Supplementary Material to “Oracle Sparse PCA via Adaptive Estimation”

Wenfu Zhong
School of Information Science and Technology
ShanghaiTech University
Shanghai, China
wenfuzhong@shanghaitech.edu.cn

Ziping Zhao
School of Information Science and Technology
ShanghaiTech University
Shanghai, China
zipingzhao@shanghaitech.edu.cn

This supplementary material contains the design details of the optimality metric mentioned in Section III and the proofs for Proposition 1, Lemma 2, and Theorem 3 in Section IV.

Appendix A
Optimality Metric

In this appendix, we design an optimality metric used in the stopping criterion of Algorithm 2. Firstly, we need to identify the variational inequality associated with problem (5), which characterizes the optimality of the solution. Suppose $\hat{\boldsymbol{\Pi}}$ and $\hat{\boldsymbol{\Psi}}$ are the primal optima of problem (5), and $\hat{\mathbf{Z}}$ is the dual optimum. Let

$$\mathcal{L}_0\left(\boldsymbol{\Pi},\boldsymbol{\Psi},\mathbf{Z}\right) := \langle \mathbf{S},\boldsymbol{\Pi}\rangle - \left\|\boldsymbol{\Lambda}\odot\boldsymbol{\Psi}\right\|_1 + \langle \mathbf{Z},\boldsymbol{\Pi}-\boldsymbol{\Psi}\rangle$$

be the Lagrangian function of problem (5). Then according to convex optimization theory [1], $\left(\hat{\boldsymbol{\Pi}},\hat{\boldsymbol{\Psi}},\hat{\mathbf{Z}}\right)$ should be a saddle point of $\mathcal{L}_0\left(\cdot\right)$, and the following inequalities hold:

$$\mathcal{L}_0\left(\hat{\boldsymbol{\Pi}},\hat{\boldsymbol{\Psi}},\mathbf{Z}\right) \geq \mathcal{L}_0\left(\hat{\boldsymbol{\Pi}},\hat{\boldsymbol{\Psi}},\hat{\mathbf{Z}}\right) \geq \mathcal{L}_0\left(\boldsymbol{\Pi},\boldsymbol{\Psi},\hat{\mathbf{Z}}\right),\ \forall \boldsymbol{\Pi}\in\mathcal{F},\boldsymbol{\Psi}\in\mathbb{R}^{d\times d},\mathbf{Z}\in\mathbb{R}^{d\times d}.$$

From the first inequality,

$$\left\langle \hat{\mathbf{Z}}-\mathbf{Z},\hat{\boldsymbol{\Pi}}-\hat{\boldsymbol{\Psi}}\right\rangle \leq 0. \tag{6}$$

From the second inequality,

$$\left\langle \mathbf{S},\boldsymbol{\Pi}-\hat{\boldsymbol{\Pi}}\right\rangle + \left\|\boldsymbol{\Lambda}\odot\hat{\boldsymbol{\Psi}}\right\|_1 - \left\|\boldsymbol{\Lambda}\odot\boldsymbol{\Psi}\right\|_1 + \left\langle \hat{\mathbf{Z}},\boldsymbol{\Pi}-\hat{\boldsymbol{\Pi}}\right\rangle - \left\langle \hat{\mathbf{Z}},\boldsymbol{\Psi}-\hat{\boldsymbol{\Psi}}\right\rangle \leq 0. \tag{7}$$

Summing up (6) and (7), we obtain the variational inequality:

$$\left\langle \mathbf{S},\boldsymbol{\Pi}-\hat{\boldsymbol{\Pi}}\right\rangle + \left\|\boldsymbol{\Lambda}\odot\hat{\boldsymbol{\Psi}}\right\|_1 - \left\|\boldsymbol{\Lambda}\odot\boldsymbol{\Psi}\right\|_1 - \left\langle \begin{bmatrix}\boldsymbol{\Pi}-\hat{\boldsymbol{\Pi}}\\ \boldsymbol{\Psi}-\hat{\boldsymbol{\Psi}}\\ \mathbf{Z}-\hat{\mathbf{Z}}\end{bmatrix}, \begin{bmatrix}-\hat{\mathbf{Z}}\\ \hat{\mathbf{Z}}\\ \hat{\boldsymbol{\Pi}}-\hat{\boldsymbol{\Psi}}\end{bmatrix}\right\rangle \leq 0,\ \forall \boldsymbol{\Pi}\in\mathcal{F},\boldsymbol{\Psi}\in\mathbb{R}^{d\times d},\mathbf{Z}\in\mathbb{R}^{d\times d}. \tag{8}$$

Now solving (5) is equivalent to finding a set of $\left(\hat{\boldsymbol{\Pi}},\hat{\boldsymbol{\Psi}},\hat{\mathbf{Z}}\right)$ satisfying (8).

Based on the variational inequality (8), we define an auxiliary function to measure the optimality of $\left(\mathring{\boldsymbol{\Pi}}^t,\mathring{\boldsymbol{\Psi}}^t,\mathring{\mathbf{Z}}^t\right)$ at iteration t:

$$V^t\left(\boldsymbol{\Pi},\boldsymbol{\Psi},\mathbf{Z}\right) := \left\langle \mathbf{S},\boldsymbol{\Pi}-\mathring{\boldsymbol{\Pi}}^t\right\rangle + \left\|\boldsymbol{\Lambda}\odot\mathring{\boldsymbol{\Psi}}^t\right\|_1 - \left\|\boldsymbol{\Lambda}\odot\boldsymbol{\Psi}\right\|_1 - \left\langle \begin{bmatrix}\boldsymbol{\Pi}-\mathring{\boldsymbol{\Pi}}^t\\ \boldsymbol{\Psi}-\mathring{\boldsymbol{\Psi}}^t\\ \mathbf{Z}-\mathring{\mathbf{Z}}^t\end{bmatrix}, \begin{bmatrix}-\mathring{\mathbf{Z}}^t\\ \mathring{\mathbf{Z}}^t\\ \mathring{\boldsymbol{\Pi}}^t-\mathring{\boldsymbol{\Psi}}^t\end{bmatrix}\right\rangle.$$

