Supplement of "An Efficient Greedy Search Algorithm for High-dimensional Linear Discriminant Analysis"

Hannan Yang, Danyu Lin and Quefeng Li

Department of Biostatistics, University of North Carolina, Chapel Hill

Supplementary Material

The online supplementary material contains the proofs of Theorems 1–3, equation (2.3) and the supporting lemmas.

S1 Proofs

Proof of Theorem 1. It follows from Lemmas 1 and 4 that

$$P\left(|(\widehat{\sigma}_{cc} - \widehat{\mathbf{\Sigma}}_{Sc}^T \widehat{\mathbf{\Sigma}}_{SS}^{-1} \widehat{\mathbf{\Sigma}}_{Sc}) - (\sigma_{cc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc})| \lesssim s^2 \sqrt{(\log p)/n}\right) \geq 1 - C_A p^{-C_B},$$

where C_A only depends on C_1 and C_4 in Lemmas 1 and 4, and C_B is an arbitrarily large constant. Since $\mathbf{\Sigma}_{S\cup\{c\},S\cup\{c\}}$ is a submatrix of $\mathbf{\Sigma}$ with row and column indices in $S \cup \{c\}$ and is positive definite, it follows from Condition 2 and Theorem 4.3.17 of Horn and Johnson (2012) that for any

$c \notin S$,

$$0 < m \leq \lambda_{\min}(\mathbf{\Sigma}_{S\cup\{c\},S\cup\{c\}}) \leq \lambda_{\max}(\mathbf{\Sigma}_{S\cup\{c\},S\cup\{c\}}) \leq M < \infty.$$

Since $\sigma_{cc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc}$ is the Schur complement of $\mathbf{\Sigma}_S$ in $\mathbf{\Sigma}_{S\cup\{c\},S\cup\{c\}}$, it follows that $\sigma_{cc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc} \geq m > 0$ for all $c \notin S$. Then we have

$$P\left(|(\widehat{\sigma}_{cc} - \widehat{\mathbf{\Sigma}}_{Sc}^T \widehat{\mathbf{\Sigma}}_{SS}^{-1} \widehat{\mathbf{\Sigma}}_{Sc})^{-1} - (\sigma_{cc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc})^{-1}| \lesssim s^2 \sqrt{(\log p)/n}\right) \geq 1 - C_A p^{-C_B}, \tag{S1.1}$$

where C_A only depends on C_1 and C_4, and C_B is an arbitrarily large constant. On the other hand, with probability at least $1 - C_A p^{-C_B}$, we have

$$\begin{aligned}
&|(\widehat{\delta}_c - \widehat{\mathbf{\Sigma}}_{Sc}^T \widehat{\mathbf{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\delta}}_S)^2 - (\delta_c - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S)^2| \\
&\leq |(\widehat{\delta}_c - \widehat{\mathbf{\Sigma}}_{Sc}^T \widehat{\mathbf{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\delta}}_S) - (\delta_c - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_{SS})|^2 \\
&\quad + 2|(\widehat{\delta}_c - \widehat{\mathbf{\Sigma}}_{Sc}^T \widehat{\mathbf{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\delta}}_S) - (\delta_c - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S)| \cdot |\delta_c - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S| \qquad \text{(S1.2)} \\
&\lesssim (s^2 \sqrt{(\log p)/n})^2 + (s^2 \sqrt{(\log p)/n}) \cdot |\delta_c - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S| \\
&\lesssim (s^2 \sqrt{(\log p)/n}) \cdot \max(s^2 \sqrt{(\log p)/n}, \sqrt{\theta_{Sc}}),
\end{aligned}$$

where the last inequality follows from Condition 2.

Therefore, (S1.1) and (S1.2) together imply that, with probability at least $1 - C_A p^{-C_B}$, we have

$$
\begin{aligned}
&|\widehat{\theta}_{Sc} - \theta_{Sc}| \\
&= |(\widehat{\delta}_c - \widehat{\boldsymbol{\Sigma}}_{Sc}^T \widehat{\boldsymbol{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\delta}}_S)^2 (\widehat{\sigma}_{cc} - \widehat{\boldsymbol{\Sigma}}_{Sc}^T \widehat{\boldsymbol{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\Sigma}}_{Sc})^{-1} - (\delta_c - \boldsymbol{\Sigma}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S)^2 (\sigma_{cc} - \boldsymbol{\Sigma}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1} \boldsymbol{\Sigma}_{Sc})^{-1}| \\
&\leq |(\widehat{\delta}_c - \widehat{\boldsymbol{\Sigma}}_{Sc}^T \widehat{\boldsymbol{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\delta}}_S)^2 - (\delta_c - \boldsymbol{\Sigma}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S)^2| \cdot |(\widehat{\sigma}_{cc} - \widehat{\boldsymbol{\Sigma}}_{Sc}^T \widehat{\boldsymbol{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\Sigma}}_{Sc})^{-1} - (\sigma_{cc} - \boldsymbol{\Sigma}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1} \boldsymbol{\Sigma}_{Sc})^{-1}| \\
&\quad + |(\widehat{\delta}_c - \widehat{\boldsymbol{\Sigma}}_{Sc}^T \widehat{\boldsymbol{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\delta}}_S)^2 - (\delta_c - \boldsymbol{\Sigma}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S)^2| (\sigma_{cc} - \boldsymbol{\Sigma}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1} \boldsymbol{\Sigma}_{Sc})^{-1} \\
&\quad + |(\widehat{\sigma}_{cc} - \widehat{\boldsymbol{\Sigma}}_{Sc}^T \widehat{\boldsymbol{\Sigma}}_{SS}^{-1} \widehat{\boldsymbol{\Sigma}}_{Sc})^{-1} - (\sigma_{cc} - \boldsymbol{\Sigma}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1} \boldsymbol{\Sigma}_{Sc})^{-1}| (\delta_c - \boldsymbol{\Sigma}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S)^2 \\
&\lesssim s^4 (\log p)/n \max(s^2 \sqrt{(\log p)/n}, \sqrt{\theta_{Sc}}) + s^2 \sqrt{(\log p)/n} \max(s^2 \sqrt{(\log p/n)}, \sqrt{\theta_{Sc}}) \\
&\quad + s^2 \sqrt{(\log p)/n} \theta_{Sc} \\
&\lesssim s^2 \sqrt{(\log p)/n} \max(s^2 \sqrt{(\log p)/n}, \sqrt{\theta_{Sc}}, \theta_{Sc}).
\end{aligned}
$$

□

Proof of Theorem 2. Let $\emptyset = \widehat{S}_0 \subset \widehat{S}_1 \subset \cdots$ be the sequence of selected indices given by the greedy search algorithm. The key of the proof is to show that, with high probability, $\widehat{S}_k \subset \mathcal{M}$ for all $k \leq K-1$, and $\widehat{\mathcal{M}} = \widehat{S}_K = \mathcal{M}$.

