Asymptotic Theory for Estimating the Singular Vectors and Values of a Partially-observed Low Rank Matrix with Noise

University of Wisconsin-Madison

Supplementary Material

S1 Proofs for Lemma 1

Proof of Lemma 1. Let

$$M_y = [(y_{kh} - p)M_{0kh}]_{1\le k\le n, 1\le h\le d} \quad \text{and} \quad \epsilon_y = [y_{kh}\epsilon_{kh}]_{1\le k\le n, 1\le h\le d}\,,$$

both in $\mathbb{R}^{n\times d}$. Then, $M = pM_0 + M_y + \epsilon_y$ and

$$\begin{aligned}\hat{\Sigma} \;&=\; p^2 M_0^T M_0 + M_y^T M_y + \epsilon_y^T \epsilon_y + pM_0^T M_y + pM_y^T M_0 \\ &\quad + pM_0^T \epsilon_y + p\epsilon_y^T M_0 + M_y^T \epsilon_y + \epsilon_y^T M_y. \end{aligned} \tag{S1.1}$$

The result (2) follows since under the model setup in Section 2,

$$\mathbb{E}M_y = 0, \;\; \mathbb{E}\epsilon_y = 0, \;\; \mathbb{E}(M_y^T M_y) = p(1-p)\,\mathrm{diag}(M_0^T M_0), \;\; \mathbb{E}(\epsilon_y^T \epsilon_y) = np\sigma^2 I_d,$$

$$\mathbb{E}(M_0^T M_y) = 0, \quad \mathbb{E}(M_0^T \epsilon_y) = 0, \quad \text{and} \quad \mathbb{E}(M_y^T \epsilon_y) = 0.$$

We can similarly show the result (3). □

S2 Proofs for Appendix A.1

Proof of Proposition 2. We only show the result (A.1), since the other result can be shown similarly.

Let

$$E = \left\{ \frac{1}{nd} |\ddot{\lambda}_{p_{m+1}}^2 - \lambda_{p_{m+1}}^2| < t \right\},$$

where $t = C_1 p \frac{\log n}{d} + C_2 p^{3/2} \sqrt{\frac{\log n}{n}}$. Note that $\frac{t}{p^2} \to 0$. By Weyl's theorem (Li (1998a)) and Lemma 3, we have for large constants $C_1, C_2 > 0$,

$$\mathbb{P}(E^c) \le \mathbb{P}\left(\frac{1}{nd} \left\| \hat{\Sigma}_p - \mathbb{E}\hat{\Sigma}_p \right\|_2 \ge t \right) = O(n^{-2}).$$

Thus, for large n,

$$\begin{aligned}
&\mathbb{E} \left\| \sin\left(V_p^{(m)}, V^{(m)}\right) \right\|_F^2 \\
&= \mathbb{E}\left\{ \left\| \sin\left(V_p^{(m)}, V^{(m)}\right) \right\|_F^2 \mathbb{1}_{E^c} \right\} + \mathbb{E}\left\{ \left\| \sin\left(V_p^{(m)}, V^{(m)}\right) \right\|_F^2 \mathbb{1}_{E} \right\} \\
&\le m\, \mathbb{P}\left(E^c\right) + \mathbb{E}\left\{ \frac{\left\| \frac{1}{nd}\frac{1}{p^2} \left(\hat{\Sigma}_p - \mathbb{E}\,\hat{\Sigma}_p \right) V^{(m)} \right\|_F^2}{\left(\frac{1}{nd}\frac{1}{p^2} |\ddot{\lambda}_{p_m}^2 - \lambda_{p_{m+1}}^2| \right)^2} \mathbb{1}_E \right\} \\
&\le m\, \mathbb{P}\left(E^c\right) + \mathbb{E}\left\{ \frac{\left\| \frac{1}{nd}\frac{1}{p^2} \left(\hat{\Sigma}_p - \mathbb{E}\,\hat{\Sigma}_p \right) V^{(m)} \right\|_F^2 \mathbb{1}_E}{\left(\frac{1}{nd}\frac{1}{p^2} \left| \ddot{\lambda}_{p_m}^2 - \ddot{\lambda}_{p_{m+1}}^2 \right| - \frac{t}{p^2} \right)^2} \right\}
\end{aligned}$$

$$\leq m\,\mathbb{P}\left(E^c\right) + \mathbb{E}\left\{ \frac{\left\| \frac{1}{nd}\frac{1}{p^2}\left(\hat{\Sigma}_p - \mathbb{E}\,\hat{\Sigma}_p\right) V^{(m)} \right\|_F^2 \ \mathbb{1}_E}{\left(\frac{1}{2nd}\frac{1}{p^2}\left|\ddot{\lambda}_{p_m}^2 - \ddot{\lambda}_{p_{m+1}}^2\right|\right)^2} \right\}$$

$$\leq cn^{-2} + \frac{Cn^{-1}}{p\,(b_m^2 - b_{m+1}^2)^2}, \tag{S2.1}$$

where $\mathbb{1}_E$ is an indicator function of an event E, the first inequality is due to the fact that $\|\sin(\hat{V}^{(m)}, V_p^{(m)})\|_F^2 \leq m$ and Davis-Kahan $\sin\theta$ theorem (Theorem 3.1 in Li (1998b)), and the last inequality holds by Lemma S.1 below. □

Lemma S.1. *Under the model setup in Section 2 and Assumption 1, we have for large n and d,*

$$\mathbb{E}\left\| \frac{1}{nd}\ \frac{1}{p^2}\left(\hat{\Sigma}_p - \mathbb{E}\hat{\Sigma}_p\right) V^{(m)} \right\|_F^2 \leq \frac{C_1}{p\,n} \tag{S2.2}$$

and

$$\mathbb{E}\left\| \frac{1}{nd}\ \frac{1}{p^2}\left(\hat{\Sigma}_{pt} - \mathbb{E}\hat{\Sigma}_{pt}\right) U^{(m)} \right\|_F^2 \leq \frac{C_2}{p\,d},$$

where $\hat{\Sigma}_p$ and $\hat{\Sigma}_{pt}$ are defined in (4) and C_1 and C_2 are generic constants free of n, d, and p.

Proof of Lemma S.1. We only show the result (S2.2) because the other result holds similarly.

From (S1.1), (4), Proposition 1, and triangle inequality, we have

$$\left\| \left(\hat{\Sigma}_p - \mathbb{E}\hat{\Sigma}_p\right) V^{(m)} \right\|_F$$
$$\leq \left\| \left[M_y^T M_y - (1-p)\mathrm{diag}(M_y^T M_y) - p^2(1-p)\,\mathrm{diag}(M_0^T M_0) \right] V^{(m)} \right\|_F$$
$$+ \left\| \left[\epsilon_y^T \epsilon_y - (1-p)\mathrm{diag}(\epsilon_y^T \epsilon_y) - np^2\sigma^2 I_d \right] V^{(m)} \right\|_F$$

$$
\begin{aligned}
&+p\left\|\left[M_y^T M_0-(1-p)\operatorname{diag}(M_y^T M_0)\right]V^{(m)}\right\|_F \\
&+p\left\|\left[M_0^T M_y-(1-p)\operatorname{diag}(M_0^T M_y)\right]V^{(m)}\right\|_F \\
&+p\left\|\left[\epsilon_y^T M_0-(1-p)\operatorname{diag}(\epsilon_y^T M_0)\right]V^{(m)}\right\|_F \\
&+p\left\|\left[M_0^T \epsilon_y-(1-p)\operatorname{diag}(M_0^T \epsilon_y)\right]V^{(m)}\right\|_F \\
&+\left\|\left[M_y^T \epsilon_y-(1-p)\operatorname{diag}(M_y^T \epsilon_y)\right]V^{(m)}\right\|_F \\
&+\left\|\left[\epsilon_y^T M_y-(1-p)\operatorname{diag}(\epsilon_y^T M_y)\right]V^{(m)}\right\|_F \\
&=(A)+(B)+p\,(C)+p\,(D)+p\,(E)+p\,(F)+(G)+(H). \qquad \text{(S2.3)}
\end{aligned}
$$

We examine the convergence rates of the above terms, (A)-(H).

