Sparse Quadratic Discriminant Analysis For High Dimensional Data

Quefeng Li and Jun Shao

University of Wisconsin-Madison and East China Normal University

Supplementary Material

S1 Proofs

Proof of Lemma 1. The first result follows from

$$\begin{aligned}
\|\boldsymbol{LCR}\|_G &= \sum_i \sum_j \left| \sum_k \sum_h l_{ik} c_{kh} r_{hj} \right| \\
&\le \sum_i \sum_j \sum_k \sum_h |l_{ik} c_{kh} r_{hj}| \\
&= \sum_i \sum_k \sum_h \left(|l_{ik} c_{kh}| \cdot \sum_j |r_{hj}| \right) \\
&\le \sum_i \sum_k \sum_h |l_{ik} c_{kh}| \left(\max_h \sum_j |r_{hj}| \right) \\
&= \|\boldsymbol{R}\|_1 \sum_h \sum_k |c_{kh}| \cdot \sum_i |l_{ik}| \\
&\le \|\boldsymbol{R}\|_1 \sum_h \left(\sum_k |c_{kh}| \right) \left(\max_k \sum_i |l_{ik}| \right) \\
&= \|\boldsymbol{R}\|_1 \|\boldsymbol{L}\|_1 \|\boldsymbol{C}\|_G,
\end{aligned}$$

where the last equality follows from the fact that $\boldsymbol{L}$ is symmetric. Represent $\boldsymbol{L}$ and $\boldsymbol{R}$ as their spectral decompositions, $\boldsymbol{L} = \boldsymbol{P}'\boldsymbol{D_L P}$ and $\boldsymbol{R} = \boldsymbol{Q}'\boldsymbol{D_R Q}$. Then, the second result follows from

$$\begin{aligned}
\|\boldsymbol{LCR}\|_F^2 &= \mathrm{tr}(\boldsymbol{RCL}^2\boldsymbol{CR}) = \mathrm{tr}(\boldsymbol{RCP}'\boldsymbol{D}_{\boldsymbol{L}}^2\boldsymbol{PCR}) \\
&= \mathrm{tr}(\boldsymbol{D}_{\boldsymbol{L}}^2\boldsymbol{PCR}^2\boldsymbol{CP}') \le \|\boldsymbol{L}\|_2^2 \mathrm{tr}(\boldsymbol{PCR}^2\boldsymbol{CP}') \\
&= \|\boldsymbol{L}\|_2^2 \mathrm{tr}(\boldsymbol{R}^2\boldsymbol{C}^2) = \|\boldsymbol{L}\|_2^2 \mathrm{tr}(\boldsymbol{Q}'\boldsymbol{D}_{\boldsymbol{R}}^2\boldsymbol{QC}^2)
\end{aligned}$$

$$= \|\boldsymbol{L}\|_2^2 \mathrm{tr}(\boldsymbol{D}_{\boldsymbol{R}}^2 \boldsymbol{Q}\boldsymbol{C}^2\boldsymbol{Q}') \le \|\boldsymbol{L}\|_2^2 \|\boldsymbol{R}\|_2^2 \mathrm{tr}(\boldsymbol{Q}\boldsymbol{C}^2\boldsymbol{Q}')$$
$$= \|\boldsymbol{L}\|_2^2 \|\boldsymbol{R}\|_2^2 \mathrm{tr}(\boldsymbol{C}^2) = \|\boldsymbol{L}\|_2^2 \|\boldsymbol{R}\|_2^2 \|\boldsymbol{C}\|_F^2.$$

Proof of Lemma 2. Write $\boldsymbol{\Lambda}$ as its spectral decomposition $\boldsymbol{\Lambda} = \boldsymbol{U}'\boldsymbol{D}_{\boldsymbol{\Lambda}}\boldsymbol{U}$, where $\boldsymbol{D}_{\boldsymbol{\Lambda}}$ is a diagonal matrix whose jth diagonal element is $\lambda_{p,j} = \lambda_{p,j}(\boldsymbol{\Lambda})$. Then,

$$T_p = \sum_{j=1}^{p} \lambda_{p,j}\tilde{z}_{p,j}^2 - 2a_{p,j}\tilde{z}_{p,j}, \tag{S1.1}$$

where $\tilde{z}_{p,j}$ is the jth component of $\boldsymbol{U}\boldsymbol{z}$ and $a_{p,j}$ is the jth component of $\boldsymbol{U}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}$. Let $\zeta_{p,j} = \lambda_{p,j}\tilde{z}_{p,j}^2 - 2a_{p,j}\tilde{z}_{p,j} - \lambda_{p,j}$. Then $\mathrm{E}\zeta_{p,j} = 0$, $\sigma_{p,j}^2 = \mathrm{E}\zeta_{p,j}^2 = 2\lambda_{p,j}^2 + 4a_{p,j}^2$, and $T_p - \mathrm{E}(T_p) = \sum_{j=1}^{p} \zeta_{p,j}$. In the following, we show that $\{\zeta_{p,j},\ j = 1, 2, ..., p\}$ satisfy condition (B_γ) on page 43 of Saulis and Statulevicius (1991) for $k \ge 3$. Actually, there exist constants $C_1 > 0$ and $M_0 > 0$ such that

$$\begin{aligned}
|\mathrm{E}\zeta_{p,j}^k| &= |\mathrm{E}(\lambda_{p,j}\tilde{z}_{p,j}^2 - 2a_{p,j}\tilde{z}_{p,j} - \lambda_{p,j})^k| \\
&\le \mathrm{E}|\lambda_{p,j}\tilde{z}_{p,j}^2 - 2a_{p,j}\tilde{z}_{p,j} - \lambda_{p,j}|^k \\
&\le \mathrm{E}\left(|\lambda_{p,j}\tilde{z}_{p,j}^2| + |2a_{p,j}\tilde{z}_{p,j}| + |\lambda_{p,j}|\right)^k \\
&\le 3^{k-1}\mathrm{E}\left(|\lambda_{p,j}|^k|\tilde{z}_{p,j}|^{2k} + |2a_{p,j}|^k|\tilde{z}_{p,j}|^k + |\lambda_{p,j}|^k\right) \\
&\le 3^{k-1}\left(|\lambda_{p,j}|^k(2k-1)!! + |2a_{p,j}|^k(k-1)!! + |\lambda_{p,j}|^k\right) \\
&\le 3^{k-1}C_1 k! 2^k(|\lambda_{p,j}|^k + |a_{p,j}|^k) \\
&= (1/3)C_1 6^k k!(|\lambda_{p,j}|^k + |a_{p,j}|^k) \\
&= (1/3)C_1 6^k k!\sigma_{p,j}^2 \frac{|\lambda_{p,j}|^k + |a_{p,j}|^k}{2\lambda_{p,j}^2 + 4a_{p,j}^2} \\
&\le (1/3)C_1 6^k k!\sigma_{p,j}^2 \cdot 1/2[\max\{|\lambda_{p,j}|, |a_{p,j}|\}]^{k-2} \\
&\le C_1 6^{k-1} M_0^{k-2} k!\sigma_{p,j}^2 \\
&\le (6M_0 \cdot \max\{6C_1, 1\})^{k-2} k!\sigma_{p,j}^2.
\end{aligned}$$

Therefore, $\{\zeta_{p,j}, j = 1, 2, ..., p\}$ satisfy condition (B_γ) with $\gamma = 0$ and $K = 6M_0 \max\{6C_1, 1\}$.

Then, Theorem 3.1 of Saulis and Statulevicius (1991) implies that

$$\sup_x \left|F_{Z_p}(x) - \Phi(x)\right| \le \frac{324\sqrt{2}K_p}{B_p},$$

where $K_p = 2\max\{K, \sqrt{6}M_0\}$, $B_p^2 = \sum_{j=1}^{p} 2\lambda_{p,j}^2 + 4a_{p,j}^2 = 2\|\boldsymbol{\Lambda}\|_F^2 + 4\boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}$, and F_{Z_p} denotes the distribution function of $Z_p = [T_p - \mathrm{E}(T_p)]/\sqrt{\mathrm{Var}(T_p)}$. In other words,

$$\sup_x \left|F_{Z_p}(x) - \Phi(x)\right| \lesssim \frac{1}{[2\|\boldsymbol{\Lambda}\|_F^2 + 4\boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}]^{1/2}} \lesssim \frac{1}{D_p},$$

which converges to 0 as $p \to \infty$, because under (C2), it holds that $\|\boldsymbol{\Lambda}\|_F \asymp \|\boldsymbol{\Delta}\|_F$ and $\boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta} \asymp \boldsymbol{\delta}'\boldsymbol{\delta}$.

