{2}{n}\mathrm{tr}\mathbf{B}_n'\mathbf{G}^2\mathbf{B}_n''\mathbf{G},
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{G}' &= -\mathbf{G}\mathbf{B}_n'\mathbf{G},\\
\mathbf{B}_n' &= \frac{1}{q}\mathbf{A}_n^{1/2}(\mathbf{e}_j\mathbf{e}_i'\mathbf{\Sigma}_y\mathbf{W} + \mathbf{W}'\mathbf{\Sigma}_y\mathbf{e}_i\mathbf{e}_j')\mathbf{A}_n^{1/2},\\
\mathbf{B}_n'' &= \frac{2}{q}\mathbf{A}_n^{1/2}\mathbf{e}_j\mathbf{e}_i'\mathbf{\Sigma}_y\mathbf{e}_i\mathbf{e}_j'\mathbf{A}_n^{1/2}.
\end{aligned}$$

and the vector $\mathbf{e}_j$ is the jth canonical basis on $\mathbb{R}^n$.

For the first derivative of f, since $\mathbf{\Sigma}_y$, $\mathbf{A}_n^{1/2}$ and $\mathbf{G}^2$ are all normal, we have

$$\begin{aligned}\sup\left|\frac{\partial f(\mathbf{w})}{\partial w_{ij}}\right| &\leq \sup\left\{\frac{1}{nq}\left|\mathrm{tr}\mathbf{A}_n^{1/2}\mathbf{e}_j\mathbf{e}_i'\mathbf{\Sigma}_y\mathbf{W}\mathbf{A}_n^{1/2}\mathbf{G}^2\right| + \frac{1}{nq}\left|\mathrm{tr}\mathbf{A}_n^{1/2}\mathbf{W}'\mathbf{\Sigma}_y\mathbf{e}_i\mathbf{e}_j'\mathbf{A}_n^{1/2}\mathbf{G}^2\right|\right\} \\ &\leq \sup\left\{\frac{K}{nq}\|\mathbf{e}_j\|\|\mathbf{e}_i'\mathbf{W}\| + \frac{K}{nq}\|\mathbf{W}'\mathbf{e}_i\|\|\mathbf{e}_j'\|\right\} \\ &\leq Kn^{-3/2}. \end{aligned} \tag{B.27}$$

For the second derivative, we have

$$\left|\frac{1}{n}\mathrm{tr}\mathbf{B}_n''\mathbf{G}^2\right| = \frac{2}{nq}\left|\mathrm{tr}\mathbf{A}_n^{1/2}\mathbf{e}_j\mathbf{e}_i'\mathbf{\Sigma}_y\mathbf{e}_i\mathbf{e}_j'\mathbf{A}_n^{1/2}\mathbf{G}^2\right| \leq \frac{K}{nq}\|\mathbf{e}_j\|\cdot\|\mathbf{e}_i'\mathbf{e}_i\mathbf{e}_j'\| \leq Kn^{-2}$$

and

$$\begin{aligned}&\left|\frac{2}{n}\mathrm{tr}\mathbf{B}_n'\mathbf{G}^2\mathbf{B}_n'\mathbf{G}\right| \\ &= \frac{2}{nq^2}\left|\mathrm{tr}\mathbf{A}_n^{1/2}(\mathbf{e}_j\mathbf{e}_i'\mathbf{\Sigma}_y\mathbf{W} + \mathbf{W}'\mathbf{\Sigma}_y\mathbf{e}_i\mathbf{e}_j')\mathbf{A}_n^{1/2}\mathbf{G}^2\mathbf{A}_n^{1/2}(\mathbf{e}_j\mathbf{e}_i'\mathbf{\Sigma}_y\mathbf{W} + \mathbf{W}'\mathbf{\Sigma}_y\mathbf{e}_i\mathbf{e}_j')\mathbf{A}_n^{1/2}\mathbf{G}\right| \\ &\leq \frac{K}{nq^2}\left(\|\mathbf{e}_j\|\|\mathbf{e}_i'\mathbf{W}\mathbf{e}_j\mathbf{e}_i'\mathbf{W}\| + \|\mathbf{e}_j\|\|\mathbf{e}_i'\mathbf{W}\mathbf{W}'\mathbf{e}_i\mathbf{e}_j'\| + \|\mathbf{W}'\mathbf{e}_i\|\|\mathbf{e}_j'\mathbf{e}_j\mathbf{e}_i'\mathbf{W}\| + \|\mathbf{W}'\mathbf{e}_i\|\|\mathbf{e}_j'\mathbf{W}'\mathbf{e}_i\mathbf{e}_j'\|\right) \\ &\leq \frac{K}{nq^2}\left(n + \sqrt{n}\cdot|w_{ij}|\right) \\ &\leq Kn^{-2}, \end{aligned}$$

which leads to the conclusion that

$$\sup\left|\frac{\partial^2 f(\mathbf{w})}{\partial w_{ij}^2}\right| \leq Kn^{-2}. \tag{B.28}$$

