ROBUST RECOMMENDATION VIA SOCIAL NETWORK ENHANCED MATRIX COMPLETION

Jingxuan Wang[1], Haipeng Shen[2] and Fei Jiang[1]

[1] *University of California, San Francisco and* [2] *University of Hong Kong*

Supplementary Material

S1 Proof of Theorem 3.1

In this section, we provide the proof of Theorem 3.1 in the paper which will follow from a succession of supporting concentration results. We will bring in some established results and then give some additional lemmas with proof to show Lemma 3.1, through which we will demonstrate Theorem 3.1 later.

Before giving out the auxiliary results, we define some frequent notations first. For a real matrix $\mathbf{H}$, we let $\|\mathbf{H}\|_\infty$ denote the infinity norm, trace($\mathbf{H}$) denote the trace, and $r_{\mathbf{H}}$ denote the rank of the matrix $\mathbf{H}$. We denote $\asymp$ as the symbol for asymptotic equivalence in order. We write $X \in \mathcal{SG}(\sigma)$ if X is sub-Gaussian with parameter σ.

Besides, we define some matrices commonly used in our proof. From (2.1) we define $\mathbf{\Pi} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$ as the connection probability matrix. Write $\mathbf{\Pi} = \sum_{i=1}^{n_1} \lambda_i(\mathbf{\Pi})\mathbf{u}_i(\mathbf{\Pi})\mathbf{u}_i(\mathbf{\Pi})^{\mathrm{T}}$ where $\lambda_i(\mathbf{\Pi})$'s are the eigenvalues ordered by its absolute magnitude and $\mathbf{u}_1(\mathbf{\Pi}), \ldots, \mathbf{u}_{n_1}(\mathbf{\Pi})$ are the corresponding eigenvectors, and let $\mathbf{U}_{\mathbf{\Pi}} = [\mathbf{u}_1(\mathbf{\Pi}), \ldots, \mathbf{u}_d(\mathbf{\Pi})]$, and $\mathbf{S}_{\mathbf{\Pi}} = \mathrm{diag}\{|\lambda_1(\mathbf{\Pi})|, \ldots, |\lambda_d(\mathbf{\Pi})|\}$. Let $\mathbf{W}_1\mathbf{\Sigma}\mathbf{W}_2^{\mathrm{T}}$ be the singular value decomposition of $\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}$. We define the maximum expected degree as $\delta(\mathbf{\Pi}) = \max_{i \leq n_1} \sum_{j=1}^{n_1} \Pi_{ij}$.

S1.1 Established bounds

To prove the final conclusion we will need some concentration results from the literature. We state a tight bound on the spectral norm of $\mathbf{M} - \mathbf{\Pi}$ which is a natural variant of Theorem 7 in Lu and Peng (2013).

Proposition S1.1. *Let* $\mathbf{M} \in \mathbb{R}^{n_1 \times n_1}$ *be the adjacency matrix of an independent-edge graph with matrix of edge probabilities* $\mathbf{\Pi}$. *Let* $\delta(\mathbf{\Pi}) = \max_{i \leq n_1} \sum_{j=1}^{n_1} \Pi_{ij}$ *and suppose* $\delta(\mathbf{\Pi}) > log^{4+a}(n_1)$ *for some positive constant* a. *Then*

$$\|\mathbf{M} - \mathbf{\Pi}\|_2 = O_p\left[\max\left\{\delta^{\frac{1}{2}}(\mathbf{\Pi}), \delta^{\frac{1}{4}}(\mathbf{\Pi}) log(n_1)\right\}\right].$$

S1.2 Guarantee of orthogonality

Lemma S1.1. *Let* $\widehat{\mathbf{B}}$ *be the solution of the following optimization*

$$\widehat{\mathbf{B}} = \underset{\mathbf{B}}{argmin}\left\{\frac{1}{n_1 n_2}\|\mathbf{B} - \mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2 + \lambda_2\{\alpha\|\mathbf{B}\|_* + (1-\alpha)\|\mathbf{B}\|_F^2\}\right\},$$

then $\widehat{\mathbf{B}} \in \mathcal{N}(\widehat{\mathbf{X}})$.

Proof. For any $\mathbf{B}$ not orthogonal to $\widehat{\mathbf{X}}$, $\mathbf{B} = \mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B} + \mathbf{P}_{\widehat{\mathbf{X}}}\mathbf{B}$, then

$$\begin{aligned}
& \frac{1}{n_1 n_2}\|\mathbf{B} - \mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2 + \lambda_2\{\alpha\|\mathbf{B}\|_* + (1-\alpha)\|\mathbf{B}\|_F^2\} \\
= & \frac{1}{n_1 n_2}\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B} + \mathbf{P}_{\widehat{\mathbf{X}}}\mathbf{B} - \mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2 \\
& +\lambda_2\{\alpha\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B} + \mathbf{P}_{\widehat{\mathbf{X}}}\mathbf{B}\|_* + (1-\alpha)\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B} + \mathbf{P}_{\widehat{\mathbf{X}}}\mathbf{B}\|_F^2\} \\
= & \frac{1}{n_1 n_2}\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{B} - \mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2 + \|\mathbf{P}_{\widehat{\mathbf{X}}}\mathbf{B}\|_F^2 \\
& +\lambda_2\{\alpha\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B}\|_* + \|\mathbf{P}_{\widehat{\mathbf{X}}}\mathbf{B}\|_* + (1-\alpha)\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B}\|_F^2 + \|\mathbf{P}_{\widehat{\mathbf{X}}}\mathbf{B}\|_F^2\} \\
\geq & \frac{1}{n_1 n_2}\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B} - \mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2 + \lambda_2\{\alpha\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B}\|_* + (1-\alpha)\|\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B}\|_F^2\}.
\end{aligned}$$

The loss function is always smaller when we replace $\mathbf{B}$ by $\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B}$ for any given $\mathbf{B}$, where $\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}\mathbf{B}$ is orthogonal to $\widehat{\mathbf{X}}$. Therefore, $\widehat{\mathbf{B}} \in \mathcal{N}(\widehat{\mathbf{X}})$. □

S1.3 Additional technical lemmas

In this subsection, we will show some technical lemmas to prove the Lemma 3.1 and Theorem 3.1. The following lemmas follow closely with the theorems in Tang et al. (2017), except we apply a tighter bound.

Lemma S1.2. *Assume Condition (C5) holds. Let* $\mathbf{M} \sim RDPG(\mathbf{X})$. *Then*

$$\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F = O_p\left\{dn_1^{-2}\delta(\mathbf{\Pi})\right\},$$

$$\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F = O_p\left[\max\left\{d^{\frac{3}{2}}n_1^{-2}\delta^2(\mathbf{\Pi}), dlog^{\frac{1}{2}}(n_1)\right\}\right],$$

and

$$\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}} - \mathbf{S}_{\mathbf{\Pi}}^{\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F = O_p\left[\max\left\{d^{\frac{3}{2}}n_1^{-\frac{5}{2}}\delta^2(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}log^{\frac{1}{2}}(n_1)\right\}\right].$$

Proof. Since $\mathbf{W}_1\mathbf{\Sigma}\mathbf{W}_2^{\mathrm{T}}$ is the singular value decomposition of $\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}$, then we define the principle angles between column spaces of $\mathbf{U}_{\mathbf{\Pi}}$ and $\mathbf{U}_{\mathbf{M}}$ to be the diagonal matrix

$$\mathbf{\Phi}(\mathbf{U}_{\mathbf{\Pi}}, \mathbf{U}_{\mathbf{M}}) := \mathrm{diag}\left[\cos^{-1}\{\sigma_1(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}})\}, \ldots, \cos^{-1}\{\sigma_d(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}})\}\right],$$

where $\sigma_i(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}})$ is the ith largest singular values of $\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}$. Then by Lemma 1 in Cai and Zhang (2018), we have

$$\|\mathbf{U}_{\mathbf{M}}\mathbf{U}_{\mathbf{M}}^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\|_F = \sqrt{2}\|\sin\mathbf{\Phi}(\mathbf{U}_{\mathbf{\Pi}}, \mathbf{U}_{\mathbf{M}})\|_F. \tag{S1.1}$$

Furthermore, with the conditions being satisfied, Theorem 2 in Yu, Wang and Samworth (2014) implies that

$$\|\sin\mathbf{\Phi}(\mathbf{U}_{\mathbf{\Pi}}, \mathbf{U}_{\mathbf{M}})\|_F \quad \leq \quad \frac{2\sqrt{d}\|\mathbf{M} - \mathbf{\Pi}\|_2}{\lambda_d(\mathbf{\Pi})}. \tag{S1.2}$$

Since $\mathbf{\Pi}$ is symmetric and has rank d, then by definition, $\sigma_d(\mathbf{\Pi}) = \lambda_d(\mathbf{\Pi})$, where $\sigma_d(\mathbf{\Pi})$ and $\lambda_d(\mathbf{\Pi})$ are the dth largest singular value and eigenvalue respectively. Since Condition (C5)(b) implies that $\lambda_d^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X}) = \lambda_d^{-1}(\mathbf{\Pi}) = O_p(n_1^{-1})$, then by (S1.1), (S1.2), and Proposition S1.1, we can get

$$\begin{aligned} & \|\mathbf{U_M}\mathbf{U_M}^{\mathrm{T}} - \mathbf{U_\Pi}\mathbf{U_\Pi}^{\mathrm{T}}\|_F \\ = \;& O_p\left[\max\left\{d^{\frac{1}{2}}n_1^{-1}\delta^{\frac{1}{2}}(\mathbf{\Pi}), d^{\frac{1}{2}}n_1^{-1}\log(n_1)\delta^{\frac{1}{4}}(\mathbf{\Pi})\right\}\right] \\ = \;& O_p\left\{d^{\frac{1}{2}}n_1^{-1}\delta^{\frac{1}{2}}(\mathbf{\Pi})\right\}. \end{aligned} \tag{S1.3}$$

where the second equality is due to Condition (C5)(c).

Since for any matrix $\mathbf{A} \in \mathbb{R}^{n_1\times n_2}$ and $\mathbf{B} \in \mathbb{R}^{n_2\times n_3}$, we let the singular value decomposition of $\mathbf{A}$ be $\mathbf{U\Sigma V}^{\mathrm{T}}$, then we have

$$\begin{aligned} \|\mathbf{AB}\|_F &= \|\mathbf{U\Sigma V}^{\mathrm{T}}\mathbf{B}\|_F \\ &= \sqrt{\mathrm{trace}\{(\mathbf{U\Sigma V}^{\mathrm{T}}\mathbf{B})^{\mathrm{T}}\mathbf{U\Sigma V}^{\mathrm{T}}\mathbf{B}\}} = \sqrt{\mathrm{trace}(\mathbf{B}^{\mathrm{T}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}\mathbf{B})} \\ &= \|\mathbf{\Sigma V}^{\mathrm{T}}\mathbf{B}\|_F \\ &= \sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_3}(\mathbf{\Sigma V}^{\mathrm{T}}\mathbf{B})_{ij}^2} \\ &= \sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_3}\{\sum_{k=1}^{n_2}\mathbf{\Sigma}_{ik}(\mathbf{V}^{\mathrm{T}}\mathbf{B})_{kj}\}^2} \\ &= \sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_3}[\sigma_i(\mathbf{A})1_{\{i\leq min(n_1,n_2)\}}(\mathbf{V}^{\mathrm{T}}\mathbf{B})_{ij}]^2} \end{aligned}$$

$$
\begin{aligned}
&\leq \sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_3}\{\sigma_{\max}(\mathbf{A})(\mathbf{V}^{\mathrm{T}}\mathbf{B})_{ij}\}^2} \\
&= \sigma_{\max}(\mathbf{A})\sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_3}(\mathbf{V}^{\mathrm{T}}\mathbf{B})_{ij}^2} \\
&= \|\mathbf{\Sigma}\|_2\sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_3}(\mathbf{V}^{\mathrm{T}}\mathbf{B})_{ij}^2} \\
&= \|\mathbf{\Sigma}\|_2\|\mathbf{V}^{\mathrm{T}}\mathbf{B}\|_F = \|\mathbf{\Sigma}\|_2\sqrt{\mathrm{trace}(\mathbf{B}^{\mathrm{T}}\mathbf{V}\mathbf{V}^{\mathrm{T}}\mathbf{B})} = \|\mathbf{\Sigma}\|_2\sqrt{\mathrm{trace}(\mathbf{B}^{\mathrm{T}}\mathbf{B})} \\
&= \|\mathbf{\Sigma}\|_2\|\mathbf{B}\|_F \\
&= \|\mathbf{A}\|_2\|\mathbf{B}\|_F. \qquad (S1.4)
\end{aligned}
$$