When $k = 0$, it follows from Corollary 1 and the union bound that

$$P\left(\max_{c \leq p} |\widehat{\theta}_{Sc} - \theta_{Sc}| \lesssim \sqrt{(\log p)/n}\right) \geq 1 - C_A p^{-C_B}, \text{ for } S = \emptyset.$$

Condition 4 implies that $\max_{c \in \mathcal{M}} \theta_{Sc} - \max_{c \notin \mathcal{M}} \theta_{Sc} \gg K^2 \sqrt{(\log p)/n} \geq \sqrt{(\log p)/n}$. These two results together imply that

$$P\left(\max_{c \in \mathcal{M}} \widehat{\theta}_{Sc} > \max_{c \notin \mathcal{M}} \widehat{\theta}_{Sc}\right) \geq 1 - C_A p^{-C_B}, \text{ for } S = \emptyset.$$

It further implies that $P\left(\widehat{S}_1 \subset \mathcal{M}\right) \geq 1 - C_A p^{-C_B}$.

When $k = 1$, we prove that

$$P\left(\max_{c \in \mathcal{M} \setminus \widehat{S}_1} \widehat{\theta}_{\widehat{S}_1 c} > \max_{c \notin \mathcal{M}} \widehat{\theta}_{\widehat{S}_1 c}\right) \geq 1 - C_A p^{-C_B}. \tag{S1.3}$$

This further gives $P\left(\widehat{S}_2 \subset \mathcal{M}\right) \geq 1 - C_A p^{-C_B}$, where C_A is treated as a generic postic constant. Let events

$$\begin{aligned}
E_1 &= \left\{\widehat{S}_1 \subset \mathcal{M}\right\}, \\
A_1 &= \left\{\max_{c \in \mathcal{M} \setminus \widehat{S}_1} \theta_{\widehat{S}_1 c} - \max_{c \notin \mathcal{M}} \theta_{\widehat{S}_1 c} \gg K^2 \sqrt{(\log p)/n}\right\}, \\
A_2 &= \left\{\max_{c \in \mathcal{M} \setminus \widehat{S}_1} |\widehat{\theta}_{\widehat{S}_{1c}} - \theta_{\widehat{S}_1 c}| \lesssim K^2 \sqrt{(\log p)/n}\right\}, \\
A_3 &= \left\{\max_{c \notin \mathcal{M}} |\widehat{\theta}_{\widehat{S}_{1c}} - \theta_{\widehat{S}_1 c}| \lesssim K^2 \sqrt{(\log p)/n}\right\}.
\end{aligned}$$

Note that $A_1 \cap A_2 \cap A_3 \subset \left\{\max_{c \in \mathcal{M} \setminus \widehat{S}_1} \widehat{\theta}_{\widehat{S}_1 c} > \max_{c \notin \mathcal{M}} \widehat{\theta}_{\widehat{S}_1 c}\right\}$. Therefore,

$$P\left(\max_{c \in \mathcal{M} \setminus \widehat{S}_1} \widehat{\theta}_{\widehat{S}_1 c} > \max_{c \notin \mathcal{M}} \widehat{\theta}_{\widehat{S}_1 c}\right) \geq 1 - P\left(\overline{A_1}\right) - P\left(\overline{A_2}\right) - P\left(\overline{A_3}\right). \tag{S1.4}$$

Under Condition 4, $E_1 \subset A_1$, therefore, $P\left(\overline{A_1}\right) \leq P\left(\overline{E_1}\right) \leq C_A p^{-C_B}$. It follows from Theorem 1, Condition 3, and the union bound that $P(\overline{A_2}) \leq C_A p^{-C_B}$, and $P(\overline{A_3}) \leq C_1 p^{-C_B}$. These three results, together with (S1.4), proves (S1.3). By the same argument, it holds that $\widehat{S}_k \subset \mathcal{M}$ for all $k \leq K$ with probability at least $1-(2k-1)C_A p^{-C_B}$. Since $\mathcal{M}$ contains K elements, we further have $\widehat{S}_K = \mathcal{M}$.

Next, we show that at the $(K+1)$th iteration, the greedy search algorithm terminates with high probability if we choose $\tau \asymp K^4(\log p)/n$. First, we show that $\theta_{\mathcal{M}c} = 0$ for all $c \notin M$. By definition, $\theta_{\mathcal{M}c} = \Delta_{\mathcal{M}\cup\{c\}} - \Delta_{\mathcal{M}} = \boldsymbol{\beta}_{\mathcal{M}\cup\{c\}}^T \boldsymbol{\Sigma}_{\mathcal{M}\cup\{c\},\mathcal{M}\cup\{c\}} \boldsymbol{\beta}_{\mathcal{M}\cup\{c\}} - \boldsymbol{\beta}_{\mathcal{M}}^T \boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}} \boldsymbol{\beta}_{\mathcal{M}} = 0$. Then, Theorem 1 implies that

$$P\left(\max_{c\notin\mathcal{M}} |\widehat{\theta}_{\mathcal{M}c}| \leq K^4(\log p)/n\right) \geq 1 - C_A p^{-C_B}.$$

