Multi-response Regression for Block-missing Multi-modal Data without Imputation

Haodong Wang, Quefeng Li and Yufeng Liu

The University of North Carolina at Chapel Hill

Supplementary Material

S1 Toy example with adaptive LASSO penalty

The advantage of joint estimation is not pertained to the choice of penalty. If we choose other penalty functions, we can still see such an advantage. To illustrate that, we did some further simulation experiments in the toy example with the adaptive LASSO penalty. In particular, we plot the estimation errors of the original adaptive LASSO method ("Separate LASSO" in Figure 1) and the adaptive LASSO penalty with precision matrix as the adjusted weight ("2-step weighted LASSO" in Figure 1). We use cross-validation to choose the tuning parameter. The resulting estimation errors are shown in Figure 1. It shows that the two-step weighted adaptive LASSO may perform worse than separate adaptive LASSO, so it also has the same

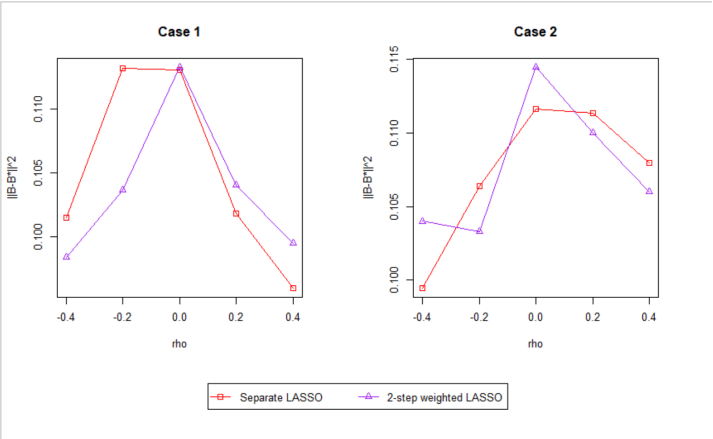

Figure 1: Plots of the estimation errors for separated adaptive LASSO and two-step weighted adaptive LASSO when $\mathbf{\Sigma}_\epsilon = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The left panel is for $\mathbf{B}^* = \begin{pmatrix} 0 & 0 \\ 2 & 3.5 \end{pmatrix}$ and the right panel is for $\mathbf{B}^* = \begin{pmatrix} 0 & 0 \\ -2 & 3.5 \end{pmatrix}$.

problem as LASSO. Jointly estimate $\mathbf{B}^*$ and $\mathbf{C}^*$ with the adaptive LASSO penalty can solve this problem.

S2 Regularity Conditions

Condition A1. *Suppose there exists two positive constants* L_1 *and* L_2 *such that for any* $\mathbf{u}_1 \in \mathbb{R}^p$, $\mathbf{u}_2 \in \mathbb{R}^q$, *and* $t \in \mathbb{R}$, $\mathbb{E}\left(\exp\left(t\mathbf{u}_1^\top \mathbf{x}_i\right)\right) \leq \exp\left(\frac{L_1^2 \|\mathbf{u}_1\|_2^2 t^2}{2}\right)$ *and* $\mathbb{E}(\exp(t\mathbf{u}_2^\top \mathbf{y}_i)) \leq \exp\left(\frac{L_2^2 \|\mathbf{u}_2\|_2^2 t^2}{2}\right)$.

Condition A2. $n_{XX} \geq 6 \log p$, $n_{XY} \geq 4 \log(pq)$ *and* $n_{YY} \geq 6 \log q$.

Under Condition A1, the predictor and the response vectors follow sub-Gaussian distributions. Condition A2 ensures that the missing pro-

portion of the data is not too large in order to get consistent estimators of $\mathbf{B}^*$ and $\mathbf{C}^*$. If we further assume that $(\log(pq))/n_0 = O(1)$, with $n_0 = \min\{n_{XX}, n_{XY}, n_{YY}\}$, Condition A2 is satisfied when n_0 is sufficiently large.

In order to prove Lemma 3.3 and 3.4, we need the following additional assumptions.

Condition A3. $\|\mathbf{B}^*\|_{L_1} \leq c_0^{\gamma_1}$ *and* $\|\mathbf{C}^*\|_{L_1} \leq c_0^{\gamma_2}$ *where* $0 < \gamma_1, \gamma_2 < \frac{1}{16}$ *and* $c_0 = \min\{\frac{n_{XY}}{\log(pq)}, \frac{n_{XX}}{\log p}, \frac{n_{YY}}{\log q}\}$. $\|\mathbf{B}^*\|_2 \leq c$ *for some positive constant* c.

Condition A4. *Suppose that* $\mathbf{\Sigma}_{XX}$ *and* $\mathbf{C}^*$ *satisfy* $c \leq \lambda_{\min}(\mathbf{\Sigma}_{XX}) \leq \lambda_{\max}(\mathbf{\Sigma}_{XX}) \leq C$ *and* $c \leq \lambda_{\min}(\mathbf{C}^*) \leq \lambda_{\max}(\mathbf{C}^*) \leq C$ *for some positive constants* c *and* C.

Condition A3 makes a weak assumption on the upper bounds of the norms of the true parameters, where the two upper bounds can diverge as $(\log(pq))/n_0 \to 0$, with $n_0 = \min\{n_{XX}, n_{XY}, n_{YY}\}$. We impose the sub-Gaussian assumption on y_i in Condition A1. We essentially assume it has bounded variance. Since it is the response from a linear model, it is reasonable to assume $\text{var}(y_i)$ is bounded. Since $var(y_i) \geq \mathbf{B}^{*\top} \text{var}(x_i)\mathbf{B}^*$, the boundness of $\text{var}(y_i)$ implies that $\|\mathbf{B}^*\|_2$ is bounded, if we assume $\lambda_{\max}(\text{var}(x_i)) < \infty$. Condition A4 ensures that the eigenvalues of $\mathbf{\Sigma}_{XX}$ and $\mathbf{C}^*$ are bounded away from 0 and infinity.

Condition A5. $\left\|(\mathbf{C}^* \otimes \mathbf{\Sigma}_{XX})_{S_B^C S_B} (\mathbf{C}^* \otimes \mathbf{\Sigma}_{XX})^{-1}_{S_B S_B}\right\|_\infty \leq 1 - \eta$ *holds for a constant* $\eta \in (0, 1)$.

Condition A5 can be viewed as a population version of the strong irrepresentable condition proposed in Zhao and Yu (2006).

S3 Proof of Proposition 2.1

We use a similar argument as the proof of Proposition 1 in Yu et al. (2020), we first decompose the objective function into the estimation error of intra-modality sample covariance matrix, the estimation error of diagonal entries and the estimation error of cross-modality sample covariance matrix. Then we find the optimal value of each term.

By using the facts that $\mathbf{\Sigma}_{XX} = \mathbf{\Sigma}_I + \mathbf{\Sigma}_C$ and $\mathbb{E}(\tilde{\mathbf{\Sigma}}_I) = \mathbf{\Sigma}_I$, we can rewrite the objective function in (2.10) as

$$\begin{aligned}
&\arg\min_{\alpha_1,\alpha_2} \mathbb{E}\|\hat{\mathbf{\Sigma}}_{XX} - \mathbf{\Sigma}_{XX}\|_F^2 \\
&= \arg\min_{\alpha_1,\alpha_2} \left\{ \alpha_1^2 \mathbb{E} \left\|\tilde{\mathbf{\Sigma}}_I - \mathbf{\Sigma}_I\right\|_F^2 + (1-\alpha_1)^2 \mathbb{E} \left\|\mathrm{diag}(\tilde{\mathbf{\Sigma}}_I) - \mathbf{\Sigma}_I\right\|_F^2 + \mathbb{E} \left\|\alpha_2 \tilde{\mathbf{\Sigma}}_C - \mathbf{\Sigma}_C\right\|_F^2 \right\}.
\end{aligned}$$

The optimal value of α_2 can be obtained by minimizing $\mathbb{E}\|\alpha_2 \tilde{\mathbf{\Sigma}}_C - \mathbf{\Sigma}_C\|_F^2$. Thus, the optimal value is $\alpha_2^* = \frac{\|\mathbf{\Sigma}_C\|_F^2}{\|\mathbf{\Sigma}_C\|_F^2 + \delta_C^2}$. Then taking the derivative of the objective function with respect to α_1, we can find that the optimal value of α_1 is $\alpha_1^* = \frac{\theta^2}{\theta^2 + \delta_I^2}$.

At the optimum, the value of the objective function is equal to $\frac{\delta_I^2 \theta^2}{\delta_I^2 + \theta^2} +$

$\frac{\delta_C^2\|\mathbf{\Sigma}_C\|_F^2}{\delta_C^2+\|\mathbf{\Sigma}_C\|_F^2} \leq \delta_I^2 + \delta_C^2$. Since $\mathbb{E}\|\tilde{\mathbf{\Sigma}}_{XX} - \mathbf{\Sigma}_{XX}\|_F^2 = \delta_I^2 + \delta_C^2$, we have $\mathbb{E}\|\hat{\mathbf{\Sigma}}_{XX}^* - \mathbf{\Sigma}_{XX}\|_F^2 \leq \mathbb{E}\|\tilde{\mathbf{\Sigma}}_{XX} - \mathbf{\Sigma}_{XX}\|_F^2$.

By taking the derivative of the objective function of (2.11) with respect to α_3, the optimal value of α_3 is $\alpha_3^* = \frac{\|\mathbf{\Sigma}_{XY}\|_F^2}{\|\mathbf{\Sigma}_{XY}\|_F^2+\delta_{XY}^2}$. At the optimum, the value of the objective function is equal to $\frac{\delta_{XY}^2\|\mathbf{\Sigma}_{XY}\|_F^2}{\delta_{XY}^2+\|\mathbf{\Sigma}_{XY}\|_F^2}$, which is less than δ_{XY}^2. Since $\mathbb{E}\|\tilde{\mathbf{\Sigma}}_{XY} - \mathbf{\Sigma}_{XY}\|_F^2 = \delta_{XY}^2$, we have $\mathbb{E}\|\hat{\mathbf{\Sigma}}_{XY}^* - \mathbf{\Sigma}_{XY}\|_F^2 \leq \mathbb{E}\|\tilde{\mathbf{\Sigma}}_{XY} - \mathbf{\Sigma}_{XY}\|_F^2$.

S4 Proof of Theorem 3.1

We first gives the large deviation bounds for the sample covariance matrices $\tilde{\mathbf{\Sigma}}_{XX}$ and $\tilde{\mathbf{\Sigma}}_{XY}$ by a similar argument as the proof of Lemma 1 in Yu et al. (2020). Then we calculate the convergence rate of entries in the estimated intra-modality sample covariance matrix, entries in the estimated cross-modality sample covariance matrix and estaimated diagnal entries usingthe previous bound, and then calculate the overall convergence rate of using the union bound.

Without loss of generality, we assume $\sigma_{jj}^{XX} = 1$ for $1 \leq j \leq p$. Then, under Condition A1, we know that X_j is sub-Gaussian with parameter L_1. Let $\delta_1 = 8\sqrt{6}\,(1+4L_1^2)\sqrt{\frac{\log p}{n_{jk}^{XX}}}$. If $n_{XX} > 6\log p$, we have $\delta_1 < 8\,(1+4L_1^2)$.

By letting $\nu_1 = 8\sqrt{6}\,(1+4L_1^2)$, it follows from Lemma S9.2 that

$$\begin{aligned} P\left(\left|\tilde{\sigma}_{jk}^{XX} - \sigma_{jk}^{XX}\right| \geq \delta_1\right) &\leq 4\exp\left\{-\frac{n_{jk}^{XX}\delta_1^2}{128\left(1+4L_1^2\right)^2}\right\} \\ &= 4\exp\left\{-\frac{\nu_1^2 \log p}{128\left(1+4L_1^2\right)^2}\right\} \\ &= 4p^{-\frac{\nu_1^2}{128\left(1+4L_1^2\right)^2}} \\ &\leq \frac{4}{p^3}. \end{aligned}$$

Hence, under Conditions A1 and A2, we have

$$\max_{j,k} P\left(\left|\tilde{\sigma}_{jk}^{XX} - \sigma_{jk}^{XX}\right| \geq \nu_1\sqrt{\tfrac{\log p}{n_{jk}^{XX}}}\right) \leq \tfrac{4}{p^3}.$$

By the union bound, we have

$$P\left(\|\tilde{\boldsymbol{\Sigma}}_{XX} - \boldsymbol{\Sigma}_{XX}\|_\infty \geq \nu_1\sqrt{\tfrac{\log p}{n_{XX}}}\right) \leq \tfrac{4}{p}.$$

Let Y_j denote the jth response. Without loss of generality, we assume that Y_j has finite variance. Under Condition A1, $Y_j/\sqrt{\mathrm{var}(Y_j)}$ is sub-Gaussian with parameter $L_2/\sqrt{\mathrm{var}(Y_j)}$. Let $\delta_3 = 16(1+4\max\{L_1^2, \frac{L_2^2}{\min_j(\mathrm{var}(Y_j))}\})$ $\sqrt{\frac{\log(pq)}{n_{jk}^{XY}}}\max\{\max_j(\mathrm{var}(Y_j)), 1\}$. When $n_{XY} > 4\log(pq)$, we have

$$\delta_3 < 8\left(1+4\max\left\{L_1^2, \frac{L_2^2}{\min_j\left(\mathrm{var}(Y_j)\right)}\right\}\right)\max_j\left(\mathrm{var}(Y_j), 1\right).$$

By choosing $\nu_2 = 16(1+4\max\{L_1^2, \frac{L_2^2}{\min_j(\mathrm{var}(Y_j))}\})\max\{\max_j(\mathrm{var}(Y_j)), 1\}$,

it follows from Lemma S9.2 that for any $1 \le j, k \le pq$, we have

$$P\left(\left|\tilde{\sigma}_{jk}^{XY} - \sigma_{jk}^{XY}\right| \ge \delta_3\right) \le 4\exp\left\{-\frac{\nu_2^2 \log(pq)}{128\left(1+4\max\left\{L_1^2, \frac{L_2^2}{\min_j\left(\mathrm{var}(\mathbf{y}_j)\right)}\right\}\right)^2 \max_j\left(\mathrm{var}(Y_j), 1\right)}\right\}$$
$$\le \frac{4}{(pq)^2}.$$

Hence, by Condition A1 and $n_{XY} > 4\log(pq)$, there exists a positive constant ν_2 such that

$$\max_{j,k} P\left(\left|\tilde{\sigma}_{jk}^{XY} - \sigma_{jk}^{XY}\right| \ge \nu_2\sqrt{\tfrac{\log(pq)}{n_{jk}^{XY}}}\right) \le \tfrac{4}{(pq)^2}.$$

By the union bound, we have

$$P\left(\|\tilde{\mathbf{\Sigma}}_{XY} - \mathbf{\Sigma}_{XY}\|_\infty \ge \nu_2\sqrt{\tfrac{\log(pq)}{n_{XY}}}\right) \le \tfrac{4}{pq}.$$

Based on the definition of $\hat{\mathbf{\Sigma}}_{XX}$, we have

$$\hat{\sigma}_{jt}^{XX} - \sigma_{jt}^{XX} = \begin{cases} \tilde{\sigma}_{jt}^{XX} - \sigma_{jt}^{XX} & \text{if } j = t; \\ \alpha_1\tilde{\sigma}_{jt}^{XX} - \sigma_{jt}^{XX} & \text{if } j \ne t, j \text{ and } t \text{ are in the same modality;} \\ \alpha_2\tilde{\sigma}_{jt}^{XX} - \sigma_{jt}^{XX} & \text{if } j \text{ and } t \text{ are in different modalities.} \end{cases}$$

Thus, if $j = t$, there exists a positive constant ν_1 such that with probability at least $1 - 4/p^3$, it holds that

$$\left|\hat{\sigma}_{jt}^{XX} - \sigma_{jt}^{XX}\right| = \left|\tilde{\sigma}_{jt}^{XX} - \sigma_{jt}^{XX}\right| \le \nu_1\sqrt{\log p / n_X}.$$

If $j \ne t$ and j and t are in the same modality, it holds with probability at

least $1-4/p^3$ that

$$\begin{aligned}\left|\hat{\sigma}_{jt}^{XX}-\sigma_{jt}^{XX}\right| = \left|\alpha_1\tilde{\sigma}_{jt}^{XX}-\sigma_{jt}^{XX}\right| &\le \alpha_1\left|\tilde{\sigma}_{jt}^{XX}-\sigma_{jt}^{XX}\right| + (1-\alpha_1)\left|\sigma_{jt}^{XX}\right| \\ &\le \alpha_1\left|\tilde{\sigma}_{jt}^{XX}-\sigma_{jt}^{XX}\right| + 1-\alpha_1 \\ &\le \alpha_1\nu_1\sqrt{\log p/n_X} + 1-\alpha_1.\end{aligned}$$

Similarly, if $j \neq t$ and j and t are in different modalities, it holds with probability at least $1-4/p^3$ that

$$\begin{aligned}\left|\hat{\sigma}_{jt}^{XX}-\sigma_{jt}^{XX}\right| = \left|\alpha_2\tilde{\sigma}_{jt}^{XX}-\sigma_{jt}^{XX}\right| &\le \alpha_2\left|\tilde{\sigma}_{jt}^{XX}-\sigma_{jt}^{XX}\right| + (1-\alpha_2)\left|\sigma_{jt}^{XX}\right| \\ &\le \alpha_2\left|\tilde{\sigma}_{jt}^{XX}-\sigma_{jt}^{XX}\right| + 1-\alpha_2 \\ &\le \alpha_2\nu_1\sqrt{\log p/n_{XX}} + 1-\alpha_2.\end{aligned}$$

