Communication-Efficient Distributed Linear Discriminant Analysis for Binary Classification

Mengyu Li[1,2] and Junlong Zhao[2*]

Renmin University of China, Beijing, China[1]

Beijing Normal University, Beijing, China[2]

Supplementary Material

In the online supplementary material, we provide proofs of the theoretical results stated within the paper. Before this, we state and prove two lemmas that are used in the following proofs.

S1. Lemmas

Lemma 1. *Let* $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n \in \mathbb{R}^p$ *be* n *i.i.d. random vectors following normal distribution* $N_p(\boldsymbol{\mu}, \Sigma)$*, and the sample mean vector* $\hat{\boldsymbol{\mu}} = n^{-1}\sum_{i=1}^n \boldsymbol{X}_i$*. Then, the difference between* $\hat{\boldsymbol{\mu}}$ *and* $\boldsymbol{\mu}$ *can be bounded by*

$$\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2 \leq \sqrt{\frac{\mathrm{tr}(\Sigma)}{n\varepsilon}}$$

with probability at least $1 - \varepsilon$*.*

*Corresponding author. (Email: zhaojunlong928@126.com)

Proof. By the fact $\hat{\boldsymbol{\mu}} \sim N_p(\boldsymbol{\mu}, n^{-1}\Sigma)$, it is easy to see that

$$E(\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2^2) = \mathrm{tr}[\mathrm{Cov}(\hat{\boldsymbol{\mu}})] = \mathrm{tr}\left(\frac{1}{n}\Sigma\right) = \frac{1}{n}\,\mathrm{tr}(\Sigma).$$

Using Markov's inequality, for any $t > 0$, we have

$$\mathrm{P}\left[\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2^2 \geq t\right] \leq \frac{\mathrm{tr}(\Sigma)}{nt}.$$

Then for any $\varepsilon \in (0, 1]$, we see that $\mathrm{tr}(\Sigma)/(nt) \leq \varepsilon$ is equivalent to $t \geq \mathrm{tr}(\Sigma)/(n\varepsilon)$. Thus, we have

$$\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2 \leq \sqrt{\frac{\mathrm{tr}(\Sigma)}{n\varepsilon}}$$

with probability at least $1 - \varepsilon$. □

Lemma 2. *Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n \in \mathbb{R}^p$ be $n > p$ i.i.d. random vectors following normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$, and the sample covariance matrix $S_n = n^{-1}\sum_{i=1}^n (\boldsymbol{X}_i - \boldsymbol{\mu})(\boldsymbol{X}_i - \boldsymbol{\mu})^\top$. Then, the difference between S_n^{-1} and Θ can be bounded by*

$$\|S_n^{-1} - \Theta\|_2 \leq \|\Theta\|_2 \left(2\sqrt{\frac{p}{n}} + \tau\right)$$

with probability at least $1 - c_3 e^{-n\tau^2}$ for all $\tau \in (0, 1]$, where c_3 is a positive constant.

Proof. See the proof of Proposition 11.19 in Wainwright (2019). □

S2. Proof of Proposition 1

Proof. We first review the results of Dobriban and Sheng (2018). For the elliptical model $Z = \Gamma^{1/2} U \Sigma^{1/2} \in \mathbb{R}^{m \times p}$, Dobriban and Sheng (2018) showed the deterministic equivalent of the sample covariance $\widetilde{\Sigma}_*^{-1} \asymp e_p \Theta$ under following assumptions:

(a) The entries of U are *i.i.d.* random variables, with zero mean, unit variance, and finite $8+\varepsilon$-th moment, for some $\varepsilon > 0$.

(b) The eigenvalues of Σ and the entries of Γ are uniformly bounded away from zero and infinity.

(c) As $m \to \infty$, $p \to \infty$ satisfies p/m bounded away from zero and infinity.

Now we prove our conclusions on $\widetilde{\Sigma}^{(l)}$ and $\widetilde{\Sigma}$.

Step 1. We prove the conclusions on $\widetilde{\Sigma}^{(l)}$ for the following cases: (1) $\gamma_p^{(l)} \to c \in (0,1)$, and (2) $\gamma_p^{(l)} \to 0$.

Step 1.1. We consider the case (1). For the model in this study, let $\{\tilde{\boldsymbol{X}}_i^{(l)} = \boldsymbol{X}_i^{(l)} - \boldsymbol{\mu}_1, i = 1, \cdots, n_{1l}\}$ and $\{\tilde{\boldsymbol{Y}}_i^{(l)} = \boldsymbol{Y}_i^{(l)} - \boldsymbol{\mu}_2, i = 1, \cdots, n_{2l}\}$ be the centralized samples on the lth machine. Thus $\tilde{\boldsymbol{X}}_i^{(l)}$ and $\tilde{\boldsymbol{Y}}_i^{(l)}$ follow the same normal distribution $N_p(0, \Sigma)$. Then, the pooled data matrix

$$\tilde{Z}^{(l)} = \left(\tilde{\boldsymbol{X}}_1^{(l)\top}, \cdots, \tilde{\boldsymbol{X}}_{n_{1l}}^{(l)\top}, \tilde{\boldsymbol{Y}}_1^{(l)\top}, \cdots, \tilde{\boldsymbol{Y}}_{n_{2l}}^{(l)\top}\right)^\top$$

follows the model $U\Sigma^{1/2}$, where the entries of U are $i.i.d.$ random variables from $N(0,1)$. Therefore, assumption (a) holds for our setting. One can see that $\tilde{Z}^{(l)}$ is a special case of the elliptical model $Z=\Gamma^{1/2}U\Sigma^{1/2}$, where $\Gamma = I_{n^{(l)}}$. This together with condition (i) of (C1) implies that the assumption (b) holds. Assumption (c) also holds according to the fact $\gamma_p^{(l)} \to c \in (0,1)$. Remind that $\Gamma = I_{n^{(l)}}$ in our model, thus $e_p = 1/(1-\gamma_p^{(l)})$ by the equation $1 = (n^{(l)})^{-1}\operatorname{tr}[e_p\Gamma(I_{n^{(l)}} + \gamma_p^{(l)} e_p \Gamma)^{-1}]$. Since the assumptions (a)-(c) in Dobriban and Sheng (2018) hold for our setting, the sample covariance

$$\begin{aligned}\widetilde{\Sigma}^{(l)} =& \frac{1}{n^{(l)}}\left[\sum_{i=1}^{n_{1l}}(\boldsymbol{X}_i^{(l)}-\boldsymbol{\mu}_1)(\boldsymbol{X}_i^{(l)}-\boldsymbol{\mu}_1)^\top + \sum_{i=1}^{n_{2l}}(\boldsymbol{Y}_i^{(l)}-\boldsymbol{\mu}_2)(\boldsymbol{Y}_i^{(l)}-\boldsymbol{\mu}_2)^\top\right]\\ =&(n^{(l)})^{-1}\tilde{Z}^{(l)\top}\tilde{Z}^{(l)}\end{aligned}$$

has the deterministic equivalent

$$(\widetilde{\Sigma}^{(l)})^{-1} \asymp \frac{1}{1-\gamma_p^{(l)}}\Theta. \tag{S2.1}$$

