SUPPLEMENT TO "MATRIX COMPLETION VIA RESIDUAL SPECTRAL MATCHING"

BY ZIYUAN CHEN[1,a], FANG YAO[1,b]

[1]School of Mathematical Sciences, Center for Statistical Science, Peking University,
[a]chenziyuan@pku.edu.cn; [b]fyao@math.pku.edu.cn

S.1. Proof of Lemma 4.1. We first present several lemmas that are instrumental in our proofs.

LEMMA S.1. *With probability larger than* $1 - cn^{-3}$,

$$\left|\frac{\widehat{\sigma}(\boldsymbol{M}_0)}{\sigma} - 1\right| \lesssim \log^{-1} m.$$

LEMMA S.2. *(Brailovskaya and van Handel (2024) Theorem 2.7) Assume that* $\boldsymbol{X} = \sum_{i=1}^{N} \boldsymbol{Z}_i$ *is sum of* N *independent random matrices* $\boldsymbol{Z}_i \in \mathbb{R}^{d\times d}$ *with zero mean and* $\mathbb{E}\|\boldsymbol{Z}_i\|^2 = 1$. *Define* $\boldsymbol{G} = \sum_{i=1}^{N} \boldsymbol{A}_i g_i$ *where* $\boldsymbol{A}_i$ *are deterministic matrices and* g_i *are i.i.d. standard Gaussian variables. Define that*

$$\sigma(\boldsymbol{X}) = \lim_{q\to\infty} \left(\operatorname{Tr}\mathbb{E}[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})^2]^{q/2}/d\right)^{1/q},$$

$$R(\boldsymbol{X}) = \lim_{q\to\infty} \left(\sum_{i=1}^{N} \mathbb{E}[\operatorname{Tr}(\boldsymbol{Z}_i'\boldsymbol{Z}_i)^{q/2}/d]\right)^{1/q},$$

$$\bar{R}(\boldsymbol{X}) = \mathbb{E}\left[\max_{1\le i\le N} \|\boldsymbol{Z}_i\|^2\right]^{1/2},$$

$$\sigma_*(\boldsymbol{X}) = \sup_{\|\boldsymbol{u}\|=\|\boldsymbol{v}\|=1} \mathbb{E}(\langle \boldsymbol{u}, (\boldsymbol{X} - \mathbb{E}\boldsymbol{X})\boldsymbol{v}\rangle^2)^{1/2}$$

We have

$$P\left(d_H(sp(\boldsymbol{X}), sp(\boldsymbol{G})) > C\varepsilon_R(t), \max_{1\le i\le n} \|\boldsymbol{Z}_i\| \le R\right) \le de^{-t}$$

for all $t \ge 0$ *and* $R \ge \bar{R}(\boldsymbol{X})^{1/2}\sigma(\boldsymbol{X})^{1/2} + \sqrt{2}\bar{R}(\boldsymbol{X})$, *where*

$$\varepsilon_R(t) = \sigma_*(\boldsymbol{X})t^{1/2} + R^{1/3}\sigma(\boldsymbol{X})^{2/3}t^{2/3} + Rt$$

and C *is a universal constant.*

LEMMA S.3. *Define the singular values of an arbitrary i.i.d. standard Gaussian random matrix* $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ *as* $\tau_1 \ge \tau_2 \ge \cdots \ge \tau_n$, *then with probability larger than* $1 - cn^{-3}$,

$$\sup_{1\le i\le n} |p^{-1/2}m^{1/2}\widehat{\lambda}_i - \tau_i| \lesssim m^{1/2}\log^{-1} m.$$

LEMMA S.4. *Define the singular values of* $P_\Omega(\boldsymbol{H})/\sqrt{m}\sigma$ *as* $\theta_1 \ge \theta_2 \ge \cdots \ge \theta_n$, *then with probability larger than* $1 - cn^{-3}$,

$$\sup_{1\le i\le n} |\theta_i - \widehat{\lambda}_i| \lesssim p^{1/2}\log^{-1} m.$$

Using these lemmas, we now provide the proof of Lemma 4.1.

PROOF OF LEMMA 4.1. Define the singular values of $P_\Omega(\boldsymbol{H})/\sqrt{m}\sigma$ as $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_n$. By substituting $\boldsymbol{M} = \boldsymbol{M}_0$, we know that

$$\mathcal{L}(\boldsymbol{M}_0; \boldsymbol{Y}, \boldsymbol{\Omega}) = p^{-1} l \left(\frac{\boldsymbol{Y} - P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0)}{\sqrt{m}\widehat{\sigma}(\boldsymbol{M}_0)}; \boldsymbol{\omega} \right) = p^{-1} \sum_{i=1}^{n} \omega_i \left(\frac{\sigma}{\widehat{\sigma}(\boldsymbol{M}_0)} \theta_i - \widehat{\lambda}_i \right)^2.$$

From Lemma S.1, we know that $|\sigma/\widehat{\sigma}(\boldsymbol{M}_0) - 1| \lesssim \log^{-1} m$ holds with probability larger than $1 - cn^{-3}$. Lemma S.4 shows the universality of the empirical spectral distribution, which means that we can approach it by the empirical spectral distributions of sparse Gaussian random matrices, i.e., with probability larger than $1 - cn^{-3}$,

$$\sup_{1 \leq i \leq n} |\theta_i - \widehat{\lambda}_i| \lesssim p^{1/2} \log^{-1} m.$$

Apply Lemma S.3, Lemma S.4 and the local convergence to the Marchenko-Panstur law for sample covariance matrices (e.g. Lemma 8.1 in Erdős et al. (2012); Theorem 1 in Cacciapuoti, Maltsev and Schlein (2013)), we have with probability larger than $1 - cn^{-3}$,

$$\begin{aligned} \theta_1 &\leq |\theta_1 - \widehat{\lambda}_1| + p^{1/2} m^{-1/2} |p^{-1/2} m^{1/2} \widehat{\lambda}_1 - \tau_1| + p^{1/2} |m^{-1/2} \tau_1 - (1 + \sqrt{\rho})| \\ &\leq p^{1/2} (1 + \sqrt{\rho} + C \log^{-1} m). \end{aligned} \tag{1}$$

Thus

$$\begin{aligned} \mathcal{L}(\boldsymbol{M}_0; \boldsymbol{Y}, \boldsymbol{\Omega}) &\leq p^{-1} \sum_{i=1}^{n} \omega_i (\theta_i - \widehat{\lambda}_i)^2 + p^{-1} \sum_{i=1}^{n} \omega_i \left(\frac{\sigma}{\widehat{\sigma}(\boldsymbol{M}_0)} - 1 \right)^2 \theta_i^2 \\ &\leq p^{-1} \sup_{1 \leq i \leq n} (\theta_i - \widehat{\lambda}_i)^2 + p^{-1} \theta_1^2 \left(\frac{\sigma}{\widehat{\sigma}(\boldsymbol{M}_0)} - 1 \right)^2 \lesssim \log^{-2} m. \end{aligned}$$

□

PROOF OF LEMMA S.1. Recall the definition of $\widehat{\sigma}(\boldsymbol{M})$, it holds that

$$\widehat{\sigma}(\boldsymbol{M}_0) = \sigma \frac{\sum_{i=n/3}^{2n/3} \theta_i}{\sum_{i=n/3}^{2n/3} \widehat{\lambda}_i} = \sigma \left(1 + \frac{\sum_{i=n/3}^{2n/3} (\theta_i - \widehat{\lambda}_i)}{\sum_{i=n/3}^{2n/3} \widehat{\lambda}_i} \right).$$

From Lemma S.4, with probability larger than $1 - cn^{-3}$,

$$\sup_{1 \leq i \leq n} |\theta_i - \widehat{\lambda}_i| \lesssim p^{1/2} \log^{-1} m,$$

and thus

$$\left| \sum_{i=n/3}^{2n/3} (\theta_i - \widehat{\lambda}_i) \right| \lesssim \frac{n}{3} p^{1/2} \log^{-1} m.$$

From Lemma S.3, with probability larger than $1 - cn^{-3}$,

$$\sup_{1 \leq i \leq n} |p^{-1/2} m^{1/2} \widehat{\lambda}_i - \tau_i| \lesssim m^{1/2} \log^{-1} m.$$

Define ν_i as the i/nth quantile for the rooted Marchenko-Pastur distribution

$$\nu(x) = \frac{1}{\pi} \frac{\sqrt{(\lambda_+^2 - x^2)(x^2 - \lambda_-^2)}}{\rho x} \mathbf{1}_{x \in [\lambda_-, \lambda_+]} \quad \lambda_\pm = 1 \pm \sqrt{\rho}.$$

Properties about partial linear eigenvalue statistics for sample covariance matrices (e.g. Theorem 1.1 in Bao, Pan and Zhou (2013); Theorem 13 in O'Rourke and Soshnikov (2015)) show that there exists some constant $C \ll 1$ with probability larger than $1 - cn^{-3}$,

$$\left| m^{-1/2} \sum_{i=n/3}^{2n/3} \tau_i - \sum_{i=n/3}^{2n/3} \nu_i \right| \le C \int_{\nu_{n/3}}^{\nu_{2n/3}} x\nu(x)\, dx.$$

So

$$\sum_{i=n/3}^{2n/3} \widehat{\lambda}_i \ge p^{1/2} m^{-1/2} \sum_{i=n/3}^{2n/3} \tau_i - Cnp^{1/2} \log^{-1} m \gtrsim p^{1/2} n. \tag{2}$$

Thus with probability larger than $1 - cn^{-3}$

$$\left| \frac{\widehat{\sigma}(\boldsymbol{M}_0)}{\sigma} - 1 \right| \le \left| \sum_{i=n/3}^{2n/3} (\theta_i - \widehat{\lambda}_i) \right| \left(\sum_{i=n/3}^{2n/3} \widehat{\lambda}_i \right)^{-1} \lesssim \log^{-1} m.$$

$\square$

PROOF OF LEMMA S.3. For any $\boldsymbol{M} \in \mathbb{R}^{m \times n}$, consider the square matrix

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{0}_{m\times m} & p^{-1/2} m^{1/2} \boldsymbol{M} \\ p^{-1/2} m^{1/2} \boldsymbol{M}' & \boldsymbol{0}_{n \times n} \end{pmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}.$$

We can regard $\boldsymbol{X}$ as

$$\boldsymbol{X} = \sum_{1\le i \le m, 1 \le j \le n} \boldsymbol{Z}_{i,j+m} + \boldsymbol{Z}_{j+m,i},$$

where

$$\boldsymbol{Z}_{i,j+m} = p^{-1/2} m^{1/2} \boldsymbol{M}_{i,j} \boldsymbol{e}_i \boldsymbol{e}_{j+m}', \ \boldsymbol{Z}_{j+m,i} = p^{-1/2} m^{1/2} \boldsymbol{M}_{i,j} \boldsymbol{e}_{j+m} \boldsymbol{e}_i'.$$

Then it can be calculated that $\sigma(\boldsymbol{X}) = \sqrt{m}$, $\sigma_*(\boldsymbol{X}) = 1$ and $\bar{R}(\boldsymbol{X}) \le p^{-1/2} \log mnp$. We choose

$$R = \bar{R}(\boldsymbol{X})^{1/2} \sigma(\boldsymbol{X})^{1/2} + \sqrt{2} \bar{R}(\boldsymbol{X})^{1/2} \le p^{-1/4} m^{1/4} \log^{1/2} mnp + \sqrt{2} p^{-1/2} \log mnp.$$

And

$$\begin{aligned}
\varepsilon_R(t) &= \sigma_*(\boldsymbol{X}) t^{1/2} + R^{1/3} \sigma(\boldsymbol{X})^{2/3} t^{2/3} + Rt \\
&\le t^{1/2} + p^{-1/12} m^{5/12} \log^{1/6}(mnp) t^{2/3} + 2^{1/6} p^{-1/6} m^{1/3} \log^{1/3}(mnp) t^{2/3} \\
&+ p^{-1/4} m^{1/4} \log^{1/2}(mnp) t + \sqrt{2} p^{-1/2} \log(mnp) t.
\end{aligned}$$

The spectral of $\boldsymbol{X}$ is

$$\mathrm{sp}(\boldsymbol{X}) = p^{-1/2} m^{1/2} (\lambda(\boldsymbol{M}_1), \cdots, \lambda(\boldsymbol{M}_n), 0, \cdots, 0, -\lambda(\boldsymbol{M}_n), \cdots, -\lambda(\boldsymbol{M}_1)).$$

From Lemma S.2, we know that with probability larger than $1 - (m+n)e^{-t}$,

$$\sup_{1 \le i \le n} |p^{-1/2} m^{1/2} \lambda(\boldsymbol{M}_i) - \tau_i| \lesssim \varepsilon_R(t).$$

Chosen $t = \log^4 n$, so

$$\varepsilon_R(t) \lesssim m^{1/2} \log^{-\alpha/12 + 17/6} m \lesssim C m^{1/2} \log^{-1} m,$$

using the assumption that $p \geq cm^{-1}\mu_0 r \log^{\alpha} m$ with large enough α. Then with probability larger than $1 - cn^{-3}$,

$$\sup_{1\leq i\leq n} |p^{-1/2}m^{1/2}\lambda(\boldsymbol{M}_i) - \tau_i| \lesssim m^{1/2}\log^{-1} m.$$

And then

$$\sup_{1\leq n} |p^{-1/2}m^{1/2}\widehat{\lambda}_i - \tau_i| \leq \frac{1}{l}\sum_{k=1}^{l}\sup_{1\leq n} |p^{-1/2}m^{1/2}\lambda(\boldsymbol{M}_i) - \tau_i| \lesssim m^{1/2}\log^{-1} m.$$

□

PROOF OF LEMMA S.4. Consider the square matrix

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{0}_{m\times m} & p^{-1/2}P_{\Omega}(\boldsymbol{H})/\sigma \\ p^{-1/2}(P_{\Omega}(\boldsymbol{H}))'/\sigma & \boldsymbol{0}_{n\times n} \end{pmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}.$$

We can regard $\boldsymbol{X}$ as

$$\boldsymbol{X} = \sum_{1\leq i\leq m, 1\leq j\leq n} \boldsymbol{Z}_{i,j+m} + \boldsymbol{Z}_{j+m,i},$$

with

$$\boldsymbol{Z}_{i,j+m} = p^{-1/2}\sigma^{-1}(P_{\Omega}(\boldsymbol{H}))_{i,j}\boldsymbol{e}_i\boldsymbol{e}'_{j+m},\ \boldsymbol{Z}_{j+m,i} = p^{-1/2}\sigma^{-1}(P_{\Omega}(\boldsymbol{H}))_{i,j}\boldsymbol{e}_{j+m}\boldsymbol{e}'_i.$$

Then it can be calculated that $\sigma(\boldsymbol{X}) = \sqrt{m},\ \sigma_*(\boldsymbol{X}) = 1$. Recall that for N i.i.d. exponential distribution $E_i, 1 \leq i \leq N$, we have $\mathbb{E}\max_{1\leq i\leq N} E_i^2 \leq (\log N + 1)^2 + \pi^2/6$. With the assumption that $h_{i,j}$ are independent sub-exponential random variables, so $\bar{R}(\boldsymbol{X}) \leq cp^{-1/2}\log mnp$. Then we choose

$$R = \bar{R}(\boldsymbol{X})^{1/2}\sigma(\boldsymbol{X})^{1/2} + \sqrt{2}\bar{R}(\boldsymbol{X})^{1/2} \leq c^{1/2}p^{-1/4}m^{1/4}\log^{1/2} mnp + \sqrt{2}cp^{-1/2}\log mnp.$$

And

$$\begin{aligned}\varepsilon_R(t) &= \sigma_*(\boldsymbol{X})t^{1/2} + R^{1/3}\sigma(\boldsymbol{X})^{2/3}t^{2/3} + Rt \\ &\leq t^{1/2} + c^{1/6}p^{-1/12}m^{5/12}\log^{1/6}(mnp)t^{2/3} + 2^{1/6}c^{1/3}p^{-1/6}m^{1/3}\log^{1/3}(mnp)t^{2/3} \\ &\quad + c^{1/2}p^{-1/4}m^{1/4}\log^{1/2}(mnp)t + \sqrt{2}cp^{-1/2}\log(mnp)t.\end{aligned}$$

Define the singular value of $P_{\Omega}(\boldsymbol{H})/\sqrt{m}\sigma$ as $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_n$ and the singular value of an arbitrary i.i.d. standard Gaussian random matrix as $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_n$. Then the spectral of $\boldsymbol{X}$ is $\mathrm{sp}(\boldsymbol{X}) = \sqrt{m}(\theta_1, \cdots, \theta_n, 0, \cdots, 0, -\theta_n, \cdots, -\theta_1)$. From Lemma S.2, we know that with probability larger than $1 - (m+n)e^{-t}$,

$$\sup_{1\leq i\leq n} |p^{-1/2}m^{1/2}\theta_i - \tau_i| \lesssim \varepsilon_R(t).$$

Chosen $t = \log^4 n$, so

$$\varepsilon_R(t) \lesssim m^{1/2}\log^{-\alpha/12+17/6} m \lesssim m^{1/2}\log^{-1} m,$$

using the assumption that $p \geq cm^{-1}\mu_0 r \log^{\alpha} m$ with large enough α. Then with probability larger than $1 - cn^{-3}$,

$$\sup_{1\leq i\leq n} |p^{-1/2}m^{1/2}\theta_i - \tau_i| \lesssim m^{1/2}\log^{-1} m.$$

With Lemma S.3 that

$$\sup_{1\le i\le n} |p^{-1/2} m^{1/2}\widehat{\lambda}_i - \tau_i| \lesssim m^{1/2}\log^{-1} m,$$

so

$$\sup_{1\le i\le n} |\theta_i - \widehat{\lambda}_i| \lesssim p^{1/2}\log^{-1} m.$$

□

S.2. Proof of Lemma 4.2. Define the singular value decomposition $\boldsymbol{M}_0 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' = \sum_{i=1}^r \sigma_i \boldsymbol{u}_i\boldsymbol{v}_i'$ and $\boldsymbol{M} = \boldsymbol{X}\boldsymbol{\Theta}\boldsymbol{Y}' = \sum_{i=1}^r \theta_i \boldsymbol{x}_i\boldsymbol{y}_i'$. There exist matrices $\widetilde{\boldsymbol{U}}, \widetilde{\boldsymbol{X}} \in \mathbb{R}^{m\times r}, \widetilde{\boldsymbol{V}}, \widetilde{\boldsymbol{Y}} \in \mathbb{R}^{n\times r}$ and coefficient matrices $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\Gamma}_3, \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\Lambda}_3 \in \mathbb{R}^{r\times r}$ such that

$$\begin{aligned}
&\boldsymbol{U}'\widetilde{\boldsymbol{U}} = \boldsymbol{V}'\widetilde{\boldsymbol{V}} = \boldsymbol{X}'\widetilde{\boldsymbol{X}} = \boldsymbol{Y}'\widetilde{\boldsymbol{Y}} = 0, \widetilde{\boldsymbol{U}}'\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{V}}' = \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{Y}}\widetilde{\boldsymbol{Y}}' = \boldsymbol{I}_r,\\
&\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2'\boldsymbol{\Gamma}_2 = \boldsymbol{\Lambda}_1'\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2'\boldsymbol{\Lambda}_2 = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1' + \boldsymbol{\Gamma}_3'\boldsymbol{\Gamma}_3 = \boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1' + \boldsymbol{\Lambda}_3'\boldsymbol{\Lambda}_3 = \boldsymbol{I}_r,\\
&\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Gamma}_1 + \widetilde{\boldsymbol{U}}\boldsymbol{\Gamma}_2, \boldsymbol{Y} = \widetilde{\boldsymbol{U}}\boldsymbol{\Lambda}_1 + \widetilde{\boldsymbol{V}}\boldsymbol{\Lambda}_2, \boldsymbol{U} = \boldsymbol{X}\boldsymbol{\Gamma}_1' + \widetilde{\boldsymbol{X}}\boldsymbol{\Gamma}_3, \boldsymbol{V} = \boldsymbol{Y}\boldsymbol{\Lambda}_1' + \widetilde{\boldsymbol{Y}}\boldsymbol{\Lambda}_3.
\end{aligned}$$

Also define the singular value decomposition $\boldsymbol{M}_0 - \boldsymbol{M} = \sum_{i=1}^{r'} \lambda_i \boldsymbol{p}_i\boldsymbol{q}_i'$, $r' \le 2r$. Then we induce some notations that

$$\begin{aligned}
w_{ij} &= \boldsymbol{u}_i'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{v}_j, r_{ij} = \boldsymbol{x}_i'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{y}_j\\
\zeta_{ij} &= \boldsymbol{x}_i'(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{y}}_j, \chi_{ij} = \widetilde{\boldsymbol{x}}_i'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{y}_j.
\end{aligned}$$

Here i, j index weights vector $i = (i_1, \cdots, i_r), j = (i_1, \cdots, j_r)$ and $\|i\|_2 = \|j\|_2 = 1$. For example, when $i = j = (r^{-1/2}, r^{-1/2}, \cdots, r^{-1/2})$, then

$$\omega_{i,j} = \left(\sum_{k=1}^r r^{-1/2}\boldsymbol{u}_k\right)' (\boldsymbol{M}_0 - \boldsymbol{M}) \left(\sum_{k=1}^r r^{-1/2}\boldsymbol{v}_k\right).$$

Some useful lemmas are presented here to proof Lemma 4.2.