Similarly, we could bound the third derivative as follows,

$$\begin{aligned}\sup\left|\frac{\partial^3 f(\mathbf{w})}{\partial w_{ij}^3}\right| &\leq \sup\left\{\frac{K}{nq^3}\left(\|\mathbf{e}_i'\mathbf{W}\||w_{ij}|^2 + 2|w_{ij}||\mathbf{e}_i'\mathbf{W}\mathbf{W}'\mathbf{e}_i| + \|\mathbf{e}_i'\mathbf{W}\||\mathbf{e}_i'\mathbf{W}\mathbf{W}'\mathbf{e}_i|\right)\right. \\ &\qquad \left. + \frac{K}{nq^2}\left(\|\mathbf{e}_i'\mathbf{W}\| + |w_{ij}|\right)\right\} \\ &\leq Kn^{-5/2}. \end{aligned} \tag{B.29}$$

Finally, combing (B.27), (B.28) and (B.29) gives

$$\lambda_3(f) = \sup\left\{\left|\frac{\partial f}{\partial w_{ij}}\right|^3, \left|\frac{\partial^2 f}{\partial w_{ij}^2}\right|^{\frac{3}{2}}, \left|\frac{\partial^3 f}{\partial w_{ij}^3}\right|\right\} = Kn^{-5/2},$$

which together with (B.26) imply

$$|\mathbb{E}(U) - \mathbb{E}(V)| \leq Kn^{-1/2} \to 0, \quad \text{as } n \to \infty.$$

The proof is done.

B.4 Proof of Theorem 4

Under our model setting (4.1), the three data matrices $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ are related as:

$$\mathbf{Z} = \Gamma\mathbf{X}\mathbf{S} + \mathbf{Y},$$

where $\Gamma = \sum_{k=1}^{m} \theta_k \mathbf{u}_k \mathbf{v}_k'$ and $\mathbf{S} = \mathrm{Diag}(\varepsilon_1, \ldots, \varepsilon_n)$. So we have

$$\begin{aligned}\frac{1}{q}\mathbf{Z}'\mathbf{Z} &= \frac{1}{q}\mathbf{Y}'\mathbf{Y} + \frac{1}{q}\mathbf{S}\mathbf{X}'\Gamma'\Gamma\mathbf{X}\mathbf{S} + \frac{1}{q}\mathbf{S}\mathbf{X}'\Gamma'\mathbf{Y} + \frac{1}{q}\mathbf{Y}'\Gamma\mathbf{X}\mathbf{S} \\ &\triangleq \frac{1}{q}\mathbf{Y}'\mathbf{Y} + \mathbf{H},\end{aligned}$$

where

$$\mathbf{H} = \frac{1}{q}\mathbf{S}\mathbf{X}'\Gamma'\Gamma\mathbf{X}\mathbf{S} + \frac{1}{q}\mathbf{S}\mathbf{X}'\Gamma'\mathbf{Y} + \frac{1}{q}\mathbf{Y}'\Gamma\mathbf{X}\mathbf{S} \tag{B.30}$$

is a matrix of finite rank, at most $2m$. Denote

$$\widetilde{\mathbf{S}}_{xz} = \mathbf{A}_n^{1/2}\left(\frac{1}{q}\mathbf{Z}'\mathbf{Z} + \gamma_z\mathbf{I}_n\right)\mathbf{A}_n^{1/2} \quad \text{and} \quad \widehat{\mathbf{S}}_{xz} = \mathbf{A}_n^{1/2}\left(\frac{1}{q}\mathbf{Y}'\mathbf{Y} + \gamma_z\mathbf{I}_n\right)\mathbf{A}_n^{1/2},$$

where

$$\gamma_z = \frac{1}{q}\mathrm{tr}(\mathbf{\Sigma}_z) = \gamma_y + \frac{1}{q}\sum_{i=1}^{m}\theta_i^2 \cdot \gamma_x = \gamma_y + o(1). \tag{B.31}$$

