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 2.1, Theorem 2.2, Theorem 3.1, Theorem 4.1 and Theorem 4.2 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 \( M \) with entries

\[
M_{i,j} = f\left( \frac{\|x_i - x_j\|_2^2}{p} \right).
\]
Let us call \( \psi \) the vector with \( i \)-th entry \( \psi_i = \|x_i\|_2^2/p - \tau/2 \), where \( \tau = 2\text{tr}(\Sigma_p)/p \). We assume that:

(a) \( n \asymp p \), that is, \( n/p \) and \( p/n \) remain bounded as \( p \to \infty \).

(b) \( \Sigma_p \) is a positive semi-definite \( p \times p \) matrix, and \( \|\Sigma_p\| = \sigma_1(\Sigma_p) \) remains bounded in \( p \), that is, there exists \( K > 0 \), such that \( \sigma_1(\Sigma_p) \leq K \), for all \( p \).

(c) There exists \( \ell \in \mathbb{R} \) such that \( \lim_{p \to \infty} \text{tr}(\Sigma_p)/p = \ell \).

(d) \( X = (x_1, \ldots, x_n) \) and \( x_i = \Sigma_p^{1/2}w_i \) for \( i = 1, \ldots, n \).

(e) The entries of \( w_i \), a \( p \)-dimensional random vector, are i.i.d. Also, denoting by \( w_{ik} \) the \( k \)th entry of \( w_i \), we assume that \( \mathbb{E}(w_{ik}) = 0 \), \( \mathbb{V}(w_{ik}) = 1 \) and \( \mathbb{E}(|w_{ik}|^{5+\varepsilon}) < \infty \) for some \( \varepsilon > 0 \).

(f) \( f \) is \( C^3 \) in a neighborhood of \( \tau \).

Then \( M \) can be approximated consistently in operator norm (and in probability) by the matrix \( \widetilde{M} \), defined by

\[
\widetilde{M} = f(\tau)11' + f'(\tau) \left[ 1 \psi' + \psi 1' - 2 \frac{X'X}{p} \right] \\
+ \frac{f''(\tau)}{2} \left[ 1(\psi \circ \psi)' + (\psi \circ \psi)1' + 2 \psi \psi' + 4 \frac{\text{tr}(\Sigma_p^2)}{p^2}11' \right] + v_p I_n,
\]

\[ v_p = f(0) + \tau f'(\tau) - f(\tau). \]

In other words,

\[ \|M - \widetilde{M}\| \to 0, \quad \text{in probability}. \]

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

\[ \| F^A - F^B \| \leq \frac{1}{n} \text{rank}(A - B) \quad \text{and} \quad L^3(F^A, F^B) \leq \frac{1}{n} \text{tr}[(A - B)(A - B)^*], \]

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 \( x = (x_1, \ldots, x_n)' \) and \( y = (y_1, \ldots, y_n)' \) be two random vectors in \( \mathbb{R}^n \) with i.i.d. elements, respectively, and set \( U = f(x) \) and \( V = f(y) \). If

\[ \gamma = \max\{E|x_i|^3, E|y_i|^3, 1 \leq i \leq n\} < \infty, \]

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

\[ |Eg(U) - Eg(V)| \leq 2C_2(g)\gamma n\lambda_3(f), \]

where \( \lambda_3(f) = \sup \{|\partial^k f(z)|^{3/k} : z = (z_\ell), z_\ell \in \{x_\ell, y_\ell\}, 1 \leq i \leq n, 1 \leq k \leq 3\} \)

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.

\[ V_x = \left( \|x_k - x_\ell\| / \sqrt{p} \right), \quad V_y = \left( \|y_k - y_\ell\| / \sqrt{q} \right), \quad P_n = I_n - \frac{1}{n}1_n1_n', \]

\[ \gamma_x = \frac{1}{p} \text{tr} \Sigma_x, \quad \gamma_y = \frac{1}{q} \text{tr} \Sigma_y, \quad \kappa_x = \frac{1}{pn} \sum_{i=1}^{n} ||x_i||^2, \quad \kappa_y = \frac{1}{qn} \sum_{i=1}^{n} ||y_i||^2, \]

\[ A_n = \frac{1}{p} X'X + \gamma_x I_n, \quad C_n = \frac{1}{q} Y'Y + \gamma_y I_n, \quad B_n = A_n^\frac{1}{2} C_n A_n^\frac{1}{2}, \]

\[ D_x = \frac{1}{p} X'X + \kappa_x I_n, \quad D_y = \frac{1}{q} Y'Y + \kappa_y I_n, \quad D_z = \frac{1}{q} Z'Z + \kappa_z I_n, \]

\[ S_{xy} = P_n D_x P_n D_y P_n, \quad S_{zz} = P_n D_z P_n D_z P_n. \]

**B.1 Proof of Theorem 2.1**

The squared sample distance covariance \( \nu^2_n(x, y) \) in (1.2) can be expressed as an inner product between the two matrices \( P_n V_x P_n \) and \( P_n V_y P_n \), that is,

\[ \nu^2_n(x, y) = \frac{\sqrt{pq}}{n^2} \text{tr} P_n V_x P_n V_y P_n. \]

Notice that the matrices \( V_x \) and \( V_y \) 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

\[ P_n V_x P_n V_y P_n \]  