Proof of Lemma 3. Under (C2), $\hat{\boldsymbol{\Sigma}}_k$ is asymptoticly invertible by (14). Hence,

$$\hat{\boldsymbol{\Sigma}}_k^{-1} = \boldsymbol{\Sigma}_k^{-1} + \hat{\boldsymbol{\Sigma}}_k^{-1} - \boldsymbol{\Sigma}_k^{-1}, \quad \hat{\boldsymbol{\Sigma}}_k^{-1} - \boldsymbol{\Sigma}_k^{-1} = \hat{\boldsymbol{\Sigma}}_k^{-1}(\boldsymbol{\Sigma}_k - \hat{\boldsymbol{\Sigma}}_k)\boldsymbol{\Sigma}_k^{-1}.$$

Then,

$$\|\hat{\boldsymbol{\Sigma}}_k^{-1}\|_1 \leq \|\boldsymbol{\Sigma}_k^{-1}\|_1 + \|\hat{\boldsymbol{\Sigma}}_k^{-1} - \boldsymbol{\Sigma}_k^{-1}\|_1 \leq (1 + \|\hat{\boldsymbol{\Sigma}}_k^{-1}\|_1 \|\boldsymbol{\Sigma}_k - \hat{\boldsymbol{\Sigma}}_k\|_1)\boldsymbol{\Sigma}_k^{-1}.$$

Since $\|\hat{\boldsymbol{\Sigma}}_k - \boldsymbol{\Sigma}_k\|_1 = O_P(a_n)$, $\|\boldsymbol{\Sigma}_k^{-1}\|_1 = O_P(v_p)$, and $a_n v_p \to 0$, it holds that

$$1/2\|\hat{\boldsymbol{\Sigma}}_k^{-1}\|_1 \leq (1 + \|\boldsymbol{\Sigma}_k - \hat{\boldsymbol{\Sigma}}_k\|_1 \|\boldsymbol{\Sigma}_k^{-1}\|_1)^{-1} \|\hat{\boldsymbol{\Sigma}}_k^{-1}\|_1 \leq \|\boldsymbol{\Sigma}_k^{-1}\|_1.$$

Hence, $\|\hat{\boldsymbol{\Sigma}}_k^{-1}\|_1 \leq 2\|\boldsymbol{\Sigma}_k^{-1}\|_1$. Then,

$$\|\hat{\boldsymbol{\Sigma}}_k^{-1} - \boldsymbol{\Sigma}_k^{-1}\|_1 = \|\hat{\boldsymbol{\Sigma}}_k^{-1}(\boldsymbol{\Sigma}_k - \hat{\boldsymbol{\Sigma}}_k)\boldsymbol{\Sigma}_k^{-1}\|_1 = O_P\left(a_n v_p^2\right).$$

Proof of Lemma 4. By Chebyshev's inequality,

$$\begin{aligned}
|(\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{x} - \boldsymbol{\mu}_1)| &= O_P\left(\sqrt{\mathrm{Var}\left((\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{x} - \boldsymbol{\mu}_1) \mid \boldsymbol{X}\right)}\right) \\
&= O_P\left(\sqrt{(\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta})'\boldsymbol{\Sigma}_1(\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta})}\right) \\
&= O_P\left(\|\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}\|\right) \\
&= O_P\left(\|(\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1})\hat{\boldsymbol{\delta}}\| + \|\boldsymbol{\Sigma}_2^{-1}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\|\right) \\
&= O_P\left(\|\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1}\|_2\|\hat{\boldsymbol{\delta}}\| + \|\boldsymbol{\Sigma}_2^{-1}\|_2\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|\right) \\
&= O_P(a_n) + O_P(\sqrt{b_n}),
\end{aligned}$$

where the third equality follows from (C2) and the last equality follows from (C2), $\|\hat{\boldsymbol{\delta}}\|^2 = O_P(1)$, and results (12) and (14). It follows from a result in the proof of Theorem 3 in Shao et al. (2011) that

$$|\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)| = O_P\left(\max\{\sqrt{b_n}, \sqrt{c_p q_n / n}\}\right).$$

Hence, the result follows from

$$\begin{aligned}
|\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1) - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1)| \leq |(\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{x} - \boldsymbol{\mu}_1)| \\
+ |\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)|.
\end{aligned}$$

Proof of Lemma 5. The result follows from

$$|\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}| \leq |\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}\hat{\boldsymbol{\delta}}| + |\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}|,$$

$$|\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}\hat{\boldsymbol{\delta}}| \leq \|\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1}\|_2\|\hat{\boldsymbol{\delta}}\|^2 = O_P\left(a_n\right)$$

and

$$\begin{aligned}|\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}| &\leq |\boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})| + |\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})| \\ &= O_P\left(\|\boldsymbol{\delta}\|\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|\right) + O_P\left(\|\hat{\boldsymbol{\delta}}\|\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|\right) \\ &= O_P(\sqrt{b_n}).\end{aligned}$$

Proof of Lemma 6. Consider

$$|(\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1)'\hat{\boldsymbol{\nabla}}(\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1) - (\boldsymbol{x} - \boldsymbol{\mu}_1)'\boldsymbol{\nabla}(\boldsymbol{x} - \boldsymbol{\mu}_1)| \leq I + II + III,$$

where $I = |(\boldsymbol{x} - \boldsymbol{\mu}_1)'(\hat{\boldsymbol{\nabla}} - \boldsymbol{\nabla})(\boldsymbol{x} - \boldsymbol{\mu}_1)|$, $II = 2|(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)'\hat{\boldsymbol{\nabla}}(\boldsymbol{x} - \boldsymbol{\mu}_1)|$, and $III = |(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)'\hat{\boldsymbol{\nabla}}(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)|$. Let y_j be the jth component of $\boldsymbol{x} - \boldsymbol{\mu}_1$ and d_{ij} be the (i,j)th element of $\hat{\boldsymbol{\nabla}} - \boldsymbol{\nabla}$. Then

$$I \leq \sum_{i=1}^{p}\sum_{j=1}^{p}|d_{ij}y_iy_j| = O_P\left(\sum_{i=1}^{p}\sum_{j=1}^{p}|d_{ij}|\right) = O_P\left(\|\hat{\boldsymbol{\nabla}} - \boldsymbol{\nabla}\|_G\right),$$

where the first equality follows from

$$\mathrm{E}\left(|y_iy_j|\big|\boldsymbol{X}\right) \leq \sqrt{\mathrm{E}(y_i^2\big|\boldsymbol{X})\mathrm{E}(y_j^2\big|\boldsymbol{X})} = \sqrt{\sigma_{1ii}}\sqrt{\sigma_{1jj}} \leq M,$$

where the last inequality follows from (C2) and Shur-Horn Theorem. It then follows from Theorem 2(ii) that $I = O_P\left(\tau_n\right)$.

