# Joint Estimation and Inference for Multiple Multi-layered Gaussian Graphical Models

Subhabrata Majumdar

**Abstract**: The rapid development of high-throughput technologies has enabled generation of data from biological processes that span multiple layers, like genomic, proteomic or metabolomic data; and pertain to multiple sources, like disease subtypes or experimental conditions. In this work we propose a general statistical framework based on graphical models for horizontal (i.e. across conditions or subtypes) and vertical (i.e. across different layers containing data on molecular compartments) integration of information in such datasets. We start with decomposing the multi-layer problem into a series of two-layer problems. For each two-layer problem, we model the outcomes at a node in the lower layer as dependent on those of other nodes in that layer, as well as all nodes in the upper layer. Following the biconvexity of our objective function, this estimation problem decomposes into two parts, where we use neighborhood selection and subsequent refitting of the precision matrix to quantify the dependency of two nodes in a single layer, and use group-penalized least square estimation to quantify the directional dependency of two nodes in different layers. Finally, to test for differences in these directional dependencies across multiple sources, we devise a hypothesis testing procedure that utilizes already computed neighborhood selection coefficients for nodes in the upper layer. We establish theoretical results for the validity of this testing procedure and the consistency of our estimates, and also evaluate their performance through simulations and a real data application.

**Keywords**: Data integration; Gaussian Graphical Models; Neighborhood selection; Group lasso

# 1 Notations

We shall denote scalars by small letters, vectors by bold small letters and matrices by bold capital letters. For any matrix $\mathbf{A}$, $(\mathbf{A})_{ij}$ denote its element in the $(i,j)^{\text{th}}$ position. For $a, b \in \mathbb{N}$, we denote the set of all $a \times b$ real matrices by $\mathbb{M}(a,b)$. For any positive integer $c$, define $\mathcal{I}_c = \{1, \ldots, c\}$.

# 2 Model

Consider the two -layered setup:

$$\mathbb{X}^k = (X_1^k, \ldots, X_p^k)^T \sim \mathcal{N}(0, \Sigma_x^k) \tag{2.1}$$

$$\mathbb{Y}^k = \mathbb{X}^k \mathbf{B}^k + \mathbb{E}^k; \quad \mathbb{E}^k = (E_1^k, \ldots, E_p^k)^T \sim \mathcal{N}(0, \Sigma_y^k) \tag{2.2}$$

$$\mathbf{B}^k \in \mathbb{M}(p,q); \quad \Omega_x^k = (\Sigma_x^k)^{-1}; \quad \Omega_y^k = (\Sigma_y^k)^{-1} \tag{2.3}$$

Want to estimate $\{(\Omega_x^k, \Omega_y^k, \mathbf{B}^k); k \in \mathcal{I}_K$ from data $\mathcal{Z}^k = \{(\mathbf{Y}^k, \mathbf{X}^k); \mathbf{Y}^k \in \mathbb{M}(n,q), \mathbf{X}^k \in \mathbb{M}(n,p), k \in \mathcal{I}_K\}$. in presence of known grouping structures $\mathcal{G}_x, \mathcal{G}_y, \mathcal{H}$ respectively.

Estimation of $\{\Omega_x^k\}$ done using JSEM. For the other part, we use the following two-step procedure:

1. Run neighborhood selection on $y$-network incorporating effects of $x$-data and an additional blockwise group penalty:

$$\min_{\mathcal{B},\Theta} \left\{ \sum_{j=1}^q \frac{1}{n_k} \left[ \sum_{k=1}^K \| \mathbf{Y}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k \mathbf{B}_{-j}^k) \boldsymbol{\theta}_j^k - \mathbf{X}^k \mathbf{B}_j^k \|^2 + \sum_{j \neq i} \sum_{g \in \mathcal{G}_y^{ij}} \lambda_{ij}^g \| \boldsymbol{\theta}_{ij}^{[g]} \| \right] + \sum_{b \in \mathcal{G}_x \times \mathcal{G}_y \times \mathcal{H}} \eta^b \| \mathbf{B}^{[b]} \| \right\} \tag{2.4}$$

$$= \min \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) + P(\Theta) + Q(\mathcal{B}) \right\} \tag{2.5}$$

where $\Theta = \{\Theta_i\}, \mathcal{B} = \{\mathbf{B}^k\}, \mathcal{Y} = \{\mathbf{Y}^k\}, \mathcal{X} = \{\mathbf{X}^k\}, \mathcal{E} = \{\mathbf{E}^k\}$.

This estimates $\mathcal{B}$ **(possibly refit and/or within-group threshold)** .

2. Step I part 2 and step II of JSEM (see 15-656 pg 6) follows to estimate $\{\Omega_y^k\}$.

The objective function is bi-convex, so we are going to do the following in step 1-

- Start with initial estimates of $\mathcal{B}$ and $\Theta$, say $\mathcal{B}^{(0)}, \Theta^{(0)}$.
- Iterate:

$$\Theta^{(t+1)} = \arg\min \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}^{(t)}, \Theta^{(t)}) + P(\Theta^{(t)}) \right\} \tag{2.6}$$

$$\mathcal{B}^{(t+1)} = \arg\min \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}^{(t)}, \Theta^{(t+1)}) + Q(\mathcal{B}^{(t)}) \right\} \tag{2.7}$$

- Continue till convergence.

# 3 Conditions

Conditions A1 from JSEM paper holds for $\mathcal{X}$ and $\mathcal{E}$. Also A2, A3 from JSEM paper.

# 4 Results

To prove the results in this section, we shall use a reparametrization of the neighborhood coefficients at the lower level. Specifically, notice that for $j \in \mathcal{I}_q, k \in \mathcal{I}_K$, the corresponding summand in $f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta)$ can be rearranged as

$$\begin{aligned}\|\mathbf{Y}_j^k - \mathbf{X}^k\mathbf{B}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k\mathbf{B}_{-j}^k)\boldsymbol{\theta}_j^k\|^2 &= \|\mathbf{Y}_j^k - \mathbf{Y}_{-j}^k\boldsymbol{\theta}_j^k - (\mathbf{X}^k\mathbf{B}_j^k - \mathbf{X}^k\mathbf{B}_{-j}^k\boldsymbol{\theta}_j^k)\|^2 \\ &= \|(\mathbf{Y} - \mathbf{X}\mathbf{B})\mathbf{T}_j^k\|^2\end{aligned}$$

where

$$T_{jj'}^k = \begin{cases} 1 \text{ if } j = j' \\ -\theta_{jj'}^k \text{ if } j \neq j' \end{cases}$$

Thus, with $\mathbf{T}^k := (\mathbf{T}_j^k)_{j \in \mathcal{I}_q}$, we have

$$f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) = \frac{1}{n}\sum_{j=1}^{p}\sum_{k=1}^{K}\|(\mathbf{Y}^k - \mathbf{X}^k\mathbf{B}^k)\mathbf{T}_j^k\|^2 = \frac{1}{n}\sum_{k=1}^{K}\|\mathbf{Y}^k - \mathbf{X}^k\mathbf{B}^k)\mathbf{T}^k\|_F^2 = \sum_{k=1}^{K}\text{Tr}(\mathbf{S}^k(\mathbf{T}^k)^2)$$

where $\mathbf{S}^k = (1/n)(\mathbf{Y}^k - \mathbf{X}^k\mathbf{B}^k)(\mathbf{Y}^k - \mathbf{X}^k\mathbf{B}^k)^T$ is the sample covariance matrix.