Step 1.2. We consider the case (2), showing that $\|(\widetilde{\Sigma}^{(l)})^{-1}-\Theta\|_2 \to 0$ with probability 1. By Borel–Cantelli Lemma, it is sufficient to show that for any small $\varepsilon > 0$, $\lim_{N\to\infty}\sum_{n^{(l)}>N} \mathrm{P}(\|(\widetilde{\Sigma}^{(l)})^{-1}-\Theta\|_2 > \varepsilon) = 0$. Now we prove this result. By Lemma 2, it is shown that for all $\tau \in (0,1]$,

$$\mathrm{P}\left\{\|(\widetilde{\Sigma}^{(l)})^{-1}-\Theta\|_2 \le \|\Theta\|_2\left(2\sqrt{\gamma_p^{(l)}}+\tau\right)\right\} \ge 1 - c_3 e^{-n^{(l)}\tau^2}.$$

Because the eigenvalues of Σ are bounded away from zero and infinity, we have $\|\Theta\|_2 < c_4$ for some positive constant c_4. Thus,

$$\mathrm{P}\left\{\|(\widetilde{\Sigma}^{(l)})^{-1} - \Theta\|_2 < c_4\left(2\sqrt{\gamma_p^{(l)}} + \tau\right)\right\} \geq 1 - c_3 e^{-n^{(l)}\tau^2}.$$

For any $\varepsilon \in (0, c_4]$, take $\tau = c_4^{-1}\varepsilon - 2(\gamma_p^{(l)})^{1/2}$. Since $\gamma_p^{(l)} \to 0$, we have $0 < \gamma_p^{(l)} < c_4^{-2}\varepsilon^2/16$ when $n^{(l)}$ is sufficient large. Then it holds that $\tau \in (c_4^{-1}\varepsilon/2, c_4^{-1}\varepsilon) \subset (0, 1]$. Thus, we have

$$\mathrm{P}\left\{\|(\widetilde{\Sigma}^{(l)})^{-1} - \Theta\|_2 < \varepsilon\right\} \geq 1 - c_3 e^{-n^{(l)}\left(c_4^{-1}\varepsilon - 2\sqrt{\gamma_p^{(l)}}\right)^2} > 1 - c_3 e^{-n^{(l)}\left(c_4^{-1}\varepsilon/2\right)^2},$$

when $n^{(l)}$ is sufficient large. It follows that

$$\sum_{n^{(l)} > N} \mathrm{P}\left\{\|(\widetilde{\Sigma}^{(l)})^{-1} - \Theta\|_2 \geq \varepsilon\right\} \leq c_3 \sum_{n^{(l)} > N} e^{-n^{(l)}\left(c_4^{-1}\varepsilon/2\right)^2},$$

when N is sufficient large. Moreover, it is easy to see that

$$\begin{aligned}\sum_{n^{(l)} > N} e^{-n^{(l)}\left(c_4^{-1}\varepsilon/2\right)^2} &= \sum_{n^{(l)}=1}^{\infty} e^{-n^{(l)}\left(c_4^{-1}\varepsilon/2\right)^2} - \sum_{n^{(l)}=1}^{N} e^{-n^{(l)}\left(c_4^{-1}\varepsilon/2\right)^2} \\ &= \frac{e^{-N\left(c_4^{-1}\varepsilon/2\right)^2}}{e^{\left(c_4^{-1}\varepsilon/2\right)^2} - 1} \to 0,\end{aligned}$$

as $N \to \infty$. Therefore, we have $\|(\widetilde{\Sigma}^{(l)})^{-1} - \Theta\|_2 \to_{a.s.} 0$ as $n^{(l)} \to \infty$. Consequently, for any matrix sequence C_n with $\sup_n \|C_n\|_* < \infty$, we have

$$\lim_{n^{(l)} \to \infty} |\operatorname{tr}[C_n((\widetilde{\Sigma}^{(l)})^{-1} - \Theta)]| \leq \lim_{n^{(l)} \to \infty} \|C_n\|_* \|(\widetilde{\Sigma}^{(l)})^{-1} - \Theta\|_2 = 0$$

almost surely. Then by Definition 1, the sample covariance $\widetilde{\Sigma}^{(l)}$ has the deterministic equivalent

$$(\widetilde{\Sigma}^{(l)})^{-1} \asymp \Theta. \tag{S2.2}$$

Combining (S2.1) and (S2.2), we conclude that $(\widetilde{\Sigma}^{(l)})^{-1} \asymp (1-\gamma_p^{(l)})^{-1}\Theta$ for $\gamma_p^{(l)} \to c \in [0,1)$.

Step 2. For the sample covariance $\widetilde{\Sigma}$, note that $\gamma_p \to 0$ according to the condition (ii) of (C1). Similarly to Step 1.2, we have $\widetilde{\Sigma}^{-1} \asymp \Theta$. This completes the proof. □

S3. Proof of Theorem 1

Proof. Denote $\widetilde{\Theta} = n^{-1}\sum_{l=1}^{k} n^{(l)}(\widetilde{\Sigma}^{(l)})^{-1}$. According to Proposition 1, we have

$$(\widetilde{\Sigma}^{(l)})^{-1} \asymp \frac{1}{1-\gamma_p^{(l)}}\Theta, \quad 1 \le l \le k.$$

Then by the calculus rules of deterministic equivalents (Theorem 4.3 in Dobriban and Sheng (2018)), we obtain that

$$\widetilde{\Theta} = \frac{1}{n}\sum_{l=1}^{k} n^{(l)}(\widetilde{\Sigma}^{(l)})^{-1} \asymp \sum_{l=1}^{k} \frac{\gamma_p}{\gamma_p^{(l)}(1-\gamma_p^{(l)})}\Theta = \frac{1}{1-k\gamma_p}\Theta. \tag{S3.1}$$

Since $n^{(l)} \equiv n/k$ for $1 \le l \le k$, we have $\widetilde{\Theta} = k^{-1}\sum_{l=1}^{k}(\widetilde{\Sigma}^{(l)})^{-1}$. To simplify $\hat{R}_{two}$, we first decompose $(\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_2)^\top \bar{\Theta}\hat{\boldsymbol{\mu}}_d$ and $\bar{\Delta}_p^2$ as follows.

$$\begin{aligned}
&(\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_2)^\top \bar{\Theta}\hat{\boldsymbol{\mu}}_d \\
=&\left[(\boldsymbol{\mu}_a - \boldsymbol{\mu}_2) + (\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_a)\right]^\top \bar{\Theta}\hat{\boldsymbol{\mu}}_d \\
=&(\boldsymbol{\mu}_a - \boldsymbol{\mu}_2)^\top \bar{\Theta}\left[\boldsymbol{\mu}_d + (\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d)\right] + (\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_a)^\top \bar{\Theta}\hat{\boldsymbol{\mu}}_d \\
=&\frac{1}{2}\boldsymbol{\mu}_d^\top \bar{\Theta}\boldsymbol{\mu}_d + \frac{1}{2}\boldsymbol{\mu}_d^\top \bar{\Theta}(\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d) + (\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_a)^\top \bar{\Theta}\hat{\boldsymbol{\mu}}_d.
\end{aligned} \tag{S3.2}$$