Therefore, by the union bound, there exists a constant ν_1' such that

$$P\left(\left\|\hat{\mathbf{\Sigma}}_{XX}-\mathbf{\Sigma}_{XX}\right\|_\infty \ge v_1'\sqrt{\frac{\log p}{n_{XX}}}\right) \le \frac{4}{p}.$$

Similarly, it holds with probability at least $1-4/(pq)^2$ that

$$\begin{aligned}\left|\hat{\sigma}_{jk}^{XY}-\sigma_{jt}^{XY}\right| = \left|\alpha_3\tilde{\sigma}_{jt}^{XY}-\sigma_{jt}^{XY}\right| &\le \alpha_3\left|\tilde{\sigma}_{jt}^{XY}-\sigma_{jt}^{XY}\right| + (1-\alpha_3)\left|\sigma_{jt}^{XY}\right| \\ &\le \alpha_3\left|\tilde{\sigma}_{jt}^{XY}-\sigma_{jt}^{XY}\right| + 1-\alpha_3 \\ &\le \alpha_3\nu_2\sqrt{\log(pq)/n_{XY}} + 1-\alpha_3.\end{aligned}$$

Therefore, by the union bound, there exists a constant ν_2' such that

$$P\left(\left\|\hat{\mathbf{\Sigma}}_{XY}-\mathbf{\Sigma}_{XY}\right\|_\infty \ge v_2'\sqrt{\frac{\log(pq)}{n_{XY}}}\right) \le \frac{4}{pq}.$$

Let $\delta_2 = 8\sqrt{6}(1 + 4\frac{L_2^2}{\min_j(\mathrm{var}(Y_j))})\sqrt{\frac{\log q}{n_{jk}^{YY}}}\max_j(\mathrm{var}(Y_j))$. If $n_{YY} > 6\log q$, we have

$$\delta_2 < 8\left(1 + 4\frac{L_2^2}{\min_j\left(\mathrm{var}(Y_j)\right)}\right)\max_j\left(\mathrm{var}(Y_j)\right).$$

By choosing $\nu_3' = 8\sqrt{6}(1 + 4\frac{L_2^2}{\min_j(\mathrm{var}(Y_j))})\max_j(\mathrm{var}(Y_j))$, it follows from Lemma S9.2 that

$$P\left(\left|\hat{\sigma}_{jk}^{YY} - \sigma_{jk}^{YY}\right| \geq \delta_2\right) \leq 4\exp\left\{-\frac{\nu_3'^2\log q}{128\left(1 + 4\frac{L_2^2}{\min_j(\mathrm{var}(Y_j))}^2\right)^2}\right\} \leq \frac{4}{q^3}.$$

Hence, under Conditions A1 and A2, we have

$$\max_{j,k} P\left(\left|\hat{\sigma}_{jk}^{YY} - \sigma_{jk}^{YY}\right| \geq \nu_3'\sqrt{\frac{\log q}{n_{jk}^{YY}}}\right) \leq \tfrac{4}{q^3},$$

where ν_3' is a positive constant. By the union bound, we have

$$P\left(\|\hat{\mathbf{\Sigma}}_{YY} - \mathbf{\Sigma}_{YY}\|_\infty \geq \nu_3'\sqrt{\frac{\log q}{n_{YY}}}\right) \leq \tfrac{4}{q}.$$

S5 Proof of Lemma 3.2

We use a similar argument as the proof of Theorem 2 in Yu et al. (2020). By Theorem 2 in Yu et al. (2020), we have $\|\hat{\mathbf{B}}_i - \mathbf{B}_i^*\|_2 = O_p(\sqrt{s_B}\lambda_B)$. In order to prove the ℓ_2–error bound, we only need to prove $\|\hat{\mathbf{\Sigma}}_{XY,i} - \hat{\mathbf{\Sigma}}_{XX}\hat{\mathbf{B}}_i\|_\infty \leq \lambda_B$, where $\hat{\mathbf{\Sigma}}_{XY,i}$ and $\hat{\mathbf{B}}_i$ are the ith column of $\hat{\mathbf{\Sigma}}_{XY}$ and $\hat{\mathbf{B}}$, respectively.

Let $\mathbf{\Delta}^{XX} = \hat{\mathbf{\Sigma}}_{XX} - \mathbf{\Sigma}_{XX}$ and $\mathbf{\Delta}^{XY} = \hat{\mathbf{\Sigma}}_{XY} - \mathbf{\Sigma}_{XY}$. Let $\mathbf{\Delta}_i^{XX}$ and $\mathbf{\Delta}_i^{XY}$ be the ith column of $\mathbf{\Delta}^{XX}$ and $\mathbf{\Delta}^{XY}$, respectively. We have

$$\begin{aligned}
&\left\|\hat{\mathbf{\Sigma}}_{XY,i} - \hat{\mathbf{\Sigma}}_{XX}\mathbf{B}_i^*\right\|_\infty \\
=&\left\|\mathbf{\Delta}_i^{XY} - \mathbf{\Delta}^{XX}\mathbf{B}_i^*\right\|_\infty \\
\leq&\left\|\mathbf{\Delta}_i^{XY}\right\|_\infty - \left\|\mathbf{\Delta}^{XX}\right\|_\infty \|\mathbf{B}_i^*\|_{L_1} \\
\leq&(\|\mathbf{B}^*\|_{L_1} v_1' + v_3')\sqrt{\frac{\log(pq)}{\min(n_{XX}, n_{XY})}} \\
\lesssim&\lambda_{B_0}.
\end{aligned}$$

Denote the ith column of $\hat{\mathbf{B}}_0$ as $\hat{\mathbf{B}}_{0,i}$. By Theorem 2 in Yu et al. (2020), we have

$$\left\|\hat{\mathbf{B}}_{0,i} - \mathbf{B}_i^*\right\|_F^2 = O_p\left(s_B \left\|\hat{\mathbf{\Sigma}}_{XY,i} - \hat{\mathbf{\Sigma}}_{XX}\mathbf{B}_i^*\right\|_\infty^2\right) = O_p\left(\|\mathbf{B}_i^*\|_1^2 \, s_B \frac{\log(pq)}{\min(n_{XX}, n_{XY})}\right).$$

Adding all q columns together, we have

$$\left\|\hat{\mathbf{B}}_0 - \mathbf{B}^*\right\|_F = O\left(\|\mathbf{B}^*\|_{L_1} \sqrt{\frac{s_B q \log(pq)}{\min(n_{XX}, n_{XY})}}\right).$$

S6 Proof of Lemma 3.3

We first verify the RSC conditions of the objective function, see (S9.29) and (S9.30) in Theorem S9.1. Then we use Theorem 1 of Loh and Wainwright (2015) to prove the convergence rate.

Recall that $\hat{\mathbf{\Sigma}}_0 = \tilde{\mathbf{\Sigma}}_{YY} - 2\hat{\mathbf{\Sigma}}_{XY}^\top \hat{\mathbf{B}}_0 + \hat{\mathbf{B}}_0^\top \hat{\mathbf{\Sigma}}_{XX} \hat{\mathbf{B}}_0$. Let $\mathcal{L}_n(\mathbf{C}) = \mathrm{tr}(\hat{\mathbf{\Sigma}}_0 \mathbf{C}) - \log\det(\mathbf{C})$. Its Hessian matrix is $\nabla^2 \mathcal{L}_n(\mathbf{C}) = (\mathbf{C} \otimes \mathbf{C})^{-1}$.

For any $\mathbf{\Delta}^{C_0}$ such that $\|\mathbf{\Delta}^{C_0}\|_F \leq 1$. By Mean Value Theorem, there exists some $t \in [0,1]$ such that

$$
\begin{aligned}
&\left\langle \nabla \mathcal{L}_n(\mathbf{C}^* + \mathbf{\Delta}^{C_0}) - \nabla \mathcal{L}_n\left(\mathbf{C}^*\right), \mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle \\
=& \mathrm{vec}(\mathbf{\Delta}^{C_0})^\top \left(\nabla^2 \mathcal{L}_n\left(\mathbf{C}^* + t\mathbf{\Delta}^{C_0}\right) \right) \mathrm{vec}(\mathbf{\Delta}^{C_0}) \\
\geq& \lambda_{\min}\left(\nabla^2 \mathcal{L}_n\left(\mathbf{C}^* + t\mathbf{\Delta}^{C_0}\right) \right) \|\mathbf{\Delta}^{C_0}\|_F^2 \\
=& \left\| \mathbf{C}^* + t\mathbf{\Delta}^{C_0} \right\|_2^{-2} \|\mathbf{\Delta}^{C_0}\|_F^2 \\
\geq& \left(\|\mathbf{C}^*\|_2 + t\|\mathbf{\Delta}^{C_0}\|_2 \right)^{-2} \|\mathbf{\Delta}^{C_0}\|_F^2 \\
\geq& \left(\|\mathbf{C}^*\|_2 + 1 \right)^{-2} \|\mathbf{\Delta}^{C_0}\|_F^2.
\end{aligned}
$$

Thus, (S9.29) holds. Moreover, since $\mathcal{L}_n$ is convex, the function $f(t) := \mathcal{L}_n(\mathbf{C}^* + t\mathbf{\Delta}^{C_0})$ is also convex. So, $f'(1) - f'(0) \geq f'(t) - f'(0)$ for all $t \in [0,1]$. Since

$$
\begin{aligned}
f'(1) - f'(0) =& \left\langle \nabla \mathcal{L}_n\left(\mathbf{C}^* + \mathbf{\Delta}^{C_0}\right), \mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle - \left\langle \nabla \mathcal{L}_n\left(\mathbf{C}^*\right), \mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle \\
=& \left\langle \nabla \mathcal{L}_n\left(\mathbf{C}^* + \mathbf{\Delta}^{C_0}\right) - \nabla \mathcal{L}_n\left(\mathbf{C}^*\right), \mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle, \\
f'(t) - f'(0) =& \left\langle \nabla \mathcal{L}_n\left(\mathbf{C}^* + t\mathbf{\Delta}^{C_0}\right), \mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle - \left\langle \nabla \mathcal{L}_n\left(\mathbf{C}^*\right), \mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle \\
=& \frac{1}{t} \left\langle \nabla \mathcal{L}_n\left(\mathbf{C}^* + t\mathbf{\Delta}^{C_0}\right) - \nabla \mathcal{L}_n\left(\mathbf{C}^*\right), t\,\mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle,
\end{aligned}
$$

we have

$$
\left\langle \nabla \mathcal{L}_n\left(\mathbf{C}^* + \mathbf{\Delta}^{C_0}\right) - \nabla \mathcal{L}_n\left(\mathbf{C}^*\right), \mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle \geq \frac{1}{t} \left\langle \nabla \mathcal{L}_n\left(\mathbf{C}^* + t\mathbf{\Delta}^{C_0}\right) - \nabla \mathcal{L}_n\left(\mathbf{C}^*\right), t\,\mathrm{vec}\left(\mathbf{\Delta}^{C_0}\right) \right\rangle.
$$

For any $\|\mathbf{\Delta}^{C_0}\|_F \geq 1$, take $t = \frac{1}{\|\mathbf{\Delta}^{C_0}\|_F} \in (0, 1]$. Since $\|t\mathbf{\Delta}^{C_0}\|_F = 1$, we have

$$\begin{aligned}
&\left\langle \nabla \mathcal{L}_n \left(\mathbf{C}^* + \mathbf{\Delta}^{C_0}\right) - \nabla \mathcal{L}_n \left(\mathbf{C}^*\right), \operatorname{vec}\left(\mathbf{\Delta}^{C_0}\right)\right\rangle \\
\geq &\|\mathbf{\Delta}^{C_0}\|_F \left\langle \nabla \mathcal{L}_n \left(\mathbf{C}^* + \frac{\mathbf{\Delta}^{C_0}}{\|\mathbf{\Delta}^{C_0}\|_F}\right) - \nabla \mathcal{L}_n \left(\mathbf{C}^*\right), \operatorname{vec}\left(\frac{\mathbf{\Delta}^{C_0}}{\|\mathbf{\Delta}^{C_0}\|_F}\right)\right\rangle \\
\geq &\|\mathbf{\Delta}^{C_0}\|_F \left(\|\mathbf{C}^*\|_2 + 1\right)^{-2}.
\end{aligned}$$

Thus, (S9.30) holds. Denote $\mathbf{\Delta}^{XX} = \mathbf{\Sigma}_{XX} - \hat{\mathbf{\Sigma}}_{XX}$, $\mathbf{\Delta}^{XY} = \mathbf{\Sigma}_{XY} - \hat{\mathbf{\Sigma}}_{XY}$ and $\mathbf{\Delta}^{YY} = \mathbf{\Sigma}_{YY} - \hat{\mathbf{\Sigma}}_{YY}$. Theorem 3.1 implies that with probability at least $1 - \frac{4}{p} - \frac{4}{pq} - \frac{4}{q}$, we have

$$\|\mathbf{\Delta}^{XX}\|_\infty \leq v_1' \sqrt{\frac{\log p}{n_{XX}}}, \quad \|\mathbf{\Delta}^{XY}\|_\infty \leq v_2' \sqrt{\frac{\log(pq)}{n_{XY}}}, \quad \|\mathbf{\Delta}^{YY}\|_\infty \leq v_3' \sqrt{\frac{\log q}{n_{YY}}}.$$

Then, we have

$$\begin{aligned}
&\left\|\nabla \mathcal{L}_n \left(\mathbf{C}^*\right)\right\|_\infty \\
= &\left\|\mathbf{\Sigma}_\epsilon - \hat{\mathbf{\Sigma}}_0\right\|_\infty \\
\leq &\|\mathbf{\Sigma}_\epsilon - \hat{\mathbf{\Sigma}}_{YY} + 2\hat{\mathbf{\Sigma}}_{XY}^\top \mathbf{B}^* - \mathbf{B}^{*\top} \hat{\mathbf{\Sigma}}_{XX} \mathbf{B}^*\|_\infty + \left\|\hat{\mathbf{\Sigma}}_{YY} - 2\hat{\mathbf{\Sigma}}_{XY}^\top \mathbf{B}^* + \mathbf{B}^{*\top} \hat{\mathbf{\Sigma}}_{XX} \mathbf{B}^* - \right. \\
&\left. \left(\hat{\mathbf{\Sigma}}_{YY} - 2\hat{\mathbf{\Sigma}}_{XY}^\top \hat{\mathbf{B}}_0 + \hat{\mathbf{B}}_0^\top \hat{\mathbf{\Sigma}}_{XX} \hat{\mathbf{B}}_0\right)\right\|_\infty \\
\leq &2\|\mathbf{B}^* - \hat{\mathbf{B}}_0\|_{L_1} \|\hat{\mathbf{\Sigma}}^{XY}\|_\infty + 2\|\mathbf{B}^* - \hat{\mathbf{B}}_0\|_{L_1} \|\mathbf{B}^*\|_{L_1} \|\hat{\mathbf{\Sigma}}^{XX}\|_\infty + \|\mathbf{B}^* - \hat{\mathbf{B}}_0\|_{L_1} \\
&\|\mathbf{B}^* - \hat{\mathbf{B}}_0\|_{L_1} \|\hat{\mathbf{\Sigma}}^{XX}\|_\infty + \|\mathbf{\Delta}^{YY}\|_\infty + 2\|\mathbf{B}^*\|_{L_1} \|\mathbf{\Delta}^{XY}\|_\infty + \|\mathbf{B}^*\|_{L_1}^2 \|\mathbf{\Delta}^{XX}\|_\infty \\
\lesssim &\|\mathbf{B}^*\|_{L_1}^2 \sqrt{\frac{\log(pq)}{\min(n_{XX}, n_{XY})}} + \|\mathbf{B}^*\|_{L_1} s_B \sqrt{\frac{q \log(pq)}{\min(n_{XX}, n_{XY})}}.
\end{aligned}$$

Then, the result follows from Theorem S9.1.

S7 Proof of Theorem 3.4

We first rely on verifying the RSC conditions of our loss function to express the upper bound of $\|\hat{\mathbf{B}} - \mathbf{B}^*\|_1$ as a function of $\|\hat{\mathbf{C}} - \mathbf{C}^*\|_1$; see (S7.15). Similarly, we show that the upper bound of $\|\hat{\mathbf{C}} - \mathbf{C}^*\|_1$ can also be expressed as a function of $\|\hat{\mathbf{B}} - \mathbf{B}^*\|_1$; see (S7.18). Combining these two results with some algebra proves the theorem.

For $\mathcal{L}_n(\mathbf{B}, \mathbf{C}) = \mathrm{tr}[\mathbf{C}\hat{\mathbf{\Sigma}}_{\mathbf{YY}} + \mathbf{B}\mathbf{C}\mathbf{B}^\top\hat{\mathbf{\Sigma}}_{\mathbf{XX}} - \mathbf{2}\mathbf{C}\mathbf{B}^\top\hat{\mathbf{\Sigma}}_{\mathbf{XY}}] - \log\det(\mathbf{C})$, we have $\nabla^2_B\mathcal{L}_n(\mathbf{B}, \mathbf{C}) = 2\hat{\mathbf{\Sigma}}_{XX} \otimes \mathbf{C}$ and $\nabla^2_C\mathcal{L}_n(\mathbf{B}, \mathbf{C}) = \mathbf{C}^{-1} \otimes \mathbf{C}^{-1}$.

For $\mathcal{L}(\mathbf{B}, \mathbf{C}) = \mathrm{tr}\left[\mathbf{C}\mathbf{\Sigma}_{\mathbf{YY}} + \mathbf{B}\mathbf{C}\mathbf{B}^\top\mathbf{\Sigma}_{\mathbf{XX}} - \mathbf{2}\mathbf{C}\mathbf{B}^\top\mathbf{\Sigma}_{\mathbf{XY}}\right] - \log\det(\mathbf{C})$, we have $\nabla^2_B\mathcal{L}(\mathbf{B}, \mathbf{C}) = 2\mathbf{\Sigma}_{XX} \otimes \mathbf{C}$ and $\nabla^2_C\mathcal{L}(\mathbf{B}, \mathbf{C}) = \mathbf{C}^{-1} \otimes \mathbf{C}^{-1}$.