Now suppose $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$, and any subscript or superscript on $\mathbf{B}$ will be passed on to $\boldsymbol{\beta}$. Denote by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Theta}$ the generic estimators given by

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{pq}}{\arg\min}\left\{-2\boldsymbol{\beta}^T\widehat{\boldsymbol{\gamma}} + \boldsymbol{\beta}^T\widehat{\boldsymbol{\Gamma}}\boldsymbol{\beta} + \lambda_n\sum_{g \in \mathcal{G}}\|\boldsymbol{\beta}^{[g]}\|\right\} \tag{4.1}$$

$$\widehat{\Theta}_j = \underset{\Theta_j \in \mathbb{M}(q-1,K)}{\arg\min}\left\{\frac{1}{n}\sum_{k=1}^{K}\|\mathbf{Y}_j^k - \mathbf{X}^k\widehat{\mathbf{B}}_j^k - (\mathbf{Y}_{-j}^k - \mathbf{X}^k\widehat{\mathbf{B}}_{-j}^k)\boldsymbol{\theta}_j^k\|^2 + \gamma_n\sum_{j \neq j'}\sum_{g \in \mathcal{G}_y^{jj'}}\|\boldsymbol{\theta}_{jj'}^{[g]}\|\right\} \tag{4.2}$$

where

$$\widehat{\boldsymbol{\Gamma}} = \begin{bmatrix} (\widehat{\mathbf{T}}^1)^2 \otimes \frac{(\mathbf{X}^1)^T\mathbf{X}^1}{n} & & \\ & \ddots & \\ & & (\widehat{\mathbf{T}}^K)^2 \otimes \frac{(\mathbf{X}^K)^T\mathbf{X}^K}{n} \end{bmatrix}; \quad \widehat{\boldsymbol{\gamma}} = \begin{bmatrix} (\widehat{\mathbf{T}}^1)^2 \otimes \frac{(\mathbf{X}^1)^T}{n} \\ \vdots \\ (\widehat{\mathbf{T}}^K)^2 \otimes \frac{(\mathbf{X}^K)^T}{n} \end{bmatrix}\begin{bmatrix} \text{vec}(\mathbf{Y}^1) \\ \vdots \\ \text{vec}(\mathbf{Y}^K) \end{bmatrix}$$

with $\widehat{\mathbf{T}}^k$ defined the same way using $\widehat{\boldsymbol{\theta}}_j^k$ as we defined $\mathbf{T}^k$ using $\boldsymbol{\theta}_j^k$.

**Theorem 4.1.** *Assume fixed $\mathcal{X}, \mathcal{E}$ and deterministic $\widehat{\mathcal{B}} = \{\widehat{\mathbf{B}}^k\}$. Also for $k = 1, \ldots, K$,*
**(T1)** $\|\widehat{\mathbf{B}}^k - \mathbf{B}_0^k\|_1 \leq v_\beta$, *where* $v_\beta = \eta_\beta \sqrt{\frac{\log(\mathbf{pq})}{\mathbf{n}}}$ *with $\eta_\beta$ being a quantity depending on $\mathcal{B}$ only;*
**(T2)** *Denote $\widehat{\mathbf{E}}^k = \mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k, k \in \mathcal{I}_K$. Then for all $j \in \mathcal{I}_q$,*

$$\frac{1}{n} \left\| (\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k \mathbf{T}_{0,j}^k \right\|_\infty \leq \mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right)$$

*where $\mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right)$ is a $O(1/\sqrt{n})$ deterministic function depending on the population parameters $\mathcal{B}, \Sigma_x^k$ and $\Sigma_y^k$.*
**(T3)** *Denote $\widehat{\mathbf{S}}^k = (\widehat{\mathbf{E}}^k)^T \widehat{\mathbf{E}}^k / n$. Then $\widehat{\mathbf{S}}^k \sim RE(\psi^k, \phi^k)$ with $Kq\phi \leq \psi/2$ where $\psi = \min_k \psi^k, \phi = \max_k \phi^k$;*
**(T4)** *Assumption (A2) holds for $\Sigma_y^k$.*

*Then, given the choice of tuning parameter*

$$\gamma_n \geq 4\sqrt{q}\mathbb{Q}_{\max}; \quad \mathbb{Q}_{\max} := \max_{k \in \mathcal{I}_K} \mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right)$$

*the following holds*

$$\frac{1}{K} \sum_{k=1}^K \|\widehat{\Omega}_y^k - \Omega_y^k\|_F^2 \leq \mathbf{O}\left(\frac{\mathbf{48c_0}\sqrt{\mathbf{q}|\mathbf{g}_{\max}|\mathbf{S}}\mathbb{Q}_{\max}}{\psi}\right) \tag{4.3}$$

*where $|g_{\max}|$ is the maximum group size.*

*Proof of Theorem 4.1.* The proof has three parts, where we prove the consistency of the neighborhood regression coefficients, selection of edge sets, and finally the refitting step, respectively. This is the same structure as the proof of Theorem 1 in Ma and Michailidis (2016), where they prove consistency of the JSEM estimates. The derivation of the first part is different from that in the JSEM proof, which we shall show in detail. The second and third parts follow similar lines, incorporating the updated quantities from part 1. For these we provide a rough sketch and leave the details to the reader.

*Step 1: consistency of neighborhood regression.* The following proposition establishes error bounds for the estimated $y$-neighborhood coefficients.

**Proposition 4.2.** *Consider the estimation problem in (4.2) and choose $\gamma_n \geq 4\sqrt{q}\mathbb{Q}_{\max}$. Given the conditions (T2) and (T3) hold, for any solution of (4.2) we shall have*

$$\|\widehat{\Theta}_j - \Theta_{0,j}\|_F \leq 12\sqrt{|g_{\max}|s_j}\gamma_n/\psi \tag{4.4}$$

$$\sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\hat{\boldsymbol{\theta}}_{jj'}^{[g]} - \boldsymbol{\theta}_{0,jj'}^{[g]}\| \leq 48|g_{\max}|s_j\gamma_n/\psi \tag{4.5}$$

*Also denote the non-zero support of $\widehat{\Theta}_j$ by $\widehat{\mathcal{S}}_j$, i.e. $\widehat{\mathcal{S}}_j = \{(j', g) : \hat{\boldsymbol{\theta}}_{jj'}^{[g]} \neq \mathbf{0}\}$. Then*

$$|\widehat{\mathcal{S}}_j| \leq 128|g_{\max}|s_j/\psi \tag{4.6}$$

*Proof of Proposition 4.2.* In its reparametrized version, (4.2) becomes

$$\widehat{\mathbf{T}}_j = \arg\min_{\mathbf{T}_j} \left\{ \frac{1}{n} \sum_{k=1}^K \|(\mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k)\mathbf{T}_j^k\|^2 + \gamma_n \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\mathbf{T}_{jj'}^{[g]}\| \right\} \tag{4.7}$$

with $\mathbf{T}_{jj'}^{[g]} := (T_{jj'}^k)_{k \in g}$. Now for any $\mathbf{T}_j \in \mathbb{M}(q, K)$ we have

$$\frac{1}{n} \sum_{k=1}^K \|(\mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k)\widehat{\mathbf{T}}_j^k\|^2 + \gamma_n \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\widehat{\mathbf{T}}_{jj'}^{[g]}\| \leq \frac{1}{n} \sum_{k=1}^K \|(\mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k)\mathbf{T}_j^k\|^2 + \gamma_n \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\mathbf{T}_{jj'}^{[g]}\|$$