LEMMA S.5. *If it holds that*

$$\max_{i,j} |\omega_{i,j}|, |r_{i,j}| \ge ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp},$$

then with probability larger than $1 - cn^{-3}$,

$$\mathcal{L}(\boldsymbol{M}; \boldsymbol{Y}, \boldsymbol{\Omega}) \ge \mathcal{C}\omega_1.$$

LEMMA S.6. *If it holds that*

$$\max_{i,j} |\zeta_{i,j}|, |\chi_{i,j}| \ge ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp},$$

then with probability larger than $1 - cn^{-3}$,

$$\mathcal{L}(\boldsymbol{M}; \boldsymbol{Y}, \boldsymbol{\Omega}) \ge \mathcal{C}\omega_1.$$

LEMMA S.7. *If there exists some* $\varepsilon \ge 0$ *that*

$$\max_{i,j} |\omega_{i,j}|, |r_{i,j}|, |\zeta_{i,j}|, |\chi_{i,j}| \le \varepsilon,$$

then

$$\|\boldsymbol{M}_0 - \boldsymbol{M}\|_F \le (3+5\kappa) r^{1/2}\varepsilon.$$

LEMMA S.8. *For $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ and $\operatorname{rank}(\boldsymbol{M}) = s < n/3$, then with probability larger than $1 - cn^{-3}$,*

$$\left|\frac{\widehat{\sigma}(\boldsymbol{M})}{\sigma} - 1\right| \le C_1 \log^{-1} m + C_2\sigma^{-1}(mnp)^{-1/2}\|P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\|_F.$$

LEMMA S.9. *With probability larger than $1 - \exp\{-c\mu_0 rn\}$, for any $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ and $r \le \operatorname{rank}(\boldsymbol{M}) = s \lesssim mp\log^{-2} m$,*

$$\left|\|p^{-1/2}P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\|_F - \|\boldsymbol{M}_0 - \boldsymbol{M}\|_F\right| \le \log^{-\alpha/2} m\|\boldsymbol{M}_0 - M\|_F.$$

Given those lemmas, the proof of Lemma 4.2 can be derived.

PROOF OF LEMMA 4.2. Define that

$$\varepsilon = ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp}. \tag{3}$$

From Lemma S.8 and S.9 that with probability larger than $1 - cn^{-3}$,

$$\widehat{\sigma} \le (1 + C_1\log^{-1} m)\sigma + C_2(1 + C_3\log^{-1} m)(mnp)^{-1/2}p^{1/2}\|\boldsymbol{M}_0 - \boldsymbol{M}\|_F.$$

For those $\boldsymbol{M} \in \mathbb{M}_{r,\mu_1}$ that there exist large enough $C > 0$,

$$\|\boldsymbol{M}\|_F \ge C\|\boldsymbol{M}_0\|_F,$$

we have

$$\begin{aligned}\|\boldsymbol{M}_0 - \boldsymbol{M}\|_F &\ge (1 - C^{-1})\|\boldsymbol{M}\|_F \\ &\ge C(\sigma\sqrt{mr/p} + rm^{-1/2}\log^{-\alpha/2+1}(m)\|\boldsymbol{M}\|_F) \ge (3+5\kappa)r^{1/2}\varepsilon.\end{aligned}$$

If there exists constant $C > 0$,

$$\|\boldsymbol{M}\|_F \le C\|\boldsymbol{M}_0\|_F,$$

and

$$\|\boldsymbol{M}_0 - \boldsymbol{M}\|_F - (3+5\kappa)(2+2\sqrt{\rho}+\mathcal{C}^{1/2})\sigma\sqrt{mr/p} \gtrsim \log^{-1}(m)\sigma\sqrt{mr/p}.$$

With assumption that

$$(mn)^{-1/2}\|\boldsymbol{M}_0\|_F \lesssim \sigma r^{-1/2}m^{1/2}\log^{-1/2} n,$$

we have

$$\begin{aligned}\|\boldsymbol{M}_0 - \boldsymbol{M}\|_F \ge\ & (3+5\kappa)r^{1/2}(2+2\sqrt{\rho}+\mathcal{C}^{1/2}+C_1\log^{-1} m)\sigma\sqrt{m/p} \\ & + 8\sqrt{(1+C)r^2(\mu_0^2+\mu_1^2)\log n/mnp}\|\boldsymbol{M}_0\|_F \\ & \ge (3+5\kappa)r^{1/2}\varepsilon.\end{aligned}$$

From Lemma S.7, it must holds that

$$\max_{i,j} |\omega_{i,j}|, |r_{i,j}|, |\zeta_{i,j}|, |\chi_{i,j}| \ge \varepsilon.$$

Then from Lemma S.5 and S.6, we know that with probability larger than $1 - cn^{-3}$ that

$$\mathcal{L}(\boldsymbol{M}; \boldsymbol{Y}, \boldsymbol{\Omega}) \ge \mathcal{C}\omega_1.$$

□

PROOF OF LEMMA S.5. If there exist some $l, h \in \mathbb{R}^r$ with $\|l\|_2 = \|h\|_2 = 1$ such that

$$|w_{l,h}| \geq ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp}.$$

Without loss of generality, we assume that $w_{l,h} \geq 0$. Then

$$\mathbb{E}\left[p^{-1}\boldsymbol{u}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{v}_h\right] = \sum_{i,j} \mathbb{E}(p^{-1}\boldsymbol{\Omega}_{ij})\boldsymbol{u}_{li}\left(\sum_{k=1}^r \sigma_k \boldsymbol{u}_{ki}\boldsymbol{v}_{kj} - \sum_{k=1}^r \theta_k \boldsymbol{x}_{ki}\boldsymbol{y}_{kj}\right)\boldsymbol{v}_{hj} = \omega_{l,h}.$$

And

$$\begin{aligned}
\operatorname{Var}\left[p^{-1}\boldsymbol{u}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{v}_h\right] &= \sum_{i,j} p^{-2}\operatorname{Var}(\boldsymbol{\Omega}_{ij})\left(\boldsymbol{u}_{li}\left(\sum_{k=1}^r \sigma_k \boldsymbol{u}_{ki}\boldsymbol{v}_{kj} - \sum_{k=1}^r \theta_k \boldsymbol{x}_{ki}\boldsymbol{y}_{kj}\right)\boldsymbol{v}_{hj}\right)^2 \\
&= \frac{1-p}{p}\sum_{i,j}\boldsymbol{u}_{li}^2\left(\sum_{s=1}^r \sigma_s \boldsymbol{u}_{si}\boldsymbol{v}_{sj} - \sum_{s=1}^r \theta_s \boldsymbol{x}_{si}\boldsymbol{y}_{sj}\right)^2 \boldsymbol{v}_{hj}^2 \\
&\leq \frac{1-p}{p}\sum_{i,j}\boldsymbol{u}_{li}^2\Bigg(\sum_{1\leq k,s\leq r}\sigma_k\sigma_s|\boldsymbol{u}_{ki}\boldsymbol{u}_{si}\boldsymbol{v}_{kj}\boldsymbol{v}_{sj}| \\
&\quad + \sum_{1\leq k,s\leq r}\theta_k\theta_s|\boldsymbol{x}_{ki}\boldsymbol{x}_{si}\boldsymbol{y}_{kj}\boldsymbol{y}_{sj}| + 2\sum_{1\leq k,s\leq r}\sigma_k\theta_s|\boldsymbol{u}_{ki}\boldsymbol{x}_{si}\boldsymbol{v}_{kj}\boldsymbol{y}_{sj}|\Bigg)\boldsymbol{v}_{hj}^2 \\
&\leq \frac{2r(1-p)}{p}(\sigma_1^2+\theta_1^2)\sum_{i,j}\boldsymbol{u}_{li}^2\left(\sum_{k=1}^r \boldsymbol{u}_{ki}^2\boldsymbol{v}_{kj}^2 + \boldsymbol{x}_{ki}^2\boldsymbol{y}_{kj}^2\right)\boldsymbol{v}_{hj}^2
\end{aligned}$$

Because that $\boldsymbol{M}_0$ is μ-incoherent and $\boldsymbol{M}$ is μ_1-incoherent, then

$$\|\boldsymbol{U}\|_{2,\infty} \leq \sqrt{\frac{\mu_0 r}{m}}, \quad \|\boldsymbol{V}\|_{2,\infty} \leq \sqrt{\frac{\mu_0 r}{n}}$$
$$\|\boldsymbol{X}\|_{2,\infty} \leq \sqrt{\frac{\mu_1 r}{m}}, \quad \|\boldsymbol{Y}\|_{2,\infty} \leq \sqrt{\frac{\mu_1 r}{n}}.$$

Then

$$\begin{aligned}
\operatorname{Var}\left[p^{-1}\boldsymbol{u}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{v}_h\right] &\leq \frac{2r(1-p)}{p}(\sigma_1^2+\theta_1^2)\sum_{i,j}\boldsymbol{u}_{li}^2\left(\sum_{k=1}^r \boldsymbol{u}_{ki}^2\sum_{k=1}^r \boldsymbol{v}_{kj}^2 + \sum_{k=1}^r \boldsymbol{x}_{ki}^2\sum_{k=1}^r \boldsymbol{y}_{kj}^2\right)\boldsymbol{v}_{hj}^2 \\
&\leq \frac{2r^3(\mu_0^2+\mu_1^2)(1-p)}{mnp}(\sigma_1^2+\theta_1^2)\sum_{i=1}^m \boldsymbol{u}_{li}^2\sum_{j=1}^n \boldsymbol{v}_{hj}^2 \\
&= \frac{2r^3(\mu_0^2+\mu_1^2)(1-p)}{mnp}(\sigma_1^2+\theta_1^2).
\end{aligned}$$

Consider that

$$\begin{aligned}
\mathbb{E}\exp\{\lambda p^{-1}\boldsymbol{u}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{v}_h\} &= \prod_{i,j}\mathbb{E}\exp\{\lambda p^{-1}\boldsymbol{u}_{li}P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0-\boldsymbol{M})_{ij}\boldsymbol{v}_{hj}\} \\
&= \prod_{i,j}\left(p\exp\{\lambda p^{-1}\boldsymbol{u}_{li}(\boldsymbol{M}_0-\boldsymbol{M})_{ij}\boldsymbol{v}_{hj}\} + (1-p)\right)
\end{aligned}$$

$$= \prod_{i,j} \left(1 + p\left(\exp\{\lambda p^{-1} \boldsymbol{u}_{li} (\boldsymbol{M}_0 - \boldsymbol{M})_{ij} \boldsymbol{v}_{hj}\} - 1\right)\right).$$

When $0 > \lambda \geq -mnp/\sqrt{8(\mu_0^2 + \mu_1^2)\mu_0^2 r^5 (\sigma_1^2 + \theta_1^2)}$, we know that

$$\begin{aligned}
\left(\lambda p^{-1} \boldsymbol{u}_{li} (\boldsymbol{M}_0 - \boldsymbol{M})_{ij} \boldsymbol{v}_{hj}\right)^2 &= \lambda^2 p^{-2} \boldsymbol{u}_{li}^2 \left(\sum_{s=1}^r \sigma_s \boldsymbol{u}_{si} \boldsymbol{v}_{sj} - \sum_{s=1}^r \theta_s \boldsymbol{x}_{si} \boldsymbol{y}_{sj}\right)^2 \boldsymbol{v}_{hj}^2 \\
&\leq 2r\lambda^2 p^{-2} (\sigma_1^2 + \theta_1^2) \boldsymbol{u}_{li}^2 \left(\sum_{k=1}^r \boldsymbol{u}_{ki}^2 \boldsymbol{v}_{kj}^2 + \boldsymbol{x}_{ki}^2 \boldsymbol{y}_{kj}^2\right) \boldsymbol{v}_{hj}^2 \\
&\leq 2r^5 (\mu_0^2 + \mu_1^2) \mu_0^2 \lambda^2 (\sigma_1^2 + \theta_1^2)(mnp)^{-2} \leq \frac{1}{4}.
\end{aligned}$$

Because $e^x \leq 1 + x + 2x^2$ when $x^2 \leq 1/4$, so

$$\begin{aligned}
&\mathbb{E} \exp\{\lambda p^{-1} \boldsymbol{u}_l' P_{\boldsymbol{\Omega}} (\boldsymbol{M}_0 - \boldsymbol{M}) \boldsymbol{v}_h\} \\
\leq &\prod_{i,j} \left(1 + p\left(\lambda p^{-1} \boldsymbol{u}_{li} (\boldsymbol{M}_0 - \boldsymbol{M})_{ij} \boldsymbol{v}_{hj} + 2\left(\lambda p^{-1} \boldsymbol{u}_{li} (\boldsymbol{M}_0 - \boldsymbol{M})_{ij} \boldsymbol{v}_{hj}\right)^2\right)\right) \\
\leq &\exp\left\{\lambda \sum_{i,j} \boldsymbol{u}_{li} (\boldsymbol{M}_0 - \boldsymbol{M})_{ij} \boldsymbol{v}_{hj} + 2\lambda^2 p^{-1} \sum_{i,j} \boldsymbol{u}_{li}^2 (\boldsymbol{M}_0 - \boldsymbol{M})_{ij}^2 \boldsymbol{v}_{hj}^2\right\} \\
= &\exp\left\{\lambda \mathbb{E}\left[p^{-1} \boldsymbol{u}_l' P_{\boldsymbol{\Omega}} (\boldsymbol{M}_0 - \boldsymbol{M}) \boldsymbol{v}_h\right] + 2\lambda^2 (1-p)^{-1} \operatorname{Var}\left[p^{-1} \boldsymbol{u}_l' P_{\boldsymbol{\Omega}} (\boldsymbol{M}_0 - \boldsymbol{M}) \boldsymbol{v}_h\right]\right\}.
\end{aligned}$$

Because that $w_{lh} \geq ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2 + \mu_1^2)(\sigma_1^2 + \theta_1^2)\log n/mnp}$, we know

$$\begin{aligned}
&P\left(p^{-1} \boldsymbol{u}_l' P_{\boldsymbol{\Omega}} (\boldsymbol{M}_0 - \boldsymbol{M}) \boldsymbol{v}_h \leq ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p}\right) \\
\leq &\exp\left\{\lambda(w_{l,h} - ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p}) + 2\lambda^2(1-p)^{-1} \operatorname{Var} p^{-1} u_l' P_\Omega (M_0 - M) v_h\right\} \\
\leq &\exp\left\{-8\lambda\sqrt{r^3(\mu_0^2 + \mu_1^2)(\sigma_1^2 + \theta_1^2)\log n/mnp} + \log n\right\} = n^{-3},
\end{aligned}$$

where chosen

$$\lambda = -\sqrt{\frac{mnp\log n}{4r^3(\mu_0^2 + \mu_1^2)(\sigma_1^2 + \theta_1^2)}} \geq -\frac{mnp}{\sqrt{8(\mu_0^2 + \mu_1^2)\mu_0^2(\sigma_1^2 + \theta_1^2)}},$$

because $p \geq m^{-1}\mu_0 r \log^\alpha m$.

Then

$$\begin{aligned}
&P\left(\boldsymbol{u}_l' P_{\boldsymbol{\Omega}} (\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H}) \boldsymbol{v}_h \geq (\mathcal{C}^{1/2} + 1 + \sqrt{\rho} + C\log^{-1} m)\widehat{\sigma}\sqrt{mp}\right) \\
\geq &1 - P\left(\boldsymbol{u}_l' P_{\boldsymbol{\Omega}} (\boldsymbol{M}_0 - \boldsymbol{M}) \boldsymbol{v}_h \leq \left((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma}\right)\sqrt{mp}\right) \\
&- P\left(\boldsymbol{u}_l' P_{\boldsymbol{\Omega}} (\boldsymbol{H}) \boldsymbol{v}_h \leq -\sigma\sqrt{mp}(1 + \sqrt{\rho} + C\log^{-1} m)\right) \\
\geq &1 - n^{-3} - P\left(\theta_1 \geq \sqrt{p}(1 + \sqrt{\rho} + C\log^{-1} m)\right).
\end{aligned}$$

From (1), we have

$$P\left(\theta_1 \geq \sqrt{p}(1 + \sqrt{\rho} + C\log^{-1} m)\right) \lesssim n^{-3}.$$

From Lemma S.3 and local convergence to the Marchenko-Panstur law for i.i.d. standard Gaussian random matrix, we have with probability larger than $1-cn^{-3}$

$$\widehat{\lambda}_1 \le |\widehat{\lambda}_1 - p^{1/2}m^{-1/2}\tau_1| + p^{1/2}|m^{-1/2}\tau_1 - (1+\sqrt{\rho})| \le p^{1/2}(1+\sqrt{\rho}+C\log^{-1} m).$$

So with probability larger than $1-cn^{-3}$,

$$\begin{aligned}\mathcal{L}(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega}) &\ge \frac{\omega_1}{\widehat{\sigma}^2 mp}\left(\|P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0-\boldsymbol{M}+\boldsymbol{H})\| - \widehat{\sigma}\sqrt{m}\widehat{\lambda}_1\right)^2\\ &\ge \frac{\omega_1}{\widehat{\sigma}^2 mp}\left(\widehat{\sigma}\sqrt{mp}(\mathcal{C}^{1/2}+1+\sqrt{\rho}+C\log^{-1} m) - \widehat{\sigma}\sqrt{mp}(1+\sqrt{\rho}+C\log^{-1} m)\right)^2\\ &\ge \mathcal{C}\omega_1.\end{aligned}$$

The proof is same if there exist some $l,h\in\mathbb{R}^r$ with $\|l\|_2=\|h\|_2=1$ such that

$$|r_{l,h}| \ge ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma)+\mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp}.$$

□

PROOF OF LEMMA S.6. The proof of Lemma S.6 is similar to the proof of Lemma S.5. The main different is that $\|\widetilde{\boldsymbol{x}}_i\|_\infty$ and $\|\widetilde{\boldsymbol{y}}_i\|_\infty$ do not have tight bound like $\sqrt{\mu_1 r}/\sqrt{m}$ and $\sqrt{\mu_1 r}/\sqrt{n}$ which holds for $\|\boldsymbol{x}_i\|_\infty$ and $\|\boldsymbol{y}_i\|_\infty$.

If there exist some $l,h\in\mathbb{R}^r$ with $\|l\|_2=\|h\|_2=1$ such that

$$|\chi_{l,h}| \ge ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma)+\mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp}.$$

Without loss of generality, we still assume that $\chi_{l,h}\ge 0$. Then

$$\mathbb{E}\left[p^{-1}\widetilde{\boldsymbol{x}}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{y}_h\right] = \chi_{l,h}$$

and

$$\begin{aligned}\mathrm{Var}\left[p^{-1}\widetilde{\boldsymbol{x}}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{y}_h\right] &\le \frac{2r(1-p)}{p}(\sigma_1^2+\theta_1^2)\sum_{i,j}\widetilde{\boldsymbol{x}}_{li}^2\left(\sum_{k=1}^r \boldsymbol{u}_{ki}^2\boldsymbol{v}_{kj}^2+\boldsymbol{x}_{ki}^2\boldsymbol{y}_{kj}^2\right)\boldsymbol{y}_{hj}^2\\ &\le \frac{2r^3(\mu_0^2+\mu_1^2)(1-p)}{mnp}(\sigma_1^2+\theta_1^2).\end{aligned}$$

When $0>\lambda\ge -m^{1/2}np/\sqrt{8(\mu_0^2+\mu_1^2)\mu_0 r^4(\sigma_1^2+\theta_1^2)}$, we know that

$$\begin{aligned}\left(\lambda p^{-1}\widetilde{\boldsymbol{x}}_{li}(\boldsymbol{M}_0-\boldsymbol{M})_{ij}\boldsymbol{y}_{hj}\right)^2 &\le 2r\lambda^2 p^{-2}(\sigma_1^2+\theta_1^2)\widetilde{\boldsymbol{x}}_{li}^2\left(\sum_{k=1}^r \boldsymbol{u}_{ki}^2\boldsymbol{v}_{kj}^2+\boldsymbol{x}_{ki}^2\boldsymbol{y}_{kj}^2\right)\boldsymbol{y}_{hj}^2\\ &\le 2r^4\lambda^2p^{-2}(\sigma_1^2+\theta_1^2)(\mu_0^2+\mu_1^2)\mu_0 m^{-1}n^{-2} \le \frac{1}{4}.\end{aligned}$$

where using $\widetilde{\boldsymbol{x}}_{li}^2\le 1$ instead of the proof of Lemma S.5 that $\boldsymbol{u}_{li}^2\le \mu_0 r/m$. Then

$$\begin{aligned}&\mathbb{E}\exp\{\lambda p^{-1}\widetilde{\boldsymbol{x}}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{y}_h\}\\ &\le \exp\left\{\lambda\mathbb{E}\left[p^{-1}\widetilde{\boldsymbol{x}}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{y}_h\right] + 2\lambda^2(1-p)^{-1}\mathrm{Var}\left[p^{-1}\widetilde{\boldsymbol{x}}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{y}_h\right]\right\}.\end{aligned}$$

Because $\chi_{l,h}\ge((1+\sqrt{\rho})(\widehat{\sigma}+\sigma)+\mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p}+8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp}$, we know that

$$P\left(p^{-1}\widetilde{\boldsymbol{x}}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{y}_h \le ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma)+\mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p}\right) \le n^{-3},$$

where chosen

$$\lambda = -\sqrt{\frac{mnp\log n}{4r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)}} \geq -m^{1/2}np/\sqrt{8(\mu_0^2+\mu_1^2)\mu_0 r^4(\sigma_1^2+\theta_1^2)},$$

because $p \geq m^{-1}\mu_0 r \log^\alpha m$. Then

$$P\left(\widetilde{\boldsymbol{x}}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H})\boldsymbol{y}_h \geq (\mathcal{C}^{1/2} + 1 + \sqrt{\rho} + C\log^{-1} m)\widehat{\sigma}\sqrt{mp}\right) \geq 1 - cn^{-3}.$$

So with probability larger than $1 - cn^{-3}$, $\mathcal{L}(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega}) \geq \mathcal{C}\omega_1$. The proof is same if there exist some $l, h \in \mathbb{R}^r$ with $\|l\|_2 = \|h\|_2 = 1$ such that

$$|\zeta_{l,h}| \geq ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp}.$$

□

PROOF OF LEMMA S.7. From the condition that $\max_{i,j}|r_{i,j}| \leq \varepsilon$, then

$$\|\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\| \leq \max_{i,j}|r_{i,j}| \leq \varepsilon.$$

With $\operatorname{rank} \boldsymbol{X}'(\boldsymbol{M}_0 - \widehat{\boldsymbol{M}})\boldsymbol{Y} \leq \operatorname{rank}(\boldsymbol{X}) = r$, so

$$\|\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\|_F \leq \sqrt{\operatorname{rank}(\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y})}\|\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\| \leq r^{1/2}\varepsilon.$$