First, consider the term (A) in (S2.3). Then, we have

$$
\begin{aligned}
&\mathbb{E}\left\|\left[M_y^T M_y-(1-p)\operatorname{diag}(M_y^T M_y)-p^2(1-p)\operatorname{diag}(M_0^T M_0)\right]V^{(m)}\right\|_F^2 \\
&=\sum_{i=1}^{d}\sum_{j=1}^{m}\mathbb{E}\Bigg\{\sum_{k=1}^{n}\sum_{h=1}^{d}\Big[p\Big((y_{ki}-p)^2-p(1-p)\Big)M_{0ki}^2 V_{ji}\,\mathbb{1}_{(h=i)} \\
&\qquad\qquad +(y_{ki}-p)(y_{kh}-p)M_{0ki}M_{0kh}V_{jh}\,\mathbb{1}_{(h\neq i)}\Big]\Bigg\}^2 \\
&=\sum_{i=1}^{d}\sum_{j=1}^{m}\Bigg\{\sum_{k=1}^{n}\sum_{h=1}^{d}\Big[p^2\,\mathbb{E}\Big((y_{ki}-p)^2-p(1-p)\Big)^2 M_{0ki}^4 V_{ji}^2\,\mathbb{1}_{(h=i)} \\
&\qquad\qquad +\mathbb{E}\Big((y_{ki}-p)^2(y_{kh}-p)^2\Big)M_{0ki}^2 M_{0kh}^2 V_{jh}^2\,\mathbb{1}_{(h\neq i)}\Big]\Bigg\} \\
&=\sum_{i=1}^{d}\sum_{j=1}^{m}\Bigg\{\sum_{k=1}^{n}\sum_{h=1}^{d}\Big[p^3(1-p)(2p-1)^2 M_{0ki}^4 V_{ji}^2\,\mathbb{1}_{(h=i)} \\
&\qquad\qquad +p^2(1-p)^2 M_{0ki}^2 M_{0kh}^2 V_{jh}^2\,\mathbb{1}_{(h\neq i)}\Big]\Bigg\} \\
&\le p^2(1-p)L^4\sum_{i=1}^{d}\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{h=1}^{d}V_{jh}^2 \\
&=p^2(1-p)L^4\sum_{i=1}^{d}\sum_{j=1}^{m}\sum_{k=1}^{n}1
\end{aligned}
$$

$$\leq Cp^2(1-p)nd. \tag{S2.4}$$

Similarly to (S2.4), we can show that the expected values of the terms (B), (D), (F), (G), and (H) squared are bounded by Cp^2nd.

Second, consider the term (C) in (S2.3). Then, we have

$$\begin{aligned}
&\mathbb{E}\left\|\left[M_y^T M_0 - (1-p)\mathrm{diag}(M_y^T M_0)\right]V^{(m)}\right\|_F^2 \\
&= \sum_{i=1}^{d}\sum_{j=1}^{m}\mathbb{E}\left\{\sum_{k=1}^{n}(y_{ki}-p)\sum_{h=1}^{d}\left[p\,M_{0ki}^2 V_{ji}\,\mathbb{1}_{(h=i)}\right.\right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad \left.\left. + M_{0ki}M_{0kh}V_{jh}\,\mathbb{1}_{(h\neq i)}\right]\right\}^2 \\
&= \sum_{i=1}^{d}\sum_{j=1}^{m}\sum_{k=1}^{n}\mathbb{E}(y_{ki}-p)^2\left\{\sum_{h=1}^{d}M_{0ki}M_{0kh}V_{jh}\left[1-(1-p)\mathbb{1}_{(h=i)}\right]\right\}^2 \\
&= p(1-p)\sum_{i=1}^{d}\sum_{j=1}^{m}\sum_{k=1}^{n}\left\{\sum_{h=1}^{d}M_{0ki}M_{0kh}V_{jh}\left[1-(1-p)\mathbb{1}_{(h=i)}\right]\right\}^2 \\
&\leq p(1-p)L^4\sum_{i=1}^{d}\sum_{j=1}^{m}\sum_{k=1}^{n}\left\{\sum_{h=1}^{d}|V_{jh}|\right\}^2 \\
&\leq Cp(1-p)nd^2,
\end{aligned} \tag{S2.5}$$

where the last inequality holds due to Cauchy-Schwarz inequality.

Lastly, for the term (E) in (S2.3),

$$\begin{aligned}
&\mathbb{E}\left\|\left[\epsilon_y^T M_0 - (1-p)\mathrm{diag}(\epsilon_y^T M_0)\right]V^{(m)}\right\|_F^2 \\
&= \sum_{i=1}^{d}\sum_{j=1}^{m}\mathbb{E}\left\{\sum_{k=1}^{n}y_{ki}\epsilon_{ki}\sum_{h=1}^{d}M_{0kh}V_{jh}\left[1-(1-p)\mathbb{1}_{(h=i)}\right]\right\}^2 \\
&= \sum_{i=1}^{d}\sum_{j=1}^{m}\sum_{k=1}^{n}\mathbb{E}\left(y_{ki}^2\epsilon_{ki}^2\right)\left\{\sum_{h=1}^{d}M_{0kh}V_{jh}\left[1-(1-p)\mathbb{1}_{(h=i)}\right]\right\}^2
\end{aligned}$$

$$= p\sigma^2 \sum_{i=1}^{d}\sum_{j=1}^{m}\sum_{k=1}^{n}\left\{\sum_{h=1}^{d} M_{0kh}V_{jh}\Big[1-(1-p)\mathbb{1}_{(h=i)}\Big]\right\}^2$$
$$\le p\sigma^2 L^2 \sum_{i=1}^{d}\sum_{j=1}^{m}\sum_{k=1}^{n}\left\{\sum_{h=1}^{d}|V_{jh}|\right\}^2$$
$$\le Cpnd^2, \tag{S2.6}$$

where last inequality holds due to Cauchy-Schwarz inequality.

The result follows from (S2.4)-(S2.6). □

Lemma S2. *Under the model setup in Section 2 and Assumption 1, we have for any given* $\xi_1 > 0$,

$$\|M_y\|_2 \le C_{\xi_1}\sqrt{p\, n \log n}$$

with probability $1 - O(n^{-\xi_1})$. *Similarly, we have for any given* $\xi_2 > 0$,

$$\|\epsilon_y\|_2 \le C_{\xi_2}\sqrt{p\, n \log n}$$

with probability $1 - O(n^{-\xi_2})$.

Proof of Lemma S.2. Let $M_y^{(i,j)} \in \mathbb{R}^{n\times d}$ be such that

$$M_{ykh}^{(i,j)} = \begin{cases} (y_{kh}-p)M_{0kh}, & (k,h) = (i,j) \\ 0, & (k,h) \ne (i,j) \end{cases} \quad \text{for } 1 \le k \le n \text{ and } 1 \le h \le d.$$

Then,

$$\frac{1}{nd}M_y = \frac{1}{nd}\sum_{i=1}^{n}\sum_{j=1}^{d} M_y^{(i,j)},$$

$\mathbb{E}(M_y^{(i,j)}) = 0$, and $\|M_y^{(i,j)}\|_2 \leq L$ for all $1 \leq k \leq n$ and $1 \leq h \leq d$. Also, we have

$$\begin{aligned}\left\|\frac{1}{nd}\mathbb{E}\left(M_y^{(i,j)}M_y^{(i,j)T}\right)\right\|_2 &= \left\|\frac{p(1-p)}{nd}\text{diag}\left(M_0 M_0^T\right)\right\|_2 \leq \frac{p\,L^2}{n} \text{ and} \\ \left\|\frac{1}{nd}\mathbb{E}\left(M_y^{(i,j)T}M_y^{(i,j)}\right)\right\|_2 &= \left\|\frac{p(1-p)}{nd}\text{diag}\left(M_0^T M_0\right)\right\|_2 \leq \frac{p\,L^2}{d}. \end{aligned} \tag{S2.7}$$

Thus, by Proposition 1 in Koltchinskii, Lounici, and Tsybakov (2011), we have

$$\left\|\frac{1}{nd}M_y\right\|_2 \leq C \max\left(\sqrt{\frac{p\,L^2}{d}}\sqrt{\frac{\log n}{nd}}, L\frac{\log n}{nd}\right) \leq C\sqrt{\frac{p\,\log n}{nd^2}}$$

with probability at least $1 - n^{-\xi_1}$.

In a similar way together with Proposition 2 in Koltchinskii, Lounici, and Tsybakov (2011), we can show that $\left\|\frac{1}{nd}\epsilon_y\right\|_2 \leq C\sqrt{\frac{p\,\log n}{nd^2}}$ with probability at least $1-n^{-\xi_2}$.

□

Proof of Lemma 3. We only show the result (A.2) because the other result holds similarly.