Then we have

$$
\begin{aligned}
& \|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F \\
= \ & \|\mathbf{W}_1(\mathbf{\Sigma}-\mathbf{I})\mathbf{W}_2^{\mathrm{T}}\|_F \\
\leq \ & \|\mathbf{W}_1\|_2\|\mathbf{\Sigma}-\mathbf{I}\|_F\|\mathbf{W}_2^{\mathrm{T}}\|_2 \\
= \ & \|\mathbf{\Sigma}-\mathbf{I}\|_F \\
= \ & \sqrt{\sum_{i=1}^{d}(1-\sigma_i(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}))^2} \\
\leq \ & \sum_{i=1}^{d}\{1-\sigma_i(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}})\}^2 \\
\leq \ & \sum_{i=1}^{d} 1-\sigma_i^2(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}) \\
= \ & \|\sin\mathbf{\Phi}(\mathbf{U}_{\mathbf{\Pi}},\mathbf{U}_{\mathbf{M}})\|_F^2
\end{aligned}
$$

$$\leq \quad \frac{4d\|\mathbf{M}-\mathbf{\Pi}\|_2^2}{\lambda_d^2(\mathbf{\Pi})},$$

where the second equality is due to $\mathbf{W}_1, \mathbf{W}_2$ are orthogonal matrix, the last inequality is from (S1.2). Therefore, we have

$$\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F = O_p\left\{dn_1^{-2}\delta(\mathbf{\Pi})\right\}. \tag{S1.5}$$

Let $\mathbf{\Gamma} = \mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}$. Since $\mathbf{U}_{\mathbf{\Pi}}$ has orthonormal columns, then we can rewrite $\mathbf{\Gamma}$ as $\{\mathbf{I} - \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\}\mathbf{U}_{\mathbf{M}}$. Then $\mathbf{\Gamma}$ can be regarded as the orthogonal projection of $\mathbf{U}_{\mathbf{M}}$ onto the column space of $\mathbf{U}_{\mathbf{\Pi}}$. And we have for any $\mathbf{W} \in \mathbb{R}^{d\times d}$,

$$\begin{aligned}
& \|\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{W}\|_F^2 \\
= \; & \|\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} + \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{W}\|_F^2 \\
= \; & \|\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F^2 + \|\mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{W}\|_F^2 \\
& +2\langle \mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}, \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{W}\rangle \\
= \; & \|\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F^2 + \|\mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{W}\|_F^2 \\
\geq \; & \|\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F^2 \\
= \; & \|\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F^2 \\
= \; & \|\mathbf{\Gamma}\|_F^2
\end{aligned}$$

where the third equality is due to

$$\langle \mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}, \mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{W}\rangle$$

$$
\begin{aligned}
&= \operatorname{trace}[\{\mathbf{U_M} - \mathbf{U_\Pi}(\mathbf{U_\Pi^T U_\Pi})^{-1}\mathbf{U_\Pi^T U_M}\}^{\mathrm{T}} \\
&\quad \{\mathbf{U_\Pi}(\mathbf{U_\Pi^T U_\Pi})^{-1}\mathbf{U_\Pi^T U_M} - \mathbf{U_\Pi W}\}] \\
&= \operatorname{trace}[\{\mathbf{I} - \mathbf{U_\Pi}(\mathbf{U_\Pi^T U_\Pi})^{-1}\mathbf{U_\Pi^T}\}\mathbf{U_\Pi}\{(\mathbf{U_\Pi^T U_\Pi})^{-1}\mathbf{U_\Pi^T U_M} - \mathbf{W}\}\mathbf{U_M^T}] \\
&= \operatorname{trace}(\mathbf{0}) \\
&= 0.
\end{aligned}
$$

By taking $\mathbf{W} = \mathbf{U_\Pi^T U_M}$ we can get

$$
\begin{aligned}
\|\mathbf{\Gamma}\|_F &\le \|\mathbf{U_M} - \mathbf{U_\Pi U_\Pi^T U_M}\|_F \\
&= \|(\mathbf{U_M U_M^T} - \mathbf{U_\Pi U_\Pi^T})\mathbf{U_M}\|_F \\
&\le \|\mathbf{U_M U_M^T} - \mathbf{U_\Pi U_\Pi^T}\|_F \|\mathbf{U_M}\|_F \\
&= \sqrt{d}\|\mathbf{U_M U_M^T} - \mathbf{U_\Pi U_\Pi^T}\|_F, \qquad \text{(S1.6)}
\end{aligned}
$$

then from (S1.3) we have $\|\mathbf{\Gamma}\|_F = O_p\{d^{3/2}n_1^{-2}\delta(\mathbf{\Pi})\}$.

Since

$$
\begin{aligned}
&\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S_M} - \mathbf{S_\Pi}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} \\
&= (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U_\Pi^T U_M})\mathbf{S_M} + \mathbf{U_\Pi^T U_M S_M} - \mathbf{S_\Pi}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} \\
&= (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U_\Pi^T U_M})\mathbf{S_M} + \mathbf{U_\Pi^T M U_M} - \mathbf{S_\Pi}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} \\
&= (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U_\Pi^T U_M})\mathbf{S_M} + \mathbf{U_\Pi^T}(\mathbf{M} - \mathbf{\Pi})\mathbf{U_M} + \mathbf{U_\Pi^T \Pi U_M} \\
&\quad - \mathbf{S_\Pi}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} \\
&= (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U_\Pi^T U_M})\mathbf{S_M} + \mathbf{U_\Pi^T}(\mathbf{M} - \mathbf{\Pi})\mathbf{\Gamma} + \mathbf{U_\Pi^T}(\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi U_\Pi^T U_M}
\end{aligned}
$$

$$
\begin{aligned}
& +\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{\Pi}\mathbf{U}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} \\
= \quad & (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}})\mathbf{S}_{\mathbf{M}} + \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{\Gamma} + \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} \\
& +\mathbf{S}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} \\
= \quad & (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}})\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}) \\
& +\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{\Gamma} + \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}},
\end{aligned}
$$

then we can get

$$
\begin{aligned}
& \|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F \\
\leq \quad & \|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F(\|\mathbf{S}_{\mathbf{M}}\|_F + \|\mathbf{S}_{\mathbf{\Pi}}\|_F) + \|\mathbf{U}_{\mathbf{\Pi}}\|_F\|(\mathbf{M}-\mathbf{\Pi})\mathbf{\Gamma}\|_F \\
& +\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F\|\mathbf{U}_{\mathbf{\Pi}}\|_F\|\mathbf{U}_{\mathbf{M}}\|_F \\
\leq \quad & \|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F(\|\mathbf{S}_{\mathbf{M}}\|_F + \|\mathbf{S}_{\mathbf{\Pi}}\|_F) + \|\mathbf{U}_{\mathbf{\Pi}}\|_F\|\mathbf{M}-\mathbf{\Pi}\|_2\|\mathbf{\Gamma}\|_F \\
& +\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F\|\mathbf{U}_{\mathbf{\Pi}}\|_F\|\mathbf{U}_{\mathbf{M}}\|_F \\
\leq \quad & \sqrt{d}\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F(\|\mathbf{M}\|_2 + \|\mathbf{\Pi}\|_2) + \sqrt{d}\|\mathbf{M}-\mathbf{\Pi}\|_2\|\mathbf{\Gamma}\|_F \\
& +d\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F \\
\leq \quad & \sqrt{d}\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F(\sqrt{\|\mathbf{M}\|_1\|\mathbf{M}\|_\infty} + \sqrt{\|\mathbf{\Pi}\|_1\|\mathbf{\Pi}\|_\infty}) \\
& +\sqrt{d}\|\mathbf{M}-\mathbf{\Pi}\|_2\|\mathbf{\Gamma}\|_F + d\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F \\
\leq \quad & \sqrt{d}\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F(\|\mathbf{M}\|_\infty + \|\mathbf{\Pi}\|_\infty) + \sqrt{d}\|\mathbf{M}-\mathbf{\Pi}\|_2\|\mathbf{\Gamma}\|_F \\
& +d\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F \\
\leq \quad & C\sqrt{d}\delta(\mathbf{\Pi})\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} - \mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F + \sqrt{d}\|\mathbf{M}-\mathbf{\Pi}\|_2\|\mathbf{\Gamma}\|_F
\end{aligned}
$$

$$+d\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F \qquad \text{(S1.7)}$$

where the first inequality is due to the triangular inequality as well as the sub-multiplicativity for Frobenius norm, the second inequality comes from (S1.4), the fourth inequality is due to the Holder's Inequality, the fifth inequality is due to the symmetry of $\mathbf{M}$ and $\mathbf{\Pi}$, and the last inequality holds with some positive constant C because of the fact $|\delta(\mathbf{M}) - \delta(\mathbf{\Pi})| = O_p\{\delta(\mathbf{\Pi})\}$ which is derived by Chebyshev's Inequality with Condition (C5)(c).

Since $\{\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\}_{ij} = \sum_{k=1}^{n_1}\sum_{l=1}^{n_1} U_{\mathbf{\Pi},ki}(M_{kl}-\Pi_{kl})U_{\mathbf{\Pi},lj}$, then by Hoeffding's Inequality we can get

$$\begin{aligned}\Pr(|\{\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\}_{ij}| \geq t) &\leq \exp\left(-\frac{t^2}{2\sum_{k=1}^{n_1}\sum_{l=1}^{n_1} U_{\mathbf{\Pi},ki}^2 U_{\mathbf{\Pi},lj}^2}\right) \\ &\leq \exp\left(-\frac{t^2}{2}\right),\end{aligned}$$

thus we have $\{\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\}_{ij} = O_p\{\log^{1/2}(n_1)\}$. Therefore we can get

$$\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F = O_p\left\{\log^{\frac{1}{2}}(n_1)\right\}. \qquad \text{(S1.8)}$$

Combining Proposition S1.1, the bounds (S1.5) and (S1.6) and the bound of $\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F$, we can get $\delta(\mathbf{\Pi})\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}-\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F = O_p\{dn_1^{-2}\delta^2(\mathbf{\Pi})\}$, $\|\mathbf{M}-\mathbf{\Pi}\|_2\|\mathbf{\Gamma}\|_F = O_p\{d^{3/2}n_1^{-2}\delta^{3/2}(\mathbf{\Pi})\}$ and $\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F = O_p\{\log^{1/2}(n_1)\}$, thus we can derive from (S1.7) that

$$\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F$$

$$= O_p\left[\max\left\{d^{\frac{3}{2}}n_1^{-2}\delta^2(\mathbf{\Pi}), d\log^{\frac{1}{2}}(n_1)\right\}\right]. \tag{S1.9}$$