Similarly we have

$$\begin{aligned}
\|\boldsymbol{U}'(M_0 - \boldsymbol{M})\boldsymbol{V}\|_F &\leq r^{1/2}\varepsilon, \\
\|\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{Y}}\|_F &\leq r^{1/2}\varepsilon, \\
\|\widetilde{\boldsymbol{X}}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\|_F &\leq r^{1/2}\varepsilon.
\end{aligned}$$

With

$$\begin{aligned}
\|\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\|_F &= \|\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\Gamma}_1' + \widetilde{\boldsymbol{X}}\boldsymbol{\Gamma}_3)\boldsymbol{\Sigma}(\boldsymbol{\Lambda}_1\boldsymbol{Y}' + \boldsymbol{\Lambda}_3'\widetilde{\boldsymbol{Y}})\boldsymbol{Y}' - \boldsymbol{\Theta}\|_F \\
&= \|\boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_1 - \boldsymbol{\Theta}\|_F
\end{aligned}$$

and

$$\begin{aligned}
\|\boldsymbol{U}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{V}\|_F &= \|\boldsymbol{\Theta} - \boldsymbol{U}'(\boldsymbol{U}\boldsymbol{\Gamma}_1 + \widetilde{\boldsymbol{U}}\boldsymbol{\Gamma}_2)\boldsymbol{\Theta}(\boldsymbol{\Lambda}_1'\boldsymbol{V}' + \boldsymbol{\Lambda}_2'\widetilde{\boldsymbol{V}})\boldsymbol{V}'\|_F \\
&= \|\boldsymbol{\Sigma} - \boldsymbol{\Gamma}_1\boldsymbol{\Theta}\boldsymbol{\Lambda}_1'\|_F,
\end{aligned}$$

then

$$\|\boldsymbol{\Sigma} - \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1'\|_F \leq 2r^{1/2}\varepsilon.$$

Similarly we have

$$\begin{aligned}
\|\boldsymbol{\Gamma}_1'\boldsymbol{\Sigma}\boldsymbol{\Lambda}_3'\|_F &= \|\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{Y}}\|_F \leq r^{1/2}\varepsilon \\
\|\boldsymbol{\Gamma}_3\boldsymbol{\Sigma}\boldsymbol{\Lambda}_1\|_F &= \|\widetilde{\boldsymbol{X}}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\|_F \leq r^{1/2}\varepsilon.
\end{aligned}$$

With

$$\begin{aligned}
\|\boldsymbol{\Gamma_1}\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{Y}}\boldsymbol{\Lambda}_3\|_2 &\leq \|\boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{Y}}\|_2 \leq \varepsilon \\
\|\boldsymbol{\Gamma}_3'\widetilde{\boldsymbol{X}}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\boldsymbol{\Lambda}_1'\|_2 &\leq \|\widetilde{\boldsymbol{X}}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\|_2 \leq \varepsilon,
\end{aligned}$$

and the rank of $\mathbf{\Gamma}_1 \boldsymbol{X}'(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{Y}}\mathbf{\Lambda}_3$ and $\mathbf{\Gamma}_3'\widetilde{\boldsymbol{X}}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{Y}\mathbf{\Lambda}_1'$ is not more than r, so

$$\|\mathbf{\Gamma}_1\mathbf{\Gamma}_1'\mathbf{\Sigma}\mathbf{\Lambda}_3'\mathbf{\Lambda}_3\|_F \le r^{1/2}\varepsilon, \|\mathbf{\Gamma}_3'\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_1\mathbf{\Lambda}_1'\|_F \le r^{1/2}\varepsilon.$$

Thus

$$\|\mathbf{\Gamma}_3'\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_3'\mathbf{\Lambda}_3\|_F \le \|\mathbf{\Sigma} - \mathbf{\Gamma}_1\mathbf{\Gamma}_1'\mathbf{\Sigma}\mathbf{\Lambda}_1\mathbf{\Lambda}_1'\|_F + \|\mathbf{\Gamma}_1\mathbf{\Gamma}_1'\mathbf{\Sigma}\mathbf{\Lambda}_3'\mathbf{\Lambda}_3\|_F + \|\mathbf{\Gamma}_3'\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_1\mathbf{\Lambda}_1'\|_F \le 4r^{1/2}\varepsilon.$$

Then

$$\|\mathbf{\Gamma}_3'\mathbf{\Gamma}_3\mathbf{\Sigma}\|_F \le \|\mathbf{\Gamma}_3'\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_1\mathbf{\Lambda}_1'\|_F + \|\mathbf{\Gamma}_3'\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_3\mathbf{\Lambda}_3'\|_F \le 5r^{1/2}\varepsilon$$
$$\|\mathbf{\Sigma}\mathbf{\Lambda}_3'\mathbf{\Lambda}_3\|_F \le \|\mathbf{\Gamma}_3'\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_3'\mathbf{\Lambda}_3\|_F + \|\mathbf{\Gamma}_1\mathbf{\Gamma}_1'\mathbf{\Sigma}\mathbf{\Lambda}_3'\mathbf{\Lambda}_3\|_F \le 5r^{1/2}\varepsilon,$$

which means that

$$\|\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_3'\|_F^2 \le \kappa^2\left(\frac{1}{2}\|\mathbf{\Gamma}_3'\mathbf{\Gamma}_3\mathbf{\Sigma}\|_F^2 + \frac{1}{2}\|\mathbf{\Sigma}\mathbf{\Lambda}_3'\mathbf{\Lambda}_3\|_F^2\right) \le 25\kappa^2 r\varepsilon^2.$$

Finally,

$$\begin{aligned}\|\boldsymbol{M}_0 - \boldsymbol{M}\|_F &= \|(\boldsymbol{X}', \widetilde{\boldsymbol{X}}')(\boldsymbol{M}_0 - \boldsymbol{M})(\boldsymbol{Y}', \widetilde{\boldsymbol{Y}}')'\|_F \\ &= \|\mathbf{\Gamma}_1'\mathbf{\Sigma}\mathbf{\Lambda}_1 - \mathbf{\Theta} + \mathbf{\Gamma}_1'\mathbf{\Sigma}\mathbf{\Lambda}_3' + \mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_1 + \mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_3'\|_F \\ &\le \|\mathbf{\Gamma}_1'\mathbf{\Sigma}\mathbf{\Lambda}_1 - \mathbf{\Theta}\|_F + \|\mathbf{\Gamma}_1'\mathbf{\Sigma}\mathbf{\Lambda}_3'\|_F + \|\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_1\|_F + \|\mathbf{\Gamma}_3\mathbf{\Sigma}\mathbf{\Lambda}_3'\|_F \\ &\le (3 + 5\kappa)r^{1/2}\varepsilon.\end{aligned}$$

□

PROOF OF LEMMA S.8. For two matrix $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ with ordered singular values $\sigma_i(\boldsymbol{A})$ and $\sigma_i(\boldsymbol{B}), 1 \le i \le n$, Thompson (1975) shows that for any positive integers $i_1 < i_2 < \cdots < i_l$,$j_1 < j_2 < \cdots < j_l$ and $k_t = i_t + j_t - t$, then

$$\sum_{t=1}^{m} \sigma_{k_t}(\boldsymbol{A} + \boldsymbol{B}) \le \sum_{t=1}^{m} \sigma_{i_t}(\boldsymbol{A}) + \sum_{t=1}^{m} \sigma_{j_t}(\boldsymbol{B}).$$

Then

$$\begin{aligned}\sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H})) &\le \sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{H})) + \sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})) \\ &\le \sigma\sqrt{m}\sum_{i=n/3}^{2n/3} \widehat{\lambda}_i + \sigma\sqrt{m}\sum_{i=n/3}^{2n/3} |\theta_i - \widehat{\lambda}_i| + \sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})),\end{aligned}$$

$$\begin{aligned}\sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H})) &\ge \sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{H})) - \sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})) \\ &\ge \sigma\sqrt{m}\sum_{i=n/3}^{2n/3} \widehat{\lambda}_i - \sigma\sqrt{m}\sum_{i=n/3}^{2n/3} |\theta_i - \widehat{\lambda}_i| - \sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})).\end{aligned}$$

In Lemma S.4 we proved that with probability larger than $1 - cn^{-3}$,

$$\sup_{1\le i\le n} |\theta_i - \widehat{\lambda}_i| \lesssim p^{1/2}\log^{-1} m.$$

So

$$\sum_{i=n/3}^{2n/3} |\theta_i - \widehat{\lambda}_i| \le \frac{n}{3} p^{1/2} \log^{-1} m.$$

From (2), with probability larger than $1 - cn^{-3}$.

$$\sum_{i=n/3}^{2n/3} \widehat{\lambda}_i \gtrsim p^{1/2} n.$$

Because that

$$\|P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\|_F^2 = \sum_{i=1}^{n} \sigma_i^2(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})) \ge \frac{n}{3}\sigma_{n/3}^2(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})),$$

we have

$$\sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})) \le \frac{n}{3}\sigma_{n/3}(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})) \le (n/3)^{1/2} \|P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\|_F.$$

So with probability larger than $1 - cn^{-3}$

$$\begin{aligned}\left|\frac{\widehat{\sigma}(\boldsymbol{M})}{\sigma} - 1\right| &\le \left(\sum_{i=n/3}^{2n/3} \widehat{\lambda}_i\right)^{-1} \left(\sum_{i=n/3}^{2n/3} |\theta_i - \widehat{\lambda}_i| + \sigma^{-1} m^{-1/2} \sum_{i=n/3}^{2n/3} \sigma_i(P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M}))\right) \\ &\le C_1 \log^{-1} m + C_2 \sigma^{-1}(mnp)^{-1/2} \|P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\|_F.\end{aligned}$$

□

PROOF OF LEMMA S.9. Define the low rank matrix class

$$\mathcal{S}_s = \left\{\boldsymbol{M} \in \mathbb{R}^{m\times n} \mid \operatorname{rank}(\boldsymbol{M}) = s, \|\boldsymbol{M}\|_F = 1\right\}.$$

Lemma 3.1 in Candès and Plan (2011) shows that there exists an ϵ-net $\bar{\mathcal{S}}_s$ for the Frobenius norm obeying

$$\left|\bar{\mathcal{S}}_s\right| \le (9/\epsilon)^{(m+n+1)s}.$$

Then for fixed $\boldsymbol{N} \in \mathcal{S}_s$,

$$\begin{aligned}&P\left(\sup_{N-\boldsymbol{M}\in\mathcal{S}_s} \left|\|p^{-1/2} P_{\mathbf{\Omega}}(\boldsymbol{N} - \boldsymbol{M})\|_F - \|\boldsymbol{N} - \boldsymbol{M}\|_F\right| \ge \epsilon/2\right) \\ &\le 2\left|\bar{\mathcal{S}}_s\right| \exp\left\{-cmnp\epsilon^2\right\} \\ &\le \exp\left\{-c\mu_0 r \log^{\alpha} m\epsilon^2 + (m+n+1)s\log(9/\epsilon)\right\}.\end{aligned}$$

With $p \ge m^{-1}\log^{\alpha} m$ for enough large α and $s \lesssim mp\log^{-2} m$, then

$$\begin{aligned}P\left(\sup_{\boldsymbol{M},\operatorname{rank}(\boldsymbol{M})\le s} \|\boldsymbol{M}_0 - M\|_F^{-1} \left|\|p^{-1/2} P_{\mathbf{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\|_F - \|\boldsymbol{M}_0 - \boldsymbol{M}\|_F\right| \ge \log^{-\alpha/2} m\right) \\ \le \exp\left\{-c\mu_0 rn\right\}.\end{aligned}$$

□

S.3. Proof of Theorem 4.1, Corollary 4.1 and Corollary 4.2.

PROOF OF THEOREM 4.1. Lemma 4.2 shows that for $M \in \mathbb{M}_{r,\mu_1}$, if $\mathcal{L}(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega}) \le \mathcal{C}\omega_1$, then with probability larger than $1 - cn^{-3}$,

$$\|\boldsymbol{M} - \boldsymbol{M}_0\|_F \le (3+5\kappa)(2+2\sqrt{\rho}+\mathcal{C}^{1/2})\sigma\sqrt{mr/p}.$$

It holds for $\widetilde{\boldsymbol{M}}$ because with the definition of $\widetilde{\boldsymbol{M}}$ and Lemma 4.1, for any constant $\mathcal{C} > C_0 \log^{-2} m\omega_1^{-1} = o(1)$, with probability larger than $1 - cn^{-3}$,

$$\mathcal{L}(\widetilde{\boldsymbol{M}};\boldsymbol{Y},\boldsymbol{\Omega}) \le \mathcal{L}(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega}) \le C_0 \log^{-2} m \ll \mathcal{C}\omega_1.$$

□

PROOF OF COROLLARY 4.1. Still define ε as

$$\begin{aligned}\varepsilon &= ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma)+\mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\theta_1^2)\log n/mnp} \\ &\le ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma)+\mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8(\mu_0+\mu_1)\frac{r\log^{1/2} n}{(mnp)^{1/2}}(\|\boldsymbol{M}_0\|_F + \|\boldsymbol{M}\|_F)\end{aligned}$$

We do the same procedure as in proof of Lemma 4.2. Now without (c) in Assumption 1, the second term in ε is not a small term anymore and we need to keep this term. Then following the proof of Theorem 4.1, we reach the result. □

PROOF OF COROLLARY 4.2. For $\boldsymbol{M} \in \mathbb{M}_{s,\mu_1}$, the results in Lemma S.5 and S.6 is similar. Follow the proof of Lemma S.6, we shows that

$$\begin{aligned}\mathrm{Var}\left[p^{-1}\widetilde{\boldsymbol{x}}_l' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{y}_h\right] &\le \frac{2s(1-p)}{p}(\sigma_1^2+\theta_1^2)\sum_{i,j}\widetilde{\boldsymbol{x}}_{li}^2\left(\sum_{k=1}^r \boldsymbol{u}_{ki}^2\boldsymbol{v}_{kj}^2 + \sum_{k=1}^s \boldsymbol{x}_{ki}^2\boldsymbol{y}_{kj}^2\right)\boldsymbol{y}_{hj}^2 \\ &\le \frac{2s(s^2\mu_1^2+r^2\mu_0^2)(1-p)}{mnp}(\sigma_1^2+\theta_1^2).\end{aligned}$$

And when $0 > \lambda \ge -m^{1/2}np/\sqrt{8(r^2\mu_0^2+s^2\mu_1^2)\mu_0 sr(\sigma_1^2+\theta_1^2)}$, we have

$$\begin{aligned}\left(\lambda p^{-1}\widetilde{\boldsymbol{x}}_{li}(\boldsymbol{M}_0-\boldsymbol{M})_{ij}\boldsymbol{y}_{hj}\right)^2 &\le 2s\lambda^2 p^{-2}(\sigma_1^2+\theta_1^2)\widetilde{\boldsymbol{x}}_{li}^2\left(\sum_{k=1}^r \boldsymbol{u}_{ki}^2\boldsymbol{v}_{kj}^2 + \boldsymbol{x}_{ki}^2\boldsymbol{y}_{kj}^2\right)\boldsymbol{y}_{hj}^2 \\ &\le 2sr\lambda^2 p^{-2}(\sigma_1^2+\theta_1^2)(r^2\mu_0^2+s^2\mu_1^2)\mu_0 m^{-1}n^{-2} \le \frac{1}{4}.\end{aligned}$$

And the chosen λ should satisfies

$$\lambda = -\sqrt{\frac{mnp\log n}{4s(r^2\mu_0^2+s^2\mu_1^2)(\sigma_1^2+\theta_1^2)}} \ge -\frac{m^{1/2}np}{\sqrt{8(r^2\mu_0^2+s^2\mu_1^2)\mu_0 sr(\sigma_1^2+\theta_1^2)}},$$

which is still hold because $p \gtrsim m^{-1}\mu_0 r\log^{\alpha} m$. So similar to Lemma S.5 and S.6, if

$$\begin{aligned}&\max_{i,j}|\omega_{i,j}|, |r_{i,j}|, |\zeta_{i,j}|, |\chi_{i,j}| \\ \ge&((1+\sqrt{\rho})(\widehat{\sigma}+\sigma)+\mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{\frac{s(s^2\mu_1^2+r^2\mu_0^2)\log n}{mnp}(\sigma_1^2+\theta_1^2)},\end{aligned}$$

then with probability larger than $1-cn^{-3}$, $\mathcal{L}(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega})\geq\mathcal{C}\omega_1$. Then consider the proof in Lemma S.7 with $\max_{i,j}|\omega_{i,j}|,|r_{i,j}|,|\zeta_{i,j}|,|\chi_{i,j}|\leq\varepsilon$. Because that $\mathrm{rank}(\boldsymbol{X}),\mathrm{rank}(\boldsymbol{Y})\leq s$, then

$$\|\boldsymbol{X}'(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{Y}\|_F\leq s^{1/2}\varepsilon.$$

and thus

$$\|\boldsymbol{\Gamma}_3'\boldsymbol{\Gamma}_3\boldsymbol{\Sigma}\boldsymbol{\Lambda}_3'\boldsymbol{\Lambda}\|_F\leq(3r^{1/2}+s^{1/2})\varepsilon,\ \|\boldsymbol{\Gamma}_3'\boldsymbol{\Gamma}_3\boldsymbol{\Sigma}\|_F,\|\boldsymbol{\Sigma}\boldsymbol{\Lambda}_3'\boldsymbol{\Lambda}_3\|_F\leq(4r^{1/2}+s^{1/2})\varepsilon.$$

Then $\|\boldsymbol{M}_0-\boldsymbol{M}\|_F\leq\left((2+4\kappa)r^{1/2}+(1+\kappa)s^{1/2}\right)\varepsilon$. Following the proof of Lemma 4.2, we know that when

$$\|\boldsymbol{M}_0-\boldsymbol{M}\|_F\geq\left((2+4\kappa)r^{1/2}+(1+\kappa)s^{1/2}\right)(2+2\sqrt{\rho}+\mathcal{C}^{1/2})\sigma\sqrt{m/p}$$

and it holds that

$$\sigma\gg\frac{s^{3/2}}{r}\frac{\|\boldsymbol{M}_0\|_F\log^{1/2}n}{mn^{1/2}},$$

the loss function valued at $\boldsymbol{M}$ is larger than $\mathcal{C}\omega_1$. Finally follow the proof of Theorem 4.1, with probability larger than $1-cn^{-3}$,

$$\|\widetilde{\boldsymbol{M}}_s-\boldsymbol{M}_0\|_F\leq\left((2+4\kappa)r^{1/2}+(1+\kappa)s^{1/2}\right)(2+2\sqrt{\rho}+o(1))\sigma\sqrt{m/p}.$$

□

S.4. Proof of Theorem 4.2 and Corollary 4.3.

PROOF OF THEOREM 4.2. Define that $\mathcal{L}^\lambda(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega})=\mathcal{L}(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega})+\lambda\|\boldsymbol{M}\|_*$. From Lemma 4.1, with probability larger than $1-cn^{-3}$,

$$\mathcal{L}^\lambda(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega})\leq\lambda\|\boldsymbol{M}_0\|_*+C\log^{-2}m.$$

If $\|\boldsymbol{M}\|_*\geq\|\boldsymbol{M}_0\|_*+(C_0+1)\lambda^{-1}\log^{-2}m$, we know that

$$\mathcal{L}^\lambda(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega})\geq\lambda\|\boldsymbol{M}\|_*\geq\lambda\|\boldsymbol{M}_0\|_*+(C_0+1)\log^{-2}m\geq\mathcal{L}^\lambda(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega})+\log^{-2}m.$$

Now consider the case $\|\boldsymbol{M}\|_*\leq\|\boldsymbol{M}_0\|_*+(C_0+1)\lambda^{-1}\log^{-2}m$. Using the similar notations as in proof of Lemma 4.2 but now M is not a r rank matrix anymore and we set the singular value decomposition of $\boldsymbol{M}$ as $\boldsymbol{M}=\sum_{i=1}^n\theta_i\boldsymbol{x}_i\boldsymbol{y}_i'$. Consider the matrix $\boldsymbol{U}'(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{V}$, we denote the singular values of it as $\tau_1,\cdots,\tau_r$. So

$$\|\boldsymbol{M}\|_*\geq\|\boldsymbol{U}'\boldsymbol{M}\boldsymbol{V}\|_*\geq\|\boldsymbol{U}'\boldsymbol{M}_0\boldsymbol{V}\|_*-\sum_{i=1}^r\tau_i=\|\boldsymbol{M}_0\|_*-\sum_{i=1}^r\tau_r.$$

Also we have

$$\begin{aligned}\sum_{i=1}^r\theta_i\geq\sum_{i=1}^r\boldsymbol{u}_i'\boldsymbol{M}\boldsymbol{v}_i'&\geq\sum_{i=1}^r\left(\boldsymbol{u}_i'\boldsymbol{M}_0\boldsymbol{v}_i-|\boldsymbol{u}_i'(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{v}_i|\right)\\&\geq\|\boldsymbol{M}_0\|_*-\sum_{i=1}^r\tau_i,\end{aligned}$$

and

$$\sum_{i=r+1}^n\theta_i\leq\sum_{i=1}^n\theta_i-\sum_{i=1}^r\theta_i\leq(C_0+1)\lambda^{-1}\log^{-2}m+\sum_{i=1}^r\tau_i.$$

Define ε as

$$\varepsilon = ((1+\sqrt{\rho})(\widehat{\sigma}+\sigma) + \mathcal{C}^{1/2}\widehat{\sigma})\sqrt{m/p} + 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\widetilde{\theta}_1^2)\log n/mnp}$$

and the second term defined as $\varepsilon_2 = 8\sqrt{r^3(\mu_0^2+\mu_1^2)(\sigma_1^2+\widetilde{\theta}_1^2)\log n/mnp}$, where $\widetilde{\theta}_i, 1 \le i \le r$ are the singular values of $\boldsymbol{U}'\boldsymbol{M}\boldsymbol{V}$. We denote $r' = \inf_{k\le r} \tau_{k+1} \le \varepsilon$. If $r' \ge 1$, there exist r' orthogonal vectors $\widetilde{\boldsymbol{u}}_1, \cdots, \widetilde{\boldsymbol{u}}_{r'} \in \mathrm{Span}(\boldsymbol{U})$ and $\widetilde{\boldsymbol{v}}_1, \cdots, \widetilde{\boldsymbol{v}}_{r'} \in \mathrm{Span}(\boldsymbol{V})$ such that

$$\begin{aligned}
&P\left(\widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H})\widetilde{\boldsymbol{v}}_i \ge \xi_i p - \sigma\sqrt{mp}(1+\sqrt{\rho}+C\log^{-1} m) - \varepsilon_2 p\right) \\
\ge& 1 - P\left(p^{-1}\widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{v}}_i \le \mathbb{E}\left[p^{-1}\widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{v}}_i\right] - \varepsilon_2\right) \\
&- P\left(\left|\widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{H})\widetilde{\boldsymbol{v}}_i\right| \ge \sigma\sqrt{mp}(1+\sqrt{\rho}+C\log^{-1} m)\right) \ge 1 - cn^{-3}.
\end{aligned}$$