From (S1.1), Proposition 1, and triangle inequality, we have

$$\begin{aligned}&\frac{1}{nd}\left\|\hat{\Sigma}_p - \mathbb{E}\hat{\Sigma}_p\right\|_2 \\ &\leq \frac{1}{nd}\left\|M_y^T M_y - (1-p)\text{diag}(M_y^T M_y) - p^2(1-p)\,\text{diag}(M_0^T M_0)\right\|_2 \\ &+\frac{1}{nd}\left\|\epsilon_y^T\epsilon_y - (1-p)\text{diag}(\epsilon_y^T\epsilon_y) - np^2\sigma^2 I_d\right\|_2 \\ &+2\,\frac{1}{nd}\left\|pM_y^T M_0 - (1-p)p\text{diag}(M_y^T M_0)\right\|_2 \\ &+2\,\frac{1}{nd}\left\|p\epsilon_y^T M_0 - (1-p)p\text{diag}(\epsilon_y^T M_0)\right\|_2\end{aligned}$$

$$
\begin{aligned}
&+2\,\frac{1}{nd}\left\|M_y^T\epsilon_y-(1-p)\mathrm{diag}(M_y^T\epsilon_y)\right\|_2\\
&=(I)+(II)+2\,(III)+2\,(IV)+2\,(V). \qquad \text{(S2.8)}
\end{aligned}
$$

Because of similarity, we provide arguments only for (I) and (IV).

Consider the term (I) in (S2.8). First, we have by Lemma S.2

$$
\frac{1}{nd}\left\|M_y^T M_y\right\|_2 = nd\left\|\frac{1}{nd}M_y\right\|_2^2 \le Cp\frac{\log n}{d} \qquad \text{(S2.9)}
$$

with probability at least $1-O(n^{-\mu_1})$. Also, we have with probability at least $1-O(n^{-\mu_1})$,

$$
\begin{aligned}
&\frac{1-p}{nd}\left\|\mathrm{diag}(M_y^T M_y)+p^2\,\mathrm{diag}(M_0^T M_0)\right\|_2\\
&\le \frac{1-p}{nd}\left\|\mathrm{diag}(M_y^T M_y)-p(1-p)\,\mathrm{diag}(M_0^T M_0)\right\|_2\\
&\qquad+\frac{p(1-p)}{nd}\left\|\mathrm{diag}(M_0^T M_0)\right\|_2\\
&=(1-p)\max_{1\le h\le d}\left|\sum_{k=1}^{n}\frac{\left[(y_{kh}-p)^2-p(1-p)\right]M_{0kh}^2}{nd}\right|\\
&\qquad+\frac{p(1-p)}{nd}\max_{1\le h\le d}\left|\sum_{k=1}^{n}M_{0kh}^2\right|\\
&\le C\sqrt{\frac{p\log n}{n}}\frac{1}{d}+\frac{p(1-p)L^2}{d}\\
&\le Cpd^{-1}, \qquad \text{(S2.10)}
\end{aligned}
$$

where the second inequality holds by (S2.11) below. Take $t^2=c\frac{\log n}{nd^2}p(1-p)(3p^2-$

$3p+1)$ for some large constant $c>0$. Then, by Bernstein's inequality,

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le h\le d}\left|\sum_{k=1}^{n}\frac{\left[(y_{kh}-p)^2-p(1-p)\right]M_{0kh}^2}{nd}\right|\ge t\right)\\
&\le \sum_{h=1}^{d}\mathbb{P}\left(\left|\sum_{k=1}^{n}\left[(y_{kh}-p)^2-p(1-p)\right]M_{0kh}^2\right|\ge ndt\right)\\
&\le 2d\ \exp\left\{-\frac{nd^2t^2}{2L^4p(1-p)(3p^2-3p+1)}\right\}\\
&= Cn^{-\mu_1}. \qquad\qquad (S2.11)
\end{aligned}$$

By (S2.9) and (S2.10), we have

$$(I)\le Cp\frac{\log n}{d} \qquad\qquad (S2.12)$$

with probability at least $1-O(n^{-\mu_1})$. Similarly, we can show that (II) and (V) are bounded by $Cp\frac{\log n}{d}$ with probability at least $1-O(n^{-\mu_1})$.

Consider the term (IV) in (S2.8). We have

$$(IV)^2\le\left\{\max_{1\le j\le d}\sum_{i=1}^{d}\left|\sum_{k=1}^{n}X_{kij}^{(IV)}\right|\right\}\left\{\max_{1\le i\le d}\sum_{j=1}^{d}\left|\sum_{k=1}^{n}X_{kij}^{(IV)}\right|\right\},$$

where $nd\,X_{kij}^{(IV)}=p\,y_{ki}\epsilon_{ki}M_{0kj}\mathbb{1}_{(i\ne j)}+p^2y_{ki}\epsilon_{ki}M_{0kj}\mathbb{1}_{(i=j)}$ and hence $X_{kij}^{(IV)}$ are centered sub-Gaussian random variables under the model setup in Section 2. Then, we have for any $\rho\in\mathbb{R}$ and for all $1\le k\le n$, $1\le i\le d$, and $1\le j\le d$,

$$\mathbb{E}\exp\left\{\rho X_{kij}^{(IV)}\right\}\le\exp\left\{\frac{\rho^2\frac{p^3\beta}{n^2d^2}}{2}\right\}\ \text{ for some constant }\beta>0.$$

Take $t^2 = cp^3 \frac{\log n}{n}$ for some large constant $c > 0$ and $\rho = \frac{t/d}{n \frac{p^3 \beta}{n^2 d^2}}$. Then, by Markov's inequality,

$$\begin{aligned}
\mathbb{P}\left(\max_{1 \le j \le d} \sum_{i=1}^{d} \left| \sum_{k=1}^{n} X_{kij}^{(IV)} \right| > t \right) &\le \sum_{j=1}^{d} \sum_{i=1}^{d} \mathbb{P}\left(\left| \sum_{k=1}^{n} X_{kij}^{(IV)} \right| > t/d \right) \\
&\le 2 \sum_{j=1}^{d} \sum_{i=1}^{d} \frac{\mathbb{E}\left(\exp\left\{ \rho \sum_{k=1}^{n} X_{kij}^{(IV)} \right\} \right)}{\exp\left\{ \rho (t/d) \right\}} \\
&\le 2d^2 \exp\left\{ -\rho \frac{t}{d} + \frac{\rho^2}{2} \frac{p^3 \beta}{n d^2} \right\} \\
&= 2d^2 \exp\left\{ -\frac{n t^2}{2 p^3 \beta} \right\} \\
&= C n^{-\mu_1}. \qquad \text{(S2.13)}
\end{aligned}$$

Similarly,

$$\mathbb{P}\left(\max_{1 \le i \le d} \sum_{j=1}^{d} \left| \sum_{k=1}^{n} X_{kij}^{(IV)} \right| > t \right) \le C n^{-\mu_1}. \qquad \text{(S2.14)}$$

By (S2.13) and (S2.14), with probability at least $1 - O\left(n^{-\mu_1}\right)$,

$$|(IV)| \le C p^{3/2} \sqrt{\frac{\log n}{n}}. \qquad \text{(S2.15)}$$

Similarly, we can show that (III) is bounded by $Cp^{3/2}\sqrt{\frac{\log n}{n}}$ with probability at least $1 - O(n^{-\mu_1})$.

The statement is showed by (S2.12) and (S2.15). □

Proof of Lemma 4. We only show the result (A.3) because the other result holds similarly.

By triangle inequality, we have

$$\begin{aligned}
&\frac{1}{nd}\left\|\hat{\Sigma}_{\hat{p}}-\hat{\Sigma}_{p}\right\|_2=\frac{1}{nd}\left\|(\hat{p}-p)\operatorname{diag}(\hat{\Sigma})\right\|_2\\
&\leq\frac{|\hat{p}-p|}{nd}\Big\{\|\operatorname{diag}(\hat{\Sigma})-\operatorname{diag}(pM_0^TM_0+np\sigma^2I_d)\|_2\\
&\qquad\qquad+\|\operatorname{diag}(pM_0^TM_0+np\sigma^2I_d)\|_2\Big\}. \qquad (S2.16)
\end{aligned}$$

We will look at the terms in (S2.16) one by one.