Since for $1 \le i, j \le d$,

$$\begin{aligned}
(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}} - \mathbf{S}_{\mathbf{\Pi}}^{\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij} &= (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\left\{\lambda_i^{\frac{1}{2}}(\mathbf{M}) - \lambda_j^{\frac{1}{2}}(\mathbf{\Pi})\right\} \\
&= (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\frac{\lambda_i(\mathbf{M}) - \lambda_j(\mathbf{\Pi})}{\lambda_i^{\frac{1}{2}}(\mathbf{M}) + \lambda_j^{\frac{1}{2}}(\mathbf{\Pi})} \\
&\le (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\frac{\lambda_i(\mathbf{M}) - \lambda_j(\mathbf{\Pi})}{\lambda_j^{\frac{1}{2}}(\mathbf{\Pi})} \\
&\le \lambda_d^{-\frac{1}{2}}(\mathbf{\Pi})(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\left\{\lambda_i(\mathbf{M}) - \lambda_j(\mathbf{\Pi})\right\},
\end{aligned}$$

and we know $(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij} = (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\{\lambda_i(\mathbf{M}) - \lambda_j(\mathbf{\Pi})\}$, then we can get $\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{1/2} - \mathbf{S}_{\mathbf{\Pi}}^{1/2}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F \le \lambda_d^{-1/2}(\mathbf{\Pi})\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F$. Then, by (S1.9) and the fact $\lambda_d^{-1}(\mathbf{\Pi}) = O_p(n_1^{-1})$, we get

$$\begin{aligned}
&\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}} - \mathbf{S}_{\mathbf{\Pi}}^{\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F \\
= \;& O_p\left[\max\left\{d^{\frac{3}{2}}n_1^{-\frac{5}{2}}\delta^2(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}\log^{\frac{1}{2}}(n_1)\right\}\right].
\end{aligned}$$

□

Lemma S1.3. *Assume Condition (C5) holds. Let* $\mathbf{M} \sim RDPG(\mathbf{X})$*. Then there exists an orthogonal matrix* $\mathbf{O}$ *such that*

$$\|\widehat{\mathbf{X}} - \mathbf{X}\mathbf{O}\|_F = \|(\mathbf{M} - \mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\|_F + O_p\left[\max\left\{dn_1^{-\frac{3}{2}}\delta(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}log^{\frac{1}{2}}(n_1)\right\}\right].$$

Proof. Let's define $\mathbf{\Gamma}_1 = \mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}$, $\mathbf{\Gamma}_2 = \mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{1/2} - \mathbf{S}_{\mathbf{\Pi}}^{1/2}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}$, and $\mathbf{\Gamma}_3 = \mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} = \mathbf{U}_{\mathbf{M}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}} + \mathbf{\Gamma}_1$. Since

$\mathbf{U_\Pi U_\Pi^T \Pi} = \mathbf{\Pi}$ and $\mathbf{U_M S_M^{1/2}} = \mathbf{M U_M S_M^{-1/2}}$, then we can get

$$
\begin{aligned}
& \widehat{\mathbf{X}} - \mathbf{U_\Pi S_\Pi^{\frac{1}{2}} W_1 W_2^T} \\
= \; & \mathbf{U_M S_M^{\frac{1}{2}}} - \mathbf{U_\Pi W_1 W_2^T S_M^{\frac{1}{2}}} + \mathbf{U_\Pi}(\mathbf{W_1 W_2^T S_M^{\frac{1}{2}}} - \mathbf{S_\Pi^{\frac{1}{2}} W_1 W_2^T}) \\
= \; & \mathbf{U_M S_M^{\frac{1}{2}}} + (\mathbf{U_\Pi U_\Pi^T U_M S_M^{\frac{1}{2}}} - \mathbf{U_\Pi U_\Pi^T U_M S_M^{\frac{1}{2}}}) - \mathbf{U_\Pi W_1 W_2^T S_M^{\frac{1}{2}}} \\
& + \mathbf{U_\Pi U_\Pi \Gamma_2} \\
= \; & (\mathbf{U_M} - \mathbf{U_\Pi U_\Pi^T U_M})\mathbf{S_M^{\frac{1}{2}}} + \mathbf{\Gamma_1 S_M^{\frac{1}{2}}} + \mathbf{U_\Pi \Gamma_2} \\
= \; & \mathbf{M U_M S_M^{-\frac{1}{2}}} - \mathbf{U_\Pi U_\Pi^T U_M S_M^{\frac{1}{2}}} + \mathbf{\Gamma_1 S_M^{\frac{1}{2}}} + \mathbf{U_\Pi \Gamma_2} \\
= \; & (\mathbf{M} - \mathbf{\Pi})\mathbf{U_M S_M^{-\frac{1}{2}}} + \mathbf{\Pi U_M S_M^{-\frac{1}{2}}} - \mathbf{U_\Pi U_\Pi^T U_M S_M^{\frac{1}{2}}} + \mathbf{\Gamma_1 S_M^{\frac{1}{2}}} + \mathbf{U_\Pi \Gamma_2} \\
= \; & (\mathbf{M} - \mathbf{\Pi})\mathbf{U_M S_M^{-\frac{1}{2}}} + \mathbf{U_\Pi U_\Pi^T}(\mathbf{\Pi U_M S_M^{-\frac{1}{2}}} - \mathbf{U_M S_M^{\frac{1}{2}}}) + \mathbf{\Gamma_1 S_M^{\frac{1}{2}}} + \mathbf{U_\Pi \Gamma_2} \\
= \; & (\mathbf{M} - \mathbf{\Pi})\mathbf{U_M S_M^{-\frac{1}{2}}} - \mathbf{U_\Pi U_\Pi^T}(\mathbf{M} - \mathbf{\Pi})\mathbf{U_M S_M^{-\frac{1}{2}}} + \mathbf{\Gamma_1 S_M^{\frac{1}{2}}} + \mathbf{U_\Pi \Gamma_2} \\
= \; & (\mathbf{I} - \mathbf{U_\Pi U_\Pi^T})(\mathbf{M} - \mathbf{\Pi})\mathbf{U_M S_M^{-\frac{1}{2}}} + \mathbf{\Gamma_1 S_M^{\frac{1}{2}}} + \mathbf{U_\Pi \Gamma_2} \\
= \; & (\mathbf{I} - \mathbf{U_\Pi U_\Pi^T})(\mathbf{M} - \mathbf{\Pi})\mathbf{U_M S_M^{-\frac{1}{2}}} - (\mathbf{I} - \mathbf{U_\Pi U_\Pi^T})(\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi W_1 W_2^T S_M^{-\frac{1}{2}}} \\
& - \mathbf{U_\Pi U_\Pi^T}(\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi W_1 W_2^T S_M^{-\frac{1}{2}}} + (\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi W_1 W_2^T S_M^{-\frac{1}{2}}} + \mathbf{\Gamma_1 S_M^{\frac{1}{2}}} \\
& + \mathbf{U_\Pi \Gamma_2} \\
= \; & (\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi W_1 W_2^T S_M^{-\frac{1}{2}}} - \mathbf{U_\Pi U_\Pi^T}(\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi W_1 W_2^T S_M^{-\frac{1}{2}}} \\
& + (\mathbf{I} - \mathbf{U_\Pi U_\Pi^T})(\mathbf{M} - \mathbf{\Pi})\mathbf{\Gamma_3 S_M^{-\frac{1}{2}}} + \mathbf{\Gamma_1 S_M^{\frac{1}{2}}} + \mathbf{U_\Pi \Gamma_2} \\
= \; & (\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi S_\Pi^{-\frac{1}{2}} W_1 W_2^T} + (\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi}(\mathbf{W_1 W_2^T S_M^{-\frac{1}{2}}} - \mathbf{S_\Pi^{-\frac{1}{2}} W_1 W_2^T}) \\
& - \mathbf{U_\Pi U_\Pi^T}(\mathbf{M} - \mathbf{\Pi})\mathbf{U_\Pi W_1 W_2^T S_M^{-\frac{1}{2}}} + (\mathbf{I} - \mathbf{U_\Pi U_\Pi^T})(\mathbf{M} - \mathbf{\Pi})\mathbf{\Gamma_3 S_M^{-\frac{1}{2}}}
\end{aligned}
$$

$$+\mathbf{\Gamma}_1\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}} + \mathbf{U}_{\mathbf{\Pi}}\mathbf{\Gamma}_2. \tag{S1.10}$$

Let $\lambda_i(\cdot)$ denote the ith largest eigenvalue. Then Weyl's Inequality implies

$$\lambda_d(\mathbf{\Pi}) + \lambda_{n_1}(\mathbf{M} - \mathbf{\Pi}) \leq \lambda_d(\mathbf{M}) \leq \lambda_d(\mathbf{\Pi}) + \lambda_1(\mathbf{M} - \mathbf{\Pi}).$$

Since $\mathbf{\Pi}$, $\mathbf{M}$ and $\mathbf{M} - \mathbf{\Pi}$ are symmetric, then $\|\mathbf{M} - \mathbf{\Pi}\|_2 \geq |\lambda_1(\mathbf{M} - \mathbf{\Pi})|$ and $\|\mathbf{M} - \mathbf{\Pi}\|_2 \geq |\lambda_n(\mathbf{M} - \mathbf{\Pi})|$, thus

$$\lambda_d(\mathbf{\Pi}) - \|\mathbf{M} - \mathbf{\Pi}\|_2 \leq \lambda_d(\mathbf{M}) \leq \lambda_d(\mathbf{\Pi}) + \|\mathbf{M} - \mathbf{\Pi}\|_2.$$

By Condition (C5) and Proposition S1.1, we can get

$$\lambda_d(\mathbf{M}) \asymp \lambda_d(\mathbf{\Pi}). \tag{S1.11}$$

Since $\|\mathbf{S}_{\mathbf{M}}^{-1/2}\|_F = \sqrt{\sum_{i=1}^d |\lambda_i^{-1}(\mathbf{M})|} \leq \sqrt{d|\lambda_d^{-1}(\mathbf{M})|}$, then by (S1.11) we can get $\|\mathbf{S}_{\mathbf{M}}^{-1/2}\|_F = O_p\{d^{1/2}n_1^{-1/2}\}$. And by Weyl's Inequality we have

$$\begin{aligned}
\|\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}}\|_F &= \sqrt{\sum_{i=1}^d |\lambda_i(\mathbf{M})|} \\
&\leq \sqrt{\sum_{i=1}^d \lambda_i(\mathbf{\Pi}) + \lambda_1(\mathbf{M} - \mathbf{\Pi})} \\
&\leq \sqrt{d\{\lambda_1(\mathbf{\Pi}) + \|\mathbf{M} - \mathbf{\Pi}\|_2\}} \\
&= \sqrt{d(\|\mathbf{\Pi}\|_2 + \|\mathbf{M} - \mathbf{\Pi}\|_2)} \\
&\leq \sqrt{d(\sqrt{\|\mathbf{\Pi}\|_1\|\mathbf{\Pi}\|_\infty} + \|\mathbf{M} - \mathbf{\Pi}\|_2)} \\
&= \sqrt{d(\|\mathbf{\Pi}\|_\infty + \|\mathbf{M} - \mathbf{\Pi}\|_2)}
\end{aligned}$$

$$= \sqrt{d\{\delta(\mathbf{\Pi}) + \|\mathbf{M} - \mathbf{\Pi}\|_2\}}$$

where the third inequality is due to the Holder's Inequality, the third equality is due to the symmetry of $\mathbf{\Pi}$, and the last equality comes from the definition of $\delta(\mathbf{\Pi})$ and the fact that elements $\mathbf{\Pi}$ are nonnegative. Therefore, from Proposition S1.1 we can get $\|\mathbf{S}_{\mathbf{M}}^{1/2}\|_F = O_p\left\{d^{1/2}\delta^{1/2}(\mathbf{\Pi})\right\}$.