Let $\boldsymbol{A} = \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Delta}\boldsymbol{\Sigma}_2^{-1}$ and $\hat{\boldsymbol{A}} = \hat{\boldsymbol{\Sigma}}_1^{-1}\hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\Sigma}}_2^{-1}$. Then, with $\hat{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{x}}_1$,

$$III = |(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)'\hat{\boldsymbol{A}}(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)| \leq |(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)'\boldsymbol{A}(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)| + |(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)'(\hat{\boldsymbol{A}} - \boldsymbol{A})(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)|$$

Let α_{ij} be the (i,j)th element of $\boldsymbol{A}$ and ϵ_i be the ith element of $\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1$. Then,

$$\mathrm{E}|(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)'\boldsymbol{A}(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)| = \mathrm{E}|\sum_{i,j=1}^{p}\alpha_{ij}\epsilon_i\epsilon_j| \leq \sum_{i,j=1}^{p}|\alpha_{ij}|\mathrm{E}|\epsilon_i\epsilon_j|$$

$$= O(\|\boldsymbol{A}\|_G/n) = O(\|\boldsymbol{\Sigma}_1^{-1}\|_1\|\boldsymbol{\Delta}\|_G\|\boldsymbol{\Sigma}_2^{-1}\|_1/n).$$

Since $\max_{k,i,j}|\sigma_{kij}| \leq M$, it holds that

$$\|\boldsymbol{\Delta}\|_G = \sum_{i,j=1}^{p}|\Delta_{ij}| \leq \max_{i,j}|\Delta_{ij}|^{1-\eta}\sum_{i,j=1}^{p}|\Delta_{ij}|^{\eta} = 2Mc_{1p}.$$

Hence, $|(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)'\boldsymbol{A}(\boldsymbol{\mu}_1 - \bar{\boldsymbol{x}}_1)| = O_p(c_{1p}v_p^2/n)$. Note that,

$$\begin{aligned}\hat{\boldsymbol{A}} - \boldsymbol{A} &= \hat{\boldsymbol{\Sigma}}_1^{-1}\hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Delta}\boldsymbol{\Sigma}_2^{-1} \\ &= (\hat{\boldsymbol{\Sigma}}_1^{-1} - \boldsymbol{\Sigma}_1^{-1})\hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\Sigma}}_2^{-1} + \boldsymbol{\Sigma}_1^{-1}(\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})\hat{\boldsymbol{\Sigma}}_2^{-1} + \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Delta}(\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1})\end{aligned}$$

By Lemma 1,

$$
\begin{aligned}
\|\hat{\boldsymbol{A}}-\boldsymbol{A}\|_G \le & \|\hat{\boldsymbol{\Sigma}}_1^{-1}-\boldsymbol{\Sigma}_1^{-1}\|_1\|\hat{\boldsymbol{\Delta}}\|_G\|\hat{\boldsymbol{\Sigma}}_2^{-1}\|_1+\|\boldsymbol{\Sigma}_1^{-1}\|_1\|\hat{\boldsymbol{\Delta}}-\boldsymbol{\Delta}\|_G\|\hat{\boldsymbol{\Sigma}}_2^{-1}\|_1 \\
& +\|\boldsymbol{\Sigma}_1^{-1}\|_1\|\boldsymbol{\Delta}\|_G\|\hat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}_2^{-1}\|_1 \\
& =O_P\left(v_p^3 c_{1p} a_n\right)+O_P\left(v_p^2 a_{1n}\right)+O_P\left(v_p c_{1p} a_n\right) \\
& =O_P\left(\tau_n\right).
\end{aligned} \tag{S1.2}
$$

Let $\hat{\theta}=(\boldsymbol{\mu}_1-\bar{\boldsymbol{x}}_1)'(\hat{\boldsymbol{A}}-\boldsymbol{A})(\boldsymbol{\mu}_1-\bar{\boldsymbol{x}}_1)=\sum_{i,j=1}^p \hat{\alpha}_{ij}\epsilon_i\epsilon_j$, where $\hat{\alpha}_{ij}$ is the (i,j)th element of $\hat{\boldsymbol{A}}-\boldsymbol{A}$. Note that,

$$
\begin{aligned}
\mathrm{E}\left[|\hat{\theta}|\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right] &= \mathrm{E}\left[|\sum_{i,j=1}^p \hat{\alpha}_{ij}\epsilon_i\epsilon_j|\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right] \le \sum_{i,j=1}^p |\hat{\alpha}_{ij}|\mathrm{E}\left[|\epsilon_i\epsilon_j|\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right] \\
&\le \sum_{i,j=1}^p |\hat{\alpha}_{ij}|(\mathrm{E}\epsilon_i^2)^{1/2}(\mathrm{E}\epsilon_j^2)^{1/2}=O(n^{-1}\sum_{i,j}|\hat{\alpha}_{ij}|),
\end{aligned}
$$

where the second inequality follows from Cauchy inequality and the independence of $\bar{\boldsymbol{x}}_1$ and $(\boldsymbol{S}_1,\boldsymbol{S}_2)$. Hence,

$$
\hat{\theta}\big|(\boldsymbol{S}_1,\boldsymbol{S}_2)=O_P\left(n^{-1}\sum_{i,j=1}^p|\hat{\alpha}_{ij}|\right). \tag{S1.3}
$$

It follows from (S1.2) and (S1.3) that, for any positive number ε_1 and ε_2, there exists $C_1>0$ and $C_2>0$, such that

$$
P\left(|\hat{\theta}|>C_1/n\sum_{i,j=1}^p|\hat{\alpha}_{ij}|\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right)<\varepsilon_1 \quad \text{and} \quad P\left(\sum_{i,j=1}^p|\hat{\alpha}_{ij}|>C_2\tau_n\right)<\varepsilon_2.
$$

Then,

$$
\begin{aligned}
P(|\hat{\theta}|>C_1C_2\tau_n/n) &= \mathrm{E}\left[P\left(|\hat{\theta}|>C_1C_2\tau_n/n\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right)\right] \\
&= \mathrm{E}\left[P\left(|\hat{\theta}|>C_1C_2\tau_n/n, |\hat{\theta}|>C_1/n\sum_{i,j=1}^p|\hat{\alpha}_{ij}|\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right)\right] \\
&+ \mathrm{E}\left[P\left(|\hat{\theta}|>C_1C_2\tau_n/n, |\hat{\theta}|\le C_1/n\sum_{i,j=1}^p|\hat{\alpha}_{ij}|\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right)\right] \\
&\le \mathrm{E}\left[P\left(|\hat{\theta}|>C_1/n\sum_{i,j=1}^p|\hat{\alpha}_{ij}|\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right)\right] \\
&+ \mathrm{E}\left[P\left(\sum_{i,j=1}^p|\hat{\alpha}_{ij}|>C_2\tau_n\right)\Big|(\boldsymbol{S}_1,\boldsymbol{S}_2)\right] \le \varepsilon_1+\varepsilon_2.
\end{aligned}
$$

Hence, $|(\boldsymbol{\mu}_1-\bar{\boldsymbol{x}}_1)'(\hat{\boldsymbol{A}}-\boldsymbol{A})(\boldsymbol{\mu}_1-\bar{\boldsymbol{x}}_1)| = O_P(\tau_n/n)$. Therefore, $III = O_P\left(c_{1p}v_p^2/n\right)$. Using a similar argument, we can also show that $II = O_P(c_{1p}v_p^2/\sqrt{n})$. Since $c_{1p}v_p^2/\sqrt{n} = O(\tau_n)$, the result follows.

Proof of Lemma 7. From the property of the trace operation, we obtain that

$$\begin{aligned}
\left|\operatorname{tr}(\hat{\boldsymbol{\Lambda}}) - \operatorname{tr}(\boldsymbol{\Lambda})\right| &= \left|\operatorname{tr}(\hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\Sigma}}_2^{-1}) - \operatorname{tr}(\boldsymbol{\Delta}\boldsymbol{\Sigma}_2^{-1})\right| \\
&\le \left|\operatorname{tr}((\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})\boldsymbol{\Sigma}_2^{-1})\right| + \left|\operatorname{tr}(\hat{\boldsymbol{\Delta}}(\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1}))\right| \\
&\le \|(\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta})\boldsymbol{\Sigma}_2^{-1}\|_G + \|\hat{\boldsymbol{\Delta}}(\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1})\|_G \\
&\le \|\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\|_G\|\boldsymbol{\Sigma}_2^{-1}\|_1 + \|\hat{\boldsymbol{\Delta}}\|_G\|\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1}\|_1 \\
&= O_P\left(a_{1n}v_p\right) + O_P\left(c_{1p}a_nv_p^2\right) \\
&= O_P\left(\tau_n\right),
\end{aligned}$$

where the third inequality follows from Lemma 1 and the second equality follows from Lemma 3 and Theorem 2. This proves the first result.

