Penalized integrative analysis under the accelerated failure time model

Qingzhao Zhang[1], Sanguo Zhang[1], Jin Liu[2], Jian Huang[3] and Shuangge Ma[4*]

[1] *School of Mathematical Sciences, University of Chinese Academy of Sciences*

[2] *Center of Quantitative Medicine, Duke-NUS School of Medicine*

[3] *Department of Statistics and Actuarial Science, University of Iowa*

[4] *Department of Biostatistics, Yale University* **email: shuangge.ma@yale.edu*

Supplementary Material

This file contains proofs (Section S1) for the theoretical results described in the main text as well as additional numerical results (Section S2).

S1 Proofs

Let

$$\boldsymbol{y}^m = {W_m}^{1/2}\boldsymbol{Y}^m \ \text{ and } X^m = {W_m}^{1/2}\boldsymbol{X}^m. \tag{S1.1}$$

Then $(\boldsymbol{Y}^m - \boldsymbol{X}^m\boldsymbol{\beta}^m)^\top W_m(\boldsymbol{Y}^m - \boldsymbol{X}^m\boldsymbol{\beta}^m)$ can be rewritten as $\|\boldsymbol{y}^m - X^m\boldsymbol{\beta}^m\|^2$, where $\|\cdot\|$ is the ℓ_2 norm. Moreover, we can easily see that

$$\boldsymbol{y}^m = X^m\boldsymbol{\beta}^m + {W_m}^{1/2}\boldsymbol{\epsilon}^m. \tag{S1.2}$$

Proof of Theorem 1. First, we prove that

$$\Pr\left\{\|\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*}\|_2 < \lambda\frac{4}{\underline{\rho}_1^m}\frac{n}{n_m}, m = 1,\cdots,M\right\} \geq 1-\tau_1,$$

where $\tau_1 = \sum_{m=1}^{M}\exp\left(-\frac{\lambda^2 n^2}{2\sigma_m^2\bar{\rho}_2^m n_m}\right)$. Recall that $\hat{\boldsymbol{\beta}}_{\mathcal{B}} = \arg\min_{\boldsymbol{\beta}_{\mathcal{B}}} L(\boldsymbol{\beta}_{\mathcal{B}})$, where

$$L(\boldsymbol{\beta}_{\mathcal{B}}) = \frac{1}{2n}\sum_{m=1}^{M}\|\boldsymbol{y}^m - X_S^m\boldsymbol{\beta}_S^m\|^2 + \lambda\sum_{j\in S}\|\boldsymbol{\beta}_j\|_2.$$

Let $r_m = \lambda\sqrt{|S|}\frac{4}{\underline{\rho}_1^m}\frac{n}{n_m}$ and $\mathfrak{I} = \{\boldsymbol{\beta}_{\mathcal{B}} : \|\boldsymbol{\beta}_S^m - \boldsymbol{\beta}_S^{m*}\|_2 = r_m, m = 1,\cdots,M\}$. It suffices to show that

$$\Pr\left(\inf_{\boldsymbol{\beta}_{\mathcal{B}}\in\mathfrak{I}} L(\boldsymbol{\beta}_{\mathcal{B}}) > L(\boldsymbol{\beta}_{\mathcal{B}}^*)\right) \geq 1-\tau_1.$$

This implies that with probability at least $1-\tau_1$, $L(\boldsymbol{\beta}_{\mathcal{B}})$ has a local minimum $\hat{\boldsymbol{\beta}}_{\mathcal{B}}$ that satisfies $\|\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*}\|_2 < \lambda\sqrt{|S|}\frac{4}{\underline{\rho}_1^m}\frac{n}{n_m}$, for $m = 1,\cdots,M$.

Let $\boldsymbol{u} \in R^{p\times M}$ with $\|\boldsymbol{u}_S^m\|_2 = 1,\ m = 1, \cdots, M$. Define $\boldsymbol{\beta}_S^m = \boldsymbol{\beta}_S^{m*} + r_m \boldsymbol{u}_S^m$. Consider $Q(\boldsymbol{u}_{\mathcal{B}}) = n\left\{L(\boldsymbol{\beta}_{\mathcal{B}}) - L(\boldsymbol{\beta}^*_{\mathcal{B}})\right\}$. Obviously, it is equivalent to show that

$$\Pr\left(\inf_{\|\boldsymbol{u}^m\|_2=1,\ m=1,\cdots,M} Q(\boldsymbol{u}_{\mathcal{B}}) > 0\right) \geq 1 - \tau_1. \tag{S1.3}$$

Together with (S1.1) and (S1.2), we have

$$\begin{aligned}
Q(\boldsymbol{u}_{\mathcal{B}}) &= \frac{1}{2}\sum_{m=1}^{M}\left(\|\boldsymbol{y}^m - X_S^m(\boldsymbol{\beta}_S^{m*} + r_m\boldsymbol{u}_S^m)\|^2 - \|\boldsymbol{y}^m - X_S^m\boldsymbol{\beta}_S^{m*}\|^2\right) \\
&\quad + n\lambda\sum_{j\in S}\left\{\left\|\boldsymbol{\beta}_j^* + \boldsymbol{r}\circ\boldsymbol{u}_j\right\|_2 - \left\|\boldsymbol{\beta}_j^*\right\|_2\right\} \\
&= -\sum_{m=1}^{M} r_m \boldsymbol{u}_S^{m\top}\boldsymbol{X}_S^{m\top}W_m\boldsymbol{\epsilon}^m + \frac{1}{2}\sum_{m=1}^{M} r_m^2\boldsymbol{u}_S^{m\top}\boldsymbol{X}_S^{m\top}W_m\boldsymbol{X}_S^m\boldsymbol{u}_S^m \\
&\quad + n\lambda\sum_{j\in S}\left\{\left\|\boldsymbol{\beta}_j^* + \boldsymbol{r}\circ\boldsymbol{u}_j\right\|_2 - \left\|\boldsymbol{\beta}_j^*\right\|_2\right\} \\
&=: Q_1 + Q_2 + Q_3,
\end{aligned} \tag{S1.4}$$