(B.1)
can be approximated by a simplified random matrix $V_n$ such that as $(n, p, q)$ tend to infinity,

$$
\|V_n - P_n V_x P_y V_y P_n\| \to 0 \quad \text{(B.2)}
$$

in probability, where

$$
V_n \triangleq \frac{1}{2\sqrt{\gamma_x \gamma_y}} P_n \left( A_n + \frac{1}{8\gamma_x} \psi_x \psi_x' \right) P_n \left( C_n + \frac{1}{8\gamma_y} \psi_y \psi_y' \right) P_n, \quad \text{(B.3)}
$$

in which

$$
\psi_x = \frac{1}{p} \begin{pmatrix} \|x_1\|^2 - \text{tr} \Sigma_x \\ \vdots \\ \|x_n\|^2 - \text{tr} \Sigma_x \end{pmatrix} \quad \text{and} \quad \psi_y = \frac{1}{q} \begin{pmatrix} \|y_1\|^2 - \text{tr} \Sigma_y \\ \vdots \\ \|y_n\|^2 - \text{tr} \Sigma_y \end{pmatrix}.
$$

Then we replace the two traces $\gamma_x$ and $\gamma_y$ in $A_n$ and $C_n$ with their unbiased sample counterparts $\kappa_x$ and $\kappa_y$, respectively, which does not affect the convergence in (B.2). Finally in (B.3), by removing the two rank-one matrices $(8\gamma_x)^{-1} \psi_x \psi_x'$ and $(8\gamma_y)^{-1} \psi_y \psi_y'$ (which have bounded spectral norm, almost surely), we get the conclusion of the theorem. The proof is thus complete.

B.2 Proof of Theorem 2.2

Recall the approximation from Theorem 2.1

$$
\gamma_n^2(x, y) = \frac{1}{2n^2} \sqrt{pq} \frac{\gamma_x \gamma_y}{\gamma_x \gamma_y} \text{tr} S_{xy} + o_p(1)
$$
and notice that

\[
\frac{1}{n} \text{tr}(S_{xy}) = \frac{1}{npq} \text{tr}(P_n X' X P_n Y' Y P_n) + \frac{\kappa_x}{np} \text{tr}(P_n X' X) + \frac{\kappa_y}{nq} \text{tr}(P_n Y' Y) \\
+ \frac{n-1}{n} \kappa_x \kappa_y \\
= \frac{1}{npq} \text{tr}(X' X Y' Y) + 3\gamma_x \gamma_y + o_a.s(1).
\]

Moreover, from Equation (21) in [Li and Yao (2018)] and the independence between X and Y,

\[
\frac{1}{npq} \text{tr}(X' X Y' Y) = \frac{1}{p} \text{tr}(\Sigma_x) - \frac{1}{q} \text{tr}(\Sigma_y) + o_a.s(1).
\]

Collecting the above results yields

\[
\mathcal{V}_n^2(x, y) = 2\sqrt{c_1 c_2 \gamma_x \gamma_y} + o_p(1).
\]

On the other hand, applying Lemma 1, we have

\[
\frac{1}{n} S_{2,n} = \frac{1}{2n} \sqrt{\frac{pq}{\gamma_x \gamma_y}} \left( \frac{1}{n^2} 1' D_x 1 - 2\gamma_x \right) \left( \frac{1}{n^2} 1' D_y 1 - 2\gamma_y \right) + o_p(1) \\
= 2\sqrt{c_1 c_2 \gamma_x \gamma_y} + o_p(1).
\]

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

The proof is complete.

\section*{B.3 Proof of Theorem 3.1}

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 $\Sigma_x$ and $\Sigma_y$ 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

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

as $(n,p,q)$ tend to $\infty$. From Lemma 2 and (B.4), we get

$$L^3(F_{S_{xy}}, F_{B_n}) \xrightarrow{a.s.} 0.$$

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

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

**Step 2:** For any fixed $z \in \mathbb{C}^+$, $\mathbb{E}s_{B_n}(z) \to s(z)$ with $s(z)$ satisfies the equations in (3.1).
Step 3: The uniqueness of the solution $s(z)$ to (3.1) on the set (3.2).

Step 1. Almost sure convergence of $s_{B_n}(z) - E s_{B_n}(z)$.

We assume $\Sigma_y$ is diagonal, having the form

$$\Sigma_y = \text{Diag}(\tau_1, \ldots, \tau_q).$$

By this and notations

$$r_k = \frac{1}{\sqrt{q}} A_n^{1/2} (w_{p+k,1}, \ldots, w_{p+k,n})', \quad k = 1, \ldots, q,$$

the matrix $B_n$ can be expressed as

$$B_n = \gamma_y A_n + \sum_{k=1}^q \tau_k r_k r_k'.$$  \hspace{1cm} (B.5)

It’s “leave-one-out” version is denoted by $B_{k,n} = B_n - \tau_k r_k r_k'$, $k = 1, \ldots, q$.

Let $E_0(\cdot)$ be expectation and $E_k(\cdot)$ be conditional expectation given $r_1, \ldots, r_k$.

From the martingale decomposition and the identity

$$r_k'(B_n - zI_n)^{-1} = \frac{r_k'(B_{k,n} - zI_n)^{-1}}{1 + \tau_k r_k'(B_{k,n} - zI_n)^{-1}r_k},$$  \hspace{1cm} (B.6)

we have

$$s_{B_n}(z) - E s_{B_n}(z) = \frac{1}{n} \sum_{k=1}^q (E_k - E_{k-1}) \left[ \text{tr}(B_n - zI_n)^{-1} - \text{tr}(B_{k,n} - zI_n)^{-1} \right]$$

$$= -\frac{1}{n} \sum_{k=1}^q (E_k - E_{k-1}) \frac{\tau_k r_k'(B_{k,n} - zI_n)^{-2}r_k}{1 + \tau_k r_k'(B_{k,n} - zI_n)^{-1}r_k}.$$  \hspace{1cm} (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 $s_{B_n}(z) - \mathbb{E}s_{B_n}(z) \to 0$, almost surely.