Then, we insert the average inverse sample covariance $\widetilde{\Theta}$ with known $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Thus, (S3.2) can be further decomposed into

$$\begin{aligned}&\frac{1}{2}\boldsymbol{\mu}_d^\top\widetilde{\Theta}\boldsymbol{\mu}_d+\frac{1}{2}\boldsymbol{\mu}_d^\top(\bar{\Theta}-\widetilde{\Theta})\boldsymbol{\mu}_d+\frac{1}{2}\boldsymbol{\mu}_d^\top\bar{\Theta}(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)+(\hat{\boldsymbol{\mu}}_a-\boldsymbol{\mu}_a)^\top\bar{\Theta}\boldsymbol{\mu}_d\\ =&\Delta_{0,1}+\frac{1}{2}\Delta_1+\frac{1}{2}\Delta_2+\Delta_3,\end{aligned}$$

where $\Delta_{0,1}=\boldsymbol{\mu}_d^\top\widetilde{\Theta}\boldsymbol{\mu}_d/2$, $\Delta_1=\boldsymbol{\mu}_d^\top(\bar{\Theta}-\widetilde{\Theta})\boldsymbol{\mu}_d$, $\Delta_2=\boldsymbol{\mu}_d^\top\bar{\Theta}(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)$, and $\Delta_3=(\hat{\boldsymbol{\mu}}_a-\boldsymbol{\mu}_a)^\top\bar{\Theta}\boldsymbol{\mu}_d$. Moreover, it holds that

$$\begin{aligned}\bar{\Delta}_p^2=&\hat{\boldsymbol{\mu}}_d^\top\bar{\Theta}\Sigma\bar{\Theta}\hat{\boldsymbol{\mu}}_d\\ =&\left[\boldsymbol{\mu}_d+(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)\right]^\top\bar{\Theta}\Sigma\bar{\Theta}\hat{\boldsymbol{\mu}}_d\\ =&\boldsymbol{\mu}_d^\top\bar{\Theta}\Sigma\bar{\Theta}\left[\boldsymbol{\mu}_d+(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)\right]+(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)^\top\bar{\Theta}\Sigma\bar{\Theta}\hat{\boldsymbol{\mu}}_d\\ =&\boldsymbol{\mu}_d^\top\bar{\Theta}\Sigma\bar{\Theta}\boldsymbol{\mu}_d+\boldsymbol{\mu}_d^\top\bar{\Theta}\Sigma\bar{\Theta}(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)+(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)^\top\bar{\Theta}\Sigma\bar{\Theta}\hat{\boldsymbol{\mu}}_d\\ =&\boldsymbol{\mu}_d^\top[\widetilde{\Theta}+(\bar{\Theta}-\widetilde{\Theta})]\Sigma[\widetilde{\Theta}+(\bar{\Theta}-\widetilde{\Theta})]\boldsymbol{\mu}_d+\boldsymbol{\mu}_d^\top\bar{\Theta}\Sigma\bar{\Theta}(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)\\ &+(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)^\top\bar{\Theta}\Sigma\bar{\Theta}\hat{\boldsymbol{\mu}}_d \qquad\qquad (S3.3)\\ =&\boldsymbol{\mu}_d^\top\widetilde{\Theta}\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d+2\boldsymbol{\mu}_d^\top(\bar{\Theta}-\widetilde{\Theta})\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d+\boldsymbol{\mu}_d^\top(\bar{\Theta}-\widetilde{\Theta})\Sigma(\bar{\Theta}-\widetilde{\Theta})\boldsymbol{\mu}_d\\ &+\boldsymbol{\mu}_d^\top\bar{\Theta}\Sigma\bar{\Theta}(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)+(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)^\top\bar{\Theta}\Sigma\bar{\Theta}\hat{\boldsymbol{\mu}}_d\\ =&\boldsymbol{\mu}_d^\top\widetilde{\Theta}\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d+2\boldsymbol{\mu}_d^\top(\bar{\Theta}-\widetilde{\Theta})\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d+\boldsymbol{\mu}_d^\top(\bar{\Theta}-\widetilde{\Theta})\Sigma(\bar{\Theta}-\widetilde{\Theta})\boldsymbol{\mu}_d\\ &+(\hat{\boldsymbol{\mu}}_d+\boldsymbol{\mu}_d)^\top\bar{\Theta}\Sigma\bar{\Theta}(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)\\ =&\Delta_{0,2}+2\Delta_4+\Delta_5+\Delta_6,\end{aligned}$$

where $\Delta_{0,2}=\boldsymbol{\mu}_d^\top\widetilde{\Theta}\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d$, $\Delta_4=\boldsymbol{\mu}_d^\top(\bar{\Theta}-\widetilde{\Theta})\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d$, $\Delta_5=\boldsymbol{\mu}_d^\top(\bar{\Theta}-\widetilde{\Theta})\Sigma(\bar{\Theta}-\widetilde{\Theta})\boldsymbol{\mu}_d$, and $\Delta_6=(\hat{\boldsymbol{\mu}}_d+\boldsymbol{\mu}_d)^\top\bar{\Theta}\Sigma\bar{\Theta}(\hat{\boldsymbol{\mu}}_d-\boldsymbol{\mu}_d)$. Now we consider the terms Δ_i

successively.

(1) Consider the terms $\Delta_{0,1}$ and $\Delta_{0,2}$. We first show that $\Delta_{0,1} = [2(1-k\gamma_p)]^{-1}\boldsymbol{\mu}_d^\top \Theta \boldsymbol{\mu}_d(1+o(1))$.

Recall that $\widetilde{\Theta} \asymp (1-k\gamma_p)^{-1}\Theta$ by (S3.1). Let $\tilde{\boldsymbol{\mu}}_d = \boldsymbol{\mu}_d/\|\boldsymbol{\mu}_d\|_2$. Noting that $\Delta_{0,1} = \mathrm{tr}(\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top)\|\boldsymbol{\mu}_d\|_2^2/2$, where $\|\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top\|_* = 1$, by Definition 1, we have

$$\frac{1}{2}\mathrm{tr}(\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top) - \frac{1}{2(1-k\gamma_p)}\mathrm{tr}(\Theta\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top) \to_{a.s.} 0.$$

Then by the assumption $k\gamma_p < 1$, it follows that

$$\Delta_{0,1} = \|\boldsymbol{\mu}_d\|_2^2\left[\frac{1}{2(1-k\gamma_p)}\mathrm{tr}(\Theta\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top) + o(1)\right] = \frac{1}{2(1-k\gamma_p)}\boldsymbol{\mu}_d^\top\Theta\boldsymbol{\mu}_d(1+o(1)).$$

Let us consider $\Delta_{0,2}$, showing that $\Delta_{0,2} = (1-k\gamma_p)^{-2}\boldsymbol{\mu}_d^\top\Theta\boldsymbol{\mu}_d(1+o(1))$. Note that $\Delta_{0,2} = \mathrm{tr}(\widetilde{\Theta}\Sigma\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top)\|\boldsymbol{\mu}_d\|_2^2$. Then,

$$\begin{aligned}
&\mathrm{tr}(\widetilde{\Theta}\Sigma\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top) - \frac{1}{(1-k\gamma_p)^2}\,\mathrm{tr}(\Theta\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top)\\
=&\,\mathrm{tr}\left(\widetilde{\Theta}\Sigma\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top - \frac{1}{1-k\gamma_p}\Theta\Sigma\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top + \frac{1}{1-k\gamma_p}\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top - \frac{1}{(1-k\gamma_p)^2}\Theta\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top\right)\\
=&\,\mathrm{tr}\left[\left(\widetilde{\Theta} - \frac{1}{1-k\gamma_p}\Theta\right)\Sigma\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top\right] + \frac{1}{1-k\gamma_p}\,\mathrm{tr}\left[\left(\widetilde{\Theta} - \frac{1}{1-k\gamma_p}\Theta\right)\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top\right].
\end{aligned}$$