where $\boldsymbol{r} = (r_1, \cdots, r_M)^\top$, and $\circ$ denotes the Hadamard (component-wise) product. Write $Q_1 = \sum_{m=1}^{M} Q_{1m}$ where $Q_{1m} = -r_m\boldsymbol{u}_S^{m\top}\boldsymbol{X}_S^{m\top}W_m\boldsymbol{\epsilon}^m$. Note that $\|W_m\boldsymbol{X}_S^m\boldsymbol{u}_S^m\|_2^2 \leq n_m\bar{\rho}_2^m$. With the sub-Gaussian tail as specified in Condition 1, we have for any given ε_m

$$\Pr(|Q_{1m}| > r_m\varepsilon_m) \leq 2\exp\left(-\frac{\varepsilon_m^2}{2\sigma_m^2\|W_m\boldsymbol{X}_S^m\boldsymbol{u}_S^m\|_2^2}\right) \leq 2\exp\left(-\frac{\varepsilon_m^2}{2n_m\bar{\rho}_2^m\sigma_m^2}\right).$$

Together with the Bonferroni's inequality, we have

$$\Pr(Q_1 < -\sum_{m=1}^{M} r_m\varepsilon_m) \leq \sum_{m=1}^{M}\Pr(Q_{1m} < -r_m\varepsilon_m) \leq \sum_{m=1}^{M}\exp\left(-\frac{\varepsilon_m^2}{2n_m\bar{\rho}_2^m\sigma_m^2}\right).$$

Set $\varepsilon_m = \frac{1}{4}\underline{\rho}_1^m n_m r_m$. Then

$$\Pr(Q_1 \geq -\frac{1}{4}\sum_{m=1}^{M} r_m^2 n_m\underline{\rho}_1^m) \geq 1 - \sum_{m=1}^{M}\exp\left(-\frac{n_m r_m^2(\underline{\rho}_1^m)^2}{32\bar{\rho}_2^m\sigma_m^2}\right). \tag{S1.5}$$

For Q_2, since $\boldsymbol{u}_S^{m\top}\boldsymbol{X}_S^{m\top}W_m\boldsymbol{X}_S^m\boldsymbol{u}_S^m \geq n_m\underline{\rho}_1^m$, we have

$$Q_2 \geq \frac{1}{2}\sum_{m=1}^{M} r_m^2 n_m\underline{\rho}_1^m. \tag{S1.6}$$

Term Q_3 can be dealt with as follows. By the Triangle inequality and $(\sum_{i=1}^{d}|v_i|)^2 \leq d\sum_{i=1}^{d} v_i^2$, for any sequence v_i, we have

$$\sum_{j\in S}\left\|\boldsymbol{\beta}_j^* + \boldsymbol{r}\circ\boldsymbol{u}_j\right\|_2 - \left\|\boldsymbol{\beta}_j^*\right\|_2 \le \sum_{j\in S}\|\boldsymbol{r}\circ\boldsymbol{u}_j\|_2$$

$$\le \sqrt{|S|}\sqrt{\sum_{j\in S}\|\boldsymbol{r}\circ\boldsymbol{u}_j\|_2^2} = \sqrt{|S|}\sqrt{\sum_{m=1}^{M} r_m^2} \le \sqrt{|S|}\sum_{m=1}^{M} r_m.$$

Therefore, we have that term Q_3 satisfies

$$|Q_3| \le n\lambda\sqrt{|S|}\sum_{m=1}^{M} r_m. \tag{S1.7}$$

Combining (S1.4), (S1.5), (S1.6), and (S1.7), we have

$$Q(\boldsymbol{u}_S) \ge \frac{1}{4}\sum_{m=1}^{M} r_m^2 n_m \underline{\rho}_1^m - n\lambda\sqrt{|S|}\sum_{m=1}^{M} r_m := L(\boldsymbol{r}) \tag{S1.8}$$

with probability at least $1-\sum_{m=1}^{M}\exp\left(-\frac{n_m r_m^2(\underline{\rho}_1^m)^2}{32\bar{\rho}_2^m\sigma_m^2}\right)$. Recall that $r_m = \lambda\sqrt{|S|}\frac{4}{\underline{\rho}_1^m}\frac{n}{n_m}$. Then $L(\boldsymbol{r})>0$ with probability at least $1-\sum_{m=1}^{M}\exp\left(-\frac{\lambda^2|S|n^2}{2\sigma_m^2\bar{\rho}_2^m n_m}\right)$. Therefore, (S1.3) is proved, and Part 1 of Theorem 1 is established.