Note that $\log(|\boldsymbol{\Sigma}_1|/|\boldsymbol{\Sigma}_2|) = \log|\boldsymbol{I}+\boldsymbol{\Lambda}|$ and $\log(|\hat{\boldsymbol{\Sigma}}_1|/|\hat{\boldsymbol{\Sigma}}_2|) = \log|\boldsymbol{I}+\hat{\boldsymbol{\Lambda}}|$. We employ a Taylor expansion of $f(t) = \log|\boldsymbol{I} + t\boldsymbol{\Lambda}|$ as appeared in equation (9) of Rothman et al. (2008),

$$\log|\boldsymbol{I} + \boldsymbol{\Lambda}| = \operatorname{tr}(\boldsymbol{\Lambda}) - \boldsymbol{l}'\boldsymbol{K}\boldsymbol{l}, \quad \boldsymbol{K} = \int_0^1 (1-v)(\boldsymbol{I} + v\boldsymbol{\Lambda})^{-1} \otimes (\boldsymbol{I} + v\boldsymbol{\Lambda})^{-1}dv,$$

where $\boldsymbol{l}$ is $\boldsymbol{\Lambda}$ vectorized to be a $p^2 \times 1$ vector and $\otimes$ denotes the Kronecker product of matrices. Let $\hat{\boldsymbol{K}}$ be $\boldsymbol{K}$ with $\boldsymbol{\Lambda}$ and $\boldsymbol{l}$ replaced by $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{l}}$. Then,

$$\left|\log|\boldsymbol{I} + \hat{\boldsymbol{\Lambda}}| - \operatorname{tr}(\hat{\boldsymbol{\Lambda}}) - \log|\boldsymbol{I} + \boldsymbol{\Lambda}| + \operatorname{tr}(\boldsymbol{\Lambda})\right| \le \left|\hat{\boldsymbol{l}}'\hat{\boldsymbol{K}}\hat{\boldsymbol{l}} - \boldsymbol{l}'\boldsymbol{K}\boldsymbol{l}\right| \le I + II + III,$$

where $I = |\boldsymbol{l}'(\hat{\boldsymbol{K}} - \boldsymbol{K})\boldsymbol{l}|$, $II = |(\hat{\boldsymbol{l}} - \boldsymbol{l})'\hat{\boldsymbol{K}}\hat{\boldsymbol{l}}|$, and $III = |\boldsymbol{l}'\hat{\boldsymbol{K}}(\hat{\boldsymbol{l}} - \boldsymbol{l})|$. Note that

$$\begin{aligned}
\|\hat{\boldsymbol{K}} - \boldsymbol{K}\|_2 \le& \int_0^1 (1-v)\left\|[(\boldsymbol{I} + v\hat{\boldsymbol{\Lambda}})^{-1} - (\boldsymbol{I} + v\boldsymbol{\Lambda})^{-1}] \otimes (\boldsymbol{I} + v\hat{\boldsymbol{\Lambda}})^{-1}\right\|_2 dv \\
&+ \int_0^1 (1-v)\left\|(\boldsymbol{I} + v\boldsymbol{\Lambda})^{-1} \otimes [(\boldsymbol{I} + v\hat{\boldsymbol{\Lambda}})^{-1} - (\boldsymbol{I} + v\boldsymbol{\Lambda})^{-1}]\right\|_2 dv \\
&\lesssim \int_0^1 (1-v)\|(\boldsymbol{I} + v\hat{\boldsymbol{\Lambda}})^{-1} - (\boldsymbol{I} + v\boldsymbol{\Lambda})^{-1}\|_2 dv \\
&\lesssim \int_0^1 (1-v)v\|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_2 dv \\
&= \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_2/6,
\end{aligned}$$

where the first $\lesssim$ follows from the fact that eigenvalues of a Kronecker product of symmetric matrices are products of the eigenvalues of the symmetric matrices, and the second $\lesssim$ follows from the fact that the spectra of $(\boldsymbol{I}+v\boldsymbol{\Lambda})^{-1}$ and $(\boldsymbol{I}+v\hat{\boldsymbol{\Lambda}})^{-1}$ are both positive and bounded. Thus,

$$I = |\boldsymbol{l}'(\hat{\boldsymbol{K}} - \boldsymbol{K})\boldsymbol{l}| \le \|\hat{\boldsymbol{K}} - \boldsymbol{K}\|_2\boldsymbol{l}'\boldsymbol{l} \lesssim \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_2\boldsymbol{l}'\boldsymbol{l} = \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_2\|\boldsymbol{\Lambda}\|_F^2.$$

Since $\hat{\boldsymbol{K}}$ is positive definite with bounded spectrum,

$$\begin{aligned} II &= |(\hat{\boldsymbol{l}} - \boldsymbol{l})'\hat{\boldsymbol{K}}\hat{\boldsymbol{l}}| \\ &\le |(\hat{\boldsymbol{l}} - \boldsymbol{l})'\hat{\boldsymbol{K}}(\hat{\boldsymbol{l}} - \boldsymbol{l})|^{1/2}\, |\hat{\boldsymbol{l}}'\hat{\boldsymbol{K}}\hat{\boldsymbol{l}}|^{1/2} \\ &\lesssim |(\hat{\boldsymbol{l}} - \boldsymbol{l})'(\hat{\boldsymbol{l}} - \boldsymbol{l})|^{1/2}\, |\hat{\boldsymbol{l}}'\hat{\boldsymbol{l}}|^{1/2} \\ &= \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_F \|\hat{\boldsymbol{\Lambda}}\|_F. \end{aligned}$$

Similarly,

$$III = |\boldsymbol{l}'\hat{\boldsymbol{K}}(\hat{\boldsymbol{l}} - \boldsymbol{l})| \lesssim \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_F \|\boldsymbol{\Lambda}\|_F.$$

From

$$\begin{aligned} \hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda} = \hat{\boldsymbol{\Sigma}}_1^{1/2}(\hat{\boldsymbol{\nabla}} - \boldsymbol{\nabla})\hat{\boldsymbol{\Sigma}}_1^{1/2} + (\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2})\boldsymbol{\nabla}\hat{\boldsymbol{\Sigma}}_1^{1/2} \\ + \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\nabla}(\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2}), \end{aligned} \tag{S1.4}$$

$$\begin{aligned} \hat{\boldsymbol{\nabla}} - \boldsymbol{\nabla} = \hat{\boldsymbol{\Sigma}}_1^{-1}(\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}})\hat{\boldsymbol{\Sigma}}_2^{-1} + (\boldsymbol{\Sigma}_1^{-1} - \hat{\boldsymbol{\Sigma}}_1^{-1})\boldsymbol{\Delta}\hat{\boldsymbol{\Sigma}}_2^{-1} \\ + \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Delta}(\boldsymbol{\Sigma}_2^{-1} - \hat{\boldsymbol{\Sigma}}_2^{-1}), \end{aligned} \tag{S1.5}$$

and Lemma 1,

$$\begin{aligned} \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_F &\lesssim \|\hat{\boldsymbol{\nabla}} - \boldsymbol{\nabla}\|_F + 2\|\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\|_2\|\boldsymbol{\nabla}\|_F \\ &\lesssim \|\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\|_F + \|\hat{\boldsymbol{\Sigma}}_1^{-1} - \boldsymbol{\Sigma}_1^{-1}\|_2\|\boldsymbol{\Delta}\|_F \\ &\quad + \|\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1}\|_2\|\boldsymbol{\Delta}\|_F + 2\|\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\|_2\|\boldsymbol{\Delta}\|_F. \end{aligned}$$

From (C2) and (10), $\|\boldsymbol{\Delta}\|_F^2 = \sum \Delta_{ij}^2 \le (2M)^{2-\eta}\sum|\Delta_{ij}|^\eta = (2M)^{2-\eta}c_{1p}$. Using this fact and a similar proof to that of Theorem 2(i), we can show that

$$\begin{aligned} \|\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\|_F &= O_P\left(\sqrt{c_{1p}}\left(n^{-1}\log p\right)^{1/2-\eta/4}\right) \\ &= O_P\left(\sqrt{c_{1p}}\left(n^{-1}\log p\right)^{(1-\eta)/2}\right). \end{aligned} \tag{S1.6}$$

Also, $\|\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\|_2 \lesssim [\|\hat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1\|_2]^{1/2} = O_p\left(\sqrt{a_n}\right)$. Hence,

$$\|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_F = O_P\left(\sqrt{c_{1p}}\left(n^{-1}\log p\right)^{(1-\eta)/2} + \sqrt{c_{1p}a_n}\right).$$

Therefore,

$$II + III = O_P\left(c_{1p}\left(n^{-1}\log p\right)^{(1-\eta)/2} + c_{1p}\sqrt{a_n}\right) = O_P(\tau_n).$$

As for term I, by (S1.4),

$$\begin{aligned} \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_2 &\lesssim \|\hat{\boldsymbol{\nabla}} - \boldsymbol{\nabla}\|_2 + 2\|\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\|_2 \\ &\lesssim \|\hat{\boldsymbol{\Sigma}}_1^{-1} - \boldsymbol{\Sigma}_1^{-1}\|_2 + \|\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1}\|_2 + 2\|\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\|_2 \end{aligned}$$

$$= O_P\left(\sqrt{a_n}\right),$$

Hence, $I = O_P(\tau_n)$. This together with the proved first result imply the second result.