Since for $1 \leq i, j \leq d$,

$$\begin{aligned}
(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}} - \mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij} &= (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\left\{\lambda_i^{-\frac{1}{2}}(\mathbf{M}) - \lambda_j^{-\frac{1}{2}}(\mathbf{\Pi})\right\} \\
&= (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\frac{\lambda_j^{\frac{1}{2}}(\mathbf{\Pi}) - \lambda_i^{\frac{1}{2}}(\mathbf{M})}{\lambda_i^{\frac{1}{2}}(\mathbf{M})\lambda_j^{\frac{1}{2}}(\mathbf{\Pi})} \\
&\leq (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\frac{\lambda_j^{\frac{1}{2}}(\mathbf{\Pi}) - \lambda_i^{\frac{1}{2}}(\mathbf{M})}{\lambda_d^{\frac{1}{2}}(\mathbf{M})\lambda_d^{\frac{1}{2}}(\mathbf{\Pi})},
\end{aligned}$$

and $(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{1/2} - \mathbf{S}_{\mathbf{\Pi}}^{1/2}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij} = (\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})_{ij}\{\lambda_i^{1/2}(\mathbf{M}) - \lambda_j^{1/2}(\mathbf{\Pi})\}$, then by (S1.11) and Lemma S1.2, we can get $\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{-1/2} - \mathbf{S}_{\mathbf{\Pi}}^{-1/2}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F = O_p\left[\max\left\{d^{3/2}n_1^{-7/2}\delta^2(\mathbf{\Pi}), dn_1^{-3/2}\log^{1/2}(n_1)\right\}\right]$. Since we have

$$\begin{aligned}
&\|(\mathbf{M} - \mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}} - \mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})\|_F \\
\leq\ & \|\mathbf{M} - \mathbf{\Pi}\|_2\|\mathbf{U}_{\mathbf{\Pi}}\|_2\|(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}} - \mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})\|_F,
\end{aligned}$$

then we can get from Proposition S1.1 that

$$\begin{aligned}
&\|(\mathbf{M} - \mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}} - \mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})\|_F \\
=\ & O_p\left[\max\left\{d^{\frac{3}{2}}n_1^{-\frac{7}{2}}\delta^{\frac{5}{2}}(\mathbf{\Pi}), dn_1^{-\frac{3}{2}}\log^{\frac{1}{2}}(n_1)\delta^{\frac{1}{2}}(\mathbf{\Pi})\right\}\right]. \qquad \text{(S1.12)}
\end{aligned}$$

Since the bound (S1.8) implies $\|\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F = O_p\{\log^{1/2}(n_1)\}$, and we have $\|\mathbf{S}_{\mathbf{M}}^{-1/2}\|_F = O_p\{d^{1/2}n_1^{-1/2}\}$, $\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_2 = 1$, and $\|\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{-1/2}\|_F \leq \|\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\|_F\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_2\|\mathbf{S}_{\mathbf{M}}^{-1/2}\|_F$, then we have

$$\|\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}}\|_F = O_p\left\{d^{\frac{1}{2}}\log^{\frac{1}{2}}(n_1)n_1^{-\frac{1}{2}}\right\}. \quad \text{(S1.13)}$$

Additionally Lemma S1.2 suggests that $\|\mathbf{\Gamma}_1\|_F = O_p\{dn_1^{-2}\delta(\mathbf{\Pi})\}$ and $\|\mathbf{\Gamma}_2\|_F = O_p\left[\max\left\{d^{3/2}n_1^{-5/2}\delta^2(\mathbf{\Pi}), dn_1^{-1/2}\log^{1/2}(n_1)\right\}\right]$. Since we have $\|\mathbf{U}_{\mathbf{M}}-\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F = O_p\{d^{1/2}n_1^{-1}\delta^{1/2}(\mathbf{\Pi})\}$ and

$$\begin{aligned}\|\mathbf{\Gamma}_3\|_F &= \|\mathbf{U}_{\mathbf{M}}-\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}+\mathbf{\Gamma}_1\|_F \\ &\leq \|\mathbf{U}_{\mathbf{M}}-\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\|_F + \|\mathbf{\Gamma}_1\|_F,\end{aligned}$$

then $\|\mathbf{\Gamma}_3\|_F = O_p\{d^{1/2}n_1^{-1}\delta^{1/2}(\mathbf{\Pi})\}$. Since $\mathbf{I}-\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}} = \mathbf{I}-\mathbf{U}_{\mathbf{\Pi}}(\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}})^{-1}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}$ is a projection matrix, then we have $\|\mathbf{I}-\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\|_2 = 1$, thus from

$$\begin{aligned}\|(\mathbf{I}-\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}})(\mathbf{M}-\mathbf{\Pi})\mathbf{\Gamma}_3\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}}\|_F &\leq \|\mathbf{I}-\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}}\|_2\|(\mathbf{M}-\mathbf{\Pi})\mathbf{\Gamma}_3\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}}\|_F \\ &= \|(\mathbf{M}-\mathbf{\Pi})\mathbf{\Gamma}_3\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}}\|_F \\ &\leq \|(\mathbf{M}-\mathbf{\Pi})\|_2\|\mathbf{\Gamma}_3\|_F\|\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}}\|_F,\end{aligned}$$

Proposition S1.1, bounds for $\|\mathbf{\Gamma}_3\|_F$ and $\|\mathbf{S}_{\mathbf{M}}^{-1/2}\|_F$, we can get that

$$\|(\mathbf{I}-\mathbf{U}_{\mathbf{\Pi}}\mathbf{U}_{\mathbf{\Pi}}^{\mathrm{T}})(\mathbf{M}-\mathbf{\Pi})\mathbf{\Gamma}_3\mathbf{S}_{\mathbf{M}}^{-\frac{1}{2}}\|_F = O_p\left\{dn_1^{-\frac{3}{2}}\delta(\mathbf{\Pi})\right\}. \quad \text{(S1.14)}$$

And the bounds on $\|\mathbf{\Gamma}_1\|_F$, $\|\mathbf{\Gamma}_2\|_F$ and $\|\mathbf{S}_{\mathbf{M}}^{1/2}\|_F$ yield

$$\|\mathbf{\Gamma}_1\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}}\|_F = O_p\left\{d^{\frac{3}{2}}n_1^{-2}\delta^{\frac{3}{2}}(\mathbf{\Pi})\right\}, \tag{S1.15}$$

and

$$\|\mathbf{U}_{\mathbf{\Pi}}\mathbf{\Gamma}_2\|_F = O_p\left[\max\left\{d^{\frac{3}{2}}n_1^{-\frac{5}{2}}\delta^2(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}\log^{\frac{1}{2}}(n_1)\right\}\right]. \tag{S1.16}$$

Thus by (S1.10) and (S1.12)-(S1.16), we can get that

$$\begin{aligned}
&\|\widehat{\mathbf{X}} - \mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F \\
=\ & \|(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F + O_p\left[\max\left\{dn_1^{-\frac{3}{2}}\delta(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}\log^{\frac{1}{2}}(n_1)\right\}\right] \\
=\ & \left(\mathrm{trace}\left[\mathbf{W}_2\mathbf{W}_1^{\mathrm{T}}\left\{(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\right\}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\right]\right)^{\frac{1}{2}} \\
& +O_p\left[\max\left\{dn_1^{-\frac{3}{2}}\delta(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}\log^{\frac{1}{2}}(n_1)\right\}\right] \\
=\ & \left(\mathrm{trace}\left[\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{W}_2\mathbf{W}_1^{\mathrm{T}}\left\{(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\right\}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\right]\right)^{\frac{1}{2}} \\
& +O_p\left[\max\left\{dn_1^{-\frac{3}{2}}\delta(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}\log^{\frac{1}{2}}(n_1)\right\}\right] \\
=\ & \left(\mathrm{trace}\left[\left\{(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\right\}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\right]\right)^{\frac{1}{2}} \\
& +O_p\left[\max\left\{dn_1^{-\frac{3}{2}}\delta(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}\log^{\frac{1}{2}}(n_1)\right\}\right] \\
=\ & \|(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\|_F + O_p\left[\max\left\{dn_1^{-\frac{3}{2}}\delta(\mathbf{\Pi}), dn_1^{-\frac{1}{2}}\log^{\frac{1}{2}}(n_1)\right\}\right].
\end{aligned}$$

Since $\mathbf{X} = \mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{1/2}\mathbf{W}$ for some orthogonal matrix $\mathbf{W}$, then there exists an orthogonal matrix $\mathbf{O} = \mathbf{W}^{\mathrm{T}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}$ such that

$$\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{1/2}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} = \mathbf{X}\mathbf{O}, \tag{S1.17}$$

which completes the proof. □

Remark 1. (S1.17) suggests $\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{1/2}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}} = \mathbf{X}\mathbf{O}$, then we obtain $\mathbf{O} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-1/2}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}$.

Lemma S1.4. *Assume Condition (C5) holds, let* $\mathbf{E} = \widehat{\mathbf{X}}\mathbf{O}^{\mathrm{T}} - \mathbf{X}$ *where* $\mathbf{O}$ *is the orthogonal matrix defined in Lemma 3.1. Then there exists a matrix* $\mathbf{R} \in \mathbb{R}^{n_1 \times d}$ *such that*

$$\mathbf{E} = \widehat{\mathbf{X}}\mathbf{O}^{\mathrm{T}} - \mathbf{X} = (\mathbf{M} - \mathbf{\Pi})\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} + \mathbf{R}$$

where $\|\mathbf{R}\|_F = O_p\left[\max\left\{dn_1^{-3/2}\delta(\mathbf{\Pi}), dn_1^{-1/2}log^{1/2}(n_1)\right\}\right]$.

Proof. By Lemma S1.3, there exists an orthogonal matrix $\mathbf{O}$ such that $\|\widehat{\mathbf{X}} - \mathbf{X}\mathbf{O}\|_F = \|(\mathbf{M}-\mathbf{\Pi})\mathbf{U}_{\mathbf{\Pi}}\mathbf{S}_{\mathbf{\Pi}}^{-\frac{1}{2}}\|_F + O_p\left[\max\left\{dn_1^{-3/2}\delta(\mathbf{\Pi}), dn_1^{-1/2}\log^{1/2}(n_1)\right\}\right]$. Therefore, there exist a matrix $\mathbf{R}$ which satisfies that its Frobenius norm is of $O_p\left[\max\left\{dn_1^{-3/2}\delta(\mathbf{\Pi}), dn_1^{-1/2}\log^{1/2}(n_1)\right\}\right]$ such that

$$\widehat{\mathbf{X}}\mathbf{O}^{\mathrm{T}} - \mathbf{X} = (\mathbf{M} - \mathbf{\Pi})\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} + \mathbf{R}.$$

□

Proof of Lemma 3.1.