Denote $\mathbf{\Delta}^B = \mathbf{B}^* - \hat{\mathbf{B}}$ and $\mathbf{\Delta}^C = \mathbf{C}^* - \hat{\mathbf{C}}$. For any $t \in [0, 1]$, denote $\hat{\mathbf{B}}^t = \mathbf{B}^* + t\mathbf{\Delta}^B$. For any vector $\mathbf{v}_{I_1} \in \mathbb{R}^{pq}$, we have

$$\begin{aligned}
&\mathbf{v}_{I_1}^\top\nabla^2_B\mathcal{L}(\hat{\mathbf{B}}^t, \hat{\mathbf{C}})\mathbf{v}_{I_1} = 2\mathbf{v}_{I_1}^\top(\mathbf{\Sigma}_{XX} \otimes \hat{\mathbf{C}})\mathbf{v}_{I_1} \\
\geq &2\|\mathbf{v}_{I_1}\|_2^2\lambda_{\min}(\mathbf{\Sigma}_{XX})\lambda_{\min}(\hat{\mathbf{C}}) \geq \lambda_{\min}(\mathbf{\Sigma}_{XX})\lambda_{\min}(\hat{\mathbf{C}})\|\mathbf{v}_{I_1}\|_2^2.
\end{aligned}$$

In addition, define

$$\begin{aligned}
\tilde{\epsilon}_n^B &= \max_{t' \in [0,1]} \left\{ \|\nabla^2_B\mathcal{L}(\hat{\mathbf{B}}^t, \hat{\mathbf{C}}) - \nabla^2_B\mathcal{L}_n(\hat{\mathbf{B}}^t, \hat{\mathbf{C}})\|_\infty \right\} \\
&= \left\| 2\mathbf{\Delta}^{XX} \otimes \hat{\mathbf{C}} \right\|_\infty,
\end{aligned}$$

where $\boldsymbol{\Delta}^{XX} = \hat{\boldsymbol{\Sigma}}_{XX} - \boldsymbol{\Sigma}_{XX}$. Then, we have

$$\begin{aligned}
&\frac{\mathbf{v}_{I_1}^\top \nabla_B^2 \mathcal{L}_n(\hat{\mathbf{B}}^t, \hat{\mathbf{C}}) \mathbf{v}_{I_1}}{\|\mathbf{v}_{I_1}\|_2^2} \\
=&\frac{\mathbf{v}_{I_1}^\top \nabla_B^2 \mathcal{L}(\hat{\mathbf{B}}^t, \hat{\mathbf{C}}) \mathbf{v}_{I_1}}{\|\mathbf{v}_{I_1}\|_2^2} + \frac{\mathbf{v}_{I_1}^\top (\nabla_B^2 \mathcal{L}_n(\hat{\mathbf{B}}^t, \hat{\mathbf{C}}) - \nabla_B^2 \mathcal{L}(\hat{\mathbf{B}}^t, \hat{\mathbf{C}})) \mathbf{v}_{I_1}}{\|\mathbf{v}_{I_1}\|_2^2} \\
\geq&\alpha_B - \frac{\tilde{\epsilon}_n^B \|\mathbf{v}_{I_1}\|_1^2}{\|\mathbf{v}_{I_1}\|_2^2},
\end{aligned}$$

where $\alpha_B = \lambda_{\min}(\boldsymbol{\Sigma}_{XX})(\lambda_{\min}(\mathbf{C}^*) - \lambda_{\min}(\boldsymbol{\Delta}^C))$.

Let $\boldsymbol{\delta}_B = \mathrm{vec}(\boldsymbol{\Delta}^B)$ and $\boldsymbol{\delta}_C = \mathrm{vec}(\boldsymbol{\Delta}^C)$. Then, we have

$$\begin{aligned}
&\left\langle \boldsymbol{\delta}_B, \mathrm{vec}\left(\nabla_B \mathcal{L}_n(\hat{\mathbf{B}}, \hat{\mathbf{C}}) - \nabla_B \mathcal{L}_n(\mathbf{B}^*, \hat{\mathbf{C}}) \right) \right\rangle \\
=&\left\langle \boldsymbol{\delta}_B, \mathrm{vec}\left(\int_0^1 \nabla_B^2 \mathcal{L}_n \left(\mathbf{B}^* + t(\hat{\mathbf{B}} - \mathbf{B}^*), \hat{\mathbf{C}} \right) \boldsymbol{\Delta}^B dt \right) \right\rangle \\
\geq&\langle \boldsymbol{\delta}_B, \alpha_B \boldsymbol{\delta}_B \rangle - \tilde{\epsilon}_n^B \|\boldsymbol{\delta}_B\|_1^2 \\
=&\alpha_B \|\boldsymbol{\delta}_B\|_2^2 - \tilde{\epsilon}_n^B \|\boldsymbol{\delta}_B\|_1^2.
\end{aligned} \tag{S7.1}$$

For any matrix $\mathbf{B} = (B_{ij}) \in \mathbb{R}^{p \times q}$, define $f_1(\mathbf{B}) = (b_{ij})$, where $b_{ij} = 1$ if $B_{ij} > 0$, $b_{ij} = -1$ if $B_{ij} < 0$ and $b_{ij} = 0$ if $B_{ij} = 0$. Similarly, for any matrix $\mathbf{C} = (C_{ij}) \in \mathbb{R}^{q \times q}$, define $f_2(\mathbf{C}) = (c_{ij})$, where $c_{ij} = 1$ if $C_{ij} > 0$, $c_{ij} = -1$ if $C_{ij} < 0$ and $c_{ij} = 0$ if $C_{ij} = 0$. Then $f_1(\mathbf{B}) \in \nabla_B(\|\mathbf{B}\|_1)$ and $f_2(\mathbf{C}) \in \nabla_C(\|\mathbf{C}\|_1)$. Since $\hat{\mathbf{B}}$ is a stationary point of $\mathcal{L}_n + \lambda_B \|\mathbf{B}\|_1$ and $\hat{\mathbf{C}}$ is a stationary point of $\mathcal{L}_n + \lambda_C \|\mathbf{C}\|_1$, we have

$$\langle \mathrm{vec}(\nabla_B \mathcal{L}_n(\hat{\mathbf{B}}, \hat{\mathbf{C}}) + \lambda_B f_1(\hat{\mathbf{B}})), \boldsymbol{\delta}_B \rangle \geq 0, \tag{S7.2}$$

and

$$\langle \mathrm{vec}(\nabla_C \mathcal{L}_n(\hat{\mathbf{B}}, \hat{\mathbf{C}}) + \lambda_C f_2(\hat{\mathbf{C}})), \boldsymbol{\delta}_C \rangle \geq 0. \tag{S7.3}$$

By (S7.1) and (S7.2), we have

$$\begin{aligned}
&\alpha_B\|\boldsymbol{\delta}_B\|_2^2 - \tilde{\epsilon}_n^B\|\boldsymbol{\delta}_B\|_1^2 \\
\leq&\langle \mathrm{vec}(\nabla_B\mathcal{L}_n(\hat{\mathbf{B}},\hat{\mathbf{C}}) - \nabla_B\mathcal{L}_n(\mathbf{B}^*,\hat{\mathbf{C}})),\boldsymbol{\delta}_B\rangle \\
=&\langle \mathrm{vec}(\nabla_B\mathcal{L}_n(\hat{\mathbf{B}},\hat{\mathbf{C}})),\boldsymbol{\delta}_B\rangle - \langle \mathrm{vec}(\nabla_B\mathcal{L}_n(\mathbf{B}^*,\hat{\mathbf{C}})),\boldsymbol{\delta}_B\rangle \qquad \text{(S7.4)}\\
\leq&\langle \mathrm{vec}(\nabla_B(\lambda_B\|\hat{\mathbf{B}}\|_1 + \lambda_C\|\hat{\mathbf{C}}\|_1)),\boldsymbol{\delta}_B\rangle - \langle \mathrm{vec}(\nabla_B\mathcal{L}_n(\mathbf{B}^*,\hat{\mathbf{C}})),\boldsymbol{\delta}_B\rangle \\
\leq&\lambda_B\|\mathbf{B}^*\|_1 - \lambda_B\|\hat{\mathbf{B}}\|_1 + \|\nabla_B\mathcal{L}_n(\mathbf{B}^*,\hat{\mathbf{C}})\|_\infty\|\boldsymbol{\delta}_B\|_1.
\end{aligned}$$

Define

$$\begin{aligned}
\tilde{\lambda}_B =& C_\lambda(\log p)^{1/2}/\min(n_{XX}^{1-\tau_1/2}, n_{XY}^{1-\tau_2/2})(\|\mathbf{B}^*\mathbf{C}^*\|_{L_1} + \|\mathbf{B}^*\|_{L_1}\|\boldsymbol{\delta}_C\|_1) \\
&+ C_\lambda \max\{\lambda_{\max}(\mathbf{C}^*), 1/\lambda_{\min}(\mathbf{C}^*)\}\{\frac{\log(pq)}{n_{XY}}\}^{1/2}(1+\|\boldsymbol{\delta}_C\|_1),
\end{aligned} \qquad \text{(S7.5)}$$

where C_λ is a constant only depending on $\lambda_{\max}(\mathbf{C}^*), 1/\lambda_{\min}(\mathbf{C}^*), L_1, L_2$. Then with a large enough constant C_λ, we have $\lambda_B < \tilde{\lambda}_B$. By Lemma S9.5, we have

$$\|\nabla_B\mathcal{L}_n(\mathbf{B}^*,\hat{\mathbf{C}})\|_\infty \lesssim \tilde{\lambda}_B. \qquad \text{(S7.6)}$$

By Lemma S9.3, we have

$$\begin{aligned}\|\boldsymbol{\delta}_B\|_1 &= \left\|\boldsymbol{\Delta}^B\right\|_1 \\ &= \left\|\mathrm{vec}\left(\boldsymbol{\Delta}^B\right)_{S_B}\right\|_1 + \left\|\mathrm{vec}\left(\boldsymbol{\Delta}^B\right)_{S_B^C}\right\|_1 \\ &\lesssim 4\left\|\mathrm{vec}\left(\boldsymbol{\Delta}^B\right)_{S_B}\right\|_1 \\ &\lesssim 4\sqrt{s_B}\left\|\mathrm{vec}\left(\boldsymbol{\Delta}^B\right)_{S_B}\right\|_2 \\ &\lesssim \sqrt{s_B}\left\|\boldsymbol{\delta}_B\right\|_2 .\end{aligned} \tag{S7.7}$$

Then by (S7.4), (S7.6) and (S7.7), it holds with probability at least $1-\frac{4}{p}-\frac{4}{pq}$ that

$$\begin{aligned}&\left\{\alpha_B - 16\tilde{\epsilon}_n^B s_B\right\}\|\boldsymbol{\delta}_B\|_2^2 \\ \leq&\lambda_B\|\mathbf{B}^*\|_1 - \lambda_B\|\hat{\mathbf{B}}\|_1 + \|\nabla_B\mathcal{L}_n(\mathbf{B}^*,\hat{\mathbf{C}})\|_\infty\|\boldsymbol{\delta}_B\|_1 \\ \lesssim&\lambda_B\|\mathbf{B}^*\|_1 - \lambda_B\|\hat{\mathbf{B}}\|_1 + \tilde{\lambda}_B\|\boldsymbol{\delta}_B\|_1 \\ \lesssim&\lambda_B\|\mathbf{B}^*\|_1 - \lambda_B\|\hat{\mathbf{B}}\|_1 + \tilde{\lambda}_B\|\,\mathrm{vec}\left(\boldsymbol{\Delta}^B\right)_{S_B}\|_1 \\ \lesssim&\tilde{\lambda}_B\|\,\mathrm{vec}(\boldsymbol{\Delta}^B)_{S_B}\|_1 - \lambda_B\|\,\mathrm{vec}(\hat{\mathbf{B}})_{S_B^C}\|_1 \\ \lesssim&\tilde{\lambda}_B\sqrt{s_B}\|\boldsymbol{\delta}_B\|_2.\end{aligned} \tag{S7.8}$$

Next we show that with large enough n_{XX} and q, $\alpha_B - 16\tilde{\epsilon}_n^B s_B$ is bounded away from 0. To show α_B is bounded away from 0, we first prove that $\mathcal{L}_n(\mathbf{B},\mathbf{C})$ satisfies the RSC condition (S9.29) and (S9.30) with respect to $\mathbf{C}$ for any $\mathbf{B}$.

For any $t' \in [0,1]$, denote $\hat{\mathbf{C}}^{t'} = \mathbf{C}^* + t'\boldsymbol{\Delta}^C$. For any vector $\mathbf{v}_{I_2} \in \mathbb{R}^{q^2}$,

we have

$$\begin{aligned}
&\mathbf{v}_{I_2}^\top \nabla_C^2 \mathcal{L}(\hat{\mathbf{B}}, \hat{\mathbf{C}}^{t'}) \mathbf{v}_{I_2} \\
=&\mathbf{v}_{I_2}^\top ((\hat{\mathbf{C}}^{t'})^{-1} \otimes (\hat{\mathbf{C}}^{t'})^{-1}) \mathbf{v}_{I_2} \\
\geq&(\|\mathbf{C}^*\|_2 + t' \|\boldsymbol{\Delta}^C\|_2)^{-2} \|\mathbf{v}_{I_2}\|_2^2,
\end{aligned}$$

where we use the Weyl's inequality that $\lambda_{\max}(\mathbf{C}^*) - t'\lambda_{\max}(\boldsymbol{\Delta}^{\mathbf{C}}) \geq \lambda_{\max}(\hat{\mathbf{C}}^{t'})$. Then, for all $\|\boldsymbol{\Delta}^C\|_F \leq 1$ and any $\mathbf{B}$, we have

$$\frac{\mathbf{v}_{I_2}^\top \nabla_C^2 \mathcal{L}_n(\mathbf{B}, \hat{\mathbf{C}}^{t'}) \mathbf{v}_{I_1}}{\|\mathbf{v}_{I_2}\|_2^2} \geq \quad (\|\mathbf{C}^*\|_2 + t' \|\boldsymbol{\Delta}^C\|_2)^{-2} \geq (\|\mathbf{C}^*\|_2 + 1)^{-2}.$$

Then, for any $\|\boldsymbol{\Delta}^C\|_F \leq 1$ and any $\mathbf{B}$ we have

$$\begin{aligned}
&\langle \boldsymbol{\delta}_C, \mathrm{vec}(\nabla_C \mathcal{L}_n(\mathbf{B}, \hat{\mathbf{C}}) - \nabla_C \mathcal{L}_n(\mathbf{B}, \mathbf{C}^*)) \rangle \\
=&\left\langle \boldsymbol{\delta}_C, \mathrm{vec}\left(\int_0^1 \nabla_C^2 \mathcal{L}_n \left(\mathbf{B}, \mathbf{C}^* + t'(\hat{\mathbf{C}} - \mathbf{C}^*) \right) \boldsymbol{\Delta}^C dt \right) \right\rangle \qquad \text{(S7.9)} \\
\geq&\alpha_C \|\boldsymbol{\delta}_C\|_2^2,
\end{aligned}$$

where $\alpha_C = (\|\mathbf{C}^*\|_2 + 1)^{-2}$. If $\|\boldsymbol{\Delta}^C\|_F > 1$, since $\mathcal{L}_n(\mathbf{B}, \mathbf{C})$ is convex with respect to $\mathbf{C}$, the function $f : [0, 1] \to \mathbb{R}$ given by $f(t) := \mathcal{L}_n(\mathbf{B}, \mathbf{C}^* + t'\boldsymbol{\Delta}^C)$ is also convex, so $f'(1) - f'(0) \geq f'(t') - f'(0)$ for all $t' \in [0, 1]$. Computing the derivatives of f yields

$$\begin{aligned}
&\langle \mathrm{vec}(\nabla_C \mathcal{L}_n(\mathbf{B}, \hat{\mathbf{C}}) - \nabla_C \mathcal{L}_n(\mathbf{B}, \mathbf{C}^*)), \boldsymbol{\delta}_C \rangle \\
\geq&\frac{1}{t'} \left\langle \mathrm{vec}(\nabla \mathcal{L}_n \left(\mathbf{B}, \mathbf{C}^* + t'\boldsymbol{\Delta}^C \right) - \nabla \mathcal{L}_n \left(\mathbf{B}, \mathbf{C}^* \right)), t'\boldsymbol{\delta}_C \right\rangle.
\end{aligned}$$

Taking $t' = \frac{1}{\|\boldsymbol{\Delta}^C\|_F} \in (0, 1]$, for any $\|\boldsymbol{\Delta}^C\|_F > 1$ and any $\mathbf{B}$, we have

$$\langle \mathrm{vec}(\nabla_C \mathcal{L}_n(\mathbf{B}, \hat{\mathbf{C}}) - \nabla_C \mathcal{L}_n(\mathbf{B}, \mathbf{C}^*)), \boldsymbol{\delta}_C \rangle \geq \alpha_C \|\boldsymbol{\delta}_C\|_2. \tag{S7.10}$$

Combining (S7.9) and (S7.10), we show that $\mathcal{L}_n(\mathbf{B}, \mathbf{C})$ satisfies the RSC conditions (S9.29) and (S9.30) with respect to $\mathbf{C}$ for any $\mathbf{B}$. Next, following the proof of Lemma S9.1 from Loh and Wainwright (2015), we can prove $\|\boldsymbol{\delta}_C\|_2 \leq 3(\|\mathbf{C}^*\|_2 + 1)^2/2$. For completeness, we prove it as follows.