Proof of Theorem 1. In this proof, $\hat{\boldsymbol{\Sigma}}_1$, $\hat{\boldsymbol{\Sigma}}_2$, $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$ denote the MLEs of the corresponding parameters without thresholding.
(i) Note that

$$|(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_1)'\hat{\boldsymbol{\nabla}}(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_1)-(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\nabla}(\boldsymbol{x}-\boldsymbol{\mu}_1)| \le I + II + III,$$

where $I = |(\boldsymbol{x}-\boldsymbol{\mu}_1)'(\hat{\boldsymbol{\nabla}}-\boldsymbol{\nabla})(\boldsymbol{x}-\boldsymbol{\mu}_1)|$, $II = 2|(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_1)'\hat{\boldsymbol{\nabla}}(\boldsymbol{x}-\boldsymbol{\mu}_1)|$, and $III = |(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_1)'\hat{\boldsymbol{\nabla}}(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_1)|$. When $p<n$, both $\|\hat{\boldsymbol{\Sigma}}_k-\boldsymbol{\Sigma}_k\|_2$ and $\|\hat{\boldsymbol{\Sigma}}_k^{-1}-\boldsymbol{\Sigma}_k^{-1}\|_2$ are $O_P(\sqrt{p/n})$. Hence,

$$\begin{aligned}
I &= O_P\left(E[|(\boldsymbol{x}-\boldsymbol{\mu}_1)'(\hat{\boldsymbol{\nabla}}-\boldsymbol{\nabla})(\boldsymbol{x}-\boldsymbol{\mu}_1)| \mid \boldsymbol{X}]\right)\\
&= O_P\left(\sum_{j=1}^{p}|\lambda_{p,j}(\boldsymbol{\Sigma}_1^{1/2}(\hat{\boldsymbol{\nabla}}-\boldsymbol{\nabla})\boldsymbol{\Sigma}_1^{1/2})|\right)\\
&= O_P(p\|\boldsymbol{\Sigma}_1^{1/2}(\hat{\boldsymbol{\nabla}}-\boldsymbol{\nabla})\boldsymbol{\Sigma}_1^{1/2}\|_2)\\
&= O_P(p\|\hat{\boldsymbol{\nabla}}-\boldsymbol{\nabla}\|_2)\\
&= O_P(p\sqrt{p/n}).
\end{aligned}$$

Also,

$$II = O_P\left(\sqrt{(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_1)'(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_1)}\right) = O_P\left(\sqrt{p/n}\right)$$

and

$$III = O_P\left((\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_1)'(\boldsymbol{\mu}_1-\hat{\boldsymbol{\mu}}_1)\right) = O_P(p/n).$$

Hence,

$$|(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_1)'\hat{\boldsymbol{\nabla}}(\boldsymbol{x}-\hat{\boldsymbol{\mu}}_1)-(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\nabla}(\boldsymbol{x}-\boldsymbol{\mu}_1)| = O_P(p\sqrt{p/n}). \tag{S1.7}$$

By Chebyshev's inequality,

$$\begin{aligned}
|(\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{x}-\boldsymbol{\mu}_1)| &= O_P\left(\sqrt{\mathrm{Var}\left((\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{x}-\boldsymbol{\mu}_1) \mid \boldsymbol{X}\right)}\right)\\
&= O_P\left(\sqrt{(\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}}-\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta})'\boldsymbol{\Sigma}_1(\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}}-\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta})}\right)\\
&= O_P\left(\|\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}}-\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}\|\right)\\
&= O_P\left(\|(\hat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}_2^{-1})\hat{\boldsymbol{\delta}}\| + \|\boldsymbol{\Sigma}_2^{-1}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta})\|\right)\\
&= O_P\left(\|\hat{\boldsymbol{\Sigma}}_2^{-1}-\boldsymbol{\Sigma}_2^{-1}\|_2\|\hat{\boldsymbol{\delta}}\| + \|\boldsymbol{\Sigma}_2^{-1}\|_2\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\|\right)\\
&= O_P\left(p/\sqrt{n}\right) + O_P\left(\sqrt{p/n}\right).
\end{aligned}$$

Also,

$$|\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)| \le [\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}}(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)'\hat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)]^{1/2} = O_P(\sqrt{p/n}). \tag{S1.8}$$

Hence, $|\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1) - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1)|$ is bounded by

$$|(\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{x} - \boldsymbol{\mu}_1)| + |\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}(\boldsymbol{\mu}_1 - \hat{\boldsymbol{\mu}}_1)| = O_P(p/\sqrt{n}).$$

From

$$|\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}\hat{\boldsymbol{\delta}}| \le \|\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1}\|_2\|\hat{\boldsymbol{\delta}}\|^2 = O_P\left(p\sqrt{p/n}\right)$$

and

$$\begin{aligned} |\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}| &\le |\boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})| + |\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})| \\ &= O_P\left(\|\boldsymbol{\delta}\|\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|\right) + O_P\left(\|\hat{\boldsymbol{\delta}}\|\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|\right) \\ &= O_P(p/\sqrt{n}), \end{aligned}$$

we obtain that

$$\begin{aligned} |\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}| &\le |\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}\hat{\boldsymbol{\delta}}| + |\hat{\boldsymbol{\delta}}'\boldsymbol{\Sigma}_2^{-1}\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}| \\ &= O_P(p\sqrt{p/n}). \end{aligned} \tag{S1.9}$$

Next,

$$\begin{aligned} |\mathrm{tr}(\hat{\boldsymbol{\Lambda}}) - \mathrm{tr}(\boldsymbol{\Lambda})| &= |\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_1\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1})| \\ &\le |\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_2^{-1/2}(\hat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1)\hat{\boldsymbol{\Sigma}}_2^{-1/2})| + |\mathrm{tr}(\boldsymbol{\Sigma}_1^{1/2}(\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1})\boldsymbol{\Sigma}_1^{1/2})| \\ &\le p\|\hat{\boldsymbol{\Sigma}}_2^{-1/2}(\hat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1)\hat{\boldsymbol{\Sigma}}_1^{-1/2}\|_2 + p\|\boldsymbol{\Sigma}_1^{1/2}(\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1})\boldsymbol{\Sigma}_1^{1/2}\|_2 \\ &= O_P(p\sqrt{p/n}). \end{aligned}$$

Similar to the proof of Lemma 7, we have

$$\left|\log|\boldsymbol{I} + \hat{\boldsymbol{\Lambda}}| - \mathrm{tr}(\hat{\boldsymbol{\Lambda}}) - \log|\boldsymbol{I} + \boldsymbol{\Lambda}| + \mathrm{tr}(\boldsymbol{\Lambda})\right| \lesssim A,$$

where $A = \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_2\|\boldsymbol{\Lambda}\|_F^2 + \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_F\|\boldsymbol{\Lambda}\|_F + \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_F\|\hat{\boldsymbol{\Lambda}}\|_F$. By (S1.4),

$$\|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_2 \lesssim \|\hat{\boldsymbol{\nabla}} - \boldsymbol{\nabla}\|_2 + 2\|\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\|_2 = O_P((p/n)^{1/4}).$$

By (S1.4)-(S1.5),

$$\begin{aligned} \|\hat{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\|_F &\lesssim \|\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\|_F + \|\hat{\boldsymbol{\Sigma}}_1^{-1} - \boldsymbol{\Sigma}_1^{-1}\|_2\|\boldsymbol{\Delta}\|_F \\ &\quad + \|\hat{\boldsymbol{\Sigma}}_2^{-1} - \boldsymbol{\Sigma}_2^{-1}\|_2\|\boldsymbol{\Delta}\|_F + 2\|\hat{\boldsymbol{\Sigma}}_1^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\|_2\|\boldsymbol{\Delta}\|_F \\ &= O_P(\sqrt{p}(p/n)^{1/4}). \end{aligned}$$

Then, $A = O_P(p(p/n)^{1/4})$, since $\|\mathbf{\Lambda}\|_F = O(\sqrt{p})$. This together with (S1.7), (S1.8) and (S1.9) prove that the difference between the quantity on the left hand side of (1) and the quantity on the left hand side of (3) is $O_P\left(p(p/n)^{1/4}\right)$, which converges to 0. The rest of the proof is similar to the proof of Theorem 3(i).
(ii) From the proof of part (i), we have $\|\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\|^2 = O_P\left(p/n\right)$ and $\|\hat{\mathbf{\Delta}} - \mathbf{\Delta}\|_F^2 = O_P\left(p^2/n\right)$. Hence,

$$\|\hat{D}_p - D_p\|^2 = O_P\left(p^2/n\right).$$

The rest of the proof is similar to the proof of Theorem 3(ii).

