Supplement to "Structured Matrix Completion With Applications to Genomic Data Integration" [1]

Tianxi Cai, T. Tony Cai and Anru Zhang

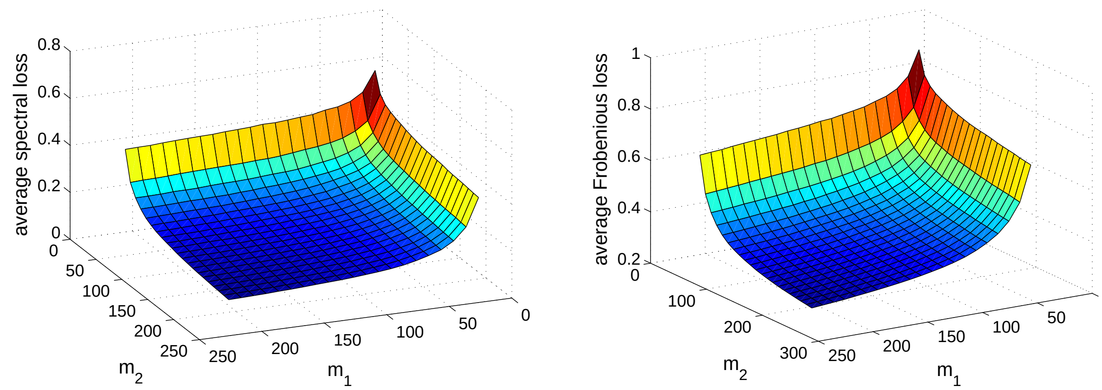
Abstract

In this supplement we provide additional simulation results and the proofs of the main theorems. Some key technical tools used in the proofs of the main results are also developed and proved.

1 Additional Simulation Results

We consider the effect of the number of the observed rows and columns on the estimation accuracy. We let $p_1 = p_2 = 1000$, let the singular values of A be $\{j^{-1}, j = 1, 2, ...\}$ and let m_1 and m_2 vary from 10 to 210. The singular spaces U and V are again generated randomly from the Haar measure. The estimation errors of $\hat{A}_{22}$ from Algorithm 2 with row thresholding and $T_R = 2\sqrt{p_1/m_1}$ over different choices of m_1 and m_2 are shown in Figure 1. As expected, the average loss decreases as m_1 or m_2 grows. Another interesting fact is that the average loss is approximately symmetric with respect to m_1 and m_2. This implies that even with different numbers of observed rows and columns, Algorithm 2 has similar performance with row thresholding or column thresholding.

[1]Tianxi Cai is Professor of Biostatistics, Department of Biostatistics, Harvard School of Public Health, Harvard University, Boston, MA (E-mail: tcai@hsph.harvard.edu); T. Tony Cai is Dorothy Silberberg Professor of Statistics, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA (E-mail: tcai@wharton.upenn.edu); Anru Zhang is a Ph.D. student, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA (E-mail: anrzhang@wharton.upenn.edu). The research of Tianxi Cai was supported in part by NIH Grants R01 GM079330 and U54 LM008748; the research of Tony Cai and Anru Zhang was supported in part by NSF Grant DMS-1208982 and NIH Grant R01 CA127334.



(a) Spectral norm loss

(b) Frobenious norm loss

Figure 8: Losses for the settings with singular values of A being $\{j^{-1}, j = 1, 2, ...\}$, $p_1 = p_2 = 1000$, $m_1, m_2 = 10, ..., 210$.

We are also interested in the performance of Algorithm 2 as p_1 and the ratio m_1/p_1 vary. To this end, we consider the setting where $p_2 = 1000$, $m_2 = 50$, and the singular values of A are chosen as $\{j^{-1}, j = 1, 2, ...\}$. The results are shown in Figure 9. It can be seen that when m_1/p_1 increases, the recovery is generally more accurate; when m_1/p_1 is kept as a constant, the average loss does decrease but not converge to zero as p_1 increases.

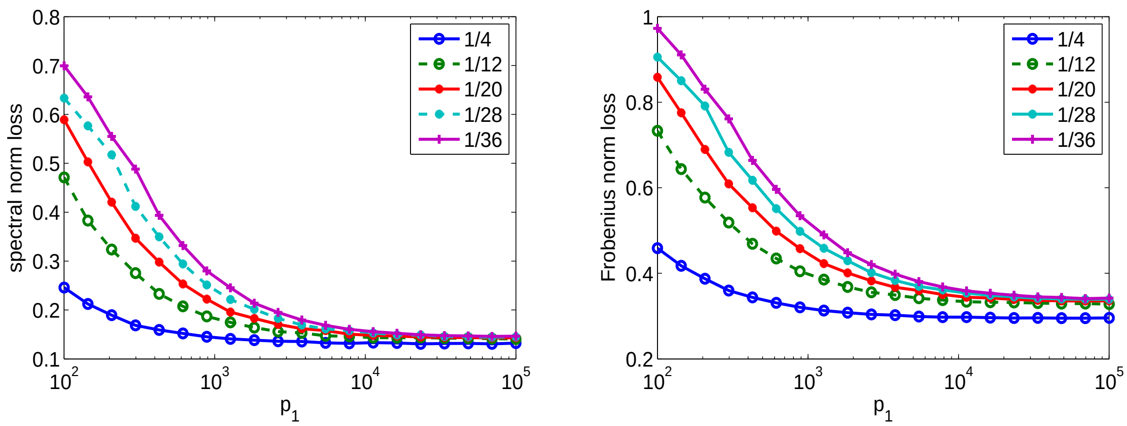
2 Technical Tools

We collect important technical tools in this section. The first lemma is about the inequalities of singular values in the perturbed matrix.

Lemma 1 *Suppose* $X \in \mathbb{R}^{p\times n}$, $Y \in \mathbb{R}^{p\times n}$, $rank(X) = a$, $rank(Y) = b$,

1. $\sigma_{a+b+1-r}(X+Y) \leq \min(\sigma_{a+1-r}(X), \sigma_{b+1-r}(Y))$ *for* $r \geq 1$;

2. *if we further have* $X^\intercal Y = 0$, *we must have* $a + b \leq n$, $\sigma_r(X+Y) \geq \max(\sigma_r(X), \sigma_r(Y))$ *for* $r \geq 1$.



(a) Spectral norm loss (b) Frobenious norm loss

Figure 9: Losses for settings with singular values of A being $\{j^{-1}, j = 1, 2, 3...\}$, $p_2 = 1000$, $m_2 = 50$, $m_1/p_1 = 1/4, 1/12, 1/20, 1/28, 1/36$, and $p_1 = 100, ..., 100,000$.

Lemma 2 *Suppose* $X \in \mathbb{R}^{p\times n}, Y \in \mathbb{R}^{n\times m}$ *are two arbitrary matrices, denote* $\|\cdot\|_q$, $\|\cdot\|$ *as the Schatten-q norm and spectral norm respectively, then we have*

$$\|XY\|_q \leq \|X\|_q \cdot \|Y\|. \tag{22}$$

The following two lemmas provide examples that illustrate NNM fails to recover $\hat{A}_{22}$.

Lemma 3 *Assume* $A = B_1 B_2^T$, *where* $B_1 \in \mathbb{R}^{p_1\times r}$ *and* $B_2 \in \mathbb{R}^{p_2\times r}$ *are two i.i.d. standard Gaussian matrices. Let* A *is divided into blocks as* (1). *Suppose*

$$r \leq \frac{1}{400}\min(p_1, p_2), \quad m_1 \leq \frac{1}{25}p_1, \quad m_2 \leq \frac{1}{25}p_2, \tag{23}$$

then the NNM (3) *fails to recover* A_{22} *with probability at least* $1 - 12\exp(-\min(p_1, p_2)/400)$.

Lemma 4 *Denote* 1_p *as the p-dimensional vector with all entries 1. Suppose* $A = 1_{p_1} \cdot 1_{p_2}^{\intercal}$, *and* A *is divided into blocks as* (1). *Then the NNM* (3) *yields*

$$\hat{A}_{22} = \min\left\{\sqrt{\frac{m_1 m_2}{(p_1 - m_1)(p_2 - m_2)}}, 1\right\} 1_{p_1-m_1} 1_{p_2-m_2}^{\intercal}.$$

The following result is on the norm of a random submatrix of a given orthonormal matrix.

Lemma 5 *Suppose $U \in \mathbb{R}^{p\times d}$ is a fixed matrix with orthonormal columns (hence $d \leq p$). Denote $W = \max_{1\leq i\leq p} \frac{p}{d} \cdot \sum_{j=1}^d u_{ij}^2$. Suppose we uniform randomly draw n rows (with or without replacement) from U and note the index as Ω and denote*

$$U_\Omega = \begin{bmatrix} U_{\Omega(1)} \\ \vdots \\ U_{\Omega(n)} \end{bmatrix}.$$

When $n \geq \frac{4Wd(\log d + c)}{(1-\alpha)^2}$ for some $0 < \alpha < 1$ and $c > 1$, we have

$$\|\sigma_{\min}(U_\Omega)\| \geq \sqrt{\frac{\alpha n}{p}}$$

with probability $1 - 2e^{-c}$.

The following results is about the spectral norm of the submatrix of a random orthonormal matrix.

Lemma 6 *Suppose $U \in \mathbb{R}^{p\times d}$ $(d \leq p)$ is with random orthonormal columns with Haar measure. For all $0 < \alpha_1 < 1 < \alpha_2$, there exists constant $C, \delta > 0$ depending only on α_1, α_2 such that when $p \geq n \geq \min\{Cd, p\}$, we have*

$$\sqrt{\frac{\alpha_1 n}{p}} \leq \sigma_{\min}(U_{[1:n,:]}) \leq \|U_{[1:n,:]}\| \leq \sqrt{\frac{\alpha_2 n}{p}} \tag{24}$$

with probability at least $1 - \exp(-\delta n)$.

Proof of the Technical Lemmas

Proof of Lemma 1.

1. First, by a well-known fact about best low-rank approximation,

$$\sigma_{a+b+1-r}(X+Y) = \min_{M\in\mathbb{R}^{p\times n}, \text{rank}(M)\leq a+b-r} \|X+Y-M\|.$$

   Hence,

$$\sigma_{a+b+1-r}(X+Y) \leq \|X+Y-(X_{\max(a-r)}+Y)\| = \|X_{-\max(a-r)}\| = \sigma_{a+1-r}(X);$$

   similarly $\sigma_{a+b+1-r}(X+Y) \leq \sigma_{b+1-r}(Y)$.