By Bernstein's inequality, we have for large constant $C>0$,

$$\begin{aligned}
&\mathbb{P}\left(|\hat{p}-p|\geq C\sqrt{\frac{p(1-p)\log n}{nd}}\right)\\
&=\mathbb{P}\left(\left|\sum_{k=1}^{n}\sum_{h=1}^{d}(y_{kh}-p)\right|\geq C\sqrt{p(1-p)nd\log n}\right)\\
&\leq 2\,\exp\left\{-\nu_1\log n\right\}\\
&=2n^{-\nu_1}. \qquad (S2.17)
\end{aligned}$$

Take $t^2=c\frac{p\log n}{nd^2}$ for some large constant $c>0$. Then, since $y_{ki}^2(M_{0ki}+\epsilon_{ki})^2-p(M_{0ki}^2+\sigma^2), k=1,\dots,n$, are independent centered sub-exponential random variables, we have by Proposition 5.16 in Vershynin (2010),

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{nd}\left\|\operatorname{diag}(\hat{\Sigma})-\operatorname{diag}(pM_0^TM_0+np\sigma^2I_d)\right\|_2\geq t\right)\\
&=\mathbb{P}\left(\frac{1}{nd}\max_{1\leq i\leq d}\left|\sum_{k=1}^{n}\left[y_{ki}^2(M_{0ki}+\epsilon_{ki})^2-p(M_{0ki}^2+\sigma^2)\right]\right|\geq t\right)\\
&\leq\sum_{i=1}^{d}\mathbb{P}\left(\left|\sum_{k=1}^{n}\left[y_{ki}^2(M_{0ki}+\epsilon_{ki})^2-p(M_{0ki}^2+\sigma^2)\right]\right|\geq ndt\right)
\end{aligned}$$

$$\leq 2d \ \exp\left\{-\frac{n^2d^2t^2}{c_1 np}\right\}$$
$$\leq Cn^{-\nu_1}. \tag{S2.18}$$

Also, note that

$$\begin{aligned}\left\|\frac{1}{nd}\mathrm{diag}(pM_0^TM_0 + np\sigma^2 I_d)\right\|_2 &= \frac{1}{nd}\max_{1\leq i\leq d} p\sum_{k=1}^n {M_0}_{ki}^2 + np\sigma^2 \\ &\leq \frac{p(L^2+\sigma^2)}{d}.\end{aligned} \tag{S2.19}$$

Combining the results in (S2.16)-(S2.19), we have

$$\frac{1}{nd}\left\|\hat{\Sigma}_{\hat{p}} - \hat{\Sigma}_p\right\|_2 \leq Cp^{3/2}\sqrt{\frac{\log n}{nd}}\ \frac{1}{d} \tag{S2.20}$$

with probability at least $1 - O\left(n^{-\nu_1}\right)$. □

Proof of Lemma 5. We only show the result (A.4) because the other result holds similarly.

We have

$$\begin{aligned}&\mathbb{E}\left\|\frac{1}{nd}\left(\hat{\Sigma}_{\hat{p}} - \hat{\Sigma}_p\right)V^{(m)}\right\|_F^2 \\ &\leq m\ \mathbb{E}\left\|\frac{1}{nd}\left(\hat{\Sigma}_{\hat{p}} - \hat{\Sigma}_p\right)\right\|_2^2 \\ &\leq m\ \mathbb{E}\left\{(\hat{p}-p)^2\left\|\frac{1}{nd}\mathrm{diag}(\hat{\Sigma})\right\|_2^2\right\} \\ &\leq 4m\ \mathbb{E}\left\{(\hat{p}-p)^2\left\|\frac{1}{nd}\mathrm{diag}(\hat{\Sigma}) - \frac{1}{nd}\mathrm{diag}(pM_0^TM_0 + np\sigma^2 I_d)\right\|_2^2\right\}\end{aligned}$$

$$
\begin{aligned}
&\quad +4m \left\| \frac{1}{nd}\mathrm{diag}(pM_0^T M_0 + np\sigma^2 I_d) \right\|_2^2 \; \mathbb{E}\left(\hat{p}-p\right)^2 \\
&\leq 4m \sqrt{\mathbb{E}\left(\hat{p}-p\right)^4 \; \mathbb{E}\left\| \frac{1}{nd}\mathrm{diag}(\hat{\Sigma}) - \frac{1}{nd}\mathrm{diag}(pM_0^T M_0 + np\sigma^2 I_d) \right\|_2^4} \\
&\quad +4m \, \frac{p^2(L^2+\sigma^2)^2}{d^2}\frac{p(1-p)}{nd} \\
&\leq C_1 \frac{p^2(1-p)}{n^2 d^{5/2}} + C_2 \frac{p^3(1-p)}{nd^3},
\end{aligned} \tag{S2.21}
$$

where the fourth inequality holds by Hölder's inequality and the fifth inequality is due to the fact that

$$
\begin{aligned}
&\mathbb{E}\left(\hat{p}-p\right)^4 \; \mathbb{E}\left\| \frac{1}{nd}\mathrm{diag}(\hat{\Sigma}) - \frac{1}{nd}\mathrm{diag}(pM_0^T M_0 + np\sigma^2 I_d) \right\|_2^4 \\
&\leq \frac{\mathbb{E}\big\{ \sum_{k=1}^n \sum_{h=1}^d (y_{kh}-p) \big\}^4}{n^4 d^4} \\
&\qquad \frac{\mathbb{E}\left\{ \max_{1\leq i\leq d} \left| \sum_{k=1}^n \left[y_{ki}^2 (M_{0ki}+\epsilon_{ki})^2 - p(M_{0ki}^2+\sigma^2) \right] \right|^4 \right\}}{n^4 d^4} \\
&\leq \frac{\mathbb{E}\big\{ \sum_{k=1}^n \sum_{h=1}^d (y_{kh}-p) \big\}^4}{n^4 d^4} \\
&\qquad \frac{d\,\mathbb{E}\left[\sum_{k=1}^n \left(y_{ki}^2 (M_{0ki}+\epsilon_{ki})^2 - p(M_{0ki}^2+\sigma^2) \right) \right]^4}{n^4 d^4} \\
&= \frac{O\left(p^2(1-p)^2 n^2 d^2\right) \; O\left(p^2 n^2 d\right)}{n^8 d^8}.
\end{aligned} \tag{S2.22}
$$

□

S3 Proofs for Appendix A.2

Lemma S.3. *Under the model setup in Section 2 and Assumption 1, we have*

$$
\begin{aligned}
&\sum_{i=1}^{m} \lambda_{p_i}^2 - p^2 \left[\sum_{i=1}^{m} \lambda_i^2 + n\sigma^2 \right] \\
&= 2p \sum_{k=1}^{n} \sum_{h=1}^{d} (y_{kh} - p) M_{0kh} \sum_{i=1}^{m} V_{ih} \left[\left(\sum_{h'=1}^{d} M_{0kh'} V_{ih'} \right) - (1-p) M_{0kh} V_{ih} \right] \\
&\quad + 2p \sum_{k=1}^{n} \sum_{h=1}^{d} y_{kh} \epsilon_{kh} \sum_{i=1}^{m} V_{ih} \left[\left(\sum_{h'=1}^{d} M_{0kh'} V_{ih'} \right) - (1-p) M_{0kh} V_{ih} \right] \\
&\quad + o_p\left(\sqrt{nd}\right) \\
&= (i) + (ii) + O_p\left(p\sqrt{n} + pd\right)
\end{aligned}
$$

and $(i)+(ii) = O_p\left(\sqrt{p^3 nd}\right)$*, where* λ_{pi} *and* λ_i *are defined in (4) and (1), respectively.*

Proof of Lemma S.3. We have

$$
\begin{aligned}
&\sum_{i=1}^{m} \lambda_{p_i}^2 - p^2 \left[\sum_{i=1}^{m} \lambda_i^2 + n\sigma^2 \right] \\
&= \mathrm{tr}({V_p}^{(m)T} \hat{\Sigma}_p {V_p}^{(m)}) - \mathrm{tr}\left(V^{(m)T} (p^2 M_0^T M_0 + np^2\sigma^2 I_d) V^{(m)} \right) \\
&= \mathrm{tr}(\mathcal{O}^T V^{(m)T} \hat{\Sigma}_p V^{(m)} \mathcal{O}) \\
&\qquad + \mathrm{tr}({V_p}^{(m)T} \hat{\Sigma}_p {V_p}^{(m)} - \mathcal{O}^T V^{(m)T} \hat{\Sigma}_p V^{(m)} \mathcal{O}) \\
&\qquad - \mathrm{tr}\left(V^{(m)T} (p^2 M_0^T M_0 + np^2\sigma^2 I_d) V^{(m)} \right) \\
&= \mathrm{tr}(V^{(m)T} \hat{\Sigma}_p V^{(m)}) + \mathrm{tr}({V_p}^{(m)T} \hat{\Sigma}_p {V_p}^{(m)} - \mathcal{O}^T V^{(m)T} \hat{\Sigma}_p V^{(m)} \mathcal{O}) \\
&\qquad - \mathrm{tr}\left(V^{(m)T} (p^2 M_0^T M_0 + np^2\sigma^2 I_d) V^{(m)} \right) \\
&= \mathrm{tr}\Big(V^{(m)T} (M_y^T M_y - (1-p)\mathrm{diag}(M_y^T M_y) \\
&\qquad\qquad - p^2(1-p)\,\mathrm{diag}(M_0^T M_0)) V^{(m)} \Big)
\end{aligned}
$$