By (S7.3) and (S7.10), we have

$$\left\langle \mathrm{vec}(-\lambda_C f_2(\hat{\mathbf{C}}) - \nabla_C \mathcal{L}_n(\hat{\mathbf{B}}, \mathbf{C}^*)), \boldsymbol{\delta}_C \right\rangle \geq \alpha_C \|\boldsymbol{\delta}_C\|_2.$$

By Hölder's inequality and the triangle inequality, we also have

$$\left\langle \mathrm{vec}(-\lambda_C f_2(\hat{\mathbf{C}}) - \nabla_C \mathcal{L}_n(\hat{\mathbf{B}}, \mathbf{C}^*)), \boldsymbol{\delta}_C \right\rangle \leq \frac{3}{2} \lambda_C \|\boldsymbol{\delta}_C\|_1.$$

Combining the above two inequalities yields

$$\|\boldsymbol{\delta}_C\|_2 \leq \frac{3\|\boldsymbol{\delta}_C\|_1 \lambda_C}{2\alpha_C} \leq \frac{3R\lambda_C}{2\alpha_C}. \tag{S7.11}$$

With our choice of λ_C and R, and large enough n_{XX}, n_{XY}, n_{YY}, we have $\|\boldsymbol{\delta}_C\|_2 \leq 3(\|\mathbf{C}^*\|_2+1)^2/2$. Since $\sqrt{\sum_{i=1}^q |\lambda_i(\boldsymbol{\Delta}^C)|^2} \leq \|\boldsymbol{\Delta}^C\|_F$, where $\lambda_i(\boldsymbol{\Delta}^C)$ denotes all the q eigenvalues of $\boldsymbol{\Delta}^C$, we have $\lambda_{\min}(\boldsymbol{\Delta}^C) \leq \frac{3(\|\mathbf{C}^*\|_2+1)^2}{2q}$. Thus with large enough q, $\alpha_B = \lambda_{\min}(\boldsymbol{\Sigma}_{XX})(\lambda_{\min}(\mathbf{C}^*) - \lambda_{\min}(\boldsymbol{\Delta}^C))$ is bounded away from 0 by Condition A4.

Denote $\mathbf{\Delta}^{YY} = \mathbf{\Sigma}_{YY} - \hat{\mathbf{\Sigma}}_{YY}$. Theorem 3.1 implies that with probability at least $1 - \frac{4}{p} - \frac{4}{pq} - \frac{4}{q}$, we have

$$\|\mathbf{\Delta}^{XX}\|_\infty \leq v_1' \sqrt{\frac{\log p}{n_{XX}}}; \tag{S7.12}$$

$$\|\mathbf{\Delta}^{XY}\|_\infty \leq v_2' \sqrt{\frac{\log(pq)}{n_{XY}}}; \tag{S7.13}$$

$$\|\mathbf{\Delta}^{YY}\|_\infty \leq v_3' \sqrt{\frac{\log q}{n_{YY}}}. \tag{S7.14}$$

By inequalities (S7.12), (S7.13), (S7.14) and Condition A3, with probability at least $1 - \frac{4}{p}$, it holds that

$$\tilde{\epsilon}_n^B = 2\|\mathbf{\Delta}^{XX}\|_\infty \|\hat{\mathbf{C}}\|_\infty \lesssim v_1' \left(\frac{\log p}{n_{XX}}\right)^{\frac{1}{2}-\gamma_2}.$$

Thus when n_{XX} and q are large enough, $\alpha_B - 16\tilde{\epsilon}_n^B s_B$ is bounded away from 0. Then by (S7.8), it holds with probability at least $1 - \frac{4}{p} - \frac{4}{pq}$ that

$$\|\boldsymbol{\delta}_B\|_2 \lesssim \tilde{\lambda}_B \sqrt{s_B}.$$

By (S7.7), it holds with probability at least $1 - \frac{4}{p} - \frac{4}{pq}$ that

$$\|\boldsymbol{\delta}_B\|_1 \lesssim \sqrt{s_B}\, \|\boldsymbol{\delta}_B\|_2 \lesssim \tilde{\lambda}_B s_B, \tag{S7.15}$$

where $\tilde{\lambda}_B$ is as stated in (S7.5). Next, we show that the upper bound of $\|\mathbf{C}^* - \hat{\mathbf{C}}\|_1$ can also be expressed as a function of $\|\hat{\mathbf{B}} - \mathbf{B}^*\|_1$. By (S7.3) and (S7.9), we have

$$\alpha_C \|\boldsymbol{\delta}_C\|_2^2 \leq \lambda_C \|\mathbf{C}^*\|_1 - \lambda_C \|\hat{\mathbf{C}}\|_1 - \langle \mathrm{vec}(\nabla_C \mathcal{L}_n(\hat{\mathbf{B}}, \mathbf{C}^*)), \boldsymbol{\delta}_C \rangle.$$

By (S7.12), (S7.13) and (S7.14), it holds with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$ that

$$
\begin{aligned}
&\left\|\nabla_C \mathcal{L}_n(\hat{\mathbf{B}}, \mathbf{C}^*)\right\|_\infty \\
=&\left\|\boldsymbol{\Sigma}_\epsilon - \hat{\boldsymbol{\Sigma}}_0\right\|_\infty \\
\leq&\|\boldsymbol{\Sigma}_\epsilon - \hat{\boldsymbol{\Sigma}}_{YY} + 2\hat{\boldsymbol{\Sigma}}_{XY}^\top \mathbf{B}^* - \mathbf{B}^{*\top}\hat{\boldsymbol{\Sigma}}_{XX}\mathbf{B}^*\|_\infty + \left\|\hat{\boldsymbol{\Sigma}}_{YY} - 2\hat{\boldsymbol{\Sigma}}_{XY}^\top \mathbf{B}^* + \mathbf{B}^{*\top}\hat{\boldsymbol{\Sigma}}_{XX}\mathbf{B}^* - \right. \\
&\left.(\hat{\boldsymbol{\Sigma}}_{YY} - 2\hat{\boldsymbol{\Sigma}}_{XY}^\top \hat{\mathbf{B}} + \hat{\mathbf{B}}^\top \hat{\boldsymbol{\Sigma}}_{XX}\hat{\mathbf{B}})\right\|_\infty \\
\lesssim&\|\mathbf{B}^*\|_{L_1}^2 \sqrt{\frac{\log(pq)}{\min(n_{XX}, n_{XY})}} + \|\boldsymbol{\delta}_B\|_1.
\end{aligned}
$$

Then, with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, it holds that

$$
\|\boldsymbol{\delta}_C\|_2 \lesssim \sqrt{s_C}\|\mathbf{C}^*\|_2^2\|\mathbf{B}^*\|_{L_1}^2 \sqrt{\frac{\log(pq)}{\min(n_{XX}, n_{XY})}} + \sqrt{s_C}\|\mathbf{C}^*\|_2^2\|\boldsymbol{\delta}_B\|_1. \quad \text{(S7.16)}
$$

By Lemma S9.3, we have

$$
\begin{aligned}
\|\boldsymbol{\delta}_C\|_1 =& \left\|\operatorname{vec}\left(\boldsymbol{\Delta}^C\right)_{S_C}\right\|_1 + \left\|\operatorname{vec}\left(\boldsymbol{\Delta}^C\right)_{S_C^C}\right\|_1 \\
\lesssim& 4\left\|\operatorname{vec}\left(\boldsymbol{\Delta}^C\right)_{S_C}\right\|_1 \\
\lesssim& \sqrt{s_C}\left\|\boldsymbol{\delta}_C\right\|_2 .
\end{aligned} \quad \text{(S7.17)}
$$

Finally, we combine (S7.15), (S7.16) and (S7.17) to show the upper bounds of the estimation errors of $\hat{\mathbf{C}}$ and $\hat{\mathbf{B}}$.

By (S7.5), (S7.16) and (S7.17, with large enough n_{XX}, n_{XY}, it holds

with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$ that

$$
\begin{aligned}
&\|\boldsymbol{\delta}_C\|_1 \\
\lesssim & s_C\|\mathbf{C}^*\|_2^2\|\mathbf{B}^*\|_{L_1}^2\sqrt{\frac{\log(pq)}{\min(n_{XX},n_{XY})}}+s_Bs_C\|\mathbf{C}^*\|_2^2\left(\frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2},n_{XY}^{1-\tau_2/2}\right)}\right. \\
&(\|\mathbf{B}^*\mathbf{C}^*\|_{L_1}+\|\mathbf{B}^*\|_{L_1}\|\boldsymbol{\delta}_C\|_1)+\max\left\{\lambda_{\max}(\mathbf{C}^*),1/\lambda_{\min}(\mathbf{C}^*)\right\}\left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2} \\
&\left.(1+\|\boldsymbol{\delta}_C\|_1)\right) \\
\lesssim & \|\mathbf{C}^*\|_2^2s_C\left(\|\mathbf{B}^*\|_{L_1}^2+\frac{\|\mathbf{B}^*\mathbf{C}^*\|_{L_1}s_B}{\min\left(n_{XX}^{1/2-\tau_1/2},n_{XY}^{1/2-\tau_2/2}\right)}\right)\sqrt{\frac{\log(pq)}{\min(n_{XX},n_{XY})}} \\
&+\|\boldsymbol{\delta}_C\|_1s_Bs_C\|\mathbf{C}^*\|_2^2\left(\frac{(\log p)^{1/2}\|\mathbf{B}^*\|_{L_1}}{\min\left(n_{XX}^{1-\tau_1/2},n_{XY}^{1-\tau_2/2}\right)}+\max\left\{\lambda_{\max}(\mathbf{C}^*),1/\lambda_{\min}(\mathbf{C}^*)\right\}\right. \\
&\left.\left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2}\right).
\end{aligned}
$$

With large enough n_{XX},n_{XY}, we have $s_Bs_C\|\mathbf{C}^*\|_2^2(\frac{(\log p)^{1/2}\|\mathbf{B}^*\|_{L_1}}{\min\left(n_{XX}^{1-\tau_1/2},n_{XY}^{1-\tau_2/2}\right)}+\max\left\{\lambda_{\max}(\mathbf{C}^*),1/\lambda_{\min}(\mathbf{C}^*)\right\}\{\frac{\log(pq)}{n_{XY}}\}^{1/2})=o(1)$, so we have

$$
\begin{aligned}
&\|\boldsymbol{\delta}_C\|_1 \\
\lesssim & \|\mathbf{C}^*\|_2^2s_C\left(\|\mathbf{B}^*\|_{L_1}^2+\frac{\|\mathbf{B}^*\mathbf{C}^*\|_{L_1}s_B}{\min\left(n_{XX}^{1/2-\tau_1/2},n_{XY}^{1/2-\tau_2/2}\right)}\right)\sqrt{\frac{\log(pq)}{\min(n_{XX},n_{XY})}} \\
\lesssim & \lambda_Cs_C.
\end{aligned}
\tag{S7.18}
$$

By choosing large enough n_{XX},n_{XY} and n_{YY}, it holds with probability

at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$ that

$$\|\boldsymbol{\delta}_C\|_1 \lesssim 1. \tag{S7.19}$$

By (S7.16) and (S7.19), it holds with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$ that

$$\begin{aligned}
&\|\boldsymbol{\delta}_C\|_2 \\
\lesssim&\sqrt{s_C}\|\mathbf{C}^*\|_2^2\|\mathbf{B}^*\|_{L_1}^2\sqrt{\frac{\log(pq)}{\min(n_{XX},n_{XY})}}+s_B\sqrt{s_C}\|\mathbf{C}^*\|_2^2\left(\frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2},n_{XY}^{1-\tau_2/2}\right)}\right. \\
&\left.(\|\mathbf{B}^*\mathbf{C}^*\|_{L_1}+\|\mathbf{B}^*\|_{L_1}\|\boldsymbol{\delta}_C\|_1)+\max\left\{\lambda_{\max}(\mathbf{C}^*),1/\lambda_{\min}(\mathbf{C}^*)\right\}\left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2}\right. \\
&\left.(1+\|\boldsymbol{\delta}_C\|_1)\right) \\
\lesssim&\|\mathbf{C}^*\|_2^2\sqrt{s_C}\left(\|\mathbf{B}^*\|_{L_1}^2+\frac{\|\mathbf{B}^*\mathbf{C}^*\|_{L_1}s_B}{\min\left(n_{XX}^{1/2-\tau_1/2},n_{XY}^{1/2-\tau_2/2}\right)}\right)\sqrt{\frac{\log(pq)}{\min(n_{XX},n_{XY})}} \\
\lesssim&\lambda_C\sqrt{s_C}.
\end{aligned}$$

By Lemma S9.5 and (S7.19), with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, we have $\|\nabla_B\mathcal{L}_n(\mathbf{B}^*,\hat{\mathbf{C}})\|_\infty\lesssim\lambda_B$. Then by (S7.15), it holds with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$ that

$$\begin{aligned}
&\|\boldsymbol{\delta}_B\|_2^2 \\
\leq&\lambda_B\|\mathbf{B}^*\|_1-\lambda_B\|\hat{\mathbf{B}}\|_1+\|\nabla_B\mathcal{L}_n(\mathbf{B}^*,\hat{\mathbf{C}})\|_\infty\|\boldsymbol{\delta}_B\|_1 \\
\lesssim&\lambda_B\|\mathbf{B}^*\|_1-\lambda_B\|\hat{\mathbf{B}}\|_1+\lambda_B\|\boldsymbol{\delta}_B\|_1 \\
\lesssim&\lambda_B\sqrt{s_B}\|\boldsymbol{\delta}_B\|_2.
\end{aligned}$$

So with probability at least $1 - \frac{4}{p} - \frac{4}{pq} - \frac{4}{q}$, it holds that

$$\|\boldsymbol{\delta}_B\|_2 \lesssim \lambda_B \sqrt{s_B},$$

and

$$\|\boldsymbol{\delta}_B\|_1 \lesssim \lambda_B s_B.$$

This completes the proof.

S8 Proof of Theorem 3.5

We use a similar argument as the proof of Theorem 6 in Yu et al. (2020). We first transfer the objective function. Then we show the upper bounds of $\|(\boldsymbol{\Gamma}_{S_B S_B})^{-1}\|_{L_\infty}$ and $\left\|\widehat{\boldsymbol{\gamma}}_{S_B^C} - \widehat{\boldsymbol{\Gamma}}_{S_B^C S_B} \boldsymbol{\beta}^*_{S_B}\right\|_\infty$. We use them to show that $\|\hat{\boldsymbol{\beta}}_{S_B} - \boldsymbol{\beta}_{S_B}{}^*\|_\infty < \min_{j \in S_B} |\boldsymbol{\beta}^*_j|$ with probability close to 1. Then we show that $\left\|\widehat{\boldsymbol{\gamma}}_{S_B^C} - \widehat{\boldsymbol{\Gamma}}_{S_B^C S_B} \hat{\boldsymbol{\beta}}_{S_B}\right\|_\infty \leq \lambda_B$ with probability close to 1.

By properties of trace and vectorization, we can rewrite (2.5) as

$$\begin{aligned}
&(\hat{\mathbf{B}}, \hat{\mathbf{C}}) \\
&= \arg \min_{\mathbf{C} \in \mathbb{S}_+^{q \times q}, \mathbf{B}} \left\{ \operatorname{tr}\left[\mathbf{C}\hat{\boldsymbol{\Sigma}}_{YY}\right] + \operatorname{tr}\left[\mathbf{B}^\top \hat{\boldsymbol{\Sigma}}_{XX} \mathbf{B}\mathbf{C}\right] - 2\operatorname{tr}\left[\mathbf{B}^\top \hat{\boldsymbol{\Sigma}}_{XY} \mathbf{C}\right] + \lambda_B \|\mathbf{B}\|_1 + \lambda_C \|\mathbf{C}\|_1 - \log\det \mathbf{C} \right\} \\
&= \arg \min_{\mathbf{C} \in \mathbb{S}_+^{q \times q}, \mathbf{B}} \left\{ \operatorname{tr}\left[\mathbf{C}\hat{\boldsymbol{\Sigma}}_{YY}\right] + \operatorname{vec}(\mathbf{B})^\top \left(\mathbf{C} \otimes \hat{\boldsymbol{\Sigma}}_{XX}\right) \operatorname{vec}(\mathbf{B}) - 2\operatorname{vec}(\mathbf{B})^\top \operatorname{vec}\left(\hat{\boldsymbol{\Sigma}}_{XY}\mathbf{C}\right) + \lambda_B \|\mathbf{B}\|_1 + \right. \\
&\left. \lambda_C \|\mathbf{C}\|_1 - \log\det \mathbf{C} \right\}.
\end{aligned}$$

Denote $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$, (2.5) is equivalent to solving

$$(\hat{\boldsymbol{\beta}}, \hat{\mathbf{C}}) = \arg \min_{\boldsymbol{\beta}, \mathbf{C} \in \mathbb{S}_+^{q \times q}} \Big\{ \text{tr}\left[\mathbf{C}\hat{\boldsymbol{\Sigma}}_{YY}\right] - \log\det \mathbf{C} - 2\boldsymbol{\beta}^\top \text{vec}(\hat{\boldsymbol{\Sigma}}_{XY}\mathbf{C}) + \boldsymbol{\beta}^\top \left(\mathbf{C} \otimes \hat{\boldsymbol{\Sigma}}_{XX}\right)\boldsymbol{\beta} + \lambda_B \|\boldsymbol{\beta}\|_1 + \lambda_C \|\mathbf{C}\|_1 \Big\}, \tag{S8.20}$$

For an optimal solution $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{C}})$ to (S8.20), $\hat{\boldsymbol{\beta}}$ should satisfy

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \left\{ -2\boldsymbol{\beta}^\top \hat{\boldsymbol{\gamma}} + \boldsymbol{\beta}^\top \hat{\boldsymbol{\Gamma}} \boldsymbol{\beta} + \lambda_B \|\boldsymbol{\beta}\|_1 \right\}, \tag{S8.21}$$

where $\widehat{\boldsymbol{\Gamma}} = \hat{\mathbf{C}} \otimes \hat{\boldsymbol{\Sigma}}_{XX}$ and $\widehat{\boldsymbol{\gamma}} = \text{vec}(\hat{\boldsymbol{\Sigma}}_{XY}\hat{\mathbf{C}})$. This can be proved by contradiction. If $\hat{\boldsymbol{\beta}}$ does not satisfy (S8.21), let $\boldsymbol{\beta}_1$ to be a solution of (S8.21). Denote $\mathcal{L}_n(\boldsymbol{\beta}, \mathbf{C}) = \text{tr}\left[\mathbf{C}\hat{\boldsymbol{\Sigma}}_{YY}\right] - \log\det \mathbf{C} - 2\boldsymbol{\beta}^\top \text{vec}(\hat{\boldsymbol{\Sigma}}_{XY}\mathbf{C}) + \boldsymbol{\beta}^\top \left(\mathbf{C} \otimes \hat{\boldsymbol{\Sigma}}_{XX}\right)\boldsymbol{\beta} + \lambda_B \|\boldsymbol{\beta}\|_1 + \lambda_C \|\mathbf{C}\|_1$. Then

$$\begin{aligned} &\mathcal{L}_n(\hat{\boldsymbol{\beta}}, \hat{\mathbf{C}}) - \mathcal{L}_n(\boldsymbol{\beta}_1, \hat{\mathbf{C}}) \\ =& \left(2\hat{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\beta}}^\top \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\beta}} + \lambda_B \|\hat{\boldsymbol{\beta}}\|_1\right) - \left(2\boldsymbol{\beta}_1^\top \hat{\boldsymbol{\gamma}} + \boldsymbol{\beta}_1^\top \hat{\boldsymbol{\Gamma}} \boldsymbol{\beta}_1 + \lambda_B \|\boldsymbol{\beta}_1\|_1\right) \\ >& 0, \end{aligned}$$

which is a contradiction. Thus $\hat{\boldsymbol{\beta}}$ should satisfy (S8.21). Since $\hat{\mathbf{C}}$ is the optimal solution to (S8.20), it is positive definite. By our construction, $\hat{\boldsymbol{\Sigma}}_{XX}$ is also positive definite. Thus (S8.21) is a strictly convex problem, which has a unique solution. Thus $\hat{\boldsymbol{\beta}}$ is the unique solution to (S8.21).