Now consider Part 2. By the Karush-Kuhn-Tucher(KKT) conditions, we need to prove that for $m = 1,\cdots,M$,

$$-X_S^{m\top}\left(\boldsymbol{y}^m - X_S^m\hat{\boldsymbol{\beta}}_S^m\right) + n\lambda\frac{\hat{\boldsymbol{\beta}}_S^m}{\|\hat{\boldsymbol{\beta}}_{\mathcal{B}}\|_2} = 0, \tag{S1.9}$$

$$\|X_{S^c}^{\top}(\boldsymbol{y}^m - X_S^m\hat{\boldsymbol{\beta}}_S^m)\|_\infty \le n\lambda. \tag{S1.10}$$

Then $\hat{\boldsymbol{\beta}}^{glasso} = \{\hat{\boldsymbol{\beta}}_{\mathcal{B}}^{glasso}, \hat{\boldsymbol{\beta}}_{\mathcal{B}^c}^{glasso}\}$ with $\hat{\boldsymbol{\beta}}_{\mathcal{B}}^{glasso} = \hat{\boldsymbol{\beta}}_{\mathcal{B}}, \hat{\boldsymbol{\beta}}_{\mathcal{B}^c}^{glasso} = 0$ is a local minimizer of (3). From Part 1, $\tilde{\boldsymbol{\beta}}_S$ minimizes

$$L(\boldsymbol{\beta}_{\mathcal{B}}) = \frac{1}{2n}\sum_{m=1}^{M}\|\boldsymbol{y}^m - X_S^m\boldsymbol{\beta}_S^m\|^2 + \lambda\sum_{j\in S}\|\boldsymbol{\beta}_j\|_2.$$

Therefore, (S1.9) holds, together with (S1.2) which also yields

$$\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*} = \left(X_S^{m\top}X_S^m\right)^{-1}\left\{X_S^{m\top}W_m{}^{1/2}\boldsymbol{\epsilon}^m - n\lambda\frac{\hat{\boldsymbol{\beta}}_S^m}{\|\hat{\boldsymbol{\beta}}_{\mathcal{B}}\|_2}\right\}. \tag{S1.11}$$

Note that

$$X_{S^c}^{m\top}(\boldsymbol{y}^m - X_S^m\hat{\boldsymbol{\beta}}_S^m) = X_{S^c}^{m\top}W_m{}^{1/2}\boldsymbol{\epsilon}^m - X_{S^c}^{m\top}X_S^m(\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*}). \tag{S1.12}$$

Substituting (S1.11) into (S1.12), we obtain

$$
\begin{aligned}
& \|X_{S^c}^{m\top}(\boldsymbol{y}^m - X_S^m \hat{\boldsymbol{\beta}}_S^m)\|_\infty \\
= & \left\| X_{S^c}^{m\top} {W_m}^{1/2} \boldsymbol{\epsilon}^m - X_{S^c}^{m\top} X_S^m \left(X_S^{m\top} X_S^m \right)^{-1} \left\{ X_S^{m\top} {W_m}^{1/2} \boldsymbol{\epsilon}^m - n\lambda \frac{\hat{\boldsymbol{\beta}}_S^m}{\|\hat{\boldsymbol{\beta}}_{\mathcal{B}}\|_2} \right\} \right\|_\infty \\
\leq & \left\| X_{S^c}^{m\top} {W_m}^{1/2} \boldsymbol{\epsilon}^m \right\|_\infty + \left\| X_{S^c}^{m\top} X_S^m \left(X_S^{m\top} X_S^m \right)^{-1} X_S^{m\top} {W_m}^{1/2} \boldsymbol{\epsilon}^m \right\|_\infty \\
& + n\lambda \left\| X_{S^c}^{m\top} X_S^m \left(X_S^{m\top} X_S^m \right)^{-1} \frac{\hat{\boldsymbol{\beta}}_S^m}{\|\hat{\boldsymbol{\beta}}_{\mathcal{B}}\|_2} \right\|_\infty \\
\leq & \left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty + \left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{X}_S^m \left(\boldsymbol{X}_S^{m\top} W_m \boldsymbol{X}_S^m \right)^{-1} \right\|_\infty \left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty \\
& + n\lambda \left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{X}_S^m \left(\boldsymbol{X}_S^{m\top} W_m \boldsymbol{X}_S^m \right)^{-1} \right\|_\infty \left\| \frac{\hat{\boldsymbol{\beta}}_S^m}{\|\hat{\boldsymbol{\beta}}_{\mathcal{B}}\|_2} \right\|_\infty \\
\leq & \left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty + \psi_m \left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty + n\lambda\psi_m
\end{aligned} \tag{S1.13}
$$

By the condition $\psi_m \leq D_m < 1$, if

$$
\left\| \boldsymbol{X}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty \leq n\lambda \frac{1 - D_m}{1 + D_m}, \tag{S1.14}
$$

then from (S1.13) it follows

$$
\begin{aligned}
\|X_{S^c}^{m\top}(\boldsymbol{y}^m - X_S^m \tilde{\boldsymbol{\beta}}_S^m)\|_\infty & \leq \left\| \boldsymbol{X}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty (1 + \psi_m) + n\lambda\psi_m \\
& \leq n\lambda(1 - D_m) + n\lambda D_m = n\lambda.
\end{aligned}
$$

We now derive the probability bounds for the event in (S1.14). By the Bonferroni's inequality and sub-Gaussian tail probability bound in Condition 1,

$$
\begin{aligned}
& \Pr\left\{ \left\| \boldsymbol{X}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty > n\lambda \frac{1 - D_m}{1 + D_m}, \text{ for } m = 1, \cdots, M \right\} \\
\leq & \; p \sum_{m=1}^{M} \Pr\left\{ |\boldsymbol{X}_j^{m\top} W_m \boldsymbol{\epsilon}^m| > n\lambda \frac{1 - D_m}{1 + D_m} \right\} \\
\leq & \; 2p \sum_{m=1}^{M} \exp\left\{ -\frac{n^2\lambda^2(1 - D_m)^2}{2n_m\Lambda_m\sigma_m^2(1 + D_m)^2} \right\}.
\end{aligned} \tag{S1.15}
$$