2. When we further have $X^\intercal Y = 0$, we know the column space of X and Y are orthogonal, then we have $\text{rank}(X+Y) = \text{rank}(X) + \text{rank}(Y) = a + b$, which means $a + b \leq n$. Next, note that

$$(X+Y)^\intercal(X+Y) = X^\intercal X + Y^\intercal Y + X^\intercal Y + Y^\intercal X = X^\intercal X + Y^\intercal Y,$$

if we note $\lambda_r(\cdot)$ as the r-th largest eigenvalue of the matrix, then we have

$$\begin{aligned}\sigma_r^2(X+Y) =& \lambda_r((X+Y)^\intercal(X+Y)) = \lambda_r(X^\intercal X + Y^\intercal Y) \\ &\geq \max(\lambda_r(X^\intercal X), \lambda_r(Y^\intercal Y)) = \max(\sigma_r^2(X), \sigma_r^2(Y)).\end{aligned}$$

□

Proof of Lemma 2. Since

$$\|XY\|_q = \sqrt[q]{\sum_i \sigma_i^q(XY)}, \quad \|X\|_q = \sqrt[q]{\sum_i \sigma_i^q(X)},$$

it suffices to show $\sigma_i(XY) \leq \sigma_i(X)\|Y\|$. To this end, we have

$$\sigma_i(X) = \min_{M \in \mathbb{R}^{p \times m}, \text{rank}(M) \leq i-1} \|XY - M\| \leq \|XY - X_{\max(i-1)}Y\| = \|X_{-\max(i-1)}Y\| \leq \sigma_i(X)\|Y\|,$$

which finishes the proof of this lemma. □

Proof of Lemma 3. Since B_1 and B_2 and their submatrices are all i.i.d. standard matrices, by the random matrix theory (Corollary 5.35 in Vershynin (2010)), for $t > 0$, we have with probability at least $1 - 12\exp(-t^2/2)$, the following inequalities hold,

$$\begin{aligned}\lambda_r(A) \geq& \lambda_{\min}(B_1)\lambda_{\min}(B_2) \geq (\sqrt{p_1} - \sqrt{r} - t)(\sqrt{p_2} - \sqrt{r} - t) \\ &\overset{(23)}{\geq} \left(\frac{19}{20}\sqrt{p_1} - t\right)\left(\frac{19}{20}\sqrt{p_2} - t\right)\end{aligned} \tag{25}$$

$$\|A_{1\bullet}\| = \|B_{1,[1:m_1,:]}B_2^T\| \leq (\sqrt{m_1} + \sqrt{r} + t)(\sqrt{p_2} + \sqrt{r} + t) \overset{(23)}{\leq} \left(\frac{1}{4}\sqrt{p_1} + t\right)\left(\frac{21}{20}\sqrt{p_2} + t\right) \tag{26}$$

and

$$\begin{aligned}\|A_{21}\| =& \|B_{1,[(m_1+1):p_1,:]}B_{2,[1:m_2,:]}^T\| \leq (\sqrt{p_1} + \sqrt{r} + t)(\sqrt{m_2} + \sqrt{r} + t) \\ &\overset{(23)}{\leq} \left(\frac{21}{20}\sqrt{p_1} + t\right)\left(\frac{1}{4}\sqrt{p_2} + t\right).\end{aligned} \tag{27}$$

Denote

$$A_0 = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix}$$

and set $t = \frac{1}{20}\min(\sqrt{p_1}, \sqrt{p_2})$. Since $\|A_0\|_* \le \|A_{1\bullet}\|_* + \|A_{21}\|_*$, , we have

$$P\left(\|A\|_* \ge \frac{326}{400}\sqrt{p_1 p_2}\right) \ge 1 - 12\exp(-\min(p_1,p_2)/400) \tag{28}$$

and

$$P\left(\|A_0\|_* \le \frac{264}{400}\sqrt{p_1 p_2}\right) \ge 1 - 12\exp(-\min(p_1,p_2)/400). \tag{29}$$

Hence, with probability at least $1 - 12\exp(-\min(p_1,p_2)/400)$, $\|A_0\|_* < \|A\|_*$, which implies that the NNM (3) fails to recover A_{22}. $\square$

Proof of Lemma 4. For convenience, we denote $x \wedge y = \min(x,y)$ for any two real numbers x, y. First, we can extend the unit vectors $\frac{1}{\sqrt{m_1}}1_{m_1}$, $\frac{1}{\sqrt{m_2}}1_{m_2}$, $\frac{1}{\sqrt{p_1-m_1}}1_{p_1-m_1}$ and $\frac{1}{\sqrt{p_2-m_2}}1_{p_2-m_2}$ into orthogonal matrices, which we denote as $U_{m_1} \in \mathbb{R}^{m_1\times m_1}$, $U_{m_2} \in \mathbb{R}^{m_2\times m_2}$, $U_{p_1-m_1} \in \mathbb{R}^{(p_1-m_1)\times(p_1-m_1)}$, $U_{p_2-m_2} \in \mathbb{R}^{(p_2-m_2)\times(p_2-m_2)}$. Next, for all $A'_{22} \in \mathbb{R}^{(p_1-m_1)\times(p_2-m_2)}$, we must have

$$\begin{aligned}\left\|\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A'_{22} \end{bmatrix}\right\|_* &= \left\|\begin{bmatrix} U_{m_1}^\intercal & 0 \\ 0 & U_{p_1-m_1}^\intercal \end{bmatrix} \cdot \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A'_{22} \end{bmatrix} \cdot \begin{bmatrix} U_{m_2} & 0 \\ 0 & U_{p_2-m_2} \end{bmatrix}\right\|_* \\ &\triangleq \left\|\begin{bmatrix} E_{11} & E_{12} \\ E_{21} & U_{p_1-m_1}^\intercal A'_{22} U_{p_2-m_2} \end{bmatrix}\right\|_*,\end{aligned}$$

where $E_{11} \in \mathbb{R}^{m_1\times m_2}, E_{12} \in \mathbb{R}^{m_1\times(p_2-m_2)}, E_{21} \in \mathbb{R}^{(p_1-m_1)\times m_2}$ are with the first entry $\sqrt{m_1 m_2}$, $\sqrt{m_1(p_2-m_2)}$ and $\sqrt{m_2(p_1-m_1)}$ respectively and other entries 0. Therefore, we can see

$$\left\|\begin{bmatrix} E_{11} & E_{12} \\ E_{21} & U_{p_1-m_1}^\intercal A'_{22} U_{p_2-m_2} \end{bmatrix}\right\|_* \ge \left\|\begin{bmatrix} \sqrt{m_1 m_2} & \sqrt{m_1(p_2-m_2)} \\ \sqrt{m_2(p_1-m_1)} & [U_{p_1-m_1}^\intercal A'_{22} U_{p_2-m_2}]_{[1,1]} \end{bmatrix}\right\|_*$$

and the equality holds if and only if $U_{p_1-m_1}^\intercal A'_{22} U_{p_2-m_2}$ is zero except the first entry.

By some calculation, we can see the nuclear norm of 2-by-2 matrix

$$\left\|\begin{bmatrix} \sqrt{m_1 m_2} & \sqrt{m_1(p_2-m_2)} \\ \sqrt{m_2(p_1-m_1)} & x \end{bmatrix}\right\|_*$$

achieves its minimum if and only if

$$x = \sqrt{m_1 m_2} \wedge \sqrt{(p_1 - m_1)(p_2 - m_2)}.$$

Hence, A'_{22} achieves the minimum of $\left\| \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A'_{22} \end{bmatrix} \right\|_*$ if and only if

$$U_{p_1-m_1}^\intercal A'_{22} U_{p_2-m_2} = \begin{bmatrix} \sqrt{m_1 m_2} \wedge \sqrt{(p_1 - m_1)(p_2 - m_2)} & 0 & \cdots \\ 0 & 0 & \\ \vdots & & \ddots \end{bmatrix},$$

which means the minimizer $A'_{22} = \left(\sqrt{\frac{m_1 m_2}{(p_1-m_1)(p_2-m_2)}} \wedge 1\right) \cdot 1_{p_1-m_1} 1_{p_2-m_2}^\intercal$. $\square$

Proof of Lemma 5. The proof of this lemma relies on operator-Bernstein's inequality for sampling (Theorem 1 in Gross and Nesme (2010)). For two symmetric matrices A, B, we say $A \preceq B$ if $B - A$ is positive definite. By assumption, $\{U_{\Omega(j)\bullet}, j = 1, \cdots, n\}$ are uniformly random samples (with or without replacement) from $\{U_{i\bullet}, i = 1, \cdots, n\}$. Suppose

$$X_i = U_{i\bullet}^\intercal U_{i\bullet} - \frac{1}{p} I_d, \quad i = 1, \cdots, p, \tag{30}$$

then X_i are symmetric matrices, $X_{\Omega(j)}, j = 1, \cdots, n$ are uniformly random samples (with or without replacement) from $\{X_1, \cdots, X_p\}$. In addition, we have

$$EX_j = \frac{1}{p}\sum_{i=1}^{p} U_{i\bullet}^\intercal U_{i\bullet} - \frac{1}{p} I_d = \frac{1}{p} U^\intercal U - \frac{1}{p} I_d = 0$$

$$\|X_j\| \leq \max_{1\leq i\leq p} \left\| U_{i\bullet}^\intercal U_{i\bullet} - \frac{1}{p} I_d \right\| \leq \max_{1\leq i\leq p} \max\left\{ \|U_{i\bullet}^\intercal U_{i\bullet}\|, \frac{1}{p}\|I_d\| \right\} \leq \frac{Wd}{p}$$

$$\begin{aligned} EX_j^2 =& \frac{1}{p}\sum_{i=1}^{p}\left(U_{i\bullet}^\intercal U_{i\bullet} - \frac{1}{p} I_d\right)^2 = \frac{1}{p}\sum_{i=1}^{p}\left(U_{i\bullet}^\intercal U_{i\bullet} U_{i\bullet}^\intercal U_{i\bullet} - \frac{2}{p} U_{i\bullet}^\intercal U_{i\bullet} + \frac{1}{p^2} I_d\right) \\ =& \frac{1}{p}\sum_{i=1}^{p} \|U_{i\bullet}\|_2^2 \cdot U_{i\bullet}^\intercal U_{i\bullet} - \frac{1}{p^2} I_d \\ \preceq& \frac{1}{p} \cdot \frac{Wd}{p} \sum_{i=1}^{p} U_{i\bullet}^\intercal U_{i\bullet} - \frac{1}{p^2} I_d \preceq \frac{Wd - 1}{p^2} I_d \end{aligned}$$

For all $0 < \alpha < 1$, by Theorem 1 in Gross and Nesme (2010),

$$\begin{aligned}
P\left(\|U_\Omega\| \le \sqrt{\frac{\alpha n}{p}}\right) &= P\left(U_\Omega^\intercal U_\Omega \preceq \frac{\alpha n}{p} I_d\right) = P\left(\sum_{j=1}^n U_{\Omega(j)\bullet}^\intercal U_{\Omega(j)\bullet} \preceq \frac{\alpha n}{p} I_d\right) \\
&= P\left(\sum_{j=1}^n X_j \preceq -\frac{(1-\alpha)n}{p} I_d\right) \le P\left(\left\|\sum_{j=1}^n X_j\right\| \ge \frac{(1-\alpha)n}{p}\right) \\
&\le 2d \exp\left(-\min\left(\frac{((1-\alpha)n/p)^2}{4n(Wd-1)/p^2}, \frac{(1-\alpha)n/p}{2Wd/p}\right)\right) \\
&\le 2d \exp\left(-\frac{n(1-\alpha)^2}{4Wd}\right) \le 2\exp(-c).
\end{aligned}$$

The last inequality is due to the assumption that

$$n \ge \frac{4Wd(\log d + c)}{(1-\alpha)^2}.$$

□

Proof of Lemma 6. By the assumption on n, we have $n \ge p$ or $n \ge Cd$. When $n \ge p$, we know $n = p$ and $U_{[1:n,:]} = U$ is an orthogonal matrix, which means (24) is clearly true. Hence, we only need to prove the theorem under the assumption that $p \ge n$ is true. In this case, we must have $n \ge Cd$.