Proof of Theorem 2. (i) Let $\tilde{\mathbf{\Delta}} = \tilde{\mathbf{\Sigma}}_2 - \tilde{\mathbf{\Sigma}}_1$ and Δ_{ij}^S be the (i,j)th element of $\boldsymbol{S}_2 - \boldsymbol{S}_1$. We adopt the technique in the proof of Theorem 1 in Bickel and Levina (2008). Note that,

$$\begin{aligned} \|\tilde{\mathbf{\Delta}} - \mathbf{\Delta}\|_G &= \sum_{i,j} |\Delta_{ij}^S I(|\Delta_{ij}^S| \geq t_{1n}) - \Delta_{ij}| \\ &\leq \sum_{i,j} \left|\Delta_{ij}^S I(|\Delta_{ij}^S| \geq t_{1n}) - \Delta_{ij} I(|\Delta_{ij}| \geq t_{1n})\right| \qquad \text{(S1.10)} \\ &\quad + \sum_{i,j} |\Delta_{ij}|\, I(|\Delta_{ij}| < t_{1n}). \qquad \text{(S1.11)} \end{aligned}$$

By (10), the sum in (S1.11) is bounded by

$$t_{1n}^{1-\eta} \sum_{i,j} |\Delta_{ij}|^{\eta} = t_{1n}^{1-\eta} c_{1p}.$$

The sum in (S1.10) is bounded by

$$\sum_{i,j} |\Delta_{ij}^S - \Delta_{ij}| I(|\Delta_{ij}^S| \geq t_{1n}, |\Delta_{ij}| \geq t_{1n}) \qquad \text{(S1.12)}$$

$$+ \sum_{i,j} |\Delta_{ij}| I(|\Delta_{ij}^S| < t_{1n}, |\Delta_{ij}| \geq t_{1n}) \qquad \text{(S1.13)}$$

$$+ \sum_{i,j} |\Delta_{ij}^S| I(|\Delta_{ij}^S| \geq t_{1n}, |\Delta_{ij}| < t_{1n}). \qquad \text{(S1.14)}$$

The quantity in (S1.12) is bounded by

$$\begin{aligned} \sum_{i,j} |\Delta_{ij}^S - \Delta_{ij}| I(|\Delta_{ij}| \geq t_{1n}) &\leq \max_{i,j} |\Delta_{ij}^S - \Delta_{ij}| \sum_{i,j} \frac{|\Delta_{ij}|}{t_{1n}} \\ &= O_P\left((n^{-1}\log p)^{1/2}\, c_{1p} t_{1n}^{-\eta} \right) \\ &= O_P\left(a_{1n}\right). \end{aligned}$$

The quantity in (S1.13) is bounded by

$$\max_{i,j} |\Delta_{ij}^S - \Delta_{ij}| \sum_{i,j} I(|\Delta_{ij}| \geq t_{1n}) + t_{1n} \sum_{i,j} I(|\Delta_{ij}| \geq t_{1n})$$

$$
\begin{aligned}
&\le \max_{i,j} |\Delta_{ij}^S - \Delta_{ij}| \, t_{1n}^{-\eta} c_{1p} + t_{1n}^{1-\eta} c_{1p} \\
&= O_P(a_{1n}).
\end{aligned}
$$

The quantity in (S1.14) is bounded by

$$
\begin{aligned}
&\sum_{i,j} |\Delta_{ij}^S - \Delta_{ij}| I(|\Delta_{ij}^S| \ge t_{1n}, |\Delta_{ij}| < t_{1n}) + \sum_{i,j} |\Delta_{ij}| I(|\Delta_{ij}| < t_{1n}) \\
\le\; & I + II + t_{1n}^{1-\eta} c_{1p}
\end{aligned}
$$

for some $\gamma \in (0,1)$, where

$$
\begin{aligned}
I &= \sum_{i,j} |\Delta_{ij}^S - \Delta_{ij}| I(|\Delta_{ij}^S| \ge t_{1n}, |\Delta_{ij}| \le \gamma t_{1n}) \\
&\le \max_{i,j} |\Delta_{ij}^S - \Delta_{ij}| \sum_{i,j} I(|\Delta_{ij}^S - \Delta_{ij}| > (1-\gamma) t_{1n})
\end{aligned}
$$

and

$$
\begin{aligned}
II &= \sum_{i,j} |\Delta_{ij}^S - \Delta_{ij}| I(|\Delta_{ij}^S| \ge t_{1n}, \gamma t_{1n} < |\Delta_{ij}| < t_{1n}) \\
&\le \max_{i,j} |\Delta_{ij}^S - \Delta_{ij}| \sum_{i,j} I(|\Delta_{ij}^S| \ge t_{1n}, \gamma t_{1n} < |\Delta_{ij}| < t_{1n}) \\
&\le \max_{i,j} |\Delta_{ij}^S - \Delta_{ij}| (\gamma t_{1n})^{-\eta} c_{1p} \\
&= O_P(a_{1n}).
\end{aligned}
$$

Note that

$$
P(I > 0) = P\left(\max_{i,j} |\Delta_{ij}^S - \Delta_{ij}| > (1-\gamma) t_{1n} \right) \le 2p^2 e^{-n\zeta(1-\gamma)^2 t_{1n}^2/4}
$$

for some $\zeta > 0$. Since $t_{1n} = M_1 \sqrt{\log p/n}$ and $0 < 1-\gamma < 1$, $2\log p - n\zeta(1-\gamma)^2 t_{1n}^2/4 \to -\infty$, if M_1 is sufficiently large. Hence, $I = 0$ with probability tending to 1. Combining these results, we conclude that

$$
\|\tilde{\boldsymbol{\Delta}} - \boldsymbol{\Delta}\|_G = O_P(a_{1n}). \tag{S1.15}
$$

Consider

$$
\begin{aligned}
\|\hat{\boldsymbol{\Delta}} - \tilde{\boldsymbol{\Delta}}\|_G &= \sum_{i,j} |\Delta_{ij}^S + s_{1ij}| I(|s_{1ij}| \ge t_{2n}, |s_{2ij}| < t_{2n}, |\Delta_{ij}^S| \ge t_{1n}) \\
&\quad + \sum_{i,j} |\Delta_{ij}^S - s_{2ij}| I(|s_{1ij}| < t_{2n}, |s_{2ij}| \ge t_{2n}, |\Delta_{ij}^S| \ge t_{1n}) \\
&\le 2t_{2n} \sum_{i,j} I(|\Delta_{ij}^S| \ge t_{1n}) \\
&\le 2t_{2n}(III + IV),
\end{aligned}
$$

where

$$III = \sum_{i,j} I(|\Delta_{ij}^S - \Delta_{ij}| \geq (1-\gamma)t_{1n})$$

and

$$IV = \sum_{i,j} I(|\Delta_{ij}| \geq \gamma t_{1n}) \leq \sum_{i,j} \frac{|\Delta_{ij}|}{\gamma t_{1n}}$$

for some $\gamma \in (0,1)$. In analogous to the aforementioned analysis of term I, $III = 0$ with probability tending to 1. On the other hand, $IV \leq c_{1p}/(\gamma t_{1n})^\eta$. Hence, $2t_{2n}IV = O_P(a_{1n})$. This shows that $\|\tilde{\mathbf{\Delta}} - \hat{\mathbf{\Delta}}\|_G = O_P(a_{1n})$, which together with result (S1.15) imply that $\|\hat{\mathbf{\Delta}} - \mathbf{\Delta}\|_G = O_P(a_{1n})$.
(ii) It follows from Lemma 1 and result (S1.5) that

$$\begin{aligned}\|\tilde{\mathbf{\nabla}} - \mathbf{\nabla}\|_G \leq & \|\hat{\mathbf{\Sigma}}_1^{-1}\|_1 \|\mathbf{\Delta} - \hat{\mathbf{\Delta}}\|_G \|\hat{\mathbf{\Sigma}}_2^{-1}\|_1 + \|\hat{\mathbf{\Sigma}}_1^{-1} - \mathbf{\Sigma}_1^{-1}\|_1 \|\mathbf{\Delta}\|_G \|\hat{\mathbf{\Sigma}}_2^{-1}\|_1 \\ & + \|\mathbf{\Sigma}_1^{-1}\|_1 \|\mathbf{\Delta}\|_G \|\hat{\mathbf{\Sigma}}_1^{-1} - \mathbf{\Sigma}_1^{-1}\|_1.\end{aligned}$$