**Step 2. Convergence of $\mathbb{E}s_{B_n}(z)$.**

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

$$z = -\frac{1}{s_A(z)} + \int t + \frac{t}{1 + tc^{-1}s_A(z)}dH_x(t). \quad \text{(B.8)}$$

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

$$w_n(z) = \frac{1}{n} \text{tr}(B_n - zI_n)^{-1}A_n \quad \text{and} \quad m_n(z) = \gamma_y + \frac{1}{q} \sum_{k=1}^{q} \frac{\tau_k}{1 + \tau_k c_n^{-1}w_n(z)}. \quad \text{(B.9)}$$

We first show that

$$m_n^{-1}(z)s_{A_n}[zm_n^{-1}(z)] - \mathbb{E}s_{B_n}(z) \to 0, \quad n \to \infty. \quad \text{(B.10)}$$

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

$$\frac{1}{n} \text{tr} [m_n(z)A_n - zI_n]^{-1} - \frac{1}{n} \text{tr}(B_n - zI_n)^{-1}$$

$$= \frac{1}{n} \text{tr} [m_n(z)A_n - zI_n]^{-1} \left( \sum_{k=1}^{q} \tau_k r_k r_k' - (m_n(z) - \gamma_y)A_n \right) (B_n - zI_n)^{-1}$$

$$= \frac{1}{n} \sum_{k=1}^{q} \frac{\tau_k r_k'(B_{k,n} - zI_n)^{-1}[(m_n(z)A_n - zI_n)^{-1} r_k}{1 + \tau_k r_k'(B_{k,n} - zI_n)^{-1}r_k}$$

$$- \frac{m_n(z) - \gamma_y}{n} \text{tr} [m_n(z)A_n - zI_n]^{-1} A_n (B_n - zI_n)^{-1}$$
B.3 Proof of Theorem 3.1

\[
= \frac{1}{n} \sum_{k=1}^{q} \frac{\tau_k d_k}{1 + \tau_k c_{n2}^{-1} w_n(z)},
\]

where

\[
d_k = \frac{1 + \tau_k c_{n2}^{-1} w_n(z)}{1 + \tau_k r_k^*(B_{k,n} - z I_n)^{-1} r_k} r_k^*(B_{k,n} - z I_n)^{-1} [m_n(z)A_n - z I_n]^{-1} r_k
- \frac{1}{q} \text{tr} [m_n(z)A_n - z I_n]^{-1} A_n (B_n - z I_n)^{-1}.
\]

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 \(\mathbb{E}s_{B_n}(z)\) and \(w_n(z)\) by proving

\[
1 + z \mathbb{E}s_{B_n}(z) - \gamma_y w_n(z) - \frac{1}{n} \sum_{k=1}^{q} \frac{\tau_k w_n(z)}{c_{n2} + \tau_k w_n(z)} \to 0. \quad (B.11)
\]

From the expression of \(B_n\) in (B.5) and the identity in (B.6), we have

\[
I_n + z (B_n - z I)^{-1} = B_n (B_n - z I_n)^{-1}
= \gamma_y A_n (B_n - z I)^{-1} + \sum_{k=1}^{q} \frac{\tau_k r_k^*(B_{k,n} - z I_n)^{-1}}{1 + \tau_k r_k^*(B_{k,n} - z I_n)^{-1} r_k}.
\]
B.3 Proof of Theorem 3.1

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

$$1 + z \frac{1}{n} \text{tr}(B_n - zI_n)^{-1} = \gamma y \frac{1}{n} \text{tr}(B_n - zI_n)^{-1} A_n + \frac{1}{n} \sum_{k=1}^{q} \frac{\tau_k r'_k(B_{k,n} - zI_n)^{-1} r_k}{1 + \tau_k r'_k(B_{k,n} - zI_n)^{-1} r_k} \tag{B.12}$$

$$= \gamma y \frac{1}{n} \text{tr}(B_n - zI_n)^{-1} A_n + \frac{1}{n} \sum_{k=1}^{q} \frac{\tau_k c_{n2}^{-1} w_n(z)}{1 + \tau_k c_{n2}^{-1} w_n(z)} + \varepsilon_n,$$

where

$$\varepsilon_n = \frac{1}{n} \sum_{k=1}^{q} \frac{\tau_k [c_{n2}^{-1} w_n(z) - r'_k(B_{k,n} - zI_n)^{-1} r_k]}{[1 + \tau_k r'_k(B_{k,n} - zI_n)^{-1} r_k][1 + \tau_k c_{n2}^{-1} w_n(z)]}.$$

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

$$\inf_n \left| [1 + \tau_k r'_k(B_{k,n} - zI_n)^{-1} r_k][1 + \tau_k c_{n2}^{-1} w_n(z)] \right| > 0.$$

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

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

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

$$m_{n_k}(z) \to \int t + \frac{t}{1 + tc_{n2}^{-1} w(z)} dH_y(t) \triangleq m(z),$$

$$s_{n_k}(z) \to \frac{1}{m(z)} s_A \left( \frac{z}{m(z)} \right),$$

$$zs_{n_k}(z) \to -1 + w(z) \int t + \frac{t}{1 + tc_{2}^{-1} w(z)} dH_y(t),$$

11
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:

\[
\begin{align*}
    s(z) &= \frac{1}{m(z)} s_A \left( \frac{z}{m(z)} \right), \\
    zs(z) &= -1 + w(z) \int t + \frac{t}{1 + tc_2^2 w(z)} dH_y(t), \\
    z &= -\frac{1}{s_A(z)} + \int t + \frac{t}{1 + tc_1^2 s(z)} dH_x(t), \\
    m(z) &= \int t + \frac{t}{1 + tc_2^1 w(z)} dH_y(t).
\end{align*}
\]