Here the first probability term $P\left(p^{-1}\widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{v}}_i \le \mathbb{E}\left[p^{-1}\widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M})\widetilde{\boldsymbol{v}}_i\right] - \varepsilon_2\right)$ can be similarly bounded following the proof of Lemma S.5 and S.6. The second probability term

$$P\left(\left|\widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{H})\widetilde{\boldsymbol{v}}_i\right| \ge \sigma\sqrt{mp}(1+\sqrt{\rho}+C\log^{-1} m)\right) \le P\left(\theta_1 \ge \sigma\sqrt{p}(1+\sqrt{\rho}+C\log^{-1} m)\right).$$

can be bounded using (1). So from the Courant-Fischer-Weyl min-max principle, we have

$$\begin{aligned}
\sigma_i(P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H})) &= \min_{\mathcal{U},\mathcal{V},\dim(\mathcal{U})=m-r,\dim(\mathcal{V})=n-r,} \max_{\boldsymbol{u}\in\mathcal{U},\boldsymbol{v}\in\mathcal{V}} \boldsymbol{u}' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H})\boldsymbol{v} \\
&\ge \widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H})\widetilde{\boldsymbol{v}}_i.
\end{aligned}$$

Thus with probability larger than $1 - cn^{-3}$,

$$\begin{aligned}
\mathcal{L}^\lambda(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega}) \ge& \sum_{i=1}^{r'} \frac{\omega_i}{\widehat{\sigma}^2 mp}\left(\widetilde{\boldsymbol{u}}_i' P_{\boldsymbol{\Omega}}(\boldsymbol{M}_0 - \boldsymbol{M} + \boldsymbol{H})\widetilde{\boldsymbol{v}}_i - \widehat{\sigma}\sqrt{m}\widehat{\lambda}_i\right)^2 + \left(\lambda\|\boldsymbol{M}_0\|_* - \sum_{i=1}^r \tau_i\right) \\
\ge& \mathcal{L}^\lambda(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega}) - \lambda\sum_{i=1}^r \tau_i - C\log^{-2} m \\
&+ \sum_{i=1}^{r'} \frac{\omega_i}{\widehat{\sigma}^2 mp}\left((\tau_i - \varepsilon_2)p - (\sigma+\widehat{\sigma})\sqrt{mp}(1+\sqrt{\rho}+C\log^{-1} m)\right)^2 \\
\ge& \mathcal{L}^\lambda(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega}) + \sum_{i=1}^{r'}\left(\frac{\omega_i}{\widehat{\sigma}^2 mp}(\omega_i p - A)^2 - \lambda\tau_i - \lambda B\right)
\end{aligned}$$

where using (c) in Assumption 1 such that $\varepsilon_2 \ll \varepsilon \le \xi_i, 1 \le i \le r'$ and notation

$$A = (\sigma+\widehat{\sigma})\sqrt{mp}(1+\sqrt{\rho}+C\log^{-1} m), \quad B = \frac{r-r'}{r'}\varepsilon.$$

When

$$\tau_i \ge \frac{\lambda\widehat{\sigma}^2 m}{\omega_i p} + \left(\frac{\lambda(B+A/p)\widehat{\sigma}^2 m}{\omega_i p}\right)^{1/2} + \frac{A}{p},$$

we have with probability larger than $1 - cn^{-3}$ that

$$\mathcal{L}^\lambda(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega}) \ge \mathcal{L}^\lambda(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega}).$$

We also know that

$$\theta_{r+s} = \min_{\mathcal{U},\dim(\mathcal{U})=m-r-s+1,} \max_{\boldsymbol{u}\in\mathcal{U},\boldsymbol{v}\in\mathbb{R}^n} \boldsymbol{u}'\boldsymbol{M}\boldsymbol{v} \le \min_{\mathcal{U}\subset U^\perp,\dim(\mathcal{U})=m-r-s+1,} \max_{\boldsymbol{u}\in\mathcal{U},\boldsymbol{v}\in\mathbb{R}^n} \boldsymbol{u}'(\boldsymbol{M}_0 - \boldsymbol{M})\boldsymbol{v},$$

and similarly

$$\theta_{r+s} = \min_{\mathcal{V},\dim(\mathcal{V})=n-r-s+1,}\ \max_{\boldsymbol{u}\in\mathbb{R}^m,\boldsymbol{v}\in\mathcal{V}} \boldsymbol{u}'\boldsymbol{M}\boldsymbol{v} \le \min_{\mathcal{V}\subset V^{\perp},\dim(\mathcal{V})=n-r-s+1,}\ \max_{\boldsymbol{u}\in\mathbb{R}^m,\boldsymbol{v}\in\mathcal{V}} \boldsymbol{u}'(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{v}.$$

Define the singular values of $(\boldsymbol{M}_0-\boldsymbol{M})-\boldsymbol{U}\boldsymbol{U}'(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{V}\boldsymbol{V}'$ as $\eta_1\ge\cdots\ge\eta_{n-r}\ge 0$ and set $\bar{r}=\inf_k \eta_{k+1}\le\varepsilon$. Similarly we have with probability larger than $1-cn^{-3}$,

$$\mathcal{L}^{\lambda}(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega}) \ge \mathcal{L}^{\lambda}(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega}) + \sum_{i=1}^{\bar{r}} \frac{\omega_{r+i}}{\widehat{\sigma}^2 mp}(\eta_i p - A)^2 - \lambda\sum_{i=1}^{r}\tau_i + \lambda\sum_{i=1}^{\bar{r}}\eta_i \ge \mathcal{L}^{\lambda}(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega}),$$

when $\sum_{i=1}^{\bar{r}}\eta_i\ge\sum_{i=1}^{r}\tau_i$ or

$$\eta_i \ge \left(\frac{\lambda\widehat{\sigma}^2 m(\bar{r}^{-1}\sum_{i=1}^{r}\tau_i - A/p)}{\omega_{r+i}p}\right)^{1/2} + \frac{A}{p}.$$

Thus we have with probability larger than $1-cn^{-3}$, for $\boldsymbol{M}$ with $\mathcal{L}(\boldsymbol{M};\boldsymbol{Y},\boldsymbol{\Omega})\le\mathcal{L}(\boldsymbol{M}_0;\boldsymbol{Y},\boldsymbol{\Omega})$, then

$$\tau_i \le \frac{\lambda\widehat{\sigma}^2 m}{\omega_i p} + \left(\frac{\lambda(B+A/p)\widehat{\sigma}^2 m}{\omega_i p}\right)^{1/2} + \frac{A}{p}.$$

With $\min_{1\le i\le 2r}\omega_i=\bar{\omega}$,

$$\begin{aligned}\sum_{i=1}^{r}\tau_i \le \sum_{i=1}^{r'}\tau_i + (r-r')\varepsilon &\le \frac{rA}{p} + r'\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \sqrt{rr'}\left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}} \\ &\le r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right).\end{aligned}$$

and

$$\sum_{i=1}^{r}\tau_i^2 \le \sum_{i=1}^{r'}\tau_i^2 + (r-r')\varepsilon^2 \le r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right)^2.$$

Thus with probability larger than $1-cn^{-3}$, $\sum_{i=1}^{\bar{r}}\eta_i\le\sum_{i=1}^{r}\tau_i$ and

$$\eta_i \le \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\left[\bar{r}^{-1}r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right) - \frac{A}{p}\right]^{1/2} + \frac{A}{p}.$$

Now we composite $\boldsymbol{M}=\sum_{i=1}^{r}\theta_i\boldsymbol{x}_i\boldsymbol{y}_i'+\sum_{i=r+1}^{n}\theta_i\boldsymbol{x}_i\boldsymbol{y}_i' := \boldsymbol{M}_1+\boldsymbol{M}_2$, then

$$\begin{aligned}\|\boldsymbol{M}_2\|_F^2 &= \sum_{i=1}^{\bar{r}}\theta_{r+i}^2 + \sum_{i=\bar{r}+1}^{n-r}\theta_{r+i}^2 \le \sum_{i=1}^{\bar{r}}\eta_i^2 + \varepsilon\sum_{i=\bar{r}+1}^{n-r}\theta_{r+i} \\ &\le \bar{r}\left(\left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\left[\bar{r}^{-1}r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right) - \frac{A}{p}\right]^{1/2} + \frac{A}{p}\right)^2 \\ &\quad + \left((C_0+1)\lambda^{-1}\log^{-2}m + r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right)\right)\varepsilon := A_2.\end{aligned}$$

Define $r_i, \zeta_i, \xi_i, 1 \le i \le r$ as the singular values of $\boldsymbol{U}'(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{V}, \boldsymbol{X}'_{[1:r]}(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{Y}_{[1:r]}, \boldsymbol{X}'_{[1:r]}(\boldsymbol{M}_0-\boldsymbol{M})\widetilde{\boldsymbol{Y}}_{[1:r]}, \widetilde{\boldsymbol{X}}'_{[1:r]}(\boldsymbol{M}_0-\boldsymbol{M})\boldsymbol{Y}_{[1:r]}$ respectively. Similarly we know that

$$\max\{r_i, \zeta_i, \xi_i\} \le \frac{\lambda\widehat{\sigma}^2 m}{\omega_i p} + \left(\frac{\lambda(B+A/p)\widehat{\sigma}^2 m}{\omega_i p}\right)^{1/2} + \frac{A}{p}$$

and

$$\max\left\{\sum_{i=1}^r r_i, \sum_{i=1}^r \zeta_i, \sum_{i=1}^r \xi_i\right\} \le r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right),$$
$$\max\left\{\sum_{i=1}^r r_i^2, \sum_{i=1}^r \zeta_i^2, \sum_{i=1}^r \xi_i^2\right\} \le r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right)^2.$$

Then define $\widetilde{\tau}_i, \widetilde{r}_i, \widetilde{\zeta}_i, \widetilde{\xi}_i, 1 \le i \le r$ as the singular values of $\boldsymbol{U}'(\boldsymbol{M}_0-\boldsymbol{M}_1)\boldsymbol{V}, \boldsymbol{X}'_{[1:r]}(\boldsymbol{M}_0-\boldsymbol{M}_1)\boldsymbol{Y}_{[1:r]}, \boldsymbol{X}'_{[1:r]}(\boldsymbol{M}_0-\boldsymbol{M}_1)\widetilde{\boldsymbol{Y}}_{[1:r]}, \widetilde{\boldsymbol{X}}'_{[1:r]}(\boldsymbol{M}_0-\boldsymbol{M}_1)\boldsymbol{Y}_{[1:r]}$ respectively, so

$$\begin{aligned}
\sum_{i=1}^r \widetilde{\tau}_i^2 &\le \sum_{i=1}^r \tau_i^2 + \sum_{i=1}^r \theta_{r+i}^2 \le \sum_{i=1}^r \tau_i^2 + \sum_{i=1}^{\bar{r}} \eta_i^2 + |r-\bar{r}|\varepsilon^2 \\
&\le r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right)^2 \\
&+ \bar{r}\left(\left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\left[\bar{r}^{-1} r\left(\frac{A}{p} + \frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p} + \left(\frac{\lambda\widehat{\sigma}^2 m}{\bar{\omega}p}\right)^{1/2}\sqrt{\frac{A}{p}}\right) - \frac{A}{p}\right]^{1/2} + \frac{A}{p}\right)^2 \\
&+ |r-\bar{r}|\varepsilon^2 \\
&:= A_3.
\end{aligned}$$

We have the same upper bounds for $\sum_{i=1}^r \widetilde{r}_i, \sum_{i=1}^r \widetilde{\zeta}_i, \sum_{i=1}^r \widetilde{\xi}_i$. Similar to lemma S.7, we know that

$$\|\boldsymbol{M}_0 - \boldsymbol{M}_1\|_F \le (3+5\kappa)A_3^{1/2}.$$

So

$$\|\boldsymbol{M}_0 - \boldsymbol{M}\|_F \le \|\boldsymbol{M}_0 - \boldsymbol{M}_1\|_F + \|\boldsymbol{M}_2\|_F \le (3+5\kappa)A_3^{1/2} + A_2^{1/2}.$$

With Lemma S.8, define that

$$R_1 = 2(1+\sqrt{\rho})\sigma\sqrt{m/p},$$
$$R_2 = R_1 + \frac{\lambda\sigma^2 m}{\bar{\omega}p} + \left(\frac{\lambda\sigma^2 m}{\bar{\omega}p}\right)^{1/2} R_1^{1/2}.$$

then $\bar{r} \le rR_2/R_1$ and

$$A_2^{1/2} \le \left(\frac{\lambda\sigma^2 m}{\bar{\omega}p}\right)^{1/2}(rR_2 - R_1)^{1/2} + r^{1/2}R_1^{1/2}R_2^{1/2} + \left((C_0+1)\lambda^{-1}\log^{-2} m + rR_2\right)^{1/2} R_1^{1/2}$$

$$A_3^{1/2} \le r^{1/2}R_2 + \left(\frac{\lambda\sigma^2 m}{\bar{\omega}p}\right)^{1/2}(rR_2 - R_1)^{1/2} + r^{1/2}R_1$$

then with probability larger than $1 - cn^{-3}$

$$\|\boldsymbol{M}_0 - \boldsymbol{M}\|_F \le (4+5\kappa)\left(\left(\frac{\lambda\sigma^2 m}{\bar{\omega} p}\right)^{1/2}(rR_2 - R_1)^{1/2} + R_1 + R_2\right)$$
$$+ \left((C_0+1)\lambda^{-1}\log^{-2} m + 2rR_2\right)^{1/2} R_1^{1/2}.$$

□

PROOF OF COROLLARY 4.3. With the upper bound shown in Theorem 4.2, we know that when $\lambda = 2(1+\sqrt{\rho})\sigma^{-1}p^{1/2}m^{-1/2}\bar{\omega}$ and $\bar{\omega} \gg \log^{-2} m$, then

$$R_2 = 3R_1 = 6(1+\sqrt{\rho})\sigma\sqrt{m/p},$$

and $r(R_1+R_2)/R_1 \le 4r$. Thus with probability larger than $1 - cn^{-3}$,

$$\|\widetilde{\boldsymbol{M}}_\lambda - \boldsymbol{M}_0\|_F \le \left((4+5\kappa)\left(\sqrt{3}r^{1/2}+4\right) + \sqrt{6r + (C_0+1)/4(1+\sqrt{\rho})^2}\right) R_1$$
$$\le 2(1+\sqrt{\rho})\left((\sqrt{6}+4\sqrt{3}+5\sqrt{3}\kappa)r^{1/2} + 4(4+5\kappa) + o(1)\right)\sigma\sqrt{m/p}$$

□

S.5. Proof of Lemma 4.3.

LEMMA S.10. *For rank r symmetric matrix $\boldsymbol{M} = \sum_{i=1}^r \sigma_i \boldsymbol{u}_i\boldsymbol{u}_i'$, if $\boldsymbol{H}$ is an independent symmetric perturbation matrix with $\|\boldsymbol{H}\|_\infty \le K$ and $\boldsymbol{M}+\boldsymbol{H} = \sum_{i=1}^n \widetilde{\sigma}_i\widetilde{\boldsymbol{u}}_i\widetilde{\boldsymbol{u}}_i'$. Assume that $\sigma_1 \ge \cdots \ge \sigma_r, |\sigma_i| \ge \sigma_0, |\sigma_i|/\sigma_0 \le \kappa$ and $\|\boldsymbol{H}\|/\sigma_0 = \epsilon_1$, with probability larger than $1 - 2rn^{-c_0}$,*

$$\|\widetilde{\boldsymbol{U}}\|_{2,\infty} \le \|\boldsymbol{U}\|_{2,\infty}\left(\left(\frac{\sqrt{2r}\epsilon_1}{1-\epsilon_1} + 2\sqrt{2r}\left(\frac{(\kappa+\epsilon_1-1)^2\epsilon_1}{(1-\epsilon_1)^3}\right)^{1/3}\right)\left(1+\frac{\sqrt{r\epsilon_1}}{1-4\epsilon_1}\right)\right.$$
$$\left. + \frac{\sqrt{2r}(1+2\epsilon_1)\epsilon_1}{1-4\epsilon_1} + 1 + 2\sqrt{5}\epsilon_1 + \frac{\kappa\sqrt{2(c_0+1)Kn}\log n}{(1-4\epsilon_1)\sigma_0}\right).$$

LEMMA S.11. *For a rank r matrix $\boldsymbol{M}$ with singular value decomposition $\boldsymbol{M} = \boldsymbol{U\Sigma V}'$ such that $\|\boldsymbol{U}\|_{2,\infty} \le \sqrt{\mu_1 r/m}$ and $\|\boldsymbol{V}\|_{2,\infty} \le \sqrt{\mu_2 r/n}$, then*

$$\|\boldsymbol{M}\|_\infty \le \frac{\|\boldsymbol{M}\|_F r\sqrt{\mu_1\mu_2}}{\sqrt{mn}}.$$

LEMMA S.12. *For $\boldsymbol{M} = \sum_{i=1}^r \sigma_i\boldsymbol{u}_i\boldsymbol{v}_i'$ satisfies μ-incoherent condition. Consider the perturbation $\boldsymbol{M}+\boldsymbol{H} = \sum_{i=1}^r \widetilde{\sigma}_i\widetilde{\boldsymbol{u}}_i\widetilde{\boldsymbol{v}}_i' + (\boldsymbol{M}+\boldsymbol{H})^{r,\perp}$ and $\mathrm{Proj}_r(\boldsymbol{M}+\boldsymbol{H}) = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{D}}\widetilde{\boldsymbol{V}}$. Set $\kappa = \sigma_1/\sigma_r, \epsilon_1 = \|\boldsymbol{H}\|_2/\sigma_r$ and $K = \|\boldsymbol{H}\|_\infty$. If $r\|\boldsymbol{H}\|/\sigma_r \le C$, then with probability larger than $1-n^{-3}$ that*

$$\|\widetilde{\boldsymbol{U}}\|_{2,\infty} \le \|\boldsymbol{U}\|_{2,\infty}(1+f), \quad \|\widetilde{\boldsymbol{V}}\|_{2,\infty} \le \|\boldsymbol{V}\|_{2,\infty}(1+f)$$

and

$$\|\mathrm{Proj}_r(\boldsymbol{M}+\boldsymbol{H}) - \boldsymbol{M}\|_\infty \le (\|\mathrm{Proj}_r(\boldsymbol{M}+\boldsymbol{H}) - \boldsymbol{M}\|_F + \sqrt{r}\|\boldsymbol{H}\|)\|\boldsymbol{U}\|_{2,\infty}\|\boldsymbol{V}\|_{2,\infty}(1+f)^2$$
$$\|\mathrm{Proj}_r(\boldsymbol{M}+\boldsymbol{H})\|_\infty \le \|\mathrm{Proj}_r(\boldsymbol{M}+\boldsymbol{H})\|_F\|\boldsymbol{U}\|_{2,\infty}\|\boldsymbol{V}\|_{2,\infty}(1+f)^2.$$

where

$$f = c_1\sqrt{r}\epsilon_1 + c_2\sqrt{r}\epsilon_1^{1/3}(\kappa-1+\epsilon_1)^{2/3} + c_3\sigma_r^{-1}\kappa\sqrt{Km}\log m.$$

PROOF OF LEMMA 4.3. For each row i, from Chernoff–Hoeffding theorem we have

$$P\left(\frac{1}{n}\sum_{j=1}^{n}\mathbf{\Omega}_{i,j}\geq 2p\right)\leq \exp\left\{-\frac{p^2n}{p(1-p)}\right\}\leq cn^{-\alpha}.$$

Similarly we have

$$P\left(\frac{1}{m}\sum_{i=1}^{m}\mathbf{\Omega}_{i,j}\geq 2p\right)\leq cn^{-\alpha}.$$

Consider the trimming procedure. Set to zero all columns with degree larger that $2|\mathbf{\Omega}|/n$ and set to 0 all rows with degree larger than $2|\mathbf{\Omega}|/m$, i.e.,

$$\widetilde{\mathbf{\Omega}}=\{(i,j)|(i,j)\in\mathbf{\Omega},|\mathbf{\Omega}_{i,.}|\leq 2|\mathbf{\Omega}|/n,|\mathbf{\Omega}_{.,j}|\leq 2|\mathbf{\Omega}|/m\}.$$

Then with probability larger than $1-cn^{-\alpha+1}\geq 1-cn^{-3}$ for large enough α, we have $\widetilde{\mathbf{\Omega}}=\mathbf{\Omega}$. With Theorem 1.1 in Keshavan, Montanari and Oh (2010), we know that

$$\begin{aligned}\|p^{-1}P_{\mathbf{\Omega}}(\boldsymbol{Y})-\boldsymbol{M}_0\|&\leq\|p^{-1}P_{\mathbf{\Omega}}(\boldsymbol{M}_0)-\boldsymbol{M}_0\|+\|p^{-1}P_{\mathbf{\Omega}}(\boldsymbol{H})\|\\&\leq C(\rho^{-1/4}\|\boldsymbol{M}_0\|_\infty+\sigma)\sqrt{\frac{m}{p}}\end{aligned}$$