Since U has random orthonormal columns with Haar measure, for any fixed vector $v \in \mathbb{R}^d$, Uv is identitical distributed as

$$\|x\|_2^{-1}(x_1, x_2, \cdots, x_p), \quad \text{where } x_1, \cdots, x_p \overset{iid}{\sim} N(0,1)$$

Hence, $U_{[1:n,:]}v$ is identical distributed with $\|x\|_2^{-1}(x_1, \cdots, x_n)$ and

$$\|U_{[1:n,:]}v\|_2 \text{ is identical distributed as } \sqrt{(\sum_{i=1}^n x_i^2)(\sum_{i=1}^p x_i^2)^{-1}}, \tag{31}$$

which is the also the square root of Beta distribution. Denote

$$\alpha_1' = \frac{1+\alpha_1}{2}, \quad \alpha_2' = \frac{1+\alpha_2}{2}. \tag{32}$$

By Lemma 1 in Laurent and Massart (2000), when $x_1, \cdots, x_p$ are i.i.d. standard normal, we have

$$1 - 2\sqrt{C'} \le \frac{\sum_{i=1}^n x_i^2}{n} \le 1 + 2\sqrt{C'} + 2C'$$

$$1 - 2\sqrt{\frac{C'n}{p}} \le \frac{\sum_{i=1}^p x_i^2}{p} \le 1 + 2\sqrt{\frac{C'n}{p}} + \frac{2C'n}{p}$$

both hold with probability at least $1 - 4\exp(-C'n)$. Here we let $C' > 0$ be small enough and only depending on α_1, α_2 such that

$$\alpha_1' \le \frac{1 - 2\sqrt{C'}}{1 + 2\sqrt{C'} + 2C'}, \quad \frac{1 + 2\sqrt{C'} + 2C'}{1 - 2\sqrt{C'}} \le \alpha_2'.$$

Combining the previous inequalities and (31), we have for any fixed unit vector $v \in \mathbb{R}^d$,

$$\frac{\alpha_1' n}{p} \le \|U_{[1:n,:]}v\|_2^2 \le \frac{\alpha_2' n}{p} \tag{33}$$

with probability at least $1 - 4\exp(-C'n)$, where C' only depends on α_1', α_2'. Next, based on Lemma 2.5 in Vershynin (2013), we can construct an ε-net on the unit sphere of $\mathbb{R}^d$ as B, such that $|B| \le (1 + 2/\varepsilon)^d$, where $\varepsilon > 0$ is to be determined later. Under the event that $\{\forall v \in B, (33) \text{ holds}\}$, we suppose

$$\kappa_1 = \min_{\|v\|_2=1} \|U_{[1:n,:]}v\|_2^2, \quad \kappa_2 = \max_{\|v\|_2=1} \|U_{[1:n,:]}v\|_2^2.$$

For any v in the unit sphere of $\mathbb{R}^d$, there must exists $v' \in B$ such that $\|v - v'\|_2 \le \varepsilon$, which yields,

$$\|U_{[1:n,:]}v\|_2 \le \|U_{[1:n,:]}v'\|_2 + \|U_{[1:n,:]}(v - v')\|_2 \le \sqrt{\alpha_2' n/p} + \kappa_2 \varepsilon$$

$$\|U_{[1:n,:]}v\|_2 \ge \|U_{[1:n,:]}v'\|_2 - \|U_{[1:n,:]}(v - v')\|_2 \ge \sqrt{\alpha_1' n/p} - \varepsilon \kappa_2$$

These implies that $\kappa_2 \le \sqrt{\alpha_2' n/p}/(1-\varepsilon)$, $\kappa_1 \ge \sqrt{\alpha_1' n/p} - \varepsilon\kappa_2 \ge \sqrt{\alpha_1' n/p} - \sqrt{\alpha_2' n/p} \cdot \varepsilon/(1-\varepsilon)$. Hence, we can take ε depending on α_1, α_2 such that $\kappa_2 \le \sqrt{\alpha_2 n/p}$, $\kappa_1 \ge \sqrt{\alpha_1 n/p}$, which implies (24).

Finally we estimate the probability that the event $\{\forall v \in B, (33) \text{ holds}\}$ happens. We choose $C \ge 4d\log(1 + 2/\varepsilon)/C'$ that only depends on α_1 and α_2. If $n \ge Cd$,

$$C'n/2 \ge d\log(1 + 2/\varepsilon) + \log 4.$$

so

$$1-(1+2/\varepsilon)^d\cdot 4\exp(-C'n)=1-\exp(d\log(1+2/\varepsilon)+\log 4-C'n)\geq 1-\exp(-nC'/2)$$

Finally, we finish the proof of the lemma by setting $\delta = C'/2$. $\square$

3 Proofs of the Results in the Main Paper

We prove Proposition 1, Theorems 1 and 2, Lemma 7, Lemma 8, Theorem 3, Corollary 1 and Corollary 2 in this section.

Proof of Proposition 1

Since $A_{1\bullet}$ is of rank r, which is the same as A, all rows of A must be linear combinations of the rows of $A_{1\bullet}$. This implies all rows of $A_{\bullet 1}$ is a linear combination of A_{11}. Since rank($A_{\bullet 1}$)$= r$, we must have rank$(A_{11}) \geq r$. Besides, rank$(A_{11}) \leq$ rank$(A) = r$ since A_{11} is a submatrix of A. So rank$(A_{11}) = r$. Simiarly, rows of $A_{\bullet 1}$ is the linear combination of A_{11}, so we have

$$A_{21}=A_{21}P_{A_{11}}=A_{21}A_{11}^\intercal(A_{11}A_{11}^\intercal)^\dagger A_{11}=A_{21}V\Sigma U^\intercal(U\Sigma^2U^\intercal)^\dagger A_{11}=\left(A_{21}V\Sigma^{-1}U^\intercal\right)A_{11},$$

namely rows of A_{21} is a linear combination of A_{11}. By the argument before, we know A_{22} can be represented as the same linear combination of A_{12} as A_{21} by A_{11}, so we have $A_{22} = (A_{21}V\Sigma^{-1}U^\intercal)\,A_{12} = A_{21}V\Sigma^{-1}U^\intercal A_{12} = A_{21}A_{11}^\dagger A_{12}$, which concludes the proof. $\square$

Proof of Theorem 1

Suppose $M \in \mathbb{R}^{m_1\times r}, N \in \mathbb{R}^{m_2\times r}$ are column orthonormalized matrices of U_{11} and V_{11}. $\hat{M} \in \mathbb{R}^{m_1\times r}$ and $\hat{N} \in \mathbb{R}^{m_2\times r}$ are the first r left singular vectors of $A_{1\bullet}$ and $A_{\bullet 1}$, respectively. Also, recall that we use $P_U = U(U^\intercal U)^\dagger U^\intercal$ to represent the projection onto the column space of U.

1. We first give the lower bound for $\sigma_{\min}(\hat{M}^\intercal M)$, $\sigma_{\min}(\hat{N}^\intercal N)$ by the unilateral perturbation bound result in Cai and Zhang (2014). Since,

   $$P_{U_{11}}A_{1\bullet}=P_{U_{11}}U_{1\bullet}\Sigma V^\intercal=[U_{11}\Sigma_1,P_{U_{11}}U_{12}\Sigma_2]V^\intercal,\quad P_{U_{11}^\perp}A_{1\bullet}=P_{U_{11}^\perp}U_{1\bullet}\Sigma V^\intercal=[0,P_{U_{11}^\perp}U_{12}\Sigma_2]V^\intercal,$$

by V is an orthogonal matrix, we can see

$$\sigma_r(P_{U_{11}}A_{1\bullet}) = \sigma_r([U_{11}\Sigma_1 \quad P_{U_{11}}U_{12}\Sigma_2]) \geq \sigma_r(U_{11}\Sigma_1) \geq \sigma_r(A)\sigma_{\min}(U_{11}),$$

$$\|P_{U_{11}^\perp}A_{1\bullet}\| = \|P_{U_{11}^\perp}U_{12}\Sigma_2\| \leq \|P_{U_{11}^\perp}U_{12}\|\|\Sigma_2\| \leq \sigma_{r+1}(A).$$

So $\sigma_r(P_{U_{11}}A_{1\bullet}) \geq \|P_{U_{11}^\perp}A_{1\bullet}\|$. Besides, $\mathrm{rank}(P_{U_{11}}A_{1\bullet}) \leq r$. Apply the unilateral perturbation bound result in Cai and Zhang (2014) by setting $X = P_{U_{11}}A_{1\bullet}$, $Y = P_{U_{11}^\perp}A_{1\bullet}$, we have

$$\sigma_{\min}^2(\hat{M}^\intercal M) \leq 1 - \left(\frac{\|Y \cdot P_{X^\intercal}\| \cdot \sigma_{r+1}(A)}{\sigma_r^2(A)\sigma_{\min}^2(U_{11}) - \sigma_{r+1}^2(A)}\right)^2. \tag{34}$$

Moreover, $A_{1\bullet} = [U_{11}\ U_{12}]\mathrm{diag}(\Sigma_1, \Sigma_2)V^\intercal = [U_{11}\Sigma_1\ U_{12}\Sigma_2]V^\intercal$, and hence,

$$\|YP_{X^\intercal}\| = \left\|P_{U_{11}^\perp}A_{1\bullet} \cdot P_{(P_{U_{11}}A_{1\bullet})^\intercal}\right\| = \left\|[0 \quad P_{U_{11}^\perp}U_{12}\Sigma_2]V^\intercal \cdot P_{V\cdot[U_{11}\Sigma_1 \quad P_{U_{11}}U_{12}\Sigma_2]^\intercal}\right\|$$

$$= \left\|[0 \quad P_{U_{11}^\perp}U_{12}\Sigma_2] \cdot P_{[U_{11}\Sigma_1 \quad P_{U_{11}}U_{12}\Sigma_2]^\intercal}\right\| = \sup_{x\in\mathbb{R}^{p_2}, \|x\|_2=1} [0 \quad P_{U_{11}^\perp}U_{12}\Sigma_2] \cdot P_{[U_{11}\Sigma_1 \quad P_{U_{11}}U_{12}\Sigma_2]^\intercal} x.$$

When $\|x\|_2 = 1$, let y denote the projection of x onto the column space of $[U_{11}\Sigma_1\ \ P_{U_{11}}U_{12}\Sigma_2]^\intercal$. Then $\|y\|_2 \leq 1$ and y is in the column space of $[U_{11}\Sigma_1\ \ P_{U_{11}}U_{12}\Sigma_2]^\intercal$. Hence,

$$\frac{\|y_{[1:m_1]}\|_2}{\|y_{[(m_1+1):p_1]}\|_2} \geq \frac{\sigma_{\min}(U_{11}\Sigma_1)}{\|P_{U_{11}}U_{12}\Sigma_2\|} \geq \frac{\sigma_{\min}(U_{11})\sigma_r(A)}{\sigma_{r+1}(A)} \text{ and } \|y_{[(m_1+1):p_1]}\|_2^2 + \|y_{[1:m_1]}\|_2^2 \leq 1,$$

which implies $\|y_{[(m_1+1):p_1]}\|_2^2 \leq \sigma_{r+1}^2(A)/\sigma_{\min}^2(U_{11})\sigma_r^2(A) + \sigma_{r+1}^2(A)$. Hence for all $x \in \mathbb{R}^{p_2}$ such that $\|x\|_2 = 1$,

$$\begin{aligned}\left\|[0 \quad P_{U_{11}^\perp}U_{12}\Sigma_2] \cdot P_{[U_{11}\Sigma_1 \quad P_{U_{11}}U_{12}\Sigma_2]^\intercal} x\right\| \leq & \|P_{U_{11}^\perp}U_{12}\Sigma_2\| \cdot \|y_{[m_1+1:p_1]}\|_2 \\ \leq & \sigma_{r+1}(A)\frac{\sigma_{r+1}(A)}{\sqrt{\sigma_{r+1}^2(A) + \sigma_{\min}^2(U_{11})\sigma_r^2(A)}}.\end{aligned}$$

This yields $\|YP_{X^\intercal}\| = \|P_{U_{11}^\perp}A_{1\bullet} \cdot P_{(P_{U_{11}}A_{1\bullet})}\| \leq \sigma_{r+1}^2(A)/\sqrt{\sigma_{r+1}^2(A) + \sigma_{\min}^2(U_{11})\sigma_r^2(A)}$. Combining (34), we have

$$\sigma_{\min}^2(\hat{M}^\intercal M) \geq 1 - \left(\frac{\sigma_{r+1}^3(A)}{\sqrt{\sigma_{r+1}^2(A) + \sigma_{\min}^2(U_{11})\sigma_r^2(A)}\left(\sigma_r^2(A)\sigma_{\min}^2(U_{11}) - \sigma_{r+1}^2(A)\right)}\right)^2. \tag{35}$$

Since $\sigma_{\min}(U_{11})\sigma_r(A) \geq 2\sigma_{r+1}(A)$, we have

$$\sigma_{\min}^2(\hat{M}^\intercal M) \geq 1 - \left(\frac{1}{\sqrt{5}\cdot 3}\right)^2 \geq \frac{44}{45}.$$

Similarly, we also have $\sigma_{\min}^2(\hat{N}^\intercal N) \geq \frac{44}{45}$.