Applying Lemma 2 to $\mathbf{B}_n$, $\widetilde{\mathbf{S}}_{xz}$ and $\widehat{\mathbf{S}}_{xz}$, we have

$$||F^{\widetilde{\mathbf{S}}_{xz}} - F^{\widehat{\mathbf{S}}_{xz}}|| \to 0 \quad \text{and} \quad L^3(F^{\mathbf{B}_n}, F^{\widehat{\mathbf{S}}_{xz}}) \to 0, \tag{B.32}$$

almost surely, as (n, p, q) tend to infinty. Combining (B.32) and the fact that $\widetilde{\mathbf{S}}_{xz}$ shares the same LSD as $\mathbf{S}_{xz}$, we conclude that $F^{\mathbf{S}_{xz}}$ converges weakly to the LSD F defined by (3). The proof is thus complete.

B.5 Proof of Theorem 5

We first note that, from the convergence in (B.4) and (B.31), asymptotically, the largest eigenvalues of $\mathbf{S}_{xz}$ are the same as those of

$$\bar{\mathbf{S}}_{xz} := \mathbf{A}_n^{1/2}\left(\frac{1}{q}\mathbf{Y}'\mathbf{Y} + \mathbf{H} + \gamma_y \mathbf{I}_n\right)\mathbf{A}_n^{1/2},$$

where $\mathbf{H}$ is given in (B.30). So it's equivalent to prove the theorem for $\bar{\mathbf{S}}_{xz}$.

Next, from Bai and Silverstein (1998) and the inequality

$$||\mathbf{A}_n^{1/2}\mathbf{C}_n\mathbf{A}_n^{1/2}|| \leq ||\mathbf{A}_n|| \cdot ||\mathbf{C}_n||,$$

we know that the spectral norm $||\mathbf{A}_n^{1/2}\mathbf{C}_n\mathbf{A}_n^{1/2}||$ is bounded in n, almost surely. Define

$$\lambda_+ = \limsup_{n\to\infty} ||\mathbf{A}_n^{1/2}\mathbf{C}_n\mathbf{A}_n^{1/2}||,$$

we consider the existence of spiked eigenvalues $(\lambda_{n,\ell})$ of $\bar{\mathbf{S}}_{xz}$ in the interval $(\lambda_+, +\infty)$. That is, for each $\ell \in \{1, \dots, k\}$, $\lambda_{n,\ell}$ is an eigenvalue of $\bar{\mathbf{S}}_{xz}$ but not an eigenvalue of $\mathbf{A}_n^{1/2}\mathbf{C}_n\mathbf{A}_n^{1/2}$, i.e.

$$\left|\lambda \mathbf{I}_n - \bar{\mathbf{S}}_{xz}\right| = 0 \quad \text{and} \quad \left|\lambda \mathbf{I}_n - \mathbf{A}_n^{1/2}\mathbf{C}_n\mathbf{A}_n^{1/2}\right| \neq 0, \tag{B.33}$$

for $\lambda \in \{\lambda_{n,1}, \dots, \lambda_{n,k}\}$.

In the following, we will show the limits of λ is defined in (4.4). Under the assumptions in (B.33), we have

$$\left|\mathbf{I}_n - \left(\lambda \mathbf{I}_n - \mathbf{A}_n^{1/2}\mathbf{C}_n\mathbf{A}_n^{1/2}\right)^{-1}\mathbf{A}_n^{1/2}\mathbf{H}\mathbf{A}_n^{1/2}\right| = 0. \tag{B.34}$$

Recall the definition of $\mathbf{H}$ in (B.30), then with a little bit calculation, this matrix can be decomposed as

$$\mathbf{H} = \frac{1}{q}\begin{pmatrix} \mathbf{a}_1 & \mathbf{b}_1 & \cdots & \mathbf{a}_m & \mathbf{b}_m \end{pmatrix} \begin{pmatrix} \theta_1\lambda_{11} & 0 & \cdots & 0 & 0 \\ 0 & \theta_1\lambda_{12} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \theta_m\lambda_{m1} & 0 \\ 0 & 0 & \cdots & 0 & \theta_m\lambda_{m2} \end{pmatrix} \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{b}_1' \\ \vdots \\ \mathbf{a}_m' \\ \mathbf{b}_m' \end{pmatrix}, \tag{B.35}$$