Then Part 2 is established by combining Part1, (S1.9), (S1.10), and (S1.15). □

Proof of Theorem 2. Recall that $\tilde{\boldsymbol{\beta}}_{\mathcal{B}} = \arg\min_{\boldsymbol{\beta}_{\mathcal{B}}} H(\boldsymbol{\beta}_{\mathcal{B}})$, where

$$
H(\boldsymbol{\beta}_{\mathcal{B}}) = \frac{1}{2n} \sum_{m=1}^{M} \|\boldsymbol{y}^m - X_S^m \boldsymbol{\beta}_S^m\|^2.
$$

Let $r_m = \sqrt{\frac{|S|}{n}} R_m$ with $R_m \in (0, \infty)$ and $\mathfrak{I} = \{\boldsymbol{\beta}_{\mathcal{B}} : \|\boldsymbol{\beta}_S^m - \boldsymbol{\beta}_S^{m*}\|_2 = r_m, m = 1, \cdots, M\}$. Similar as the proof of part 1 in Theorem 1, if we can prove

$$\Pr\left(\inf_{\boldsymbol{\beta}_{\mathcal{B}}\in\mathfrak{I}} H(\boldsymbol{\beta}_{\mathcal{B}}) > H(\boldsymbol{\beta}^*_{\mathcal{B}})\right) \geq 1 - \sum_{m=1}^{M} \exp\left\{-R_m^2 \frac{|S|(\underline{\rho}_1^m)^2}{8\bar{\rho}_2^m \sigma_m^2}\right\}, \tag{S1.16}$$

then $H(\boldsymbol{\beta}_{\mathcal{B}})$ has a local minimum $\hat{\boldsymbol{\beta}}_{\mathcal{B}}$ that satisfies $\|\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*}\|_2 < r_m, m = 1, \cdots, M$ with probability at least $1 - \sum_{m=1}^{M} \exp\left\{-R_m^2 \frac{|S|(\underline{\rho}_1^m)^2}{8\bar{\rho}_2^m \sigma_m^2}\right\}$.

Together with (S1.1) and (S1.2), we have

$$\begin{aligned} H(\boldsymbol{\beta}_{\mathcal{B}}) - H(\boldsymbol{\beta}^*_{\mathcal{B}}) &= -\sum_{m=1}^{M} (\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*})^\top \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \\ &\quad + \frac{1}{2}\sum_{m=1}^{M} (\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*})^\top \boldsymbol{X}_S^{m\top} W_m \boldsymbol{X}_S^m (\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*}) \\ &=: H_1 + H_2, \end{aligned} \tag{S1.17}$$

For H_2, since $\lambda_{\min}\left\{n_m^{-1}\boldsymbol{X}_S^{m\top} W_m \boldsymbol{X}_S^m\right\} = \underline{\rho}_1^m$ and $\|\boldsymbol{\beta}_S^m - \boldsymbol{\beta}_S^{m*}\|_2 = r_m$, we have

$$H_2 \geq \frac{1}{2}\sum_{m=1}^{M} r_m^2 n_m \underline{\rho}_1^m. \tag{S1.18}$$

For H_1 we have for any ε_m,

$$\begin{aligned} \Pr(H_1 \leq -\sum_{m=1}^{M} r_m \varepsilon_m) &\leq \sum_{m=1}^{M} \exp\left(-\frac{r_m^2 \varepsilon_m^2}{2\sigma_m^2 \|W_m \boldsymbol{X}_S^m (\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*})\|_2^2}\right) \\ &\leq \sum_{m=1}^{M} \exp\left(-\frac{\varepsilon_m^2}{2 n_m \bar{\rho}_2^m \sigma_m^2}\right). \end{aligned}$$

The first inequality holds due to the sub-Gaussian tail probability under Condition 1, and the last inequality holds due to the fact that $\|W_m \boldsymbol{X}_S^m (\hat{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*})\|_2^2 \leq n_m \bar{\rho}_2^m r_m^2$. Set $\varepsilon_m = \frac{1}{2}\underline{\rho}_1^m n_m r_m$. Then

$$\Pr(H_1 > -\frac{1}{2}\sum_{m=1}^{M} r_m^2 n_m \underline{\rho}_1^m) \geq 1 - \sum_{m=1}^{M} \exp\left(-\frac{n_m r_m^2 (\underline{\rho}_1^m)^2}{8\bar{\rho}_2^m \sigma_m^2}\right). \tag{S1.19}$$

Recall that $r_m = \sqrt{\frac{|S|}{n}} R_m$. Combining (S1.17), (S1.18) and (S1.19), we have (S1.16) holds. This complete the proof of Part 1.

Next, we prove Part 2. By the Karush-Kuhn-Tucher(KKT) conditions, we need to prove that $\hat{\boldsymbol{\beta}}^{oracle}$ satisfies

$$-X_S^{m\top}\left(\boldsymbol{y}^m - X_S^m \tilde{\boldsymbol{\beta}}_S^m\right) + n p'_\lambda(\|\tilde{\boldsymbol{\beta}}_{\mathcal{B}}\|_2) \circ \frac{\tilde{\boldsymbol{\beta}}_S^m}{\|\tilde{\boldsymbol{\beta}}_{\mathcal{B}}\|_2} = 0, \tag{S1.20}$$

$$\|X_{S^c}^{m\top}(\boldsymbol{y}^m - X_S^m \tilde{\boldsymbol{\beta}}_S^m)\|_\infty \leq n p'_\lambda(0+). \tag{S1.21}$$