Cancelling the function \( s_A(z) \) from the above system yields an equivalent but simpler system of equations as shown in (3.1). 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 (3.1).**

The system of equations in (3.1) is equivalent to

\[
\begin{align*}
    1 + zs &= wm, \\
    m &= \int t + \frac{t}{1 + tc_2^1 w} dH_y(t),
    \tag{B.13} \\
    w &= s \int t + \frac{t}{1 + tc_1^1 (1 + zs)w^{-1} s} dH_x(t).
\end{align*}
\]

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

\[
w = \int \frac{t}{z} \left( \frac{wm - 1}{z + c_1^{-1} tm(wm - 1)} \right) dH_x(t). \tag{B.14}
\]

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.3) Proof of Theorem 3.1

(B.13) and (B.14), we can obtain

\[ w - \tilde{w} = (wm - \tilde{\omega}\tilde{m}) \]

\[ \times \int \left[ \frac{t}{z} + \frac{tz}{(z + c_1^{-1}tm(wm - 1))(z + c_1^{-1}t\tilde{m}(\tilde{\omega}\tilde{m} - 1))} \right] dH_x(t) \]

\[ + (\tilde{m} - m) \int \frac{t^2c_1^{-1}(wm - 1)(\tilde{\omega}\tilde{m} - 1)}{(z + c_1^{-1}tm(wm - 1))(z + c_1^{-1}t\tilde{m}(\tilde{\omega}\tilde{m} - 1))} dH_x(t), \]

(B.15)

\[ \tilde{m} - m = (w - \tilde{w}) \int \frac{t^2c_2^{-1}}{(1 + tc_2^{-1}w)(1 + tc_2^{-1}\tilde{w})} dH_y(t), \]

(B.16)

\[ wm - \tilde{w}\tilde{m} = (w - \tilde{w}) \int \left( t + \frac{t}{(1 + tc_2^{-1}w)(1 + tc_2^{-1}\tilde{w})} \right) dH_y(t). \]

(B.17)

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

\[ B_1B_2 + C_1C_2 = 1, \]

(B.18)

where

\[ B_1 = \int \frac{t}{z} + \frac{tz}{(z + c_1^{-1}tm(wm - 1))(z + c_1^{-1}t\tilde{m}(\tilde{\omega}\tilde{m} - 1))} dH_x(t), \]

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

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

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

By the Cauchy-Schwarz inequality, we have

\[ |B_1B_2|^2 \leq \int \left| \frac{t}{z} \right| + \frac{|tz|}{|z + c_1^{-1}tm(wm - 1)|^2} dH_x(t) \]
\[ \times \left\{ \int \frac{|t/z|}{|z + c_1^{-1} t \tilde{m}(\tilde{w} \tilde{m} - 1)|^2} dH_x(t) \right\} \]

\[ \times \left\{ \int \frac{t + |t|}{|1 + tc_2^{-1} w|^2} dH_y(t) \right\} \int \frac{t + |t|}{|1 + tc_2^{-1} \tilde{w}|^2} dH_y(t) \]

\[ = \left\{ \int \frac{|t/z|}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \right\} \int \frac{t + |t|}{|1 + tc_2^{-1} w|^2} dH_y(t) \]

\[ \times \left\{ \int \frac{|t/z|}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \right\} \int \frac{t + |t|}{|1 + tc_2^{-1} \tilde{w}|^2} dH_y(t) \]

\[ \vdash (\tilde{B}_1 \tilde{B}_2)^2, \]

\[ |C_1 C_2|^2 \leq \left\{ \int \frac{t^2 c_1^{-1} |wm - 1|^2}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \right\} \int \frac{t^2 c_1^{-1} |\tilde{w} \tilde{m} - 1|^2}{|z + c_1^{-1} \tilde{m}(\tilde{w} \tilde{m} - 1)|^2} dH_x(t) \]

\[ \times \left\{ \int \frac{t^2 c_2^{-1}}{|1 + tc_2^{-1} w|^2} dH_y(t) \right\} \int \frac{t^2 c_2^{-1}}{|1 + tc_2^{-1} \tilde{w}|^2} dH_y(t) \]

\[ = \left\{ \int \frac{t^2 c_1^{-1} |wm - 1|^2}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \right\} \int \frac{t^2 c_2^{-1}}{|1 + tc_2^{-1} w|^2} dH_y(t) \]

\[ \times \left\{ \int \frac{t^2 c_1^{-1} |\tilde{w} \tilde{m} - 1|^2}{|z + c_1^{-1} \tilde{m}(\tilde{w} \tilde{m} - 1)|^2} dH_x(t) \right\} \int \frac{t^2 c_2^{-1}}{|1 + tc_2^{-1} \tilde{w}|^2} dH_y(t) \]

\[ \vdash (\tilde{C}_1 \tilde{C}_2)^2. \]

Then (B.18) implies

\[ 1 = |B_1 B_2 + C_1 C_2| \]

\[ \leq \sqrt{(\tilde{B}_1^2 + \tilde{C}_1^2)(\tilde{B}_2^2 + \tilde{C}_2^2)} \]

\[ = \left\{ \int |t/z| + \frac{|tz|}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \right\} \int \frac{t + |t|}{|1 + tc_2^{-1} w|^2} dH_y(t) \]

\[ + \left\{ \int \frac{t^2 c_1^{-1} |wm - 1|^2}{|z + c_1^{-1} tm(wm - 1)|^2} dH_x(t) \right\} \int \frac{t^2 c_2^{-1}}{|1 + tc_2^{-1} w|^2} dH_y(t) \}^{1/2} \]
Further, if it holds (B.22) will lead to