$$
\begin{aligned}
&+\mathrm{tr}\left(V^{(m)T}(\epsilon_y^T\epsilon_y-(1-p)\mathrm{diag}(\epsilon_y^T\epsilon_y)-np^2\sigma^2 I_d)V^{(m)}\right)\\
&+\mathrm{tr}\Big(V^{(m)T}(pM_0^TM_y+pM_y^TM_0\\
&\qquad\qquad -(1-p)p\mathrm{diag}(M_0^TM_y+M_y^TM_0))V^{(m)}\Big)\\
&+\mathrm{tr}\Big(V^{(m)T}(pM_0^T\epsilon_y+p\epsilon_y^TM_0\\
&\qquad\qquad -(1-p)p\mathrm{diag}(M_0^T\epsilon_y+\epsilon_y^TM_0))V^{(m)}\Big)\\
&+\mathrm{tr}\Big(V^{(m)T}(M_y^T\epsilon_y+\epsilon_y^TM_y\\
&\qquad\qquad -(1-p)\mathrm{diag}(M_y^T\epsilon_y+\epsilon_y^TM_y))V^{(m)}\Big)\\
&+\mathrm{tr}({V_p}^{(m)T}\hat{\Sigma}_p{V_p}^{(m)}-\mathcal{O}^TV^{(m)T}\hat{\Sigma}_pV^{(m)}\mathcal{O})\\
&=(a)+(b)+(c)+(d)+(e)+(f),
\end{aligned}
\tag{S3.1}
$$

where $\mathcal{O}\in\mathbb{V}_{m,m}$ is a solution to $\inf_{\mathcal{Q}\in\mathbb{V}_{m,m}}\|V_p^{(m)}-V^{(m)}\mathcal{Q}\|_F^2$ and the fourth equality holds by (4) and (S1.1). Below, we examine the six terms (a)-(f) one by one.

The term (a) in (S3.1) is

$$
\begin{aligned}
(a) &= \sum_{i=1}^m V_i^T\left(M_y^TM_y-(1-p)\mathrm{diag}(M_y^TM_y)-p^2(1-p)\,\mathrm{diag}(M_0^TM_0)\right)V_i\\
&= \sum_{i=1}^m\Bigg\{\sum_{k=1}^n\left(\sum_{h=1}^d(y_{kh}-p)M_{0kh}V_{ih}\right)^2\\
&\qquad -(1-p)\sum_{k=1}^n\sum_{h=1}^d(y_{kh}-p)^2M_{0kh}^2V_{ih}^2\\
&\qquad -p^2(1-p)\sum_{k=1}^n\sum_{h=1}^d M_{0kh}^2V_{ih}^2\Bigg\}\\
&= \sum_{k=1}^n\sum_{h=1}^d p\Big[(y_{kh}-p)^2-p(1-p)\Big]M_{0kh}^2\sum_{i=1}^m V_{ih}^2\\
&\qquad +2\sum_{k=1}^n\sum_{h<h'}^{1\sim d}(y_{kh}-p)(y_{kh'}-p)M_{0kh}M_{0kh'}\sum_{i=1}^m V_{ih}V_{ih'}.
\end{aligned}
\tag{S3.2}
$$

Note that the two terms in (S3.2) are centered and uncorrelated with each other. So, the variance is

$$\begin{aligned}
\text{var}(a) &= \left\{ \sum_{k=1}^{n} \sum_{h=1}^{d} p^3(1-p)(2p-1)^2 M_{0kh}^4 \left(\sum_{i=1}^{m} V_{ih}^2 \right)^2 \right\} \\
&\qquad + \left\{ 4 \sum_{k=1}^{n} \sum_{h<h'}^{1\sim d} p^2(1-p)^2 M_{0kh}^2 M_{0kh'}^2 \left(\sum_{i=1}^{m} V_{ih} V_{ih'} \right)^2 \right\} \\
&\le m \sum_{i=1}^{m} \sum_{k=1}^{n} \sum_{h=1}^{d} p^3(1-p)(2p-1)^2 M_{0kh}^4 V_{ih}^4 \\
&\quad +4m \sum_{i=1}^{m} \sum_{k=1}^{n} \sum_{h<h'}^{1\sim d} p^2(1-p)^2 M_{0kh}^2 M_{0kh'}^2 V_{ih}^2 V_{ih'}^2 \\
&\le mL^4 p^3(1-p)(2p-1)^2 \sum_{i=1}^{m} \sum_{k=1}^{n} \sum_{h=1}^{d} V_{ih}^4 \\
&\quad +4mL^4 p^2(1-p)^2 \sum_{i=1}^{m} \sum_{k=1}^{n} \sum_{h,h'}^{1\sim d} V_{ih}^2 V_{ih'}^2 \\
&\le Cp^2(1-p)n, \qquad \text{(S3.3)}
\end{aligned}$$

where the first inequality is due to Jensen's inequality. This shows that the term (a) is $O_p(p\sqrt{n})$. Similarly, we can show that the terms (b) and (e) are $O_p(p\sqrt{n})$.

The term (c) in (S3.1) is

$$\begin{aligned}
&\frac{1}{2p}(c) \\
&= \sum_{i=1}^{m} V_i^T \left(M_0^T M_y - (1-p)\text{diag}(M_0^T M_y) \right) V_i \\
&= \sum_{i=1}^{m} \left\{ \sum_{k=1}^{n} \left(\sum_{h=1}^{d} M_{0kh} V_{ih} \right) \left(\sum_{h'=1}^{d} (y_{kh'} - p) M_{0kh'} V_{ih'} \right) \right. \\
&\qquad \left. -(1-p) \sum_{k=1}^{n} \sum_{h=1}^{d} (y_{kh} - p) M_{0kh}^2 V_{ih}^2 \right\}
\end{aligned}$$

$$= \sum_{k=1}^{n}\sum_{h=1}^{d}(y_{kh}-p)M_{0kh}\sum_{i=1}^{m}V_{ih}\left[\left(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\right)\right.$$
$$\left. -(1-p)M_{0kh}V_{ih}\right].$$

Then, its variance is

$$\left(\frac{1}{2p}\right)^2 \mathrm{var}(c)$$
$$= \sum_{k=1}^{n}\sum_{h=1}^{d}p(1-p)M_{0kh}^2\left\{\sum_{i=1}^{m}V_{ih}\left[\left(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\right)\right.\right.$$
$$\left.\left. -(1-p)M_{0kh}V_{ih}\right]\right\}^2$$
$$\leq Cp(1-p)nd,$$

where the last inequality is due to Assumption 1(1) and the fact that

$$\sum_{k=1}^{n}\sum_{h=1}^{d}M_{0kh}^2\left\{\sum_{i=1}^{m}V_{ih}\left[\left(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\right)-(1-p)M_{0kh}V_{ih}\right]\right\}^2$$
$$= \sum_{k=1}^{n}\sum_{h=1}^{d}M_{0kh}^2\left\{\sum_{i=1}^{m}\lambda_i U_{ik}V_{ih}-(1-p)\sum_{i=1}^{m}M_{0kh}V_{ih}^2\right\}^2$$
$$= \sum_{k=1}^{n}\sum_{h=1}^{d}M_{0kh}^2\left\{\sum_{i=1}^{m}\lambda_i U_{ik}V_{ih}\right\}^2$$
$$+(1-p)^2\sum_{k=1}^{n}\sum_{h=1}^{d}\left\{\sum_{i=1}^{m}M_{0kh}^2V_{ih}^2\right\}^2$$
$$-2(1-p)\sum_{k=1}^{n}\sum_{h=1}^{d}M_{0kh}\left\{\sum_{i=1}^{m}\lambda_i U_{ik}V_{ih}\right\}\left\{\sum_{i=1}^{m}M_{0kh}^2V_{ih}^2\right\}$$
$$= \sum_{k=1}^{n}\sum_{h=1}^{d}M_{0kh}^2\left\{\sum_{i=1}^{m}\lambda_i U_{ik}V_{ih}\right\}^2+O(n)$$

$$= O(nd). \tag{S3.4}$$

The term (d) in (S3.1) is

$$\begin{aligned}
\frac{1}{2p}(d) &= \sum_{i=1}^{m} V_i^T \left(M_0^T \epsilon_y - (1-p)\mathrm{diag}(M_0^T \epsilon_y) \right) V_i \\
&= \sum_{i=1}^{m} \Bigg\{ \sum_{k=1}^{n} \left(\sum_{h=1}^{d} M_{0kh} V_{ih} \right) \left(\sum_{h'=1}^{d} y_{kh'} \epsilon_{kh'} V_{ih'} \right) \\
&\qquad\qquad - (1-p) \sum_{k=1}^{n} \sum_{h=1}^{d} y_{kh} \epsilon_{kh} M_{0kh} V_{ih}^2 \Bigg\} \\
&= \sum_{k=1}^{n} \sum_{h=1}^{d} y_{kh} \epsilon_{kh} \left\{ \sum_{i=1}^{m} V_{ih} \left[\left(\sum_{h'=1}^{d} M_{0kh'} V_{ih'} \right) - (1-p) M_{0kh} V_{ih} \right] \right\}.
\end{aligned}$$