By Lemma 3,

$$\|\hat{\mathbf{\Sigma}}_1^{-1} - \mathbf{\Sigma}_1^{-1}\|_1 = O_P\left(a_n v_p^2\right).$$

Then, it holds that

$$\begin{aligned}\|\hat{\mathbf{\nabla}} - \mathbf{\nabla}\|_G &= O_P\left(\|\hat{\mathbf{\Delta}} - \mathbf{\Delta}\|_G v_p^2 + a_n c_{1p} v_p^3\right) \\ &= O_P\left(c_{1p}(n^{-1}\log p)^{(1-\eta)/2} v_p^2 + a_n c_{1p} v_p^3\right) \\ &= O_P(\tau_n),\end{aligned}$$

where the second equality follows from part (i).

Proof of Theorem 3. (i) When D_p is bounded. Let T_p be defined as in Lemma 2 and

$$\begin{aligned}\hat{T}_p|\boldsymbol{X} &= (\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1)'\hat{\mathbf{\nabla}}(\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1) - 2\hat{\boldsymbol{\delta}}'\hat{\mathbf{\Sigma}}_2^{-1}(\boldsymbol{x} - \hat{\boldsymbol{\mu}}_1) \\ \hat{\varphi}_p|\boldsymbol{X} &= \mathrm{tr}(\hat{\mathbf{\Lambda}}) + \hat{\boldsymbol{\delta}}'\hat{\mathbf{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}} - \log(|\hat{\mathbf{\Sigma}}_1|/|\hat{\mathbf{\Sigma}}_2|) \\ \varphi_p &= \mathrm{tr}(\mathbf{\Lambda}) + \boldsymbol{\delta}'\mathbf{\Sigma}_2^{-1}\boldsymbol{\delta} - \log(|\mathbf{\Sigma}_1|/|\mathbf{\Sigma}_2|)\end{aligned}$$

Denote $F_p(\cdot)$ the cumulative distribution function (c.d.f) of $T_p - \mathrm{E}(T_p)$ and $\hat{F}_p(\cdot)$ the conditional c.d.f of $\hat{T}_p|\boldsymbol{X} - \mathrm{E}(\hat{T}_p|\boldsymbol{X})$. From Lemma 4-7, we have

$$[\hat{T}_p|\boldsymbol{X} - E(\hat{T}_p|\boldsymbol{X})] - [T_p - \mathrm{E}(T_p)] \xrightarrow{P} 0 \quad \text{and} \quad \hat{\varphi}_p|\boldsymbol{X} - \varphi_p \xrightarrow{P} 0.$$

It follows from (3) that

$$R_2 - R_{B2} = \hat{F}_p(-\hat{\varphi}_p) - F_p(-\varphi_p),$$

where R_{B2} and R_2 are defined in (2) and (4).

Next, we show that if (15) holds,

$$\hat{F}_p(-\hat{\varphi}_p) - F_p(-\varphi_p) \xrightarrow{P} 0. \tag{S1.16}$$

Similarly, we can show that $R_1 - R_{B1} \xrightarrow{P} 0$. Then, (i) of Theorem 3 is proved.

We prove (S1.16) by a subsequence argument. For any subsequence $\{p_k\} \subset \{p\}$, there is a further subsequence $\{p_{k_t}\} \subset \{p_k\}$ such that

$$[\hat{T}_{p_{k_t}}|\boldsymbol{X} - E(\hat{T}_{p_{k_t}}|\boldsymbol{X})] - [T_{p_{k_t}} - \mathrm{E}(T_{p_{k_t}})] \xrightarrow{a.s.} 0, \tag{S1.17}$$

$$\hat{\varphi}_{p_{k_t}}|\boldsymbol{X} - \varphi_{p_{k_t}} \xrightarrow{a.s.} 0. \tag{S1.18}$$

Then,

$$\begin{aligned}
&|\hat{F}_{p_{k_t}}(-\hat{\varphi}_{p_{k_t}}) - F_{p_{k_t}}(-\hat{\varphi}_{p_{k_t}}) + F_{p_{k_t}}(-\hat{\varphi}_{p_{k_t}}) - F_{p_{k_t}}(-\varphi_{p_{k_t}})| \\
&\le \sup_x |\hat{F}_{p_{k_t}}(x) - F_{p_{k_t}}(x)| + \sup_x |F'_{p_{k_t}}(x)|(\varphi_{p_{k_t}} - \hat{\varphi}_{p_{k_t}}) \\
&\le (1 + \sup_x |F'_{p_{k_t}}(x)|) d_L(\hat{F}_{p_{k_t}}, F_{p_{k_t}}) + \sup_x |F'_{p_{k_t}}(x)|(\varphi_{p_{k_t}} - \hat{\varphi}_{p_{k_t}}),
\end{aligned}$$

where the last inequality follows from a well-known inequality on page 43 of Petrov (1995) and $d_L(\hat{F}_{p_{k_t}}, F_{p_{k_t}})$ is the Levy metric between $\hat{F}_{p_{k_t}}$ and $F_{p_{k_t}}$.

Under (C4), it holds that

$$\sup_x |F'_{p_{k_t}}(x)| \le C, \tag{S1.19}$$

where C does not depend on the index p_{k_t}. In addition, (S1.17) implies that $d_L(\hat{F}_{p_{k_t}}, F_{p_{k_t}}) \to 0$. Then, this together with (S1.18) and (S1.19) proves that

$$|\hat{F}_{p_{k_t}}(-\hat{\varphi}_{p_{k_t}}) - F_{p_{k_t}}(-\varphi_{p_{k_t}})| \xrightarrow{a.s.} 0.$$

By the above subsequence argument, we prove (S1.16).

In the following, we show that (S1.19) holds in some meaningful cases.
Case 1: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Note that,

$$\begin{aligned}
T_p &= \sum_{j=1}^{p} \lambda_{p,j}\tilde{z}_{p,j}^2 - 2a_{p,j}\tilde{z}_{p,j} \\
&= \sum_{\lambda_{p,j}\neq 0} \lambda_{p,j}\left(\tilde{z}_{p,j} - \frac{a_{p,j}}{\lambda_{p,j}}\right)^2 - \sum_{\lambda_{p,j}\neq 0} \frac{a_{p,j}^2}{\lambda_{p,j}} - \sum_{\lambda_{p,j}=0} a_{p,j}\tilde{z}_{p,j},
\end{aligned} \tag{S1.20}$$

where $\tilde{z}_{p,j}, j = 1, \ldots, p$ are i.i.d from $N(0,1)$, $\lambda_{p,j}$ is the jth smallest eigenvalue of $\boldsymbol{\Lambda}$ and $a_{p,j}$ is the jth component of $\boldsymbol{U}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta}$, where $\boldsymbol{U}'\boldsymbol{\Lambda}\boldsymbol{U} = \mathrm{diag}(\lambda_{p,1}, \ldots, \lambda_{p,p})$.

If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, T_p reduces to a normal random variable with mean 0 and

$$\mathrm{Var}(T_p) = \boldsymbol{\delta}'\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\delta},$$

which is bounded below by a constant under (C3). Then, (S1.19) holds.

Case 2: There are at least two eigenvalues of $\mathbf{\Lambda}$ in $(-\infty, m]$ or $[m, \infty)$. Without loss of generality, assume the two eigenvalues $\lambda_{p,1} \geq m$ and $\lambda_{p,2} \geq m$.