By the Karush–Kuhn–Tucker (KKT) conditions of (S8.21), we know that $\hat{\boldsymbol{\beta}}$ is a solution to (S8.21) if there exists a subgradient $\boldsymbol{\omega}^B \in \mathbb{R}^{pq}$ such

that

$$\widehat{\boldsymbol{\gamma}} - \widehat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\beta}} = \lambda_B \boldsymbol{\omega}^B, \tag{S8.22}$$

where $\boldsymbol{\omega}_j^B = \text{sign}(\hat{\beta}_j)$ if $\hat{\beta}_j \neq 0$, and $\boldsymbol{\omega}_j^B \in [-1, 1]$ if $\hat{\beta}_j = 0$.

First, we show that for $j \in S_B$, with high probability, there exists a solution $\hat{\boldsymbol{\beta}}$ to (S8.21) s.t.

$$\left\| \hat{\boldsymbol{\beta}}_{S_B} - \boldsymbol{\beta}^*_{S_B} \right\|_\infty < \min_{j \in S_B} |\beta_j^*|,$$

where $\hat{\boldsymbol{\beta}}_{S_B}$ is the sub-vector of $\hat{\boldsymbol{\beta}}$ with indices in S_B. Then letting $\hat{\boldsymbol{\beta}}_{S_B^C} = \mathbf{0}$, we show that $\hat{\boldsymbol{\beta}}$ also satisfies the KKT conditions with high probability for $j \notin S_B$. Then, by construction, $\text{sign}(\hat{\boldsymbol{\beta}}) = \text{sign}(\boldsymbol{\beta}^*)$. Define events $\mathcal{A}_1 = \{\|\hat{\boldsymbol{\beta}}_{S_B} - \boldsymbol{\beta}^*_{S_B}\|_\infty < \min_{j \in S_B} |\beta_j^*|\}$ and $\mathcal{A}_2 = \{\|\hat{\boldsymbol{\gamma}}_{S_B^C} - \hat{\boldsymbol{\Gamma}}_{S_B^C S_B} \hat{\boldsymbol{\beta}}_{S_B}\|_\infty \leq \lambda_B\}$, where $\boldsymbol{\beta}^* = \text{vec}(\mathbf{B}^*)$. We show that $P(\mathcal{A}_1)$ and $P(\mathcal{A}_2)$ are close to 1.

Denote $V = \|(\boldsymbol{\Gamma}_{S_B S_B})^{-1}\|_{L_\infty}$, where $\boldsymbol{\Gamma} := \mathbf{C}^* \otimes \boldsymbol{\Sigma}_{XX}$. Since

$$\begin{aligned}
&\left\| \left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B} \right)^{-1} - (\boldsymbol{\Gamma}_{S_B S_B})^{-1} \right\|_{L_\infty} \\
\leq &\left\| (\boldsymbol{\Gamma}_{S_B S_B})^{-1} \right\|_{L_\infty} \left\| \left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B} \right)^{-1} \right\|_{L_\infty} \left\| \widehat{\boldsymbol{\Gamma}}_{S_B S_B} - \boldsymbol{\Gamma}_{S_B S_B} \right\|_{L_\infty} \\
\leq &\left\| (\boldsymbol{\Gamma}_{S_B S_B})^{-1} \right\|_{L_\infty} \left(\left\| (\boldsymbol{\Gamma}_{S_B S_B})^{-1} \right\|_{L_\infty} + \left\| \left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B} \right)^{-1} - (\boldsymbol{\Gamma}_{S_B S_B})^{-1} \right\|_{L_\infty} \right) \\
&\left\| \widehat{\boldsymbol{\Gamma}}_{S_B S_B} - \boldsymbol{\Gamma}_{S_B S_B} \right\|_{L_\infty} \\
= &V \left(V + \left\| \left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B} \right)^{-1} - (\boldsymbol{\Gamma}_{S_B S_B})^{-1} \right\|_{L_\infty} \right) \left\| \widehat{\boldsymbol{\Gamma}}_{S_B S_B} - \boldsymbol{\Gamma}_{S_B S_B} \right\|_{L_\infty},
\end{aligned}$$

by some algebra, we have,

$$\begin{aligned}\left\|\left(\widehat{\mathbf{\Gamma}}_{S_BS_B}\right)^{-1}-\left(\mathbf{\Gamma}_{S_BS_B}\right)^{-1}\right\|_{L_\infty} \leq & \frac{V^2\left\|\widehat{\mathbf{\Gamma}}_{S_BS_B}-\mathbf{\Gamma}_{S_BS_B}\right\|_{L_\infty}}{1-V\left\|\widehat{\mathbf{\Gamma}}_{S_BS_B}-\mathbf{\Gamma}_{S_BS_B}\right\|_{L_\infty}} \\ \leq & \frac{s_BV^2\left\|\widehat{\mathbf{\Gamma}}_{S_BS_B}-\mathbf{\Gamma}_{S_BS_B}\right\|_{\infty}}{1-s_BV\left\|\widehat{\mathbf{\Gamma}}_{S_BS_B}-\mathbf{\Gamma}_{S_BS_B}\right\|_{\infty}},\end{aligned}$$

and

$$\begin{aligned}\left\|\left(\widehat{\mathbf{\Gamma}}_{S_BS_B}\right)^{-1}\right\|_{L_\infty} \leq & V+\frac{s_BV^2\left\|\widehat{\mathbf{\Gamma}}_{S_BS_B}-\mathbf{\Gamma}_{S_BS_B}\right\|_{\infty}}{1-s_BV\left\|\widehat{\mathbf{\Gamma}}_{S_BS_B}-\mathbf{\Gamma}_{S_BS_B}\right\|_{\infty}} \\ = & \frac{V}{1-s_BV\left\|\widehat{\mathbf{\Gamma}}_{S_BS_B}-\mathbf{\Gamma}_{S_BS_B}\right\|_{\infty}}.\end{aligned}$$

By Theorem 3.4, with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, it holds that

$$\left\|\hat{\mathbf{C}}-\mathbf{C}^*\right\|_1 \lesssim \lambda_C s_C, \tag{S8.23}$$

where $\lambda_C = C\|\mathbf{C}^*\|_2^2[\|\mathbf{B}^*\|_{L_1}^2 + s_B\|\mathbf{B}^*\mathbf{C}^*\|_{L_1}/\min(n_{XX}^{1/2-\tau_1/2}, n_{XY}^{1/2-\tau_2/2})](\log(p$ $q)/\min(n_{XX}, n_{XY}))^{1/2}$. Denote $\mathbf{\Delta}^C = \hat{\mathbf{C}} - \mathbf{C}^*$. By (S8.23) and Condition A3, with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, it holds that

$$\left\|\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\right\|_\infty \leq \left(\|\mathbf{C}^*\|_\infty + \|\mathbf{\Delta}^C\|_\infty\right)\|\hat{\mathbf{\Sigma}}_{XX}-\mathbf{\Sigma}_{XX}\|_\infty \lesssim \left(\frac{\log p}{n_{XX}}\right)^{\frac{1}{2}-\gamma_2}. \tag{S8.24}$$

Define $\boldsymbol{\gamma} := \mathrm{vec}(\mathbf{\Sigma}_{XY}\mathbf{C}^*)$. Then, with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$,

it holds that

$$\begin{aligned}\left\|\widehat{\boldsymbol{\gamma}}_{S_B}-\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\boldsymbol{\beta}^*_{S_B}\right\|_\infty &\le \left\|\widehat{\boldsymbol{\gamma}}_{S_B}-\boldsymbol{\gamma}_{S_B}\right\|_\infty+\left\|\left(\boldsymbol{\Gamma}_{S_BS_B}-\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\right)\boldsymbol{\beta}^*_{S_B}\right\|_\infty\\ &\le \left\|\widehat{\boldsymbol{\gamma}}_{S_B}-\boldsymbol{\gamma}_{S_B}\right\|_\infty+\left\|\boldsymbol{\Gamma}_{S_BS_B}-\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\right\|_\infty\left\|\boldsymbol{\beta}^*_{S_B}\right\|_\infty\\ &\le \left\|\widehat{\boldsymbol{\gamma}}_{S_B}-\boldsymbol{\gamma}_{S_B}\right\|_\infty+s_B\|\mathbf{B}^*\|_\infty\left\|\widehat{\boldsymbol{\Gamma}}_{S_BS_B}-\boldsymbol{\Gamma}_{S_BS_B}\right\|_\infty.\end{aligned}$$

By (S8.23), with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, it holds that

$$\begin{aligned}&\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_\infty\\ &\le\left(\|\mathbf{C}^*\|_{L_\infty}+\|\boldsymbol{\Delta}^C\|_{L_\infty}\right)\|\hat{\boldsymbol{\Sigma}}_{XY}-\boldsymbol{\Sigma}_{XY}\|_\infty\\ &\lesssim\left(\frac{\log(pq)}{n_{XY}}\right)^{\frac12-\gamma_2}.\end{aligned}\tag{S8.25}$$

Since $\hat{\boldsymbol{\beta}}_{S_B}=(\widehat{\boldsymbol{\Gamma}}_{S_BS_B})^{-1}\widehat{\boldsymbol{\gamma}}_{S_B}-\lambda_B(\hat{\boldsymbol{\Sigma}}_{XX,S_BS_B})^{-1}\operatorname{sign}(\hat{\boldsymbol{\beta}}_{S_B})$, $\frac{s_B}{\lambda_B}(\frac{\log p}{n_{XX}})^{\frac12-\gamma_1-\gamma_2}=O(1)$, $\frac{1}{\lambda_B}(\frac{\log(p+q)}{n_{XY}})^{\frac12-\gamma_2}=O(1)$, $s_BV(\frac{\log p}{n_{XX}})^{\frac12-\gamma_2}=O(1)$, with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, it holds that

$$\begin{aligned}\left\|\hat{\boldsymbol{\beta}}_{S_B}-\boldsymbol{\beta}^*_{S_B}\right\|_\infty&=\left\|\left(\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{S_B}-\lambda_B\left(\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\right)^{-1}\cdot\operatorname{sign}\left(\hat{\boldsymbol{\beta}}_{S_B}\right)-\boldsymbol{\beta}^*_{S_B}\right\|_\infty\\ &\le\left\|\left(\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{S_B}-\boldsymbol{\beta}^*_{S_B}\right\|_\infty+\lambda_B\left\|\left(\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\right)^{-1}\right\|_{L_\infty}\\ &\le\left(\left\|\widehat{\boldsymbol{\gamma}}_{S_B}-\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\boldsymbol{\beta}^*_{S_B}\right\|_\infty+\lambda_B\right)\cdot\left\|\left(\widehat{\boldsymbol{\Gamma}}_{S_BS_B}\right)^{-1}\right\|_{L_\infty}\\ &\le\left(\left\|\widehat{\boldsymbol{\gamma}}_{S_B}-\boldsymbol{\gamma}_{S_B}\right\|_\infty+s_B\|\mathbf{B}^*\|_\infty\left\|\widehat{\boldsymbol{\Gamma}}_{S_BS_B}-\boldsymbol{\Gamma}_{S_BS_B}\right\|_\infty+\lambda_B\right)\\ &\quad\frac{V}{1-s_BV\left\|\widehat{\boldsymbol{\Gamma}}_{S_BS_B}-\boldsymbol{\Gamma}_{S_BS_B}\right\|_\infty}\\ &\le\frac{2\lambda_BV}{1-s_BV\left\|\widehat{\boldsymbol{\Gamma}}_{S_BS_B}-\boldsymbol{\Gamma}_{S_BS_B}\right\|_\infty}\le4\lambda_BV<\min_{j\in S_B}|\boldsymbol{\beta}^*_j|,\end{aligned}$$

for sufficiently large p, q, n_{XX},n_{XY} and n_{YY}. The last step holds because we assume that $\lambda_B V / \min_{j\in S_B} |\boldsymbol{\beta}_j^*| = o(1)$. Thus we have $P(\mathcal{A}_1) \geq 1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$ for sufficiently large p, q, n_{XX},n_{XY} and n_{YY}.

For $\|\widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1} - \boldsymbol{\Gamma}_{S_B^C S_B}\left(\boldsymbol{\Gamma}_{S_B S_B}\right)^{-1}\|_{L_\infty}$, we have

$$
\begin{aligned}
&\left\|\widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1} - \boldsymbol{\Gamma}_{S_B^C S_B}\left(\boldsymbol{\Gamma}_{S_B S_B}\right)^{-1}\right\|_{L_\infty} \\
\leq &\left\|\boldsymbol{\Gamma}_{S_B^C S_B}\left(\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1} - \left(\boldsymbol{\Gamma}_{S_B S_B}\right)^{-1}\right)\right\|_{L_\infty} + \left\|\left(\widehat{\boldsymbol{\Gamma}}_{S_B^C S_B} - \boldsymbol{\Gamma}_{S_B^C S_B}\right)\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\right\|_{L_\infty} \\
\leq &\left\|\boldsymbol{\Gamma}_{S_B^C S_B}\left(\boldsymbol{\Gamma}_{S_B S_B}\right)^{-1}\right\|_{L_\infty} \cdot \left\|\boldsymbol{\Gamma}_{S_B S_B} - \widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right\|_{L_\infty} \cdot \left\|\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\right\|_{L_\infty} \\
&+ \left\|\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\right\|_{L_\infty} \cdot \left\|\widehat{\boldsymbol{\Gamma}}_{S_B^C S_B} - \boldsymbol{\Gamma}_{S_B^C S_B}\right\|_{L_\infty} \\
\leq &\left\|\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\right\|_{L_\infty} \cdot \left(\left\|\widehat{\boldsymbol{\Gamma}}_{S_B S_B} - \boldsymbol{\Gamma}_{S_B S_B}\right\|_{L_\infty} + \left\|\widehat{\boldsymbol{\Gamma}}_{S_B^C S_B} - \boldsymbol{\Gamma}_{S_B^C S_B}\right\|_{L_\infty}\right) \\
\leq &\frac{2 s_B V \|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_\infty}{1 - s_B V \|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_\infty}.
\end{aligned} \tag{S8.26}
$$