Proof. Let $\mathbf{E} = \widehat{\mathbf{X}}\mathbf{O}^{\mathrm{T}} - \mathbf{X}$ and $\mathbf{G} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$. Since by Lemma S1.4 , for all i, j, we have $E_{ij} = R_{ij} + \sum_{k=1}^{n_1}(M - \Pi)_{ik}G_{kj}$, then we have $(E_{ij} - R_{ij})^2 =$

$\{\sum_{k=1}^{n_1}(M-\Pi)_{ik}G_{kj}\}^2$. By (2.1) we have

$$\begin{aligned}
E\left\{(E_{ij}-R_{ij})^2\right\} &= E\left[\left\{\sum_{k=1}^{n_1}(M_{ik}-\Pi_{ik})G_{kj}\right\}^2\right] \\
&= E\left\{\sum_{k=1}^{n_1}(M_{ik}-\Pi_{ik})^2G_{kj}^2\right\}+2E\left\{\sum_{k<l}(M_{ik}-\Pi_{ik})G_{lj}\right\} \\
&= \sum_{k=1}^{n_1}G_{kj}^2E\left\{(M_{ik}-\Pi_{ik})^2\right\}+2\left\{\sum_{k<l}G_{lj}E(M_{ik}-\Pi_{ik})\right\} \\
&= \sum_{k=1}^{n_1}G_{kj}^2E\left\{(M_{ik}-\Pi_{ik})^2\right\} \\
&= \sum_{k=1}^{n_1}G_{kj}^2\Pi_{ik}(1-\Pi_{ik}),
\end{aligned}$$

then we can get

$$\begin{aligned}
\Pr\left\{(E_{ij}-R_{ij})^2 \ge n_1^{-\frac{1}{2}}\right\} &\le n_1^{\frac{1}{2}}E\left\{(E_{ij}-R_{ij})^2\right\} \\
&= n_1^{\frac{1}{2}}\sum_{k=1}^{n_1}G_{kj}^2\Pi_{ik}(1-\Pi_{ik}) \\
&\le \frac{n_1^{\frac{1}{2}}}{4}\sum_{k=1}^{n_1}G_{kj}^2 \\
&\le \frac{n_1^{\frac{1}{2}}}{4}\|\mathbf{G}\|_F^2 \\
&= \frac{n_1^{\frac{1}{2}}}{4}\|\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_F^2 \\
&\le \frac{n_1^{\frac{1}{2}}}{4}\|\mathbf{X}\|_F^2\|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_F^2 \\
&\le \frac{d^2}{c_0^2 n_1^{\frac{1}{2}}} \to 0,
\end{aligned}$$

where the first inequality is due to the Markov's Inequality, the second inequality is due to the fact that $\Pi_{ik}\in[0,1]$, and the last inequality comes from Condition (C5) and the fact that $\|\mathbf{X}\|_F \leq \sqrt{dn_1}$, therefore, we can get for all i, j, $(E_{ij} - R_{ij})^2 = O_p(n_1^{-1/2})$ thus $\|\mathbf{E} - \mathbf{R}\|_F = O_p(d^{1/2}n_1^{1/4})$. Thus, by Lemma S1.4, we get $\|\mathbf{E}\|_F = O_p\left[\max\left\{dn_1^{-1}\delta(\mathbf{\Pi}), dn_1^{-1/2}\log^{1/2}(n_1), d^{1/2}n_1^{1/4}\right\}\right]$ $= O_p\left[\max\left\{dn_1^{-1}\delta(\mathbf{\Pi}), d^{1/2}n_1^{1/4}\right\}\right]$. Since $\delta(\mathbf{\Pi}) \leq n_1$, then we can get $\|\mathbf{E}\|_F = O_p(n_1^{1/4})$. □

Lemma S1.5. *Assume Condition (C5) holds, then there exists an orthogonal matrix* $\mathbf{O}$ *such that*

$$\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\|_F = O_p\left[\max\left\{d^2 n_1^{-3}\delta^{\frac{3}{2}}(\mathbf{\Pi}), d^2 log^{\frac{1}{2}}(n_1)n_1^{-2}\right\}\right],$$

and for all vectors $\mathbf{v} \in \mathbb{R}^{n_1}$ *s.t.* $\|\mathbf{v}\|_2 = 1$,

$$\|(\mathbf{P}_{\widehat{\mathbf{X}}} - \mathbf{P}_{\mathbf{X}})\mathbf{v}\|_2 = O_p\left[\max\left\{n_1^{-\frac{1}{4}}, n_1^{-2}\delta^{\frac{3}{2}}(\mathbf{\Pi}), log^{\frac{1}{2}}(n_1)n_1^{-1}\right\}\right].$$

Proof. Let $\mathbf{\Delta} = \widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O}$, by Lemma 11 in Loh and Wainwright (2017), we have

$$\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - (\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1}\|_F \leq \frac{\|(\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1}\|_F^2\|\mathbf{\Delta}\|_F}{1 - \|(\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1}\|_F\|\mathbf{\Delta}\|_F}. \quad \text{(S1.18)}$$

Additionally, since $(\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1} = \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}$ and $n_1(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \to \mathbf{S}_x^{-1}$, then we have

$$\begin{aligned}\|(\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1}\|_F &= \|\mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\|_F \\ &= \{1 + o(1)\}\frac{1}{n_1}\|\mathbf{O}^{\mathrm{T}}\mathbf{S}_x^{-1}\mathbf{O}\|_F\end{aligned}$$

$$
\begin{aligned}
&\leq \{1+o(1)\}\frac{1}{n_1}\|\mathbf{O}^{\mathrm{T}}\|_2\|\mathbf{S}_x^{-1}\|_F\|\mathbf{O}\|_2 \\
&\leq d^{\frac{1}{2}}\{1+o(1)\}\frac{1}{n_1}\|\mathbf{O}^{\mathrm{T}}\|_2\|\mathbf{S}_x^{-1}\|_2\|\mathbf{O}\|_2 \\
&= d^{\frac{1}{2}}\{1+o(1)\}\frac{1}{n_1}|\mathbf{S}_x^{-1}\|_2 \\
&= d^{\frac{1}{2}}\{1+o(1)\}\frac{1}{n_1\sigma_{min}(\mathbf{S}_x)} \\
&\leq \frac{2d^{\frac{1}{2}}}{c_0 n_1}, \qquad \text{(S1.19)}
\end{aligned}
$$

where the first inequality is due to the sub-multiplicativity for Frobenius norm, the second inequality is due to the norm equivalence, and the third equality comes from the fact that $\mathbf{O}$ is an orthogonal matrix.

Since by definition $\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}} = (\mathbf{U}_{\mathbf{M}}\mathbf{S}_{\mathbf{M}}^{1/2})^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\mathbf{S}_{\mathbf{M}}^{1/2} = \mathbf{S}_{\mathbf{M}}$ and by (S1.17) we have $\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O} = \mathbf{W}_2\mathbf{W}_1^{\mathrm{T}}\mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}$, then we can get

$$
\begin{aligned}
\|\mathbf{\Delta}\|_F &= \|\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O}\|_F \\
&= \|\mathbf{S}_{\mathbf{M}} - \mathbf{W}_2\mathbf{W}_1^{\mathrm{T}}\mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F \\
&= \|\mathbf{W}_2\mathbf{W}_1^{\mathrm{T}}(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})\|_F \\
&\leq \|\mathbf{W}_2\mathbf{W}_1^{\mathrm{T}}\|_2\|(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})\|_F \\
&= \|(\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}})\|_F
\end{aligned}
$$

where the third and forth equalities are due to $\mathbf{W}_1, \mathbf{W}_2$ are orthogonal and the inequality is from (S1.4). By Lemma S1.2 we have $\|\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\mathbf{S}_{\mathbf{M}} - \mathbf{S}_{\mathbf{\Pi}}\mathbf{W}_1\mathbf{W}_2^{\mathrm{T}}\|_F = O_p\left[\max\left\{dn_1^{-1}\delta^{3/2}(\mathbf{\Pi}), d\log^{1/2}(n_1)\right\}\right]$, then we can get that

$\|\boldsymbol{\Delta}\|_F = O_p\left[\max\left\{dn_1^{-1}\delta^{3/2}(\boldsymbol{\Pi}), d\log^{1/2}(n_1)\right\}\right]$. Thus, from (S1.18) we get

$$\begin{aligned} & \|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - (\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1}\|_F \\ = \; & O_p\left[\max\left\{d^2 n_1^{-3}\delta^{\frac{3}{2}}(\boldsymbol{\Pi}), d^2\log^{\frac{1}{2}}(n_1)n_1^{-2}\right\}\right]. \end{aligned} \quad \text{(S1.20)}$$

Since

$$\begin{aligned} & \|(\mathbf{P}_{\widehat{\mathbf{X}}} - \mathbf{P}_{\mathbf{X}})\mathbf{v}\|_2 \\ \leq \; & \|\mathbf{P}_{\widehat{\mathbf{X}}} - \mathbf{P}_{\mathbf{X}}\|_F\|\mathbf{v}\|_2 \\ = \; & \|\widehat{\mathbf{X}}(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\|_F \\ = \; & \|(\widehat{\mathbf{X}} - \mathbf{X}\mathbf{O})(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}} + \mathbf{X}\mathbf{O}\left\{(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - (\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1}\right\}\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\|_F \\ & + \|\widehat{\mathbf{X}}(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}(\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}})\|_F \\ \leq \; & \|\widehat{\mathbf{X}} - \mathbf{X}\mathbf{O}\|_F\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\|_F\|\mathbf{X}^{\mathrm{T}}\|_F + \|\mathbf{X}\|_F^2\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - (\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1}\|_F \\ & + \|\widehat{\mathbf{X}}(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\|_F\|\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\|_F, \end{aligned}$$

where the last inequality is due to the sub-multiplicativity for Frobenius norm. By definition, we have $\widehat{\mathbf{X}}(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} = \mathbf{U}_{\mathbf{M}}\mathbf{S}_{\mathbf{M}}^{1/2}\{(\mathbf{U}_{\mathbf{M}}\mathbf{S}_{\mathbf{M}}^{1/2})^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\mathbf{S}_{\mathbf{M}}^{1/2}\}^{-1} = \mathbf{U}_{\mathbf{M}}\mathbf{S}_{\mathbf{M}}^{-1/2}$ and $(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} = \{(\mathbf{U}_{\mathbf{M}}\mathbf{S}_{\mathbf{M}}^{1/2})^{\mathrm{T}}\mathbf{U}_{\mathbf{M}}\mathbf{S}_{\mathbf{M}}^{1/2}\}^{-1} = \mathbf{S}_{\mathbf{M}}^{-1}$. And we can get from Lemma S1.3 that $\|\mathbf{S}_{\mathbf{M}}^{-1/2}\|_F = O_p(n_1^{-1/2})$, from Lemma 3.1 that $\|\widehat{\mathbf{X}} - \mathbf{X}\mathbf{O}\|_F = O_p(n_1^{1/4})$, and (S1.20) that $\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - (\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{O})^{-1}\|_F = O_p[\max\{d^2n_1^{-3}\delta^{3/2}(\boldsymbol{\Pi}), d^2\log^{1/2}(n_1)n_1^{-2}\}]$, then we can get $\|(\mathbf{P}_{\widehat{\mathbf{X}}} - \mathbf{P}_{\mathbf{X}})\mathbf{v}\|_2 = O_p[\max\{n_1^{-1/4}, n_1^{-2}\delta^{3/2}(\boldsymbol{\Pi}), \log^{1/2}(n_1)n_1^{-1}\}]$. □

Proof of Theorem 3.1. Since $\lambda_1 = o(n_2^{-1})$, then we have

$$\begin{aligned} & (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}} + n_1 n_2 \lambda_1 \mathbf{I})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}} \\ = \; & (n_1^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}} + n_2 \lambda_1 \mathbf{I})^{-1} n_1^{-1}\widehat{\mathbf{X}}^{\mathrm{T}} \\ = \; & (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}} \{1 + o_p(1)\}, \end{aligned} \tag{S1.21}$$

and since $|\widehat{\theta}_{ij} - \theta_{ij}| = O_p(n_1^{-1})$, then we have

$$\mathbf{W} \circ \widehat{\mathbf{\Omega}} \circ \mathbf{Y} = \mathbf{W} \circ \left\{1 + O_p(n_1^{-\frac{1}{2}})\right\} \mathbf{\Omega}_0 \circ \mathbf{Y}. \tag{S1.22}$$