where

$$\mathbf{a}_i = u_{i1}\mathbf{S}\mathbf{X}'\mathbf{v}_i + w_{i1}\mathbf{Y}'\mathbf{u}_i,$$
$$\mathbf{b}_i = u_{i2}\mathbf{S}\mathbf{X}'\mathbf{v}_i + w_{i2}\mathbf{Y}'\mathbf{u}_i,$$
$$\lambda_{i1} = \|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|\|\mathbf{Y}'\mathbf{u}_i\| \left\{ \frac{\sqrt{4\|\mathbf{Y}'\mathbf{u}_i\|^2 + \theta_i^2\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|^2} + \theta_i\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|}{\sqrt{4\|\mathbf{Y}'\mathbf{u}_i\|^2 + \theta_i^2\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|^2} - \theta_i\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|} \right\}^{1/2},$$
$$\lambda_{i2} = -\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|\|\mathbf{Y}'\mathbf{u}_i\| \left\{ \frac{\sqrt{4\|\mathbf{Y}'\mathbf{u}_i\|^2 + \theta_i^2\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|^2} - \theta_i\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|}{\sqrt{4\|\mathbf{Y}'\mathbf{u}_i\|^2 + \theta_i^2\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|^2} + \theta_i\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|} \right\}^{1/2},$$

with

$$u_{i1} = \frac{1}{\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|} \left\{ \frac{1}{2} + \frac{\theta_i\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|}{2\sqrt{4\|\mathbf{Y}'\mathbf{u}_i\|^2 + \theta_i^2\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|^2}} \right\}^{1/2},$$
$$u_{i2} = \frac{1}{\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|} \left\{ \frac{1}{2} - \frac{\theta_i\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|}{2\sqrt{4\|\mathbf{Y}'\mathbf{u}_i\|^2 + \theta_1^2\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|^2}} \right\}^{1/2},$$
$$w_{i1} = \frac{1}{\|\mathbf{Y}'\mathbf{u}_i\|} \left\{ \frac{1}{2} - \frac{\theta_i\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|}{2\sqrt{4\|\mathbf{Y}'\mathbf{u}_i\|^2 + \theta_i^2\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|^2}} \right\}^{1/2},$$
$$w_{i2} = -\frac{1}{\|\mathbf{Y}'\mathbf{u}_i\|} \left\{ \frac{1}{2} + \frac{\theta_i\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|}{2\sqrt{4\|\mathbf{Y}'\mathbf{u}_i\|^2 + \theta_i^2\|\mathbf{S}\mathbf{X}'\mathbf{v}_i\|^2}} \right\}^{1/2}.$$

In addition, it's straightforward to verify the following relations,

$$\begin{cases} \lambda_{i1}u_{i1}^2 + \lambda_{i2}u_{i2}^2 = \theta_i, \\ \lambda_{i1}w_{i1}^2 + \lambda_{i2}w_{i2}^2 = 0, \\ \lambda_{i1}u_{i1}w_{i1} + \lambda_{i2}u_{i2}w_{i2} = 1. \end{cases} \tag{B.36}$$

Denote $\mathbf{D}_n = \mathbf{A}_n^{1/2}\left(\lambda\mathbf{I}_n - \mathbf{A}_n^{1/2}\mathbf{C}_n\mathbf{A}_n^{1/2}\right)^{-1}\mathbf{A}_n^{1/2}$ and

$$\mathbf{M}_n = \frac{1}{q} \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{b}_1' \\ \vdots \\ \mathbf{a}_m' \\ \mathbf{b}_m' \end{pmatrix} \mathbf{D}_n \begin{pmatrix} \mathbf{a}_1 & \mathbf{b}_1 & \cdots & \mathbf{a}_m & \mathbf{b}_m \end{pmatrix} \begin{pmatrix} \theta_1\lambda_{11} & 0 & \cdots & 0 & 0 \\ 0 & \theta_1\lambda_{12} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \theta_m\lambda_{m1} & 0 \\ 0 & 0 & \cdots & 0 & \theta_m\lambda_{m2} \end{pmatrix}.$$