For $\mathbf{T}_j = \mathbf{T}_{0,j}$ this reduces to

$$\sum_{k=1}^K (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{d}_j^k \leq -2 \sum_{k=1}^K (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{T}_{0,j}^k + \gamma_n \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \left( \|\mathbf{T}_{jj'}^{[g]}\| - \|\mathbf{T}_{jj'}^{[g]} + \mathbf{d}_{jj'}^{[g]}\| \right) \tag{4.8}$$

with $\widehat{\mathbf{T}}_j^k := \mathbf{T}_{0,j}^k + \mathbf{d}_j^k$ etc. For the $k^{\text{th}}$ summand in the first term on the right hand side, since $d_{jj}^k = 0$, $\widehat{\mathbf{E}}^k \mathbf{d}_j^k = \widehat{\mathbf{E}}_{-j}^k \mathbf{d}_{-j}^k$. Thus by cauchy-schwarz inequality

$$\left| (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{T}_{0,j}^k \right| \leq \|(\mathbf{d}_j^k\| \left\| \frac{(\widehat{\mathbf{E}}^k)^T \widehat{\mathbf{E}}^k}{n} \mathbf{T}_{0,j}^k \right\| \leq \sqrt{q} \|\mathbf{d}_j^k\| \left\| \frac{1}{n} (\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k \mathbf{T}_{0,j}^k \right\|_\infty$$

which is $\leq \|\mathbf{d}_j^k\| \sqrt{q} \mathbb{Q}(v_\beta, \Sigma_x^k, \Sigma_y^k) \leq \|\mathbf{d}_j^k\| \gamma_n / 4$ by assumption (T2) and choice of $\gamma_n$. For the second term, suppose $\mathcal{S}_{0,j}$ is the support of $\Theta_{0,j}$, i.e. $\mathcal{S}_{0,j} = \{(j', g) : \boldsymbol{\theta}_{jj'}^{[g]} \neq 0\}$. Then

$$\begin{aligned}
\sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \left( \|\mathbf{T}_{jj'}^{[g]}\| - \|\mathbf{T}_{jj'}^{[g]} + \mathbf{d}_{jj'}^{[g]}\| \right) &\leq \sum_{(j', g) \in \mathcal{S}_{0,j}} \left( \|\mathbf{T}_{jj'}^{[g]}\| - \|\mathbf{T}_{jj'}^{[g]} + \mathbf{d}_{jj'}^{[g]}\| \right) - \sum_{(j', g) \notin \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \\
&\leq \sum_{(j', g) \in \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| - \sum_{(j', g) \notin \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\|
\end{aligned}$$

so that (4.8) reduces to

$$\begin{aligned}
\sum_{k=1}^K (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{d}_j^k &\leq \frac{\gamma_n}{2} \left[ \sum_{(j', g) \in \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| + \sum_{(j', g) \notin \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \right] + \gamma_n \left[ \sum_{(j', g) \in \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| - \sum_{(j', g) \notin \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \right] \\
&= \frac{3\gamma_n}{2} \sum_{(j', g) \in \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| - \frac{\gamma_n}{2} \sum_{(j', g) \notin \mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \\
&\leq \frac{3\gamma_n}{2} \sum_{j \neq j', g \in \mathcal{G}_y^{jj'}} \|\mathbf{d}_{jj'}^{[g]}\| \qquad (4.9)
\end{aligned}$$

Since the left hand side is $\geq 0$, this also implies

$$\sum_{(j',g)\notin\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \leq 3 \sum_{(j',g)\in\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \quad \Rightarrow \sum_{j\neq j', g\in\mathcal{G}_y^{jj'}} \|\mathbf{d}_{jj'}^{[g]}\| \leq 4 \sum_{(j',g)\in\mathcal{S}_{0,j}} \|\mathbf{d}_{jj'}^{[g]}\| \leq 4\sqrt{|g_{\max}|s_j}\|\mathbf{D}_j\|_F$$

with $\mathbf{D}_j = (\mathbf{d}_j^k)_{k\in\mathcal{I}_K}$. Now the RE condition on $\widehat{\mathbf{S}}^k$ means that

$$\sum_{k=1}^K (\mathbf{d}_j^k)^T \widehat{\mathbf{S}}^k \mathbf{d}_j^k \geq \sum_{k=1}^K \left(\psi_k\|\mathbf{d}_j^k\|^2 - \phi_k\|\mathbf{d}_j^k\|_1^2\right) \geq \psi\|\mathbf{D}_j\|_F^2 - \phi\|\mathbf{D}_j\|_1^2 \geq (\psi - Kq\phi)\|\mathbf{D}_j\|_F^2 \geq \frac{\psi}{2}\|\mathbf{D}_j\|_F^2$$

by assumption (T3). Thus we finally have

$$\frac{\psi}{3}\|\mathbf{D}_j\|_F^2 \leq \gamma_n \sum_{j\neq j', g\in\mathcal{G}_y^{jj'}} \|\mathbf{d}_{jj'}^{[g]}\| \leq 4\gamma_n\sqrt{|g_{\max}|s_j}\|\mathbf{D}_j\|_F \tag{4.10}$$

Since

$$(\mathbf{D}_j)_{j',k} = \widehat{T}_{jj'}^k - T_{0,jj'}^k = \begin{cases} 0 \text{ if } j = j' \\ -(\widehat{\theta}_{jj'}^k - \theta_{0,jj'}^k) \text{ if } j \neq j' \end{cases}$$

The bounds in (4.4) and (4.5) are obtained by replacing the corresponding elements in (4.10).

For the bound on $|\widehat{\mathcal{S}}_j|$, notice that if $\hat{\boldsymbol{\theta}}_{jj'}^{[g]} \neq 0$ for some $(j', g)$,

$$\begin{aligned}\frac{1}{n}\sum_{k\in g}\left|((\widehat{\mathbf{E}}_{-j}^k)^T\widehat{\mathbf{E}}^k(\widehat{\mathbf{T}}_j^k - \mathbf{T}_{0,j}^k))^{j'}\right| &\geq \frac{1}{n}\sum_{k\in g}\left|((\widehat{\mathbf{E}}_{-j}^k)^T\widehat{\mathbf{E}}^k\widehat{\mathbf{T}}_j^k)^{j'}\right| - \frac{1}{n}\sum_{k\in g}\left|((\widehat{\mathbf{E}}_{-j}^k)^T\widehat{\mathbf{E}}^k\mathbf{T}_{0,j}^k)^{j'}\right| \\ &\geq |g|\gamma_n - \sum_{k\in g}\mathbb{Q}(v_\beta, \Sigma_x^k, \Sigma_y^k)\end{aligned}$$

using the KKT condition for (4.2) and assumption (T2). The choice of $\gamma_n$ now ensures that the right hand side is $\geq 3|g|\gamma_n/4$. Hence

$$\begin{aligned}|\hat{\mathcal{S}}_j| &\leq \sum_{(j',g)\in\widehat{\mathcal{S}}_j} \frac{16}{9n^2|g|^2\gamma_n^2}\sum_{k\in g}\left|((\widehat{\mathbf{E}}_{-j}^k)^T\widehat{\mathbf{E}}^k(\widehat{\mathbf{T}}_j^k - \mathbf{T}_{0,j}^k))^{j'}\right|^2 \\ &\leq \frac{16}{9\gamma_n^2}\sum_{k=1}^K\frac{1}{n}\left\|(\widehat{\mathbf{E}}_{-j}^k)^T\widehat{\mathbf{E}}^k(\widehat{\mathbf{T}}_j^k - \mathbf{T}_{0,j}^k)\right\|^2 \\ &= \frac{16}{9\gamma_n^2}\sum_{k=1}^K(\mathbf{d}_j^k)^T\widehat{\mathbf{S}}^k\mathbf{d}_j^k \\ &\leq \frac{8}{3\gamma_n}\sum_{j\neq j', g\in\mathcal{G}_y^{jj'}}\|\mathbf{d}_{jj'}^{[g]}\| \leq \frac{1}{\psi}128|g_{\max}|s_j\end{aligned}$$

using (4.9) and (4.10).