Hence, by choosing $\tau \asymp K^4(\log p)/n$, the greedy search program terminates with high probability, i.e., $P(\widehat{\mathcal{M}} = \widehat{S}_K | \widehat{S}_K = \mathcal{M}) \geq 1 - C_A p^{-C_B}$. Then,

$$\begin{aligned} P\left(\widehat{\mathcal{M}} = \mathcal{M}\right) &= P\left(\widehat{\mathcal{M}} = \widehat{S}_K, \widehat{S}_K = M\right) = P\left(\widehat{\mathcal{M}} = \widehat{S}_K | \widehat{S}_K = \mathcal{M}\right) P\left(\widehat{S}_K = \mathcal{M}\right) \\ &\geq (1 - C_A p^{-C_B})(1 - (2K-1)C_A p^{-C_B}) \geq 1 - C_A K p^{-C_B}. \end{aligned}$$

□

Proof of Theorem 3. We prove the result conditioning on the event that $\{\widehat{\mathcal{M}} = \mathcal{M}\}$, which holds with probability tending to 1. We first bound $\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T \widehat{\boldsymbol{\Omega}}_{\mathcal{M}} \widehat{\boldsymbol{\delta}}_{\mathcal{M}}$. By Lemma 3, we have

$$\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T \boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}^{-1} \widehat{\boldsymbol{\delta}}_{\mathcal{M}} - \boldsymbol{\delta}_{\mathcal{M}}^T \boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}^{-1} \boldsymbol{\delta}_{\mathcal{M}} = O_P\left(K\sqrt{(\log p)/n}\right).$$

By Condition 3, $K \lesssim \Delta_p = \boldsymbol{\delta}_{\mathcal{M}}^T \boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}^{-1} \boldsymbol{\delta}_{\mathcal{M}}$. Therefore,

$$\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T \boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}^{-1} \widehat{\boldsymbol{\delta}}_{\mathcal{M}} - \boldsymbol{\delta}_{\mathcal{M}}^T \boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}^{-1} \boldsymbol{\delta}_{\mathcal{M}} = O_P\left(\Delta_p\sqrt{(\log p)/n}\right). \tag{S1.5}$$

Then, by Lemma 4 we have

$$|\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T(\widehat{\boldsymbol{\Omega}}_{\mathcal{M}} - \boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}^{-1})\widehat{\boldsymbol{\delta}}_{\mathcal{M}}| = O_P\left(\Delta_p K\sqrt{(\log p)/n}\right). \tag{S1.6}$$

It follows from the triangular inequality and (S1.5) and (S1.6) that

$$\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}} = \Delta_p\left\{1 + O_P(K\sqrt{(\log p)/n})\right\}. \tag{S1.7}$$

Next, we bound $\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}$. It follows from Lemma 2 that $\|\widehat{\boldsymbol{\Omega}}_{\mathcal{M}} - \boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}^{-1}\| = O_P\left(K\sqrt{(\log p)/n}\right)$. This result, together with Condition 2, imply that $\|\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\| = O_P(1)$. Then, using the same argument as in the proof of Lemma 4, we have

$$\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T(\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}} - \widehat{\boldsymbol{\Omega}}_{\mathcal{M}})\widehat{\boldsymbol{\delta}}_{\mathcal{M}} = O_P\left(\Delta_p K\sqrt{(\log p)/n}\right).$$

This result, together with (S1.7), gives

$$\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}} = \Delta_p\left\{1 + O_P(K\sqrt{(\log p)/n})\right\}. \tag{S1.8}$$

Then, we have

$$\begin{aligned}\frac{\widehat{\boldsymbol{\beta}}_{\mathcal{M}}^T(\bar{\boldsymbol{x}}_{1\mathcal{M}} - \boldsymbol{\mu}_{1\mathcal{M}})}{\sqrt{\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}}} &= \frac{(\widehat{\boldsymbol{\beta}}_{\mathcal{M}} - \boldsymbol{\beta}_{\mathcal{M}})^T(\bar{\boldsymbol{x}}_{1\mathcal{M}} - \boldsymbol{\mu}_{1\mathcal{M}})}{\sqrt{\Delta_p\{1 + O_P(K\sqrt{(\log p)/n})\}}} \\ &\quad + \frac{\boldsymbol{\beta}_{\mathcal{M}}^T(\bar{\boldsymbol{x}}_{1\mathcal{M}} - \boldsymbol{\mu}_{1\mathcal{M}})}{\sqrt{\Delta_p\{1 + O_P(K\sqrt{(\log p)/n})\}}}.\end{aligned}$$

Since the leading term $\Delta_p^{-1/2}\boldsymbol{\beta}_{\mathcal{M}}^T(\bar{\boldsymbol{x}}_{1\mathcal{M}} - \boldsymbol{\mu}_{1\mathcal{M}}) \sim N(0, 1/n_1)$, we have

$$\frac{\boldsymbol{\beta}_{\mathcal{M}}^T(\bar{\boldsymbol{x}}_{1\mathcal{M}} - \boldsymbol{\mu}_{1\mathcal{M}})}{\sqrt{\Delta_p\{1 + O_P(K\sqrt{(\log p)/n})\}}} = \frac{O_P(1/\sqrt{n})}{\sqrt{1 + O_P(K\sqrt{(\log p)/n})}}.$$

Since $K\sqrt{(\log p)/n} \leq K^2\sqrt{(\log p)/n} = o(1)$, the leading term can be simplified as

$$\frac{\boldsymbol{\beta}_{\mathcal{M}}^T(\bar{\boldsymbol{x}}_{1\mathcal{M}} - \boldsymbol{\mu}_{1\mathcal{M}})}{\sqrt{\Delta_p\{1 + O_P(K\sqrt{(\log p)/n})\}}} = O_P(1/\sqrt{n})(1 + O_P(K\sqrt{(\log p)/n}))$$

$$= O_P(1/\sqrt{n}) + O_P(K\sqrt{\log p}/n).$$

Since $1/\sqrt{n} = o(\sqrt{K/n})$ and $K\sqrt{\log p}/n = o(\sqrt{K/n})$, we have

$$\frac{\widehat{\boldsymbol{\beta}}_{\mathcal{M}}^T(\bar{\boldsymbol{x}}_{1\mathcal{M}} - \boldsymbol{\mu}_{1\mathcal{M}})}{\sqrt{\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}}} = O_P\left(\sqrt{K/n}\right). \tag{S1.9}$$