Such inequality also holds true if we replace \( w \) then for \( t \neq 0 \), combining the above three equations (B.20), (B.21) and (B.22) will lead to

\[
1 > \frac{t}{|z|} \left| \frac{t|z|}{|z + c_1^{-1}tm(\tilde{w}\tilde{m} - 1)|^2} \right|^2 dH_x(t) \int t + \frac{t}{|1 + tc_2^{-1}w|^2} dH_y(t) \\
+ \int \frac{t^2c_1^{-1}|\tilde{w}\tilde{m} - 1|^2}{|z + c_1^{-1}tm(\tilde{w}\tilde{m} - 1)|^2} dH_x(t) \int \frac{t^2c_2^{-1}}{|1 + tc_2^{-1}w|^2} dH_y(t). \tag{B.23}
\]

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

\[
1 > \frac{t}{|z|} \left| \frac{t|z|}{|z + c_1^{-1}tm(\tilde{w}\tilde{m} - 1)|^2} \right|^2 dH_x(t) \int t + \frac{t}{|1 + tc_2^{-1}w|^2} dH_y(t) \\
+ \int \frac{t^2c_1^{-1}|\tilde{w}\tilde{m} - 1|^2}{|z + c_1^{-1}tm(\tilde{w}\tilde{m} - 1)|^2} dH_x(t) \int \frac{t^2c_2^{-1}}{|1 + tc_2^{-1}w|^2} dH_y(t). \tag{B.24}
\]
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

\[
\Im(zm - z) = |z|^2 \Im(s),
\]

\[
\Im(zs) = \Im(wm) = \int t + \frac{t}{|1 + tc_z|w|2} dH_y(t) \Im(w).
\]

Then assertion (B.22) is equivalent to

\[
\Im(s) > \frac{1}{|z|} \Im(zs).
\] (B.25)

Actually, for any subsequence \(\{n_k\}\) such that

\[
s_{n_k}(z) = \frac{1}{n_k} \text{Etr}(B_{n_k} - zI_n)^{-1}
\]

converges, the empirical distribution \(F^{B_{n_k}}\) 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 \(\{x_i\}\) and \(\{y_i\}\) are independent, we first fix the sequence of matrices \((A_n)\) and show that, without the Gaussian assumption, the empirical spectral distribution \(F_{S_{xy}}\) will still converge weakly to the same spectral distribution \(F\) under Assumptions (a)-(c). Next, the same trick can be applied to \(\{x_i\}\), 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
\[ B_n = A_n^{1/2}\left(\frac{1}{q}Y'Y + \gamma_y I\right)A_n^{1/2} = A_n^{1/2}\left(\frac{1}{q}W'\Sigma_y W + \gamma_y I\right)A_n^{1/2}, \]
where the table \( W \) consists i.i.d. standard Gaussian random variables and we vectorize it as a \( qn \)-dimensional random vector, denoted as \( w = (w_{ij}) \). Therefore, the Stieltjes transform \( s_n(z) \) of \( F^{B_n} \) can be viewed as a function of the random vector \( w \), defined as
\[ U := f(w) = \frac{1}{n} \text{tr}(B_n - zI)^{-1}, \]
Similarly, we denote by
\[ V := f(\tilde{w}) \]
the non-Gaussian counterpart of \( U \), where \( \tilde{w} = (\tilde{w}_{ij}) \) have the same first two moments as \( \{w_{ij}\} \) and finite fourth moment. Let \( \tilde{w} = (\tilde{w}_{ij}) \) be a mixture of \( w \) and \( \tilde{w} \) by taking \( \tilde{w}_{ij} \in \{w_{ij}, \tilde{w}_{ij}\} \) for \( i = p + 1, \ldots, p+q \) and \( j = 1, \ldots, n \), whose matrix form is denoted by \( \tilde{W} \). Applying Lemma 3, one gets
\[ |\mathbb{E}(U) - \mathbb{E}(V)| \leq Kqn\lambda_3(f), \] (B.26)
where
\[ \lambda_3(f) = \sup \left\{ \left| \frac{\partial^k f(\tilde{w})}{\partial \tilde{w}_{ij}^k} \right|^{3/k} \right. : p + 1 \leq i \leq p + q, 1 \leq j \leq n, 1 \leq k \leq 3, \tilde{w} \in \mathbb{R}^{qn} \right\}. \]
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 \( \tilde{w}_{ij} \).
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:

$$\mathbb{E}(\bar{w}_{ij}) = 0, \quad \text{Var}(\bar{w}_{ij}) = 1, \quad |\bar{w}_{ij}| \leq \sqrt{n}, \quad e'_i \bar{W} \bar{W}' e_i \leq Kn,$$

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

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

$$\frac{\partial f(w)}{\partial w_{ij}} = \frac{1}{n} \text{tr} G' = -\frac{1}{n} \text{tr} B_n' G^2,$$

$$\frac{\partial^2 f(w)}{\partial w_{ij}^2} = -\frac{1}{n} \text{tr} (B_n'' G^2 + 2B_n' G G') = -\frac{1}{n} \text{tr} B_n'' G^2 + \frac{2}{n} \text{tr} B_n' G^2 B_n' G,$$

$$\frac{\partial^3 f(w)}{\partial w_{ij}^3} = \frac{4}{n} \text{tr} B_n'' G^2 B_n' - \frac{6}{n} \text{tr} B_n' G^2 B_n' GB_n' G + \frac{2}{n} \text{tr} B_n' G^2 B_n'' G,$$

where

$$G' = -GB_n' G,$$

$$B_n' = \frac{1}{q} A_n^{1/2} (e_j e'_i \Sigma_y W + W' \Sigma_y e_i e'_j) A_n^{1/2},$$

$$B_n'' = \frac{2}{q} A_n^{1/2} e_j e'_i \Sigma_y e_i e'_j A_n^{1/2}.$$
B.3 Proof of Theorem 3.1