Our design approach is to identify the maximum value of the function V^t within a subset of set $\mathcal{F}\times\mathbb{R}^{d\times d}\times\mathbb{R}^{d\times d}$ and use it as the optimality metric, i.e.,

$$\omega^t := \underset{\boldsymbol{\Pi}\in\mathcal{F},\boldsymbol{\Psi}\in\mathbb{R}^{d\times d},\mathbf{Z}=\widetilde{\mathbf{Z}}^t}{\text{maximize}}\ V^t\left(\boldsymbol{\Pi},\boldsymbol{\Psi},\mathbf{Z}\right), \tag{9}$$

where $\widetilde{\mathbf{Z}}^t := -\boldsymbol{\Lambda}\odot\operatorname{sgn}\left(\mathring{\boldsymbol{\Pi}}^t-\mathring{\boldsymbol{\Psi}}^t\right)$. If there exists a small $\varepsilon>0$ such that $\omega^t\leq\varepsilon$, then $\left(\mathring{\boldsymbol{\Pi}}^t,\mathring{\boldsymbol{\Psi}}^t,\mathring{\mathbf{Z}}^t\right)$ can be regarded as an approximate solution of (8) with tolerance ε [2].

Since there is no coupling between variables $\boldsymbol{\Pi}$, $\boldsymbol{\Psi}$, and $\mathbf{Z}$ in problem (9), we can analyze their respective optimization processes separately. For the variable $\boldsymbol{\Pi}$, we get the following maximization subproblem:

$$\underset{\boldsymbol{\Pi}\in\mathcal{F}}{\text{maximize}}\quad \left\langle \mathbf{S}+\mathring{\mathbf{Z}}^t,\boldsymbol{\Pi}\right\rangle \tag{10}$$

The optimal value of (10) takes $\sum_{i=1}^{r} \lambda_i \left(\mathbf{S} + \mathring{\mathbf{Z}}^t\right)$ by applying the basic properties of Fantope projection (see Lemma 1 in [3]). For the variable $\mathbf{\Psi}$, we have

$$\underset{\mathbf{\Psi}\in\mathbb{R}^{d\times d}}{\text{maximize}} \quad -\left\|\mathbf{\Lambda}\odot\mathbf{\Psi}\right\|_1 - \left\langle \mathring{\mathbf{Z}}^t, \mathbf{\Psi}\right\rangle. \tag{11}$$

By the first order optimality condition derived from Line 5 of Algorithm 2 and the update of dual variable in Line 6, there exists

$$-\mathbf{\Lambda}\odot\mathbf{\Xi}^{t'} = \mathbf{Z}^{t'-1} - \rho\left(\mathbf{\Pi}^{t'} - \mathbf{\Psi}^{t'}\right) = \mathbf{Z}^{t'},\ \exists \mathbf{\Xi}^{t'} \in \partial\left\|\mathbf{\Psi}^{t'}\right\|_1,\ t'\geq 1,$$

and thus the following relation holds:

$$\left|\mathring{Z}^t_{ij}\right| \leq \frac{1}{t}\sum_{t'=1}^{t}\left|Z^{t'}_{ij}\right| = \frac{1}{t}\sum_{t'=1}^{t}\left|\Lambda_{ij}\right|\left|\Xi^{t'}_{ij}\right| \leq \frac{1}{t}\sum_{t'=1}^{t}\left|\Lambda_{ij}\right| = \Lambda_{ij},\ 1\leq i,j\leq d,$$

in which the second inequality is due to $\Xi^{t'}_{ij}\in[-1,1]$. Consequently, it is straightforward to deduce that problem (11) must have a zero solution. For the variable $\mathbf{Z}$, as the feasible set contains only a single point, it is trivial to calculate the optimal value by substituting $\widetilde{\mathbf{Z}}^t$. Based on the above analysis, we can derive the explicit expression of the optimality metric:

$$\omega^t = \sum_{i=1}^{r}\lambda_i\left(\mathbf{S}+\mathring{\mathbf{Z}}^t\right) - \left\langle \mathbf{S}, \mathring{\mathbf{\Pi}}^t\right\rangle + \left\|\mathbf{\Lambda}\odot\mathring{\mathbf{\Psi}}^t\right\|_1 + \left\|\mathbf{\Lambda}\odot\left(\mathring{\mathbf{\Pi}}^t - \mathring{\mathbf{\Psi}}^t\right)\right\|_1.$$

APPENDIX B
PROOF OF MAIN RESULTS

This appendix contains the proofs of main theoretical results, including Proposition 1 and Theorem 3.

A. *Proof of Proposition 1*

Proof: Let

$$\left(\mathbf{\Pi}_M, \mathbf{\Psi}_M, \widetilde{\mathbf{Z}}^t\right) = \underset{\mathbf{\Pi}\in\mathcal{F},\mathbf{\Psi}\in\mathbb{R}^{d\times d},\mathbf{Z}=\widetilde{\mathbf{Z}}^t}{\arg\max} V^t\left(\mathbf{\Pi},\mathbf{\Psi},\mathbf{Z}\right).$$