Since $\hat{\boldsymbol{\beta}}_{S_B} = (\widehat{\boldsymbol{\Gamma}}_{S_B S_B})^{-1}\widehat{\boldsymbol{\gamma}}_{S_B} - \lambda_B(\hat{\boldsymbol{\Sigma}}_{XX,S_B S_B})^{-1}\cdot \mathrm{sign}(\hat{\boldsymbol{\beta}}_{S_B})$, with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, it holds that

$$
\begin{aligned}
&\left\|\widehat{\boldsymbol{\gamma}}_{S_B^C} - \widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\hat{\boldsymbol{\beta}}_{S_B}\right\|_\infty \\
\leq &\left\|\widehat{\boldsymbol{\gamma}}_{S_B^C} - \widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\widehat{\boldsymbol{\gamma}}_{S_B}\right\|_\infty + \lambda_B\left\|\widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\right\|_{L_\infty} \\
\leq &\left\|\widehat{\boldsymbol{\gamma}}_{S_B^C} - \boldsymbol{\gamma}_{S_B^C}\right\|_\infty + \left\|\left(\boldsymbol{\Gamma}_{S_B^C S_B}\left(\boldsymbol{\Gamma}_{S_B S_B}\right)^{-1} - \widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\right)\boldsymbol{\gamma}_{S_B}\right\|_\infty \\
&+ \left\|\widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\left(\boldsymbol{\gamma}_{S_B} - \widehat{\boldsymbol{\gamma}}_{S_B}\right)\right\|_\infty + \lambda_B\left\|\widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\right\|_{L_\infty} \\
\leq &\underbrace{\left\|\widehat{\boldsymbol{\gamma}}_{S_B^C} - \boldsymbol{\gamma}_{S_B^C}\right\|_\infty}_{\text{(I)}} + \underbrace{\left\|\left(\boldsymbol{\Gamma}_{S_B^C S_B}\left(\boldsymbol{\Gamma}_{S_B S_B}\right)^{-1} - \widehat{\boldsymbol{\Gamma}}_{S_B^C S_B}\left(\widehat{\boldsymbol{\Gamma}}_{S_B S_B}\right)^{-1}\right)\boldsymbol{\Gamma}_{S_B S_B}\boldsymbol{\beta}_{S_B}^*\right\|_\infty}_{\text{(II)}}
\end{aligned}
$$

$$+ \underbrace{\left\| \widehat{\mathbf{\Gamma}}_{S_B^C S_B} \left(\widehat{\mathbf{\Gamma}}_{S_B S_B} \right)^{-1} \left(\gamma_{S_B^C} - \widehat{\boldsymbol{\gamma}}_{S_B^C} \right) \right\|_\infty + \lambda_B \left\| \widehat{\mathbf{\Gamma}}_{S_B^C S_B} \left(\widehat{\mathbf{\Gamma}}_{S_B S_B} \right)^{-1} \right\|_{L_\infty}}_{\text{(III)}}.$$

By Condition A3, Condition A5, (S8.24), (S8.25) and (S8.26), with probability at least $1 - \frac{4}{p} - \frac{4}{pq} - \frac{4}{q}$, it holds that

$$(I) \lesssim \left(\frac{\log(pq)}{n_{XY}} \right)^{\frac{1}{2}-\gamma_2},$$

$$\begin{aligned}(II) \leq & s_B \|\mathbf{B}^*\|_\infty \|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_\infty \left(1 + \left\| \widehat{\mathbf{\Gamma}}_{S_B^C S_B} \left(\widehat{\mathbf{\Gamma}}_{S_B S_B} \right)^{-1} \right\|_{L_\infty} \right) \\ \lesssim & s_B \left(\frac{\log p}{n_{XX}} \right)^{\frac{1}{2}-\gamma_1-\gamma_2} \left(2 - \eta + \frac{2 s_B V \|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_\infty}{1 - s_B V \|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_\infty} \right),\end{aligned}$$

$$\begin{aligned}(III) \leq & \left(\lambda_B + \left\| \boldsymbol{\gamma}_{S_B^C} - \widehat{\boldsymbol{\gamma}}_{S_B^C} \right\|_\infty \right) \left\| \widehat{\mathbf{\Gamma}}_{S_B^C S_B} \left(\widehat{\mathbf{\Gamma}}_{S_B S_B} \right)^{-1} \right\|_{L_\infty} \\ \lesssim & \left(\lambda_B + \left(\frac{\log(pq)}{n_{XY}} \right)^{\frac{1}{2}-\gamma_2} \right) \left(1 - \eta + \frac{2 s_B V \|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_\infty}{1 - s_B V \|\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_\infty} \right).\end{aligned}$$

Since $\frac{s_B}{\lambda_B}(\frac{\log p}{n_{XX}})^{\frac{1}{2}-\gamma_1-\gamma_2} = O(1)$, $\frac{1}{\lambda_B}(\frac{\log(p+q)}{n_{XY}})^{\frac{1}{2}-\gamma_2} = O(1)$, and $s_B V(\frac{\log p}{n_{XX}})^{\frac{1}{2}-\gamma_2} = O(1)$, when p, q, n_{XY}, n_{XX} and n_{YY} are sufficiently large, with probability at least $1 - \frac{4}{p} - \frac{4}{pq} - \frac{4}{q}$, it holds that

$$\frac{(I)}{\lambda_B} \leq \frac{\eta}{4}, \qquad \frac{(II)}{\lambda_B} \leq \frac{\eta}{4},$$

$$\begin{aligned}\frac{(III)}{\lambda_B} \leq & \frac{1}{\lambda_B}\left(\lambda_B + \left\|\boldsymbol{\gamma}_{S_B^C} - \widehat{\boldsymbol{\gamma}}_{S_B^C}\right\|_\infty\right)\left(1-\eta+\frac{2s_BV\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_\infty}{1-s_BV\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_\infty}\right)\\ =&1-\eta+\frac{1}{\lambda_B}\frac{2s_BV\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_\infty}{1-s_BV\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_\infty}+\frac{1}{\lambda_B}\left\|\boldsymbol{\gamma}_{S_B^C} - \widehat{\boldsymbol{\gamma}}_{S_B^C}\right\|_\infty(1-\eta+\\ &\left(\frac{2s_BV\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_\infty}{1-s_BV\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\|_\infty}\right)\\ \leq&1-\frac{\eta}{2}.\end{aligned}$$

Thus, with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, it holds that

$$\frac{\left\|\widehat{\boldsymbol{\gamma}}_{S_B^C}-\widehat{\boldsymbol{\Gamma}}_{S_B^CS_B}\hat{\boldsymbol{\beta}}_{S_B}\right\|_\infty}{\lambda_B}=\frac{(I)+(II)+(III)}{\lambda_B}\leq\frac{\eta}{4}+\frac{\eta}{4}+1-\frac{\eta}{2}=1.$$

Therefore, $P\left(\mathcal{A}_2\right)=P\left(\|\hat{\boldsymbol{\gamma}}^{\mathbf{C}}_{S_B^C}-\hat{\boldsymbol{\Gamma}}^{\mathbf{C}}_{S_B^CS_B}\hat{\boldsymbol{\beta}}_{S_B}\|_\infty\leq\lambda_B\right)\geq 1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$.

Since $P(\|\hat{\boldsymbol{\beta}}_{S_B}-\boldsymbol{\beta}^*_{S_B}\|_\infty<\min_{j\in S_B}|\boldsymbol{\beta}^*_j|)\geq 1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$ it holds that $|\hat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}^*_j|<|\boldsymbol{\beta}^*_j|$ for $j\in S_B$. Thus, we have $P(\text{sign}(\hat{\boldsymbol{\beta}}_{S_B})=\text{sign}(\boldsymbol{\beta}^*_{S_B}))\geq 1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$. Let $\hat{\boldsymbol{\beta}}\in\mathbb{R}^{pq}$ which satisfies $\hat{\boldsymbol{\beta}}_{S_B^C}=\mathbf{0}$ and $\hat{\boldsymbol{\beta}}_{S_B}=\hat{\boldsymbol{\beta}}_{S_B}$. Since $P(\|\hat{\boldsymbol{\gamma}}^{\mathbf{C}}_{S_B^C}-\hat{\boldsymbol{\Gamma}}^{\mathbf{C}}_{S_B^CS_B}\hat{\boldsymbol{\beta}}_{S_B}\|_\infty\leq\lambda_B)\geq 1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, $\hat{\boldsymbol{\beta}}$ satisfies (S8.22) with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$. Thus, we have verified that $\text{sign}(\hat{\boldsymbol{\beta}})=\text{sign}(\boldsymbol{\beta}^*)$ with high probability. This completes the proof.

S9 Supporting lemmas

Lemma S9.1. *(Lemma 1 from Cai et al. (2013)) Let $\xi_1, \ldots, \xi_n$ be independent random variables with mean zero. Suppose that there exists some $t > 0$ and $\bar{B}_n$ such that $\sum_{k=1}^n E\left\{\xi_k^2 e^{t|\xi_k|}\right\} \leqslant \bar{B}_n^2$. Then uniformly for $0 < x \leqslant \bar{B}_n$,*

$$\mathrm{pr}\left(\sum_{k=1}^n \xi_k \geqslant C_t \bar{B}_n x\right) \leqslant \exp\left(-x^2\right),$$

where $C_t = t + t^{-1}$.

Lemma S9.2. *(Lemma 1 from Ravikumar et al. (2011)) Consider a zero-mean random vector $\mathbf{X} = (X_1, \ldots, X_p)^\top$ with covariance $\mathbf{\Sigma} = (\sigma_{ij})$ such that $X_j/\sqrt{\sigma_{jj}}$ is sub-Gaussian with parameter L for $1 \leq j \leq p$. Let $\{\mathbf{X}_i\}_{i=1}^n$ be i.i.d. samples of $\mathbf{X}$, the sample covariance $\hat{\mathbf{\Sigma}} = (\hat{\sigma}_{ij})$ satisfies the tail bound that*

$$P\left(|\hat{\sigma}_{jt} - \sigma_{jt}| \geq \delta\right) \leq 4\exp\left\{-\frac{n\delta^2}{128\left(1+4L^2\right)^2 \max_j\left(\sigma_{jj}\right)^2}\right\},$$

for all $\delta \in (0, 8\max_j(\sigma_{jj})(1+4L^2))$.

Lemma S9.3. *(Lemma 1 of Negahban et al. (2012)) Suppose that $\mathcal{L}$ is a convex and differentiable function and consider any optimal solution $\widehat{\boldsymbol{\theta}}_{\lambda_n}$ to the following optimization problem*

$$\widehat{\boldsymbol{\theta}}_{\lambda_n} \in \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^p}\left\{\mathcal{L}\left(\boldsymbol{\theta}; \mathbf{Z}_1^n\right) + \lambda_n \mathcal{R}(\boldsymbol{\theta})\right\},$$

where $\lambda_n > 0$ *is a constant and* $\mathcal{R} : \mathbb{R}^p \to \mathbb{R}_+$ *is a decomposable norm. For a given inner product* $\langle \cdot, \cdot \rangle$*, define the dual norm of* $\mathcal{R}$ *as*

$$\mathcal{R}^*(\mathbf{v}) := \sup_{\mathbf{u} \in \mathbb{R}^p \setminus \{0\}} \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\mathcal{R}(\mathbf{u})} = \sup_{\mathcal{R}(\mathbf{u}) \leq 1} \langle \mathbf{u}, \mathbf{v} \rangle.$$

If $\lambda_n \geq 2\mathcal{R}^* \left(\nabla \mathcal{L} \left(\boldsymbol{\theta}^*; \mathbf{Z}_1^n \right) \right)$ *and for any pair of sets* $(\mathcal{M}, \overline{\mathcal{M}}^{\perp})$ *over which* $\mathcal{R}$ *is decomposable, the error* $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\theta}}_{\lambda_n} - \boldsymbol{\theta}^*$ *belongs to the set*

$$S\left(\mathcal{M}, \overline{\mathcal{M}}^{\perp}; \boldsymbol{\theta}^*\right) := \left\{ \boldsymbol{\Delta} \in \mathbb{R}^p \mid \mathcal{R}\left(\boldsymbol{\Delta}_{\overline{\mathcal{M}}^{\perp}}\right) \leq 3\mathcal{R}\left(\boldsymbol{\Delta}_{\overline{\mathcal{M}}}\right) + 4\mathcal{R}\left(\boldsymbol{\theta}^*_{\mathcal{M}^{\perp}}\right) \right\}.$$

Lemma S9.4. *Under assumptions of Theorem 3.4,* $\boldsymbol{\Delta}^B = \hat{\mathbf{B}} - \mathbf{B}^*$ *belongs to the set*

$$\mathcal{C}_B := \left\{ \boldsymbol{\Delta}^B \in \mathbb{R}^{p \times q} \left| \left\| \boldsymbol{\Delta}^B_{S_B^C} \right\|_1 \leq 3 \left\| \boldsymbol{\Delta}^B_{S_B} \right\|_1 \right. \right\}, \tag{S9.27}$$

and $\boldsymbol{\Delta}^C = \hat{\mathbf{C}} - \mathbf{C}^*$ *belongs to the set*

$$\mathcal{C}_C := \left\{ \boldsymbol{\Delta}^C \in \mathbb{R}^{q \times q} \left| \left\| \boldsymbol{\Delta}^C_{S_C^C} \right\|_1 \leq 3 \left\| \boldsymbol{\Delta}^C_{S_C} \right\|_1 \right. \right\}. \tag{S9.28}$$

Proof of Lemma S9.4. Since $\mathbf{Y} = \mathbf{X}\mathbf{B}^* + \mathcal{E}$, we have

$$\begin{aligned}
\boldsymbol{\Sigma}_{YY} &= \mathrm{Cov}(\mathbf{Y}, \mathbf{Y}) = \mathrm{Cov}(\mathbf{X}\mathbf{B}^* + \mathbf{E}, \mathbf{X}\mathbf{B}^* + \mathcal{E}) = {\mathbf{B}^*}^{\top} \boldsymbol{\Sigma}_{XX} \mathbf{B}^* + \mathrm{Cov}(\mathcal{E}, \mathcal{E}) \\
&= {\mathbf{B}^*}^{\top} \boldsymbol{\Sigma}_{XX} \mathbf{B}^* + {\mathbf{C}^*}^{-1}, \\
\boldsymbol{\Sigma}_{XY} &= \mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = \mathrm{Cov}(\mathbf{X}, \mathbf{X}\mathbf{B}^* + \mathcal{E}) = \boldsymbol{\Sigma}_{XX} \mathbf{B}^*.
\end{aligned}$$

Thus, by Theorem 3.1 and Condition A3, it holds with probability at least

$1 - \frac{4}{p} - \frac{4}{pq} - \frac{4}{q}$ that

$$\begin{aligned}
&\|\nabla_B \mathcal{L}_n(\mathbf{B}^*, \mathbf{C}^*)\|_\infty \\
=&\left\|2\mathbf{C}^{*\top}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{XX} - 2\mathbf{C}^{*\top}\hat{\mathbf{\Sigma}}_{XY}^\top\right\|_\infty \\
=&\left\|2\mathbf{C}^{*\top}\mathbf{B}^{*\top}\mathbf{\Delta}^{XX} - 2\mathbf{C}^{*\top}\mathbf{\Delta}^{XY\top}\right\|_\infty \\
\leq& 2\|\mathbf{C}^*\|_{L_1}(\|\mathbf{B}^*\|_{L_1}\|\mathbf{\Delta}^{XX}\|_\infty + \|\mathbf{\Delta}^{XY}\|_\infty) \\
\lesssim& \max\left\{\left(\frac{\log p}{n_{XX}}\right)^{\frac{1}{2}-\gamma_1-\gamma_2}, \left(\frac{\log(pq)}{n_{XY}}\right)^{\frac{1}{2}-\gamma_2}\right\} \\
\lesssim& \lambda_B,
\end{aligned}$$

and

$$\begin{aligned}
&\|\nabla_C \mathcal{L}_n(\mathbf{B}^*, \mathbf{C}^*)\|_\infty \\
=&\left\|\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{XX}\mathbf{B}^* + \hat{\mathbf{\Sigma}}_{YY} - \mathbf{C}^{*-1} - 2\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{XY}\right\|_\infty \\
=&\left\|\mathbf{B}^{*\top}\mathbf{\Delta}^{XX}\mathbf{B}^* + \mathbf{\Delta}^{YY} - 2\mathbf{B}^{*\top}\mathbf{\Delta}^{XY}\right\|_\infty \\
\leq& \|\mathbf{B}^*\|_{L_1}^2\|\mathbf{\Delta}^{XX}\|_\infty + \|\mathbf{\Delta}^{YY}\|_\infty + 2\|\mathbf{B}^*\|_{L_1}\|\mathbf{\Delta}^{XY}\|_\infty \\
\lesssim& \max\left\{\left(\frac{\log p}{n_{XX}}\right)^{\frac{1}{2}-2\gamma_1}, \left(\frac{\log(pq)}{n_{XY}}\right)^{\frac{1}{2}-\gamma_1}, \left(\frac{\log q}{n_{YY}}\right)^{\frac{1}{2}}\right\} \\
\lesssim& \lambda_C.
\end{aligned}$$

Since L_1 penalty is decomposable, by applying Lemma S9.3, we have

$$\left\|\mathbf{\Delta}^B_{S_B^C}\right\|_1 \leq 3\|\mathbf{\Delta}^B_{S_B}\|_1 + 4\|\mathbf{B}^*_{S_B^C}\|_1 = 3\|\mathbf{\Delta}^B_{S_B}\|_1,$$

$$\left\|\mathbf{\Delta}^C_{S_C^C}\right\|_1 \leq 3\|\mathbf{\Delta}^C_{S_C}\|_1 + 4\|\mathbf{C}^*_{S_C^C}\|_1 = 3\|\mathbf{\Delta}^B_{S_C}\|_1.$$

□

Theorem S9.1. *(Theorem 1 of Loh and Wainwright (2015)) Consider the optimization problem*

$$\tilde{\boldsymbol{\beta}} = \arg \min_{\|\boldsymbol{\beta}\|_1 \leq R, \boldsymbol{\beta} \in \Omega} \mathcal{L}_n(\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1,$$

where Ω is some convex set and the empirical loss $\mathcal{L}_n$ satisfies the RSC conditions

$$\langle \nabla \mathcal{L}_n (\boldsymbol{\beta}^* + \boldsymbol{\Delta}) - \nabla \mathcal{L}_n (\boldsymbol{\beta}^*), \boldsymbol{\Delta} \rangle \geq \begin{cases} \alpha_1 \|\boldsymbol{\Delta}\|_2^2 - \tau_1 \dfrac{\log p}{n} \|\boldsymbol{\Delta}\|_1^2, & \forall \|\boldsymbol{\Delta}\|_2 \leq 1; \text{(S9.29)} \\ \alpha_2 \|\boldsymbol{\Delta}\|_2 - \tau_2 \sqrt{\dfrac{\log p}{n}} \|\boldsymbol{\Delta}\|_1, & \forall \|\boldsymbol{\Delta}\|_2 \geq 1; \text{(S9.30)} \end{cases}$$

α_1, α_2 are positive constants and τ_1, τ_2 are non-negative constants. Suppose $n \geq \frac{16 R^2 \max(\tau_1^2, \tau_2^2)}{\alpha_2^2} \log p$, $\|\boldsymbol{\beta}^*\|_1 \leq R$ and

$$\frac{4}{L} \max \left\{ \|\nabla \mathcal{L}_n (\boldsymbol{\beta}^*)\|_\infty, \alpha_2 \sqrt{\frac{\log p}{n}} \right\} \leq \lambda \leq \frac{\alpha_2}{6RL},$$

where L is a constant. Then for any vector $\tilde{\boldsymbol{\beta}}$ with $\|\tilde{\boldsymbol{\beta}}\|_1 \leq R$ and satisfies the first-order necessary condition

$$\left\langle \nabla \mathcal{L}_n(\widetilde{\boldsymbol{\beta}}) + \nabla \|\tilde{\boldsymbol{\beta}}\|_1, \boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}} \right\rangle \geq 0, \quad \text{for all } \|\boldsymbol{\beta}\|_1 \leq 1,$$

it holds that

$$\left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \right\|_2 \leq \frac{6\lambda\sqrt{k}}{4\alpha_1}, \quad \text{and} \quad \left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \right\|_1 \leq \frac{24\lambda k}{4\alpha_1},$$

where $k = \|\boldsymbol{\beta}^*\|_0$.