2. Following by (8),

$$\hat{A}_{22} = U_{2\bullet}\Sigma V_{1\bullet}^\intercal \hat{N}\left(\hat{M}^\intercal(U_{1\bullet}\Sigma V_{1\bullet}^\intercal)\hat{N}\right)^{-1}\hat{M}^\intercal U_{1\bullet}\Sigma V_{2\bullet}^\intercal$$
$$= \left(U_{21}\Sigma_1 V_{11}^\intercal \hat{N} + U_{22}\Sigma_2 V_{12}^\intercal \hat{N}\right)\left(\hat{M}^\intercal U_{11}\Sigma_1 V_{11}^\intercal \hat{N} + \hat{M}^\intercal U_{12}\Sigma_2 V_{12}^\intercal \hat{N}\right)^{-1}\left(\hat{M}^\intercal U_{11}\Sigma_1 V_{21}^\intercal + \hat{M}^\intercal U_{12}\Sigma_2 V_{22}^\intercal\right).$$

Let "L", "M", "R" stand for "Left", "Middle" and "Right",

$$B_L = U_{21}\Sigma_1 V_{11}^\intercal \hat{N}, \quad E_L = U_{22}\Sigma_2 V_{12}^\intercal \hat{N}; \tag{36}$$

$$B_M = \hat{M}^\intercal U_{11}\Sigma_1 V_{11}^\intercal \hat{N}, \quad E_M = \hat{M}^\intercal U_{12}\Sigma_2 V_{12}^\intercal \hat{N}; \tag{37}$$

$$B_R = \hat{M}^\intercal U_{11}\Sigma_1 V_{21}^\intercal, \quad E_R = \hat{M}^\intercal U_{12}\Sigma_2 V_{22}^\intercal. \tag{38}$$

By Lemma 2 in the Supplement, we can see the following properties of these matrices,

$$\|E_L\| \le \sigma_{r+1}(A), \quad \|E_M\| \le \sigma_{r+1}(A), \quad \|E_R\| \le \sigma_{r+1}(A), \tag{39}$$

$$\|E_L\|_q \le \|\Sigma_2\|_q, \quad \|E_M\|_q \le \|\Sigma_2\|_q, \quad \|E_R\|_q \le \|\Sigma_2\|_q, \tag{40}$$

$$\sigma_{\min}(B_M) = \sigma_{\min}\left(\hat{M}^\intercal(P_M U_{11})\Sigma_1(V_{11}^\intercal P_N)\hat{N}\right) = \sigma_{\min}\left((\hat{M}^\intercal M)(M^\intercal U_{11})\Sigma_1(V_{11}^\intercal N)(N^\intercal \hat{N})\right)$$
$$\ge \sigma_{\min}(\Sigma_1)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11})\sigma_{\min}(\hat{M}^\intercal M)\sigma_{\min}(\hat{N}^\intercal N) \ge \frac{44}{45}\sigma_r(A)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11}), \tag{41}$$

$$\|B_M^{-1}\| = \sigma_{\min}^{-1}(B_M) \le \frac{45}{44\sigma_r(A)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11})}, \tag{42}$$

$$\hat{A}_{22} = (B_L + E_L)(B_M + E_M)^{-1}(B_R + E_R), \quad B_L B_M^{-1} B_R = U_{21}\Sigma_1 V_{21}^\intercal, \tag{43}$$

$$\begin{aligned}\|B_L B_M^{-1}\| =& \|U_{21}\Sigma_1(V_{11}^\intercal\hat{N})(V_{11}^\intercal\hat{N})^{-1}\Sigma^{-1}(\hat{M}^\intercal U_{11})^{-1}\| = \|U_{21}(\hat{M}^\intercal U_{11})^{-1}\| \\ \le& \|(\hat{M}^\intercal M M^\intercal U_{11})^{-1}\| \le \frac{1}{\sigma_{\min}(M^\intercal U_{11})\sigma_{\min}(\hat{M}^\intercal M)} \le \frac{\sqrt{45/44}}{\sigma_{\min}(U_{11})},\end{aligned} \tag{44}$$

$$\|B_M^{-1} B_R\| = \|(V_{11}\hat{N})^{-1}V_{21}^\intercal\| \le \frac{\sqrt{45/44}}{\sigma_{\min}(V_{11})}. \tag{45}$$

By (39), (41) and the assumption (10), we can see $\sigma_{\min}(B_M) > \|E_M\|$, so

$$\hat{A}_{22} \overset{(43)}{=} (B_L + E_L)(B_M^{-1} - B_M^{-1}E_M B_M^{-1} + B_M^{-1}E_M B_M^{-1}E_M B_M^{-1} - \cdots)(B_R + E_R);$$

$$
\begin{aligned}
\|\hat{A}_{22} - B_L B_M^{-1} B_R\|_q \leq &\left\| B_L B_M^{-1} E_M \sum_{i=0}^{\infty} (-B_M^{-1} E_M)^i B_M^{-1} B_R \right\|_q + \left\| E_L \sum_{i=0}^{\infty} (-B_M^{-1} E_M)^i B_M^{-1} B_R \right\|_q \\
&+ \left\| B_L B_M^{-1} \sum_{i=0}^{\infty} (-E_M B_M^{-1})^i E_R \right\|_q + \left\| E_L B_M^{-1} \sum_{i=0}^{\infty} (-E_M B_M^{-1})^i E_R \right\|_q \\
\leq & \|B_L B_M^{-1}\| \|E_M\|_q \sum_{i=0}^{\infty} \|E_M\|^i \|B_M^{-1}\|^i \|B_M^{-1} B_R\| + \|E_L\|_q \sum_{i=0}^{\infty} \|B_M^{-1}\|^i \|E_M\|^i \|B_M^{-1} B_R\| \\
&+ \|B_L B_M^{-1}\| \sum_{i=0}^{\infty} \|E_M\|^i \|B_M^{-1}\|^i \|E_R\|_q + \|E_L\| \sum_{i=0}^{\infty} \|B_M^{-1}\|^{i+1} \|E_M\|^i \|E_R\|_q \\
\overset{(39)(40)}{\leq} & \frac{\|B_L B_M^{-1}\| \|B_M^{-1} B_R\| + \|B_M^{-1} B_R\| + \|B_L B_M^{-1}\| + \|B_M^{-1}\| \sigma_{r+1}(A)}{1 - \sigma_{r+1}(A) \|B_M^{-1}\|} \|\Sigma_2\|_q \\
\overset{(44)(45)}{\leq} & \frac{1}{1 - \sigma_{r+1}(A) \|B_M^{-1}\|} \left(\frac{45/44}{\sigma_{\min}(U_{11}) \sigma_{\min}(V_{11})} + \frac{\sqrt{45/44}}{\sigma_{\min}(U_{11})} + \frac{\sqrt{45/44}}{\sigma_{\min}(V_{11})} + \frac{45}{88} \right) \|\Sigma_2\|_q \\
\leq & \frac{\|A_{-\max(r)}\|_q}{1 - \frac{45 \sigma_{r+1}(A)}{44 \sigma_r(A) \sigma_{\min}(U_{11}) \sigma_{\min}(V_{11})}} \left(\frac{45/44}{\sigma_{\min}(U_{11}) \sigma_{\min}(V_{11})} + \frac{\sqrt{45/44}}{\sigma_{\min}(U_{11})} + \frac{\sqrt{45/44}}{\sigma_{\min}(V_{11})} + \frac{45}{88} \right) \\
\leq & \frac{88}{43} \|A_{-\max(r)}\|_q \left(\frac{45/44}{\sigma_{\min}(U_{11}) \sigma_{\min}(V_{11})} + \frac{\sqrt{45/44}}{\sigma_{\min}(U_{11})} + \frac{\sqrt{45/44}}{\sigma_{\min}(V_{11})} + \frac{45}{88} \right).
\end{aligned}
$$

Finally, since $A_{22} = U_{21} \Sigma_1 V_{21}^\intercal + U_{22} \Sigma_2 V_{22}^\intercal \overset{(43)}{=} B_L B_M^{-1} B_R + U_{22} \Sigma_2 V_{22}^\intercal$, we have

$$
\begin{aligned}
\|\hat{A}_{22} - A_{22}\|_q \leq & \|\hat{A}_{22} - B_L B_M^{-1} B_R\|_q + \|U_{22} \Sigma_2 V_{22}^\intercal\|_q \\
\leq & 3 \|A_{-\max(r)}\|_q \left(1 + \frac{1}{\sigma_{\min}(U_{11})}\right) \left(1 + \frac{1}{\sigma_{\min}(V_{11})}\right). \qquad \square
\end{aligned}
$$

Proof of Theorem 2

We only present proof for row thresholding as the column thresholding is essentially the same by working with A^T. Suppose M, N are orthonormal basis of column vectors of U_{11}, V_{11}. We denote $U^{(1)}_{[:,1:r]} = \hat{M}$, $V^{(2)}_{[:,1:r]} = \hat{N}$, which are exactly the same as the $\hat{M}$ and $\hat{N}$ in Algorithm 1. Similarly to the proof of Theorem 1, we have (35). Due to the assumption that $\sigma_r(A) \sigma_{\min}(U_{11}) \sigma_{\min}(V_{11}) \geq 4 \sigma_{r+1}(A)$, (35) yields

$$
\sigma^2_{\min}(\hat{M}^\intercal M) \geq 3824/3825, \quad \sigma^2_{\min}(\hat{N}^\intercal N) \geq 3824/3825. \tag{46}
$$

As shown in the Supplementary material, we have

Lemma 7 *Under the assumption of Theorem 2, we have $\hat{r} \geq r$.*

We next show (13) with the condition that $\hat{r} \geq r$ in steps.