Then we have

$$\begin{aligned}
V^t\left(\mathbf{\Pi}_M,\mathbf{\Psi}_M,\widetilde{\mathbf{Z}}^t\right) &= \left\langle \mathbf{S}, \mathbf{\Pi}_M - \mathring{\mathbf{\Pi}}^t\right\rangle + \left\|\mathbf{\Lambda}\odot\mathring{\mathbf{\Psi}}^t\right\|_1 - \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}_M\right\|_1 - \left\langle \begin{bmatrix}\mathbf{\Pi}_M - \mathring{\mathbf{\Pi}}^t\\ \mathbf{\Psi}_M - \mathring{\mathbf{\Psi}}^t\\ \widetilde{\mathbf{Z}}^t - \mathring{\mathbf{Z}}^t\end{bmatrix}, \begin{bmatrix}-\mathring{\mathbf{Z}}^t\\ \mathring{\mathbf{Z}}^t\\ \mathring{\mathbf{\Pi}}^t - \mathring{\mathbf{\Psi}}^t\end{bmatrix}\right\rangle \\
&= \left\langle \mathbf{S}, \mathbf{\Pi}_M - \mathring{\mathbf{\Pi}}^t\right\rangle + \left\|\mathbf{\Lambda}\odot\mathring{\mathbf{\Psi}}^t\right\|_1 - \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}_M\right\|_1 - \left\langle \begin{bmatrix}\mathbf{\Pi}_M\\ \mathbf{\Psi}_M\\ \widetilde{\mathbf{Z}}^t\end{bmatrix}, \begin{bmatrix}-\mathring{\mathbf{Z}}^t\\ \mathring{\mathbf{Z}}^t\\ \mathring{\mathbf{\Pi}}^t - \mathring{\mathbf{\Psi}}^t\end{bmatrix}\right\rangle \\
&\leq \frac{1}{t}\sum_{i=1}^{t}\left\{\left\langle \mathbf{S}, \mathbf{\Pi}_M - \mathbf{\Pi}^i\right\rangle + \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}^i\right\|_1 - \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}_M\right\|_1 - \left\langle \begin{bmatrix}\mathbf{\Pi}_M\\ \mathbf{\Psi}_M\\ \widetilde{\mathbf{Z}}^t\end{bmatrix}, \begin{bmatrix}-\mathbf{Z}^i\\ \mathbf{Z}^i\\ \mathbf{\Pi}^i - \mathbf{\Psi}^i\end{bmatrix}\right\rangle\right\} \\
&= \frac{1}{t}\sum_{i=1}^{t}\left\{\left\langle \mathbf{S}, \mathbf{\Pi}_M - \mathbf{\Pi}^i\right\rangle + \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}^i\right\|_1 - \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}_M\right\|_1 - \left\langle \begin{bmatrix}\mathbf{\Pi}_M - \mathbf{\Pi}^i\\ \mathbf{\Psi}_M - \mathbf{\Psi}^i\\ \widetilde{\mathbf{Z}}^t - \mathbf{Z}^i\end{bmatrix}, \begin{bmatrix}-\mathbf{Z}^i\\ \mathbf{Z}^i\\ \mathbf{\Pi}^i - \mathbf{\Psi}^i\end{bmatrix}\right\rangle\right\},
\end{aligned} \tag{12}$$

in which the inequality is due to the convexity of $\left\|\cdot\right\|_1$. For each term $i\in[1,t]$ in the summation of the last equality in (12), there should be

$$\begin{aligned}
&\left\langle \mathbf{S}, \mathbf{\Pi}_M - \mathbf{\Pi}^i\right\rangle + \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}^i\right\|_1 - \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}_M\right\|_1 - \left\langle \begin{bmatrix}\mathbf{\Pi}_M - \mathbf{\Pi}^i\\ \mathbf{\Psi}_M - \mathbf{\Psi}^i\\ \widetilde{\mathbf{Z}}^t - \mathbf{Z}^i\end{bmatrix}, \begin{bmatrix}-\mathbf{Z}^i\\ \mathbf{Z}^i\\ \mathbf{\Pi}^i - \mathbf{\Psi}^i\end{bmatrix}\right\rangle \\
=& \underbrace{\left\langle \mathbf{S}, \mathbf{\Pi}_M - \mathbf{\Pi}^i\right\rangle + \left\langle \mathbf{Z}^i, \mathbf{\Pi}_M - \mathbf{\Pi}^i\right\rangle}_{\text{I}} + \underbrace{\left\|\mathbf{\Lambda}\odot\mathbf{\Psi}^i\right\|_1 - \left\|\mathbf{\Lambda}\odot\mathbf{\Psi}_M\right\|_1 + \left\langle \mathbf{Z}^i, \mathbf{\Psi}^i - \mathbf{\Psi}_M\right\rangle}_{\text{II}} + \underbrace{\left\langle \mathbf{\Pi}^i - \mathbf{\Psi}^i, \mathbf{Z}^i - \widetilde{\mathbf{Z}}^t\right\rangle}_{\text{III}}.
\end{aligned} \tag{13}$$

We bound the terms I, II and III respectively.