Then, its variance is

$$\begin{aligned}
&\left(\frac{1}{2p} \right)^2 \mathrm{var}(d) \\
&= \sum_{k=1}^{n} \sum_{h=1}^{d} p\sigma^2 \left\{ \sum_{i=1}^{m} V_{ih} \left[\left(\sum_{h'=1}^{d} M_{0kh'} V_{ih'} \right) - (1-p) M_{0kh} V_{ih} \right] \right\}^2 \\
&\leq Cpnd,
\end{aligned}$$

where the last inequality is due to Assumption 1(1) and the fact that

$$\begin{aligned}
&\sum_{k=1}^{n} \sum_{h=1}^{d} \left\{ \sum_{i=1}^{m} V_{ih} \left[\left(\sum_{h'=1}^{d} M_{0kh'} V_{ih'} \right) - (1-p) M_{0kh} V_{ih} \right] \right\}^2 \\
&= \sum_{k=1}^{n} \sum_{h=1}^{d} \left\{ \sum_{i=1}^{m} \lambda_i U_{ik} V_{ih} - (1-p) \sum_{i=1}^{m} M_{0kh} V_{ih}^2 \right\}^2 \\
&= \sum_{i=1}^{m} \lambda_i^2 + (1-p)^2 \sum_{k=1}^{n} \sum_{h=1}^{d} \left(\sum_{i=1}^{m} M_{0kh} V_{ih}^2 \right)^2
\end{aligned}$$

$$
\begin{aligned}
&-2(1-p)\sum_{i=1}^{m}\lambda_i^2\sum_{h=1}^{d}V_{ih}^2\sum_{i'=1}^{m}V_{i'h}^2 \\
&= \sum_{i=1}^{m}\lambda_i^2 + O(n). \qquad \text{(S3.5)}
\end{aligned}
$$

The term (f) in (S3.1) is

$$
\begin{aligned}
|(f)| &= \left|\mathrm{tr}({V_p}^{(m)T}\hat{\Sigma}_p {V_p}^{(m)} - \mathcal{O}^T V^{(m)T}\hat{\Sigma}_p V^{(m)}\mathcal{O})\right| \\
&\leq \sum_{i=1}^{m}\left|\mathcal{O}_i^T V^T\hat{\Sigma}_p V\mathcal{O}_i - {V_{p_i}}^T\hat{\Sigma}_p V_{p_i}\right| \\
&= \sum_{i=1}^{m}\left\{\left|(V\mathcal{O}_i - V_{p_i})^T\hat{\Sigma}_p(V\mathcal{O}_i - V_{p_i}) + 2\lambda_{p_i}^2 {V_{p_i}}^T(V\mathcal{O}_i - V_{p_i})\right|\right\} \\
&\leq \sum_{i=1}^{m}\lambda_{p_1}^2\left(\left\|V\mathcal{O}_i - V_{p_i}\right\|_2^2 + 2\left|{V_{p_i}}^T(V\mathcal{O}_i - V_{p_i})\right|\right) \\
&= \sum_{i=1}^{m}\lambda_{p_1}^2\Big(\left\|V\mathcal{O}_i - V_{p_i}\right\|_2^2 \\
&\qquad + \left|\mathcal{O}_i^T V^T V\mathcal{O}_i - \mathcal{O}_i^T V^T V_{p_i} - {V_{p_i}}^T V\mathcal{O}_i + {V_{p_i}}^T V_{p_i})\right|\Big) \\
&= \sum_{i=1}^{m}2\lambda_{p_1}^2\left\|V\mathcal{O}_i - V_{p_i}\right\|_2^2 \\
&= 2\lambda_{p_1}^2\left\|V^{(m)}\mathcal{O} - {V_p}^{(m)}\right\|_F^2 \\
&= O_p(pd), \qquad \text{(S3.6)}
\end{aligned}
$$

where $\mathcal{O}_i$ is the i-th column of $\mathcal{O}$ and the last equality holds by Proposition 2, (9), and (A.10).

Therefore, the result follows from (S3.1)-(S3.6). □

Proof of Proposition 3. By Cramèr-Wold device, it is enough to show that for any given $(c_1, c_2)^T \in \mathbb{R}^2 \setminus (0,0)^T$,

$$\frac{1}{\sqrt{nd}\gamma_{c_1,c_2}} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^T \left[\begin{pmatrix} p^{-2} \sum_{i=1}^m \lambda_{p_i}^2 \\ p^2 \sum_{i=1}^m (\lambda_i^2 + n\sigma^2)\, \hat{p} \end{pmatrix} - \begin{pmatrix} \sum_{i=1}^m [\lambda_i^2 + n\sigma^2] \\ p^3 \sum_{i=1}^m (\lambda_i^2 + n\sigma^2) \end{pmatrix} \right]$$
$$\to \mathcal{N}(0,1) \text{ in distribution}, \quad \text{as } n, d \to \infty,$$

where $\gamma_{c_1,c_2}^2 = \big(c_1\, c_2\big)\Gamma_{nd} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$. When $c_1 = 0$, this can be directly showed by CLT. Thus, we only consider the case where $c_1 \neq 0$.

We have

$$\begin{aligned}
&\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^T \left[\begin{pmatrix} p^{-2} \sum_{i=1}^m \lambda_{p_i}^2 \\ p^2 \sum_{i=1}^m (\lambda_i^2 + n\sigma^2)\, \hat{p} \end{pmatrix} - \begin{pmatrix} \sum_{i=1}^m [\lambda_i^2 + n\sigma^2] \\ p^3 \sum_{i=1}^m (\lambda_i^2 + n\sigma^2) \end{pmatrix} \right] \\
&= c_1 \frac{1}{p^2} \sum_{i=1}^m \left[\lambda_{p_i}^2 - p^2 \left(\lambda_i^2 + n\sigma^2 \right) \right] + c_2\, p^2 \sum_{i=1}^m (\lambda_i^2 + n\sigma^2)\, (\hat{p} - p) \\
&= \frac{2c_1}{p} \sum_{k=1}^n \sum_{h=1}^d (y_{kh} - p) M_{0kh} \sum_{i=1}^m V_{ih} \Bigg[\Bigg(\sum_{h'=1}^d M_{0kh'} V_{ih'} \Bigg) \\
&\qquad\qquad - (1-p) M_{0kh} V_{ih} \Bigg] \\
&\quad + \frac{2c_1}{p} \sum_{k=1}^n \sum_{h=1}^d y_{kh} \epsilon_{kh} \sum_{i=1}^m V_{ih} \Bigg[\Bigg(\sum_{h'=1}^d M_{0kh'} V_{ih'} \Bigg) - (1-p) M_{0kh} V_{ih} \Bigg] \\
&\quad + o_p \left(\sqrt{\frac{nd}{p}} \right) + \frac{c_2\, p^2}{nd} \sum_{k=1}^n \sum_{h=1}^d (y_{kh} - p) \sum_{i=1}^m (\lambda_i^2 + n\sigma^2) \\
&= \sum_{k=1}^n \sum_{h=1}^d (y_{kh} - p) \Bigg\{ \frac{2c_1}{p} M_{0kh} \sum_{i=1}^m V_{ih} \Bigg[\Bigg(\sum_{h'=1}^d M_{0kh'} V_{ih'} \Bigg) \\
&\qquad\qquad - (1-p) M_{0kh} V_{ih} \Bigg] + \frac{c_2\, p^2}{nd} \sum_{i=1}^m (\lambda_i^2 + n\sigma^2) \Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{2c_1}{p}\sum_{k=1}^{n}\sum_{h=1}^{d} y_{kh}\epsilon_{kh}\sum_{i=1}^{m} V_{ih}\left[\left(\sum_{h'=1}^{d} M_{0kh'}V_{ih'}\right)-(1-p)M_{0kh}V_{ih}\right]\\
&+o_p\left(\sqrt{\frac{nd}{p}}\right)\\
=\;&(a)+(b)+o_p\left(\sqrt{\frac{nd}{p}}\right), \qquad \text{(S3.7)}
\end{aligned}
$$

where the second equality holds by Lemma S.3. Since the terms (a) and (b) are centered and not correlated with each other under the model setup in Section 2, we have