1. Note that $A_{11} = U_{11}\Sigma_1 V_{11}^{\intercal} + U_{12}\Sigma_2 V_{12}^{\intercal}$, we consider the decompositions of Z and let

$$Z_{11} = U^{(2)\intercal} U_{11}\Sigma_1 V_{11}^{\intercal} V^{(1)} + U^{(2)\intercal} U_{12}\Sigma_2 V_{12}^{\intercal} V^{(1)},$$

$$Z_{11,[1:\hat{r},1:\hat{r}]} = U_{[:,1:\hat{r}]}^{(2)\intercal} U_{11}\Sigma_1 V_{11}^{\intercal} V_{[:,1:\hat{r}]}^{(1)} + U_{[:,1:\hat{r}]}^{(2)\intercal} U_{12}\Sigma_2 V_{12}^{\intercal} V_{[:,1:\hat{r}]}^{(1)} \triangleq B_{M,\hat{r}} + E_{M,\hat{r}}, \tag{47}$$

$$Z_{21,[:,1:\hat{r}]} = U_{21}\Sigma_1 V_{11}^{\intercal} V_{[:,1:\hat{r}]}^{(1)} + U_{22}\Sigma_2 V_{12}^{\intercal} V_{[:,1:\hat{r}]}^{(1)} \triangleq B_{L,\hat{r}} + E_{L,\hat{r}}, \tag{48}$$

$$Z_{12,[1:\hat{r},:]} = U_{[:,1:\hat{r}]}^{(2)\intercal} U_{11}\Sigma_1 V_{21}^{\intercal} + U_{[:,1:\hat{r}]}^{(2)\intercal} U_{12}\Sigma_2 V_{22}^{\intercal} \triangleq B_{R,\hat{r}} + E_{R,\hat{r}}. \tag{49}$$

Note that the square matrix $U_{[:,1:r]}^{(2)\intercal} M \in \mathbb{R}^{r\times r}$ is a submatrix of $U_{[:,1:\hat{r}]}^{(2)\intercal} M \in \mathbb{R}^{\hat{r}\times r}$, we know

$$\sigma_{\min}(U_{[:,1:\hat{r}]}^{(2)\intercal} M) \geq \sigma_{\min}(U_{[:,1:r]}^{(2)\intercal} M) = \sigma_{\min}(\hat{M}M) \overset{(46)}{\geq} \sqrt{\frac{3824}{3825}}. \tag{50}$$

Similarly, $\sigma_{\min}(V_{[:,1:\hat{r}]}^{(1)\intercal} N) \geq \sqrt{\frac{3824}{3825}}$. By M, N are the orthonormal basis of column vectors of U_{11}, V_{11}, we have $P_M = MM^{\intercal}$, $P_N = NN^{\intercal}$, and

$$\sigma_{\min}(U_{[:,1:\hat{r}]}^{(2)\intercal} U_{11}) \geq \sigma_{\min}(U_{[:,1:\hat{r}]}^{(2)\intercal} M)\sigma_{\min}(M^{\intercal} U_{11}) \geq \sqrt{\frac{3824}{3825}}\sigma_{\min}(U_{11}); \tag{51}$$

similarly, we also have

$$\sigma_{\min}(V_{[:,1:\hat{r}]}^{(1)\intercal} V_{11}) \geq \sqrt{\frac{3824}{3825}}\sigma_{\min}(V_{11}). \tag{52}$$

(51) and (52) immediately yield

$$\sigma_r(B_{M,\hat{r}}) \geq \frac{3824}{3825}\sigma_{\min}(U_{11})\sigma_{\min}(\Sigma_1)\sigma_{\min}(V_{11}) = \frac{3824}{3825}\sigma_r(A)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11}). \tag{53}$$

Besides, we also have

$$\|E_{M,\hat{r}}\| \overset{(47)}{\leq} \|\Sigma_2\| = \sigma_{r+1}(A) \tag{54}$$

2. Next, we consider the SVD of $Z_{11,[1:\hat{r},1:\hat{r}]}$

$$Z_{11,[1:\hat{r},1:\hat{r}]} = J\Lambda K^{\intercal}, \quad J, \Lambda, K \in \mathbb{R}^{\hat{r}\times\hat{r}}. \tag{55}$$

For convenience, we denote $\Lambda_1 = \Lambda_{[1:r,1:r]}, \Lambda_2 = \Lambda_{[(r+1):\hat{r},(r+1):\hat{r}]}$,

$$J_1 = J_{[:,1:r]}, \quad J_2 = J_{[:,(r+1):\hat{r}]}, \quad K_1 = K_{[:,1:r]}, \quad K_2 = K_{[:,(r+1):\hat{r}]}, \tag{56}$$

Suppose $M_Z \in \mathbb{R}^{\hat{r}\times r}$ is an orthonormal basis of the column space of $B_{M,\hat{r}}$; $N_Z \in \mathbb{R}^{\hat{r}\times r}$ is an orthonormal basis of the column space of $B_{M,\hat{r}}^\intercal$. Denote span$(\cdot)$ as the linear span of the column space of the matrix. We want to show span(M_Z) is close to span(J_1); while span(N_Z) is close to span(K_1). So in the rest of this step, we try to establish bounds for $\sigma_{\min}(J_1^\intercal M_Z)$ and $\sigma_{\min}(K_1^\intercal N_Z)$. Actually,

$$Z_{11,[1:\hat{r},1:\hat{r}]} = B_{M,\hat{r}} + E_{M,\hat{r}} = (B_{M,\hat{r}} + P_{M_Z} E_{M,\hat{r}}) + P_{M_Z^\perp} E_{M,\hat{r}}.$$

Now we set $X = (B_{M,\hat{r}} + P_{M_Z} E_{M,\hat{r}})$, $Y = P_{M_Z^\perp} E_{M,\hat{r}}$, then we have

$$\begin{aligned}\sigma_r(X) \geq & \sigma_r(B_{M,\hat{r}}) - \|P_{M_Z} E_{M,\hat{r}}\| \overset{(53)}{\geq} \frac{3824}{3825}\sigma_r(A)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11}) - \sigma_{r+1}(A),\\ & \overset{(12)}{\geq} \sigma_{r+1}(A) \overset{(54)}{\geq} \|E_{M,\hat{r}}\| \geq \|Y\|.\end{aligned}$$

Besides, by the definition of $B_{M,\hat{r}}$ and M_Z we know rank$(X) \leq r$. Also based on the definition of Y, we know $P_X Y = 0$. Now the unilateral perturbation bound in Cai and Zhang (2014) yields

$$\sigma_{\min}^2(M_Z^\intercal J_1) \geq 1 - \left(\frac{\sigma_r(X)\cdot\|Y\|}{\sigma_r^2(X) - \|Y\|^2}\right)^2. \tag{57}$$

The right hand side of the inequality above is an increasing function of $\sigma_r(X)$. Since $\sigma_r(X) \geq \frac{3824}{3825}\sigma_r(A)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11}) - \sigma_{r+1}(A) \geq (3 - \frac{4}{3825})\sigma_{r+1}(A) \geq (3 - \frac{4}{3825})\|Y\|$,

$$\sigma_{\min}^2(J_1^\intercal M_Z) \geq 1 - \left(\frac{3 - 4/3825}{(3-4/3825)^2 - 1}\right)^2 \geq 0.859. \tag{58}$$

Similarly, we also have

$$\sigma_{\min}^2(K_1^\intercal N_Z) \geq 0.859. \tag{59}$$

3. We next derive useful expressions of A_{22} and $\hat{A}_{22}$. First we introduce the following quantities,

$$J_1^\intercal Z_{11,[1:\hat{r},1:\hat{r}]} K_1 \overset{(47)}{=} J_1^\intercal B_{M,\hat{r}} K_1 + J_1^\intercal E_{M,\hat{r}} K_1 \triangleq B_{M1} + E_{M1}, \tag{60}$$

$$J_2^\intercal Z_{11,[1:\hat{r},1:\hat{r}]} K_2 \overset{(47)}{=} J_2^\intercal B_{M,\hat{r}} K_2 + J_2^\intercal E_{M,\hat{r}} K_2 \triangleq B_{M2} + E_{M2}, \tag{61}$$

$$Z_{21,[:,1:\hat{r}]} K_1 \overset{(48)}{=} B_{L,\hat{r}} K_1 + E_{L,\hat{r}} K_1 \triangleq B_{L1} + E_{L1}, \tag{62}$$

$$Z_{21,[:,1:\hat{r}]} K_2 \overset{(48)}{=} B_{L,\hat{r}} K_2 + E_{L,\hat{r}} K_2 \triangleq B_{L2} + E_{L2}, \tag{63}$$

$$J_1^\intercal Z_{12,[1:\hat{r},:]} \overset{(49)}{=} J_1^\intercal B_{R,\hat{r}} + J_1^\intercal E_{R,\hat{r}} \triangleq B_{R1} + E_{R1}, \tag{64}$$

$$J_2^\intercal Z_{11,[1:\hat{r},:]} \overset{(49)}{=} J_2^\intercal B_{R,\hat{r}} + J_2^\intercal E_{R,\hat{r}} \triangleq B_{R2} + E_{R2}. \tag{65}$$

Since

$$\begin{aligned} B_{L1} B_{M1}^{-1} B_{R1} &= B_{L,\hat{r}} K_1 \left(J_1^\intercal B_{M,\hat{r}} K_1\right)^{-1} J_1^\intercal B_{R,\hat{r}} \\ &= U_{21}\Sigma_1 V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)} K_1 \left(J_1^\intercal U_{[:,1:\hat{r}]}^{(2)\intercal} U_{11}\Sigma_1 V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)} K_1\right)^{-1} J_1^\intercal U_{[:,1:\hat{r}]}^{(2)\intercal} U_{11}\Sigma_1 V_{21}^\intercal = U_{21}\Sigma_1 V_{21}^\intercal, \end{aligned} \tag{66}$$

we can characterize $A_{22}, \hat{A}_{22}$ by these new notations as

$$A_{22} = U_{21}\Sigma_1 V_{21}^\intercal + U_{22}\Sigma_2 V_{22}^\intercal \overset{(66)}{=} B_{L1} B_{M1}^{-1} B_{R1} + U_{22}\Sigma_2 V_{22}^\intercal, \tag{67}$$

$$\begin{aligned} \hat{A}_{22} &= Z_{21,[:,1:\hat{r}]} Z_{11,[1:\hat{r},1:\hat{r}]}^{-1} Z_{12,[1:\hat{r},:]} \overset{(55)}{=} Z_{21,[:,1:\hat{r}]} K \left(J^\intercal Z_{11,[1:\hat{r},1:\hat{r}]} K\right)^{-1} J^\intercal Z_{12,[1:\hat{r},:]} \\ &= \left(Z_{21,[1:\hat{r}]} K_1 + Z_{21,[1:\hat{r}]} K_2\right) \left(J_1^\intercal Z_{11,[1:\hat{r},1:\hat{r}]} K_1 + J_2^\intercal Z_{11,[1:\hat{r},1:\hat{r}]} K_2\right)^{-1} \left(J_1^\intercal Z_{12,[1:\hat{r}]} + J_2^\intercal Z_{12,[1:\hat{r}]}\right) \\ &\overset{(60)-(65)}{=} \sum_{k=1}^{2} (B_{Lk} + E_{Lk})(B_{Mk} + E_{Mk})^{-1}(B_{Rk} + E_{Rk}) \end{aligned} \tag{68}$$

4. We now establish a number of bounds for the terms on the right hand side of (60)-(65).

Lemma 8 *Based on the assumptions above, we have*

$$\sigma_{\min}(B_{M1}) \geq 3.43\sigma_{r+1}(A); \tag{69}$$

$$\|B_{L1} B_{M1}^{-1}\| \leq \frac{\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(U_{11})}, \quad \|B_{M1}^{-1} B_{R1}\| \leq \frac{\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(V_{11})}, \tag{70}$$

$$\|E_{Mt}\|_q \leq \|A_{-\max(r)}\|_q, \ \|E_{Lt}\|_q \leq \|A_{-\max(r)}\|_q, \ \|E_{Rt}\|_q \leq \|A_{-\max(r)}\|_q, \quad t = 1, 2, \tag{71}$$

$$\|(B_{L2} + E_{L2})(B_{M2} + E_{M2})^{-1}\| \leq T_R + \frac{1}{1 - 1/3.43}\left(\frac{\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(U_{11})} + \frac{1}{3.43}\right), \tag{72}$$

$$\|B_{R2}\|_q \leq \frac{2\sqrt{3825/3824}}{\sqrt{0.859\sigma_{\min}(V_{11})}} \|A_{-\max(r)}\|_q. \tag{73}$$

The proof of Lemma 8 is given in the Supplement.