Then, it follows from (S1.7), (S1.8), and (S1.9) that

$$\begin{aligned}
&\frac{-\widehat{\boldsymbol{\beta}}_{\mathcal{M}}^T(\boldsymbol{\mu}_{1\mathcal{M}} - \bar{\boldsymbol{x}}_{1\mathcal{M}}) - \widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}/2}{\sqrt{\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}\mathcal{M}}}} \\
&= \frac{-\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}/2}{\sqrt{\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}}} - \frac{\widehat{\boldsymbol{\beta}}_{\mathcal{M}}^T(\boldsymbol{\mu}_{1\mathcal{M}} - \bar{\boldsymbol{x}}_{1\mathcal{M}})}{\sqrt{\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^T\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}}} \\
&= \frac{-\Delta_p\left(1 + O_P(K\sqrt{(\log p)/n})\right)}{2\sqrt{\Delta_p\left(1 + O_P(K\sqrt{(\log p)/n})\right)}} + O_P(\sqrt{K/n}) \\
&= -\frac{\sqrt{\Delta_p}\left(1 + O_P(K\sqrt{(\log p)/n})\right)}{2} + O_P(\sqrt{K/n}) \\
&= -\frac{\sqrt{\Delta_p}\left(1 + O_P(K\sqrt{(\log p)/n})\right)}{2},
\end{aligned} \tag{S1.10}$$

where in the second-to-last equation, we use the fact that $\{1+O_P(K\sqrt{(\log p)/n})\}^{-1/2}$ $= 1+O_P(K\sqrt{(\log p)/n})$, and in the last equation, we use $\sqrt{K/n} = o(K\{\Delta_p(\log p)/n\}^{1/2})$.

Using the same argument, we also have

$$\frac{\widehat{\boldsymbol{\beta}}_{\mathcal{M}}^{T}(\boldsymbol{\mu}_{0\mathcal{M}}-\bar{\boldsymbol{x}}_{0\mathcal{M}})-\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^{T}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}/2}{\sqrt{\widehat{\boldsymbol{\delta}}_{\mathcal{M}}^{T}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M}\mathcal{M}}\widehat{\boldsymbol{\Omega}}_{\mathcal{M}}\widehat{\boldsymbol{\delta}}_{\mathcal{M}}}}=-\frac{\sqrt{\Delta_p}\left(1+O_P(K\sqrt{(\log p)/n})\right)}{2}. \tag{S1.11}$$

Equations (S1.10) and (S1.11) together prove statement (a).

To prove (b), we use the fact that $R_{Bayes}=\Phi(-\sqrt{\Delta_p}/2)$ and a well-known result of the normal cumulative distribution function (Shao et al., 2011): that

$$\frac{x}{1+x^2}e^{-x^2/2}\leq\Phi(-x)\leq\frac{1}{x}e^{-x^2/2},\quad\text{for all } x>0. \tag{S1.12}$$

First, when $\Delta_p<\infty$, by the Mean Value Theorem, we have

$$R_{GS\text{-}LDA}(\boldsymbol{X})=R_{Bayes}+\phi(\widetilde{x})O_P\left(K\sqrt{\Delta_p(\log p)/n}\right)=R_{Bayes}+\phi(\widetilde{x})O_p(\sqrt{(\log p)/n}),$$

where $\widetilde{x}$ is a number between $-\sqrt{\Delta_p}/2$ and $-\sqrt{\Delta_p}(1+O_p(K\sqrt{(\log p)/n}))/2$. In the last equation, we use the fact that $K\asymp\Delta_p<\infty$, which is implied by Conditions 1, 2 and 4, since $\Delta_p<\infty$, R_{Bayes} is bounded away from 0. Then, we have

$$\frac{R_{GS\text{-}LDA}(\boldsymbol{X})}{R_{Bayes}}=1+\frac{\phi(\widetilde{x})}{R_{Bayes}}O_p(\sqrt{(\log p)/n}).$$

Then, the boundedness of the normal density function and R_{Bayes} implies that

$$\frac{R_{GS\text{-}LDA}(\boldsymbol{X})}{R_{Bayes}}-1=O_p(\sqrt{(\log p)/n}).$$

This proves statement (b).

When $\Delta_p \to \infty$, let $a_n = K\sqrt{(\log p)/n}$. Noting that $a_n = o(K^2\sqrt{(\log p)/n}) = o(1)$, it follows from statement (a) and (S1.12) that

$$\begin{aligned}\frac{R_{GS\text{-}LDA}(\boldsymbol{X})}{R_{Bayes}} &\le \frac{\frac{1}{\sqrt{\Delta_p}/2(1+O_p(a_n))} e^{-(\frac{\sqrt{\Delta_p}}{2}(1+O_p(a_n)))^2/2}}{\frac{\sqrt{\Delta_p}/2}{1+(\sqrt{\Delta_p}/2)^2} e^{-(\frac{\sqrt{\Delta_p}}{2})^2/2}} \\ &\le \frac{4+\Delta_p}{\Delta_p\{1+O_p(a_n)\}} e^{-\frac{\Delta_p}{8}(1-(1+O_p(a_n))^2)} \\ &\le \frac{4+\Delta_p}{\Delta_p\{1+O_p(a_n)\}} e^{O_p(\Delta_p a_n)}.\end{aligned}$$

Since $\Delta_p a_n \lesssim K^2\sqrt{(\log p)/n} = o(1)$, by the Taylor expansion, we have

$$\begin{aligned}\frac{R_{GS\text{-}LDA}(\boldsymbol{X})}{R_{Bayes}} &\le \frac{4+\Delta_p}{\Delta_p}(1+O_P(a_n))(1+O_P(\Delta_p a_n)) \le \frac{4+\Delta_p}{\Delta_p}(1+O_P(\Delta_p a_n)) \\ &= (1+\frac{4}{\Delta_p})(1+O_P(\Delta_p a_n)) \le 1 + O_P\left(\Delta_p^{-1}\right) + O_P\left(\Delta_p a_n\right).\end{aligned}$$