Thus, by (2.6), we can get there exists an orthogonal matrix $\mathbf{O}$ such that

$$\begin{aligned} & \widehat{\boldsymbol{\beta}} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0 \\ = \; & (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}} + n_1 n_2 \lambda_1 \mathbf{I})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}(\mathbf{W} \circ \mathbf{\Omega}_0 \circ \mathbf{Y}) - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0 \\ = \; & \{1 + o(1)\} (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}(\mathbf{W} \circ \mathbf{\Omega}_0 \circ \mathbf{Y}) - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0 \\ = \; & \{1 + o(1)\} \left\{1 + O_p(n_1^{-\frac{1}{2}})\right\} (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}(\mathbf{W} \circ \mathbf{\Omega}_0 \circ \mathbf{Y}) - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0 \\ = \; & \left\{1 + O_p(n_1^{-\frac{1}{2}})\right\} (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}(\mathbf{W} \circ \mathbf{\Omega}_0 \circ \mathbf{Y}) - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}_0 \\ = \; & \left\{1 + O_p(n_1^{-\frac{1}{2}})\right\} (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\{(\mathbf{W} \circ \mathbf{\Omega}_0 \circ \mathbf{Y}) - \mathbf{A}_0\} \\ & + (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_0 - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{A}_0 + O_p(n_1^{-\frac{1}{2}})(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_0 \\ = \; & \left\{1 + O_p(n_1^{-\frac{1}{2}})\right\} (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\{(\mathbf{W} \circ \mathbf{\Omega}_0 \circ \mathbf{Y}) - \mathbf{A}_0\} \\ & + \{(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_0 \\ & + \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\{\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\mathbf{A}_0 + O_p(n_1^{-\frac{1}{2}})(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_0, \end{aligned}$$

where the second equality is due to (S1.21) and the third equality comes from (S1.22), then for all j we have

$$
\begin{aligned}
& \widehat{\boldsymbol{\beta}}_{\cdot j} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_{0,\cdot j} \\
= \; & \left\{1 + O_p(n_1^{-\frac{1}{2}})\right\} (\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\{(\mathbf{W} \circ \boldsymbol{\Omega}_0 \circ \mathbf{Y})_{\cdot j} - \mathbf{A}_{0,\cdot j}\} \\
& + \{(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j} \\
& + \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\{\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\mathbf{A}_{0,\cdot j} + O_p(n_1^{-\frac{1}{2}})(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}
\end{aligned}
$$

Let $\mathbf{D} = \mathbf{X}^{\mathrm{T}}\{(\mathbf{W} \circ \boldsymbol{\Omega}_0 \circ \mathbf{Y}) - \mathbf{A}_0\}$, then $D_{ij} = \sum_{k=1}^{n_1} X_{ki}\omega_{kj}Y_{kj}/\theta_{kj} - X_{ki}A_{0,kj} = \sum_{k=1}^{n_1} X_{ki}(\omega_{kj} - \theta_{kj})A_{0,kj}/\theta_{kj} - X_{ki}\omega_{kj}\epsilon_{kj}/\theta_{kj}$ for $1 \le i \le d$ and $1 \le j \le n_2$. By condition (C1) and (C3) we know that $X_{ki}(\omega_{kj} - \theta_{kj})A_{0,kj}/\theta_{kj} \in \mathcal{SG}(|A_{0,kj}/\theta_{kj}|)$ and $X_{ki}\omega_{kj}\epsilon_{kj}/\theta_{kj} \in \mathcal{SG}(c_\sigma/\theta_{kj})$, then $X_{ki}(\omega_{kj} - \theta_{kj})A_{0,kj}/\theta_{kj} - X_{ki}\omega_{kj}\epsilon_{kj}/\theta_{kj} \in \mathcal{SG}(|A_{0,kj}/\theta_{kj}| + c_\sigma/\theta_{kj})$ with zero mean. Therefore, Hoeffding's inequality gives us

$$
\begin{aligned}
\Pr(D_{ij} \ge t) & \le \exp\left\{-\frac{t^2}{2\sum_{k=1}^{n_1}(|A_{0,kj}/\theta_{kj}| + c_\sigma/\theta_{kj})^2}\right\} \\
& \le \exp\left\{-\frac{t^2\theta_L^2}{2\sum_{k=1}^{n_1}(|A_{0,kj}| + c_\sigma)^2}\right\} \\
& \le \exp\left\{-\frac{t^2\theta_L^2}{4\sum_{k=1}^{n_1}A_{0,kj}^2 + c_\sigma^2}\right\} \\
& \le \exp\left\{-\frac{t^2\theta_L^2}{4\|\mathbf{A}_0^{\mathrm{T}}\|_{2\to\infty}^2 + 4c_\sigma^2 n_1}\right\} \\
& \le \exp\left\{-\frac{t^2\theta_L^2}{4(a_2^2 + c_\sigma^2)n_1}\right\}.
\end{aligned}
$$

Then there exists a constant c such that $D_{ij} = O_p\{n_1^{1/2}\log^c(n_1)\theta_L^{-1}\}$. Since

$$\begin{aligned}
&(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\{(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\}\\
=\;&(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\{\mathbf{O}^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}+\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\{(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\}\\
=\;&(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\{\mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})+\mathbf{O}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}+\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\\
&\{(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\}\\
=\;&\mathbf{S}_{\mathbf{M}}^{-1}\{\mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})+\mathbf{O}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}+\mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\\
&\{(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\}\\
=\;&\mathbf{S}_{\mathbf{M}}^{-1}\{\mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{M}-\mathbf{\Pi})+\mathbf{O}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\}\{(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\}\\
&+\mathbf{S}_{\mathbf{M}}^{-1}\mathbf{O}^{\mathrm{T}}\mathbf{D}_{\cdot j},
\end{aligned}$$

then we have

$$\begin{aligned}
&\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\{(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\}\|_2\\
\le\;&\|\mathbf{S}_{\mathbf{M}}^{-1}\|_F\|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_F\|\mathbf{X}^{\mathrm{T}}\|_F\|\mathbf{M}-\mathbf{\Pi}\|_2\|(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\|_2\\
&+\|\mathbf{S}_{\mathbf{M}}^{-1}\|_F\|\mathbf{R}^{\mathrm{T}}\|_F\|(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\|_2+\|\mathbf{S}_{\mathbf{M}}^{-1}\|_F\|\mathbf{D}_{\cdot j}\|_2.
\end{aligned}$$

Now, because we know $D_{ij} = O_p\{n_1^{1/2}\log^c(n_1)\theta_L^{-1}\}$ thus $\|\mathbf{D}_{\cdot j}\|_2 = O_p\{n_1^{1/2}\log^c(n_1)\theta_L^{-1}\}$, $\|(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\|_2 = O_p\{n_1^{1/2}\log^c(n_1)\theta_L^{-1}\}$, $\|\mathbf{S}_{\mathbf{M}}^{-1}\|_F = O_p\{n_1^{-1}\}$ in Lemma S1.3, $\|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_F = O(n_1^{-1})$ in (S1.19), $\|\mathbf{M}-\mathbf{\Pi}\|_2 = O_p\{\delta^{1/2}(\mathbf{\Pi})\}$ in Proposition S1.1 and condition (C5), and $\|\mathbf{R}\|_F = O_p\left[\max\left\{dn_1^{-1}\delta(\mathbf{\Pi}), dn_1^{-1/2}\log^{1/2}(n_1)\right\}\right]$, we have $\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\{(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j}-\mathbf{A}_{0,\cdot j}\}\|_2 = O_p\{n_1^{-1/2}\log^c(n_1)\theta_L^{-1}\}$.

Since

$$\begin{aligned}
& \|\{(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}\|_2 \\
= & \|\{(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\}\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}}\mathbf{U}_{\mathbf{M}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}\|_2 \\
\leq & \|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\|_F\|\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}}\|_F\|\mathbf{A}_{0,\cdot j}\|_2 \\
\leq & \|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\|_F\|\mathbf{S}_{\mathbf{M}}^{\frac{1}{2}}\|_F\|\mathbf{A}_0^{\mathrm{T}}\|_{2\to\infty},
\end{aligned}$$

and we know $\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\|_F = O_p\left[\max\left\{n_1^{-3}\delta^{3/2}(\mathbf{\Pi}), \log^{1/2}(n_1)n_1^{-2}\right\}\right]$ in Lemma S1.20, $\|\mathbf{S}_{\mathbf{M}}^{1/2}\|_F = O_p\{\delta^{1/2}(\mathbf{\Pi})\}$ in Lemma S1.3, and $\|\mathbf{A}_0^{\mathrm{T}}\|_{2\to\infty} = O(n_1^{1/2})$ in condition (C2), then $\|\{(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}\|_2 = O_p\left[\max\left\{n_1^{-5/2}\delta^2(\mathbf{\Pi}), \log^{1/2}(n_1)n_1^{-3/2}\delta^{1/2}(\mathbf{\Pi})\right\}\right]$.

Since

$$\begin{aligned}
& \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\{\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\mathbf{A}_{0,\cdot j} \\
= & \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{M} - \mathbf{\Pi}) + \mathbf{R}^{\mathrm{T}}\}\mathbf{A}_{0,\cdot j},
\end{aligned}$$

then we have

$$\begin{aligned}
& \|\mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\{\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\mathbf{A}_{0,\cdot j}\|_2 \\
= & \|\mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{M} - \mathbf{\Pi}) + \mathbf{R}^{\mathrm{T}}\}\mathbf{A}_{0,\cdot j}\|_2 \\
\leq & \|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-2}\|_F\|\mathbf{X}^{\mathrm{T}}\|_F\|\mathbf{M} - \mathbf{\Pi}\|_2\|\mathbf{A}_{0,\cdot j}\|_2 + \|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_F\|\mathbf{R}^{\mathrm{T}}\|_F\|\mathbf{A}_{0,\cdot j}\|_2 \\
\leq & \|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-2}\|_F\|\mathbf{X}^{\mathrm{T}}\|_F\|\mathbf{M} - \mathbf{\Pi}\|_2\|\mathbf{A}_0^{\mathrm{T}}\|_{2\to\infty} + \|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_F\|\mathbf{R}^{\mathrm{T}}\|_F\|\mathbf{A}_0^{\mathrm{T}}\|_{2\to\infty}
\end{aligned}$$

and we know $\|(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\|_F = O(n_1^{-1})$ in (S1.19), $\|\mathbf{M} - \mathbf{\Pi}\|_2 = O_p\{\delta^{1/2}(\mathbf{\Pi})\}$

in Proposition S1.1 and condition (C5), and $\|\mathbf{R}\|_F = O_p\left[\max\left\{dn_1^{-1}\delta(\mathbf{\Pi}), dn_1^{-1/2}\log^{1/2}(n_1)\right\}\right]$, we have $\|\mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\{\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\mathbf{A}_{0,\cdot j}\|_2 = O_p\{n_1^{-1}\delta^{1/2}(\mathbf{\Pi})\}$.

Since $(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j} = \mathbf{S}_{\mathbf{M}}^{-1/2}\mathbf{U}_{\mathbf{M}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}$, then $\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}\|_2 \leq \|\mathbf{S}_{\mathbf{M}}^{-1/2}\|_F\|\mathbf{A}_0^{\mathrm{T}}\|_{2\to\infty}$, thus we have $\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}\|_2 = O_p(1)$. Since

$$
\begin{aligned}
&\|\widehat{\boldsymbol{\beta}}_{\cdot j} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_{0,\cdot j}\|_2 \\
\leq\ & \left\{1 + O_p(n_1^{-\frac{1}{2}})\right\}\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\{(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j} - \mathbf{A}_{0,\cdot j}\}\|_F \\
&+\|\{(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1} - \mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}\|_2 \\
&+\|\mathbf{O}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{O}\{\widehat{\mathbf{X}}^{\mathrm{T}} - \mathbf{O}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\}\mathbf{A}_{0,\cdot j}\|_2 + O_p(n_1^{-\frac{1}{2}})\|(\widehat{\mathbf{X}}^{\mathrm{T}}\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^{\mathrm{T}}\mathbf{A}_{0,\cdot j}\|_2,
\end{aligned}
$$

then we can get that $\|\widehat{\boldsymbol{\beta}}_{\cdot j} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_{0,\cdot j}\|_2 = O_p[\max\{n_1^{-1/2}\log^c(n_1)\theta_L^{-1}, n_1^{-1}\delta^{1/2}(\mathbf{\Pi})\}]$.