Treating $p^{-1}P_{\mathbf{\Omega}}(\boldsymbol{Y})$ as perturbation of $\boldsymbol{M}_0$, Weyl theorem tells us that

$$|\sigma_i^{(0)}-\sigma_i|\leq\|p^{-1}P_{\mathbf{\Omega}}(\boldsymbol{Y})-\boldsymbol{M}_0\|_2,\quad i=1,2,\cdots,r.$$

and Wedin theorem shows that

$$\|\sin\theta(\boldsymbol{U}^{(0)},\boldsymbol{U})\|_F^2+\|\sin\theta(\boldsymbol{V}^{(0)},\boldsymbol{V})\|_F^2\leq\frac{Cr\|p^{-1}P_{\mathbf{\Omega}}(\boldsymbol{Y})-\boldsymbol{M}_0\|_2^2}{\sigma_r^2}$$

So

$$\begin{aligned}&\|\boldsymbol{M}^{(0)}-\boldsymbol{M}_0\|_F\\\leq&\left(\sum_{i=1}^{r}|\sigma_i(\boldsymbol{M}^{(0)})-\sigma_i|^2\right)^{1/2}+\|\Sigma\sin\theta(\boldsymbol{U}(\boldsymbol{M}^{(0)}),\boldsymbol{U}(\boldsymbol{M}_0))\|_F+\|\Sigma\sin\theta(\boldsymbol{V}(\boldsymbol{M}^{(0)}),\boldsymbol{V}(\boldsymbol{M}_0))\|_F\\\leq&C(\rho^{-1/4}\|\boldsymbol{M}_0\|_\infty+\sigma)\kappa\sqrt{\frac{mr}{p}}.\end{aligned}$$

Apply lemma S.12 with $\boldsymbol{M}^{(0)}=\mathrm{Proj}_r(\boldsymbol{M}_0+(p^{-1}P_{\mathbf{\Omega}}(\boldsymbol{Y})-\boldsymbol{M}_0))$, we know that

$$P\left(\|\boldsymbol{H}\|_\infty\geq 6\sigma c\log m\right)\leq\sum_{i,j}P\left(|\boldsymbol{H}_{ij}|\geq 6\sigma c\log m\right)\leq mn\exp\left(-6\log\right)\leq cn^{-4}.$$

Set the high probability set

$$\mathcal{B}_1=\{\|H\|_\infty\leq 6\sigma c\log m\},P(\mathcal{B}_1)\geq 1-cn^{-4}.$$

Then

$$\|p^{-1}P_{\mathbf{\Omega}}(\boldsymbol{Y})-\boldsymbol{M}_0\|_\infty\leq p^{-1}((1+p)\|\boldsymbol{M}_0\|_\infty+\|\boldsymbol{H}\|_\infty)$$

So with probability larger than $1-cn^{-4}$, we can choose $K=p^{-1}((1+p)\|\boldsymbol{M}_0\|_\infty+6\sigma c^{-1}\log m)$ and

$$\|\boldsymbol{U}^{(0)}\|_{2,\infty}\leq\|\boldsymbol{U}\|_{2,\infty}(1+f^{(0)}),\quad\|\boldsymbol{V}^{(0)}\|_{2,\infty}\leq\|\boldsymbol{V}\|_{2,\infty}(1+f^{(0)})$$

and

$$\|\boldsymbol{M}^{(0)}-\boldsymbol{M}_0\|_\infty \le \|\boldsymbol{U}\|_{2,\infty}\|\boldsymbol{V}\|_{2,\infty}(\|\boldsymbol{M}^{(0)}-\boldsymbol{M}_0\|_F+\sqrt{r}\|\boldsymbol{H}\|)(1+f^{(0)})^2$$
$$\|\boldsymbol{M}^{(0)}\|_\infty \le \|\boldsymbol{U}\|_{2,\infty}\|\boldsymbol{V}\|_{2,\infty}\|\boldsymbol{M}^{(0)}\|_F(1+f^{(0)})^2,$$

where

$$\begin{aligned}
f^{(0)} &= c_1\sqrt{r}\epsilon_1 + c_2\sqrt{r}\epsilon_1^{1/3}(\kappa-1+\epsilon_1)^{2/3} + c_3\sigma_r^{-1}\kappa\sqrt{Km}\log m \\
&\le Cc_1\sqrt{r}\sigma_r^{-1}(\rho^{-1/4}\|\boldsymbol{M}_0\|_\infty+\sigma)\sqrt{\frac{m}{p}} \\
&+ Cc_2\sqrt{r}\sigma_r^{-1/3}(\rho^{-1/4}\|\boldsymbol{M}_0\|_\infty+\sigma)^{1/3}\sqrt{\frac{m}{p}}^{1/3}(\kappa-1+\sigma_r^{-1}(\rho^{-1/4}\|\boldsymbol{M}_0\|_\infty+\sigma)\sqrt{\frac{m}{p}})^{2/3} \\
&+ c_3\kappa\log m\sigma_r^{-1}\sqrt{p^{-1}m((1+p)\|\boldsymbol{M}_0\|_\infty+6\sigma c\log m)} \\
&\le C_1\sqrt{r}(\kappa-1)+C_2(\kappa+\sqrt{r})(\|\boldsymbol{M}_0\|_\infty+6\sigma c\log m)^{1/2}\left(\frac{m}{p\sigma_r^2}\right)^{1/2}\log m.
\end{aligned}$$

□

PROOF OF LEMMA S.10. The proof is inspired by the technology used in Theorem 6 of Bhardwaj and Vu (2024), in which the absolute value perturbation bound is given for each element of top eigenvectors.

Let $\boldsymbol{H}^{\{l\}}$ be the perturbation obtained from $\boldsymbol{H}$ by leaving out the lth row and column and let $\boldsymbol{H}_{\{l\}}=\boldsymbol{H}-\boldsymbol{H}^{\{l\}}$. Also we set $\boldsymbol{M}^{\{l\}}=\boldsymbol{M}+\boldsymbol{H}^{\{l\}}$ and the eigenvalue and eigenvector as $\sigma_i^{\{l\}}$ and $\boldsymbol{u}_i^{\{l\}}$. Define that $\alpha_{ij}=\boldsymbol{u}_i^{\{l\}\prime}\boldsymbol{u}_j$, $\beta_{ij}=\widetilde{\boldsymbol{u}}_i'\boldsymbol{u}_j^{\{l\}}$ and $\kappa_i=\|\boldsymbol{M}\|/|\sigma_i|$, $\epsilon_1(i)=\|\boldsymbol{H}\|/|\sigma_i|$, we want to prove that

$$|\boldsymbol{u}_{il}^{\{l\}}-\sum_{j=1}^r\alpha_{ij}\boldsymbol{u}_{jl}| \le \|\boldsymbol{U}_{l,\cdot}\|_2\left(\frac{\epsilon_1(i)}{1-\epsilon_1(i)}+2\left(\frac{(\kappa_i+\epsilon_1(i)-1)^2\epsilon_1(i)}{(1-\epsilon_1(i))^3}\right)^{1/3}\right), \tag{4}$$

and with probability larger than $1-2n^{-c_0-1}$,

$$\|\widetilde{\boldsymbol{u}}_i-\sum_{j=1}^r\beta_{ij}\boldsymbol{u}_j^{\{l\}}\|_2 \le \frac{\kappa_i\sqrt{2(c_0+1)Kn}\log n\|\boldsymbol{U}\|_{2,\infty}+\|\boldsymbol{H}\|_2\|\boldsymbol{U}_{l,\cdot}^{\{l\}}\|_2}{\sigma_i-4\|\boldsymbol{H}\|}. \tag{5}$$

To prove (4), from

$$\boldsymbol{u}_{il}^{\{l\}}=\sigma_i^{\{l\}-1}\langle\boldsymbol{M}_{l,\cdot}^{\{l\}},\boldsymbol{u}_i^{\{l\}}\rangle=\sigma_i^{\{l\}-1}\langle\boldsymbol{M}_{l,\cdot},\boldsymbol{u}_i^{\{l\}}\rangle=\sigma_i^{\{l\}-1}\sum_{j=1}^r\sigma_j\boldsymbol{u}_{jl}\boldsymbol{u}_j'\boldsymbol{u}_i^{\{l\}}=\sigma_i^{\{l\}-1}\sum_{j=1}^r\sigma_j\alpha_{ij}\boldsymbol{u}_{jl},$$

so

$$|\boldsymbol{u}_{il}^{\{l\}}-\sum_{j=1}^r\alpha_{ij}\boldsymbol{u}_{jl}|=|\sigma_i^{\{l\}-1}\sum_{j=1}^r\alpha_{ij}(\sigma_j-\sigma_i^{\{l\}})\boldsymbol{u}_{jl}|.$$

From

$$\sum_{j=1}^r\alpha_{ij}\boldsymbol{u}_j'\boldsymbol{M}^{\{l\}}\sum_{j=1}^r\alpha_{ij}\boldsymbol{u}_j=\sum_{j=1}^r\alpha_{ij}\boldsymbol{u}_j'(\boldsymbol{M}+\boldsymbol{H}^{\{l\}})\sum_{j=1}^r\alpha_{ij}\boldsymbol{u}_j=\sum_{j=1}^r\alpha_{ij}^2\sigma_j+\sum_{j=1}^r\alpha_{ij}\boldsymbol{u}_j'\boldsymbol{H}^{\{l\}}\sum_{j=1}^r\alpha_{ij}\boldsymbol{u}_j$$

and

$$\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j'\boldsymbol{M}^{\{l\}}\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j = (\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j - \boldsymbol{u}_i^{\{l\}} + \boldsymbol{u}_i^{\{l\}})'\boldsymbol{M}^{\{l\}}(\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j - \boldsymbol{u}_i^{\{l\}} + \boldsymbol{u}_i^{\{l\}})$$
$$= \sigma_i^{\{l\}} + (\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j - \boldsymbol{u}_i^{\{l\}})'\boldsymbol{M}^{\{l\}}(\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j - \boldsymbol{u}_i^{\{l\}}),$$

so

$$|\sum_{j=1}^{r}\alpha_{ij}^2\sigma_j - \sigma_i^{\{l\}}| = |(\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j - \boldsymbol{u}_i^{\{l\}})'\boldsymbol{M}^{\{l\}}(\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j - \boldsymbol{u}_i^{\{l\}}) - \alpha_{ij}\boldsymbol{u}_j'\boldsymbol{H}^{\{l\}}\sum_{j=1}^{r}\alpha_{ij}\boldsymbol{u}_j|$$
$$= |\boldsymbol{u}_i^{\{l\}\prime}\boldsymbol{H}^{\{l\}}\boldsymbol{u}_i^{\{l\}}| \le \|\boldsymbol{H}\|.$$

Because $|\sigma_i - \sigma_i^{\{l\}}| \le \|\boldsymbol{H}^{\{l\}}\| \le \|\boldsymbol{H}\|$, we know that

$$\sum_{j=1}^{k}\alpha_{ij}^2 \le \frac{\|\boldsymbol{H}\|}{\sigma_k - \sigma_i}, \quad \sum_{j=k}^{r}\alpha_{ij}^2 \le \frac{\|\boldsymbol{H}\|}{\sigma_i - \sigma_k}.$$

Choose $k_1 = \max_k \sigma_k - \sigma_i \ge \alpha$ and $k_2 = \min_k \sigma_i - \sigma_k \ge \alpha$. So

$$\begin{aligned}
&|\sigma_i^{\{l\}-1}\sum_{j=1}^{r}\alpha_{ij}(\sigma_j - \sigma_i^{\{l\}})\boldsymbol{u}_{jl}| \\
\le&|\sigma_i^{\{l\}-1}\sum_{j\le k_1, j\ge k_2}\alpha_{ij}(\sigma_j - \sigma_i^{\{l\}})\boldsymbol{u}_{jl}| + |\sigma_i^{\{l\}-1}\sum_{j=k_1+1}^{k_2-1}\alpha_{ij}(\sigma_j - \sigma_i^{\{l\}})\boldsymbol{u}_{jl}| \\
\le&\left(\sum_{j\le k_1, j\ge k_2}\alpha_{ij}^2\frac{(\sigma_j - \sigma_i^{\{l\}})^2}{\sigma_i^{\{l\}2}}\sum_{j\le k_1, j\ge k_2}\boldsymbol{u}_{jl}^2\right)^{1/2} + \left(\sum_{j=k_1+1}^{k_2-1}\alpha_{ij}^2\frac{(\sigma_j - \sigma_i^{\{l\}})^2}{\sigma_i^{\{l\}2}}\sum_{j=k_1+1}^{k_2-1}\boldsymbol{u}_{jl}^2\right)^{1/2} \\
\le&\|\boldsymbol{U}_{l,\cdot}\|_2\left(\sum_{j\le k_1, j\ge k_2}\alpha_{ij}^2\frac{(\sigma_j - \sigma_i^{\{l\}})^2}{\sigma_i^{\{l\}2}} + \sum_{j=k_1+1}^{k_2-1}\alpha_{ij}^2\frac{(\sigma_j - \sigma_i^{\{l\}})^2}{\sigma_i^{\{l\}2}}\right) \\
\le&\|\boldsymbol{U}_{l,\cdot}\|_2\left(\|\boldsymbol{H}\|^{1/2}\alpha^{-1/2}\frac{\kappa_i + \epsilon_1(i) - 1}{1 - \epsilon_1(i)} + \frac{\alpha + \|\boldsymbol{H}\|}{\sigma_i - \|\boldsymbol{H}\|}\right) \\
\le&\|\boldsymbol{U}_{l,\cdot}\|_2\left(\frac{\epsilon_1(i)}{1 - \epsilon_1(i)} + 2\left(\frac{(\kappa_i + \epsilon_1(i) - 1)^2\epsilon_1(i)}{(1 - \epsilon_1(i))^3}\right)^{1/3}\right),
\end{aligned}$$

where we set

$$\alpha = \|\boldsymbol{H}\|\left(\frac{\kappa_i + \epsilon_1(i) - 1}{\epsilon_1(i)}\right)^{2/3}.$$

To prove (5), we denote that $P^{\{l\}}\widetilde{\boldsymbol{u}}_i = \widetilde{\boldsymbol{u}}_i - \sum_{i=1}^{r}\beta_{ij}\boldsymbol{u}_j^{\{l\}}$. So

$$|(P^{\{l\}}\widetilde{\boldsymbol{u}}_i)'\boldsymbol{M}^{\{l\}}\widetilde{\boldsymbol{u}}_i| = |\langle\widetilde{\boldsymbol{u}}_i, P^{\{l\}}\boldsymbol{M}^{\{l\}}\widetilde{\boldsymbol{u}}_i\rangle| \le \sigma_{r+1}^{\{l\}}\|P^{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2^2$$

and

$$(P^{\{l\}}\widetilde{\boldsymbol{u}}_i)'\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{u}}_i = \widetilde{\sigma}_i\|P^{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2^2.$$

Then

$$
\begin{aligned}
|(P^{\{l\}}\widetilde{\boldsymbol{u}}_i)'\boldsymbol{H}_{\{l\}}\widetilde{\boldsymbol{u}}_i| &\geq (P^{\{l\}}\widetilde{\boldsymbol{u}}_i)'\widetilde{\boldsymbol{M}}\widetilde{\boldsymbol{u}}_i - |(P^{\{l\}}\widetilde{\boldsymbol{u}}_i)'\boldsymbol{M}^{\{l\}}\widetilde{\boldsymbol{u}}_i| \\
&\geq \widetilde{\sigma}_i\|P^{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2^2 - \sigma_{r+1}^{\{l\}}\|P^{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2^2 \\
&\geq (\sigma_i - 2\|\boldsymbol{H}\|)\|P^{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2^2.
\end{aligned}
$$

So

$$
\begin{aligned}
\|P^{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2 \leq \frac{\|\boldsymbol{H}_{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2}{\sigma_i - 2\|\boldsymbol{H}\|} &\leq \frac{\|\boldsymbol{H}_{\{l\}}\sum_{j=1}^r \beta_{ij}\boldsymbol{u}_j^{\{l\}}\|_2 + \|\boldsymbol{H}_{\{l\}}\|\|\widetilde{\boldsymbol{u}}_i - \sum_{j=1}^r \beta_{ij}\boldsymbol{u}_j^{\{l\}}\|_2}{\sigma_i - 2\|\boldsymbol{H}\|} \\
&\leq \frac{\|\boldsymbol{H}_{\{l\}}\sum_{j=1}^r \beta_{ij}\boldsymbol{u}_j^{\{l\}}\|_2 + 2\|\boldsymbol{H}\|\|P^{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2}{\sigma_i - 2\|\boldsymbol{H}\|}.
\end{aligned}
$$

Rewrite that $\boldsymbol{H}_{\{l\}} = \boldsymbol{h}_{\{l\}}\boldsymbol{e}_l' + \boldsymbol{e}_l\boldsymbol{h}_{\{l\}}'$, we have

$$
\begin{aligned}
\|\boldsymbol{H}_{\{l\}}\sum_{j=1}^r \beta_{ij}\boldsymbol{u}_j^{\{l\}}\|_2 &\leq |\langle \boldsymbol{h}_{\{l\}}, \sum_{j=1}^r \beta_{ij}\boldsymbol{u}_j^{\{l\}}\rangle| + \|\boldsymbol{h}_{\{l\}}\|_2|\sum_{j=1}^r \beta_{ij}\boldsymbol{u}_{jl}^{\{l\}}| \\
&\leq |\langle \boldsymbol{h}_{\{l\}}, \sum_{j=1}^r \beta_{ij}\boldsymbol{u}_j^{\{l\}}\rangle| + \|\boldsymbol{H}\|\|\boldsymbol{U}_{l,\cdot}^{\{l\}}\|_2.
\end{aligned}
$$

Then

$$
\|P^{\{l\}}\widetilde{\boldsymbol{u}}_i\|_2 \leq \frac{|\langle \boldsymbol{h}_{\{l\}}, \sum_{j=1}^r \beta_{ij}\boldsymbol{u}_j^{\{l\}}\rangle| + \|\boldsymbol{H}\|\|\boldsymbol{U}_{l,\cdot}^{\{l\}}\|_2}{\sigma_i - 4\|\boldsymbol{H}\|}.
$$

Follow the proof of Lemma 17 and Theorem 15 in Bhardwaj and Vu (2024) with the fact that $\boldsymbol{h}_{\{l\}}$ is independent to $\boldsymbol{u}_j^{\{l\}}$, we finally have

$$
P\left(|\langle \boldsymbol{h}_{\{l\}}, \sum_{j=1}^r \beta_{ij}\boldsymbol{u}_j^{\{l\}}\rangle| \geq \|\boldsymbol{U}\|_{2,\infty}\kappa_i\sqrt{2(c_0+1)Kn}\log n\right) \leq 2n^{-c_0-1}.
$$

Then we finish the proof of (4) and (5).

Using

$$
\min_i \sum_{j=1}^r \alpha_{ij}^2 \geq 1 - \|\boldsymbol{H}^{\{l\}}\|^2/\sigma_0^2 \geq 1 - \|\boldsymbol{H}\|^2/\sigma_0^2,
$$

$$
\min_i \sum_{j=1}^r \beta_{ij}^2 \geq 1 - \|\boldsymbol{H}_{\{l\}}\|^2/\sigma_0^2 \geq 1 - 4\|\boldsymbol{H}\|^2/\sigma_0^2,
$$

we have

$$
\begin{aligned}
\|(\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 &\leq \|\boldsymbol{U}_{l,\cdot}\|_2 + \|((\mathcal{I} - \mathcal{P}_{\boldsymbol{U}^{\{l\}}})\boldsymbol{U})_{l,\cdot}\|_2 \\
&\leq \|\boldsymbol{U}\|_{2,\infty} + \|\mathcal{I} - \mathcal{P}_{\boldsymbol{U}^{\{l\}}}\|\|\boldsymbol{U}\|_{2,\infty} \\
&\leq \|\boldsymbol{U}\|_{2,\infty}\left(1 + 2\|\boldsymbol{H}\|\sigma_0^{-1}\right), \\
\|(\mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 &\leq \|\boldsymbol{U}_{l,\cdot}\|_2 + \|((\mathcal{I} - \mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{\{l\}}})\boldsymbol{U})_{l,\cdot}\|_2 \\
&\leq \|\boldsymbol{U}\|_{2,\infty} + \|\mathcal{I} - \mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\|\|\boldsymbol{U}\|_{2,\infty} \\
&\leq \|\boldsymbol{U}\|_{2,\infty}\left(1 + 2\sqrt{5}\|\boldsymbol{H}\|\sigma_0^{-1}\right).
\end{aligned}
$$

It can be checked that

$$\begin{aligned}
\|\boldsymbol{U}^{\{l\}}_{l,\cdot}\|_2 &\le \|\boldsymbol{U}^{\{l\}}_{l,\cdot} - (\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 + \|(\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 \\
&\le \left(\sum_{i=1}^r (\boldsymbol{u}^{\{l\}}_{il} - \sum_{j=1}^r \alpha_{ij}\boldsymbol{u}_{jl})^2\right)^{1/2} + \|(\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 \\
&\le \|\boldsymbol{U}\|_{2,\infty}\left(\frac{\sqrt{r}\epsilon_1}{1-\epsilon_1} + 2\sqrt{r}\left(\frac{(\kappa+\epsilon_1-1)^2\epsilon_1}{(1-\epsilon_1)^3}\right)^{1/3} + 1 + 2\epsilon_1\right).
\end{aligned}$$