5. We finally give the upper bound of $\|\hat{A}_{22} - A_{22}\|_q$. By (67) and (68), we can split the loss as,

$$\begin{aligned}\hat{A}_{22} - A_{22} = &\left((B_{L1} + E_{L1}) (B_{M1} + E_{M1})^{-1} (B_{R1} + E_{R1}) - B_{L1} B_{M1}^{-1} B_{R1} \right) \\ &+ (B_{L2} + E_{L2}) (B_{M2} + E_{M2})^{-1} (B_{R2} + E_{R2}) - U_{22} \Sigma_2 V_{22}^{\intercal}.\end{aligned} \tag{74}$$

We will analyze them separately. First, $\|U_{22}\Sigma_2 V_{22}^{\intercal}\|_q \leq \|A_{-\max(r)}\|_q$; second,

$$\begin{aligned}&\|(B_{L2} + E_{L2})(B_{M2} + E_{M2})^{-1}(B_{R2} + E_{M2})\|_q \\ \leq &\|(B_{L2} + E_{L2})(B_{M2} + E_{M2})^{-1}\| \cdot (\|B_{R2}\|_q + \|E_{M2}\|_q) \\ \overset{(72)(73)}{\leq} &\left(T_R + \frac{3.43}{2.43} \left(\frac{\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(U_{11})} + \frac{1}{3.43} \right) \right) \left(\frac{2\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(V_{11})} + 1 \right) \|A_{-\max(r)}\|_q \\ \leq &\left(T_R + \frac{1.524}{\sigma_{\min}(U_{11})} + 0.412 \right) \left(\frac{2.16}{\sigma_{\min}(V_{11})} + 1 \right) \|A_{-\max(r)}\|_q.\end{aligned} \tag{75}$$

The analysis of $\left((B_{L1} + E_{L1}) (B_{M1} + E_{M1})^{-1} (B_{R1} + E_{R1}) - B_{L1} B_{M1}^{-1} B_{R1} \right)$ is similar to the proof of Theorem 1. We have

$$\begin{aligned}&\left\|(B_{L1} + E_{L1})(B_{M1} + E_{M1})^{-1}(B_{R1} + E_{R1}) - B_{L1}B_{M1}^{-1}B_{R1}\right\|_q \\ \leq &\left\| B_{L1}(B_{M1}^{-1}E_{M1}\sum_{i=0}^{\infty}(-B_{M1}^{-1}E_{M1})^i B_{M1}^{-1})B_{R1} \right\|_q + \left\| E_{L1}\left(\sum_{i=0}^{\infty}(-B_{M1}^{-1}E_{M1})^i B_{M1}^{-1} \right) B_{R1} \right\|_q \\ &+ \left\| B_{L1}\left(B_{M1}^{-1}\sum_{i=0}^{\infty}(-E_{M1}B_{M1}^{-1})^i \right) E_{R1} \right\|_q + \left\| E_{L1}\left(B_{M1}^{-1}\sum_{i=0}^{\infty}(-E_{M1}B_{M1}^{-1})^i \right) E_{R1} \right\|_q \\ \leq &\|B_{L1}B_{M1}^{-1}\|\|E_{M1}\|_q \sum_{i=0}^{\infty} \|E_{M1}\|^i \|B_{M1}^{-1}\|^i \|B_{M1}^{-1}B_{R1}\| + \|E_{L1}\|_q \sum_{i=0}^{\infty} \|B_{M1}^{-1}\|^i \|E_{M1}\|^i \|B_{M1}^{-1}B_{R1}\| \\ &+ \|B_{L1}B_{M1}^{-1}\| \sum_{i=0}^{\infty} \|E_{M1}\|^i \|B_{M1}^{-1}\|^i \|E_{R1}\|_q + \|E_{L1}\| \sum_{i=0}^{\infty} \|B_{M1}^{-1}\|^{i+1} \|E_{M1}\|^i \|E_{R1}\|_q \\ \overset{(71)}{\leq} &\frac{\|\Sigma_2\|_q}{1 - \sigma_{r+1}(A)\|B_{M1}^{-1}\|} \left(\|B_{L1}B_{M1}^{-1}\|\|B_{M1}^{-1}B_{R1}\| + \|B_{M1}^{-1}B_{R1}\| + \|B_{L1}B_{M1}^{-1}\| + \|B_{M1}^{-1}\|\sigma_{r+1}(A) \right) \\ \overset{(70)(69)}{\leq} &\left(\frac{1.65}{\sigma_{\min}(U_{11})\sigma_{\min}(V_{11})} + \frac{1.53}{\sigma_{\min}(V_{11})} + \frac{1.53}{\sigma_{\min}(V_{11})} + 0.42 \right) \|A_{-\max(r)}\|_q.\end{aligned} \tag{76}$$

From (75), (76), (74), and the fact that $\sigma_{\min}(U_{11}) \leq 1$ and $T_R \geq \frac{1.36}{\sigma_{\min}(U_{11})} + 0.35$,

$$\begin{aligned}
\|\hat{A}_{22} - A_{22}\|_q \leq& \left(2.16T_R + \left(\frac{4.95}{\sigma_{\min}(U_{11})} + 2.42\right)\right)\left(\frac{1}{\sigma_{\min}(V_{11})} + 1\right)\|A_{-\max(r)}\|_q \\
\leq& \left(2.16T_R + 4.31\left(\frac{1.36}{\sigma_{\min}(U_{11})} + 0.35\right)\right)\left(\frac{1}{\sigma_{\min}(V_{11})} + 1\right)\|A_{-\max(r)}\|_q \quad (77)\\
\leq& 6.5T_R\left(\frac{1}{\sigma_{\min}(V_{11})} + 1\right)\|A_{-\max(r)}\|_q.
\end{aligned}$$

This concludes the proof. □

Proof of Lemma 7.

In order to prove this lemma, we just need to prove that the for-loop in Algorithm 2 will break for some $s \geq r$. This can be shown by proving the break condition

$$\|D_{R,s}\| = \|Z_{21,[1:s]}Z^{-1}_{11,[1:s,1:s]}\| \leq T_R, \tag{78}$$

hold for $s = r$.

We adopt the definitions in (36), (37), (38), then we have

$$\begin{aligned}
Z_{11,[1:r,1:r]} &= U^{(2)\intercal}_{[:,1:r]}A_{11}V^{(1)}_{[:,1:r]} = \hat{M}^\intercal A_{11}\hat{N} \\
&= \hat{M}^\intercal U_{11}\Sigma_1 V_{11}^\intercal \hat{N} + \hat{M}^\intercal U_{12}\Sigma_2 V_{12}^\intercal \hat{N} \\
&= B_M + E_M, \\
Z_{21,[:,1:r]} &= A_{21}V^{(1)}_{[:,1:r]} = \left(U_{21}\Sigma_1 V_{11}^\intercal + U_{22}\Sigma_2 V_{12}^\intercal\right)\hat{N} = B_L + E_L.
\end{aligned}$$

Hence,

$$\begin{aligned}
\left\|Z_{21,[:,1:r]}Z^{-1}_{11,[1:r,1:r]}\right\| =& \|(B_L + E_L)(B_M + E_M)^{-1}\| \\
\leq& \left\|B_L B_M^{-1}\sum_{i=0}^{\infty}(-E_M B_M^{-1})^i\right\| + \left\|E_L B_M^{-1}\sum_{i=0}^{\infty}(-E_M B_M^{-1})^i\right\| \\
\leq& \left(\|B_L B_M^{-1}\| + \|E_L\|\|B_M^{-1}\|\right)\frac{1}{1 - \|E_M B_M^{-1}\|} \\
\overset{(41),(70)}{\leq}& \left(\frac{\sqrt{45/44}}{\sigma_{\min}(U_{11})} + \frac{45\sigma_{r+1}(A)}{44\sigma_r(A)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11})}\right)\frac{1}{1 - \frac{45\sigma_{r+1}(A)}{44\sigma_r(A)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11})}} \\
\leq& \frac{1.36}{\sigma_{\min}(U_{11})} + 0.35 \leq T_R,
\end{aligned}$$

which finished the proof of the lemma. $\square$

Proof of Lemma 8.

First, since $M_Z \in \mathbb{R}^{\hat{r}\times r}$ and $N_Z \in \mathbb{R}^{\hat{r}\times r}$ are an orthonormal basis of $B_{M,\hat{r}}$ and $B_{M,\hat{r}}^\intercal$, we have $P_{M_Z} = M_Z M_Z^\intercal$ and $P_{N_Z} = N_Z N_Z^\intercal$ and

$$\begin{aligned}\sigma_{\min}(B_{M1}) =& \sigma_{\min}(J_1^\intercal B_{M,\hat{r}} K_1) = \sigma_{\min}(J_1^\intercal M_Z M_Z^\intercal B_{M,\hat{r}} N_Z N_Z^\intercal K_1)\\ \geq& \sigma_{\min}(J_1^\intercal M_Z)\sigma_{\min}(M_Z^\intercal B_{M,\hat{r}} N_Z)\sigma_{\min}(N_Z^\intercal K_1) \\ \overset{(58)(59)}{\geq}& 0.859\sigma_r(B_{M,\hat{r}}) \overset{(53)}{\geq} \frac{0.859\cdot 3824}{3825}\sigma_r(A)\sigma_{\min}(U_{11})\sigma_{\min}(V_{11}) \overset{(12)}{\geq} 3.43\sigma_{r+1}(A).\end{aligned} \tag{79}$$

which gives (69).

$$\begin{aligned}&\|B_{L1}B_{M1}^{-1}\| = \left\|B_{L,\hat{r}}K_1\left(J_1^\intercal B_{M,\hat{r}}K_1\right)^{-1}\right\| \\ =& \left\|U_{21}\Sigma_1 V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)} K_1\left(J_1^\intercal U_{[:,1:\hat{r}]}^{(2)\intercal} U_{11}\Sigma_1 V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)} K_1\right)^{-1}\right\| = \left\|U_{21}\left(J_1^\intercal U_{[:,1:\hat{r}]}^{(2)\intercal} U_{11}\right)^{-1}\right\| \\ \leq& \frac{1}{\sigma_{\min}(J_1^\intercal U_{[:,1:\hat{r}]}^{(2)\intercal}U_{11})} = \frac{1}{\sigma_{\min}(J_1^\intercal P_{M_Z}(U_{[:,1:\hat{r}]}^{(2)\intercal}U_{11}))} = \frac{1}{\sigma_{\min}((J_1^\intercal M_Z)(M_Z^\intercal U_{[:,1:\hat{r}]}^{(2)\intercal}U_{11}))} \\ \leq& \frac{1}{\sigma_{\min}(J_1^\intercal M_Z)}\cdot\frac{1}{\sigma_{\min}(U_{[:,1:\hat{r}]}^{(2)\intercal}U_{11})} \overset{(51)(58)}{\leq} \frac{\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(U_{11})},\end{aligned} \tag{80}$$

which gives the first part of (70). Here we used the fact that $\Sigma_1 V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)} K_1$ is a square matrix; M_Z is the orthonormal basis of the column space of $Z_{11,[1:\hat{r},1:\hat{r}]} = U_{[:,1:\hat{r}]}^{(2)\intercal} U_{11}\Sigma_1 V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)}$. Similarly we have the later part of (70),

$$\|B_{M1}^{-1}B_{R1}\| \leq \frac{\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(V_{11})}. \tag{81}$$