If $\min_{j\in S}\|\tilde{\boldsymbol{\beta}}_j\|_2 > \theta\lambda$, $p'_\lambda(\|\tilde{\boldsymbol{\beta}}_{\mathcal{B}}\|_2) = 0$, and certainly (S1.20) holds. Define

$$R_m^\dagger \le \frac{\min_{j\in S}\|\boldsymbol{\beta}_j^*\|_2}{2\sqrt{M}}\sqrt{\frac{n_m}{|S|}}.$$

Note that $\lambda < \frac{\min_{j\in S}\|\boldsymbol{\beta}_j^*\|_2}{2\theta}$. Therefore, we can conclude the event

$$\left\{\|\tilde{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*}\|_2 \le \sqrt{\frac{|S|}{n_m}}R_m^\dagger,\ m = 1,\cdots,M\right\}$$

belongs to the event $\left\{\min_{j\in S}\|\tilde{\boldsymbol{\beta}}_j\|_2 > \theta\lambda\right\}$. That is,

$$\begin{aligned}\Pr\left\{\min_{j\in S}\|\tilde{\boldsymbol{\beta}}_j\|_2 > \theta\lambda\right\} &\ge \Pr\left(\|\tilde{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*}\|_2 \le \sqrt{\frac{|S|}{n_m}}R_m^\dagger,\ m = 1,\cdots,M\right)\\ &\ge 1 - \sum_{m=1}^M \exp\left\{-\frac{|S|(\underline{\rho}_1^m)^2}{8\sigma_m^2\bar{\rho}_2^m}R_m^{\dagger 2}\right\}. \qquad \text{(S1.22)}\end{aligned}$$

Now consider the probability of

$$\|X_{S^c}^{m\top}(\boldsymbol{y}^m - X_S^m\tilde{\boldsymbol{\beta}}_S^m)\|_\infty \le np'_\lambda(0+),\ \text{for } m = 1,\cdots,M. \qquad \text{(S1.23)}$$

Note that

$$X_{S^c}^{m\top}(\boldsymbol{y}^m - X_S^m\tilde{\boldsymbol{\beta}}_S^m) = \boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{\epsilon}^m - \boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{X}_S^m(\tilde{\boldsymbol{\beta}}_S^m - \boldsymbol{\beta}_S^{m*}). \qquad \text{(S1.24)}$$

Combining (S1.23) and (S1.24), we can obtain

$$\begin{aligned}&\|X_{S^c}^{m\top}(\boldsymbol{y}^m - X_S^m\tilde{\boldsymbol{\beta}}_S^m)\|_\infty\\ = &\left\|\boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{\epsilon}^m - \boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{X}_S^m\left(\boldsymbol{X}_S^{m\top}W_m\boldsymbol{X}_S^m\right)^{-1}\boldsymbol{X}_S^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty\\ \le &\left\|\boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty + \left\|\boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{X}_S^m\left(\boldsymbol{X}_S^{m\top}W_m\boldsymbol{X}_S^m\right)^{-1}\boldsymbol{X}_S^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty\\ \le &\left\|\boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty + \left\|\boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{X}_S^m\left(\boldsymbol{X}_S^{m\top}W_m\boldsymbol{X}_S^m\right)^{-1}\right\|_\infty\left\|\boldsymbol{X}_S^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty\\ \le &\left\|\boldsymbol{X}_{S^c}^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty + \psi_m\left\|\boldsymbol{X}_S^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty\\ \le &\left\|\boldsymbol{X}^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty(1+\psi_m). \qquad \text{(S1.25)}\end{aligned}$$

If

$$\left\|\boldsymbol{X}^{m\top}W_m\boldsymbol{\epsilon}^m\right\|_\infty \le \frac{np'_\lambda(0+)}{(1+\psi_m)}, \qquad \text{(S1.26)}$$

then from (S1.25) it follows

$$\|X_{S^c}^{m\top}(\boldsymbol{y}^m - X_S^m \tilde{\boldsymbol{\beta}}_S^m)\|_\infty \quad \leq \quad \frac{np'_\lambda(0+)}{(1+\psi_m)}(1+\psi_m) \leq np'_\lambda(0+),$$

which proves (S1.21). We now derive the probability bounds for the event in (S1.26). In fact, by the Bonferroni's inequality and sub-Gaussian tail probability bound under Condition 1,

$$\begin{aligned}
& \Pr\left\{\left\|\boldsymbol{X}^{m\top} W_m \boldsymbol{\epsilon}^m\right\|_\infty > \frac{np'_\lambda(0+)}{(1+\psi_m)},\ \exists\, m \in \{1,\cdots,M\}\right\} \\
\leq \quad & p \sum_{m=1}^M \Pr\left\{|\boldsymbol{X}_j^{m\top} W_m \boldsymbol{\epsilon}^m| > \frac{np'_\lambda(0+)}{(1+\psi_m)}\right\} \\
\leq \quad & 2p \sum_{m=1}^M \exp\left\{-\frac{n^2 p'^2_\lambda(0+)}{2n_m \Lambda_m \sigma_m^2 (1+\psi_m)^2}\right\}. \qquad \text{(S1.27)}
\end{aligned}$$

Part (2) is proved by combining (S1.20), (S1.21), (S1.22), and (S1.27). □

Proof of Theorem 3. The proof is similar to that of Part 1 of Theorem 2 and is omitted here. □