Then for any l, with probability larger than $1-2rn^{-c_0-1}$,

$$\begin{aligned}
&\|\widetilde{\boldsymbol{U}}_{l,\cdot}\|_2 \le \|\widetilde{\boldsymbol{U}}_{l,\cdot} - (\mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 + \|(\mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 \\
&\le \left(\sum_{i=1}^r (\widetilde{\boldsymbol{u}}_{il} - \sum_{j=1}^r \beta_{ij}\sum_{k=1}^r \alpha_{jk}\boldsymbol{u}_{kl})^2\right)^{1/2} + \|(\mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 \\
&\le \left(\sum_{i=1}^r \left(\widetilde{\boldsymbol{u}}_{il} - \sum_{j=1}^r \beta_{ij}\boldsymbol{u}^{\{l\}}_{jl} + \sum_{j=1}^r \beta_{ij}\left(\boldsymbol{u}^{\{l\}}_{jl} - \sum_{k=1}^r \alpha_{jk}\boldsymbol{u}_{kl}\right)\right)^2\right)^{1/2} + \|(\mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 \\
&\le \sqrt{2}\left(\sum_{i=1}^r (\widetilde{\boldsymbol{u}}_{il} - \sum_{j=1}^r \beta_{ij}\boldsymbol{u}^{\{l\}}_{jl})^2 + \sum_{j=1}^r \left(\sum_{i=1}^r \beta_{ij}^2\right)\left(\boldsymbol{u}^{\{l\}}_{jl} - \sum_{k=1}^r \alpha_{jk}\boldsymbol{u}_{kl}\right)^2\right)^{1/2} + \|(\mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{\{l\}}}\boldsymbol{U})_{l,\cdot}\|_2 \\
&\le \|\boldsymbol{U}\|_{2,\infty}\left(\left(\frac{\sqrt{2r}\epsilon_1}{1-\epsilon_1} + 2\sqrt{2r}\left(\frac{(\kappa+\epsilon_1-1)^2\epsilon_1}{(1-\epsilon_1)^3}\right)^{1/3}\right)\left(1+\frac{\sqrt{r\epsilon_1}}{1-4\epsilon_1}\right)\right. \\
&\qquad\qquad \left. + \frac{\sqrt{2r}(1+2\epsilon_1)\epsilon_1}{1-4\epsilon_1} + 1 + 2\sqrt{5}\epsilon_1 + \frac{\kappa\sqrt{2(c_0+1)Kn}\log n}{(1-4\epsilon_1)\sigma_0}\right).
\end{aligned}$$

So with probability larger than $1-2rn^{-c_0}$,

$$\begin{aligned}
\|\widetilde{\boldsymbol{U}}\|_{2,\infty} \le \|\boldsymbol{U}\|_{2,\infty}&\left(\left(\frac{\sqrt{2r}\epsilon_1}{1-\epsilon_1} + 2\sqrt{2r}\left(\frac{(\kappa+\epsilon_1-1)^2\epsilon_1}{(1-\epsilon_1)^3}\right)^{1/3}\right)\left(1+\frac{\sqrt{r\epsilon_1}}{1-4\epsilon_1}\right)\right. \\
&\qquad \left. + \frac{\sqrt{2r}(1+2\epsilon_1)\epsilon_1}{1-4\epsilon_1} + 1 + 2\sqrt{5}\epsilon_1 + \frac{\kappa\sqrt{2(c_0+1)Kn}\log n}{(1-4\epsilon_1)\sigma_0}\right).
\end{aligned}$$

□

PROOF OF LEMMA S.11. It is easy to check that

$$\begin{aligned}
|\boldsymbol{M}_{ij}| = |\sum_{k=1}^r \sigma_k \boldsymbol{u}_{ki}\boldsymbol{v}_{kj}| &\le \left(\sum_{k=1}^r \sigma_k^2 \sum_{k=1}^r \boldsymbol{u}_{ki}^2\boldsymbol{v}_{kj}^2\right)^{1/2} \\
&\le \|\boldsymbol{M}\|_F\|\boldsymbol{U}\|_{2,\infty}\|\boldsymbol{V}\|_{2,\infty} \le \frac{\|\boldsymbol{M}\|_F r\sqrt{\mu_1\mu_2}}{\sqrt{mn}}.
\end{aligned}$$

□

PROOF OF LEMMA S.12. Consider the matrix perturbation problem for

$$\begin{pmatrix} 0 & \boldsymbol{M} \\ \boldsymbol{M}' & 0 \end{pmatrix} + \begin{pmatrix} 0 & \boldsymbol{H} \\ \boldsymbol{H}' & 0 \end{pmatrix}$$

and apply lemma S.10, we know that with probability at least $1 - cn^{-3}$,

$$\|\widetilde{\boldsymbol{U}}\|_{2,\infty} \le \|\boldsymbol{U}\|_{2,\infty} \left(\left(\frac{\sqrt{2r}\epsilon_1}{1-\epsilon_1} + 2\sqrt{2r} \left(\frac{(\kappa + \epsilon_1 - 1)^2 \epsilon_1}{(1-\epsilon_1)^3} \right)^{1/3} \right) \left(1 + \frac{\sqrt{r\epsilon_1}}{1 - 4\epsilon_1} \right) \right.$$
$$\left. + \frac{\sqrt{2r}(1+2\epsilon_1)\epsilon_1}{1-4\epsilon_1} + 1 + 2\sqrt{5}\epsilon_1 + \frac{\kappa\sqrt{10Km}\log m}{(1-4\epsilon_1)\sigma_r} \right)$$

where $\kappa = \sigma_1/\sigma_r, \epsilon_1 = \|\boldsymbol{H}\|/\sigma_r$ and $K = \|\boldsymbol{H}\|_\infty$. With assumption that $r\|\boldsymbol{H}\|/\sigma_r \le C$,

$$\|\widetilde{\boldsymbol{U}}\|_{2,\infty} \le \|\boldsymbol{U}\|_{2,\infty} \left(1 + c_1\sqrt{r}\epsilon_1 + c_2\sqrt{r}\epsilon_1^{1/3}(\kappa - 1 + \epsilon_1)^{2/3} + c_3\sigma_r^{-1}\kappa\sqrt{Km}\log m \right)$$

Similarly we have

$$\|\widetilde{\boldsymbol{V}}\|_{2,\infty} \le \|\boldsymbol{V}\|_{2,\infty} \left(1 + c_1\sqrt{r}\epsilon_1 + c_2\sqrt{r}\epsilon_1^{1/3}(\kappa - 1 + \epsilon_1)^{2/3} + c_3\sigma_r^{-1}\kappa\sqrt{Km}\log m \right)$$

Thus from lemma S.11,

$$\begin{aligned} &\|\operatorname{Proj}_r(\boldsymbol{M}+\boldsymbol{H})\|_\infty \\ \le& \frac{\|\boldsymbol{M}+\boldsymbol{H}\|_F r\mu_0}{\sqrt{mn}} \left(1 + c_1\sqrt{r}\epsilon_1 + c_2\sqrt{r}\epsilon_1^{1/3}(\kappa - 1 + \epsilon_1)^{2/3} + c_3\sigma_r^{-1}\kappa\sqrt{Km}\log m \right)^2. \end{aligned}$$

In Lemma S.10, we also know that

$$\|\widetilde{\boldsymbol{U}}_{l,\cdot} - (\mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{(l)}}\boldsymbol{U})_{l,\cdot}\|_2 \le \|\boldsymbol{U}\|_{2,\infty} \left(c_1\sqrt{r}\epsilon_1 + c_2\sqrt{r}\epsilon_1^{1/3}(\kappa - 1 + \epsilon_1)^{2/3} + c_3\sigma_r^{-1}\kappa\sqrt{Km}\log m \right),$$
$$\|\widetilde{\boldsymbol{V}}_{l,\cdot} - (\mathcal{P}_{\widetilde{\boldsymbol{V}}}\mathcal{P}_{\boldsymbol{V}^{(l)}}\boldsymbol{V})_{l,\cdot}\|_2 \le \|\boldsymbol{V}\|_{2,\infty} \left(c_1\sqrt{r}\epsilon_1 + c_2\sqrt{r}\epsilon_1^{1/3}(\kappa - 1 + \epsilon_1)^{2/3} + c_3\sigma_r^{-1}\kappa\sqrt{Km}\log m \right).$$

Then

$$\begin{aligned} &|[\operatorname{Proj}_r(\boldsymbol{M}+\boldsymbol{H}) - \boldsymbol{M}]_{i,j}| = |\sum_{k=1}^r \sigma_k \boldsymbol{u}_{ki}\boldsymbol{v}_{kj} - \sum_{k=1}^r \widetilde{\sigma}_k \widetilde{\boldsymbol{u}}_{ik}\widetilde{\boldsymbol{v}}_{jk}| \\ \le& \|\operatorname{Proj}_r(\boldsymbol{M}+\boldsymbol{H}) - \boldsymbol{M}\|_F \|\widetilde{\boldsymbol{U}}\|_{2,\infty}\|\widetilde{\boldsymbol{V}}\|_{2,\infty} + |\sum_{k=1}^r \sigma_k(\widetilde{\boldsymbol{u}}_{ik} - (\mathcal{P}_{\widetilde{\boldsymbol{U}}}\mathcal{P}_{\boldsymbol{U}^{(l)}}\boldsymbol{u})_{ik})|\|\widetilde{\boldsymbol{V}}\|_{2,\infty} \\ &+ |\sum_{k=1}^r \sigma_k(\widetilde{\boldsymbol{v}}_{jk} - (\mathcal{P}_{\widetilde{\boldsymbol{V}}}\mathcal{P}_{\boldsymbol{V}^{(l)}}\boldsymbol{v})_{jk})|\|\boldsymbol{U}\|_{2,\infty} \\ \le& \left(\|\operatorname{Proj}_r(\boldsymbol{M}+\boldsymbol{H}) - \boldsymbol{M}\|_F + \sqrt{r}\|\boldsymbol{H}\| \right) \|\boldsymbol{U}\|_{2,\infty}\|\boldsymbol{V}\|_{2,\infty} \Big(1 + c_1\sqrt{r}\epsilon_1 + c_2\sqrt{r}\epsilon_1^{1/3}(\kappa - 1 + \epsilon_1)^{2/3} \\ &+ c_3\sigma_r^{-1}\kappa\sqrt{Km}\log m \Big)^2. \end{aligned}$$

□

S.6. Proof of Theorem 4.3 and Corollary 4.4.

LEMMA S.13. *Define that rank r matrix $\boldsymbol{M}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$ and $\widehat{\boldsymbol{M}}=\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{V}}'$, set $\boldsymbol{x}_1,\boldsymbol{y}_1$ as the top singular vector of $\boldsymbol{M}-\widehat{\boldsymbol{M}}$ and $\kappa_r=\sigma_1/\sigma_r$. If $\|\boldsymbol{M}-\widehat{\boldsymbol{M}}\|_F\le C_1$ and $\|\boldsymbol{x}-\boldsymbol{x}_1\|_2,\|\boldsymbol{y}-\boldsymbol{y}_1\|_2\le C_2<1$, then*

$$\|\widehat{\boldsymbol{M}}+\eta\boldsymbol{x}\boldsymbol{y}'-\mathrm{Proj}_r(\widehat{\boldsymbol{M}}+\eta\boldsymbol{x}\boldsymbol{y}')\|_F\le\eta\left(3C_2+\left(\frac{\kappa_r^2\sigma_r^2(2-2rC_1^2\sigma_r^{-2})^2}{C_1^2}+1\right)^{-1/2}\right).$$

PROOF OF THEOREM 4.3. Denote the singular value decomposition

$$P_{\boldsymbol{\Omega}}(\boldsymbol{M}^{(k)}-\boldsymbol{Y})=\sum_{i=1}^{n}\theta_i^{(k)}\boldsymbol{x}_i^{(k)}\boldsymbol{y}_i^{(k)'}$$

and

$$\boldsymbol{M}^{(k)}-\boldsymbol{M}_0=\sum_{i=1}^{2r}\widetilde{\theta}_i^{(k)}\widetilde{\boldsymbol{x}}_i^{(k)}\widetilde{\boldsymbol{y}}_i^{(k)}=\sum_{i=1}^{r(k)}\widetilde{\theta}_i^{(k)}\widetilde{\boldsymbol{x}}_i^{(k)}\widetilde{\boldsymbol{y}}_i^{(k)}+\Delta\boldsymbol{M}^{(k)},$$

where $r(k)$ is the largest i satisfies that $\widetilde{\theta}_i^{(k)}\ge C(\rho^{-1/4}\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_\infty+\sigma)\sqrt{m/p}$. It must holds that $r(k)\ge 1$ when $\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|\ge\kappa\sigma\sqrt{m/p}$. Define that $\boldsymbol{X}^{(k)}=(\boldsymbol{x}_1^{(k)},\cdots,\boldsymbol{x}_{r(k)}^{(k)}),\boldsymbol{Y}^{(k)}=(\boldsymbol{y}_1^{(k)},\cdots,\boldsymbol{y}_{r(k)}^{(k)})$ and similarly $\widetilde{\boldsymbol{X}}^{(k)}=(\widetilde{\boldsymbol{x}}_1^{(k)},\cdots,\widetilde{\boldsymbol{x}}_{r(k)}^{(k)}),\widetilde{\boldsymbol{Y}}^{(k)}=(\widetilde{\boldsymbol{y}}_1^{(k)},\cdots,\widetilde{\boldsymbol{y}}_{r(k)}^{(k)})$. It is easy to check that $\|\widetilde{\boldsymbol{X}}^{(k)}\|_{2,\infty}^2\le\mu_0 r/m+\|\boldsymbol{U}^{(k)}\|_{2,\infty}^2$ and $\|\widetilde{\boldsymbol{Y}}^{(k)}\|_{2,\infty}^2\le\mu_0 r/n+\|\boldsymbol{V}^{(k)}\|_{2,\infty}^2$. From

$$\begin{aligned}\|P_{\boldsymbol{\Omega}}(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0-\boldsymbol{H})-p(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)\|&\le\|P_{\boldsymbol{\Omega}}(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)-p(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)\|+\|P_{\boldsymbol{\Omega}}(\boldsymbol{H})\|\\&\le C(\rho^{-1/4}\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_\infty+\sigma)\sqrt{mp},\end{aligned}$$

with Weyl and Wedin theorem, we have

$$|\theta_i^{(k)}-p\widetilde{\theta}_i^{(k)}|\le C(\rho^{-1/4}\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_\infty+\sigma)\sqrt{mp},$$

$$\|\sin\theta(\boldsymbol{X}^{(k)},\widetilde{\boldsymbol{X}}^{(k)})\|_F^2+\|\sin\theta(\boldsymbol{Y}^{(k)},\widetilde{\boldsymbol{Y}}^{(k)})\|_F^2\le\frac{Cr(k)(C(\rho^{-1/4}\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_\infty+\sigma)\sqrt{mp})^2}{p^2\widetilde{\theta}_{r(k)}^2}.$$

We can calculate that in $\mathcal{B}_1$,

$$\begin{aligned}K^{(k)}&:=\|p^{-1}P_{\boldsymbol{\Omega}}(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0-\boldsymbol{H})-(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)\|_\infty\\&\le p^{-1}\left(\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_\infty+\|\boldsymbol{H}\|_\infty\right)\\&\le p^{-1}(\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_\infty+6\sigma c\log m),\end{aligned}$$

and

$$\begin{aligned}\|p^{-1}P_{\boldsymbol{\Omega}}(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0-\boldsymbol{H})-(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)\|&\le\|p^{-1}P_{\boldsymbol{\Omega}}(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)-(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)\|+\|p^{-1}P_{\boldsymbol{\Omega}}\boldsymbol{H}\|\\&\le C(\rho^{-1/4}\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_\infty+\sigma)\sqrt{\frac{m}{p}}.\end{aligned}$$

Apply Lemma S.12 to

$$p^{-1}P_\Omega(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0-\boldsymbol{H})=(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)+\left[p^{-1}P_\Omega(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0-\boldsymbol{H})-(\boldsymbol{M}^{(k)}-\boldsymbol{M}_0)\right],$$

with choosing $\sigma^{(k)} \geq Cr(\rho^{-1/4}\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_\infty + \sigma)\sqrt{\frac{m}{p}}$, then

$$\|\boldsymbol{X}^{(k)}\|_{2,\infty} \leq \|\widetilde{\boldsymbol{X}}^{(k)}\|_{2,\infty}(1+f^{(k)})^2, \quad \|\boldsymbol{Y}^{(k)}\|_{2,\infty} \leq \|\widetilde{\boldsymbol{Y}}^{(k)}\|_{2,\infty}(1+f^{(k)})^2,$$

where

$$f^{(k)} = c_1\sqrt{r(k)}\epsilon_1^{(k)} + c_2\sqrt{r(k)}(\epsilon_1^{(k)})^{1/3}(\kappa^{(k)} - 1 + \epsilon_1^{(k)})^{2/3} + c_3(\widetilde{\sigma}_{r(k)}^{(k)})^{-1}\kappa^{(k)}\sqrt{K^{(k)}m}\log m.$$

So

$$\begin{aligned} f^{(k)} &\leq C_1\sqrt{r}(\kappa^{(k)} - 1) + C_2(\kappa^{(k)} + \sqrt{r})(\|M^{(k)} - M_0\|_\infty + 6\sigma c\log m)^{1/2}\left(\frac{m}{p(\widetilde{\sigma}_{r(k)}^{(k)})^2}\right)^{1/2}\log m \\ &\lesssim \mathcal{O}(\log^{3/2} m). \end{aligned}$$

Denote that $\boldsymbol{X}^{(k)} = \cos\theta_x^{(k)}\widetilde{\boldsymbol{X}}^{(k)} + \sin\theta_x^{(k)}\widetilde{\boldsymbol{X}}^{(k),\perp}$ and $\boldsymbol{Y}^{(k)} = \cos\theta_y^{(k)}\widetilde{\boldsymbol{Y}}^{(k)} + \sin\theta_y^{(k)}\widetilde{\boldsymbol{Y}}^{(k),\perp}$, then

$$\begin{aligned} &\|\boldsymbol{M}^{(k)} - \eta_{k+1}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega}) - \boldsymbol{M}_0\|_F \\ \leq&\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0 - \eta_{k+1}\sum_{i=1}^{r(k)}\widetilde{\boldsymbol{x}}_i^{(k)}\widetilde{\boldsymbol{x}}_i^{(k)\prime}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega})\widetilde{\boldsymbol{y}}^{(k)}\widetilde{\boldsymbol{y}}_i^{(k)\prime}\|_F \\ +&\|\eta\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega}) - \eta_{k+1}\sum_{i=1}^{r(k)}\widetilde{\boldsymbol{x}}_i^{(k)}\widetilde{\boldsymbol{x}}_i^{(k)\prime}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega})\widetilde{\boldsymbol{y}}^{(k)}\widetilde{\boldsymbol{y}}_i^{(k)\prime}\|_F \\ =&\left(\sum_{i=1}^{r(k)}\left(\widetilde{\theta}_i^{(k)} - \frac{2\eta_{k+1}\omega_i}{\widehat{\sigma}^2}(\theta_i^{(k)} - \widehat{\sigma}\lambda_i)\cos\theta(\boldsymbol{x}_i^{(k)},\widetilde{\boldsymbol{x}}_i^{(k)})\cos\theta(\boldsymbol{y}_i^{(k)},\widetilde{\boldsymbol{y}}_i^{(k)})\right)^2 + \|\Delta\boldsymbol{M}^{(k)}\|_F^2\right)^{1/2} \\ +&\left(\sum_{i=1}^{r(k)}\left(\frac{2\eta_{k+1}\omega_i}{\widehat{\sigma}^2}(\theta_i^{(k)} - \widehat{\sigma}\lambda)\right)^2(1 - \cos^2\theta(\boldsymbol{x}_i^{(k)},\widetilde{\boldsymbol{x}}_i^{(k)})\cos^2\theta(\boldsymbol{y}_i^{(k)},\widetilde{\boldsymbol{y}}_i^{(k)}))\right. \\ &\left.\qquad\qquad + \sum_{i=r(k)+1}^{n}\left(\frac{2\eta_{k+1}w_i}{\widehat{\sigma}^2}(\theta_i^{(k)} - \widehat{\sigma}\lambda)\right)^2\right)^{1/2}. \end{aligned}$$

Without loss of generality, let $\omega_1 = 1$ and $\omega_i = 0,$ for $i = 2,\cdots,n$. This simplification does not affect generality, and can be extended to the general case where $\omega_1 \geq \log^{-2} m$. Then

$$\begin{aligned} &\|\boldsymbol{M}^{(k)} - \eta_{k+1}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega}) - \boldsymbol{M}_0\|_F \\ \leq&\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F\left(1 - \frac{2\eta_{k+1}}{\widehat{\sigma}^2}\frac{\theta_1^{(k)} - \widehat{\sigma}\lambda_1}{\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F}\left(1 - \frac{C(\rho^{-1/4}\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_\infty + \sigma)^2 mp}{p^2\widetilde{\theta}_1^2}\right)\right) \\ +&\frac{2\eta_{k+1}(\theta_1^{(k)} - \widehat{\sigma}\lambda_1)}{\widehat{\sigma}^2}\left(\frac{C(\rho^{-1/4}\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_\infty + \sigma)^2 mp}{p^2\widetilde{\theta}_1^2}\right)^{1/2} \\ \leq&\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F\left(1 - \frac{2\eta_{k+1}}{\widehat{\sigma}^2}\frac{\theta_1^{(k)} - \widehat{\sigma}\lambda_1}{\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F}\left(1 - \frac{2C(\rho^{-1/4}\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_\infty + \sigma)\sqrt{mp}}{p\widetilde{\theta}_1}\right)\right). \end{aligned}$$

Regard $\boldsymbol{M}^{(k+1)} = \mathrm{Proj}_r(\boldsymbol{M}^{(k)} - 2\eta_{k+1}/\widehat{\sigma}^2(\theta_1^{(k)} - \widehat{\sigma}\lambda_1)\widetilde{\boldsymbol{x}}_1^{(k)}\widetilde{\boldsymbol{y}}_1^{(k)\prime})$. With lemma S.13, denoting that

$$C_2^{(k)} = \|\boldsymbol{x}_1^{(k)} - \widetilde{\boldsymbol{x}}_1^{(k)}\|_2 \le C(\rho^{-1/4}\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_\infty + \sigma)\left(m/p\widetilde{\theta}^2_{r(k)}\right)^{1/2}$$
$$\le c(\sigma^{-1}\rho^{-1/4}\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_\infty + 1),$$

and

$$\left(\frac{(\sigma_r^{(k)}\kappa_r^{(k)})^2(2-2r(C_1^{(k)}/\sigma_r^{(k)})^{-2})^2}{(C_1^{(k)})^2} + 1\right)^{-1/2} \le \|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F^2/(\sigma_r^{(k)}\kappa_r^{(k)})^2.$$