Denote $B_n = \Sigma\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top$, then $\|B_n\|_* = \tilde{\boldsymbol{\mu}}_d^\top\widetilde{\Theta}\Sigma\tilde{\boldsymbol{\mu}}_d$. Recall that $\widetilde{\Theta} \asymp (1-k\gamma_p)^{-1}\Theta$, which together with $\|\tilde{\boldsymbol{\mu}}_d\|_2 = 1$ and $\|\Sigma\|_* < c_5$ for some constant c_5 implies that

$$\|B_n\|_* \to_{a.s.} \frac{1}{1-k\gamma_p}\tilde{\boldsymbol{\mu}}_d^\top\tilde{\boldsymbol{\mu}}_d.$$

Since $\tilde{\boldsymbol{\mu}}_d^\top \tilde{\boldsymbol{\mu}}_d = \|\tilde{\boldsymbol{\mu}}_d\|_2^2 = 1$, we have $\sup_n \|B_n\|_* < \infty$. Therefore, by Definition 1, we obtain that

$$\operatorname{tr}\left[\left(\widetilde{\Theta} - \frac{1}{1-k\gamma_p}\Theta\right)\Sigma\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top\right] \to_{a.s.} 0.$$

Similarly, by $\|\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top\|_* = \|\tilde{\boldsymbol{\mu}}_d\|_2^2 = 1$, we have

$$\operatorname{tr}\left[\left(\widetilde{\Theta} - \frac{1}{1-k\gamma_p}\Theta\right)\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top\right] \to_{a.s.} 0.$$

Noting that $k\gamma_p \to c \in [0,1)$, it follows that

$$\operatorname{tr}(\widetilde{\Theta}\Sigma\widetilde{\Theta}\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top) - \frac{1}{(1-k\gamma_p)^2}\operatorname{tr}(\Theta\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top) \to_{a.s.} 0.$$

Thus,

$$\Delta_{0,2} = \|\boldsymbol{\mu}_d\|_2^2\left[\frac{1}{(1-k\gamma_p)^2}\operatorname{tr}(\Theta\tilde{\boldsymbol{\mu}}_d\tilde{\boldsymbol{\mu}}_d^\top) + o(1)\right] = \frac{1}{(1-k\gamma_p)^2}\boldsymbol{\mu}_d^\top\Theta\boldsymbol{\mu}_d(1+o(1)).$$

(2) We bound the terms $\Delta_1, \cdots, \Delta_6$.

(i) Bound $\Delta_1 = \boldsymbol{\mu}_d^\top(\bar{\Theta} - \widetilde{\Theta})\boldsymbol{\mu}_d$.

It is easy to see that

$$\widehat{\Sigma}_{two}^{(l)} = \widetilde{\Sigma}^{(l)} - \frac{n_1}{n}(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)^\top - \frac{n_2}{n}(\hat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2)(\hat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2)^\top.$$

Then we have

$$\begin{aligned}\|\widetilde{\Sigma}^{(l)} - \widehat{\Sigma}_{two}^{(l)}\|_2 &\le \frac{n_1}{n}\|(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)(\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1)^\top\|_2 + \frac{n_2}{n}\|(\hat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2)(\hat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2)^\top\|_2 \\ &= \frac{n_1}{n}\|\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1\|_2^2 + \frac{n_2}{n}\|\hat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2\|_2^2.\end{aligned} \tag{S3.4}$$

Denote $r_n = n_1/n \in (0,1)$, then $1 - r_n = n_2/n$. By Lemma 1, we have

$$\|\hat{\boldsymbol{\mu}}_1 - \boldsymbol{\mu}_1\|_2 = O_p\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{r_n n}}\right), \quad \|\hat{\boldsymbol{\mu}}_2 - \boldsymbol{\mu}_2\|_2 = O_p\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{(1-r_n)n}}\right).$$

Since the eigenvalues of Σ satisfy $0 < c_1 < \lambda_{\min}(\Sigma) < \lambda_{\max}(\Sigma) < c_2$, we have $\mathrm{tr}(\Sigma) \leq p\lambda_{\max}(\Sigma) < pc_2$. Combining this with (S3.4) and noting that $r_n \in (0,1)$, we see that

$$\|\widetilde{\Sigma}^{(l)} - \widehat{\Sigma}^{(l)}_{two}\|_2 = O_p\left(\frac{\mathrm{tr}(\Sigma)}{n}\right) = O_p\left(\frac{c_2 p}{n}\right) = O_p\left(\gamma_p\right) = o_p(1). \qquad \text{(S3.5)}$$

By Lemma 2, we have

$$\|(\widetilde{\Sigma}^{(l)})^{-1} - \Theta\|_2 = \|\Theta\|_2 O_p\left(\sqrt{k\gamma_p}\right). \qquad \text{(S3.6)}$$

Therefore, $(\widetilde{\Sigma}^{(l)})^{-1}$ can be bounded as

$$\|(\widetilde{\Sigma}^{(l)})^{-1}\|_2 \leq \|(\widetilde{\Sigma}^{(l)})^{-1} - \Theta\|_2 + \|\Theta\|_2 = \|\Theta\|_2\left[O_p\left(\sqrt{k\gamma_p}\right) + 1\right] = O_p(1),$$

where we use the fact that $\|\Theta\|_2$ is upper bounded according to (i) of (C1), and the fact $k\gamma_p \to c$. Combining with (S3.5) and reminding $\widehat{\Theta}^{(l)}_{two} = (\widehat{\Sigma}^{(l)}_{two})^{-1}$, it is easy to see that $\|\widehat{\Theta}^{(l)}_{two}\|_2 = O_p(1)$. Then noticing (S3.5) and the fact that $\widehat{\Theta}^{(l)}_{two} - (\widetilde{\Sigma}^{(l)})^{-1} = (\widetilde{\Sigma}^{(l)})^{-1}(\widetilde{\Sigma}^{(l)} - \widehat{\Sigma}^{(l)}_{two})\widehat{\Theta}^{(l)}_{two}$, we have

$$\|\widehat{\Theta}^{(l)}_{two} - (\widetilde{\Sigma}^{(l)})^{-1}\|_2 \leq \|(\widetilde{\Sigma}^{(l)})^{-1}\|_2\|\widehat{\Theta}^{(l)}_{two}\|_2\|\widetilde{\Sigma}^{(l)} - \widehat{\Sigma}^{(l)}_{two}\|_2 = O_p\left(\gamma_p\right), \qquad \text{(S3.7)}$$

for $1 \leq l \leq k$. Thus,

$$\|\bar{\Theta} - \widetilde{\Theta}\|_2 \leq \frac{1}{k}\sum_{l=1}^{k}\|\widehat{\Theta}^{(l)}_{two} - (\widetilde{\Sigma}^{(l)})^{-1}\|_2 = O_p\left(\gamma_p\right). \qquad \text{(S3.8)}$$

Then by (S3.8), we conclude that

$$\Delta_1 = \boldsymbol{\mu}_d^\top(\bar{\Theta} - \widetilde{\Theta})\boldsymbol{\mu}_d \leq \|\boldsymbol{\mu}_d\|_2^2\|\bar{\Theta} - \widetilde{\Theta}\|_2 = \|\boldsymbol{\mu}_d\|_2^2 O_p\left(\gamma_p\right).$$

(ii) Bound $\Delta_2 = \boldsymbol{\mu}_d^\top\bar{\Theta}(\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d)$.