*Step 2: Edge set selection.* We denote the selected edge set for the $k^{\text{th}}$ Y-network by $\hat{E}^k$. Denote its population version by $E_0^k$. Further, let

$$\tilde{\Omega}_y^k = \operatorname{diag}(\Omega_y^k) + \Omega_{y,E_0^k \cap \hat{E}^k}^k$$

With similar derivations to the proof of Corollary A.1 in Ma and Michailidis (2016), The following two upper bounds can be established:

$$|\hat{E}^k| \leq \frac{128|g_{\max}|S}{\psi} \tag{4.11}$$

$$\frac{1}{K}\sum_{k=1}^{K} \|\tilde{\Omega}_y^k - \Omega_y^k\|_F \leq \frac{12c_0\sqrt{|g_{\max}|S}\gamma_n}{\sqrt{K}\psi} \tag{4.12}$$

following which, taking $\gamma_n = 4\sqrt{q}\mathbb{Q}_{\max}$ and

$$\Lambda_{\min}(\tilde{\Omega}_y^k) \geq d_0 - 12\sqrt{|g_{\max}|S}\gamma_n/\psi > 0; \quad \Lambda_{\max}(\tilde{\Omega}_y^k) \leq c_0 + 12\sqrt{|g_{\max}|S}\gamma_n/\psi < \infty \tag{4.13}$$

with

$$t_1 = \textbf{tbd}$$

*Step 3: Refitting.*

□

*Part II.* Proof of Thm 2 in 15-656 follows. We only need a new bound for $Var(\mathbf{Y}_i^k|\mathbf{Y}_{-i}^k, \mathbf{X}^k, \widehat{\mathbf{B}}_i^k)$. For this we have

$$Var(\mathbf{Y}_i^k|\mathbf{Y}_{-i}^k, \mathbf{X}^k, \widehat{\mathbf{B}}_i^k) = \mathbb{E}(\widehat{\boldsymbol{\epsilon}}_i^k)^2 = \mathbb{E}(\boldsymbol{\epsilon}_i^k + \boldsymbol{\delta}_i^k)^2 \leq \left(\frac{1}{d_0} + \frac{c(v_\beta)}{n}\right)^2$$

applying cauchy-schwarz inequality followed by assumption (A2). Now Replace $1/\sqrt{nd_0}$ in choice of $\lambda, \alpha_n$ in Thm 2 statement with $1/\sqrt{n}(\sqrt{1/d_0} + \sqrt{c(v_\beta)/n})$.

□

**Proposition 4.3.** *Consider deterministic $\widehat{\mathcal{B}}$ satisfying assumption (T1). Then for sample size $n \gtrsim \log(pq)$ and $k \in \mathcal{I}_K$,*

1. $\widehat{\mathbf{S}}^k$ *satisfies the RE condition:* $\widehat{\mathbf{S}}^k \sim RE(\psi^k, \phi^k)$, *where*

$$\psi^k = \frac{\Lambda_{\min}(\Sigma_x^k)}{2}; \quad \phi^k = \frac{\psi^k \log p}{n} + 2v_\beta c_2[\Lambda_{\max}(\Sigma_x^k)\Lambda_{\max}(\Sigma_y^k)]^{1/2}\sqrt{\frac{\log(pq)}{n}}$$

*with probability* $\geq 1 - 6c_1\exp[-(c_2^2 - 1)\log(pq)] - 2\exp(-c_3 n), c_1, c_3 > 0, c_2 > 1.$

*2. The following deviation bound is satisfied for any $j \in \mathcal{I}_q$*

$$\left\| \frac{1}{n} (\widehat{\mathbf{E}}_{-j}^k)^T \widehat{\mathbf{E}}^k \mathbf{T}_{0,j}^k \right\|_\infty \leq \mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right)$$

*with probability* $\geq 1 - 1/p^{\tau_1 - 2} - 6c_1 \exp[-(c_2^2 - 1)\log(pq)] - 6c_4 \exp[-(c_5^2 - 1)\log(pq)], c_4 > 0, c_5 > 1,$ *where*

$$\begin{aligned}
\mathbb{Q}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right) = 2v_\beta^2 & \left[ \sqrt{\frac{\log 4 + \tau_1 \log p}{c_x^k n}} + \max_i \sigma_{x,ii}^k \right] + \\
& 4 v_\beta c_2 [\Lambda_{\max}(\Sigma_x^k) \Lambda_{\max}(\Sigma_y^k)]^{1/2} \sqrt{\frac{\log(pq)}{n}} + \\
& c_5 \left[ \Lambda_{\max}(\Sigma_{y,-j}^k)(\sigma_{y,jj}^k + (\boldsymbol{\theta}_{0,j}^k)^T \Sigma_{y,-j}^k \boldsymbol{\theta}_{0,j}^k - 2(\boldsymbol{\sigma}_{y,j,-j}^k)^T \boldsymbol{\theta}_{0,j}^k) \right]^{1/2} \sqrt{\frac{\log(q-1)}{n}}
\end{aligned}$$

*Proof of Proposition 4.3.* We drop the superscript $k$ since there is no scope of ambiguity. For part 1, we start with an auxiliary lemma:

**Lemma 4.4.** *For a sub-gaussian design matrix $\mathbf{X} \in \mathbb{M}(n,p)$ with columns having mean $\mathbf{0}_p$ and covariance matrix $\Sigma_x$, the sample covariance matrix $\widehat{\Sigma}_x = \mathbf{X}^T \mathbf{X}/n$ satisfies the RE condition*

$$\widehat{\Sigma}_x \sim RE\left( \frac{\Lambda_{\min}(\Sigma_x)}{2}, \frac{\Lambda_{\min}(\Sigma_x) \log p}{2n} \right)$$

*with probability $\geq 1 - 2\exp(-c_3 n)$ for some $c_3 > 0$.*

This is same as Lemma 2 in Appendix B of Lin et al. (2016) and its proof can be found there. Now denote $\widehat{\mathbf{E}} = \mathbf{Y} - \mathbf{X}\widehat{\mathbf{B}}$. For $\mathbf{v} \in \mathbb{R}^q$, we have

$$\begin{aligned}
\mathbf{v}^T \widehat{\mathbf{S}} \mathbf{v} &= \frac{1}{n} \|\widehat{\mathbf{E}} \mathbf{v}\|^2 \\
&= \frac{1}{n} \|(\mathbf{E} + \mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}}))\mathbf{v}\|^2 \\
&= \mathbf{v}^T \mathbf{S} \mathbf{v} + \frac{1}{n} \|\mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}})\mathbf{v}\|^2 + 2\mathbf{v}^T (\mathbf{B}_0 - \widehat{\mathbf{B}})^T \left( \frac{(\mathbf{X})^T \mathbf{E}}{n} \right) \mathbf{v}
\end{aligned} \tag{4.14}$$

For the first summand, $\mathbf{v}^T \mathbf{S}^k \mathbf{v} \geq \psi_y \|\mathbf{v}\|^2 - \phi_y \|\mathbf{v}\|_1^2$ with $\psi_y = \Lambda_{\min}(\Sigma_y)/2, \phi_y = \psi_y \log p/n$ by applying Lemma 4.4 on $\mathbf{S}$. The second summand is greater than or equal to 0. For the third summand,

$$2\mathbf{v}^T (\mathbf{B}_0 - \widehat{\mathbf{B}})^T \left( \frac{(\mathbf{X})^T \mathbf{E}}{n} \right) \mathbf{v} \geq -2v_\beta \left\| \frac{(\mathbf{X})^T \mathbf{E}}{n} \right\|_\infty \|\mathbf{v}\|_1^2$$

by assumption (T1). Now we use another lemma:

**Lemma 4.5.** *For zero-mean independent sub-gaussian matrices* $\mathbf{X} \in \mathbb{M}(n,p), \mathbf{E} \in \mathbb{M}(n,q)$ *with parameters* $(\Sigma_x, \sigma_x^2)$ *and* $(\Sigma_e, \sigma_e^2)$ *respectively, given that* $n \gtrsim \log(pq)$ *the following holds with probability* $\geq 1 - 6c_1 \exp[-(c_2^2 - 1)\log(pq)]$ *for some* $c_1 > 0, c_2 > 1$*:*

$$\frac{1}{n}\|\mathbf{X}^T\mathbf{E}\|_\infty \leq c_2[\Lambda_{\max}(\Sigma_x)\Lambda_{\max}(\Sigma_e)]^{1/2}\sqrt{\frac{\log(pq)}{n}}$$

This is a part of Lemma 3 of Appendix B in Lin et al. (2016), and is proved therein. Subsequently we collect all summands in (4.14) and get

$$\mathbf{v}^T\widehat{\mathbf{S}}\mathbf{v} \geq \psi_y\|\mathbf{v}\|^2 - \left(\phi_y + 2v_\beta c_2[\Lambda_{\max}(\Sigma_x)\Lambda_{\max}(\Sigma_y)]^{1/2}\sqrt{\frac{\log(pq)}{n}}\right)\|\mathbf{v}\|_1^2$$

with probability $\geq 1 - 2\exp(-c_3 n) - 6c_1\exp[-(c_2^2-1)\log(pq)]$. This concludes the proof of part 1.

To prove part 2, we decompose the quantity in question:

$$\begin{aligned}
\left\|\frac{1}{n}\widehat{\mathbf{E}}_{-j}^T\widehat{\mathbf{E}}\mathbf{T}_{0,j}\right\|_\infty &= \left\|\frac{1}{n}\left[\mathbf{E}_{-j} + \mathbf{X}(\mathbf{B}_{0,j} - \widehat{\mathbf{B}}_j)\right]^T\left[\mathbf{E} + \mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}})\right]\mathbf{T}_{0,j}\right\|_\infty \\
&\leq \left\|\frac{1}{n}\mathbf{E}_{-j}^T\mathbf{E}\mathbf{T}_{0,j}\right\|_\infty + \left\|\frac{1}{n}\mathbf{E}_{-j}^T\mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}})\mathbf{T}_{0,j}\right\|_\infty \\
&+ \left\|\frac{1}{n}(\mathbf{B}_{0,j} - \widehat{\mathbf{B}}_j)^T\mathbf{X}^T\mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}})\mathbf{T}_{0,j}\right\|_\infty + \left\|\frac{1}{n}(\mathbf{B}_{0,j} - \widehat{\mathbf{B}}_j)^T\mathbf{X}^T\mathbf{E}\mathbf{T}_{0,j}\right\|_\infty \\
&= \|\mathbf{W}_1\|_\infty + \|\mathbf{W}_2\|_\infty + \|\mathbf{W}_3\|_\infty + \|\mathbf{W}_4\|_\infty \qquad (4.15)
\end{aligned}$$

Now

$$\mathbf{W}_1 = \frac{1}{n}\mathbf{E}_{-j}^T(\mathbf{E}_j - \mathbf{E}_{-j}\boldsymbol{\theta}_{0,j})$$

For node $j$ in the $y$-network, $\mathbb{E}_{-j}$ and $E_j - \mathbb{E}_{-j}\boldsymbol{\theta}_{0,j}$ are the neighborhood regression coefficients and residuals, respectively. Thus they are orthogonal, so we can apply Lemma 4.5 on $\mathbf{E}_{-j}$ and $\mathbf{E}_j - \mathbf{E}_{-j}\boldsymbol{\theta}_{0,j}$ to obtain that for $n \gtrsim \log(q-1)$,

$$\|\mathbf{W}_1\|_\infty \leq c_5\left[\Lambda_{\max}(\Sigma_{y,-j})(\sigma_{y,jj} + \boldsymbol{\theta}_{0,j}^T\Sigma_{y,-j}\boldsymbol{\theta}_{0,j} - 2\boldsymbol{\sigma}_{y,j,-j}^T\boldsymbol{\theta}_{0,j})\right]^{1/2}\sqrt{\frac{\log(q-1)}{n}} \qquad (4.16)$$

holds with probability $\geq 1 - 6c_4\exp[-(c_5^2-1)\log(pq)]$ for some $c_4 > 0, c_5 > 1$.

The same bounds hold for $\mathbf{W}_2$ and $\mathbf{W}_4$:

$$\begin{aligned}
\|\mathbf{W}_2\|_\infty &\leq \left\|\frac{1}{n}\mathbf{E}_{-j}^T\mathbf{X}(\mathbf{B}_0 - \widehat{\mathbf{B}})\right\|_\infty\|\mathbf{T}_{0,j}\|_1 \leq \left\|\frac{1}{n}\mathbf{E}^T\mathbf{X}\right\|_\infty\|\mathbf{B}_0 - \widehat{\mathbf{B}}\|_1\|\mathbf{T}_{0,j}\|_1 \\
\|\mathbf{W}_4\|_\infty &\leq \left\|\frac{1}{n}(\mathbf{B}_{0,j} - \widehat{\mathbf{B}}_j)^T\mathbf{X}^T\mathbf{E}\right\|_\infty\|\mathbf{T}_{0,j}\|_1 \leq \left\|\frac{1}{n}\mathbf{E}^T\mathbf{X}\right\|_\infty\|\mathbf{B}_0 - \widehat{\mathbf{B}}\|_1\|\mathbf{T}_{0,j}\|_1
\end{aligned}$$

Now since $\Omega_y$ is diagonally dominant, $|\omega_{y,jj}| \geq \sum_{j \neq j'} |\omega_{y,jj'}|$ for any $j \in \mathcal{I}_q$. Hence

$$\|\mathbf{T}_{0,j}\|_1 = \sum_{j'=1}^{q} |T_{jj'}| = 1 + \sum_{j \neq j'} |\theta_{jj'}| = 1 + \frac{1}{\omega_{y,jj}} \sum_{j \neq j'} |\omega_{y,jj'}| \leq 2$$

so that for $n \gtrsim \log(pq)$,

$$\|\mathbf{W}_2\|_\infty + \|\mathbf{W}_4\|_\infty \leq 4 v_\beta c_2 [\Lambda_{\max}(\Sigma_x)\Lambda_{\max}(\Sigma_y)]^{1/2} \sqrt{\frac{\log(pq)}{n}} \tag{4.17}$$

with probability $\geq 1 - 6c_1 \exp[-(c_2^2 - 1)\log(pq)]$ by applying Lemma 4.5 and assumption (T1).