Let

$$Y_{p,1} = \lambda_{p,1}\left(\tilde{z}_{p,1} - \frac{a_{p,1}}{\lambda_{p,1}}\right)^2$$
$$Y_{p,2} = \lambda_{p,2}\left(\tilde{z}_{p,2} - \frac{a_{p,2}}{\lambda_{p,2}}\right)^2$$

and $Y_p = Y_{p,1} + Y_{p,2}$. Denote $f_{p,1}$ the p.d.f of $Y_{p,1}$, $f_{p,2}$ the p.d.f of $Y_{p,2}$; $\tilde{f}_{p,1}$ the p.d.f of the noncentral χ^2-distribution with 1 degree of freedom and noncentrality parameter $a_{p,1}^2/\lambda_{p,1}^3$, $\tilde{f}_{p,2}$ the p.d.f of the noncentral χ^2-distribution with 1 degree of freedom and noncentrality parameter $a_{p,2}^2/\lambda_{p,2}^3$.

Then,

$$\frac{f_{p,1}(y)}{\tilde{f}_{p,1}(y)} = \frac{\frac{1}{2\lambda_{p,1}} e^{-(y/\lambda_{p,1}+a_{p,1}^2/\lambda_{p,1}^2)/2}(\lambda_{p,1}a_{p,1}^{-2}y)^{-1/4}}{\frac{1}{2}e^{-(y+a_{p,1}^2/\lambda_{p,1}^3)/2}(\lambda_{p,1}^3 a_{p,1}^{-2}y)^{-1/4}}$$
$$= \frac{1}{\lambda_{p,1}^{1/2}} e^{-y(1+\lambda_{p,1}^{-1})/2} e^{a_{p,1}^2\lambda_{p,1}^{-3}(1-\lambda_{p,1})/2}.$$

Under (C1) and (C2), $a_{p,1} = O(1)$. This together with $\lambda_{p,1} \geq m$ shows that

$$\sup_{y>0} \frac{f_{p,1}(y)}{\tilde{f}_{p,1}(y)} \leq c_1,$$

where c_1 does not depend on p. Similarly, we have

$$\sup_{y>0} \frac{f_{p,2}(y)}{\tilde{f}_{p,2}(y)} \leq c_2.$$

Let $f_{p,0}$ denote the p.d.f of Y_p. By convolution formula,

$$f_{p,0}(y) = \int_0^y f_{p,1}(y-t) f_{p,2}(t)dt.$$

Hence,

$$\sup_{y>0} f_{p,0}(y) \leq \sup_{y>0} \int_0^y c_1 c_2 \tilde{f}_{p,1}(y-t)\tilde{f}_{p,2}(t)dt$$
$$= c_1 c_2 \sup_{y>0} \tilde{f}_0(y),$$

where $\tilde{f}_0(y)$ is the p.d.f of noncentral χ^2-distribution with 2 degrees of freedom and noncentrality parameter $(a_{p,1}^2\lambda_{p,1}^{-3} + a_{p,2}^2\lambda_{p,2}^{-3})$. Therefore,

$$\tilde{f}_0(y) = \frac{1}{2} e^{-(y+a_{p,1}^2\lambda_{p,1}^{-3}+a_{p,2}^2\lambda_{p,2}^{-3})/2} I_0\left(\sqrt{y(a_{p,1}^2\lambda_{p,1}^{-3} + a_{p,2}^2\lambda_{p,2}^{-3})}\right),$$

where $I_0(\cdot)$ is the Bessel function of the first kind. By the property of $I_0(\cdot)$, we have $\sup_{y>0} I_0\left(\sqrt{y(a_{p,1}^2\lambda_{p,1}^{-3}+a_{p,2}^2\lambda_{p,2}^{-3})}\right) \le 1$ and $a_{p,1}^2\lambda_{p,1}^{-3}+a_{p,2}^2\lambda_{p,2}^{-3} = O(1)$ by (C1) and (C2). Then, it holds that

$$\sup_{y>0} f_{p,0}(y) \le c_1 c_2 \sup_{y>0} \tilde{f}_0(y) \le C,$$

where C does not depend on p.

Then, it follows from (S1.20) that T_p is a sum of p independent random variables. By the property of convolution, the density of T_p

$$\sup_x |F_p'(x)| \le \sup_x |\tilde{f}_{p,0}(x)| \le C.$$

Therefore, (S1.19) holds.

(ii) When $D_p \to \infty$, by Lemma 2, the misclassification rate of Bayes rule $R_B \to 0$. Hence, it suffices to show $R_{\mathrm{SQDA}}(\boldsymbol{X}) \xrightarrow{P} 0$. Let

$$\hat{Z}_p|\boldsymbol{X} = \frac{\hat{T}_p|\boldsymbol{X} - \mathrm{E}(\hat{T}_p|\boldsymbol{X})}{[\mathrm{Var}(\hat{T}_p|\boldsymbol{X})]^{1/2}} \quad \text{and} \quad \hat{\xi}_p = \frac{\hat{\varphi}_p|\boldsymbol{X}}{[\mathrm{Var}(\hat{T}_p|\boldsymbol{X})]^{1/2}},$$

where $\hat{T}_p|\boldsymbol{X}$ and $\hat{\varphi}_p|\boldsymbol{X}$ are as defined in (i). Note that

$$\mathrm{E}(\hat{T}_p|\boldsymbol{X}) = \mathrm{tr}(\hat{\boldsymbol{\Lambda}}) \quad \text{and} \quad \mathrm{Var}(\hat{T}_p|\boldsymbol{X}) = 2\mathrm{tr}(\hat{\boldsymbol{\Lambda}}^2) + 4\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\Sigma}}_1\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}}.$$

For $\hat{D}_p = \sqrt{\|\hat{\boldsymbol{\delta}}\|^2 + \|\hat{\boldsymbol{\Delta}}\|_F^2}$,

$$\begin{aligned}\hat{D}_p &\ge D_p - \sqrt{\|\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\|^2 + \|\hat{\boldsymbol{\Delta}}-\boldsymbol{\Delta}\|_F^2}\\ &= D_p - \sqrt{O_P\left(\max\{b_n, a_{1n}\}\right)}\\ &= D_p\left(1 - \sqrt{O_P\left(\max\{b_n, a_{1n}\}/D_p^2\right)}\right),\end{aligned}$$

where the first identity follows by (12), (S1.6) and (C5). Therefore, $\hat{D}_p \to \infty$ if $\max\{b_n, a_{1n}\}/D_p^2 \to 0$. Using the subsequence argument as in (i) and a proof analogous to Lemma 2,

$$\left|F_{\hat{Z}_p|\boldsymbol{X}}(-\hat{\xi}_p) - \Phi(-\hat{\xi}_p)\right| \xrightarrow{P} 0.$$

Since

$$\hat{\xi}_p = \frac{\hat{\varphi}_p|\boldsymbol{X}}{[\mathrm{Var}(\hat{T}_p)]^{1/2}} = \frac{\mathrm{tr}(\hat{\boldsymbol{\Lambda}}) - \log(|\hat{\boldsymbol{\Sigma}}_1|/|\hat{\boldsymbol{\Sigma}}_2|) + \hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}}}{[2\mathrm{tr}(\hat{\boldsymbol{\Lambda}})^2 + 4\hat{\boldsymbol{\delta}}'\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\Sigma}}_1\hat{\boldsymbol{\Sigma}}_2^{-1}\hat{\boldsymbol{\delta}}]^{1/2}} \asymp \hat{D}_p,$$

there exists a constant c_2 such that

$$\Phi(-\hat{\xi}_p) \le \Phi(-c_2\hat{D}_p) \le \Phi\left(-c_2 D_p\left(1 - \sqrt{O_P\left(\max\{b_n, a_{1n}\}/D_p^2\right)}\right)\right).$$

Hence, if $\max\{b_n, a_{1n}\}/D_p^2 \to 0$, $F_{\hat{Z}_p|\boldsymbol{X}}(-\hat{\xi}_p) = R_2(\boldsymbol{X}) \xrightarrow{P} 0$, where $R_2(\boldsymbol{X})$ is defined in (i). Similarly, we can prove that $R_1(\boldsymbol{X}) \to 0$, and the result follows.

References

Bickel, P. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics* **36,** 2577–2604.

Petrov, V. (1995). *Limit Theorems of Probability Theory.* Oxford Science Publications.

Rothman, A., Bickel, P., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics* **2,** 494–515.

Saulis, L. and Statulevicius, V. (1991). *Limit Theorems for Large Deviations.* Springer.