Using a similar argument, we can show that

$$\begin{aligned}\frac{R_{GS\text{-}LDA}(\boldsymbol{X})}{R_{Bayes}} &\ge \frac{\Delta_p}{4+\Delta_p}(1+O_p(\Delta_p a_n)) = (1-\frac{4}{4+\Delta_p})(1+O_p(\Delta_p a_n)) \\ &\ge 1 - O_P\left(\Delta_p^{-1}\right) - O_P\left(\Delta_p a_n\right).\end{aligned}$$

Combining the lower and upper bounds for $R_{GS\text{-}LDA}(\boldsymbol{X})/R_{Bayes}$, we obtain

$$\frac{R_{GS\text{-}LDA}(\boldsymbol{X})}{R_{Bayes}} - 1 = O_P\left(\max\{\Delta_p^{-1}, \Delta_p a_n\}\right) = O_P\left(\max\{\Delta_p^{-1}, K^2\sqrt{(\log p)/n}\}\right).$$

This proves statement (c). □

Proof of (2.3). We use a similar argument to the proof of Proposition 1 given by Li and Li (2018). Letting $\alpha = (\sigma_{cc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc})^{-1}$, we have

$$\begin{aligned}\Delta_{s+1} &= \begin{pmatrix} \boldsymbol{\delta}_S^T & \delta_c \end{pmatrix} \begin{pmatrix} \mathbf{\Sigma}_{SS} & \mathbf{\Sigma}_{Sc} \\ \mathbf{\Sigma}_{Sc}^T & \sigma_{cc} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\delta}_S \\ \delta_c \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\delta}_S^T & \delta_c \end{pmatrix} \begin{pmatrix} (\mathbf{\Sigma}_{SS} - \sigma_{cc}^{-1} \mathbf{\Sigma}_{Sc} \mathbf{\Sigma}_{Sc}^T)^{-1} & -\alpha \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc} \\ -\alpha \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} & \alpha \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_S \\ \delta_c \end{pmatrix}.\end{aligned}$$

By the Sherman–Morrison–Woodbury formula,

$$(\mathbf{\Sigma}_{SS} - \sigma_{cc}^{-1} \mathbf{\Sigma}_{Sc} \mathbf{\Sigma}_{Sc}^T)^{-1} = \mathbf{\Sigma}_{SS}^{-1} + \alpha \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc} \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1}.$$

Then we have

$$\begin{aligned}\Delta_{s+1} &= \begin{pmatrix} \boldsymbol{\delta}_S^T & \delta_c \end{pmatrix} \begin{pmatrix} \mathbf{\Sigma}_{SS}^{-1} + \alpha \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc} \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} & -\alpha \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc} \\ -\alpha \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} & \alpha \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_S \\ \delta_c \end{pmatrix} \\ &= \boldsymbol{\delta}_S^T \mathbf{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S + \alpha \boldsymbol{\delta}_S^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc} \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \delta_S - 2\alpha \delta_c \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S + \alpha \delta_c^2 \\ &= \Delta_s + \alpha(\delta_c - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \boldsymbol{\delta}_S)^2.\end{aligned}$$

Hence, we have

$$\theta_{Sc} = \Delta_{s+1} - \Delta_s = \frac{(\delta_c - \mathbf{\Sigma}_{Sc}^T \mathbf{\Omega}_{SS} \boldsymbol{\delta}_S)^2}{\sigma_{cc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Omega}_{SS} \mathbf{\Sigma}_{Sc}},$$

where $\mathbf{\Omega}_{SS} = \mathbf{\Sigma}_{SS}^{-1}$. With same argument as in the proof of Theorem 1, $\sigma_{cc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Omega}_{SS} \mathbf{\Sigma}_{Sc} > 0$ for any $c \notin S$. Thus, $\theta_{Sc} \geq 0$.

□

S2 Supporting Lemmas and their Proofs

Lemma 1. *Under Conditions 1 and 2, there exists a constant t_0 such that for all $0 < t < t_0$, the following results hold.*

(a) $P\left(\max_{i,j\le p} |\widehat{\sigma}_{ij} - \sigma_{ij}| \ge t\right) \le p^2 C_1 e^{-C_2 n t^2}$, where C_1 and C_2 are some generic positive constants.

(b) $P\left(\max_{j\le p} |\widehat{\delta}_j - \delta_j| \ge t\right) \le pC_1 e^{-C_2 n t^2}$, where C_1 and C_2 are some generic positive constants.

Proof of Lemma 1. These are standard concentration inequalities that follow from the normality assumption. The proof of (a) can be found in the proof of Lemma 3 of Bickel and Levina (2008), and (b) is a result obtained by applying the Chernoff method. □

Lemma 2. *Under Condition 2 and if $s\sqrt{\log(p)/n} = o(1)$, it holds that*

$$P\left(\|\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS}\| \lesssim s\sqrt{(\log p)/n}\right) \ge 1 - C_1 p^{-C_0};$$
$$P\left(\|\widehat{\mathbf{\Sigma}}_{SS}^{-1} - \mathbf{\Sigma}_{SS}^{-1}\| \lesssim s\sqrt{(\log p)/n}\right) \ge 1 - C_1 p^{-C_0},$$

where C_1 is some generic positive constant and C_0 is a sufficiently large constant.