□

S2 Proof of Theorem 3.2

In this section, we provide the proof of Lemma 3.2-3.3 and Theorem 3.2 which will follow from a succession of supporting results in Section S1.

Proof of Lemma 3.2. Under Conditions (C1)–(C5), Lemma 3.1 shows that $\|\widehat{\mathbf{X}}\mathbf{O}^{\mathrm{T}} - \mathbf{X}\|_F = O_p(n_1^{1/4})$ and Theorem 3.1 exhibits that $\|\widehat{\boldsymbol{\beta}}_{\cdot j} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_{0,\cdot j}\|_2 = O_p[\max\{n_1^{-1/2}\log^c(n_1)\theta_L^{-1}, n_1^{-1}\delta^{1/2}(\mathbf{\Pi})\}]$ for each j, thus $\|\widehat{\boldsymbol{\beta}} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0\|_F = O_p[\max\{n_1^{-1/2}\log^c(n_1)\theta_L^{-1}n_2^{1/2}, n_1^{-1}n_2^{1/2}\delta^{1/2}(\mathbf{\Pi})\}]$. Furthermore, we can get from Lemma S1.3 that $\|\mathbf{S}_{\mathbf{M}}^{1/2}\|_F = O_p\{\delta^{1/2}(\mathbf{\Pi})\}$ thus $\|\widehat{\mathbf{X}}\|_F = O_p\{\delta^{1/2}(\mathbf{\Pi})\}$ because $\|\widehat{\mathbf{X}}\|_F \leq \|\mathbf{S}_{\mathbf{M}}^{1/2}\|_F$.

Since

$$\begin{aligned}
& \frac{1}{n_1 n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}_0\|_F^2 \\
= \quad & \frac{1}{n_1 n_2}\|\widehat{\mathbf{X}}(\widehat{\boldsymbol{\beta}} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0) + \widehat{\mathbf{X}}\mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0 - \mathbf{X}\boldsymbol{\beta}_0\|_F^2 \\
\leq \quad & \frac{2}{n_1 n_2}\|\widehat{\mathbf{X}}(\widehat{\boldsymbol{\beta}} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0)\|_F^2 + \frac{2}{n_1 n_2}\|\widehat{\mathbf{X}}\mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0 - \mathbf{X}\boldsymbol{\beta}_0\|_F^2 \\
\leq \quad & \frac{2}{n_1 n_2}\|\widehat{\mathbf{X}}\|_2^2\|\widehat{\boldsymbol{\beta}} - \mathbf{O}^{\mathrm{T}}\boldsymbol{\beta}_0\|_F^2 + \frac{2}{n_1 n_2}\|\widehat{\mathbf{X}}\mathbf{O}^{\mathrm{T}} - \mathbf{X}\|_F^2\|\boldsymbol{\beta}_0\|_F^2
\end{aligned}$$

where the first inequality is due to the matrix norm sub-additivity, the last inequality is due to the matrix norm sub-multiplicativity and the inequality (S1.4), then by the bounds above, we get that $\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}_0\|_F^2/(n_1 n_2) = O_p[\max\{n_1^{-2}\log^{2c}(n_1)\theta_L^{-2}\delta(\mathbf{\Pi}), n_1^{-3}\delta^2(\mathbf{\Pi}), n_1^{-1/2}n_2^{-1}\|\boldsymbol{\beta}_0\|_F^2\}]$. □

Proof of Lemma 3.3. Under Conditions (C1)–(C5), we know from the proof of Theorem 3.1 that $\|(\mathbf{W}\circ\mathbf{\Omega}_0\circ\mathbf{Y})_{\cdot j} - \mathbf{A}_{0,\cdot j}\|_2 = O_p\{n_1^{1/2}\log^c(n_1)\theta_L^{-1}\}$. Since $|\widehat{\theta}_{ij} - \theta_{ij}| = O_p(n_1^{-1})$, then we have $\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y} = \mathbf{W}\circ\{1+O_p(n_1^{-1/2})\}\mathbf{\Omega}_0\circ\mathbf{Y}$. Therefore, there exists positive constant c_Y such that

$$\|\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y} - \mathbf{A}_0\|_F \leq c_Y\sqrt{n_1 n_2}\log^c(n_1)\theta_L^{-1} \qquad \text{(S2.1)}$$

holds with probability 1.

By (2.5), we know from the fact $\widehat{\mathbf{B}}$ is the minimizer that

$$\begin{aligned}&\frac{1}{n_1n_2}\|\widehat{\mathbf{B}}-\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2+\lambda_2\alpha\|\widehat{\mathbf{B}}\|_*+\lambda_2(1-\alpha)\|\widehat{\mathbf{B}}\|_F^2 \qquad (S2.2)\\ \le\ &\frac{1}{n_1n_2}\|\mathbf{B}_0-\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2.\end{aligned}$$

Furthermore, from $\widehat{\mathbf{B}}\in\mathcal{N}(\widehat{\mathbf{X}})$ we get that $\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\widehat{\mathbf{B}}-\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y}\|_F^2=\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2+\|\widehat{\mathbf{B}}-\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2$. But for $\mathbf{B}_0$ we have $\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0-\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y}\|_F^2=\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2+\|\mathbf{B}_0-\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2+2\langle\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y}),\mathbf{B}_0\rangle$, then combining (S2.2) we have

$$\begin{aligned}&\frac{1}{n_1n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2+\frac{1}{n_1n_2}\|\widehat{\mathbf{B}}-\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2\\ &+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2\\ \le\ &\frac{1}{n_1n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2+\frac{1}{n_1n_2}\|\mathbf{B}_0-\mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y})\|_F^2\\ &+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2,\end{aligned}$$

then

$$\begin{aligned}&\frac{1}{n_1n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\widehat{\mathbf{B}}-\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y}\|_F^2+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2\\ \le\ &\frac{1}{n_1n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0-\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y}\|_F^2-\frac{2}{n_1n_2}\langle\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y}),\mathbf{B}_0\rangle\\ &+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2,\end{aligned}$$

then

$$\frac{1}{n_1n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0+\widehat{\mathbf{B}}-\mathbf{B}_0-\mathbf{W}\circ\widehat{\mathbf{\Omega}}\circ\mathbf{Y}\|_F^2+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2$$

$$\leq \frac{1}{n_1n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0-\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}\|_F^2-\frac{2}{n_1n_2}\langle\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}),\mathbf{B}_0\rangle$$
$$+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2,$$

then

$$\begin{aligned}&\frac{1}{n_1n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0-\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}\|_F^2+\frac{1}{n_1n_2}\|\widehat{\mathbf{B}}-\mathbf{B}_0\|_F^2\\&+\frac{2}{n_1n_2}\langle\widehat{\mathbf{B}}-\mathbf{B}_0,\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0-\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}\rangle+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2\\\leq\;&\frac{1}{n_1n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0-\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}\|_F^2-\frac{2}{n_1n_2}\langle\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}),\mathbf{B}_0\rangle\\&+\lambda_2\alpha\|\mathbf{B}_0\|_*+\lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2,\end{aligned}$$

then

$$\begin{aligned}\frac{1}{n_1n_2}\|\widehat{\mathbf{B}}-\mathbf{B}_0\|_F^2\;\leq\;&\lambda_2\alpha(\|\mathbf{B}_0\|_*-\|\widehat{\mathbf{B}}\|_*)+\lambda_2(1-\alpha)(\|\mathbf{B}_0\|_F^2-\|\widehat{\mathbf{B}}\|_F^2)\\&-\frac{2}{n_1n_2}\langle\widehat{\mathbf{B}}-\mathbf{B}_0,\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0-\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}\rangle\\&-\frac{2}{n_1n_2}\langle\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}-\mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}),\mathbf{B}_0\rangle.\end{aligned}\tag{S2.3}$$

Then by(S2.3) and with $\alpha\in(0,1]$ and $\lambda_2\geq\{2c_Y\log^c(n_1)\theta_L^{-1}+a_2\}/\sqrt{n_1n_2\alpha^2}+2\|\mathbf{B}_0\|_2/(n_1n_2\alpha)$, we have with probability 1 that

$$\begin{aligned}&\frac{1}{n_1n_2}\|\widehat{\mathbf{B}}-\mathbf{B}_0\|_F^2\\\leq\;&\lambda_2\left\{\alpha(\|\mathbf{B}_0\|_*-\|\widehat{\mathbf{B}}\|_*)+(1-\alpha)(\|\mathbf{B}_0\|_F^2-\|\widehat{\mathbf{B}}\|_F^2)\right\}\\&-\frac{2}{n_1n_2}\langle\widehat{\mathbf{B}}-\mathbf{B}_0,\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}}+\mathbf{B}_0-\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}\rangle\end{aligned}$$

$$
\begin{aligned}
& -\frac{2}{n_1n_2}\langle \widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}} - \mathbf{P}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y}), \mathbf{B}_0\rangle \\
= \ & \lambda_2\left\{\alpha(\|\mathbf{B}_0\|_* - \|\widehat{\mathbf{B}}\|_*) + (1-\alpha)(\|\mathbf{B}_0\|_F^2 - \|\widehat{\mathbf{B}}\|_F^2)\right\} \\
& +\frac{2}{n_1n_2}\left\{\langle \widehat{\mathbf{B}}, \mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y} - \widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}} - \mathbf{B}_0\rangle + \langle \mathbf{B}_0, \mathbf{B}_0 - \mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y})\rangle\right\} \\
= \ & \lambda_2\left\{\alpha(\|\mathbf{B}_0\|_* - \|\widehat{\mathbf{B}}\|_*) + (1-\alpha)(\|\mathbf{B}_0\|_F^2 - \|\widehat{\mathbf{B}}\|_F^2)\right\} \\
& +\frac{2}{n_1n_2}\left\{\langle \widehat{\mathbf{B}}, \mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y} - \mathbf{A}_0 + \mathbf{X}\boldsymbol{\beta}_0\rangle\right. \\
& \left.+\langle \mathbf{B}_0, \mathbf{B}_0 - \mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y} - \mathbf{A}_0) + \mathbf{P}_{\widehat{\mathbf{X}}}\mathbf{A}_0\rangle\right\} \\
\leq \ & \lambda_2\left\{\alpha(\|\mathbf{B}_0\|_* - \|\widehat{\mathbf{B}}\|_*) + (1-\alpha)(\|\mathbf{B}_0\|_F^2 - \|\widehat{\mathbf{B}}\|_F^2)\right\} \\
& +\frac{2}{n_1n_2}\left\{\|\widehat{\mathbf{B}}\|_*(\|\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y} - \mathbf{A}_0\|_2 + \|\mathbf{X}\boldsymbol{\beta}_0\|_2)\right. \\
& \left.+\|\mathbf{B}_0\|_*\{\|\mathbf{B}_0 - \mathbf{P}^{\perp}_{\widehat{\mathbf{X}}}(\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y} - \mathbf{A}_0)\|_2 + \|\mathbf{P}_{\widehat{\mathbf{X}}}\|_2\|\mathbf{A}_0\|_2\}\right\} \\
\leq \ & \lambda_2\left\{\alpha(\|\mathbf{B}_0\|_* - \|\widehat{\mathbf{B}}\|_*) + (1-\alpha)(\|\mathbf{B}_0\|_F^2 - \|\widehat{\mathbf{B}}\|_F^2)\right\} \\
& +\frac{2}{n_1n_2}\left\{\|\widehat{\mathbf{B}}\|_*(\|\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y} - \mathbf{A}_0\|_F + \|\mathbf{A}_0\|_F + \|\mathbf{B}_0\|_2)\right. \\
& \left.+\|\mathbf{B}_0\|_*(\|\mathbf{B}_0\|_2 + \|\mathbf{W}\circ\widehat{\boldsymbol{\Omega}}\circ\mathbf{Y} - \mathbf{A}_0\|_F + \|\mathbf{A}_0\|_F)\right\} \\
\leq \ & \lambda_2\left\{\alpha(\|\mathbf{B}_0\|_* - \|\widehat{\mathbf{B}}\|_*) + (1-\alpha)(\|\mathbf{B}_0\|_F^2 - \|\widehat{\mathbf{B}}\|_F^2)\right\} \\
& +\frac{2}{n_1n_2}\|\widehat{\mathbf{B}}\|_*\left\{c_Y\sqrt{n_1n_2}\log^c(n_1)\theta_L^{-1} + a_2\sqrt{n_1n_2} + \|\mathbf{B}_0\|_2\right\} \\
& +\frac{2}{n_1n_2}\|\mathbf{B}_0\|_*\{\|\mathbf{B}_0\|_2 + c_Y\sqrt{n_1n_2}\log^c(n_1)\theta_L^{-1} + a_2\sqrt{n_1n_2}\} \\
= \ & \left\{\lambda_2\alpha + \frac{2c_Y\log^c(n_1)\theta_L^{-1} + a_2}{\sqrt{n_1n_2}} + \frac{2\|\mathbf{B}_0\|_2}{n_1n_2}\right\}\|\mathbf{B}_0\|_* + \lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2 \\
& +\left\{\frac{2c_Y\log^c(n_1)\theta_L^{-1} + a_2}{\sqrt{n_1n_2}} + \frac{2\|\mathbf{B}_0\|_2}{n_1n_2} - \lambda_2\alpha\right\}\|\widehat{\mathbf{B}}\|_* - \lambda_2(1-\alpha)\|\widehat{\mathbf{B}}\|_F^2 \\
\leq \ & 2\lambda_2\alpha\|\mathbf{B}_0\|_* + \lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2,
\end{aligned}
$$