For term I in (13), we have

$$
\begin{aligned}
\mathrm{I} = &\left\langle \mathbf{S}, \boldsymbol{\Pi}_M - \boldsymbol{\Pi}^i \right\rangle + \left\langle \mathbf{Z}^{i-1} - \rho\left(\boldsymbol{\Pi}^i - \boldsymbol{\Psi}^i\right), \boldsymbol{\Pi}_M - \boldsymbol{\Pi}^i \right\rangle \\
= &\left\langle \mathbf{S}, \boldsymbol{\Pi}_M - \boldsymbol{\Pi}^i \right\rangle + \left\langle \mathbf{Z}^{i-1} + \rho\left(\boldsymbol{\Psi}^{i-1} - \boldsymbol{\Pi}^i\right), \boldsymbol{\Pi}_M - \boldsymbol{\Pi}^i \right\rangle + \left\langle \rho\left(\boldsymbol{\Psi}^i - \boldsymbol{\Psi}^{i-1}\right), \boldsymbol{\Pi}_M - \boldsymbol{\Pi}^i \right\rangle \\
= &\left\langle \mathbf{S} + \mathbf{Z}^{i-1} + \rho\left(\boldsymbol{\Psi}^{i-1} - \boldsymbol{\Pi}^i\right), \boldsymbol{\Pi}_M - \boldsymbol{\Pi}^i \right\rangle + \left\langle \rho\left(\boldsymbol{\Psi}^i - \boldsymbol{\Psi}^{i-1}\right), \boldsymbol{\Pi}_M - \boldsymbol{\Pi}^i \right\rangle \\
\leq &\left\langle \rho\left(\boldsymbol{\Psi}^i - \boldsymbol{\Psi}^{i-1}\right), \boldsymbol{\Pi}_M - \boldsymbol{\Pi}^i \right\rangle \\
= &\frac{\rho}{2}\left(\left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^{i-1}\right\|_{\mathsf{F}}^2 - \left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^i\right\|_{\mathsf{F}}^2\right) + \frac{\rho}{2}\left(\left\|\boldsymbol{\Pi}^i - \boldsymbol{\Psi}^i\right\|_{\mathsf{F}}^2 - \left\|\boldsymbol{\Pi}^i - \boldsymbol{\Psi}^{i-1}\right\|_{\mathsf{F}}^2\right) \\
\leq &\frac{\rho}{2}\left(\left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^{i-1}\right\|_{\mathsf{F}}^2 - \left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^i\right\|_{\mathsf{F}}^2\right) + \frac{\rho}{2}\left\|\boldsymbol{\Pi}^i - \boldsymbol{\Psi}^i\right\|_{\mathsf{F}}^2 \\
= &\frac{\rho}{2}\left(\left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^{i-1}\right\|_{\mathsf{F}}^2 - \left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^i\right\|_{\mathsf{F}}^2\right) + \frac{1}{2\rho}\left\|\mathbf{Z}^{i-1} - \mathbf{Z}^i\right\|_{\mathsf{F}}^2,
\end{aligned}
$$

in which the first and last equality follows from the update of dual variable in Line 6 of Algorithm 2, and the first inequality is derived from the first-order optimality condition for Line 4.

For term II in (13), we have

$$
\begin{aligned}
\mathrm{II} = &\left\|\boldsymbol{\Lambda} \odot \boldsymbol{\Psi}^i\right\|_1 - \left\|\boldsymbol{\Lambda} \odot \boldsymbol{\Psi}_M\right\|_1 + \left\langle \mathbf{Z}^{i-1} - \rho\left(\boldsymbol{\Pi}^i - \boldsymbol{\Psi}^i\right), \boldsymbol{\Psi}^i - \boldsymbol{\Psi}_M \right\rangle \\
= &\left\|\boldsymbol{\Lambda} \odot \boldsymbol{\Psi}^i\right\|_1 - \left\|\boldsymbol{\Lambda} \odot \boldsymbol{\Psi}_M\right\|_1 - \left\langle \boldsymbol{\Lambda} \odot \boldsymbol{\Xi}^i, \boldsymbol{\Psi}^i - \boldsymbol{\Psi}_M \right\rangle, \ \exists \boldsymbol{\Xi}^i \in \partial\left\|\boldsymbol{\Psi}^i\right\|_1 \\
\leq &0,
\end{aligned}
$$

where the first equality follows from Line 6 of Algorithm 2, the second equality holds due to the first-order optimality condition for Line 5, and the last inequality is derived from the convexity of $\|\cdot\|_1$.

For term III in (13), utilizing the update of dual variable in Line 6 of Algorithm 2, we can get

$$
\begin{aligned}
\mathrm{III} = &\frac{1}{\rho}\left\langle \mathbf{Z}^{i-1} - \mathbf{Z}^i, \mathbf{Z}^i - \widetilde{\mathbf{Z}}^t \right\rangle \\
= &\frac{1}{2\rho}\left(\left\|\widetilde{\mathbf{Z}}^t - \mathbf{Z}^{i-1}\right\|_{\mathsf{F}}^2 - \left\|\widetilde{\mathbf{Z}}^t - \mathbf{Z}^i\right\|_{\mathsf{F}}^2 - \left\|\mathbf{Z}^i - \mathbf{Z}^{i-1}\right\|_{\mathsf{F}}^2\right).
\end{aligned}
$$

Finally, by substituting I, II, and III back into (12), we obtain

$$
\begin{aligned}
\omega^t = &V^t\left(\boldsymbol{\Pi}_M, \boldsymbol{\Psi}_M, \widetilde{\mathbf{Z}}^t\right) \\
\leq &\frac{1}{t}\sum_{i=1}^{t}\left\{\frac{\rho}{2}\left(\left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^{i-1}\right\|_{\mathsf{F}}^2 - \left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^i\right\|_{\mathsf{F}}^2\right) + \frac{1}{2\rho}\left(\left\|\widetilde{\mathbf{Z}}^t - \mathbf{Z}^{i-1}\right\|_{\mathsf{F}}^2 - \left\|\widetilde{\mathbf{Z}}^t - \mathbf{Z}^i\right\|_{\mathsf{F}}^2\right)\right\} \\
= &\frac{1}{t}\left(\frac{\rho}{2}\left(\left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^0\right\|_{\mathsf{F}}^2 - \left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^t\right\|_{\mathsf{F}}^2\right) + \frac{1}{2\rho}\left(\left\|\widetilde{\mathbf{Z}}^t - \mathbf{Z}^0\right\|_{\mathsf{F}}^2 - \left\|\widetilde{\mathbf{Z}}^t - \mathbf{Z}^t\right\|_{\mathsf{F}}^2\right)\right) \\
\leq &\frac{1}{t}\left(\frac{\rho}{2}\left\|\boldsymbol{\Pi}_M - \boldsymbol{\Psi}^0\right\|_{\mathsf{F}}^2 + \frac{1}{2\rho}\left\|\widetilde{\mathbf{Z}}^t - \mathbf{Z}^0\right\|_{\mathsf{F}}^2\right) \\
= &\frac{1}{t}\left(\frac{\rho}{2}\left\|\boldsymbol{\Pi}_M\right\|_{\mathsf{F}}^2 + \frac{1}{2\rho}\left\|\widetilde{\mathbf{Z}}^t\right\|_{\mathsf{F}}^2\right) \\
\leq &\frac{1}{t}\left(\frac{\rho r}{2} + \frac{\lambda^2 d^2}{2\rho}\right),
\end{aligned}
$$