Then (B.34) and (B.35) imply

$$f_n(\lambda) := |\mathbf{I}_{2m} - \mathbf{M}_n| = 0.$$

We next find the limit of $f_n(\lambda)$. Let

$$\alpha_n = \frac{1}{n}\mathrm{tr}\mathbf{S}\mathbf{D}_n\mathbf{S}(\mathbf{A}_n - \gamma_x\mathbf{I}_n) \quad \text{and} \quad \beta_n = \frac{1}{n}\mathrm{tr}\mathbf{D}_n(\mathbf{C}_n - \gamma_y\mathbf{I}_n),$$

one may get for any $i \in \{1, \ldots, m\}$,

$$\begin{aligned}
\frac{\mathbf{a}_i'\mathbf{D}_n\mathbf{a}_i}{q} &= \frac{u_{i1}^2}{c_{n2}}\alpha_n + \frac{w_{i1}^2}{c_{n2}}\beta_n + o_{a.s.}(1), \\
\frac{\mathbf{a}_i'\mathbf{D}_n\mathbf{b}_i}{q} &= \frac{u_{i1}u_{i2}}{c_{n2}}\alpha_n + \frac{w_{i1}w_{i2}}{c_{n2}}\beta_n + o_{a.s.}(1), \\
\frac{\mathbf{b}_i'\mathbf{D}_n\mathbf{b}_i}{q} &= \frac{u_{i2}^2}{c_{n2}}\alpha_n + \frac{w_{i2}^2}{c_{n2}}\beta_n + o_{a.s.}(1) \ ,
\end{aligned}$$

and for any $i \neq j \in \{1, \ldots, m\}$,

$$\frac{\mathbf{a}_i'\mathbf{D}_n\mathbf{a}_j}{q} = o_{a.s.}(1) \ , \quad \frac{\mathbf{a}_i'\mathbf{D}_n\mathbf{b}_j}{q} = o_{a.s.}(1) \ .$$

¿From the above approximations and the identities in (B.36), we have

$$f_n(\lambda) = \prod_{k=1}^{m} \left|\mathbf{I}_2 - \mathbf{M}_{nk}\right| + o_{a.s}(1)$$

where

$$\mathbf{M}_{nk} = \frac{\theta_k}{c_{n2}} \begin{pmatrix} \alpha_n & 0 \\ 0 & \beta_n \end{pmatrix} \begin{pmatrix} \theta_k & 1 \\ 1 & 0 \end{pmatrix}. \tag{B.37}$$

Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$, then

$$\alpha_n = \frac{1}{n}\mathrm{tr}\mathbf{S}\mathbf{D}_n\mathbf{S}\mathbf{A}_n - \frac{\gamma_x}{n}\mathrm{tr}\mathbf{S}\mathbf{D}_n\mathbf{S} = \frac{1}{n}\boldsymbol{\varepsilon}'\left(\mathbf{D}_n \circ \mathbf{A}_n\right)\boldsymbol{\varepsilon} - \frac{\gamma_x}{n}\boldsymbol{\varepsilon}'\mathrm{Diag}(\mathbf{D}_n)\boldsymbol{\varepsilon}, \tag{B.38}$$

where "$\circ$" denotes the Hadamard product of two matrices. According to Theorem 1 of Varberg (1968), we have

$$\frac{1}{n}\boldsymbol{\varepsilon}'\left(\mathbf{D}_n \circ \mathbf{A}_n\right)\boldsymbol{\varepsilon} - \frac{1}{n}\mathbb{E}\left[\boldsymbol{\varepsilon}'\left(\mathbf{D}_n \circ \mathbf{A}_n\right)\boldsymbol{\varepsilon}\right] \xrightarrow{a.s.} 0, \tag{B.39}$$

$$\frac{1}{n}\boldsymbol{\varepsilon}'\mathrm{Diag}(\mathbf{D}_n)\boldsymbol{\varepsilon} - \frac{1}{n}\mathbb{E}\mathrm{tr}\mathbf{D}_n \xrightarrow{a.s.} 0. \tag{B.40}$$