Proof of Theorem 4. By the Karush-Kuhn-Tucher(KKT) conditions, we need to prove that $\check{\boldsymbol{\beta}}$ satisfies

$$-X_{S_m}^{m\top}\left(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m\right) + np'_{O,\lambda_O}\Big(\sum_{m=1}^M p_{I,\lambda_I}(|\check{\boldsymbol{\beta}}_{S_m}^m|)\Big) \circ p'_{I,\lambda_I}(|\check{\boldsymbol{\beta}}_{S_m}^m|) = 0, \qquad \text{(S1.28)}$$

$$|X_{S-S_m}^{m\top}(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m)| \leq np'_{I,\lambda_I}(0+)p'_{O,\lambda_O}\Big(\sum_{m=1}^M p_{I,\lambda_I}(|\check{\boldsymbol{\beta}}_{S-S_m}^m|)\Big), \qquad \text{(S1.29)}$$

$$\|X_{S^c}^{m\top}(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m)\|_\infty \leq np'_{O,\lambda_O}(0+)p'_{I,\lambda_I}(0+). \qquad \text{(S1.30)}$$

If $\min_{j\in S_m} |\check{\beta}_j^m| > \theta_I \lambda_I$, then $p'_{I,\lambda_I}(|\check{\boldsymbol{\beta}}_{S_m}^m|) = 0$. Recall the definition of the estimator $\check{\boldsymbol{\beta}}_{S_m}^m$. We can easily get (S1.28). Set

$$C_m^\dagger \leq \frac{\min_{(j,m)\in\mathcal{A}} |\beta_j^{m*}|}{2}\sqrt{\frac{n_m}{|S_m|}}.$$

Note that $\lambda_I < \frac{\min_{(j,m)\in\mathcal{A}} |\beta_j^{m*}|}{2\theta_I}$. Therefore,

$$\begin{aligned}
& \Pr\left\{\min_{(j,m)\in\mathcal{A}} |\check{\beta}_j^m| > \theta_I \lambda_I\right\} \geq \Pr\left(\|\check{\boldsymbol{\beta}}_{S_m}^m - \boldsymbol{\beta}_{S_m}^{m\ *}\|_2 \leq \sqrt{\frac{|S_m|}{n_m}} C_m^\dagger,\ m = 1,\cdots,M\right) \\
\geq \quad & 1 - 2\sum_{m=1}^M \exp\left\{-C_m^{\dagger 2} \frac{|S_m|(\underline{\rho}_1^{*m})^2}{8\bar{\rho}_2^{*m}\sigma_m^2}\right\}. \qquad \text{(S1.31)}
\end{aligned}$$

In fact,

$$X_{S-S^m}^{m\top}(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m) = \boldsymbol{X}_{S-S^m}^{m\top} W_m \boldsymbol{\epsilon}^m - \boldsymbol{X}_{S-S^m}^{m\top} W_m X_{S_m}^m (\check{\boldsymbol{\beta}}_{S_m}^m - \boldsymbol{\beta}_{S_m}^{m\ *}),$$

and $\check{\boldsymbol{\beta}}_{S_m}^m - \boldsymbol{\beta}_{S_m}^{m\ *} = \left(\boldsymbol{X}_{S_m}^{m\top} W_m \boldsymbol{X}_{S_m}^m\right)^{-1} X_{S_m}^{m\top} W_m \boldsymbol{\epsilon}^m$. Then we have

$$
\begin{aligned}
& |X_{S-S_m}^{m\top}(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m)| \\
\leq\ & |\boldsymbol{X}_{S-S^m}^{m\top} W_m \boldsymbol{\epsilon}^m| + |\boldsymbol{X}_{S-S^m}^{m\top} W_m X_{S_m}^m \left(\boldsymbol{X}_{S_m}^{m\top} W_m \boldsymbol{X}_{S_m}^m\right)^{-1} X_{S_m}^{m\top} W_m \boldsymbol{\epsilon}^m| \\
\leq\ & |\boldsymbol{X}_{S-S^m}^{m\top} W_m \boldsymbol{\epsilon}^m| + \left\| \boldsymbol{X}_{S-S^m}^{m\top} W_m X_{S_m}^m \left(\boldsymbol{X}_{S_m}^{m\top} W_m \boldsymbol{X}_{S_m}^m\right)^{-1} \right\|_\infty |X_{S_m}^{m\top} W_m \boldsymbol{\epsilon}^m| \\
\leq\ & \left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty + \psi_m^* \left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty \\
\leq\ & \left\| \boldsymbol{X}_S^{m\top} W_n \boldsymbol{\epsilon}^m \right\|_\infty (1 + \psi_m^*).
\end{aligned} \tag{S1.32}
$$

Hence (S1.29) holds when

$$
\left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty \leq \frac{n p'_{I,\lambda_I}(0+) p'_{O,\lambda_O}(J^{-m} f_I^{max})}{(1 + \psi_m^*)}. \tag{S1.33}
$$

That is because for $m = 1, \cdots, M$,

$$
\begin{aligned}
& |X_{S-S_m}^{m\top}(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m)| \leq \frac{n p'_{I,\lambda_I}(0+) p'_{O,\lambda_O}(J^{-m} f_I^{max})}{(1 + \psi_m^*)} (1 + \psi_m^*) \\
\leq\ & n p'_{I,\lambda_I}(0+) p'_{O,\lambda_O}(J^{-m} f_I^{max})(1 + \psi_m^*) \\
\leq\ & n p'_{I,\lambda_I}(0+) p'_{O,\lambda_O} \left(\sum_{m=1}^{M} p_{I,\lambda_I}(|\check{\boldsymbol{\beta}}_{S-S_m}^m|) \right).
\end{aligned}
$$