When we set

$$\alpha_k = \left(5C_2^{(k)} + \left(\frac{(\sigma_r^{(k)}\kappa_r^{(k)})^2(2-2r(C_1^{(k)}/\sigma_r^{(k)})^{-2})^2}{(C_1^{(k)})^2} + 1\right)^{-1/2}\right),$$

and

$$\gamma_k = \frac{2\eta_{k+1}}{\widehat{\sigma}^2}\frac{(\theta_1^{(k)} - \widehat{\sigma}\lambda_1)}{\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F},$$

then

$$\begin{aligned}
\|\boldsymbol{M}^{(k+1)} - \boldsymbol{M}_0\|_F &\le \|\boldsymbol{M}^{(k+1)} - \boldsymbol{N}^{(k+1)}\|_F + \|\boldsymbol{N}^{(k+1)} - \boldsymbol{M}_0\|_F \\
&\le \|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F\left(1 - \frac{2\eta}{\widehat{\sigma}^2}\frac{\theta_1^{(k)} - \widehat{\sigma}\lambda_1}{\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F}(1- \right. \\
&\left.\left.\frac{5C(\rho^{-1/4}\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_\infty + \sigma)\sqrt{mp}}{p\widetilde{\theta}_1} - \|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F^2/(\sigma_r^{(k)}\kappa_r^{(k)})^2\right)\right) \\
&\le \|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F\left(1 - \gamma_k(1-\alpha_k)\right).
\end{aligned}$$

Denote that $\boldsymbol{N}^{(k+1)} = \sum_{i=1}^{r+1}\widetilde{\sigma}_i^{(k+1)}\widetilde{\boldsymbol{u}}_i^{(k+1)}\widetilde{\boldsymbol{v}}_i^{(k+1)\prime} = \sum_{i=1}^{r}\widetilde{\sigma}_i^{(k+1)}\widetilde{\boldsymbol{u}}_i^{(k+1)}\widetilde{\boldsymbol{v}}_i^{(k+1)\prime} + \boldsymbol{N}^{(k+1),\perp}$. Treat

$$\begin{aligned}
\boldsymbol{N}^{(k+1)} = \boldsymbol{M}^{(k)} - \eta\left(\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega}) - \mathcal{P}_{\boldsymbol{U}^{(k)},\perp}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega})\mathcal{P}_{\boldsymbol{V}^{(k)},\perp\prime}\right) \\
+\eta\mathcal{P}_{\boldsymbol{U}^{(k)},\perp}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega})\mathcal{P}_{\boldsymbol{V}^{(k)},\perp\prime}
\end{aligned}$$

as the perturbation for $\boldsymbol{M}^{(k)}$, from lemma S.13,

$$\begin{aligned}
&\|\mathcal{P}_{\boldsymbol{U}^{(k)},\perp}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega})\mathcal{P}_{\boldsymbol{V}^{(k)},\perp\prime}\|_F \\
\le&\frac{2\eta_{k+1}}{\widehat{\sigma}^2}(\theta_1^{(k)} - \widehat{\sigma}\lambda_1)\left(3C_3^{(k)} + \left(\frac{(\sigma_r^{(k)}\kappa_r^{(k)})^2(2-2r(C_4^{(k)}/\sigma_r^{(k)})^{-2})^2}{(C_4^{(k)})^2} + 1\right)^{-1/2}\right) \\
\le&\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F\gamma_k\alpha_k.
\end{aligned}$$

We then use incoherent condition to show that

$$\begin{aligned}
&[\mathcal{P}_{\boldsymbol{U}^{(k)},\perp}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega})\mathcal{P}_{\boldsymbol{V}^{(k)},\perp\prime}]_{i,j} \\
\le&\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F\gamma_k\left(\boldsymbol{x}_1^{(k)} - \langle\boldsymbol{x}_1^{(k)},\boldsymbol{u}_1^{(k)}\rangle\boldsymbol{u}_1^{(k)}\right)'\left(\boldsymbol{y}_1^{(k)} - \langle\boldsymbol{y}_1^{(k)},\boldsymbol{v}_1^{(k)}\rangle\boldsymbol{v}_1^{(k)}\right) \\
\le&\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F\gamma_k(\|\boldsymbol{x}_1^{(k)}\|_\infty + \|\boldsymbol{u}_1^{(k)}\|_\infty)(\|\boldsymbol{y}_1^{(k)}\|_\infty + \|\boldsymbol{v}_1^{(k)}\|_\infty).
\end{aligned}$$

So

$$\|\mathcal{P}_{\boldsymbol{U}^{(k)},\perp}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega})\mathcal{P}_{\boldsymbol{V}^{(k)},\perp'}\|_1 \le \|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_F\gamma_k\alpha_k n^{1/2}(\|\boldsymbol{y}_1^{(k)}\|_\infty+\|\boldsymbol{v}_1^{(k)}\|_\infty)$$
$$\|(\mathcal{P}_{\boldsymbol{U}^{(k)},\perp}\widehat{\nabla}\mathcal{L}(\boldsymbol{M}^{(k)};\boldsymbol{Y},\boldsymbol{\Omega})\mathcal{P}_{\boldsymbol{V}^{(k)},\perp'})'\|_1 \le \|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_F\gamma_k\alpha_k m^{1/2}(\|\boldsymbol{x}_1^{(k)}\|_\infty+\|\boldsymbol{u}_1^{(k)}\|_\infty).$$

Theorem 4 in Fan, Wang and Zhong (2018) shows that

$$\|\widetilde{\boldsymbol{U}}^{(k+1)}-\boldsymbol{U}^{(k)}\boldsymbol{R}_U\|_{2,\infty} \le C\|\boldsymbol{U}^{(k)}\|_{2,\infty}r^{7/2}\sigma_r^{(k)-1}\gamma_k\alpha_k m^{1/2}(\|\boldsymbol{x}_1^{(k)}\|_\infty+\|\boldsymbol{u}_1^{(k)}\|_\infty).$$

So

$$\|\boldsymbol{U}^{(k+1)}\|_{2,\infty} \le \|\boldsymbol{U}^{(k)}\|_{2,\infty}\left(1+Cr^4\|\boldsymbol{M}^{(k)}-\boldsymbol{M}_0\|_F\sigma_r^{(k)-1}\gamma_k\alpha_k m^{1/2}\|\boldsymbol{U}^{(k)}\|_{2,\infty})\log^{3/2}m\right).$$

Consider the perturbation $\boldsymbol{M}^{(k+1)}=\mathrm{Proj}_r(\boldsymbol{M}_0+(\boldsymbol{N}^{(k+1)}-\boldsymbol{M}_0))-\boldsymbol{M}_0$, then similar to proof in Lemma S.12, we have

$$\|\boldsymbol{M}^{(k+1)}-\boldsymbol{M}_0\|_\infty \le 2\|\boldsymbol{M}^{(k+1)}-\boldsymbol{M}_0\|_F\|\widetilde{\boldsymbol{U}}^{(k+1)}\|_{2,\infty}\|\widetilde{\boldsymbol{V}}_{2,\infty}\|_{2,\infty}.$$

□

PROOF OF LEMMA S.13. Define that $\boldsymbol{U}=\widehat{\boldsymbol{U}}\boldsymbol{L}_1+\boldsymbol{U}^\perp\boldsymbol{L}_2$ and $\boldsymbol{V}=\widehat{\boldsymbol{V}}\boldsymbol{R}_1+\boldsymbol{V}^\perp\boldsymbol{L}_2$, then

$$\begin{aligned}\boldsymbol{x}_1'(\boldsymbol{M}-\widehat{\boldsymbol{M}})\boldsymbol{y}_1 &= \boldsymbol{x}_1'\widehat{\boldsymbol{U}}(\boldsymbol{L}_1\boldsymbol{\Sigma}\boldsymbol{R}_1-\widehat{\boldsymbol{\Sigma}})\widehat{\boldsymbol{V}}'\boldsymbol{y}_1+\boldsymbol{x}_1'\widehat{\boldsymbol{U}}\boldsymbol{L}_1\boldsymbol{\Sigma}\boldsymbol{R}_2\boldsymbol{V}^{\perp'}\boldsymbol{y}_1\\ &\quad+\boldsymbol{x}_1'\boldsymbol{U}^\perp\boldsymbol{L}_2\boldsymbol{\Sigma}\boldsymbol{R}_1\widehat{\boldsymbol{V}}'\boldsymbol{y}_1+\boldsymbol{x}_1'\boldsymbol{U}^\perp\boldsymbol{L}_2\boldsymbol{\Sigma}\boldsymbol{R}_2\boldsymbol{V}^{\perp'}\boldsymbol{y}_1.\end{aligned}$$

Because $\boldsymbol{x}_1\in\mathrm{span}\{\widehat{\boldsymbol{U}},\boldsymbol{U}^\perp\}$ and $\boldsymbol{y}_1\in\mathrm{span}\{\widehat{\boldsymbol{V}},\boldsymbol{V}^\perp\}$, so

$$\|\boldsymbol{x}_1'\widehat{\boldsymbol{U}}\|_F^2+\|\boldsymbol{x}_1'\boldsymbol{U}^\perp\|_F^2=1,\|\boldsymbol{y}_1'\widehat{\boldsymbol{V}}\|_F^2+\|\boldsymbol{y}_1'\boldsymbol{V}^\perp\|_F^2=1$$

We know that

$$\begin{aligned}\|\widehat{\boldsymbol{M}}+\eta\boldsymbol{x}\boldsymbol{y}'-\mathrm{Proj}_r(\widehat{\boldsymbol{M}}+\eta\boldsymbol{x}\boldsymbol{y}')\|_F &= \sigma_{r+1}(\widehat{\boldsymbol{M}}+\eta\boldsymbol{x}\boldsymbol{y}')\\ &\le \eta\|\boldsymbol{x}'\boldsymbol{U}^\perp\|_F\|\boldsymbol{x}'\boldsymbol{U}^\perp\|_F\\ &\le \eta(\|\boldsymbol{x}-\boldsymbol{x}_1\|_F+\|\boldsymbol{x}_1'\boldsymbol{U}^\perp\|_F)(\|\boldsymbol{y}-\boldsymbol{y}_1\|_F+\|\boldsymbol{y}_1'\boldsymbol{V}^\perp\|_F)\\ &\le \eta(C_2+\|\boldsymbol{x}_1'\boldsymbol{U}^\perp\|_F)(C_2+\|\boldsymbol{y}_1'\boldsymbol{V}^\perp\|_F).\end{aligned}$$

From $\|\boldsymbol{M}-\widehat{\boldsymbol{M}}\|_F\le C_1$, we know that

$$\|\boldsymbol{U}'\widehat{\boldsymbol{U}}\|_F^2+\|\boldsymbol{V}'\widehat{\boldsymbol{V}}\|_F^2 \ge 2-\frac{r\|\boldsymbol{M}-\widehat{\boldsymbol{M}}\|_2^2}{\sigma_r^2}\ge 2-\frac{rC_1^2}{\sigma_r^2}.$$

So

$$\|\boldsymbol{L}_1\|_F^2+\|\boldsymbol{R}_1\|_F^2\ge 2-\frac{rC_1^2}{\sigma_r^2},\|\boldsymbol{L}_2\|_F^2+\|\boldsymbol{R}_2\|_F^2\le\frac{rC_1^2}{\sigma_r^2}.$$

Also we know

$$\begin{aligned}\|\boldsymbol{L}_1\boldsymbol{\Sigma}\boldsymbol{R}_1-\widehat{\boldsymbol{\Sigma}}\|_2 &\le \|\boldsymbol{\Sigma}-\widehat{\boldsymbol{\Sigma}}\|_2+\|\boldsymbol{L}_1\boldsymbol{\Sigma}\boldsymbol{R}_1-\boldsymbol{\Sigma}\|_2\\ &\le C_1+\sigma_1(2-\|\boldsymbol{L}_1\|_F^2-\|\boldsymbol{R}_1\|_F^2)^{1/2}\\ &\le (1+\kappa_r\sqrt{r})C_1.\end{aligned}$$

and

$$\|\boldsymbol{L}_2\boldsymbol{\Sigma}\boldsymbol{R}_2\|_2 \leq \sigma_1\|\boldsymbol{L}_2\|_F\|\boldsymbol{R}_2\|_F$$

$$\|\boldsymbol{L}_1\boldsymbol{\Sigma}\boldsymbol{R}_2\|_2 + \|\boldsymbol{L}_2\boldsymbol{\Sigma}\boldsymbol{R}_1\|_2 \geq r^{1/2}\sigma_r(\|\boldsymbol{L}_2\|_F + \|\boldsymbol{R}_2\|_F)\left(2 - \frac{2rC_1^2}{\sigma_r^2}\right)$$

Because $\boldsymbol{x}_1, \boldsymbol{y}_1$ are the top singular vectors of $\boldsymbol{M} - \widehat{\boldsymbol{M}}$, we have

$$\|\boldsymbol{x}_1'\boldsymbol{U}^\perp\|_F\|\boldsymbol{y}_1'\boldsymbol{V}^\perp\|_F \leq \left(\frac{\kappa_r^2\sigma_r^2(2-2rC_1^2\sigma_r^{-2})^2}{C_1^2} + 1\right)^{-1}.$$

Then

$$\|\widehat{\boldsymbol{M}} + \eta\boldsymbol{x}\boldsymbol{y}' - \mathrm{Proj}_r(\widehat{\boldsymbol{M}} + \eta\boldsymbol{x}\boldsymbol{y}')\|_F \leq \eta\left(3C_2 + \left(\frac{\kappa_r^2\sigma_r^2(2-2rC_1^2\sigma_r^{-2})^2}{C_1^2} + 1\right)^{-1/2}\right).$$

□

PROOF OF COROLLARY 4.4. When $\|\boldsymbol{M}_0\|_\infty \leq C\sigma$, the initial $\boldsymbol{M}^{(0)}$ is enough, so we consider the case that $\|\boldsymbol{M}_0\|_\infty \gg \sigma$. We first control the increasing of $\|\boldsymbol{U}^{(k)}\|_{2,\infty}$ and $\|\boldsymbol{V}^{(k)}\|_{2,\infty}$ by

$$\begin{aligned}
\|\boldsymbol{U}^{(k)}\|_{2,\infty} &\leq \|\boldsymbol{U}^{(0)}\|_{2,\infty}\prod_{i=0}^{k-1}\left(1 + C\mu_0 r^4\frac{\|\boldsymbol{M}^{(0)} - \boldsymbol{M}_0\|_F}{\sigma_r - \|\boldsymbol{M}^{(0)} - \boldsymbol{M}_0\|}\gamma_i\alpha_i\log^{3/2}m\right)\\
&\leq \|\boldsymbol{U}^{(0)}\|_{2,\infty}\prod_{i=0}^{k-1}\left(1 + C\mu_0^2\kappa^2 r^{11/2}(np)^{-1/2}\gamma_i\alpha_i\log^{3/2}m\right)\\
&\leq \|\boldsymbol{U}^{(0)}\|_{2,\infty}\exp\left\{C\mu_0^2\kappa^2 r^{11/2}\sum_{i=0}^{k-1}\gamma_i\alpha_i(np)^{-1/2}\log^{3/2}m\right\}.
\end{aligned}$$

When $\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F \geq C\kappa\sigma\sqrt{mr/p}$ for some large enough $C > 0$, it can be checked that $\alpha_k \leq \alpha^*$ for some $\alpha^* < 1/2$. If $\sigma_r \geq \sigma\sqrt{m/p}$, we can check that $f^{(0)}$ in Lemma 4.3 is bounded, i.e., there exist C_5 such that $f^{(0)} \leq C_5\log^{3/2}m$. So

$$\begin{aligned}
\|\boldsymbol{U}^{(k)}\|_{2,\infty} &\leq \|\boldsymbol{U}^{(0)}\|_{2,\infty}\left(1 + 2C\mu_0^2\kappa^2 r^{11/2}(np)^{-1/2}\log^{3/2}m\sum_{i=0}^{k-1}\gamma_i\right)\\
&\leq C_5\|\boldsymbol{U}\|_{2,\infty}\left(1 + 2C\mu_0^2\kappa^2 r^{11/2}(np)^{-1/2}\log^{3/2}m\sum_{i=0}^{k-1}\gamma_i\right)(1+\log^{3/2}m).
\end{aligned}$$

When $\sum_{i=0}^{k-1}\gamma_i \leq C_6\mu_0^{-2}\kappa^{-2}r^{-11/2}(np)^{1/2}\log^{3/2}m$, we know that

$$\|\boldsymbol{U}^{(k)}\|_{2,\infty} \leq 2C_5\|\boldsymbol{U}\|_{2,\infty}(1+\log^{3/2}m).$$

Similarly it holds $\|\boldsymbol{V}^{(k)}\|_{2,\infty} \leq 2C_5\|\boldsymbol{V}\|_{2,\infty}(1+\log^{3/2}m)$. Then we have

$$\begin{aligned}
\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_F &\leq \|\boldsymbol{M}^{(0)} - \boldsymbol{M}_0\|_F\left(1 - \frac{1}{4}\sum_{i=0}^{k-1}\gamma_i\right)\\
&\leq C(\rho^{-1/4}\|\boldsymbol{M}_0\|_\infty + \sigma)\kappa\sqrt{mr/p}\left(1 - \frac{1}{4}\sum_{i=0}^{k-1}\gamma_i\right),
\end{aligned}$$

TABLE 1

The Spectral norm error $\mathcal{E}_{sp}$ of proposed methods and corresponding baselines for $\sigma = 1$, observation probability $p = 0.2$, rank of $\boldsymbol{M}_0$ as $5, 10, 20$ and matrix size m from 200 to 800 with $n/m = 0.5$.

Rank	Estimator	m=200	m=300	m=400	m=500	m=600	m=700	m=800
r = 5	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	3.6164	3.2146	2.8639	2.7558	2.7238	2.6751	2.6568
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	3.9775	3.2868	2.9585	2.8330	2.7604	2.7558	2.7099
	Method 2: $\widetilde{\boldsymbol{M}}_{\text{nuc}}$	3.4224	2.9430	2.7407	2.6548	2.6232	2.6078	2.6085
	Baseline 2: $\widehat{\boldsymbol{M}}_{\text{nuc}}$	4.0011	3.2486	3.0007	2.8820	2.8113	2.7528	2.7090
r = 10	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	10.152	4.2386	3.4315	3.1461	3.0247	2.9254	2.8573
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	12.341	4.5312	3.5790	3.2445	3.0827	3.0089	2.9314
	Method 2: $\widetilde{\boldsymbol{M}}_{\text{nuc}}$	7.0500	4.1107	3.2822	3.0645	2.9224	2.8483	2.8208
	Baseline 2: $\widehat{\boldsymbol{M}}_{\text{nuc}}$	10.103	4.5285	3.6350	3.2978	3.1384	3.0054	2.9318
r = 20	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	9.2373	9.0893	6.4189	4.7241	3.9109	3.5694	3.4231
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	9.4738	11.814	8.6139	5.1034	4.1124	3.8285	3.5524
	Method 2: $\widetilde{\boldsymbol{M}}_{\text{nuc}}$	9.3505	8.1293	5.5021	4.1959	3.6838	3.4241	3.3716
	Baseline 2: $\widehat{\boldsymbol{M}}_{\text{nuc}}$	11.657	14.994	9.6236	6.2824	5.8162	4.9592	4.2519

and

$$\|\boldsymbol{M}^{(k)} - \boldsymbol{M}_0\|_\infty \le C\|\boldsymbol{U}\|_{2,\infty}\|\boldsymbol{V}\|_{2,\infty}(\rho^{-1/4}\|\boldsymbol{M}_0\|_\infty + \sigma)\kappa\sqrt{mr/p}\left(1 - \frac{1}{4}\sum_{i=0}^{k-1}\gamma_i\right)\log^3 m.$$

Then we can find some K satisfies that

$$1/4 \le \sum_{i=0}^{K-1}\gamma_i \le C_6\mu_0^{-2}\kappa^{-2}r^{-11/2}(np)^{1/2}\log^{3/2} m$$

with sufficient small step-size η_k and $p \ge \mu_0^4\kappa^4 r^{11} m\log^\alpha m$. Then $\|\boldsymbol{M}^{(K)} - \boldsymbol{M}_0\| \ge C\kappa\sigma\sqrt{m/p}$ is not hold anymore and thus

$$\|\boldsymbol{M}^{(K)} - \boldsymbol{M}_0\|_F \lesssim \kappa\sigma\sqrt{mr/p},$$
$$\|\boldsymbol{M}^{(K)} - \boldsymbol{M}_0\|_\infty \lesssim \kappa\sigma\mu_0\sqrt{r^3\log^3 m/np}.$$

□

S.7. Additional Numerical Results.

S.7.1. *The spectral and maximum norm errors in simulations.* We present the spectral norm error $\mathcal{E}_{sp}$ and the maximum norm error $\mathcal{E}_\infty$ for the proposed estimators and baselines across various simulation scenarios. Tables 1 and 2, as well as Figure 1 illustrate the spectral and maximum norm errors for different matrix sizes. Tables 3 and 4, along with Figure 2, display the spectral and maximum norm errors for the proposed Method 1 and Baseline 1 as the selected rank $s \ge r$ increases. Finally, Tables 5, 6, 7 and Figure 3 show the errors for proposed Method 1 and Baseline 1 under different noise levels.

TABLE 2

The Maximum norm error $\mathcal{E}_\infty$ of proposed methods and corresponding baselines for $\sigma = 1$, observation probability $p = 0.2$, rank of $\boldsymbol{M}_0$ as $5, 10, 20$ and matrix size m from 200 to 800 with $n/m = 0.5$.