where we use the settings $\boldsymbol{\Psi}^0 = \mathbf{0}$ and $\mathbf{Z}^0 = \mathbf{0}$ for the last equality. The last inequality follows from two facts:1) $\forall \boldsymbol{\Pi} \in \mathcal{F}$, $\|\boldsymbol{\Pi}\|_{\mathsf{F}}^2 \leq r$, and 2) $\left\|\widetilde{\mathbf{Z}}^t\right\|_{\mathsf{F}}^2 = \left\|\boldsymbol{\Lambda} \odot \operatorname{sgn}\left(\mathring{\boldsymbol{\Pi}}^t - \mathring{\boldsymbol{\Psi}}^t\right)\right\|_{\mathsf{F}}^2 \leq \lambda^2 d^2$.

Furthermore, if we set

$$
\rho = \frac{\lambda d}{\sqrt{r}} = \underset{z \in \mathbb{R}}{\arg\min}\ \frac{zr}{2} + \frac{\lambda^2 d^2}{2z},
$$

the stopping criterion is satisfied as $t \geq \frac{\lambda d \sqrt{r}}{\varepsilon}$. ■

B. Technical Lemmata

Before proving the next theorem, we first introduce several lemmata that will be used. For notational simplicity, we define $\|\mathbf{A}\|_{\max} := \max_{1 \leq i,j \leq d} |A_{ij}|$ and $\|\mathbf{A}\|_{\min} := \min_{1 \leq i,j \leq d} |A_{ij}|$ for a matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$. We also use $\mathbf{A}_{\mathcal{E}}$ to denote a matrix whose (i,j)-th entry is equal to A_{ij} if (i,j) is contained within an index set $\mathcal{E}$ and zero otherwise. Let $\overline{\mathcal{E}}$ denote the complement of $\mathcal{E}$.

Lemma 4. *Suppose that Assumption 1 holds. Let*

$$\mathcal{E}^k := \mathcal{S}^* \cup \mathcal{S}^k \text{ and } \mathcal{S}^k := \left\{(i,j) : \Lambda_{ij}^{k-1} < p_\lambda'(u)\right\}, \text{ with } u := c\lambda \text{ and } k \geq 1,$$

where $c = \frac{2+\sqrt{2}}{\delta}$ *is specified in Assumption 1. If* $\lambda \geq 2\left\|\mathbf{W}\right\|_{\max} + \sqrt{2\delta\varepsilon}$ *holds, then there must be*

$$\left|\mathcal{E}^k\right| \leq 2s^2, \left\|\mathbf{\Lambda}_{\mathcal{E}^k}^{k-1}\right\|_{\min} \geq \frac{\lambda}{2}, \tag{14}$$

and

$$\left\|\widetilde{\mathbf{\Pi}}^k - \mathbf{\Pi}^*\right\|_{\mathsf{F}} \leq \frac{2\left(\left\|\mathbf{W}_{\mathcal{E}^k}\right\|_{\mathsf{F}} + \left\|\mathbf{\Lambda}_{\mathcal{S}^*}^{k-1}\right\|_{\mathsf{F}}\right) + \sqrt{2\delta\varepsilon}}{\delta} \leq \frac{2+\sqrt{2}}{\delta}\lambda s. \tag{15}$$

Lemma 4 provides deterministic estimation error bounds for the approximate solutions.

Lemma 5. *Under Assumption 2, there exists a constant* $C > 0$ *depending on* L *such that*

$$\max_{i,j} \mathbb{P}\left(\left|W_{ij}\right| \geq z\right) \leq 2\exp\left(-\frac{4nz^2}{\left(C\lambda_1\right)^2}\right)$$

for $0 \leq z \leq C\lambda_1$.

Lemma 6. *Under Assumption 2, the following relation holds:*

$$\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_2 \lesssim \lambda_1\sqrt{\frac{s}{n}},$$

with probability at least $1 - \frac{2}{e^s}$.

Lemma 5 and Lemma 6 are concentration inequalities, serving as the foundations for the statistical analysis. They bound $\mathbf{W}$ in probability in terms of its elements magnitude and operator norm, respectively.

C. Proof of Theorem 3

Proof: We take

$$\lambda = 2C\lambda_1\sqrt{\frac{\log d}{n}} + \sqrt{\frac{\delta}{n}} \asymp \sqrt{\frac{\log d}{n}} \quad \text{and} \quad \varepsilon \leq \frac{1}{2n},$$

in which C is defined in Lemma 5. By Lemma 5, we have $\lambda \geq 2\left\|\mathbf{W}\right\|_{\max} + \sqrt{2\delta\varepsilon}$ holds with probability at least $1 - \frac{2}{d^2}$. Applying Lemma 2, we obtain