Since $\hat{\boldsymbol{\mu}}_1 \sim N_p\left(\boldsymbol{\mu}_1, n_1^{-1}\Sigma\right)$, $\hat{\boldsymbol{\mu}}_2 \sim N_p\left(\boldsymbol{\mu}_2, n_2^{-1}\Sigma\right)$, and $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\mu}}_2$ are independent with each other, we have $\hat{\boldsymbol{\mu}}_d \sim N_p\left(\boldsymbol{\mu}_d, \left(n_1^{-1} + n_2^{-1}\right)\Sigma\right)$. By Lemma 1, we have

$$\|\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d\|_2 = O_p\left(\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathrm{tr}(\Sigma)}\right).$$

Recall that $n_1 = r_n n$, so $n_1^{-1} + n_2^{-1} = (r_n^{-1} + (1-r_n)^{-1})n^{-1}$. Since $\mathrm{tr}(\Sigma) \leq p\lambda_{\max}(\Sigma) < pc_2$ for some constant c_2, we have

$$\|\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d\|_2 = O_p\left(\sqrt{\left(\frac{1}{r_n} + \frac{1}{1-r_n}\right)c_2 \cdot \frac{p}{n}}\right) = O_p\left(\sqrt{\gamma_p}\right). \tag{S3.9}$$

Note that $\|\bar{\Theta} - \Theta\|_2 \leq k^{-1}\sum_{l=1}^k \|\widehat{\Theta}_{two}^{(l)} - \Theta\|_2$. Moreover, for $l = 1, \cdots, k$, it holds by (S3.6) and (S3.7) that

$$\|\widehat{\Theta}_{two}^{(l)} - \Theta\|_2 \leq \|\widehat{\Theta}_{two}^{(l)} - (\widetilde{\Sigma}^{(l)})^{-1}\|_2 + \|(\widetilde{\Sigma}^{(l)})^{-1} - \Theta\|_2 = O_p\left(\gamma_p\right) + O_p\left(\sqrt{k\gamma_p}\right).$$

Thus,

$$\|\bar{\Theta} - \Theta\|_2 = O_p\left(\gamma_p\right) + O_p\left(\sqrt{k\gamma_p}\right) = O_p(1),$$

where we use the fact $\gamma_p \to 0$ and $k\gamma_p \to c$, according to (C1) and (C2). Then we have

$$\|\bar{\Theta}\|_2 \leq \|\bar{\Theta} - \Theta\|_2 + \|\Theta\|_2 = O_p(1). \tag{S3.10}$$

By Cauchy–Schwarz inequality, it is easy to see that, for any symmetric matrix A and vectors $\boldsymbol{x}, \boldsymbol{y}$ with compatible dimensions, we have $|\boldsymbol{x}^\top A \boldsymbol{y}| \leq (\boldsymbol{x}^\top A \boldsymbol{x})^{1/2} (\boldsymbol{y}^\top A \boldsymbol{y})^{1/2} \leq \|\boldsymbol{x}\|_2 \|A\|_2 \|\boldsymbol{y}\|_2$. Hence, we obtain that

$$\Delta_2 = \boldsymbol{\mu}_d^\top \bar{\Theta} (\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d) \leq \|\boldsymbol{\mu}_d\|_2 \|\bar{\Theta}\|_2 \|\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d\|_2 = \|\boldsymbol{\mu}_d\|_2 O_p \left(\sqrt{\gamma_p}\right)$$

by (S3.9) and (S3.10).

(iii) Bound Δ_3. Since $\hat{\boldsymbol{\mu}}_a \sim N_p \left(\boldsymbol{\mu}_a, ((4n_1)^{-1} + (4n_2)^{-1}) \, \Sigma\right)$, we have

$$\|\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_a\|_2 = O_p \left(\sqrt{ \left(\frac{1}{4n_1} + \frac{1}{4n_2} \right) \operatorname{tr}(\Sigma) } \right)$$

by Markov's inequality. Similarly to (S3.9), we have

$$\|\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_a\|_2 = O_p \left(\sqrt{\gamma_p}\right). \tag{S3.11}$$

By (S3.10) and (S3.11), it follows that

$$\Delta_3 = (\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_a)^\top \bar{\Theta} \boldsymbol{\mu}_d \leq \|\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_a\|_2 \|\bar{\Theta}\|_2 \|\boldsymbol{\mu}_d\|_2 = \|\boldsymbol{\mu}_d\|_2 O_p \left(\sqrt{\gamma_p}\right).$$

(iv) Bound Δ_4, Δ_5 and Δ_6. Since $\|\Sigma\|_2$ and $\|\widetilde{\Theta}\|_2$ are bounded away from zero and infinity, by (S3.8) we have

$$\Delta_4 = \boldsymbol{\mu}_d^\top (\bar{\Theta} - \widetilde{\Theta}) \Sigma \widetilde{\Theta} \boldsymbol{\mu}_d \leq \|\boldsymbol{\mu}_d\|_2^2 \|\bar{\Theta} - \widetilde{\Theta}\|_2 \|\Sigma\|_2 \|\widetilde{\Theta}\|_2 = \|\boldsymbol{\mu}_d\|_2^2 O_p \left(\gamma_p\right).$$

Similarly, since $\|\Sigma\|_2$ is bounded away from zero and infinity, by (S3.8) we have

$$\Delta_5 = \boldsymbol{\mu}_d^\top (\bar{\Theta} - \widetilde{\Theta}) \Sigma (\bar{\Theta} - \widetilde{\Theta}) \boldsymbol{\mu}_d \leq \|\boldsymbol{\mu}_d\|_2^2 \|\bar{\Theta} - \widetilde{\Theta}\|_2^2 \|\Sigma\|_2 = \|\boldsymbol{\mu}_d\|_2^2 O_p \left(\gamma_p^2\right).$$

In addition, by (S3.9) we have

$$\|\hat{\boldsymbol{\mu}}_d + \boldsymbol{\mu}_d\|_2 \leq \|\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d\|_2 + 2\|\boldsymbol{\mu}_d\|_2 = O_p\left(\sqrt{\gamma_p}\right) + \|\boldsymbol{\mu}_d\|_2 O(1). \tag{S3.12}$$

This together with (S3.9) and (S3.10) yields that

$$\begin{aligned}\Delta_6 =& (\hat{\boldsymbol{\mu}}_d + \boldsymbol{\mu}_d)^\top \bar{\Theta}\Sigma\bar{\Theta}(\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d) \leq \|\hat{\boldsymbol{\mu}}_d + \boldsymbol{\mu}_d\|_2 \|\bar{\Theta}\|_2^2 \|\Sigma\|_2 \|\hat{\boldsymbol{\mu}}_d - \boldsymbol{\mu}_d\|_2 \\ =& O_p\left(\gamma_p\right) + \|\boldsymbol{\mu}_d\|_2 O_p\left(\sqrt{\gamma_p}\right).\end{aligned}$$

(3) We prove the final conclusion. By (S3.2), (S3.3) and limits of Δ_i, we have

$$(\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_2)^\top \bar{\Theta}\hat{\boldsymbol{\mu}}_d = \frac{1}{2(1-k\gamma_p)}\boldsymbol{\mu}_d^\top \Theta \boldsymbol{\mu}_d \left(1 + o(1) + O_p\left(\gamma_p\right) + O_p\left(\sqrt{\gamma_p}\right)\right),$$

$$\bar{\Delta}_p^2 = \frac{1}{(1-k\gamma_p)^2}\boldsymbol{\mu}_d^\top \Theta \boldsymbol{\mu}_d \left(1 + o(1) + O_p\left(\gamma_p^2\right) + O_p\left(\gamma_p\right) + O_p\left(\sqrt{\gamma_p}\right)\right) + O_p\left(\gamma_p\right).$$