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

For the first derivative of $f$, since $\Sigma_y$, $A_n^{1/2}$ and $G^2$ are all normal, we have

\[
\sup \left| \frac{\partial f(w)}{\partial w_{ij}} \right| \leq \sup \left\{ \frac{1}{nq} \left| \text{tr} A_n^{1/2} \mathbf{e}_j \Sigma_y W A_n^{1/2} G^2 \right| + \frac{1}{nq} \left| \text{tr} A_n^{1/2} W' \Sigma_y \mathbf{e}_j \mathbf{e}_i' A_n^{1/2} G^2 \right| \right\} \\
\leq \sup \left\{ K \frac{n}{nq} \| \mathbf{e}_j \| \| \mathbf{e}_i' W \| + K \frac{n}{nq} \| W' \mathbf{e}_i \| \| \mathbf{e}_j' \| \right\} \\
\leq Kn^{-3/2}. \tag{B.27}
\]

For the second derivative, we have

\[
\left| \frac{1}{n} \text{tr} B''_n G^2 \right| = \frac{2}{nq} \left| \text{tr} A_n^{1/2} \mathbf{e}_j \Sigma_y \mathbf{e}_i' A_n^{1/2} G^2 \right| \leq \frac{K}{nq} \| \mathbf{e}_j \| \cdot \| \mathbf{e}_i' \mathbf{e}_j' \| \leq Kn^{-2}
\]

and

\[
\left| \frac{2}{n} \text{tr} B'_n G^2 B'_n G \right| \\
= \frac{2}{nq^2} \left| \text{tr} A_n^{1/2} (\mathbf{e}_j \mathbf{e}_i' \Sigma_y W + W' \Sigma_y \mathbf{e}_i' \mathbf{e}_j') A_n^{1/2} G^2 A_n^{1/2} (\mathbf{e}_j \mathbf{e}_i' \Sigma_y W + W' \Sigma_y \mathbf{e}_i' \mathbf{e}_j') A_n^{1/2} G \right| \\
\leq \frac{K}{nq^2} \left( \| \mathbf{e}_j \| \| \mathbf{e}_i' \mathbf{e}_j' W \| + \| \mathbf{e}_j \| \| \mathbf{e}_i' W W' \mathbf{e}_i' \mathbf{e}_j' \| + \| W' \mathbf{e}_i \| \| \mathbf{e}_j' \mathbf{e}_i' W \| + \| W' \mathbf{e}_i \| \| \mathbf{e}_j' W \mathbf{e}_i' \mathbf{e}_j' \| \right) \\
\leq \frac{K}{nq^2} \left( n + \sqrt{n} \cdot |w_{ij}| \right) \\
\leq Kn^{-2},
\]

which leads to the conclusion that

\[
\sup \left| \frac{\partial^2 f(w)}{\partial w_i^2} \right| \leq Kn^{-2}. \tag{B.28}
\]
Similarly, we could bound the third derivative as follows,

\[
\sup \left| \frac{\partial^3 f(w)}{\partial w^3_{ij}} \right| \leq \sup \left\{ \frac{K}{nq^3} \left( \|e'_iW\| \|w_{ij}\|^2 + 2|w_{ij}|\|e'_iWW'\|e_i\| + \|e'_iW\|\|e'_iWW'\|e_i\| \right) + \frac{K}{nq^2} \left( \|e'_iW\| + |w_{ij}| \right) \right\}
\]

\[
\leq Kn^{-5/2}. \tag{B.29}
\]

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

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

which together with (B.26) imply

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

The proof is done.

### B.4 Proof of Theorem 4.1

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

\[
Z = \Gamma XS + Y,
\]

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

\[
\frac{1}{q}Z'Z = \frac{1}{q}Y'Y + \frac{1}{q}SXT'TXS + \frac{1}{q}SXT'Y + \frac{1}{q}Y'TXS
\]

\[
\triangleq \frac{1}{q}Y'Y + H,
\]

21
where

\[ H = \frac{1}{q} X^T \Gamma X + \frac{1}{q} X^T \Gamma Y + \frac{1}{q} Y^T \Gamma X \]  \hspace{1cm} (B.30)

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

\[ \tilde{S}_{xz} = A_n^{1/2} \left( \frac{1}{q} Z^T Z + \gamma_z I_n \right) A_n^{1/2} \quad \text{and} \quad \hat{S}_{xz} = A_n^{1/2} \left( \frac{1}{q} Y^T Y + \gamma_z I_n \right) A_n^{1/2}, \]

where

\[ \gamma_z = \frac{1}{q} \text{tr}(\Sigma_z) = \gamma_y + \frac{1}{q} \sum_{i=1}^{m} \theta_i^2 \cdot \gamma_x = \gamma_y + o(1). \]  \hspace{1cm} (B.31)

Applying Lemma 2 to \(B_n, \tilde{S}_{xz}\) and \(\hat{S}_{xz}\), we have

\[ || F^{\tilde{S}_{xz}} - F^{\hat{S}_{xz}} || \to 0 \quad \text{and} \quad L^3(F^{B_n}, F^{\hat{S}_{xz}}) \to 0, \]  \hspace{1cm} (B.32)

almost surely, as \((n, p, q)\) tend to infinity. Combining (B.32) and the fact that \(\tilde{S}_{xz}\) shares the same LSD as \(S_{xz}\), we conclude that \(F^{S_{xz}}\) converges weakly to the LSD \(F\) defined by (3.1). The proof is thus complete.