Based on the definitions, we have the bound for all " E" terms in (60)-(65), i.e. (71). Now we move on to (72). By the SVD of $Z_{11,[1:\hat{r},1:\hat{r}]}$ (55) and the partition (56), we know

$$\left([J_1\ J_2]^\intercal Z_{11,[1:\hat{r},1:\hat{r}]}[K_1\ K_2]\right)^{-1} = \begin{bmatrix}\Lambda_1 & 0\\ 0 & \Lambda_2\end{bmatrix}^{-1} = \begin{bmatrix}\left(J_1^\intercal Z_{11,[1:\hat{r},1:\hat{r}]}K_1\right)^{-1} & 0 \\ 0 & \left(J_2^\intercal Z_{11,[1:\hat{r},1:\hat{r}]}K_2\right)^{-1}\end{bmatrix}.$$

Hence, we have

$$
\begin{aligned}
&\left\|(B_{L2}+E_{L2})(B_{M2}+E_{M2})^{-1}\right\| = \left\|Z_{21,[:,1:\hat{r}]}K_2\left(J_2^\intercal Z_{11,[1:\hat{r},1:\hat{r}]}K_2\right)^{-1}\right\| \\
=&\left\|Z_{21,[:,1:\hat{r}]}[K_1\ K_2]\left([J_1\ J_2]^\intercal Z_{11,[1:\hat{r},1:\hat{r}]}[K_1\ K_2]\right)^{-1} - Z_{21,[1:\hat{r}]}K_1\left(J_1^\intercal Z_{11,[1:\hat{r},1:\hat{r}]}K_1\right)^{-1}\right\| \\
\le&\left\|Z_{21,[:,1:\hat{r}]}\left(Z_{11,[1:\hat{r},1:\hat{r}]}\right)^{-1}\right\| + \left\|(B_{L1}+E_{L1})(B_{M1}+E_{M1})^{-1}\right\| \\
\le& T_R + \left\|B_{L1}\cdot B_{M1}^{-1}\sum_{i=0}^{\infty}(-E_{M1}B_{M1}^{-1})^i\right\| + \left\|E_{L1}\cdot B_{M1}^{-1}\sum_{i=0}^{\infty}(-E_{M1}B_{M1}^{-1})^i\right\| \\
\le& T_R + \left(\|B_{L1}B_{M1}^{-1}\| + \|E_{L1}\|\|B_{M1}^{-1}\|\right)\frac{1}{1-\|E_{M1}\|\|B_{M1}^{-1}\|} \\
\overset{(69)(70)(71)}{\le}& T_R + \left(\frac{\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(U_{11})} + \frac{1}{3.43}\right)\cdot\frac{1}{1-1/3.43},
\end{aligned}
\tag{82}
$$

which proves (72). Since $Z_{11,[1:\hat{r},1:\hat{r}]} = B_{M,\hat{r}} + E_{M,\hat{r}}$ and by definition, $\mathrm{rank}(B_{M,\hat{r}}) \le r$, by Lemma 1, we know

$$
\sigma_{r+i}(Z_{11,[1:\hat{r},1:\hat{r}]}) \le \sigma_i(E_{M,\hat{r}}), \quad \forall i \ge 1. \tag{83}
$$

Then

$$
\begin{aligned}
\|B_{M2}\|_q \le&\|B_{M2}+E_{M2}\|_q + \|E_{M2}\|_q \le \|J_2^\intercal Z_{11,[1:\hat{r},1:\hat{r}]}K_2\|_q + \|E_{M2}\|_q \\
=&\sqrt[q]{\sum_{i=r+1}^{\hat{r}}\sigma_i^q(Z_{11,[1:\hat{r},1:\hat{r}]})} + \|E_{M2}\|_q \le \sqrt[q]{\sum_{i=1}^{\hat{r}-r}\sigma_i^q(E_{M,\hat{r}})} + \|E_{M2}\|_q \\
\le&\|E_{M,\hat{r}}\|_q + \|E_{M2}\|_q \overset{(71)}{\le} 2\|A_{-\max(r)}\|_q.
\end{aligned}
\tag{84}
$$

Same to the process of (80), we know

$$
\frac{1}{\sigma_{\min}(V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)}K_1)} \le \frac{\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(V_{11})}. \tag{85}
$$

Also, $\|V_{21}^\intercal\| \le 1$. Hence,

$$
\begin{aligned}
\|B_{R2}\|_q \overset{(65)}{=}&\|J_2^\intercal B_{R,\hat{r}}\|_q = \|J_2^\intercal U_{[:,1:\hat{r}]}^{(2)\intercal}U_{11}\Sigma_1 V_{21}^\intercal\|_q \\
=&\|J_2^\intercal U_{[:,1:\hat{r}]}^{(2)\intercal}U_{11}\Sigma_1(V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)}K_1)(V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)}K_1)^{-1}V_{21}^\intercal\|_q \\
\le&\|B_{M2}\|_q\cdot\|(V_{11}^\intercal V_{[:,1:\hat{r}]}^{(1)}K_1)^{-1}\|\cdot\|V_{21}^\intercal\| \\
\overset{(84)(85)}{\le}& \frac{2\sqrt{3825/3824}}{\sqrt{0.859}\sigma_{\min}(V_{11})}\|A_{-\max(r)}\|_q.
\end{aligned}
\tag{86}
$$

which proves (73). $\square$

Proof of Theorem 3.

The idea of proof is to construct two matrices $A^{(1)}, A^{(2)}$ both in $\mathcal{F}_c(M_1, M_2)$ such that they have the identical first m_1 rows and m_2 columns, but differ much in the remaining block. Suppose $a, b, c > 0$ are fixed numbers, ε is a small real number. We first consider the following 2-by-2 matrix

$$B(\varepsilon) = \begin{bmatrix} a & c \\ b & \frac{bc}{a} + \varepsilon \end{bmatrix}. \tag{87}$$

Suppose the larger and smaller singular value of $B(\varepsilon)$ are $\lambda_{\max}(\varepsilon)$ and $\lambda_{\min}(\varepsilon)$, then we have

$$\lambda_{\max}(\varepsilon) \to \|B(0)\| = \frac{\sqrt{(a^2+b^2)(a^2+c^2)}}{a} \tag{88}$$

as $\varepsilon \to 0$; since $\lambda_{\max}(\varepsilon) \cdot \lambda_{\min}(\varepsilon) = |\det(B)| = a|\varepsilon|$, we also have

$$\lambda_{\min}(\varepsilon)/|\varepsilon| \to \frac{a^2}{\sqrt{(a^2+b^2)(a^2+c^2)}} \tag{89}$$

as $\varepsilon \to 0$. If $B(\varepsilon)$ defined in (87) has SVD

$$B(\varepsilon) = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{21} \end{bmatrix} \cdot \begin{bmatrix} \lambda_{\max}(\varepsilon) & 0 \\ 0 & \lambda_{\min}(\varepsilon) \end{bmatrix} \cdot \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{21} \end{bmatrix}^{\intercal} \tag{90}$$

then we also have

$$u_{11} \to \frac{a}{\sqrt{a^2+b^2}}, u_{21} \to \frac{b}{\sqrt{a^2+b^2}}, \quad v_{11} \to \frac{a}{\sqrt{a^2+c^2}}, \quad v_{21} \to \frac{c}{\sqrt{a^2+c^2}}. \tag{91}$$

as $\varepsilon \to 0$.

Now we set $a = 1$, $b = \sqrt{1-M_1^2}/M_1 - \eta$, $c = \sqrt{1-M_2^2}/M_2 - \eta$, $d = bc/a$, where η is some small positive number to be specify later. We construct $A_{11}, A_{12}, A_{21}, A_{22}^{(1)}$ and $A_{22}^{(2)}$ such that,

$$A_{11} = \begin{bmatrix} aI_r & 0 \\ 0 & 0 \end{bmatrix}_{m_1\times m_2}, \quad A_{12} = \begin{bmatrix} cI_r & 0 \\ 0 & 0 \end{bmatrix}_{m_1\times(p_2-m_2)}, \quad A_{21} = \begin{bmatrix} bI_r & 0 \\ 0 & 0 \end{bmatrix}_{(p_1-m_1)\times m_2}; \tag{92}$$

$$A_{22}^{(1)} = \begin{bmatrix} (d+\varepsilon)I_r & 0 \\ 0 & 0 \end{bmatrix}_{(p_1-m_1)\times(p_2-m_2)}, \quad A_{22}^{(2)} = \begin{bmatrix} (d-\varepsilon)I_r & 0 \\ 0 & 0 \end{bmatrix}_{(p_1-m_1)\times(p_2-m_2)}. \tag{93}$$

Here we use I_r to note the identity matrix of dimension r. Then we construct $A^{(1)}$ and $A^{(2)}$ as

$$A^{(1)} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}^{(1)} \end{bmatrix}, \quad A^{(2)} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}^{(2)} \end{bmatrix}, \tag{94}$$

where $A^{(1)}$ and $A^{(2)}$ are with identical first m_1 rows and m_2 columns. Since the SVD of $B(\varepsilon)$ is given as (90), the SVD of $A^{(1)}$ can be written as

$$A^{(1)} = \begin{bmatrix} U_{11}^{(1)} & U_{12}^{(1)} \\ U_{21}^{(1)} & U_{22}^{(1)} \end{bmatrix} \cdot \begin{bmatrix} \Sigma_1^{(1)} & 0 \\ 0 & \Sigma_2^{(1)} \end{bmatrix} \cdot \begin{bmatrix} V_{11}^{(1)} & V_{12}^{(1)} \\ V_{21}^{(1)} & V_{22}^{(1)} \end{bmatrix}^\intercal,$$

where

$$U_{11} = \begin{bmatrix} u_{11}I_r \\ 0 \end{bmatrix}_{m_1\times r}, \quad U_{12} = \begin{bmatrix} u_{12}I_r \\ 0 \end{bmatrix}_{m_1\times r}, \quad U_{21} = \begin{bmatrix} u_{21}I_r \\ 0 \end{bmatrix}_{(p_1-m_1)\times r}, \quad U_{22} = \begin{bmatrix} u_{22}I_r \\ 0 \end{bmatrix}_{(p_1-m_1)\times r};$$

$$V_{11} = \begin{bmatrix} v_{11}I_r \\ 0 \end{bmatrix}_{m_2\times r}, \quad V_{12} = \begin{bmatrix} v_{12}I_r \\ 0 \end{bmatrix}_{m_2\times r}, \quad V_{21} = \begin{bmatrix} v_{21}I_r \\ 0 \end{bmatrix}_{(p_2-m_2)\times r}, \quad V_{22} = \begin{bmatrix} v_{22}I_r \\ 0 \end{bmatrix}_{(p_2-m_2)\times r};$$

$$\Sigma_1 = \lambda_{\max}(\varepsilon)I_r, \quad \Sigma_2 = \lambda_{\min}(\varepsilon)I_r.$$

Hence,

$$\sigma_{\min}(U_{11}) = u_{11} = \frac{a}{\sqrt{a^2+b^2}} \to \frac{1}{1+\left(\frac{\sqrt{1-M_1^2}}{M_1} - \eta\right)^2} > M_1, \quad \text{as } \varepsilon \to 0$$

$$\sigma_{\min}(V_{11}) = v_{11} = \frac{a}{\sqrt{a^2+c^2}} \to \frac{1}{1+\left(\frac{\sqrt{1-M_2^2}}{M_2} - \eta\right)^2} > M_2, \quad \text{as } \varepsilon \to 0.$$

Also, $\|\Sigma_2^{(1)}\| \to 0$ as $\varepsilon \to 0$. So we have $A^{(1)} \in \mathcal{F}_r(M_1, M_2)$ when ε is small enough. Similarly $A^{(2)} \in \mathcal{F}_r(M_1, M_2)$ when ε is small enough. Now we also have $\|A^{(1)}_{-\max(r)}\|_q = (q\lambda_{\min}(\varepsilon)^q)^{1/q} = q^{1/q}\lambda_{\min}(\varepsilon)$, $\|A^{(2)}_{-\max(r)}\|_q = (q\lambda_{\min}(-\varepsilon)^q)^{1/q} = q^{1/q}\lambda_{\min}(-\varepsilon)$. $\|A_{22}^{(1)} - A_{22}^{(2)}\|_q = (q(2|\varepsilon|)^q)^{1/q} = 2|\varepsilon|q^{1/q}$.