According to the condition (ii) of (C1), we have $\gamma_p \to 0$. Note that $r_n \to \pi_1$, $1 - r_n \to \pi_2$, thus as $n \to \infty$,

$$\Phi\left(\frac{(\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_2)^\top \bar{\Theta}\hat{\boldsymbol{\mu}}_d + \log(\pi_2/\pi_1)}{\bar{\Delta}_p}\right) \to_p \Phi\left(\frac{\delta}{2} + \frac{(1-k\gamma_p)\log(\pi_2/\pi_1)}{\delta}\right), \tag{S3.13}$$

where $\delta^2 = \boldsymbol{\mu}_d^\top \Theta \boldsymbol{\mu}_d$. Similarly, we also have

$$\Phi\left(-\frac{(\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_1)^\top \bar{\Theta}\hat{\boldsymbol{\mu}}_d + \log(\pi_2/\pi_1)}{\bar{\Delta}_p}\right) \to_p \Phi\left(\frac{\delta}{2} - \frac{(1-k\gamma_p)\log(\pi_2/\pi_1)}{\delta}\right). \tag{S3.14}$$

According to the condition (ii) of (C2), we have $k\gamma_p \to c$. Combining (S3.13) and (S3.14) shows that $\hat{A}_{two}$ converges to A_{two} in probability as $n \to \infty$.

Now we show the convergence of $\hat{A}_{cen}$. Taking $k = 1$ yields that $\bar{\Theta} = \widehat{\Theta}$ and $\bar{\Delta}_p = \hat{\Delta}_p$. Thus,

$$\Phi\left(\frac{(\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_2)^\top \widehat{\Theta} \hat{\boldsymbol{\mu}}_d + \log(\pi_2/\pi_1)}{\hat{\Delta}_p}\right) \rightarrow_p \Phi\left(\frac{\delta}{2} + \frac{(1-\gamma_p)\log(\pi_2/\pi_1)}{\delta}\right),$$

$$\Phi\left(-\frac{(\hat{\boldsymbol{\mu}}_a - \boldsymbol{\mu}_1)^\top \widehat{\Theta} \hat{\boldsymbol{\mu}}_d + \log(\pi_2/\pi_1)}{\hat{\Delta}_p}\right) \rightarrow_p \Phi\left(\frac{\delta}{2} - \frac{(1-\gamma_p)\log(\pi_2/\pi_1)}{\delta}\right).$$

Recall that $\gamma_p \to 0$, so we obtain that $\hat{A}_{cen}$ converges to A_{cen} in probability.

Therefore, we conclude that $\hat{R}_{two}$ converges to A_{two}/A_{cen} in probability as $n \to \infty$. This completes the proof. □

S4. Proof of Theorem 2

Proof. As for $\hat{A}_{one}$, we need to show

$$\begin{aligned}&\Phi\left(\frac{n^{-1}\sum_{l=1}^{k} n^{(l)}(\hat{\boldsymbol{\mu}}_a^{(l)} - \boldsymbol{\mu}_2)^\top \widehat{\Theta}^{(l)} \hat{\boldsymbol{\mu}}_d^{(l)} + \log(\pi_2/\pi_1)}{\sqrt{(n^{-1}\sum_{l=1}^{k} n^{(l)}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)})^\top \Sigma (n^{-1}\sum_{l=1}^{k} n^{(l)}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)})}}\right) \\ &\rightarrow_p \Phi\left(\frac{\delta}{2} + \frac{(1-k\gamma_p)\log(\pi_2/\pi_1)}{\delta}\right).\end{aligned} \tag{S4.1}$$

The proof is similar to that of Theorem 1. To begin with, we decompose $n^{-1}\sum_{l=1}^{k} n^{(l)}(\hat{\boldsymbol{\mu}}_a^{(l)} - \boldsymbol{\mu}_2)^\top \widehat{\Theta}^{(l)} \hat{\boldsymbol{\mu}}_d^{(l)}$ and $(n^{-1}\sum_{l=1}^{k} n^{(l)}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)})^\top \Sigma (n^{-1}\sum_{l=1}^{k} n^{(l)}$

$\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)}$) as follows. Recalling that $n^{(l)} \equiv n/k$, it holds that

$$
\begin{aligned}
&\frac{1}{n}\sum_{l=1}^{k} n^{(l)}(\hat{\boldsymbol{\mu}}_a^{(l)} - \boldsymbol{\mu}_2)^\top \widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)} \\
=&\frac{1}{k}\sum_{l=1}^{k}[(\boldsymbol{\mu}_a - \boldsymbol{\mu}_2) + (\hat{\boldsymbol{\mu}}_a^{(l)} - \boldsymbol{\mu}_a)]^\top \widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)} \\
=&\frac{1}{k}\sum_{l=1}^{k}(\boldsymbol{\mu}_a - \boldsymbol{\mu}_2)^\top \widehat{\Theta}^{(l)}[\boldsymbol{\mu}_d + (\hat{\boldsymbol{\mu}}_d^{(l)} - \boldsymbol{\mu}_d)] + \frac{1}{k}\sum_{l=1}^{k}(\hat{\boldsymbol{\mu}}_a^{(l)} - \boldsymbol{\mu}_a)^\top \widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)} \\
=&\frac{1}{2k}\sum_{l=1}^{k}\boldsymbol{\mu}_d^\top \widehat{\Theta}^{(l)}\boldsymbol{\mu}_d + \frac{1}{2k}\sum_{l=1}^{k}\boldsymbol{\mu}_d^\top \widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)} - \boldsymbol{\mu}_d) + \frac{1}{k}\sum_{l=1}^{k}(\hat{\boldsymbol{\mu}}_a^{(l)} - \boldsymbol{\mu}_a)^\top \widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)} \\
=&\frac{1}{2}\boldsymbol{\mu}_d^\top \widetilde{\Theta}\boldsymbol{\mu}_d + \frac{1}{2k}\sum_{l=1}^{k}\boldsymbol{\mu}_d^\top(\widehat{\Theta}^{(l)} - \widetilde{\Theta})\boldsymbol{\mu}_d + \frac{1}{2k}\sum_{l=1}^{k}\boldsymbol{\mu}_d^\top \widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)} - \boldsymbol{\mu}_d) \\
+&\frac{1}{k}\sum_{l=1}^{k}(\hat{\boldsymbol{\mu}}_a^{(l)} - \boldsymbol{\mu}_a)^\top \widehat{\Theta}^{(l)}\boldsymbol{\mu}_d \\
=&\frac{1}{2}\boldsymbol{\mu}_d^\top \widetilde{\Theta}\boldsymbol{\mu}_d + \frac{1}{2}\boldsymbol{\mu}_d^\top(\bar{\Theta}_{one} - \widetilde{\Theta})\boldsymbol{\mu}_d + \boldsymbol{\mu}_d^\top\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1)\right) \\
=&\Delta_{0,1} + \frac{1}{2}\Delta_1^{(2)} + \Delta_7,
\end{aligned} \tag{S4.2}
$$

where $\Delta_{0,1}$ is defined in the proof of Theorem 1, $\Delta_1^{(2)} = \boldsymbol{\mu}_d^\top(\bar{\Theta}_{one} - \widetilde{\Theta})\boldsymbol{\mu}_d$, $\Delta_7 = \boldsymbol{\mu}_d^\top[k^{-1}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1)]$, and $\bar{\Theta}_{one} = k^{-1}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}$. Recall that $\widetilde{\Theta} = k^{-1}\sum_{l=1}^{k}(\widetilde{\Sigma}^{(l)})^{-1}$ is the average of local inverse sample covariances with known $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. In addition, due to $n^{(l)} \equiv n/k$, it holds that

$$
\begin{aligned}
&\left(\frac{1}{n}\sum_{l=1}^{k} n^{(l)}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)}\right)^\top \Sigma \left(\frac{1}{n}\sum_{l=1}^{k} n^{(l)}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)}\right) \\
=&\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}[\boldsymbol{\mu}_d + (\hat{\boldsymbol{\mu}}_d^{(l)} - \boldsymbol{\mu}_d)]\right)^\top \Sigma \left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)}\right).
\end{aligned}
$$