$$
\begin{aligned}
&\operatorname{var}[(a)+(b)]=\operatorname{var}[(a)]+\operatorname{var}[(b)]\\
&=\sum_{k=1}^{n}\sum_{h=1}^{d}\mathbb{E}(y_{kh}-p)^2\left\{\frac{2c_1}{p}M_{0kh}\sum_{i=1}^{m}V_{ih}\left[\left(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\right)\right.\right.\\
&\qquad\left.\left.-(1-p)M_{0kh}V_{ih}\right]+\frac{c_2\,p^2}{nd}\sum_{i=1}^{m}(\lambda_i^2+n\sigma^2)\right\}^2\\
&\quad+\frac{4c_1^2}{p^2}\sum_{k=1}^{n}\sum_{h=1}^{d}\mathbb{E}\left(y_{kh}^2\epsilon_{kh}^2\right)\left\{\sum_{i=1}^{m}V_{ih}\left[\left(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\right)\right.\right.\\
&\qquad\left.\left.-(1-p)M_{0kh}V_{ih}\right]\right\}^2\\
&=p(1-p)\sum_{k=1}^{n}\sum_{h=1}^{d}\left\{\frac{2c_1M_{0kh}}{p}\sum_{i=1}^{m}\lambda_iU_{ik}V_{ih}+c_2p^2\sum_{i=1}^{m}b_i^2\right\}^2\\
&\quad+\frac{4\sigma^2c_1^2}{p}\sum_{i=1}^{m}\lambda_i^2+O\left(\frac{n}{p}\right)\\
&=\frac{4c_1^2(1-p)}{p}\sum_{k=1}^{n}\sum_{h=1}^{d}M_{0kh}^2\left\{\sum_{i=1}^{m}\lambda_iU_{ik}V_{ih}\right\}^2+\frac{4\sigma^2c_1^2}{p}\sum_{i=1}^{m}\lambda_i^2\\
&\quad+4c_1c_2\,nd\,p^2(1-p)\left(\sum_{i=1}^{m}b_i^2\right)^2+c_2^2\,nd\,p^5(1-p)\left(\sum_{i=1}^{m}b_i^2\right)^2\\
&\quad+O\left(\frac{n}{p}\right)
\end{aligned}
$$

$$= nd\left(c_1\, c_2\right)\Gamma_{nd}\begin{pmatrix} c_1 \\ \\ c_2 \end{pmatrix} + O\left(\frac{n}{p}\right), \tag{S3.8}$$

where the third equality is due to (S3.4), (S3.5) and Assumption 1(1). Note that

$$nd\left(c_1\, c_2\right)\Gamma_{nd}\begin{pmatrix} c_1 \\ \\ c_2 \end{pmatrix} \geq \frac{4c_1^2\sigma^2}{p}\sum_{i=1}^{m}\lambda_i^2 \geq \frac{c\, nd}{p}. \tag{S3.9}$$

Thus, Liapunov's condition is satisfied with $(a)+(b)$ because we have

$$\begin{aligned}
&\sum_{k=1}^{n}\sum_{h=1}^{d}\mathbb{E}\Bigg|(y_{kh}-p)\Bigg\{\frac{2c_1}{p}M_{0kh}\sum_{i=1}^{m}V_{ih}\Bigg[\Bigg(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\Bigg) \\
&\qquad\qquad -(1-p)M_{0kh}V_{ih}\Bigg] + \frac{c_2\, p^2}{nd}\sum_{i=1}^{m}(\lambda_i^2+n\sigma^2)\Bigg\} \\
&\quad + y_{kh}\epsilon_{kh}\Bigg\{\frac{2c_1}{p}\sum_{i=1}^{m}V_{ih}\Bigg[\Bigg(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\Bigg)-(1-p)M_{0kh}V_{ih}\Bigg]\Bigg\}\Bigg|^3 \\
&\leq 8\sum_{k=1}^{n}\sum_{h=1}^{d}\Bigg\{\mathbb{E}|y_{kh}-p|^3\Bigg|\frac{2c_1}{p}M_{0kh}\sum_{i=1}^{m}V_{ih}\Bigg[\Bigg(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\Bigg) \\
&\qquad\qquad -(1-p)M_{0kh}V_{ih}\Bigg] + O\left(1\right)\Bigg|^3 \\
&\qquad + \mathbb{E}|y_{kh}\epsilon_{kh}|^3\Bigg|\frac{2c_1}{p}\sum_{i=1}^{m}V_{ih}\Bigg[\Bigg(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\Bigg) \\
&\qquad\qquad -(1-p)M_{0kh}V_{ih}\Bigg]\Bigg|^3\Bigg\} \\
&\leq \frac{C}{p^2}\sum_{k=1}^{n}\sum_{h=1}^{d}\Bigg\{\Bigg|\sum_{i=1}^{m}V_{ih}\Bigg[\Bigg(\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\Bigg) \\
&\qquad\qquad -(1-p)M_{0kh}V_{ih}\Bigg]\Bigg|^3 + O\left(1\right)\Bigg\} \\
&\leq \frac{C}{p^2}\sum_{k=1}^{n}\sum_{h=1}^{d}\Bigg\{\sum_{i=1}^{m}|V_{ih}|^3\Bigg(\Bigg|\sum_{h'=1}^{d}M_{0kh'}V_{ih'}\Bigg|^3+|V_{ih}|^3\Bigg)+O\left(1\right)\Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{C}{p^2}\sum_{i=1}^{m}\sum_{k=1}^{n}\left|\sum_{h'=1}^{d} M_{0kh'}V_{ih'}\right|^3 + O\left(nd\right)\\
&= O\left(\frac{nd^{3/2}}{p^2}\right),
\end{aligned} \tag{S3.10}
$$

where the first inequality holds by Assumption 1(1), and the last two lines are due to Cauchy-Schwarz inequality.

By (S3.7)-(S3.10), Liapunov CLT and Slutsky theorem, we have

$$
\frac{1}{\sqrt{nd}\gamma_{c_1,c_2}}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}^T \left[\begin{pmatrix} p^{-2}\sum_{i=1}^{m}\lambda_{p_i}^2 \\ p^2\sum_{i=1}^{m}(\lambda_i^2+n\sigma^2)\,\hat{p} \end{pmatrix} - \begin{pmatrix} \sum_{i=1}^{m}\left[\lambda_i^2+n\sigma^2\right] \\ p^3\sum_{i=1}^{m}(\lambda_i^2+n\sigma^2) \end{pmatrix}\right]
$$

$$
\to \mathcal{N}\left(0,\, 1\right) \text{ in distribution}, \quad \text{as } n, d \to \infty.
$$

□

Proof of Proposition 4. Similarly to the proof of (S3.1), we have

$$
\begin{aligned}
&\hat{\tau}_p - np^2\sigma^2\\
&= \frac{1}{d-r}\mathrm{tr}\left(V_{pc}^T\hat{\Sigma}_p V_{pc}\right) - \frac{1}{d-r}\mathrm{tr}\left(V_c^T\mathbb{E}\,\hat{\Sigma}_p V_c\right)\\
&= \frac{1}{d-r}\mathrm{tr}\left(V_c^T\hat{\Sigma}_p V_c\right)\\
&\qquad + \frac{1}{d-r}\mathrm{tr}\left(V_{pc}^T\hat{\Sigma}_p V_{pc} - \mathcal{O}^T V_c^T\hat{\Sigma}_p V_c\mathcal{O}^T\right) - \frac{1}{d-r}\mathrm{tr}\left(V_c^T\mathbb{E}\,\hat{\Sigma}_p V_c\right)\\
&= \frac{1}{d-r}\mathrm{tr}\Big(V_c^T\big(M_y^T M_y - (1-p)\mathrm{diag}(M_y^T M_y)\\
&\qquad\qquad\qquad\qquad - p^2(1-p)\,\mathrm{diag}(M_0^T M_0)\big)V_c\Big)\\
&\qquad + \frac{1}{d-r}\mathrm{tr}\Big(V_c^T\left(\epsilon_y^T\epsilon_y - (1-p)\mathrm{diag}(\epsilon_y^T\epsilon_y)\right)V_c - np^2\sigma^2 I_{d-r}\Big)
\end{aligned}
$$

$$
\begin{aligned}
&-2p(1-p)\frac{1}{d-r}\mathrm{tr}\Big(V_c^T\left(\mathrm{diag}(M_y^T M_0)\right)V_c\Big)\\
&-2p(1-p)\frac{1}{d-r}\mathrm{tr}\Big(V_c^T\left(\mathrm{diag}(\epsilon_y^T M_0)\right)V_c\Big)\\
&+2\frac{1}{d-r}\mathrm{tr}\Big(V_c^T\left(M_y^T\epsilon_y-(1-p)\mathrm{diag}(M_y^T\epsilon_y)\right)V_c\Big)\\
&+\frac{1}{d-r}\mathrm{tr}\left(V_{pc}^T\hat{\Sigma}_p V_{pc}-\mathcal{O}^T V_c^T\hat{\Sigma}_p V_c\mathcal{O}\right)\\
=\ &(A)+(B)-2p(1-p)\cdot(C)-2p(1-p)\cdot(D)+2\cdot(E)+(F),
\end{aligned}
$$

where $\mathcal{O}\in\mathbb{V}_{d-r,d-r}$ is a solution to $\inf_{\mathcal{Q}\in\mathbb{V}_{d-r,d-r}}\|V_{pc}-V_c\mathcal{Q}\|_F^2$, and the third equality is due to the fact that $M_0V_c=U\Lambda V^TV_c=0$. We will show that (A)-(F) are $O_p\left(p\sqrt{n}\right)$.