Finally for any estimate $\hat{A}_{22}$, we must have

$$\max\left\{\frac{\|\hat{A}_{22}-A_{22}^{(1)}\|_q}{\|A^{(1)}_{-\max(r)}\|_q}, \frac{\|\hat{A}_{22}-A_{22}^{(2)}\|_q}{\|A^{(2)}_{-\max(r)}\|_q}\right\} \geq \frac{\frac{1}{2}\left\|\left(\hat{A}_{22}-A_{22}^{(1)}\right)-\left(\hat{A}_{22}-A_{22}^{(2)}\right)\right\|_q}{\min\left\{\|A^{(1)}_{-\max(r)}\|_q, \|A^{(2)}_{-\max(r)}\|_q\right\}}$$

$$\geq \frac{2|\varepsilon|}{2\min\left\{\lambda_{\min}(\varepsilon), \lambda_{\min}(-\varepsilon)\right\}} \overset{(89)}{\to} \frac{\sqrt{(a^2+b^2)(a^2+c^2)}}{a^2} \tag{95}$$

$$=\sqrt{\left(1+(\frac{\sqrt{1-M_1^2}}{M_1}-\eta)^2\right)\left(1+(\frac{\sqrt{1-M_2^2}}{M_2}-\eta)^2\right)}$$

as $\varepsilon \to 0$. Since $A^{(1)}, A^{(2)} \in \mathcal{F}_r(M_1, M_2)$ and are with identical first m_1 rows and m_2 columns, we must have

$$\inf_{\hat{A}_{22}} \sup_{A\in\mathcal{F}_r(M_1,M_2)} \frac{\|\hat{A}_{22}-A_{22}\|_q}{\|A_{-\max(r)}\|_q} \geq \sqrt{\left(1+(\frac{\sqrt{1-M_1^2}}{M_1}-\eta)^2\right)\left(1+(\frac{\sqrt{1-M_2^2}}{M_2}-\eta)^2\right)}.$$

Let $\eta \to 0$, since $M_1, M_2 < 1$, we have

$$\inf_{\hat{A}_{22}} \sup_{A\in\mathcal{F}_r(M_1,M_2)} \frac{\|\hat{A}_{22}-A_{22}\|_q}{\|A_{-\max(r)}\|_q} \geq \frac{1}{M_1M_2} \geq \frac{1}{4}\left(\frac{1}{M_1}+1\right)\left(\frac{1}{M_2}+1\right), \tag{96}$$

which finished the proof of theorem. □

Proof of Corollary 1.

We first prove the second part of the corollary. We set $\alpha = (136/165)^2$. Since $U_{[:,1:r]} \in \mathbb{R}^{p_1\times r}$ is with orthonormal columns, by Lemma 5 and

$$m_1 \geq 12.5W_r^{(1)}r(\log r + c) \geq \frac{4}{(1-\alpha)^2}\cdot W_r^{(1)}r(\log r + c),$$

we have

$$\sigma_{\min}(U_{11}) = \sigma_{\min}(U_{[\Omega_1,1:r]}) \geq \sqrt{\frac{\alpha m_1}{p_1}} \tag{97}$$

with probability at least $1-2\exp(-c)$. When (97) holds, by the condition, we know

$$\sigma_{r+1}(A) \leq \sigma_r(A)\sigma_{\min}(V_{11})\frac{1}{5}\sqrt{\frac{m_1}{p_1}} \leq \sigma_r(A)\sigma_{\min}(V_{11})\frac{1}{5\sqrt{\alpha}}\cdot\sigma_{\min}(U_{11}) \leq \frac{1}{4}\sigma_r(A)\sigma_{\min}(V_{11})\sigma_{\min}(U_{11}).$$

When $T_R \geq 2\sqrt{p_1/m_1}$, we have

$$\frac{1.36}{\sigma_{\min}(U_{11})} + 0.35 \leq 1.36\sqrt{\frac{p_1}{\alpha m_1}} + 0.35 \leq 2\sqrt{\frac{p_1}{m_1}} \leq T_R$$

Hence we can apply Theorem 2, for $1 \leq q \leq \infty$ we must have

$$\left\|\hat{A}_{22} - A_{22}\right\|_q \leq 6.5T_R \left\|A_{-\max(r)}\right\|_q \left(\frac{1}{\sigma_{\min}(V_{11})} + 1\right), \tag{98}$$

which finishes the proof of the second part of Corollary 1. Besides, the proof for the third part is the same as the second part after we take the transpose of the matrix.

For the first part, the proof is also similar. Again we set $\alpha = (136/165)^2$. Then we have

$$m_1 \geq \frac{4}{(1-\alpha)^2}W_r^{(1)}r(\log r + c), \quad m_2 \geq \frac{4}{(1-\alpha)^2}W_r^{(2)}r(\log r + c),$$

so

$$\sigma_{\min}(U_{11}) = \sigma_{\min}(U_{[\Omega_1,1:r]}) \geq \sqrt{\frac{\alpha m_1}{p_1}}, \quad \sigma_{\min}(V_{11}) = \sigma_{\min}(V_{[\Omega_2,1:r]}) \geq \sqrt{\frac{\alpha m_2}{p_2}} \tag{99}$$

with probability at least $1 - 4\exp(-c)$. When (99) holds, we have

$$\sigma_{r+1}(A) \leq \sigma_r(A)\frac{1}{6}\sqrt{\frac{m_1 m_2}{p_1 p_2}} \leq \sigma_r(A)\frac{1}{6\alpha}\sigma_{\min}(U_{11})\sigma_{\min}(V_{11}) \leq \frac{1}{4}\sigma_r(A)\sigma_{\min}(V_{11})\sigma_{\min}(U_{11}).$$

When $T_R = 2\sqrt{p_1/m_1}$ or $T_C = 2\sqrt{p_2/m_2}$, similarly to the first part we have

$$\frac{1.36}{\sigma_{\min}(U_{11})} + 0.35 \leq T_R, \quad \text{or} \quad \frac{1.36}{\sigma_{\min}(V_{11})} + 0.35 \leq T_C.$$

Hence we can apply Theorem 2 and get

$$\begin{aligned}\left\|\hat{A}_{22} - A_{22}\right\|_q \leq & 6.5T_R\|A_{-\max(r)}\|_q\left(\frac{1}{\sigma_{\min}(V_{11})} + 1\right) \leq 6.5 \cdot 2\sqrt{\frac{p_1}{m_1}} \cdot \left(\sqrt{\frac{p_2}{\alpha m_2}} + 1\right)\|A_{-\max(r)}\|_q \\ \leq & 29\|A_{-\max(r)}\|_q\sqrt{\frac{p_1 p_2}{m_1 m_2}}.\end{aligned}$$

□

Proof of Corollary 2.

Suppose $0 < \alpha_1 < 1$, since $U_{[:,1:r]} \in \mathbb{R}$ is with random orthonormal columns of Haar measure, we can apply Lemma 6 and find some $c > 0$ and $\delta > 0$ such that when $p_1 \geq m_1 \geq cr$,

$$\sigma_{\min}(U_{11}) = \sigma_{\min}(U_{[1:m_1,1:r]}) \geq \frac{136}{165}\sqrt{\frac{m_1}{p_1}} \tag{100}$$

with probability at least $1 - \exp(-\delta m_1)$. When (100) happen, we have

$$\sigma_{r+1}(A) \leq \sigma_r(A)\sigma_{\min}(V_{11})\frac{1}{5}\sqrt{\frac{m_1}{p_1}} \leq \sigma_r(A)\sigma_{\min}(V_{11})\sigma_{\min}(U_{11}),$$

$$\frac{1.36}{\sigma_{\min}(U_{11})} + 0.35 \leq 1.36 \cdot \frac{165}{136}\sqrt{\frac{p_1}{m_1}} + 0.35 \leq 2\sqrt{\frac{p_1}{m_1}}.$$

Hence we can apply Theorem 2, for $1 \leq q \leq \infty$, we have

$$\left\|\hat{A}_{22} - A_{22}\right\|_q \leq 6.5T_R \left\|A_{-\max(r)}\right\|_q \left(\frac{1}{\sigma_{\min}(V_{11})} + 1\right), \tag{101}$$

which finishes the proof of the corollary. □

3.1 Description of Cross-Validation

In this section, we describe the cross-validation used in penalized nuclear norm minimization (4) in the numerical comparison in Sections 4 and 5.

First, we construct a grid T of non-negative numbers based on a pre-selected positive integer N. Denote

$$t_{\max}^{PN} = \left\| \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix} \right\|,$$

i.e. the largest singular value of the observed blocks. For penalized nuclear norm minimization, we let $T = \left\{t_{\max}^{PN}, t_{\max}^{PN} \cdot 10^{-3(1/N)}, \cdots, t_{\max}^{PN} \cdot 10^{-3(N/N)}\right\}$.

Next, for a given positive integer K, we randomly divide the integer set $\{1, \cdots, m_1\}$ into two groups of size $m^{(1)} \approx \frac{(K-1)n}{K}$, $m^{(2)} \approx \frac{n}{K}$ for H times. For $h = 1, \cdots, H$, we denote by J_1^h and $J_2^h \subseteq \{1, 2, \cdots, m_1\}$ the index sets of the two groups for the h-th split. Then the penalized nuclear norm minimization estimator (4) is applied to the first group of data: $A_{11}, A_{21}, (A_{12})_{[J_1^h,:]}$, i.e. the data of the observation set $\Omega = \{(i,j) : 1 \leq j \leq m_2, \text{ or } i \in J_1^h, m_2 + 1 \leq j \leq p_2\}$, with each value of the tuning parameter $t \in T$ and denote the result by $\hat{A}_h^{PN}(t)$. Note that we did not use the observed block $A_{[J_2^h,(m_2+1):p_2]}$ in calculating $\hat{A}_h^{PN}(t)$. Instead, $A_{[J_2^h,(m_2+1):p_2]}$ is used to evaluate the performance of the tunning parameter $t \in T$.

Set

$$\hat{R}(t) = \frac{1}{H} \sum_{h=1}^{H} \left\| \left[\hat{A}_h^{PN}(t) \right]_{[J_2^h,(m_2+1):p_2]} - A_{[J_2^h,(m_2+1):p_2]} \right\|_F^2 . \tag{102}$$

Finally, the tuning parameter is chosen as

$$t_* = \arg\min_{t \in T} \hat{R}(t)$$

and the final estimator $\hat{A}^{PN}$ is calculated using this choice of the tuning parameter t_*.

In all the numerical studies with penalized nuclear norm minimization in Sections 4 and 5, we use 5-cross-validation (i.e., $K = 5$), $N = 10$ to select the tuning parameter.