Finally for $\mathbf{W}_3$, we apply Lemma 8 of Ravikumar et al. (2011) on the (sub-gaussian) design matrix $\mathbf{X}$ to obtain that for sample size

$$n \geq 512(1 + 4\Lambda_{\max}(\Sigma_x^k))^4 \max_i (\sigma_{x,ii}^k)^4 \log(4p^{\tau_1})$$

we get that with probability $\geq 1 - 1/p^{\tau_1 - 2}, \tau_1 > 2$,

$$\left\| \frac{\mathbf{X}^T\mathbf{X}}{n} \right\|_\infty \leq \sqrt{\frac{\log 4 + \tau_1 \log p}{c_x n}} + \max_i \sigma_{x,ii}; \quad c_x = \left[128(1 + 4\Lambda_{\max}(\Sigma_x))^2 \max_i (\sigma_{x,ii})^2\right]^{-1}$$

Thus with the same probability,

$$\|\mathbf{W}_4\|_\infty \leq \left\| \frac{\mathbf{X}^T\mathbf{X}}{n} \right\|_\infty \|\widehat{\mathbf{B}} - \mathbf{B}_0\|_1^2 \|\mathbf{T}_{0,j}\|_1 \leq 2v_\beta^2 \left[ \sqrt{\frac{\log 4 + \tau_1 \log p}{c_x n}} + \max_i \sigma_{x,ii} \right] \tag{4.18}$$

We now bound the right hand side of (4.15) using (4.16), (4.17) and (4.18) to complete the proof, with the leading term of the sample size requirement being $n \gtrsim \log(pq)$.

$\square$

---

Now concentrate on the $k$-population estimation problem. We want to obtain

$$\widehat{\boldsymbol{\beta}} = \operatorname*{arg\,min}_{\boldsymbol{\beta} \in \mathbb{R}^{pq}} \left\{ -2\boldsymbol{\beta}^T \widehat{\boldsymbol{\gamma}} + \boldsymbol{\beta}^T \widehat{\boldsymbol{\Gamma}} \boldsymbol{\beta} + \lambda_n \sum_{g \in \mathcal{G}} \|\boldsymbol{\beta}^{[g]}\| \right\}$$

**Theorem 4.6.** *Assume fixed $\mathcal{X}, \mathcal{E}$, and deterministic $\widehat{\Theta} = \{\widehat{\Theta}_j\}$. Also for $j \in \mathcal{I}_q$,*
**(B1)** $\|\widehat{\Theta}_j - \Theta_{0,j}\|_F \leq v_\Theta \sqrt{\frac{\log q}{n}}$ *for some $v_\Theta$ dependent on $\Theta$.*
**(B2)** *Denote $\widehat{\boldsymbol{\Gamma}}^k = (\widehat{\mathbf{T}}^k)^2 \otimes (\mathbf{X}^k)^T \mathbf{X}^k / n, \widehat{\boldsymbol{\gamma}}^k = (\widehat{\mathbf{T}}^k)^2 \otimes (\mathbf{X}^k)^T \mathbf{Y}^k / n$. Then the deviation bound holds:*

$$\left\| \widehat{\boldsymbol{\gamma}}^k - \widehat{\boldsymbol{\Gamma}}^k \boldsymbol{\beta}_0 \right\|_\infty \leq \mathbb{R}(v_\Theta, \Sigma_x^k, \Sigma_y^k)$$

**tbd**

**(B3)** $\widehat{\mathbf{\Gamma}} \sim RE(\psi_*, \phi_*)$ *with* $Kpq\phi_* \leq \psi_*/2$.

*Then, given the choice of tuning parameter*

$$\lambda \geq 4\sqrt{pq}\mathbb{R}_{\max}; \quad \mathbb{R}_{\max} := \max_{k \in \mathcal{I}_K} \mathbb{R}\left(v_\Theta, \Sigma_x^k, \Sigma_y^k\right)$$

*the following holds*

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \leq 12\sqrt{s_\beta}\lambda_n/\psi^* \tag{4.19}$$

$$\sum_{g \in \mathcal{G}} \|\boldsymbol{\beta}^{[g]} - \boldsymbol{\beta}_0^{[g]}\| \leq 48 s_\beta \lambda_n/\psi^* \tag{4.20}$$

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \widehat{\mathbf{\Gamma}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq 72 s_\beta \lambda_n^2/\psi^* \tag{4.21}$$

*Proof of Theorem 4.6.* The proof follows that of Theorem 4.1, with a different group norm structure. We only point our the differences.

Putting $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ in (4.1) we get

$$-2\widehat{\boldsymbol{\beta}}^T \widehat{\boldsymbol{\gamma}} + \boldsymbol{\beta}^T \widehat{\mathbf{\Gamma}} \widehat{\boldsymbol{\beta}} + \lambda_n \sum_{g \in \mathcal{G}} \|\widehat{\boldsymbol{\beta}}^{[g]}\| \leq -2\boldsymbol{\beta}_0^T \widehat{\boldsymbol{\gamma}} + \boldsymbol{\beta}_0^T \widehat{\mathbf{\Gamma}} \boldsymbol{\beta}_0 + \lambda_n \sum_{g \in \mathcal{G}} \|\boldsymbol{\beta}_0^{[g]}\|$$

Denote $\mathbf{b} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$. Then we have

$$\mathbf{b}^T \widehat{\mathbf{\Gamma}} \mathbf{b} \leq 2\mathbf{b}^T (\widehat{\boldsymbol{\gamma}} - \widehat{\mathbf{\Gamma}} \boldsymbol{\beta}_0) + \lambda_n \sum_{g \in \mathcal{G}} (\|\boldsymbol{\beta}_0^{[g]}\| - \|\boldsymbol{\beta}_0^{[g]} + \mathbf{b}^{[g]}\|)$$

Proceeding similarly as the proof of Theorem 4.1, with a different deviation bound and choice of $\lambda_n$ based on that, we get expressions equivalent to (4.9) and (4.10) respectively:

$$\mathbf{b}^T \widehat{\mathbf{\Gamma}} \mathbf{b} \leq \frac{3}{2} \sum_{g \in \mathcal{G}} \|\mathbf{b}^{[g]}\| \tag{4.22}$$

$$\frac{\psi^*}{3} \|\mathbf{b}\|^2 \leq \lambda_n \sum_{g \in \mathcal{G}} \|\mathbf{b}^{[g]}\| \leq 4\lambda_n \sqrt{s_\beta} \|\mathbf{b}\| \tag{4.23}$$

The bounds in (4.19), (4.20), (4.21) follow.

□

**Proposition 4.7.** *Consider deterministic* $\widehat{\Theta}$ *satisfying assumption (B1). Then for sample size* $n \gtrsim \log(pq)$,

1. $\widehat{\mathbf{\Gamma}}$ *satisfies the RE condition:* $\widehat{\mathbf{\Gamma}} \sim RE(\psi_*, \phi_*)$, *where*

$$\psi_* = \min_k \psi^k \left( \min_i \psi_t^i - d v_\Theta \right), \phi_* = \max_k \phi^k \left( \min_i \phi_t^i + d v_\Theta \right)$$

*with probability* $\geq 1 - 2\exp(c_3 n), c_3 > 0$.