Lemma S9.5. *Let $n_{XX/Y} \asymp n_{XX}^{\tau_1}$ and $n_{XY/X} \asymp n_{XY}^{\tau_2}$ with $\tau_1, \tau_2 \in \{-\infty\} \cup [0, 1]$, $\boldsymbol{\delta}_C = \text{vec}(\mathbf{C}^* - \hat{\mathbf{C}})$, $1 - \alpha_1 = O(\sqrt{\log p / n_X})$, $1 - \alpha_2 = O(\sqrt{\log p / n_{XX}})$*

and $1-\alpha_3 = O(\sqrt{\log(pq)/n_{XY}})$. *With probability at least* $1-\frac{4}{p}-\frac{4}{pq}$, *we have*

$$\begin{aligned}
&\left\|\nabla_B\left\{\mathrm{tr}[\hat{\mathbf{C}}\hat{\mathbf{\Sigma}}_{\mathbf{YY}}+\mathbf{B}^*\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XX}}-\mathbf{2}\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XY}}]-\log\det(\hat{\mathbf{C}})\right\}\right\|_\infty \\
\lesssim&\frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2},n_{XY}^{1-\tau_2/2}\right)}(\|\mathbf{B}^*\mathbf{C}^*\|_{L_1}+\|\mathbf{B}^*\|_{L_1}\|\boldsymbol{\delta}_C\|_1) \\
&+\max\{\lambda_{\max}(\mathbf{C}^*),1/\lambda_{\min}(\mathbf{C}^*)\}\left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2}(1+\|\boldsymbol{\delta}_C\|_1).
\end{aligned}$$

Proof. Denote $\mathbf{\Delta}^{YY}=\mathbf{\Sigma}_{YY}-\hat{\mathbf{\Sigma}}_{YY}$, Theorem 3.1 implies that with probability at least $1-\frac{4}{p}-\frac{4}{pq}-\frac{4}{q}$, we have

$$\|\mathbf{\Delta}^{XX}\|_\infty \le v_1'\sqrt{\frac{\log p}{n_{XX}}}; \tag{S9.31}$$

$$\|\mathbf{\Delta}^{XY}\|_\infty \le v_2'\sqrt{\frac{\log(pq)}{n_{XY}}}; \tag{S9.32}$$

$$\|\mathbf{\Delta}^{YY}\|_\infty \le v_3'\sqrt{\frac{\log q}{n_{YY}}}. \tag{S9.33}$$

Define $\tilde{y}_{ij}$ and $\tilde{x}_{ij}$ to be the underlying complete data without missing entries. Define the observed-data indicator matrix as $M^X=(m_{ij}^X)$ and $M^Y=(m_{ij}^Y)$ such that $m_{ij}^X=1$ when x_{ij} is observed, $m_{ij}^X=0$ when x_{ij} is missing, $m_{ij}^Y=1$ when y_{ij} is observed and $m_{ij}^Y=0$ when y_{ij} is missing. Then we can write the observed data as $x_{ij}=m_{ij}^X\tilde{x}_{ij}$, $y_{ij}=m_{ij}^Y\tilde{y}_{ij}$. Define

α_{ij}^{XX} to be the adjusting weight we use to estimate $(\hat{\mathbf{\Sigma}}_{XX})_{ij}$, that is

$$\alpha_{ij}^{XX} = \begin{cases} 1 & \text{if } i = j; \\ \alpha_1 & \text{if } i \neq j, i \text{ and } j \text{ are in the same modality;} \\ \alpha_2 & \text{if } i \text{ and } j \text{ are in different modalities.} \end{cases}$$

Then we have

$$\begin{aligned} &\left\| \nabla_B \left\{ \mathrm{tr}[\hat{\mathbf{C}}\hat{\mathbf{\Sigma}}_{\mathbf{YY}} + \mathbf{B}^*\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XX}} - \mathbf{2}\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XY}}] - \log\det(\hat{\mathbf{C}}) \right\} \right\|_\infty \\ =&\left\| 2\hat{\mathbf{C}}^\top \mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{XX} - 2\hat{\mathbf{C}}^\top \hat{\mathbf{\Sigma}}_{XY}^\top \right\|_\infty . \end{aligned} \tag{S9.34}$$

When either $\mathbf{Y}$ or $\mathbf{X}$ has missing entries, we have

$$\begin{aligned} &(\hat{\mathbf{\Sigma}}_{XX}\mathbf{B}^* - \hat{\mathbf{\Sigma}}_{XY})_{ij} \\ =&\sum_{l=1}^{p} \frac{\alpha_{il}^{XX} \sum_{k\in S_{il}^{XX}} x_{ki}x_{kl}}{n_{il}^{XX}} \mathbf{B}_{lj}^* - \frac{\alpha_3 \sum_{k\in S_{ij}^{XY}}^{n} x_{ki} m_{kj}^Y \tilde{y}_{kj}}{n_{ij}^{XY}} \\ =&(\hat{\mathbf{\Sigma}}_{XX}\mathbf{B}^* - \hat{\mathbf{\Sigma}}_{X\tilde{X}}\mathbf{B}^*)_{ij} - (\hat{\mathbf{\Sigma}}_{X\epsilon})_{ij}, \end{aligned}$$

where $(\hat{\mathbf{\Sigma}}_{X\tilde{X}})_{ij} = \alpha_3 \sum_{k\in S_{ij}^{XY}} x_{ki}\tilde{x}_{kj}/n_{ij}^{XY}$, and $(\hat{\mathbf{\Sigma}}_{X\epsilon})_{ij} = \alpha_3 \sum_{k\in S_{ij}^{XY}} x_{ki}\epsilon_{kj}/n_{ij}^{XY}$. Then by (S9.34) we have

$$\begin{aligned} &\left\| \nabla_B \left\{ \mathrm{tr}[\hat{\mathbf{C}}\hat{\mathbf{\Sigma}}_{\mathbf{YY}} + \mathbf{B}^*\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XX}} - \mathbf{2}\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XY}}] - \log\det(\hat{\mathbf{C}}) \right\} \right\|_\infty \\ =&2 \left\| (\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})\mathbf{B}^*\hat{\mathbf{C}} - \hat{\mathbf{\Sigma}}_{X\epsilon}\hat{\mathbf{C}} \right\|_\infty \\ \leq&2 \left\| (\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})\mathbf{B}^*\mathbf{C}^* \right\|_\infty + 2 \left\| (\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})\mathbf{B}^*(\mathbf{C}^* - \hat{\mathbf{C}}) \right\|_\infty \\ &+2 \left\| \hat{\mathbf{\Sigma}}_{X\epsilon}\mathbf{C}^* \right\|_\infty + 2 \left\| \hat{\mathbf{\Sigma}}_{X\epsilon}(\mathbf{C}^* - \hat{\mathbf{C}}) \right\|_\infty . \end{aligned}$$

We first derive an upper bound for the first term $\|(\hat{\mathbf{\Sigma}}_{XX}-\hat{\mathbf{\Sigma}}_{X\tilde{X}})\mathbf{B}^*\mathbf{C}^*\|_\infty$.

Define $S_{jkl}^{XXY}=\{i: x_{ij}, x_{ik} \text{ and } y_{il} \text{ are not missing}\}$ and $n_{jkl}^{XXY}=|S_{jkl}^{XXY}|$.

When $n_{ijl}^{XX/Y}\neq 0$ and $n_{ilj}^{XY/X}\neq 0$, for each entry in matrix $\hat{\mathbf{\Sigma}}_{XX}-\hat{\mathbf{\Sigma}}_{X\tilde{X}}$ and $1\leq l\leq q$, with probability at least $1-\frac{4}{p^3}$, we have

$$
\begin{aligned}
&(\hat{\mathbf{\Sigma}}_{XX}-\hat{\mathbf{\Sigma}}_{X\tilde{X}})_{ij}\\
=&\frac{1}{n_{ijl}^{XXY}}\sum_{k\in S_{ijl}^{XXY}} X_{ik}X_{jk}\frac{n_{ijl}^{XXY}(\alpha_{ij}^{XX}n_{ij}^{XX}-\alpha_3 n_{il}^{XY})}{n_{ij}^{XX}n_{il}^{XY}}\\
&+\frac{1}{n_{ijl}^{XX/Y}}\sum_{k\in S_{ijl}^{XX/Y}} X_{ik}X_{jk}\frac{\alpha_{ij}^{XX}n_{ijl}^{XX/Y}}{n_{ij}^{XX}}-\frac{1}{n_{ilj}^{XY/X}}\sum_{k\in S_{ilj}^{XY/X}} X_{ik}X_{jk}\frac{\alpha_3 n_{ilj}^{XY/X}}{n_{il}^{XY}}\\
\leq&\left(\frac{1}{n_{ijl}^{XXY}}\sum_{k\in S_{ijk}^{XXY}} X_{ik}X_{jk}-\mathbf{\Sigma}_{XX,ij}\right)\frac{n_{ijl}^{XXY}(\alpha_{ij}^{XX}n_{ij}^{XX}-\alpha_3 n_{il}^{XY})}{n_{ij}^{XX}n_{il}^{XY}}\\
&+\left(\frac{1}{n_{ijl}^{XX/Y}}\sum_{k\in S_{ijl}^{XX/Y}} X_{ik}X_{jk}-\mathbf{\Sigma}_{XX,ij}\right)\frac{\alpha_{ij}^{XX}n_{ijl}^{XX/Y}}{n_{ij}^{XX}}\\
&-\left(\frac{1}{n_{ilj}^{XY/X}}\sum_{k\in S_{ilj}^{XY/X}} X_{ik}X_{jk}-\mathbf{\Sigma}_{XX,ij}\right)\frac{\alpha_3 n_{ilj}^{XY/X}}{n_{il}^{XY}}+2(1-\alpha_{ij}^{XX})+2(1-\alpha_3)\\
\lesssim&\sqrt{\frac{\log p}{n_{ijl}^{XXY}}}\frac{n_{ijl}^{XXY}(n_{ij}^{XX}-n_{il}^{XY})}{n_{ij}^{XX}n_{il}^{XY}}+\sqrt{\frac{\log p}{n_{ijl}^{XX/Y}}}\frac{n_{ijl}^{XX/Y}}{n_{ij}^{XX}}+\sqrt{\frac{\log p}{n_{ilj}^{XY/X}}}\frac{n_{ilj}^{XY/X}}{n_{il}^{XY}}+\alpha_{ij}^{XX}-\alpha_3\\
\lesssim&\sqrt{\log p}\max\left[\max_{ijl}\left(\frac{(n_{ijl}^{XX/Y})^{1/2}}{n_{ij}^{XX}}\right),\max_{ijl}\left(\frac{(n_{ilj}^{XY/X})^{1/2}}{n_{il}^{XY}}\right)\right]-(1-\alpha_{ij}^{XX})+(1-\alpha_3)\\
\lesssim&\frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2},n_{XY}^{1-\tau_2/2}\right)},
\end{aligned}
$$

where $\tau_1,\tau_2\in[0,1]$.

When $n_{ijl}^{XX/Y}=0$ and $n_{ilj}^{XY/X}\neq 0$, for each entry in matrix $\hat{\mathbf{\Sigma}}_{XX}-\hat{\mathbf{\Sigma}}_{X\tilde{X}}$

and $1 \le l \le q$, with probability at least $1 - \frac{4}{p^3}$, we have

$$\begin{aligned}
&(\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})_{ij} \\
=&\frac{1}{n_{ijl}^{XXY}} \sum_{k \in S_{ijl}^{XXY}} X_{ik}X_{jk} \frac{n_{ijk}^{XXY}(\alpha_{ij}^{XX} n_{ij}^{XX} - \alpha_3 n_{il}^{XY})}{n_{ij}^{XX} n_{il}^{XY}} \\
&- \frac{1}{n_{ilj}^{XY/X}} \sum_{k \in S_{ilj}^{XY/X}} X_{ik}X_{jk} \frac{\alpha_3 n_{ilj}^{XY/X}}{n_{il}^{XY}} \\
\lesssim&\frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2}, n_{XY}^{1-\tau_2/2}\right)},
\end{aligned}$$

where $\tau_2 \in [0,1]$, $\tau_1 \in \{-\infty\} \cup [0,1]$.

When $n_{ijl}^{XX/Y} \neq 0$ and $n_{ijl}^{XY/X} = 0$, for each entry in matrix $\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}}$ and $1 \le l \le q$, with probability at least $1 - \frac{4}{p^3}$, we have

$$\begin{aligned}
&(\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})_{ij} \\
=&\frac{1}{n_{ijl}^{XXY}} \sum_{k \in S_{ijl}^{XXY}} X_{ik}X_{jk} \frac{n_{ijk}^{XXY}(\alpha_{ij}^{XX} n_{ij}^{XX} - \alpha_3 n_{il}^{XY})}{n_{ij}^{XX} n_{il}^{XY}} \\
&+ \frac{1}{n_{ijl}^{XX/Y}} \sum_{k \in S_{ijl}^{XX/Y}} X_{ik}X_{jk} \frac{\alpha_{ij}^{XX} n_{ijl}^{XX/Y}}{n_{ij}^{XX}} \\
\lesssim&\frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2}, n_{XY}^{1-\tau_2/2}\right)},
\end{aligned}$$

where $\tau_1 \in [0,1]$, $\tau_2 \in \{-\infty\} \cup [0,1]$.

When $n_{ijl}^{XX/Y} = n_{ijl}^{XY/X} = 0$, $n_{ijl}^{XXY} = n_{ij}^{XX} = n_{il}^{XY}$. Then for each entry in matrix $\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}}$ and $1 \le l \le q$, with probability at least $1 - \frac{4}{p^3}$, we

have

$$\begin{aligned}
&(\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})_{ij} \\
=&\frac{(\alpha_{ij}^{XX} - \alpha_3)}{n_{ijl}^{XXY}} \sum_{k \in S_{ijl}^{XXY}} X_{ik} X_{jk} \\
\lesssim& \max(1-\alpha_1, 1-\alpha_2, 1-\alpha_3) \sqrt{\frac{\log p}{n_{ijl}^{XXY}}} \\
\lesssim& \frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2}, n_{XY}^{1-\tau_2/2}\right)},
\end{aligned}$$

where $\tau_1, \tau_2 \in \{-\infty\} \cup [0,1]$.

If we combine the above four cases, for each entry in matrix $\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}}$ and $1 \le l \le q$, with probability at least $1 - \frac{4}{p^3}$, we have

$$(\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})_{ij} \lesssim \frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2}, n_{XY}^{1-\tau_2/2}\right)},$$

where $\tau_1, \tau_2 \in \{-\infty\} \cup [0,1]$.