Proof of Lemma 2. We have

$$\|\widehat{\mathbf{\Sigma}}_{SS}^{-1} - \mathbf{\Sigma}_{SS}^{-1}\| = \|\widehat{\mathbf{\Sigma}}_{SS}^{-1}(\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS})\mathbf{\Sigma}_{SS}^{-1}\| \leq \|\widehat{\mathbf{\Sigma}}_{SS}^{-1}\|\|\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS}\|\|\mathbf{\Sigma}_{SS}^{-1}\|. \tag{S2.1}$$

First, we bound $\|\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS}\|$. By definition,

$$\|\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS}\| \leq \|\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS}\|_1 = \max_{i \in S} \sum_{j \in S} |\widehat{\sigma}_{ij} - \sigma_{ij}|.$$

Then, it follows from Lemma 1 that

$$\begin{aligned} P\left(\|\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS}\| \geq t\right) &\leq P\left(\max_{i \in S} \sum_{j \in S} |\widehat{\sigma}_{ij} - \sigma_{ij}| \geq t\right) \leq P\left(\max_{i,j} |\widehat{\sigma}_{ij} - \sigma_{ij}| \geq t/s\right) \\ &\leq p^2 C_1 e^{-C_2 n t^2/s^2}. \end{aligned} \tag{S2.2}$$

Letting $t = C_D s\sqrt{(\log p)/n}$ for some large generic positive constant C_D and $C_0 = C_2 C_D$, we have

$$P\left(\|\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS}\| \geq C_D s\sqrt{(\log p)/n}\right) \leq C_1 p^{2-C_2C_D} \leq C_1 p^{-C_0}.$$

Next, we bound $\|\widehat{\mathbf{\Sigma}}_{SS}^{-1}\|_2$. Note that $\|\widehat{\mathbf{\Sigma}}_{SS}^{-1}\|_2 = 1/\lambda_{\min}(\widehat{\mathbf{\Sigma}}_{SS})$. By Weyl's inequality,

$$\lambda_{\min}(\mathbf{\Sigma}_{SS}) \leq \lambda_{\min}(\widehat{\mathbf{\Sigma}}_{SS}) + \lambda_{\max}(\mathbf{\Sigma}_{SS} - \widehat{\mathbf{\Sigma}}_{SS}) \leq \lambda_{\min}(\widehat{\mathbf{\Sigma}}_{SS}) + \|\widehat{\mathbf{\Sigma}}_{SS} - \mathbf{\Sigma}_{SS}\|$$

Then, it follows from Condition 2 and (S2.2) that

$$P\left(\lambda_{\max}(\widehat{\mathbf{\Sigma}}_{SS}^{-1}) \leq \frac{1}{m - C_0 s\sqrt{(\log p)/n}}\right) \geq 1 - C_1 p^{2-C_2C_D} \geq 1 - C_1 p^{-C_0}.$$

By Condition 2 and (S2.1), we have

$$P\left(\|\widehat{\mathbf{\Sigma}}_{SS}^{-1}-\mathbf{\Sigma}_{SS}^{-1}\|\leq\frac{C_0s\sqrt{(\log p)/n}}{m(m-C_0s\sqrt{(\log p)/n})}\right)=P\left(\|\widehat{\mathbf{\Sigma}}_{SS}^{-1}-\mathbf{\Sigma}_{SS}^{-1}\|\lesssim s\sqrt{(\log p)/n}\right)$$
$$\geq 1-C_1p^{2-C_2C_D}\geq 1-C_1p^{-C_0},$$

where in the first equality, we use the fact that as $s\sqrt{(\log p)/n}=o(1)$, $m-C_0s\sqrt{(\log p)/n}\geq m/2$. □

Lemma 3. *Under Condition 1–2, and if $s\sqrt{(\log p)/n}=o(1)$, the following results hold.*

$$P\left(|\widehat{\boldsymbol{\delta}}_S^T\mathbf{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S^T\mathbf{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S|\lesssim s\sqrt{(\log p)/n}\right)\geq 1-C_3p^{-C_0},$$
$$P\left(|\widehat{\mathbf{\Sigma}}_{Sc}^T\mathbf{\Sigma}_{SS}^{-1}\widehat{\mathbf{\Sigma}}_{Sc}-\mathbf{\Sigma}_{Sc}^T\mathbf{\Sigma}_{SS}^{-1}\mathbf{\Sigma}_{Sc}|\lesssim s\sqrt{(\log p)/n}\right)\geq 1-C_3p^{-C_0},$$
$$P\left(|\widehat{\mathbf{\Sigma}}_{Sc}^T\mathbf{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S-\mathbf{\Sigma}_{Sc}^T\mathbf{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S|\lesssim s\sqrt{(\log p)/n}\right)\geq 1-C_3p^{-C_0}.$$

where C_3 is a positive constant depending on the C_1, and C_0 is a sufficiently large constant.

Proof of Lemma 3. To prove the first result, we have

$$\widehat{\boldsymbol{\delta}}_S^T\mathbf{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S=\boldsymbol{\delta}_S^T\mathbf{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S+2\boldsymbol{\delta}_S^T\mathbf{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)+(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)^T\mathbf{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S).$$

Then, we have

$$\begin{aligned}
&P\left(|\widehat{\boldsymbol{\delta}}_S^T\boldsymbol{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S|\geq t\right)\\
&=P\left(|2{\boldsymbol{\delta}_S}^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)+(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)|\geq t\right)\\
&\leq P\left(|2\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)|+(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)\geq t\right)\\
&\leq P\left(|2\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)|\geq t/2\right)+P\left((\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)\geq t/2\right).
\end{aligned}$$

By Cauchy-Schwarz inequality and Conditions 1 and 2, we have

$$\begin{aligned}
|\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)| &\leq (\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S)^{1/2}\{(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)\}^{1/2}\\
&\leq (1/m)(\boldsymbol{\delta}_S^T\boldsymbol{\delta}_S)^{1/2}\{(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)^T(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)\}^{1/2}\\
&\leq (sM/m)\max_{i,j\leq p}|\widehat{\delta}_{ij}-\delta_{ij}|.
\end{aligned}$$

We also have

$$(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S)\leq (s/m)(\max_{j\leq p}|\widehat{\delta}_j-\delta_j|)^2.$$

Then, we have

$$\begin{aligned}
&P\left(|\widehat{\boldsymbol{\delta}}_S^T\boldsymbol{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S|\geq t\right)\\
&\leq P\left((sM/m)\max_{j\leq p}|\widehat{\delta}_j-\delta_j|\geq t/4\right)+P\left((s/m)(\max_{j\leq p}|\widehat{\delta}_j-\delta_j|)^2\geq t/2\right).
\end{aligned}$$

Letting $t=C_0 s\sqrt{(\log p)/n}$ for some large enough constant C_0, then it follows from Lemma 1 that

$$P\left(|\widehat{\boldsymbol{\delta}}_S^T\boldsymbol{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S-\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S|\lesssim s\sqrt{(\log p)/n}\right)\geq 1-C_3p^{-C_0},$$

where C_3 is some positive constant depending on the C_1.