*2. The following deviation bound is satisfied:*

$$\left\| \widehat{\gamma} - \widehat{\Gamma} \boldsymbol{\beta}_0 \right\|_\infty \leq \mathbb{R}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right) \sqrt{\frac{\log(pq)}{n}}$$

*with probability* $1 - 12c_1 \exp[(c_2^2 - 1)\log(pq)], c_1 > 0, c_2 > 1,$ *where*

$$\mathbb{R}\left(v_\beta, \Sigma_x^k, \Sigma_y^k\right) = c_2 \sqrt{\Lambda_{\max}(\Sigma_x^k)} \left( d v_\Theta \Lambda_{\min}(\Sigma_y^k) + \frac{1}{\Lambda_{\min}(\Sigma_y^k)} \right)$$

*Proof of Proposition 4.7.* For part 1 it is enough to prove that with $\widehat{\Sigma}_x^k := (\mathbf{X}^k)^T \mathbf{X}^k / n$,

$$\widehat{\mathbf{T}}_k^2 \otimes \widehat{\Sigma}_x^k \sim RE(\psi_*^k, \phi_*^k) \tag{4.24}$$

with high enough probability. because then we can take $\psi_* = \min_k \psi_*^k, \phi_* = \max_k \phi_*^k$. The proof of (4.24) follows similar lines of the proof of Proposition 1 in Lin et al. (2016), only replacing $\Theta_\epsilon, \widehat{\Theta}_\epsilon, \mathbf{X}$ therein with $(\mathbf{T}^k)^2, (\widehat{\mathbf{T}}^k)^2, \mathbf{X}^k$, respectively. We omit the details. □

---

# 5 Two-sample testing

Suppose there are two disease subtypes: $k = 1, 2$, and we are interested in testing whether the downstream effect of a predictor in X-data is same across both subtypes, i.e. if $\mathbf{b}_i^1 = \mathbf{b}_i^2$ for some $i \in \mathcal{I}_p$. For this we consider the modified optimization problem:

$$\begin{aligned} &\min_{\mathcal{B}, \Theta} \frac{1}{n} \left\{ \sum_{j=1}^q \sum_{k=1}^2 \| \mathbf{Y}_j^k - \mathbf{Y}_{-j}^k \boldsymbol{\theta}_j^k - \mathbf{X}^k \mathbf{B}_j^k \|^2 + \sum_{j \neq j'} \lambda_{jj'} \| \boldsymbol{\theta}_{jj'}^* \| + \sum_{i=1}^p \eta_i \| \mathbf{B}_{i*}^* \| \right\} \\ &= \min \left\{ f(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \Theta) + P(\Theta) + Q(\mathcal{B}) \right\} \end{aligned} \tag{5.1}$$

with $n_1 = n_2 = n$ for simplicity; and $\mathbf{B}^k = (\mathbf{b}_1^k, \ldots, \mathbf{b}_q^k), (\mathbf{B}_{i*}^*) \in \mathbb{R}^{q \times K}$.

In this setup, define the desparsified estimate of $\mathbf{b}_i^k$ as

$$\widehat{\mathbf{c}}_i^k = \widehat{\mathbf{b}}_i^k + \frac{1}{n t_i^k} \left( \mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\boldsymbol{\zeta}}_i^k \right)^T (\mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k) \tag{5.2}$$

for $k = 1, 2$, where $t_i^k = (\mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\boldsymbol{\zeta}}_i^k)^T \mathbf{X}_{-i}^k / n$. Then we have the asymptotic joint distribution of a scaled version of the debiased coefficients for the $i^{\text{th}}$ predictor effect.

**Theorem 5.1.** *Define* $\widehat{s}_i^k = \sqrt{\| \mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\boldsymbol{\zeta}}_i^k \|^2 / n}$, *and* $m_i^k = \sqrt{n} t_i^k / \widehat{s}_i^k$. *Assume the following:*

**(C1)** *For the X-neighborhood estimators we have* $\|\widehat{\boldsymbol{\zeta}}^k - \boldsymbol{\zeta}_0^k\|_1 \leq v_\zeta$.
**(C2)** *The precision matrix estimators satisfy*

$$\left\| \widehat{(\Omega_y^k)}^{1/2} - (\Omega_y^k)^{1/2} \right\|_\infty \leq v_\Omega$$

*Then for the debiased estimators in* (5.2) *we have*

$$\begin{bmatrix} \widehat{\Omega}_y^1 & \\ & \widehat{\Omega}_y^2 \end{bmatrix}^{1/2} \begin{bmatrix} m_i^1(\widehat{\mathbf{c}}_i^1 - \mathbf{b}_i^1) & \\ & m_i^2(\widehat{\mathbf{c}}_i^2 - \mathbf{b}_i^2) \end{bmatrix} \sim \mathcal{N}_{2q}(\mathbf{0}, \mathbf{I}) + \mathbf{S}_n \tag{5.3}$$

*where*

$$\|\mathbf{S}_n\|_\infty \leq \textbf{tbd}$$

*with probability* $\geq$ **tbd**.

*Proof of Theorem 5.1.* Let us define the following:

$$\begin{aligned} \mathbf{M}_i &= \text{diag}(m_i^1, m_i^2) \\ \widehat{\mathbf{C}}_i &= \text{diag}(\widehat{\mathbf{c}}_i^1, \widehat{\mathbf{c}}_i^2) \\ \widehat{\mathbf{D}}_i &= \text{diag}(\widehat{\mathbf{b}}_i^1, \widehat{\mathbf{b}}_i^2) \\ \mathbf{D}_i &= \text{diag}(\mathbf{b}_{0,i}^1, \mathbf{b}_{0,i}^2) \\ \mathbf{R}_i^k &= \mathbf{X}_i^k - \mathbf{X}_{-i}^k \widehat{\boldsymbol{\zeta}}_i^k; k = 1, 2 \end{aligned}$$

Then from (5.2) we have

$$\mathbf{M}_i(\widehat{\mathbf{C}}_i - \widehat{\mathbf{D}}_i)^T = \frac{1}{\sqrt{n}} \begin{bmatrix} \frac{1}{\widehat{s}_i^1}(\mathbf{R}_i^1)^T \widehat{\mathbf{E}}^1 \\ \frac{1}{\widehat{s}_i^2}(\mathbf{R}_i^2)^T \widehat{\mathbf{E}}^2 \end{bmatrix} \tag{5.4}$$

We now decompose $\widehat{\mathbf{E}}^k$ :

$$\begin{aligned} \widehat{\mathbf{E}}^k &= \mathbf{Y}^k - \mathbf{X}^k \widehat{\mathbf{B}}^k \\ &= \mathbf{E}^k + \mathbf{X}^k(\mathbf{B}_0^k - \widehat{\mathbf{B}}^k) \\ &= \mathbf{E}^k + \mathbf{X}_i^k(\mathbf{b}_{0,i}^k - \widehat{\mathbf{b}}_i^k) + \mathbf{X}_{-i}^k(\mathbf{B}_{0,-i}^k - \widehat{\mathbf{B}}_{-i}^k) \end{aligned}$$

Putting them back in (5.4) and using $t^k = (\mathbf{R}^k)^T \mathbf{X}^k / n$,

$$\mathbf{M}_i(\widehat{\mathbf{C}}_i - \widehat{\mathbf{D}}_i)^T = \frac{1}{\sqrt{n}} \begin{bmatrix} \frac{1}{\widehat{s}_i^1}(\mathbf{R}_i^1)^T \mathbf{E}^1 \\ \frac{1}{\widehat{s}_i^2}(\mathbf{R}_i^2)^T \mathbf{E}^2 \end{bmatrix} + \mathbf{M}_i(\mathbf{D}_i - \widehat{\mathbf{D}}_i)^T + \frac{1}{\sqrt{n}} \begin{bmatrix} \frac{1}{\widehat{s}_i^1}(\mathbf{R}_i^1)^T \mathbf{X}_{-i}^1(\mathbf{B}_{0,-i}^1 - \widehat{\mathbf{B}}_{-i}^1) \\ \frac{1}{\widehat{s}_i^2}(\mathbf{R}_i^2)^T \mathbf{X}_{-i}^2(\mathbf{B}_{0,-i}^2 - \widehat{\mathbf{B}}_{-i}^2) \end{bmatrix} \tag{5.5}$$

$$\Rightarrow \mathbf{M}_i(\widehat{\mathbf{C}}_i - \mathbf{D}_i)^T = \frac{1}{\sqrt{n}} \begin{bmatrix} \frac{1}{\widehat{s}_i^1}(\mathbf{R}_i^1)^T \mathbf{E}^1 \\ \frac{1}{\widehat{s}^2}(\mathbf{R}_i^2)^T \mathbf{E}^2 \end{bmatrix} + \frac{1}{\sqrt{n}} \begin{bmatrix} \frac{1}{\widehat{s}_i^1}(\mathbf{R}_i^1)^T \mathbf{X}_{-i}^1(\mathbf{B}_{0,-i}^1 - \widehat{\mathbf{B}}_{-i}^1) \\ \frac{1}{\widehat{s}_i^2}(\mathbf{R}_i^2)^T \mathbf{X}_{-i}^2(\mathbf{B}_{0,-i}^2 - \widehat{\mathbf{B}}_{-i}^2) \end{bmatrix} \tag{5.6}$$