Similarly to the decomposition of (S3.3), the above equation is equal to

$$\begin{aligned}
&\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\boldsymbol{\mu}_d\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}[\boldsymbol{\mu}_d+(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)]\right)\\
&+\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)}\right)\\
&=\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\boldsymbol{\mu}_d\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\boldsymbol{\mu}_d\right)\\
&+\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\boldsymbol{\mu}_d\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\right)\\
&+\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)}\right)\\
&=\left(\frac{1}{k}\sum_{l=1}^{k}[\widetilde{\Theta}+(\widehat{\Theta}^{(l)}-\widetilde{\Theta})]\boldsymbol{\mu}_d\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}[\widetilde{\Theta}+(\widehat{\Theta}^{(l)}-\widetilde{\Theta})]\boldsymbol{\mu}_d\right)\\
&+\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}+\boldsymbol{\mu}_d)\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\right)\\
&=\boldsymbol{\mu}_d^{\top}\widetilde{\Theta}\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d+2\boldsymbol{\mu}_d^{\top}\left(\frac{1}{k}\sum_{l=1}^{k}(\widehat{\Theta}^{(l)}-\widetilde{\Theta})\right)\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d\\
&+\boldsymbol{\mu}_d^{\top}\left(\frac{1}{k}\sum_{l=1}^{k}(\widehat{\Theta}^{(l)}-\widetilde{\Theta})\right)\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}(\widehat{\Theta}^{(l)}-\widetilde{\Theta})\right)\boldsymbol{\mu}_d\\
&+\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}+\boldsymbol{\mu}_d)\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\right)\\
&=\boldsymbol{\mu}_d^{\top}\widetilde{\Theta}\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d+2\boldsymbol{\mu}_d^{\top}(\bar{\Theta}_{one}-\widetilde{\Theta})\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d+\boldsymbol{\mu}_d^{\top}(\bar{\Theta}_{one}-\widetilde{\Theta})\Sigma(\bar{\Theta}_{one}-\widetilde{\Theta})\boldsymbol{\mu}_d\\
&+\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}+\boldsymbol{\mu}_d)\right)^{\top}\Sigma\left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\right)\\
&=\Delta_{0,2}+2\Delta_4^{(2)}+\Delta_5^{(2)}+\Delta_8,
\end{aligned}\tag{S4.3}$$

where $\Delta_{0,2}$ is defined in the proof of Theorem 1, $\Delta_4^{(2)} = \boldsymbol{\mu}_d^\top(\bar{\Theta}_{one} - \widetilde{\Theta})\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d$, $\Delta_5^{(2)} = \boldsymbol{\mu}_d^\top(\bar{\Theta}_{one} - \widetilde{\Theta})\Sigma(\bar{\Theta}_{one} - \widetilde{\Theta})\boldsymbol{\mu}_d$, and $\Delta_8 = [k^{-1}\sum_{l=1}^k \widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)} + \boldsymbol{\mu}_d)]^\top\Sigma$ $[k^{-1}\sum_{l=1}^k \widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)} - \boldsymbol{\mu}_d)]$.

Remind that the terms $\Delta_{0,1}$ and $\Delta_{0,2}$ have been studied in the proof of Theorem 1. Then we bound the terms $\Delta_i^{(2)}$, $i = 1, 4, 5$, Δ_7, and Δ_8.

(i) Bound $\Delta_1^{(2)}, \Delta_4^{(2)}, \Delta_5^{(2)}$. The procedure is similar to that of bounding $\Delta_1, \Delta_4, \Delta_5$, where the only difference is that $\bar{\Theta}$ in Δ_i is replaced by $\bar{\Theta}_{one}$ in $\Delta_i^{(2)}$.

Noting the fact

$$\widehat{\Sigma}^{(l)} = \widetilde{\Sigma}^{(l)} - \frac{n_1}{n}(\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1)(\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1)^\top - \frac{n_2}{n}(\hat{\boldsymbol{\mu}}_2^{(l)} - \boldsymbol{\mu}_2)(\hat{\boldsymbol{\mu}}_2^{(l)} - \boldsymbol{\mu}_2)^\top,$$

it follows that

$$\begin{aligned}\|\widetilde{\Sigma}^{(l)} - \widehat{\Sigma}^{(l)}\|_2 &\le \frac{n_1}{n}\|(\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1)(\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1)^\top\|_2 + \frac{n_2}{n}\|(\hat{\boldsymbol{\mu}}_2^{(l)} - \boldsymbol{\mu}_2)(\hat{\boldsymbol{\mu}}_2^{(l)} - \boldsymbol{\mu}_2)^\top\|_2 \\ &= \frac{n_1}{n}\|\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1\|_2^2 + \frac{n_2}{n}\|\hat{\boldsymbol{\mu}}_2^{(l)} - \boldsymbol{\mu}_2\|_2^2.\end{aligned}$$

According to Lemma 1, it holds that

$$\|\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1\|_2 = O_p\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{r_n n^{(l)}}}\right), \quad \|\hat{\boldsymbol{\mu}}_2^{(l)} - \boldsymbol{\mu}_2\|_2 = O_p\left(\sqrt{\frac{\mathrm{tr}(\Sigma)}{(1-r_n)n^{(l)}}}\right).$$

Since $\mathrm{tr}(\Sigma) < pc_2$ and $r_n \in (0,1)$, we have

$$\|\widetilde{\Sigma}^{(l)} - \widehat{\Sigma}^{(l)}\|_2 = O_p\left(\frac{\mathrm{tr}(\Sigma)}{n^{(l)}}\right) = O_p\left(\frac{c_2 p}{n^{(l)}}\right) = O_p\left(\gamma_p^{(l)}\right) = o_p(1), \qquad \text{(S4.4)}$$

where we use the assumption $\gamma_p^{(l)} = o(1)$ in (C2). Remind that $\|(\widetilde{\Sigma}^{(l)})^{-1}\|_2 = O_p(1)$. Due to (S4.4) and the fact $\widehat{\Theta}^{(l)} = (\widehat{\Sigma}^{(l)})^{-1}$, it is easy to see that

$\|\widehat{\Theta}^{(l)}\|_2 = O_p(1)$. Then by $\widehat{\Theta}^{(l)} - (\widetilde{\Sigma}^{(l)})^{-1} = (\widetilde{\Sigma}^{(l)})^{-1}(\widetilde{\Sigma}^{(l)} - \widehat{\Sigma}^{(l)})\widehat{\Theta}^{(l)}$ and (S4.4), we have

$$\|\widehat{\Theta}^{(l)} - (\widetilde{\Sigma}^{(l)})^{-1}\|_2 \leq \|(\widetilde{\Sigma}^{(l)})^{-1}\|_2\|\widehat{\Theta}^{(l)}\|_2\|\widetilde{\Sigma}^{(l)} - \widehat{\Sigma}^{(l)}\|_2 = O_p\left(\gamma_p^{(l)}\right). \quad \text{(S4.5)}$$