$$\begin{aligned}
\left\|\widetilde{\mathbf{\Pi}}^K - \mathbf{\Pi}^*\right\|_{\mathsf{F}} &\leq \frac{2}{\delta}\left(\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}} + \frac{\sqrt{2\delta\varepsilon}}{2}\right) + \tau\left\|\widetilde{\mathbf{\Pi}}^{K-1} - \mathbf{\Pi}^*\right\|_{\mathsf{F}} \\
&\leq \frac{2}{\left(1-\tau\right)\delta}\left(\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}} + \frac{\sqrt{2\delta\varepsilon}}{2}\right) + \tau^{K-1}\left\|\widetilde{\mathbf{\Pi}}^1 - \mathbf{\Pi}^*\right\|_{\mathsf{F}} \\
&\leq \frac{2}{\left(1-\tau\right)\delta}\left(\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}} + \frac{\sqrt{2\delta\varepsilon}}{2}\right) + \frac{2+\sqrt{2}}{\delta}\tau^{K-1}\lambda s,
\end{aligned}$$

where the last inequality follows from Lemma 4 for the special case $k = 1$. If $K \geq 1 + \frac{\log\frac{\lambda\sqrt{n}}{\lambda_1}}{\log\tau^{-1}} \gtrsim \log\log d$, we have

$$\tau^{K-1}\lambda s \leq \lambda_1 s\sqrt{\frac{1}{n}}.$$

We also have $\frac{\sqrt{2\delta\varepsilon}}{2} \lesssim \sqrt{\frac{1}{n}}$ since $\varepsilon \leq \frac{1}{2n}$, and $\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}} \leq \sqrt{s}\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_2 \lesssim \lambda_1 s\sqrt{\frac{1}{n}}$ with probability at least $1 - \frac{2}{e^s}$ by Lemma 6. With the above results, we complete the proof. ∎

Appendix C
Proofs of Technical Lemmata

In this appendix, we will prove all lemmata, including those already presented and some newly introduced fundamental lemmata. In these newly introduced lemmata, Lemma 7, proposed in [4], bounds the curvature of the objective function along the Fantope and away from the truth. Lemma 8 and Lemma 9 characterize estimation error of the approximate solution $\tilde{\mathbf{\Pi}}$ to the reduced problem (4).

Lemma 7 (Lemma 3.1 in [4])**.** *For any $\mathbf{\Pi} \in \mathcal{F}$, the following relation holds:*

$$\frac{\delta}{2} \left\| \mathbf{\Pi} - \mathbf{\Pi}^* \right\|_{\mathsf{F}}^2 \le \left\langle \mathbf{\Sigma}, \mathbf{\Pi}^* - \mathbf{\Pi} \right\rangle .$$

Lemma 8. *For an approximate solution $\tilde{\mathbf{\Pi}}$ to problem (4), there should be*

$$\frac{\delta}{2} \left\| \tilde{\mathbf{\Pi}} - \mathbf{\Pi}^* \right\|_{\mathsf{F}}^2 \le \left\langle \mathbf{W}, \tilde{\mathbf{\Pi}} - \mathbf{\Pi}^* \right\rangle + \left\| \mathbf{\Lambda} \odot \mathbf{\Pi}^* \right\|_1 - \left\| \mathbf{\Lambda} \odot \tilde{\mathbf{\Pi}} \right\|_1 + \varepsilon.$$

Proof: Suppose Algorithm 2 executes T iterations in total, i.e., $\tilde{\mathbf{\Pi}} = \mathring{\mathbf{\Pi}}^T$. From Line 8 of Algorithm 2, we know that

$$\omega^T \le \varepsilon. \tag{16}$$

By the definition, we have

$$V^T \left(\mathbf{\Pi}^*, \mathbf{\Pi}^*, \widetilde{\mathbf{Z}}^T \right) \le \underset{\mathbf{\Pi} \in \mathcal{F}, \mathbf{\Psi} \in \mathbb{R}^{d \times d}, \mathbf{Z} = \widetilde{\mathbf{Z}}^T}{\text{maximize}} V^T \left(\mathbf{\Pi}, \mathbf{\Psi}, \mathbf{Z} \right) = \omega^T. \tag{17}$$

Combining (16) with (17), and expanding $V^T \left(\mathbf{\Pi}^*, \mathbf{\Pi}^*, \widetilde{\mathbf{Z}}^T \right)$, we obtain

$$\underbrace{\left\langle \mathbf{S}, \mathbf{\Pi}^* - \mathring{\mathbf{\Pi}}^T \right\rangle}_{\mathrm{I}} + \underbrace{\left\| \mathbf{\Lambda} \odot \mathring{\mathbf{\Psi}}^T \right\|_1 - \left\| \mathbf{\Lambda} \odot \mathbf{\Pi}^* \right\|_1}_{\mathrm{II}} - \underbrace{\left\langle \begin{bmatrix} \mathbf{\Pi}^* - \mathring{\mathbf{\Pi}}^T \\ \mathbf{\Pi}^* - \mathring{\mathbf{\Psi}}^T \\ \widetilde{\mathbf{Z}}^T - \mathring{\mathbf{Z}}^T \end{bmatrix}, \begin{bmatrix} -\mathring{\mathbf{Z}}^T \\ \mathring{\mathbf{Z}}^T \\ \mathring{\mathbf{\Pi}}^T - \mathring{\mathbf{\Psi}}^T \end{bmatrix} \right\rangle}_{\mathrm{III}} \le \varepsilon. \tag{18}$$

Now we bound terms $\mathrm{I}, \mathrm{II}, \mathrm{III}$ in (18) respectively.

For term I,

$$\begin{aligned} \mathrm{I} &= \left\langle \mathbf{W}, \mathbf{\Pi}^* - \mathring{\mathbf{\Pi}}^T \right\rangle + \left\langle \mathbf{\Sigma}, \mathbf{\Pi}^* - \mathring{\mathbf{\Pi}}^T \right\rangle \\ &\ge \left\langle \mathbf{W}, \mathbf{\Pi}^* - \mathring{\mathbf{\Pi}}^T \right\rangle + \frac{\delta}{2} \left\| \mathring{\mathbf{\Pi}}^T - \mathbf{\Pi}^* \right\|_{\mathsf{F}}^2 , \end{aligned} \tag{19}$$

in which the inequality holds by applying Lemma 7 and setting $\mathbf{\Pi} = \mathring{\mathbf{\Pi}}^T \in \mathcal{F}$.