Since the first five terms, (A)-(E), are centered, we only need to check their variances to find their rates. The variances of the terms (A), (B), and (E) are $O\left(p^2n\right)$, which can be shown similarly to the proof of (S3.3). The variance of the term (C) is

$$
\begin{aligned}
\mathrm{var}(C) &\le \frac{1}{d-r}\sum_{i=1}^{d-r}\mathbb{E}\left[V_{ci}^T\left(\mathrm{diag}(M_y^TM_0)\right)V_{ci}\right]^2\\
&= \frac{1}{d-r}\sum_{i=1}^{d-r}\mathrm{var}\left[\sum_{k=1}^{n}\sum_{h=1}^{d}(y_{kh}-p)M_{0\,kh}^{\,2}V_{cih}^{\,2}\right]\\
&= \frac{1}{d-r}\sum_{i=1}^{d-r}\left[L^4\sum_{k=1}^{n}O(p(1-p))\right]\\
&= O(pn),
\end{aligned}
$$

where the inequality is due to Jensen's inequality. Similarly, the variance of the term (D) is $O(pn)$.

Now, consider the term (F). Similarly to the proof of (S3.6),

$$
|(F)| \le \frac{1}{d-r}\left|\mathrm{tr}\left(V_{pc}^T\hat{\Sigma}_pV_{pc}-\mathcal{O}^TV_c^T\hat{\Sigma}_pV_c\mathcal{O}\right)\right|
$$

$$
\begin{aligned}
&\leq \frac{1}{d-r}\sum_{i=1}^{d-r}\left|{V_{pc}}_i^T\hat{\Sigma}_p V_{pc_i} - \mathcal{O}_i^T V_c^T \hat{\Sigma}_p V_c \mathcal{O}_i\right| \\
&\leq \frac{1}{d-r}\cdot 2\lambda_{p1}^2 \left\|V_{pc} - V_c\mathcal{O}\right\|_F^2 \\
&\leq \frac{1}{d-r}\cdot 4\lambda_{p1}^2 \left\|\sin\left(V_{pc}, V_c\right)\right\|_F^2 \\
&= \frac{1}{d-r}\cdot 2\lambda_{p1}^2 \left\|V_{pc}V_{pc}^T - V_c V_c^T\right\|_F^2 \\
&= \frac{1}{d-r}\cdot 2\lambda_{p1}^2 \left\|(I_d - V_p V_p^T) - (I_d - VV^T)\right\|_F^2 \\
&= \frac{1}{d-r}\cdot 2\lambda_{p1}^2 \left\|V_p V_p^T - VV^T\right\|_F^2 \\
&= \frac{1}{d-r}\cdot 4\lambda_{p1}^2 \left\|\sin(V_p, V)\right\|_F^2 \\
&= O_p(p),
\end{aligned}
$$

where $\mathcal{O}_i$ is the i-th column of $\mathcal{O}$, the third inequality can be derived similarly to the proof of (A.9), and the last equality holds by Proposition 2 and (A.10). □

S4 Proofs for Appendix A.3

Proof of Proposition 5. Let $\Delta_{\lambda_i} = \hat{\lambda}_i - \lambda_i$, $\Delta_{U_i} = \text{sign}(\langle \hat{U}_i, U_i\rangle)\hat{U}_i - U_i$, and $\Delta_{V_i} = \text{sign}(\langle \hat{V}_i, V_i\rangle)\hat{V}_i - V_i$ for all $i \in \{1, \ldots, r\}$. Similarly to the proof of Theorem 2, we can show that for all $i = 1, \ldots, r$,

$$
|\Delta_{\lambda_i}| = O_p\left(\frac{1}{\sqrt{p}} + \frac{1}{p}\sqrt{\frac{d}{n}}\right). \tag{S4.1}
$$

Then,

$$
\begin{aligned}
&\left\| \hat{M}(s_0) - M_0 \right\|_F^2 \\
&= \left\| \sum_{i=1}^{r} s_{0i}\ \hat{\lambda}_i \hat{U}_i \hat{V}_i^T - \sum_{i=1}^{r} \lambda_i U_i V_i^T \right\|_F^2 \\
&\leq r^2 \sum_{i=1}^{r} \left\| s_{0i}\ \hat{\lambda}_i \hat{U}_i \hat{V}_i^T - \lambda_i U_i V_i^T \right\|_F^2 \\
&= r^2 \sum_{i=1}^{r} \left\| \left(\lambda_i + \Delta_{\lambda_i}\right)\left(U_i + \Delta_{U_i}\right)\left(V_i + \Delta_{V_i}\right)^T - \lambda_i U_i V_i^T \right\|_F^2 \\
&\leq C r^2 \sum_{i=1}^{r} \left\{ \left\| \Delta_{\lambda_i} U_i V_i^T \right\|_F^2 + \left\| \lambda_i \Delta_{U_i} V_i^T \right\|_F^2 + \left\| \lambda_i U_i \Delta_{V_i}^T \right\|_F^2 \right\} \\
&= C r^2 \sum_{i=1}^{r} \left\{ O_p\left(\frac{1}{\sqrt{p}} + \frac{1}{p}\sqrt{\frac{d}{n}} \right) + O\left(nd\right) \frac{1}{p\, b_r^4} O_p\left(\frac{1}{d}\right) \right. \\
&\qquad\qquad \left. + O\left(nd\right) \frac{1}{p\, b_r^4} O_p\left(\frac{1}{n}\right) \right\} \\
&= \frac{1}{p\, b_r^4}\, O_p(n),
\end{aligned}
$$

where the third equality holds due to (S4.1) and Theorem 1. □

S5 Proofs for Lemma 2

Proof of Lemma 2. By Weyl's theorem (Li (1998a)), Lemma 3, and Lemma 4, for any given $\delta > 0$, there exists a large constant $C_\delta > 0$ such that

$$
\begin{aligned}
&\max\left\{ \left| \lambda_{\hat{p}r}^2 - p^2(\lambda_r^2 + n\sigma^2) \right|, \left| \lambda_{\hat{p}\,r+1}^2 - p^2 n\sigma^2 \right| \right\} \\
&\leq \left\| \hat{\Sigma}_{\hat{p}} - \mathbb{E}(\hat{\Sigma}_p) \right\|_2 \\
&\leq \left\| \hat{\Sigma}_{\hat{p}} - \hat{\Sigma}_p \right\|_2 + \left\| \hat{\Sigma}_p - \mathbb{E}(\hat{\Sigma}_p) \right\|_2
\end{aligned}
$$

$$\leq C_\delta\, p^{3/2}\sqrt{\frac{n\log n}{d}} \tag{S5.1}$$

with probability at least $1 - O(n^{-\delta})$. Also, by definition of $\hat{r}$, we have

$$\begin{aligned}\{\hat{r} = r\} &= \{\lambda^2_{\hat{p}r} \geq p^2 n\, C_d,\ \lambda^2_{\hat{p}\,r+1} < p^2 n\, C_d\} \\ &= \Big\{\big[\lambda^2_{\hat{p}r} - p^2(\lambda_r^2 + n\sigma^2)\big] + p^2(\lambda_r^2 + n\sigma^2) \geq p^2 n\, C_d, \\ &\qquad \big[\lambda^2_{\hat{p}\,r+1} - p^2 n\sigma^2\big] + p^2 n\sigma^2 < p^2 n\, C_d\Big\},\end{aligned} \tag{S5.2}$$

where $\lambda_r^2 = b_r^2\, nd$ by Assumption 1(1). The result follows by (S5.1) and (S5.2).

□

References

Koltchinskii, V., Lounici, K., and Tsybakov, A. B.. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329.

Li, R.-C. (1998a). Relative perturbation theory: I. eigenvalue and singular value variations. *SIAM Journal on Matrix Analysis and Applications*, 19(4):956–982.

Li, R.-C. (1998b). Relative perturbation theory: Ii. eigenspace and singular subspace variations. *SIAM Journal on Matrix Analysis and Applications*, 20(2):471–492.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.