Further,

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\left[\boldsymbol{\varepsilon}'\left(\mathbf{D}_n \circ \mathbf{A}_n\right)\boldsymbol{\varepsilon}\right] = \frac{1}{n}\mathbb{E}\mathrm{tr}\big[\mathbf{D}_n \mathrm{Diag}(\mathbf{A}_n)\big] \\
=& \frac{1}{n}\mathbb{E}\mathrm{tr}\big[\mathbf{D}_n\left(\mathrm{Diag}(\mathbf{A}_n) - 2\gamma_x \mathbf{I}_n\right)\big] + \frac{2\gamma_x}{n}\mathbb{E}\mathrm{tr}\mathbf{D}_n \\
=& \frac{2\gamma_x}{n}\mathbb{E}\mathrm{tr}\mathbf{D}_n + o(1), \qquad \text{(B.41)}
\end{aligned}$$

where the last equality is due to the following convergence,

$$\left|\frac{1}{n}\mathrm{tr}\big[\mathbf{D}_n \cdot \left(\mathrm{Diag}(\mathbf{A}_n) - 2\gamma_x \mathbf{I}_n\right)\big]\right| \le \frac{1}{n}\|\mathbf{D}_n\| \cdot \mathrm{tr}\big|\mathbf{A}_n - 2\gamma_x \mathbf{I}_n\big| \xrightarrow{a.s.} 0.$$

Collecting results in (B.38)-(B.41), we get

$$\alpha_n = -\gamma_x w_n(\lambda) + o_{a.s.}(1) \xrightarrow{a.s.} \alpha \triangleq -w(\lambda)\int t dH_x(t), \qquad \text{(B.42)}$$

where $w_n(z)$ is defined in (B.9), whose domain can be expanded to $(\lambda_+, +\infty)$ for all large n. For β_n, we have

$$\begin{aligned}
\beta_n &= \frac{1}{n}\mathrm{tr}(\mathbf{D}_n \mathbf{C}_n) - \frac{\gamma_y}{n}\mathrm{tr}\mathbf{D}_n \\
&= -1 + \frac{\lambda}{n}\mathrm{tr}\left(\lambda \mathbf{I}_n - \mathbf{A}_n^{1/2}\mathbf{C}_n \mathbf{A}_n^{1/2}\right)^{-1} - \frac{\gamma_y}{n}\mathrm{tr}\mathbf{D}_n \\
&= -\frac{1}{n}\sum_{k=1}^{q}\frac{\tau_k w_n(\lambda)}{c_{n2} + \tau_k w_n(\lambda)} + o_{a.s.}(1) \\
&\xrightarrow{a.s.} \beta \triangleq -c_2 \int \frac{t w(\lambda) dH_y(t)}{c_2 + t w(\lambda)}, \qquad \text{(B.43)}
\end{aligned}$$

where the third equality is from (B.11) with (τ_k) being the eigenvalues of $\boldsymbol{\Sigma}_y$. Collecting results in (B.37),(B.42) and (B.43), we get

$$f_n(\lambda) \xrightarrow{a.s.} f(\lambda) \triangleq \prod_{k=1}^{m}\left(1 - \theta_k^2 g(\lambda)\right),$$

where the function g is given in (4.2). With the definition of the critical value θ_0 in (4.3), we find that for any $k \in \{1, \ldots, m\}$ and $\theta_k > \theta_0$, there are k zeros $\lambda_1 > \cdots > \lambda_k$ of $f(\lambda)$ on (λ_+, ∞). By continuity arguments, see Lemma 6.1 in Benaych-Georges and Nadakuditi (2011), we verify the existence of the spikes $\lambda_{n,1}, \ldots, \lambda_{n,k}$ whose limits are $\lambda_1, \ldots, \lambda_k$, respectively. The proof is then complete.

References

Bai, Z.D. and Silverstein, J.W. (1998) No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices. *Ann. Probab.*, **26**, 316–345.

Bai, Z.D. and Silverstein, J.W. (2010) *Spectral analysis of large dimensional random matrices*, 2nd ed., Springer, New York.

Bai, Z.D. and Zhou, W. (2008) Large sample covariance matrices without independence structures in columns. *Stat. Sinica*, **18**, 425–442.

Benaych-georges, F. and Nadakuditi, R.R. (2011) The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, **227**(1), 494–521.

Chatterjee, S. (2008) A simple invariance theorem. *Preprint*, available at arXiv: http://arxiv.org/pdf/math/0508213.pdfmath/0508213.

El Karoui, N. (2010) The spectrum of kernel random matrices. *Ann. Stat.*, **38**, 1–50.

Li, W.M. and Yao, J.F. (2018) On structure testing for component covariance matrices of a high dimensional mixture. *J. Roy. Stat. Soc. Ser. B*, **80**(2), 293-318.

Silverstein, J.W. (1995) Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.*, **55**, 331–339.

Varberg, Dale E. (1968) Almost sure convergence of quadratic forms in independent random variable. *Ann. Math. Statist.*, **39**(5), 1502-1506.