where the second equality is due to the fact that $\widehat{\mathbf{B}} \in \mathcal{N}(\widehat{\mathbf{X}})$, the second inequality is due to the trace duality property, the third inequality is due to the matrix norm sub-additivity and the fact that $\mathbf{P}_{\widehat{\mathbf{X}}}^{\perp}$ is a projection matrix, the fifth inequality comes from the high-probability bounds (S2.1), and the last inequality is due to $\lambda_2 \geq \{2c_Y \log^c(n_1)\theta_L^{-1} + a_2\}/\sqrt{n_1 n_2 \alpha^2} + 2\|\mathbf{B}_0\|_2/(n_1 n_2 \alpha)$ that is assumed in the statement of Lemma 3.3.

Therefore, we can get $\|\widehat{\mathbf{B}} - \mathbf{B}_0\|_F^2/(n_1 n_2) = O_p[\max\{\lambda_2 \alpha \|\mathbf{B}_0\|_*, \lambda_2(1 - \alpha)\|\mathbf{B}_0\|_F^2\}]$.

□

Proof of Theorem 3.2. Under Conditions (C1)–(C5), with $\alpha \in (0, 1]$, $\lambda_1 = o(n_2^{-1})$ and $\lambda_2 \geq \{2c_Y \log^c(n_1)\theta_L^{-1} + a_2\}/\sqrt{n_1 n_2 \alpha^2} + 2\|\mathbf{B}_0\|_2/(n_1 n_2 \alpha)$, Lemma 3.2 and Lemma 3.3 show that

$$\frac{1}{n_1 n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}_0\|_F^2 = O_p\left\{\max\left(\frac{\log^{2c}(n_1)\delta(\mathbf{\Pi})}{n_1^2\theta_L^2}, \frac{\delta^2(\mathbf{\Pi})}{n_1^3}, \frac{\|\boldsymbol{\beta}_0\|_F^2}{n_1^{1/2} n_2}\right)\right\}$$

and

$$\frac{1}{n_1 n_2}\|\widehat{\mathbf{B}} - \mathbf{B}_0\|_F^2 = O_p\left[\max\left\{\lambda_2 \alpha \|\mathbf{B}_0\|_*, \lambda_2(1 - \alpha)\|\mathbf{B}_0\|_F^2\right\}\right].$$

Since

$$d^2(\widehat{\mathbf{A}}, \mathbf{A}_0) = \frac{1}{n_1 n_2}\|\widehat{\mathbf{A}} - \mathbf{A}_0\|_F^2$$

$$
\begin{aligned}
&= \frac{1}{n_1 n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{B}} - \mathbf{X}\boldsymbol{\beta}_0 - \mathbf{B}_0\|_F^2 \\
&\le \frac{2}{n_1 n_2}\|\widehat{\mathbf{X}}\widehat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}_0\|_F^2 + \|\widehat{\mathbf{B}} - \mathbf{B}_0\|_F^2,
\end{aligned}
$$

then we have

$$
\begin{aligned}
& d^2(\widehat{\mathbf{A}}, \mathbf{A}_0) \\
=\ & O_p\left[\max\left\{\frac{\log^{2c}(n_1)\delta(\boldsymbol{\Pi})}{n_1^2\theta_L^2}, \frac{\delta^2(\boldsymbol{\Pi})}{n_1^3}, \frac{\|\boldsymbol{\beta}_0\|_F^2}{n_1^{1/2}n_2}, \lambda_2\alpha\|\mathbf{B}_0\|_*, \lambda_2(1-\alpha)\|\mathbf{B}_0\|_F^2\right\}\right].
\end{aligned}
$$

□

S2.1 Performance as d varies

In this section, we investigate the performance of MCNet, SoftImpute, TopN, and NetRec when d varies. We fix $n_1 = n_2 = 500$ and vary d from 1 to 10 by stepwise 1. The parameters $\boldsymbol{\beta}_0, \mathbf{B}_0$ and $\mathbf{X}$ are the same as those in Section 4.1. Then the adjacency matrix is generated from a Bernoulli distribution with success rate $\mathbf{X}\mathbf{X}^{\mathrm{T}}$. Let $\mathbf{A}_0 = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{B}_0$ and $\mathbf{Y} = \mathbf{A}_0 + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is an error matrix with independent zero mean normal entries. The standard deviation of the errors is chosen to achieve 0.5 signal-to-noise ratio. We adopt model I for the missing mechanism with an 80% missing rate.

In Figure 1, we plot $d(\widehat{\mathbf{A}}, \mathbf{A}_0)/\|\mathbf{A}_0\|_F$ with 95% confidence intervals versus the dimension d. Figure 1 shows that $d(\widehat{\mathbf{A}}, \mathbf{A}_0)/\|\mathbf{A}_0\|_F$ decreases along with the decrease of d, which implies the proposed method has better performance when the latent position matrix is exactly low-rank.

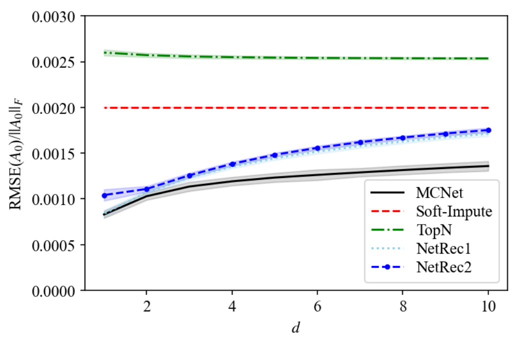

Figure 1: Dimension d. Performance of MCNet and other methods under model I for 100 repetitions.

S3 Robustness to misspecification of d

To evaluate the robustness of our method with respect to the misspecification of d, we added one simulation under misspecification of d. In this simulation, we fix $n_1 = n_2 = 100$ and true $d = 10$. Each entry of $\mathbf{X}$ is generated from a beta distribution with parameters $(3, 1)$. We then scale each entry by a chosen constant to ensure $\max_{i,j} \mathbf{X}_{i\cdot}\mathbf{X}_{j\cdot}^{\mathrm{T}} = 0.95$. We generate $M_{ij} = M_{ji}, i \neq j$ from a Bernoulli distribution with success rate $\mathbf{X}_{i\cdot}\mathbf{X}_{j\cdot}^{\mathrm{T}}$ and generate each entry in $\boldsymbol{\beta}_0$ from a mean zero normal distribution with variance d/n_2. Moreover, we define $\mathbf{B}_0 = \mathbf{P}_{\mathbf{X}}^{\perp}\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} d/\sqrt{n_1 n_2}$ where $\mathbf{U}_0 \in \mathbb{R}^{n_1 \times 10}$ and $\mathbf{V}_0 \in \mathbb{R}^{n_2 \times 10}$ are matrices with standard normal entries. Let $\mathbf{A}_0 = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{B}_0$ and $\mathbf{Y} = \mathbf{A}_0 + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is an error matrix with independent mean zero normal entries where the standard deviation is chosen to achieve 0.5 signal-to-noise ratio. We adopt the uniform missingness (Model I) with $\theta_{ij} = 0.2$. When estimating $\widehat{\mathbf{X}}$, we vary the dimension d from 3 to 50. And we select the tuning parameters using the error perturbation method in Section 4.1. The same procedure is adopted for selecting tuning parameters in the SoftImpute, TopN, and NetRec procedures and we replicate the simulation for 100 times. In Figure 2, we plot $d(\widehat{\mathbf{A}}, \mathbf{A}_0)$ with 95% confidence intervals versus the misspecified dimension d. It shows that our algorithm is insensitive to the selection of d.

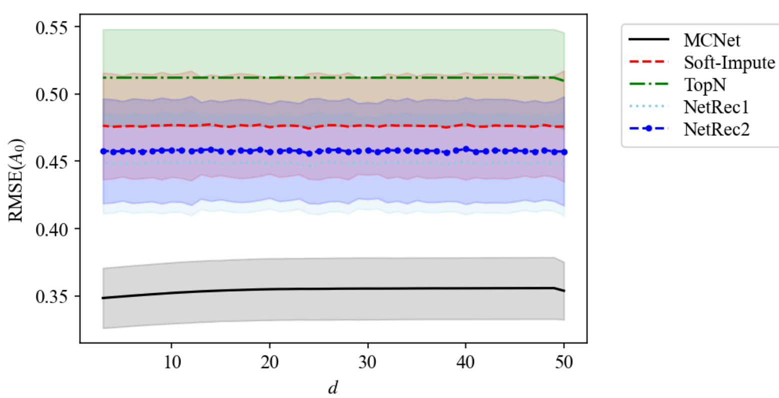

Figure 2: Misspecified dimension d. Performance of MCNet and other methods under model I for 100 repetitions.

References

Cai, T. T. and Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics* **46**, 60-89.

Loh, P.-L. and Wainwright, M. J. (2017). Support recovery without incoherence: a case for nonconvex regularization. *The Annals of Statistics* **45**, 2455-2482.

Lu, L. and Peng, X. (2013). Spectra of edge-independent random graphs. *Electronic Journal of Combinatorics* **20**, 27.

Mao, X., Chen, S. X. and Wong, R. K. (2019). Matrix completion with covariate information. *Journal of the American Statistical Association* **114**, 198-210.

Oliveira, R. I. (2009). Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600.*

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., Park, Y. and Priebe, C. E. (2017). A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics* **26**, 344-354.

Yu, Y., Wang, T. and Samworth, R. J. (2014). A useful variant of the davis–kahan theorem for statisticians. *Biometrika* **102**, 315-323.