At this point, to prove (5.3), it is enough to show the following:

$$\frac{1}{\sqrt{n\widehat{s}_i}}\widehat{\Omega}_y^{1/2}\mathbf{E}^T\mathbf{R}_i \sim \mathcal{N}_q(\mathbf{0},\mathbf{I}) + \mathbf{S}_{1n}; \quad \|\mathbf{S}_{1n}\|_\infty \leq \textbf{tbd} \tag{5.7}$$

$$\left\| \frac{1}{\sqrt{n\widehat{s}_i}}\mathbf{R}_i^T\mathbf{X}_{-i}(\mathbf{B}_{0,-i} - \widehat{\mathbf{B}}_{-i})\widehat{\Omega}_y^{1/2} \right\|_\infty \leq \textbf{tbd} \tag{5.8}$$

dropping $k$ in the superscripts. To show (5.9) we have

$$\frac{1}{\sqrt{n\widehat{s}_i}}\widehat{\Omega}_y^{1/2}\mathbf{E}^T\mathbf{R}_i = \frac{1}{\sqrt{n\widehat{s}_i}}(\widehat{\Omega}_y^{1/2} - \Omega_y^{1/2})\mathbf{E}^T\mathbf{R}_i + \frac{1}{\sqrt{n\widehat{s}_i}}\Omega_y^{-1/2}\mathbf{E}^T\mathbf{R}_i \tag{5.9}$$

The second summand is distributed as $\mathcal{N}_q(\mathbf{0},\mathbf{I})$. For the first summand,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\left\|(\widehat{\Omega}_y^{1/2} - \Omega_y^{1/2})\mathbf{E}^T\mathbf{R}_i\right\|_\infty &\leq \frac{1}{\sqrt{n}}\left\|\widehat{\Omega}_y^{1/2} - \Omega_y^{1/2}\right\|_\infty \|\mathbf{E}^T\mathbf{R}_i\|_\infty \\
&\leq \sqrt{n}v_\Omega \frac{1}{n}\left[\|\mathbf{E}^T(\mathbf{X}_i - \mathbf{X}_{-i}\boldsymbol{\zeta}_i)\|_\infty + \|\mathbf{E}^T\mathbf{X}_{-i}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_{0,i})\|_\infty\right] \\
&\leq \sqrt{n}v_\Omega\left[\frac{1}{n}\|\mathbf{E}^T(\mathbf{X}_i - \mathbf{X}_{-i}\boldsymbol{\zeta}_i)\|_\infty + \frac{v_\zeta^2}{n}\|\mathbf{E}^T\mathbf{X}_{-i}\|_\infty\right] \qquad (5.10)
\end{aligned}$$

Applying Lemma 4.5 twice we have for $n \gtrsim \log(pq)$,

$$\frac{1}{n}\|\mathbf{E}^T(\mathbf{X}_i - \mathbf{X}_{-i}\boldsymbol{\zeta}_i)\|_\infty \leq c_7[\sigma_{x,i,-i}\Lambda_{\max}(\Sigma_e)]^{1/2}\sqrt{\frac{\log q}{n}} \tag{5.11}$$

$$\frac{1}{n}\|\mathbf{E}^T\mathbf{X}_{-i}\|_\infty \leq c_9[\Lambda_{\max}(\Sigma_{x,-i})\Lambda_{\max}(\Sigma_e)]^{1/2}\sqrt{\frac{\log((p-1)q)}{n}} \tag{5.12}$$

with probability $\geq 1 - 6c_6\exp[-(c_7^2-1)\log q] - 6c_8\exp[-(c_9^2-1)\log((p-1)q)]$ for some $c_6, c_8 > 0, c_7, c_9 > 1$.

On the other hand

$$\begin{aligned}
s_i := \frac{1}{n}\left\|\mathbf{X}_i - \mathbf{X}_{-i}\widehat{\boldsymbol{\zeta}}_i\right\|^2 &\leq \widehat{s}_i + \frac{1}{n}\left\|\mathbf{X}_{-i}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_{0,i})\right\|^2 \\
&\leq s_i + v_\zeta^2\left[\sqrt{\frac{\log 4 + \tau_{1i}\log(p-1)}{c_{xi}n}} + \max_{i' \neq i}\sigma_{x,i'i'}\right] \qquad (5.13)
\end{aligned}$$

with probability $\geq 1 - 1/p^{\tau_{1i}-2}, \tau_{1i} > 2$, and

$$n \geq 512(1 + 4\Lambda_{\max}(\Sigma_{x,-i}^k))^4 \max_{i' \neq i}(\sigma_{x,i'i'}^k)^4 \log(4(p-1)^{\tau_{1i}}) \tag{5.14}$$

where

$$c_x = \left[128(1 + 4\Lambda_{\max}(\Sigma_{x,-i}))^2 \max_{i' \neq i}(\sigma_{x,i'i'}^k)^2\right]^{-1}$$

by applying Lemma 8 of Ravikumar et al. (2011) on $\mathbf{X}_{-i}^T\mathbf{X}_{-i}/n$.

Denote the second term in the right hand side of (5.13) by $V_i$. Then, for $n$ satisfying (5.14) and $s_i > V_i$, we get the bound:

$$\frac{1}{\widehat{s}_i} \leq \frac{1}{s_i - V_i} \tag{5.15}$$

Combining (5.10), (5.11), (5.12) and (5.15) gives the upper bound for the first summand of (5.9) that holds with probability larger than or equal to

$$1 - 6c_6 e^{-(c_7^2-1)\log q} - 6c_8 e^{-(c_9^2-1)\log((p-1)q)} - \frac{1}{p^{\tau_{1i}-2}}$$

for some $c_6, c_8 > 0, c_7, c_9 > 1, \tau_{1i} > 2$.

To prove (5.8) we have

$$\frac{1}{n}\|\mathbf{R}_i^T\mathbf{X}_{-i}\|_\infty \leq \frac{1}{n}\|(\mathbf{X}_i - \mathbf{X}_{-i}\boldsymbol{\zeta}_{0,i})^T\mathbf{X}_{-i}\|_\infty + \frac{1}{n}\|\mathbf{X}_{-i}^T\mathbf{X}_{-i}(\widehat{\boldsymbol{\zeta}}_i - \boldsymbol{\zeta}_{0,i})\|_\infty$$

□

# References


Lin, J., Basu, S., Banerjee, M., and Michailidis, G. (2016). Penalized Maximum Likelihood Estimation of Multi-layered Gaussian Graphical Models. *J. Mach. Learn. Res.*, 17:5097–5147.

Ma, J. and Michailidis, G. (2016). Joint Structural Estimation of Multiple Graphical Models. *J. Mach. Learn. Res.*, 17:5777–5824.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing $\ell_1$-penalized log-determinant divergence. *Electron. J. Stat.*, 5:935–980.