Then by Holder's inequality and the union bound, with probability at least $1 - 4/p$ we have

$$\|(\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})\mathbf{B}^*\mathbf{C}^*\|_\infty \lesssim \frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2}, n_{XY}^{1-\tau_2/2}\right)} \|\mathbf{B}^*\mathbf{C}^*\|_{L_1}. \tag{S9.35}$$

Similarly, for the second term $\left\|(\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})\mathbf{B}^*(\mathbf{C}^* - \hat{\mathbf{C}})\right\|_\infty$, with probability at least $1 - 4/p$ we have

$$\|(\hat{\mathbf{\Sigma}}_{XX} - \hat{\mathbf{\Sigma}}_{X\tilde{X}})\mathbf{B}^*(\mathbf{C}^* - \hat{\mathbf{C}})\|_\infty \lesssim \frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2}, n_{XY}^{1-\tau_2/2}\right)} \|\mathbf{B}^*\|_{L_1} \|\boldsymbol{\delta}_C\|_1. \tag{S9.36}$$

Next we focus on the third term $\|\hat{\mathbf{\Sigma}}_{X\epsilon}\mathbf{C}^*\|_\infty$. Each entry of the matrix $(\hat{\mathbf{\Sigma}}_{X\epsilon}\mathbf{C}^*)_{ij}$ can be written as $\frac{\alpha_3}{n_{ij}^{XY}}\sum_{k\in S_{ij}^{XY}} X_{ki}(\boldsymbol{\epsilon}_k\mathbf{C}^*)_j$. By Condition A1 and monotone convergence theorem, for any $t\in\mathbb{R}$, we have

$$\mathbb{E}\left[\exp\left(\frac{X_{ki}^2}{8L_1^2}\right)\right]=\mathbb{E}\left[\sum_{l=0}^{\infty}\frac{X_{ki}^{2l}}{(4L_1^2)^l\, l!}\frac{1}{2^l}\right]\le\sum_{l=0}^{\infty}\frac{1}{2^l}=2.$$

By Condition A1, the error vectors also follow sub-Gaussian distribution. Assume $\mathbb{E}(\exp(t\mathbf{u}_2^\top\boldsymbol{\epsilon}_i))\le\exp\left(\frac{L_3^2\|\mathbf{u}_2\|_2^2t^2}{2}\right)$. Then we have

$$\mathbb{E}\left[\exp\left(\frac{(\boldsymbol{\epsilon}_k\mathbf{C}^*)_j^2}{8L_3^2\|\mathbf{C}_j^*\|_2^2}\right)\right]\le 2.$$

By Young's inequality and the simple inequality $s^2e^s\le e^{2s}$ for $s>0$, we have

$$\begin{aligned}
&\mathbb{E}\left[(X_{ki}(\boldsymbol{\epsilon}_k\mathbf{C}^*)_j|)^2\exp\left(\frac{|X_{ki}(\boldsymbol{\epsilon}_k\mathbf{C}^*)_j|}{8L_1L_3\lambda_{\max}(\mathbf{C}^*)}\right)\right]\\
\le&\mathbb{E}\left[\exp\left(\frac{|X_{ki}(\boldsymbol{\epsilon}_k\mathbf{C}^*)_j|}{4L_1L_3\lambda_{\max}(\mathbf{C}^*)}\right)\right]\\
\le&\mathbb{E}\left[\exp\left(\frac{X_{ki}^2}{8L_1^2}\right)\exp\left(\frac{(\boldsymbol{\epsilon}_k\mathbf{C}^*)_j^2}{8L_3^2\|\mathbf{C}^*\|_2^2}\right)\right]\\
\le&\frac{1}{2}\left[\mathbb{E}\exp\left(\frac{X_{ki}^2}{8L_1^2}\right)\right]^2+\frac{1}{2}\left[\mathbb{E}\exp\left(\frac{(\boldsymbol{\epsilon}_k\mathbf{C}_j^*)^2}{8L_3^2\|\mathbf{C}^*\|_2^2}\right)\right]^2\\
\le&4.
\end{aligned}$$

By Lemma S9.1, let $\bar{B}_n=2\sqrt{n_{XY}}$, $t=\frac{1}{8L_1L_3\lambda_{\max}(\mathbf{C}^*)}$ and $x=\sqrt{2\log(pq)}$, we have

$$\max_{i,j}\mathbb{P}\left[\left|\frac{1}{n_{ij}^{XY}}\sum_{k\in S_{ij}^{XY}}\left(X_{ki}\left(\epsilon_k^\top\mathbf{C}^*\right)_j\right)\right|\ge C_1\left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2}\right]\le 2(pq)^{-2}. \tag{S9.37}$$

where $C_1 = \frac{\sqrt{2}}{8L_1L_3\lambda_{\max}(\mathbf{C}^*)} + 8\sqrt{2}L_1L_3\lambda_{\max}(\mathbf{C}^*)$. So with probability at least $1 - 2(pq)^{-1}$, we can bound the third term by

$$\|\hat{\mathbf{\Sigma}}_{X\epsilon}\mathbf{C}^*\|_\infty \lesssim C_1 \left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2}. \tag{S9.38}$$

Similarly, for the last term $\left\|\hat{\mathbf{\Sigma}}_{X\epsilon}(\mathbf{C}^* - \hat{\mathbf{C}})\right\|_\infty$, we have

$$\left\|\hat{\mathbf{\Sigma}}_{X\epsilon}(\mathbf{C}^* - \hat{\mathbf{C}})\right\|_\infty \lesssim C_2 \left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2} \|\boldsymbol{\delta}_C\|_1, \tag{S9.39}$$

where $C_2 = \frac{\sqrt{2}}{8L_1L_3\lambda_{\min}(\mathbf{C}^*)^{-1}} + 8\sqrt{2}L_1L_3\lambda_{\min}(\mathbf{C}^*)^{-1}$.

By (S9.35), (S9.36), (S9.38), (S9.39), with probability at least $1 - \frac{4}{p} - \frac{4}{pq}$ we have

$$\begin{aligned}
&\left\|\nabla_B \left\{\mathrm{tr}[\hat{\mathbf{C}}\hat{\mathbf{\Sigma}}_{\mathbf{YY}} + \mathbf{B}^*\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XX}} - \mathbf{2}\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XY}}] - \log\det(\hat{\mathbf{C}})\right\}\right\|_\infty \\
\lesssim &\frac{(\log p)^{1/2}}{\min\left(n_{XX}^{1-\tau_1/2}, n_{XY}^{1-\tau_2/2}\right)}(\|\mathbf{B}^*\mathbf{C}^*\|_{L_1} + \|\mathbf{B}^*\|_{L_1}\|\boldsymbol{\delta}_C\|_1) \\
&+ \max\{\lambda_{\max}(\mathbf{C}^*), 1/\lambda_{\min}(\mathbf{C}^*)\}\left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2}(1 + \|\boldsymbol{\delta}_C\|_1).
\end{aligned}$$

We remark that, when both $\mathbf{X}$ and $\mathbf{Y}$ are complete, we can set $\alpha_1 = \alpha_2 = \alpha_3 = 1$. Then by (S9.34) we have

$$\begin{aligned}
&\left\|\nabla_B \left\{\mathrm{tr}[\hat{\mathbf{C}}\hat{\mathbf{\Sigma}}_{\mathbf{YY}} + \mathbf{B}^*\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XX}} - \mathbf{2}\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XY}}] - \log\det(\hat{\mathbf{C}})\right\}\right\|_\infty \\
\le &\left\|2\mathbf{C}^{*\top}\tilde{\mathbf{\Sigma}}_{X\epsilon}\right\|_\infty + \left\|2(\mathbf{C}^* - \hat{\mathbf{C}})^\top\tilde{\mathbf{\Sigma}}_{X\epsilon}\right\|_\infty,
\end{aligned}$$

where $(\tilde{\mathbf{\Sigma}}_{X\epsilon})_{ij} = \sum_{k=1}^n x_{ki}\epsilon_{kj}/n$. By (S9.37), with probability at least $1 -$

$2(pq)^{-1}$, we have

$$\|\mathbf{C}^{*\top}\tilde{\mathbf{\Sigma}}_{X\epsilon}\|_\infty \lesssim C_1\left\{\frac{\log(pq)}{n}\right\}^{1/2},$$

and

$$\left\|2(\mathbf{C}^*-\hat{\mathbf{C}})^\top\tilde{\mathbf{\Sigma}}_{X\epsilon}\right\|_\infty \lesssim C_2\left\{\frac{\log(pq)}{n}\right\}^{1/2}\|\boldsymbol{\delta}_C\|.$$

Hence with probability at least $1-\frac{4}{p}-\frac{4}{pq}$ we also have

$$\begin{aligned}&\left\|\nabla_B\left\{\mathrm{tr}[\hat{\mathbf{C}}\hat{\mathbf{\Sigma}}_{\mathbf{YY}}+\mathbf{B}^*\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XX}}-\mathbf{2}\hat{\mathbf{C}}\mathbf{B}^{*\top}\hat{\mathbf{\Sigma}}_{\mathbf{XY}}]-\log\det(\hat{\mathbf{C}})\right\}\right\|_\infty\\ \lesssim&\max\{\lambda_{\max}(\mathbf{C}^*),1/\lambda_{\min}(\mathbf{C}^*)\}\left\{\frac{\log(pq)}{n_{XY}}\right\}^{1/2}(1+\|\boldsymbol{\delta}_C\|_1),\end{aligned}$$

which is the same as stated in Lemma S9.5 if we set $\tau_1=\tau_2=-\infty$ as both $\mathbf{X}$ and $\mathbf{Y}$ are complete. □

S10 Numerical study

In this section, we show some additional results of our numerical studies. The complete results for Example 1 are shown in Table 3. The results for Example 2 are shown in Table 4. The results for Example 3 are shown in Table 5.

S11 Data processing details in the ADNI study

In section 5, we are interested in predicting Mini-Mental State Examination (MMSE), ADAS1 and ADAS2 in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (Mueller et al., 2005). These scores are commonly used diagnotic scores of AD. We extract biomarkers from three complementary data sources: serial magnetic resonance imaging (MRI), positron emission tomography (PET) and CerebroSpinal Fluid (CSF). Note that, as Xue and Qu (2021) stated, our sparsity assumption of the proposed method might not be suitable for raw imaging data or imaging data at small scales since images have to show some visible atrophy for AD. However, the sparsity assumption can still be reasonable for the region of interest (ROI) level data. Thus, we apply the Multi-DISCOM to the ROI level data in ADNI instead of the raw data.

We process the image data following the similar procedure as Yu et al. (2020). For the MRI, after correction, spatial segmentation and registration steps, we obtain the image for each subject based on the Jacob template with 93 manually labeled ROIs. For each of the 93 ROIs in the labeled MRI, we compute the volume of gray matter as a feature. For each PET image, we first align the PET image to its respective MRI using affine registration. Then, we calculate the average intensity of every ROI in the PET image as

a feature. For the CSF modality, five biomarkers were used in this study, namely amyloid $\beta(A\beta 42)$, CSF total tau (t-tau), tau hyperphosphorylated at threonine 181 (p-tau), and two tau ratios with respective to $A\beta 42$ (i.e., t-tau/$A\beta 42$ and p-tau /$A\beta 42$).

References

Cai, T. T., H. Li, W. Liu, and J. Xie (2013). Covariate-adjusted precision matrix estimation with an application in genetical genomics. *Biometrika 100*(1), 139–156.

Loh, P.-L. and M. J. Wainwright (2015). Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *The Journal of Machine Learning Research 16*(1), 559–616.

Mueller, S. G., M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett (2005). The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics 15*(4), 869–877.

Negahban, S. N., P. Ravikumar, M. J. Wainwright, and B. Yu (2012). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science 27*(4), 538–557.

	Method	$\|\hat{\mathbf{B}} - \mathbf{B}^*\|_F$	MSE	FPR	FNR
$\rho = -0.4$	Lasso	1.51(0.06)	3.70(0.06)	0.09(0.02)	**0.00(0.00)**
	Imputed-Lasso	1.73(0.06)	3.57(0.06)	0.11(0.01)	**0.00(0.00)**
	MBI	2.10(0.08)	4.26(0.09)	0.12(0.02)	0.11(0.03)
	DISCOM	1.44(0.04)	3.56(0.06)	0.05(0.00)	0.05(0.01)
	Imputed-MRCE	1.53(0.05)	3.72(0.08)	0.17(0.03)	0.08(0.02)
	Multi-DISCOM	**1.40(0.04)**	**3.39(0.08)**	**0.02(0.01)**	0.09(0.02)
$\rho = -0.2$	Lasso	1.50(0.06)	3.73(0.06)	0.10(0.02)	**0.00(0.00)**
	Imputed-Lasso	1.71(0.06)	3.59(0.06)	0.11(0.01)	**0.00(0.00)**
	MBI	2.15(0.08)	4.25(0.09)	0.12(0.02)	0.11(0.03)
	DISCOM	1.43(0.04)	3.52(0.06)	0.05(0.00)	0.05(0.01)
	Imputed-MRCE	1.52(0.05)	3.78(0.08)	0.16(0.03)	0.07(0.02)
	Multi-DISCOM	**1.41(0.04)**	**3.40(0.08)**	**0.02(0.01)**	0.09(0.02)
$\rho = 0$	Lasso	1.49(0.06)	3.67(0.06)	0.08(0.02)	**0.00(0.00)**
	Imputed-Lasso	1.71(0.06)	3.55(0.06)	0.10(0.01)	**0.00(0.00)**
	MBI	2.05(0.08)	4.21(0.09)	0.10(0.02)	0.09(0.03)
	DISCOM	1.42(0.04)	3.53(0.06)	0.04(0.00)	0.05(0.01)
	Imputed-MRCE	1.51(0.05)	3.70(0.08)	0.15(0.03)	0.09(0.02)
	Multi-DISCOM	**1.42(0.04)**	**3.43(0.08)**	**0.03(0.01)**	0.10(0.02)
$\rho = 0.2$	Lasso	1.54(0.06)	3.75(0.06)	0.10(0.02)	**0.00(0.00)**
	Imputed-Lasso	1.74(0.06)	3.59(0.06)	0.13(0.01)	**0.00(0.00)**
	MBI	2.10(0.08)	4.29(0.09)	0.11(0.02)	0.10(0.03)
	DISCOM	1.43(0.04)	3.57(0.06)	0.05(0.00)	0.05(0.01)
	Imputed-MRCE	1.53(0.05)	3.73(0.08)	0.19(0.03)	0.08(0.02)
	Multi-DISCOM	**1.41(0.04)**	**3.42(0.08)**	**0.04(0.01)**	0.07(0.02)
$\rho = 0.4$	Lasso	1.55(0.06)	3.77(0.06)	0.11(0.02)	**0.00(0.00)**
	Imputed-Lasso	1.75(0.06)	3.61(0.06)	0.13(0.01)	**0.00(0.00)**
	MBI	2.14(0.08)	4.30(0.09)	0.13(0.02)	0.11(0.03)
	DISCOM	1.46(0.04)	3.59(0.06)	0.06(0.00)	0.05(0.01)
	Imputed-MRCE	1.54(0.05)	3.73(0.08)	0.19(0.03)	0.09(0.02)
	Multi-DISCOM	**1.43(0.04)**	**3.44(0.08)**	**0.04(0.01)**	0.07(0.02)

Table 3: Performance comparison of different methods for Example 1 with different ρ's. The values in the parentheses are the standard errors of the measures.

	Method	$\|\hat{\mathbf{B}} - \mathbf{B}^*\|_F$	MSE	FPR	FNR
$\alpha = 1$	Lasso	1.33(0.08)	2.19(0.06)	0.12(0.02)	**0.00(0.00)**
	Imputed-Lasso	1.44(0.06)	2.28(0.06)	0.15(0.01)	**0.00(0.00)**
	MBI	1.68(0.19)	3.56(0.07)	0.14(0.02)	0.13(0.03)
	DISCOM	**1.29(0.06)**	**1.86(0.06)**	0.05(0.00)	0.05(0.01)
	Imputed-MRCE	1.49(0.05)	2.13(0.08)	0.18(0.03)	0.07(0.02)
	Multi-DISCOM	**1.26(0.04)**	**1.77(0.09)**	**0.03(0.02)**	0.07(0.01)
$\alpha = 3$	Lasso	1.51(0.06)	3.70(0.06)	0.09(0.02)	**0.00(0.00)**
	Imputed-Lasso	1.73(0.06)	3.57(0.06)	0.11(0.01)	**0.00(0.00)**
	MBI	2.10(0.08)	4.26(0.09)	0.12(0.02)	0.11(0.03)
	DISCOM	1.44(0.04)	3.56(0.06)	0.05(0.00)	0.05(0.01)
	Imputed-MRCE	1.53(0.05)	3.72(0.08)	0.17(0.03)	0.08(0.02)
	Multi-DISCOM	**1.40(0.04)**	**3.39(0.08)**	**0.02(0.01)**	0.09(0.02)
$\alpha = 5$	Lasso	1.81(0.06)	5.70(0.06)	0.11(0.02)	**0.01(0.00)**
	Imputed-Lasso	1.89(0.06)	5.77(0.06)	0.15(0.01)	**0.01(0.00)**
	MBI	2.37(0.10)	5.95(0.12)	0.15(0.03)	0.12(0.02)
	DISCOM	1.71(0.04)	5.41(0.08)	0.06(0.02)	0.07(0.01)
	Imputed-MRCE	1.93(0.05)	5.66(0.09)	0.18(0.03)	0.10(0.02)
	Multi-DISCOM	**1.64(0.05)**	**5.19(0.12)**	**0.04(0.03)**	0.10(0.02)

Table 4: Performance comparison of different methods for Example 2 with different signal-to-noise ratios. The values in the parentheses are the standard errors of the measures.

Lasso	1.50(0.06)	3.68(0.06)	0.09(0.02)	**0.00(0.00)**
Imputed-Lasso	1.72(0.06)	3.56(0.06)	0.12(0.01)	**0.00(0.00)**
MBI	2.11(0.08)	4.26(0.09)	0.12(0.02)	0.11(0.03)
DISCOM	1.45(0.04)	3.56(0.06)	0.05(0.00)	0.05(0.01)
Imputed-MRCE	1.55(0.05)	3.74(0.08)	0.18(0.03)	0.08(0.02)
Multi-DISCOM	**1.41(0.04)**	**3.42(0.08)**	**0.03(0.01)**	0.09(0.02)

Table 5: Performance comparison of different methods for Example 3 with heavy-tailed error. The values in the parentheses are the standard errors of the measures.

Ravikumar, P., M. J. Wainwright, G. Raskutti, and B. Yu (2011). High-dimensional covariance estimation by minimizing 1-penalized log-determinant divergence. *Electronic Journal of Statistics 5*, 935–980.

Xue, F. and A. Qu (2021). Integrating multisource block-wise missing data in model selection. *Journal of the American Statistical Association 116*(536), 1914–1927.

Yu, G., Q. Li, D. Shen, and Y. Liu (2020). Optimal sparse linear prediction for block-missing multi-modality data without imputation. *Journal of the American Statistical Association 115*(531), 1406–1419.

Zhao, P. and B. Yu (2006). On model selection consistency of lasso. *Journal of Machine Learning Research 7*(Nov), 2541–2563.