To prove the second result, note that

$$\widehat{\mathbf{\Sigma}}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \widehat{\mathbf{\Sigma}}_{Sc} = \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc} + 2\mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc}) + (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc}).$$

Then, we have

$$\begin{aligned} & P\left(|\widehat{\mathbf{\Sigma}}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc}| \geq t \right) \\ \leq & P\left(|2\mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})| + |(\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})| \geq t \right) \\ \leq & P\left(|\mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})| \geq t/4 \right) + P\left(|(\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})| \geq t/2 \right). \end{aligned}$$

By Cauchy-Schwarz inequality and Condition 2,

$$\begin{aligned} |\mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})| &\leq (\mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc})^{1/2} \{(\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})\}^{1/2} \\ &\leq (1/m)(\mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{Sc})^{1/2} \{(\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})^T (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})\}^{1/2} \\ &\leq (sM/m) \max_{i,j \leq p} |\widehat{\sigma}_{ij} - \sigma_{ij}|, \end{aligned}$$

where in the last inequality, we use the fact that $|\sigma_{ij}| \leq \sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}} \leq \lambda_{\max}(\mathbf{\Sigma}) \leq M$, for all $i, j \leq p$.

Also under Condition 2, we have

$$(\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc})^T \mathbf{\Sigma}_{SS}^{-1} (\widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc}) \leq (s/m)(\max_{j \leq p} |\widehat{\sigma}_j - \sigma_j|)^2.$$

Then we have

$$\begin{aligned} & P\left(|\widehat{\mathbf{\Sigma}}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \widehat{\mathbf{\Sigma}}_{Sc} - \mathbf{\Sigma}_{Sc}^T \mathbf{\Sigma}_{SS}^{-1} \mathbf{\Sigma}_{Sc}| \geq t \right) \\ \leq & P\left((sM/m) \max_{i,j \leq p} |\widehat{\sigma}_{ij} - \sigma_{ij}| \geq t/4 \right) + P\left((s/m)(\max_{i,j \leq p} |\widehat{\sigma}_{ij} - \sigma_{ij}|)^2 \geq t/2 \right). \end{aligned}$$

Letting $t = C_0 s\sqrt{(\log p)/n}$, for some large constant C_0. Then, it follows from Lemma 1 that

$$P\left(|\widehat{\boldsymbol{\Sigma}}_{Sc}^T \boldsymbol{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\Sigma}}_{Sc} - \boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\Sigma}_{Sc}| \lesssim s\sqrt{(\log p)/n}\right) \geq 1 - C_3 p^{-C_0},$$

where C_3 is some positive constant depending on the C_1 .

To prove the third result, note that

$$\begin{aligned}
&\widehat{\boldsymbol{\Sigma}}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S \\
&= \boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S + \boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\Sigma}}_{Sc} - \boldsymbol{\Sigma}_{Sc}) + \boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S) + (\widehat{\boldsymbol{\Sigma}}_{Sc} - \boldsymbol{\Sigma}_{Sc})^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S).
\end{aligned}$$

Then, we have

$$\begin{aligned}
&P\left(|\widehat{\boldsymbol{\Sigma}}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S| \geq t\right) \\
&\leq P\left(|\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\Sigma}}_{Sc} - \boldsymbol{\Sigma}_{Sc})| \geq t/3\right) + P\left(|\boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S)| \geq t/3\right) \\
&\quad + P\left(|(\widehat{\boldsymbol{\Sigma}}_{Sc} - \boldsymbol{\Sigma}_{Sc})^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S)| \geq t/3\right).
\end{aligned}$$

By Cauchy-Schwarz inequality and Conditions 1 and 2, we have

$$\begin{aligned}
|\boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\Sigma}}_{Sc} - \boldsymbol{\Sigma}_{Sc})| &\leq (sM/m)\max_{i,j\leq p}|\widehat{\sigma}_{ij} - \sigma_{ij}|; \\
|\boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S)| &\leq (sM/m)\max_{j\leq p}|\widehat{\delta}_j - \delta_j|; \\
(\widehat{\boldsymbol{\Sigma}}_{Sc} - \boldsymbol{\Sigma}_{Sc})^T\boldsymbol{\Sigma}_{SS}^{-1}(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S) &\leq (s/m)(\max_{i,j\leq p}|\widehat{\sigma}_{ij} - \sigma_{ij}|)(\max_{j\leq p}|\widehat{\delta}_j - \delta_j|).
\end{aligned}$$

Then, we have

$$
\begin{aligned}
&P\left(|\widehat{\boldsymbol{\Sigma}}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S| \geq t\right)\\
&\leq P\left((sM/m)\max_{i,j\leq p}|\widehat{\sigma}_{ij}-\sigma_{ij}| \geq t/3\right) + P\left((sM/m)\max_{j\leq p}|\widehat{\delta}_j-\delta_j| \geq t/3\right)\\
&\quad + P\left((s/m)(\max_{i,j\leq p}|\widehat{\sigma}_{ij}-\sigma_{ij}|)(\max_{j\leq p}|\widehat{\delta}_j-\delta_j|) \geq t/3\right)\\
&\leq P\left((sM/m)\max_{i,j\leq p}|\widehat{\sigma}_{ij}-\sigma_{ij}| \geq t/3\right) + P\left((sM/m)\max_{j\leq p}|\widehat{\delta}_j-\delta_j| \geq t/3\right)\\
&\quad + P\left(\max_{i,j\leq p}|\widehat{\sigma}_{ij}-\sigma_{ij}| \geq \sqrt{mt/(3s)}\right) + P\left(\max_{j\leq p}|\widehat{\delta}_j-\delta_j| \geq \sqrt{mt/(3s)}\right).
\end{aligned}
$$