We now derive the probability bounds for the event in (S1.33). In fact, by Bonferroni's inequality and sub-Gaussian tail probability bound in Condition 1,

$$
\begin{aligned}
& \Pr\left\{ \left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty > \frac{n p'_{I,\lambda_I}(0+) p'_{O,\lambda_O}(J^{-m} f_I^{max})}{(1 + \psi_m^*)},\ \exists\, m \in \{1, \cdots, M\} \right\} \\
\leq\ & 2|S| \sum_{m=1}^{M} \exp\left\{ -\frac{n^2 p'^2_{I,\lambda_I}(0+) p'^2_{O,\lambda_O}(J^{-m} f_I^{max})}{2 n_m \bar{\rho}_2^{*m} \sigma_m^2 (1 + \psi_m^*)^2} \right\}.
\end{aligned} \tag{S1.34}
$$

Similarly, we can prove (S1.30). Actually,

$$
\begin{aligned}
& |X_{S^c}^{m\top}(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m)|_\infty \leq \left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty + \psi_m^* \left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty \\
<\ & \left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty + \frac{\psi_m^*}{(1 + \psi_m^*)} n p'_{O,\lambda_O}(0+) p'_{I,\lambda_I}(0+).
\end{aligned}
$$

Based on the above discussions, (S1.30) follows if $\left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty < \frac{n p'_{O,\lambda_O}(0+) p'_{I,\lambda_I}(0+)}{(1+\psi_m^*)}$. The probability bound is derived as

$$
\begin{aligned}
& \Pr\left\{ \left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty > \frac{n p'_{O,\lambda_O}(0+) p'_{I,\lambda_I}(0+)}{(1 + \psi_m^*)},\ \exists\, m \in \{1, \cdots, M\} \right\} \\
\leq\ & 2(p - |S|) \sum_{m=1}^{M} \exp\left\{ -\frac{n^2 p'^2_{I,\lambda_I}(0+) p'^2_{O,\lambda_O}(0+)}{2 n_m \Lambda_m \sigma_m^2 (1 + \psi_m^*)^2} \right\}.
\end{aligned} \tag{S1.35}
$$

The theorem is proved by combining (S1.28), (S1.29), (S1.29), (S1.31),(S1.34) and (S1.35) . $\square$

Proof of Theorem 5. By the Karush-Kuhn-Tucher(KKT) conditions, we need to prove that $\check{\boldsymbol{\beta}}$ satisfies

$$
\begin{aligned}
-X_{S_m}^{m\top}\left(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m\right) + np'_{1,\lambda_1}(\|\check{\boldsymbol{\beta}}_{S_m}\|_2) \circ \frac{\check{\boldsymbol{\beta}}_{S_m}^m}{\|\check{\boldsymbol{\beta}}_{S_m}\|_2} & \\
+ np'_{2,\lambda_2}(|\check{\boldsymbol{\beta}}_{S_m}^m|) \circ \mathrm{sgn}(\check{\boldsymbol{\beta}}_{S_m}^m) = 0, & \qquad \text{(S1.36)} \\
\|X_{S-S_m}^{m\top}(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m)\|_\infty \le np'_{2,\lambda_2}(0+), & \qquad \text{(S1.37)} \\
\|X_{S^c}^{m\top}(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m)\|_\infty \le np'_{1,\lambda_1}(0+) + np'_{2,\lambda_2}(0+). & \qquad \text{(S1.38)}
\end{aligned}
$$

Note that $\check{\boldsymbol{\beta}}_{S_m}^m$ satisfies $-X_{S_m}^{m\top}\left(\boldsymbol{y}^m - X_{S_m}^m \check{\boldsymbol{\beta}}_{S_m}^m\right) = 0$. If

$$
\min_{j \in S_m} |\check{\beta}_j^m| > \theta_2\lambda_2 \ \text{ and } \min_{j\in S} \|\check{\boldsymbol{\beta}}_j\|_2 > \theta_1\lambda_1,
$$

then we have (S1.36). Set $C_m^\dagger \le \frac{\min_{(j,m)\in\mathcal{A}} |\beta_j^{m*}|}{2}\sqrt{\frac{n_m}{|S_m|}}$. Note that

$$
\lambda_1 < \frac{\min_{j\in S} \|\boldsymbol{\beta}_j{}^*\|_2}{2\theta_1}, \ \lambda_2 < \frac{\min_{(j,m)\in\mathcal{A}} |\beta_j^{m*}|}{2\theta_2}.
$$

Therefore, if $\min_{(j,m)\in\mathcal{A}} |\check{\beta}_j^m| > \theta_2\lambda_2$, then we must have $\min_{j\in S} \|\check{\boldsymbol{\beta}}_j\|_2 > \theta_1\lambda_1$.

$$
\begin{aligned}
& \Pr\left\{ \min_{(j,m)\in\mathcal{A}} |\check{\beta}_j^m| > \theta_1\lambda_1, \ \min_{j\in S} \|\check{\boldsymbol{\beta}}_j\|_2 > \theta_2\lambda_2 \right\} \\
\ge \ & \Pr\left(\|\check{\boldsymbol{\beta}}_{S_m}^m - \boldsymbol{\beta}_{S_m}^m{}^*\|_2 \le \sqrt{\frac{|S_m|}{n_m}} C_m^\dagger, \ m = 1, \cdots, M \right) \\
\ge \ & 1 - 2\sum_{m=1}^M \exp\left\{ -C_m^{\dagger 2} \frac{|S_m|(\rho_1^{*m})^2}{8\bar{\rho}_2^{*m}\sigma_m^2} \right\}. \qquad \text{(S1.39)}
\end{aligned}
$$