For term II,

$$\begin{aligned} \mathrm{II} &= \left\| \mathbf{\Lambda} \odot \left(\mathring{\mathbf{\Pi}}^T + \mathring{\mathbf{\Psi}}^T - \mathring{\mathbf{\Pi}}^T \right) \right\|_1 - \left\| \mathbf{\Lambda} \odot \mathbf{\Pi}^* \right\|_1 \\ &\ge \left\| \mathbf{\Lambda} \odot \mathring{\mathbf{\Pi}}^T \right\|_1 - \left\| \mathbf{\Lambda} \odot \mathbf{\Pi}^* \right\|_1 - \left\| \mathbf{\Lambda} \odot \left(\mathring{\mathbf{\Psi}}^T - \mathring{\mathbf{\Pi}}^T \right) \right\|_1 , \end{aligned} \tag{20}$$

where the inequality follows from the triangle inequality.

For term III,

$$\begin{aligned} \mathrm{III} &= \left\langle \mathring{\mathbf{\Pi}}^T - \mathbf{\Pi}^*, \mathring{\mathbf{Z}}^T \right\rangle + \left\langle \mathbf{\Pi}^* - \mathring{\mathbf{\Psi}}^T, \mathring{\mathbf{Z}}^T \right\rangle + \left\langle \widetilde{\mathbf{Z}}^T - \mathring{\mathbf{Z}}^T, \mathring{\mathbf{\Pi}}^T - \mathring{\mathbf{\Psi}}^T \right\rangle \\ &= \left\langle \widetilde{\mathbf{Z}}^T, \mathring{\mathbf{\Pi}}^T - \mathring{\mathbf{\Psi}}^T \right\rangle \\ &= - \left\| \mathbf{\Lambda} \odot \left(\mathring{\mathbf{\Pi}}^T - \mathring{\mathbf{\Psi}}^T \right) \right\|_1 , \end{aligned} \tag{21}$$

where the last equality holds by substituting $\widetilde{\mathbf{Z}}^T$.

Plugging (19), (20) and (21) into (18), we conclude the proof. ■

Lemma 9. *Assume there exists a set $\mathcal{E}$ such that*

$$\mathcal{S}^* \subseteq \mathcal{E}, \left| \mathcal{E} \right| \le 2 \left| \mathcal{S}^* \right| \le 2s^2 \text{ and } \left\| \mathbf{\Lambda}_{\overline{\mathcal{E}}} \right\|_{\min} \ge \frac{\lambda}{2}.$$

If $\lambda \geq 2\|\mathbf{W}\|_{\max} + \sqrt{2\delta\varepsilon}$*, then we have*

$$\left\|\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right\|_{\mathsf{F}} \leq \frac{2\left(\|\mathbf{W}_{\mathcal{E}}\|_{\mathsf{F}} + \|\mathbf{\Lambda}_{\mathcal{S}^*}\|_{\mathsf{F}}\right) + \sqrt{2\delta\varepsilon}}{\delta} \leq \frac{2+\sqrt{2}}{\delta}\lambda s.$$

Proof: By Lemma 8, we have

$$\frac{\delta}{2}\left\|\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right\|_{\mathsf{F}}^2 \leq \underbrace{\left\langle \mathbf{W}, \widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^* \right\rangle}_{\mathrm{I}} + \underbrace{\|\mathbf{\Lambda}\odot\mathbf{\Pi}^*\|_1 - \left\|\mathbf{\Lambda}\odot\widetilde{\mathbf{\Pi}}\right\|_1}_{\mathrm{II}} + \varepsilon. \tag{22}$$

For term I in (22),

$$\begin{aligned}
\mathrm{I} &= \left\langle \mathbf{W}_{\mathcal{E}}, \left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\mathcal{E}} \right\rangle + \left\langle \mathbf{W}_{\overline{\mathcal{E}}}, \left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\overline{\mathcal{E}}} \right\rangle \\
&\leq \|\mathbf{W}_{\mathcal{E}}\|_{\mathsf{F}} \left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\mathcal{E}}\right\|_{\mathsf{F}} + \|\mathbf{W}_{\overline{\mathcal{E}}}\|_{\max} \left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\overline{\mathcal{E}}}\right\|_1 \\
&\leq \|\mathbf{W}_{\mathcal{E}}\|_{\mathsf{F}} \left\|\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right\|_{\mathsf{F}} + \|\mathbf{W}\|_{\max} \left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\overline{\mathcal{E}}}\right\|_1,
\end{aligned} \tag{23}$$

where the first inequality follows from the Hölder's inequality.