### B.5 Proof of Theorem 4.2

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

\[ \tilde{S}_{xz} := A_n^{1/2} \left( \frac{1}{q} Y^T Y + H + \gamma_y I_n \right) A_n^{1/2}, \]

where \(H\) is given in (B.30). So it’s equivalent to prove the theorem for \(\tilde{S}_{xz}\).
Next, from Bai and Silverstein [1998] and the inequality

\[ \| A_n^{1/2} C_n A_n^{1/2} \| \leq \| A_n \| \cdot \| C_n \|, \]

we know that the spectral norm \( \| A_n^{1/2} C_n A_n^{1/2} \| \) is bounded in \( n \), almost surely. Define

\[ \lambda_+ = \limsup_{n \to \infty} \| A_n^{1/2} C_n A_n^{1/2} \|, \]

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

\[ \| \lambda I_n - \bar{S}_{xz} \| = 0 \quad \text{and} \quad \| \lambda I_n - A_n^{1/2} C_n A_n^{1/2} \| \neq 0, \quad \text{(B.33)} \]

for \( \lambda \in \{\lambda_{n,1}, \ldots, \lambda_{n,k}\} \).

In the following, we will show the limits of \( \lambda \) is defined in (4.4). Under the assumptions in \( \text{(B.33)} \), we have

\[ \left| I_n - \left( \lambda I_n - A_n^{1/2} C_n A_n^{1/2} \right)^{-1} A_n^{1/2} HA_n^{1/2} \right| = 0. \quad \text{(B.34)} \]

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

\[
H = \frac{1}{q} \begin{pmatrix}
    a_1 & b_1 & \cdots & a_m & b_m \\
\end{pmatrix}
\begin{pmatrix}
    \theta_1 \lambda_{11} & 0 & \cdots & 0 & 0 \\
    0 & \theta_1 \lambda_{12} & \cdots & 0 & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    0 & 0 & \cdots & \theta_m \lambda_{m1} & 0 \\
    0 & 0 & \cdots & 0 & \theta_m \lambda_{m2} \\
\end{pmatrix}
\begin{pmatrix}
    a'_1 \\
    b'_1 \\
    \vdots \\
    a'_m \\
    b'_m \\
\end{pmatrix}
\]

(B.35)

where

\[
a_i = u_{i1}SX'v_i + w_{i1}Y'u_i,
\]

\[
b_i = u_{i2}SX'v_i + w_{i2}Y'u_i,
\]

\[
\lambda_{i1} = \frac{\|SX'v_i\|\|Y'u_i\|}{\sqrt{4\|Y'u_i\|^2 + \theta_i^2\|SX'v_i\|^2 + \theta_i\|SX'v_i\|^2}} + \frac{\theta_i\|SX'v_i\|^2 - \theta_i\|SX'v_i\|^2}{\sqrt{4\|Y'u_i\|^2 + \theta_i^2\|SX'v_i\|^2 + \theta_i\|SX'v_i\|^2}}
\]

\[
\lambda_{i2} = -\frac{\|SX'v_i\|\|Y'u_i\|}{\sqrt{4\|Y'u_i\|^2 + \theta_i^2\|SX'v_i\|^2 + \theta_i\|SX'v_i\|^2}} + \frac{\theta_i\|SX'v_i\|^2 + \theta_i\|SX'v_i\|^2}{\sqrt{4\|Y'u_i\|^2 + \theta_i^2\|SX'v_i\|^2 + \theta_i\|SX'v_i\|^2}}
\]

with

\[
u_{i1} = \frac{1}{\|SX'v_i\|} \left\{ \frac{1}{2} + \frac{\theta_i\|SX'v_i\|^2}{2\sqrt{4\|Y'u_i\|^2 + \theta_i^2\|SX'v_i\|^2}} \right\}^{1/2}
\]

\[
u_{i2} = \frac{1}{\|SX'v_i\|} \left\{ \frac{1}{2} - \frac{\theta_i\|SX'v_i\|^2}{2\sqrt{4\|Y'u_i\|^2 + \theta_i^2\|SX'v_i\|^2}} \right\}^{1/2}
\]

\[
w_{i1} = \frac{1}{\|Y'u_i\|} \left\{ \frac{1}{2} - \frac{\theta_i\|SX'v_i\|^2}{2\sqrt{4\|Y'u_i\|^2 + \theta_i^2\|SX'v_i\|^2}} \right\}^{1/2}
\]

\[
w_{i2} = -\frac{1}{\|Y'u_i\|} \left\{ \frac{1}{2} + \frac{\theta_i\|SX'v_i\|^2}{2\sqrt{4\|Y'u_i\|^2 + \theta_i^2\|SX'v_i\|^2}} \right\}^{1/2}
\]
B.5 Proof of Theorem 4.2

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

\[
\begin{align*}
\lambda_i u_{1i}^2 + \lambda_i u_{2i}^2 &= \theta_i, \\
\lambda_i w_{1i}^2 + \lambda_i w_{2i}^2 &= 0, \\
\lambda_i u_{1i} w_{1i} + \lambda_i u_{2i} w_{2i} &= 1.
\end{align*}
\]