Similar as the proof of Theorem 4, (S1.37) holds when

$$
\left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty \le \frac{np'_{2,\lambda_2}(0+)}{(1+\psi_m^*)}. \qquad \text{(S1.40)}
$$

Then we have

$$
\begin{aligned}
& \Pr\left\{ \left\| \boldsymbol{X}_S^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty > \frac{np'_{2,\lambda_2}(0+)}{(1+\psi_m^*)}, \ \exists\, m \in \{1, \cdots, M\} \right\} \\
\le \ & 2|S| \sum_{m=1}^M \exp\left\{ -\frac{n^2 p'^2_{2,\lambda_2}(0+)}{2n_m \bar{\rho}_2^{*m}\sigma_m^2(1+\psi_m^*)^2} \right\}. \qquad \text{(S1.41)}
\end{aligned}
$$

Similarly, we can show (S1.38) holds when $\left\| \boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{\epsilon}^m \right\|_\infty < \frac{np'_{1,\lambda_1}(0+) + np'_{2,\lambda_2}(0+)}{(1+\psi_m^*)}$. The probability bound is derived as

$$\begin{aligned}
& \Pr\left\{\left\|\boldsymbol{X}_{S^c}^{m\top} W_m \boldsymbol{\epsilon}^m\right\|_\infty > \frac{np'_{1,\lambda_1}(0+) + np'_{2,\lambda_2}(0+)}{(1+\psi_m^*)},\ \exists\, m \in \{1,\cdots,M\}\right\} \\
\leq\ & 2(p-|S|)\sum_{m=1}^{M}\exp\left\{-\frac{n^2[p'_{1,\lambda_1}(0+) + p'_{2,\lambda_2}(0+)]^2}{2n_m\Lambda_m\sigma_m^2(1+\psi_m^*)^2}\right\}. \qquad \text{(S1.42)}
\end{aligned}$$

The theorem is proved by combining (S1.36), (S1.37), (S1.37), (S1.39), (S1.41), and (S1.42). □

S2 Additional Numerical Results

Table S2.1: Analysis of lung cancer data using SGMCP: identified genes and their estimates.

Probe	Gene	UM	HLM	DFCI	MSKCC
201462_at	SCRN1		0.0034		0.0020
202831_at	GPX2		-0.0022		-0.0021
203917_at	CXADR	0.0021		0.0004	0.0066
205776_at	FMO5	0.0005	0.0035	0.0038	
206754_s_at	CYP2B6	0.0012		0.0020	
207850_at	CXCL3		-0.0216		0.0120
208025_s_at	HMGA2	-0.0028	0.0001	-0.0037	-0.0012
219654_at	PTPLA	-0.0025	-0.0145		0.0055
219764_at	FZD10	-0.0005	-0.0019	-1.6E-05	-0.0022

Table S2.2: Analysis of lung cancer data using meta-analysis: identified genes and their estimates.

Probe	Gene	UM	HLM	DFCI	MSKCC
201462_at	SCRN1		0.0101		
203559_s_at	ABP1		0.0005		
203876_s_at	MMP11				-0.0066
203921_at	CHST2	0.0051			
204855_at	SERPINB5	-0.0012			
206754_s_at	CYP2B6			0.0104	
206994_at	CST4		-0.0037		
207850_at	CXCL3		-0.0246		
208025_s_at	HMGA2	-0.0021		-0.0010	
209343_at	EFHD1				0.0096
212328_at	LIMCH1			0.0050	
213703_at	LINC00342	0.0008			
215867_x_at	CA12		-0.0026		
218677_at	S100A14				-0.0257
218824_at	PNMAL1	0.0003			
219654_at	PTPLA		-0.0240		
219747_at	NDNF	0.0002			
220952_s_at	PLEKHA5		-0.0047		
221841_s_at	KLF4	-0.0047			
222043_at	CLU		0.0049		

Table S2.3: Analysis of lung cancer data using pooled analysis: identified genes and their estimates.

Probe	Gene	UM	HLM	DFCI	MSKCC
201462_at	SCRN1		0.0101		
203559_s_at	ABP1		0.0005		
203876_s_at	MMP11				-0.0066
203921_at	CHST2	0.0051			
204855_at	SERPINB5	-0.0012			
206754_s_at	CYP2B6			0.0104	
206994_at	CST4		-0.0037		
207850_at	CXCL3		-0.0246		
208025_s_at	HMGA2	-0.0021		-0.0010	
209343_at	EFHD1				0.0096
212328_at	LIMCH1			0.0050	
213703_at	LINC00342	0.0008			
215867_x_at	CA12		-0.0026		
218677_at	S100A14				-0.0257
218824_at	PNMAL1	0.0003			
219654_at	PTPLA		-0.0240		
219747_at	NDNF	0.0002			
220952_s_at	PLEKHA5		-0.0047		
221841_s_at	KLF4	-0.0047			
222043_at	CLU		0.0049		

Table S2.4: Analysis of lung cancer data using GMCP: identified genes and their estimates.

Probe	Gene	UM	HLM	DFCI	MSKCC
202503_s_at	KIAA0101	-0.0009	-0.0020	-0.0021	-0.0019
205776_at	FMO5	0.0001	0.0002	0.0002	-0.0001
207850_at	CXCL3	-0.0017	-0.0139	0.0029	0.0095
208025_s_at	HMGA2	-3.2E-05	1.1E-05	-3.8E-05	-2.2E-05
219654_at	PTPLA	-0.0036	-0.0092	-0.0024	0.0060
219764_at	FZD10	-0.0014	-0.0036	-0.0014	-0.0036