Thus,

$$\|\bar{\Theta}_{one} - \widetilde{\Theta}\|_2 \leq \frac{1}{k}\sum_{l=1}^{k}\|\widehat{\Theta}^{(l)} - (\widetilde{\Sigma}^{(l)})^{-1}\|_2 = O_p\left(\gamma_p^{(l)}\right). \quad \text{(S4.6)}$$

By (S4.6), we conclude that

$$\Delta_1^{(2)} = \boldsymbol{\mu}_d^\top(\bar{\Theta}_{one} - \widetilde{\Theta})\boldsymbol{\mu}_d \leq \|\boldsymbol{\mu}_d\|_2^2\|\bar{\Theta}_{one} - \widetilde{\Theta}\|_2 = \|\boldsymbol{\mu}_d\|_2^2 O_p\left(\gamma_p^{(l)}\right).$$

Since $\|\Sigma\|_2$ and $\|\widetilde{\Theta}\|_2$ are bounded away from zero and infinity, by (S4.6), we have

$$\Delta_4^{(2)} = \boldsymbol{\mu}_d^\top(\bar{\Theta}_{one} - \widetilde{\Theta})\Sigma\widetilde{\Theta}\boldsymbol{\mu}_d \leq \|\boldsymbol{\mu}_d\|_2^2\|\bar{\Theta}_{one} - \widetilde{\Theta}\|_2\|\Sigma\|_2\|\widetilde{\Theta}\|_2 = \|\boldsymbol{\mu}_d\|_2^2 O_p\left(\gamma_p^{(l)}\right),$$

$$\Delta_5^{(2)} = \boldsymbol{\mu}_d^\top(\bar{\Theta}_{one} - \widetilde{\Theta})\Sigma(\bar{\Theta} - \widetilde{\Theta})\boldsymbol{\mu}_d \leq \|\boldsymbol{\mu}_d\|_2^2\|\bar{\Theta}_{one} - \widetilde{\Theta}\|_2^2\|\Sigma\|_2 = \|\boldsymbol{\mu}_d\|_2^2 O_p\left((\gamma_p^{(l)})^2\right).$$

(ii) Bound Δ_7. Similarly to (S3.9), we have

$$\|\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1\|_2 = O_p\left(\sqrt{\gamma_p^{(l)}}\right).$$

Recalling that $\|\widehat{\Theta}^{(l)}\|_2 = O_p(1)$, it follows that

$$\begin{aligned}\left\|\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1)\right\|_2 &\leq \frac{1}{k}\sum_{l=1}^{k}\|\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1)\|_2 \\ &\leq \frac{1}{k}\sum_{l=1}^{k}\|\widehat{\Theta}^{(l)}\|_2\|\hat{\boldsymbol{\mu}}_1^{(l)} - \boldsymbol{\mu}_1\|_2 = O_p\left(\sqrt{\gamma_p^{(l)}}\right).\end{aligned} \quad \text{(S4.7)}$$

By (S4.7), we obtain that

$$\begin{aligned}\Delta_7 =&\boldsymbol{\mu}_d^\top \left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_1^{(l)}-\boldsymbol{\mu}_1)\right)\\ \le&\|\boldsymbol{\mu}_d\|_2\left\|\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_1^{(l)}-\boldsymbol{\mu}_1)\right\|_2 = \|\boldsymbol{\mu}_d\|_2 O_p\left(\sqrt{\gamma_p^{(l)}}\right).\end{aligned}$$

(iii) Bound Δ_8. Similarly to (S3.9) and (S3.12), we have

$$\|\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d\|_2 = O_p\left(\sqrt{\gamma_p^{(l)}}\right), \quad \|\hat{\boldsymbol{\mu}}_d^{(l)}+\boldsymbol{\mu}_d\|_2 = O_p\left(\sqrt{\gamma_p^{(l)}}\right)+\|\boldsymbol{\mu}_d\|_2 O(1).$$

Therefore, by $\|\widehat{\Theta}^{(l)}\|_2 = O_p(1)$, we have

$$\|\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\|_2 \le \|\widehat{\Theta}^{(l)}\|_2\|\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d\|_2 = O_p\left(\sqrt{\gamma_p^{(l)}}\right)$$

and

$$\|\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}+\boldsymbol{\mu}_d)\|_2 \le \|\widehat{\Theta}^{(l)}\|_2\|\hat{\boldsymbol{\mu}}_d^{(l)}+\boldsymbol{\mu}_d\|_2 = O_p\left(\sqrt{\gamma_p^{(l)}}\right)+\|\boldsymbol{\mu}_d\|_2 O(1).$$

Consequently, we conclude that

$$\begin{aligned}\Delta_8 =& \left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}+\boldsymbol{\mu}_d)\right)^\top \Sigma \left(\frac{1}{k}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\right)\\ \le& \left(\frac{1}{k}\sum_{l=1}^{k}\|\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}+\boldsymbol{\mu}_d)\|_2\right)\|\Sigma\|_2\left(\frac{1}{k}\sum_{l=1}^{k}\|\widehat{\Theta}^{(l)}(\hat{\boldsymbol{\mu}}_d^{(l)}-\boldsymbol{\mu}_d)\|_2\right)\\ =&O_p\left(\gamma_p^{(l)}\right)+\|\hat{\boldsymbol{\mu}}_d\|_2 O_p\left(\sqrt{\gamma_p^{(l)}}\right).\end{aligned}$$

Remind that $\gamma_p^{(l)} = k\gamma_p \to c = 0$ in Theorem 2, which implies that these remaining terms (i.e., $\Delta_i^{(2)}$, $i = 1,4,5$, Δ_7, Δ_8) are lower order terms of $\Delta_{0,1}$ and $\Delta_{0,2}$. This completes the proof of (S4.1). Similarly, we also have

$$\Phi\left(-\frac{k^{-1}\sum_{l=1}^{k}(\hat{\boldsymbol{\mu}}_a^{(l)}-\boldsymbol{\mu}_1)^\top\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)}+\log(\pi_2/\pi_1)}{\sqrt{(k^{-1}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)})^\top\Sigma(k^{-1}\sum_{l=1}^{k}\widehat{\Theta}^{(l)}\hat{\boldsymbol{\mu}}_d^{(l)})}}\right)$$
$$\to_p \Phi\left(\frac{\delta}{2}-\frac{(1-k\gamma_p)\log(\pi_2/\pi_1)}{\delta}\right).$$

Since $k\gamma_p \to 0$, $\hat{A}_{one}$ converges to

$$A_{cen}=\pi_1\Phi\left(\frac{\delta}{2}-\frac{\log(\pi_2/\pi_1)}{\delta}\right)+\pi_2\Phi\left(\frac{\delta}{2}+\frac{\log(\pi_2/\pi_1)}{\delta}\right)$$

in probability as $n\to\infty$, and $\hat{A}_{cen}$ also has the same limit. Therefore, $\hat{R}_{one}$ converges to one in probability as $n\to\infty$. This completes the proof. □

References

Dobriban, E. and Y. Sheng (2018). Distributed linear regression by averaging. *ArXiv preprint arXiv:1810.00412*.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press.