Letting $t = C_0 s\sqrt{(\log p)/n}$ for some large constant C_0, it follows from Lemma 1 that

$$
P\left(|\widehat{\boldsymbol{\Sigma}}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S| \lesssim s\sqrt{(\log p)/n}\right) \geq 1 - C_3 p^{-C_0},
$$

where C_3 is some positive constant depending on the C_1. □

Lemma 4. *Under Condition 1–2 and if $s\sqrt{(\log p)/n} = o(1)$, the following results hold.*

$$
\begin{aligned}
P\left(|\widehat{\boldsymbol{\delta}}_S^T\widehat{\boldsymbol{\Sigma}}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S| \lesssim s^2\sqrt{(\log p)/n}\right) &\geq 1 - C_4 p^{-C_0};\\
P\left(|\widehat{\boldsymbol{\Sigma}}_{Sc}^T\widehat{\boldsymbol{\Sigma}}_{SS}^{-1}\widehat{\boldsymbol{\Sigma}}_{Sc} - \boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\Sigma}_{Sc}| \lesssim s^2\sqrt{(\log p)/n}\right) &\geq 1 - C_4 p^{-C_0};\\
P\left(|\widehat{\boldsymbol{\Sigma}}_{Sc}^T\widehat{\boldsymbol{\Sigma}}_{SS}^{-1}\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{SS}^{-1}\boldsymbol{\delta}_S| \lesssim s^2\sqrt{(\log p)/n}\right) &\geq 1 - C_4 p^{-C_0};
\end{aligned}
$$

where C_4 is some positive constant depending on the C_1 and C_3 and C_0 is a sufficiently large constant.

Proof of Lemma 4. By definition,

$$
\begin{aligned}
&|\widehat{\boldsymbol{\delta}}_S^T(\widehat{\boldsymbol{\Sigma}}_{SS}^{-1} - \boldsymbol{\Sigma}_{SS}^{-1})\widehat{\boldsymbol{\delta}}_S| \leq \|\widehat{\boldsymbol{\Sigma}}_{SS}^{-1} - \boldsymbol{\Sigma}_{SS}^{-1}\|\widehat{\boldsymbol{\delta}}_S^T\widehat{\boldsymbol{\delta}}_S \\
&\leq \|\widehat{\boldsymbol{\Sigma}}_{SS}^{-1} - \boldsymbol{\Sigma}_{SS}^{-1}\|\{\boldsymbol{\delta}_S^T\boldsymbol{\delta}_S + 2(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S)^T\boldsymbol{\delta}_S + (\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S)^T(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S)\}.
\end{aligned}
$$

By Condition 1, $\boldsymbol{\delta}_S^T\boldsymbol{\delta}_S = O(s)$. It follows from Lemma 1 that $(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S)^T\boldsymbol{\delta}_S = o_P(s)$ and $(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S)^T(\widehat{\boldsymbol{\delta}}_S - \boldsymbol{\delta}_S) = o_P(s)$. Then, it follows from Lemma 2 that

$$
P\left(|\widehat{\boldsymbol{\delta}}_S^T(\widehat{\boldsymbol{\Sigma}}_{SS}^{-1} - \boldsymbol{\Sigma}_{SS}^{-1})\widehat{\boldsymbol{\delta}}_S| \lesssim s^2\sqrt{(\log p)/n}\right) \geq 1 - C_4 p^{-C_0}.
$$

This result, together with Lemma 3 and the triangular inequality, prove the first result. The other two results can be proved by a similar argument, noting that $\boldsymbol{\Sigma}_{Sc}^T\boldsymbol{\Sigma}_{Sc} = O(s)$.

□

S3 Additional Results in Cancer Subtype Analysis

Figure S1 shows the variable selection performance of the GS-LDA, ROAD and Logistic-L1 in cancer subtype analysis.

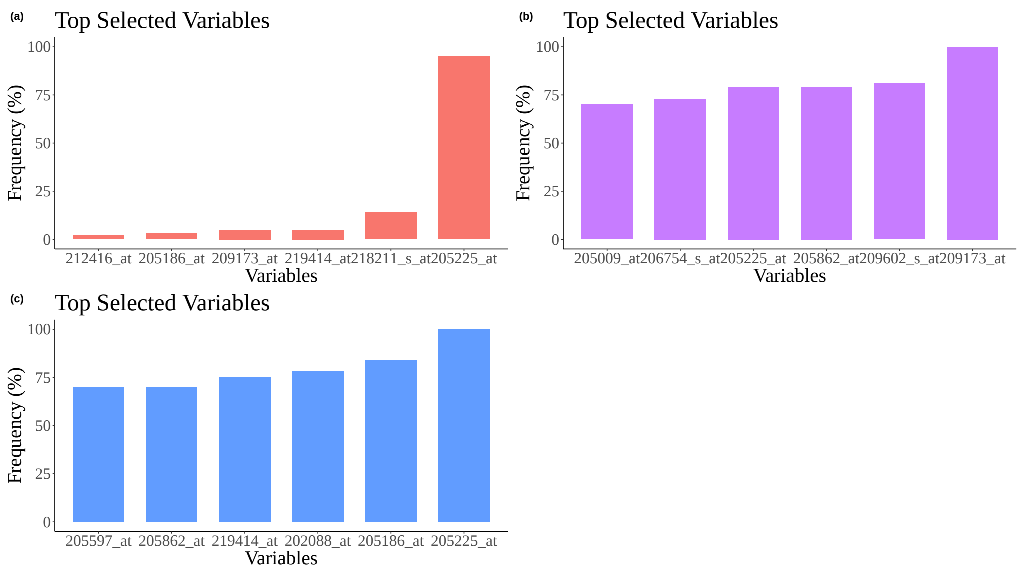

Figure S1: Variable selection performance of the three classifiers in classifying cancer subtypes: panel (a) for the GS-LDA; panel (b) for the ROAD; and panel (c) for the Logistic-L1.

References

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36,** 199–227.

Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis.* Cambridge University Press

Li, Q. and Li, L. (2018). Integrative linear discriminant analysis with guaranteed error rate improvement. *Biometrika* **105,** 917–930.

Shao, J., Wang, Y., Deng, X., and Wang, S. (2011). Sparse linear discriminant analysis by thresholding for high dimensional data. *The Annals of statistics* **39,** 1241–1265.