Rank	Estimator	m=200	m=300	m=400	m=500	m=600	m=700	m=800
r = 5	Method 1: $\widetilde{M}_{\text{fac}}$	5.4840	4.0595	3.1306	2.7745	2.5459	2.3530	2.3230
	Baseline 1: $\widehat{M}_{\text{fac}}$	6.8792	4.5146	3.3296	2.8791	2.5952	2.5432	2.3115
	Method 2: $\widetilde{M}_{\text{nuc}}$	5.0670	3.6151	3.0609	2.7152	2.4953	2.3278	2.2609
	Baseline 2: $\widehat{M}_{\text{nuc}}$	6.8921	4.3188	3.4660	3.0155	2.7053	2.5351	2.3150
r = 10	Method 1: $\widetilde{M}_{\text{fac}}$	26.208	7.1464	5.0404	4.0922	3.6330	3.2837	3.0710
	Baseline 1: $\widehat{M}_{\text{fac}}$	31.113	8.0032	5.3792	4.2277	3.6947	3.4374	3.0834
	Method 2: $\widetilde{M}_{\text{nuc}}$	13.441	6.4329	4.6935	3.9230	3.5917	3.2375	3.0532
	Baseline 2: $\widehat{M}_{\text{nuc}}$	22.711	8.0118	5.5153	4.3735	3.8346	3.4204	3.0794
r = 20	Method 1: $\widetilde{M}_{\text{fac}}$	19.785	18.694	12.356	8.0614	6.1160	5.2333	4.6936
	Baseline 1: $\widehat{M}_{\text{fac}}$	20.352	24.228	18.713	9.3175	6.5457	5.8246	4.9744
	Method 2: $\widetilde{M}_{\text{nuc}}$	19.789	16.335	9.9773	7.1124	5.6855	5.0173	4.6038
	Baseline 2: $\widehat{M}_{\text{nuc}}$	23.773	34.253	23.498	11.936	8.3179	6.3343	5.1824

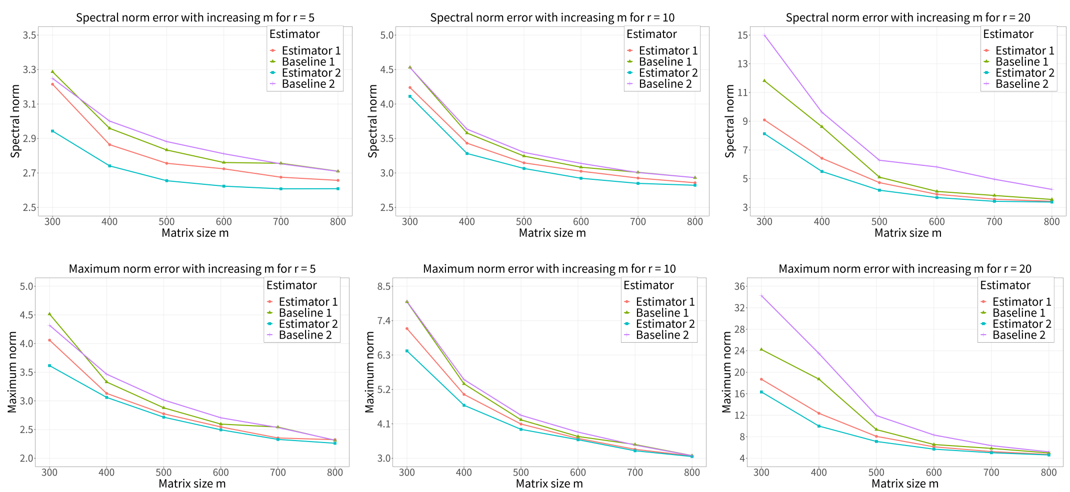

FIG 1. *The trend curve of the Spectral norm error $\mathcal{E}_{sp}$ and the Maximum norm error $\mathcal{E}_\infty$ for the proposed methods and baseline when matrix size approaches infinity. The top plots from left to right correspond to the spectral norm error in scenarios $r = 5, 10, 20$ and bottom plots correspond to the maximum norm error.*

TABLE 3

The Spectral norm error $\mathcal{E}_{sp}$ of proposed method 1 and corresponding baseline 1 for true rank of $\boldsymbol{M}_0$ as $5, 10, 20$, used rank s from r to $r + 6$ and matrix size 500×250.

Rank	Estimator	s = r	s = r+1	s = r+2	s = r+3	s = r+4	s = r+5	s = r+6
r = 5	Method 1: $\widetilde{M}_{\text{fac}}$	2.7558	3.3607	3.5294	3.6596	3.7617	3.8525	3.9245
	Baseline 1: $\widehat{M}_{\text{fac}}$	2.8330	3.9678	4.1876	4.3689	4.5215	4.6699	4.8235
r = 10	Method 1: $\widetilde{M}_{\text{fac}}$	3.1461	3.7383	3.9000	4.0422	4.1598	4.2755	4.3704
	Baseline 1: $\widehat{M}_{\text{fac}}$	3.2445	4.3699	4.6486	4.8993	5.1182	5.3356	5.5563
r = 20	Method 1: $\widetilde{M}_{\text{fac}}$	4.7241	4.8996	5.0187	5.1531	5.2604	5.3678	5.4581
	Baseline 1: $\widehat{M}_{\text{fac}}$	5.1034	6.0600	6.5084	6.8839	7.2079	7.4541	7.6527

TABLE 4

The Maximum norm error $\mathcal{E}_\infty$ of proposed method 1 and corresponding baseline 1 for true rank of $\boldsymbol{M}_0$ as $5, 10, 20$, used rank s from r to $r+6$ and matrix size 500×250.

Rank	Estimator	s = r	s = r+1	s = r+2	s = r+3	s = r+4	s = r+5	s = r+6
r = 5	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	2.7745	5.3443	5.8544	6.0946	6.3080	6.5313	6.6143
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	2.8791	4.3297	4.8918	5.5223	6.0566	6.5089	7.2443
r = 10	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	4.0922	6.5112	7.1861	7.7858	7.9974	8.3921	8.7817
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	4.2277	5.9000	6.8240	7.5652	8.3208	9.0423	9.8867
r = 20	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	8.0614	8.3689	8.8749	9.3711	9.8049	10.137	10.480
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	9.3175	10.493	11.567	12.842	13.710	14.420	14.918

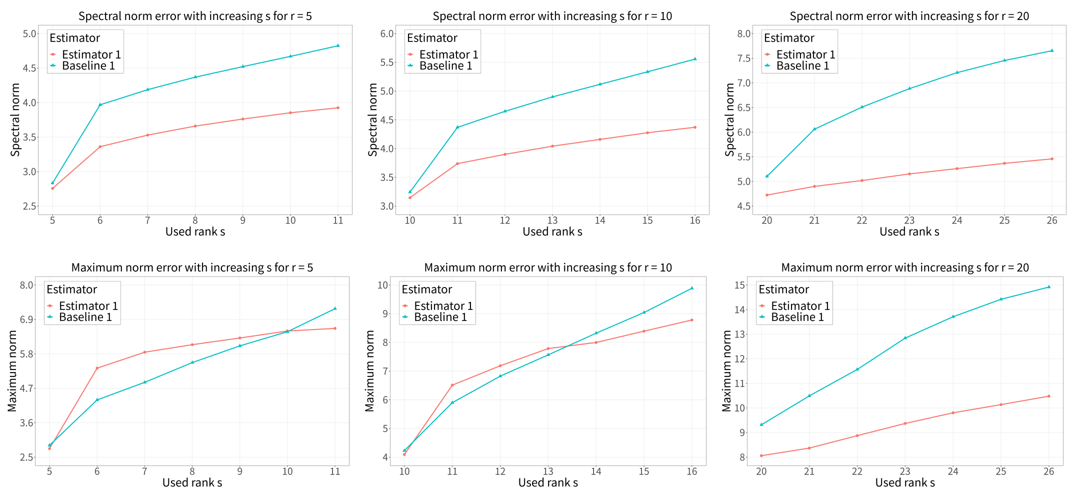

FIG 2. *The trend curve of the Spectral norm error $\mathcal{E}_{sp}$ and the Maximum norm error $\mathcal{E}_\infty$ for Method 1 and Baseline 1 when $s \geq r$ increases. The top plots from left to right correspond to the spectral norm error in scenarios $r = 5, 10, 20$ and bottom plots correspond to the maximum norm error.*

TABLE 5

The Frobenius norm error $\mathcal{E}_F$ of proposed method 1 and corresponding baseline 1 for noise level ratio R from 0.007 to 3.728, rank of $\boldsymbol{M}_0$ as $5, 10, 20$, and matrix size 500×250.

Rank	Estimator	R=0.007	R=0.038	R=0.150	R =0.745	R=1.118	R=2.236	R=3.728
r = 5	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	0.0418	0.0792	0.1401	0.4403	0.7184	1.8532	3.1496
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	0.0332	0.0388	0.0884	0.4487	0.7456	1.9943	3.3711
r = 10	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	0.1364	0.1544	0.2108	0.6764	1.1150	2.9067	5.0015
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	0.0533	0.0615	0.1366	0.6928	1.1831	3.3132	5.6503
r = 20	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	0.4044	0.4097	0.4632	1.2918	2.2253	4.9385	7.9604
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	0.1577	0.1662	0.2701	1.4555	2.8760	6.2395	10.245

S.7.2. *The spectral distributions of residual matrices in Netflix rating data.* Figure 4 displays the singular value distributions of the residual matrices for both our estimators and the baselines in Netflix rating data.

S.7.3. *The detail implementation of Amazon review data.* We choose the reviews for users about the goods in movies, books, electronics and automotive. We denote four scenarios for the four different categories of goods. In data preprocess, we select the most reviewed 1000 goods. For users selection, we randomly sample 3000 users from those with more than $14, 12, 12, 7$ reviews in scenarios 1 to 4 respectively. Figure 5 shows the singular value distributions of the observed rating matrices for each scenario. In all cases, the distribution of sin-

TABLE 6

The Spectral norm error $\mathcal{E}_{sp}$ of proposed method 1 and corresponding baseline 1 for noise level ratio R from 0.007 to 3.728, rank of $\boldsymbol{M}_0$ as $5, 10, 20$, and matrix size 500×250.

Rank	Estimator	R=0.007	R=0.038	R=0.150	R =0.745	R=1.118	R=2.236	R=3.728
r = 5	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	0.2813	0.5366	0.9624	2.7806	4.7128	12.789	22.514
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	0.2615	0.2890	0.5643	2.8839	4.9444	14.050	25.184
r = 10	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	0.7194	0.8208	1.1022	3.1768	5.4944	14.945	26.570
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	0.3348	0.3599	0.6512	3.3127	5.9508	18.268	33.094
r = 20	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	1.7929	1.8137	1.9924	4.7946	8.3744	19.265	32.086
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	0.7329	0.7501	1.0291	5.4629	11.520	26.814	47.490

TABLE 7

The Maximum norm error $\mathcal{E}_{\infty}$ of proposed method 1 and corresponding baseline 1 for noise level ratio R from 0.007 to 3.728, rank of $\boldsymbol{M}_0$ as $5, 10, 20$, and matrix size 500×250.

Rank	Estimator	R=0.007	R=0.038	R=0.150	R =0.745	R=1.118	R=2.236	R=3.728
r = 5	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	0.4458	0.8286	1.4800	2.7706	4.4942	13.379	28.391
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	0.6267	0.6383	0.7707	2.9169	5.1805	21.987	54.724
r = 10	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	1.3629	1.5458	2.0058	4.1128	6.8393	22.376	47.718
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	0.8745	0.8879	1.0883	4.3189	8.3873	45.266	104.00
r = 20	Method 1: $\widetilde{\boldsymbol{M}}_{\text{fac}}$	3.7737	3.8161	4.1130	8.0829	13.875	35.633	62.728
	Baseline 1: $\widehat{\boldsymbol{M}}_{\text{fac}}$	2.1135	2.1158	2.4135	10.090	24.468	75.605	167.57

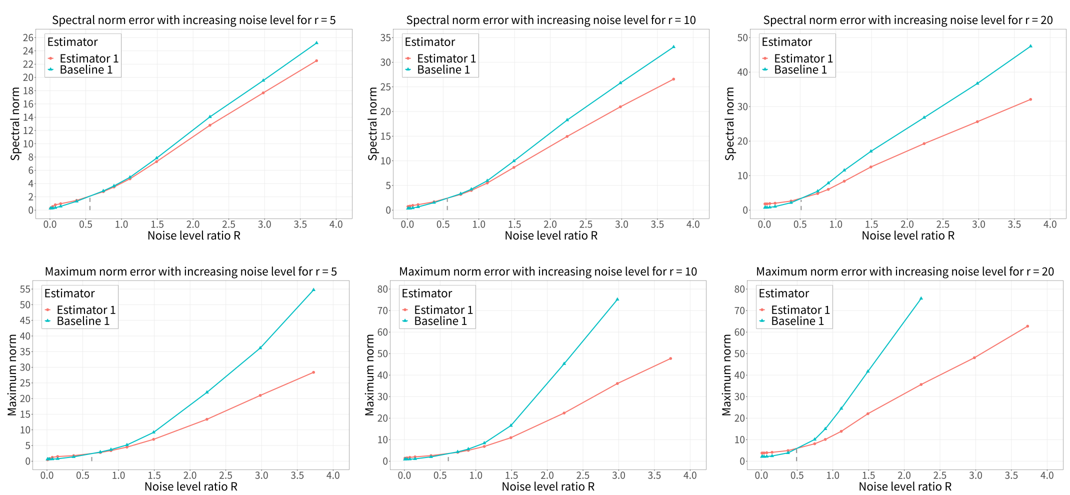

FIG 3. *The Spectral norm error $\mathcal{E}_{sp}$ and Maximum norm error $\mathcal{E}_{\infty}$ for Method 1 and Baseline 1 when noise level ratio R increases. The top plots from left to right correspond to the spectral norm error in scenarios $r = 5, 10, 20$ and bottom plots correspond to the maximum norm error.*

gular values aligns closely with the expected distribution of the corresponding sparse random matrices after variance adjustment, with only a few outliers due to low-rank perturbations.

The proposed and baseline estimators are implemented in the four scenarios. The data is divided into three parts: training, validation, and testing, with a ratio of $4:1:1$. Hyperparameters s in Estimator 1 and Baseline 1, and λ in Estimator 2 and Baseline 2, are tuned by minimizing the rooted mean squared error (RMSE) on the validation set. For scenarios 1 to 4, Estimator 1 and Baseline1 both choose $s = 1$. Given that ratings fall within the range $[1, 5]$, we truncate the estimated matrices to this interval, referring to them as modified versions. Figure 6 displays the singular value distributions of the residual matrices for both our estimators and the baselines.

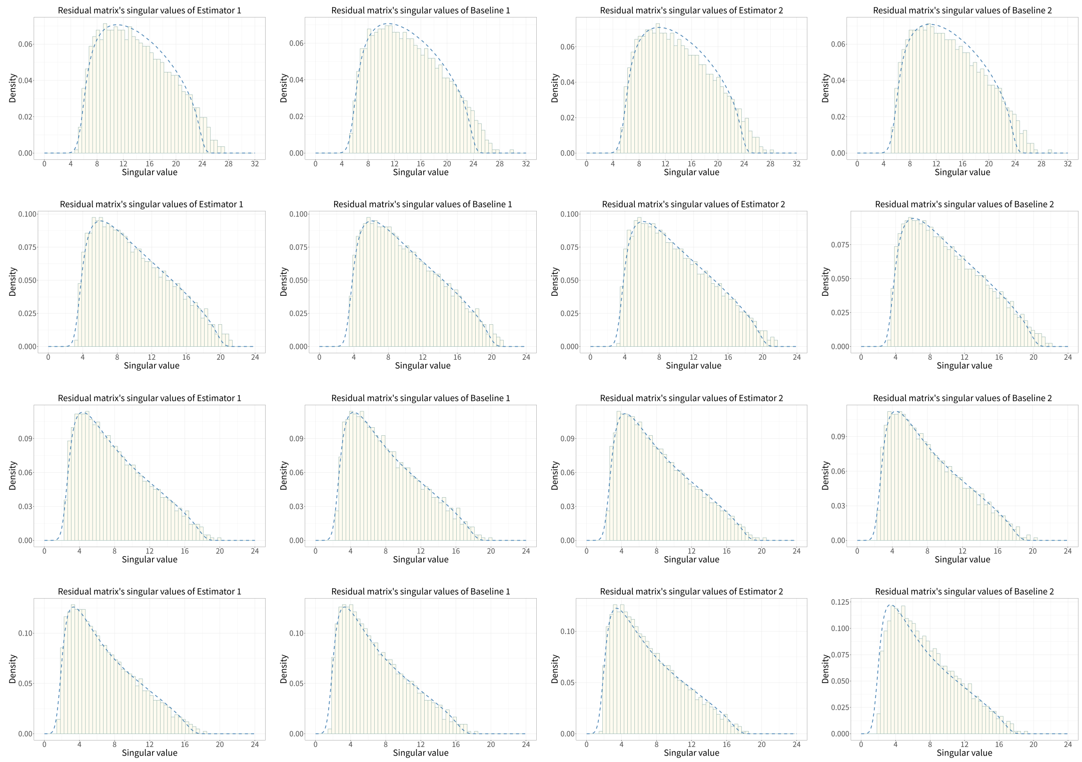

FIG 4. *The distribution of singular values of residual matrix for proposed and baseline methods for scenario 1 to 4. The left two plots correspond to the results of Estimator 1 and Baseline 1, and the right two plots correspond to the results of Estimator 2 and Baseline 2.*

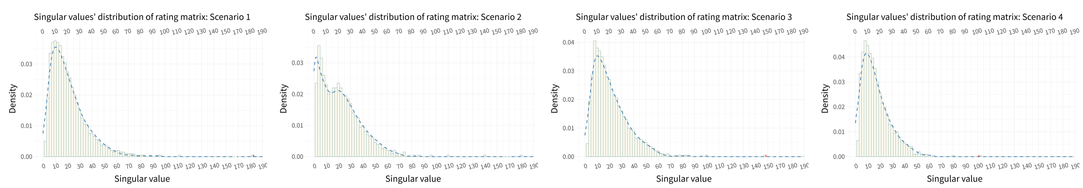

FIG 5. *The spectral distribution of the observed rating matrix and selected rank s in each scenarios respectively. The dash lines are densities of expectation of singular values of corresponding variance-adjusted sparse random matrices. The red lines shows the selected rank for Estimator 1.*

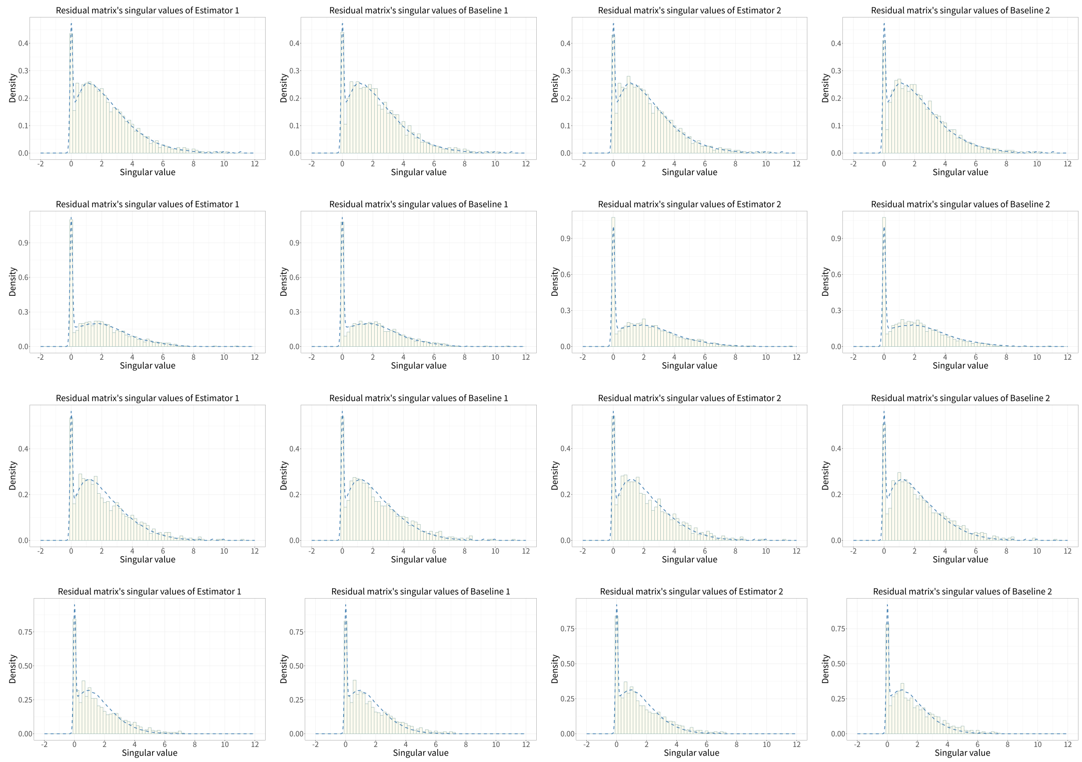

FIG 6. *The distribution of singular values of residual matrix for proposed and baseline methods for scenario 1 to 4. The left two plots correspond to the results of Estimator 1 and Baseline 1, and the right two plots correspond to the results of Estimator 2 and Baseline 2.*

REFERENCES

BAO, Z., PAN, G. and ZHOU, W. (2013). Central limit theorem for partial linear eigenvalue statistics of Wigner matrices. *Journal of Statistical Physics* **150** 88–129.

BHARDWAJ, A. and VU, V. (2024). Matrix Perturbation: Davis-Kahan in the Infinity Norm. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* 880–934. SIAM.

BRAILOVSKAYA, T. and VAN HANDEL, R. (2024). Universality and sharp matrix concentration inequalities. *Geometric and Functional Analysis* 1–105.

CACCIAPUOTI, C., MALTSEV, A. and SCHLEIN, B. (2013). Local Marchenko-Pastur law at the hard edge of sample covariance matrices. *Journal of Mathematical Physics* **54** 043302.

CANDÈS, E. J. and PLAN, Y. (2011). Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory* **57** 2342–2359.

ERDŐS, L., SCHLEIN, B., YAU, H.-T. and YIN, J. (2012). The local relaxation flow approach to universality of the local statistics for random matrices. In *Annales de l’IHP Probabilités et statistiques* **48** 1–46.

FAN, J., WANG, W. and ZHONG, Y. (2018). An l_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research* **18** 1–42.

KESHAVAN, R. H., MONTANARI, A. and OH, S. (2010). Matrix completion from a few entries. *IEEE transactions on information theory* **56** 2980–2998.

O’ROURKE, S. and SOSHNIKOV, A. (2015). Partial linear eigenvalue statistics for Wigner and sample covariance random matrices. *Journal of Theoretical Probability* **28** 726–744.

THOMPSON, R. (1975). Singular value inequalities for matrix sums and minors. *Linear Algebra and its Applications* **11** 251–269.