Denote \( D_n = A_n^{1/2} \left( \lambda I_n - A_n^{1/2} C_n A_n^{1/2} \right)^{-1} A_n^{1/2} \) and

\[
M_n = \frac{1}{q} \begin{pmatrix}
a'_1 \\ b'_1 \\ \vdots \\ a'_m \\ b'_m
\end{pmatrix}
D_n \left( \begin{pmatrix} a_1 & b_1 & \cdots & a_m & b_m \end{pmatrix} \right)
\begin{pmatrix}
\theta_1 \lambda_{11} & 0 & \cdots & 0 & 0 \\
0 & \theta_1 \lambda_{12} & \cdots & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & \cdots & \theta_m \lambda_{m1} & 0 \\
0 & 0 & \cdots & 0 & \theta_m \lambda_{m2}
\end{pmatrix}.
\]

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

\[
f_n(\lambda) := |I_{2m} - M_n| = 0.
\]

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

\[
\alpha_n = \frac{1}{n} \text{tr} SD_n S (A_n - \gamma_x I_n) \quad \text{and} \quad \beta_n = \frac{1}{n} \text{tr} D_n (C_n - \gamma_y I_n),
\]

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

\[
\begin{align*}
\frac{a'^i D_n a_i}{q} &= \frac{u_{i1}^2 c_n}{\alpha_n} + \frac{u_{i2}^2 c_n}{\beta_n} + o_{a.s.}(1), \\
\frac{a'^i D_n b_i}{q} &= \frac{u_{i1} u_{i2} c_n}{\alpha_n} + \frac{w_{i1} w_{i2} c_n}{\beta_n} + o_{a.s.}(1), \\
\frac{b'^i D_n b_i}{q} &= \frac{u_{i2}^2 c_n}{\alpha_n} + \frac{w_{i2}^2 c_n}{\beta_n} + o_{a.s.}(1),
\end{align*}
\]

25
and for any $i \neq j \in \{1, \ldots, m\}$,
\[
\frac{a'_i D_n a_j}{q} = o_{a.s.}(1), \quad \frac{a'_i D_n b_j}{q} = o_{a.s.}(1).
\]

From the above approximations and the identities in (B.36), we have
\[
f_n(\lambda) = \prod_{k=1}^{m} |I_2 - M_{nk}| + o_{a.s}(1)
\]

where
\[
M_{nk} = \frac{\theta_k}{c_n} \begin{pmatrix} \alpha_n & 0 \\ 0 & \beta_n \end{pmatrix} \begin{pmatrix} \theta_k & 1 \\ 1 & 0 \end{pmatrix}.
\]

Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$, then
\[
\alpha_n = \frac{1}{n} \text{tr} S D_n S A_n - \frac{\gamma_x}{n} \text{tr} S D_n S = \frac{1}{n} \varepsilon' (D_n \circ A_n) \varepsilon - \frac{\gamma_x}{n} \varepsilon' \text{Diag}(D_n) \varepsilon,
\]

where “$\circ$” denotes the Hadamard product of two matrices. According to Theorem 1 of Varberg (1968), we have
\[
\frac{1}{n} \varepsilon' (D_n \circ A_n) \varepsilon - \frac{1}{n} \mathbb{E} [\varepsilon' (D_n \circ A_n) \varepsilon] \overset{a.s.}{\longrightarrow} 0, \quad \text{(B.39)}
\]
\[
\frac{1}{n} \varepsilon' \text{Diag}(D_n) \varepsilon - \frac{1}{n} \mathbb{E} \text{tr} D_n \overset{a.s.}{\longrightarrow} 0. \quad \text{(B.40)}
\]

Further,
\[
\frac{1}{n} \mathbb{E} [\varepsilon' (D_n \circ A_n) \varepsilon] = \frac{1}{n} \mathbb{E} \text{tr} [D_n \text{Diag}(A_n)]
\]
\[
= \frac{1}{n} \mathbb{E} \text{tr} [D_n (\text{Diag}(A_n) - 2\gamma_x I_n)] + \frac{2\gamma_x}{n} \mathbb{E} \text{tr} D_n.
\]
= \frac{2\gamma_x}{n} \text{tr} D_n + o(1), \quad (B.41)

where the last equality is due to the following convergence,

\[ \left| \frac{1}{n} \text{tr} \left[ D_n \cdot \left( \text{Diag}(A_n) - 2\gamma_x I_n \right) \right] \right| \leq \frac{1}{n} \| D_n \| \cdot \text{tr} \left| A_n - 2\gamma_x I_n \right| \xrightarrow{a.s.} 0. \]

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

\[ \alpha_n = -\gamma_x w_n(\lambda) + o_a.s.(1) \xrightarrow{a.s.} \alpha \triangleq -w(\lambda) \int tdH_x(t), \quad (B.42) \]

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

\[ \beta_n = \frac{1}{n} \text{tr}(D_n C_n) - \frac{\gamma_y}{n} \text{tr} D_n \]

\[ = -1 + \frac{\lambda}{n} \text{tr} \left( \lambda I_n - A_n^{1/2} C_n A_n^{1/2} \right)^{-1} - \frac{\gamma_y}{n} \text{tr} D_n \]

\[ = -\frac{1}{n} \sum_{k=1}^{q} \frac{\tau_k w_n(\lambda)}{c_{n2} + \tau_k w_n(\lambda)} + o_a.s.(1) \]

\[ \xrightarrow{a.s.} \beta \triangleq -c_2 \int \frac{tw(\lambda)dH_y(t)}{c_2 + tw(\lambda)}, \quad (B.43) \]

where the third equality is from (B.11) with \( (\tau_k) \) being the eigenvalues of \( \Sigma_y \). 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 \( \theta_0 \) 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 \((\lambda_+, \infty)\). 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