For II in (22),

$$\begin{aligned}
\mathrm{II} &= \|(\mathbf{\Lambda}\odot\mathbf{\Pi}^*)_{\mathcal{S}^*}\|_1 - \left\|\mathbf{\Lambda}\odot\widetilde{\mathbf{\Pi}}\right\|_1 \\
&= \|(\mathbf{\Lambda}\odot\mathbf{\Pi}^*)_{\mathcal{S}^*}\|_1 - \left\|\left(\mathbf{\Lambda}\odot\widetilde{\mathbf{\Pi}}\right)_{\mathcal{S}^*}\right\|_1 - \left\|\left(\mathbf{\Lambda}\odot\widetilde{\mathbf{\Pi}}\right)_{\overline{\mathcal{S}^*}}\right\|_1 \\
&\leq \left\|\left(\mathbf{\Lambda}\odot\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)\right)_{\mathcal{S}^*}\right\|_1 - \left\|\left(\mathbf{\Lambda}\odot\widetilde{\mathbf{\Pi}}\right)_{\overline{\mathcal{S}^*}}\right\|_1 \\
&\leq \left\|\left(\mathbf{\Lambda}\odot\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)\right)_{\mathcal{S}^*}\right\|_1 - \left\|\left(\mathbf{\Lambda}\odot\widetilde{\mathbf{\Pi}}\right)_{\overline{\mathcal{E}}}\right\|_1 \\
&\leq \|\mathbf{\Lambda}_{\mathcal{S}^*}\|_{\mathsf{F}} \left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\mathcal{S}^*}\right\|_{\mathsf{F}} - \left\|\left(\mathbf{\Lambda}\odot\widetilde{\mathbf{\Pi}}\right)_{\overline{\mathcal{E}}}\right\|_1 \\
&\leq \|\mathbf{\Lambda}_{\mathcal{S}^*}\|_{\mathsf{F}} \left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\mathcal{S}^*}\right\|_{\mathsf{F}} - \|\mathbf{\Lambda}_{\overline{\mathcal{E}}}\|_{\min} \left\|\widetilde{\mathbf{\Pi}}_{\overline{\mathcal{E}}}\right\|_1 \\
&= \|\mathbf{\Lambda}_{\mathcal{S}^*}\|_{\mathsf{F}} \left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\mathcal{S}^*}\right\|_{\mathsf{F}} - \|\mathbf{\Lambda}_{\overline{\mathcal{E}}}\|_{\min} \left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\overline{\mathcal{E}}}\right\|_1 \\
&\leq \|\mathbf{\Lambda}_{\mathcal{S}^*}\|_{\mathsf{F}} \left\|\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right\|_{\mathsf{F}} - \|\mathbf{\Lambda}_{\overline{\mathcal{E}}}\|_{\min} \left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\overline{\mathcal{E}}}\right\|_1.
\end{aligned} \tag{24}$$

The first equality and the last equality are due to the fact that $\mathbf{\Pi}^*$ has non-zero elements only on its support $\mathcal{S}^*$. The first inequality follows from the triangular inequality and the third inequality follows from the Hölder's inequality.

Plugging (23) and (24) into (22), we obtain

$$\begin{aligned}
\frac{\delta}{2}\left\|\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right\|_{\mathsf{F}}^2 &\leq \left(\|\mathbf{W}_{\mathcal{E}}\|_{\mathsf{F}} + \|\mathbf{\Lambda}_{\mathcal{S}^*}\|_{\mathsf{F}}\right)\left\|\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right\|_{\mathsf{F}} + \left(\|\mathbf{W}\|_{\max} - \|\mathbf{\Lambda}_{\overline{\mathcal{E}}}\|_{\min}\right)\left\|\left(\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right)_{\overline{\mathcal{E}}}\right\|_1 + \varepsilon \\
&\leq \left(\|\mathbf{W}_{\mathcal{E}}\|_{\mathsf{F}} + \|\mathbf{\Lambda}_{\mathcal{S}^*}\|_{\mathsf{F}}\right)\left\|\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right\|_{\mathsf{F}} + \varepsilon,
\end{aligned}$$

in which we use the condition that $\|\mathbf{\Lambda}_{\overline{\mathcal{E}}}\|_{\min} \geq \frac{\lambda}{2} \geq \|\mathbf{W}\|_{\max}$ for the last inequality. Finally, by solving the quadratic inequality, we have

$$\begin{aligned}
\left\|\widetilde{\mathbf{\Pi}} - \mathbf{\Pi}^*\right\|_{\mathsf{F}} &\leq \frac{2\left(\|\mathbf{W}_{\mathcal{E}}\|_{\mathsf{F}} + \|\mathbf{\Lambda}_{\mathcal{S}^*}\|_{\mathsf{F}}\right) + \sqrt{2\delta\varepsilon}}{\delta} \\
&\leq \frac{2}{\delta}\left(\sqrt{|\mathcal{E}|}\,\|\mathbf{W}\|_{\max} + \sqrt{|\mathcal{E}|}\frac{\sqrt{2\delta\varepsilon}}{2} + \sqrt{|\mathcal{S}^*|}\lambda\right) \\
&\leq \frac{2+\sqrt{2}}{\delta}\lambda s,
\end{aligned}$$

in which the last two inequalities are derived from the conditions that $\sqrt{2|\mathcal{S}^*|} \geq \sqrt{|\mathcal{E}|} \geq \sqrt{|\mathcal{S}^*|} = s \geq 1$ and $\lambda \geq 2\|\mathbf{W}\|_{\max} + \sqrt{2\delta\varepsilon}$. ■

Proof of Lemma 4: Firstly, we prove (14) by induction. For $k=1$, since $\Lambda_{ij}^0 = p_\lambda'\left(\left|\widetilde{\Pi}_{ij}^0\right|\right) = p_\lambda'(0) = \lambda \geq p_\lambda'(u)$, we have $\mathcal{S}^1 = \emptyset$ and $\mathcal{E}^1 = \mathcal{S}^*$ such that

$$\left|\mathcal{E}^1\right| \leq 2s^2 \quad \text{and} \quad \left\|\mathbf{\Lambda}_{\overline{\mathcal{E}^1}}^0\right\|_{\min} \geq \lambda.$$

Assume (14) holds for some $k-1$ with $k \geq 2$, we show that (14) also holds for k. According to the monotonicity of $p_\lambda'(\cdot)$, i.e., property (b) in Assumption 1, for any $(i,j) \in \mathcal{S}^k$, we have $\left|\widetilde{\Pi}_{ij}^{k-1}\right| \geq u$. Then there should be

$$
\begin{aligned}
\sqrt{\left|\mathcal{S}^k \backslash \mathcal{S}^*\right|} \leq & \sqrt{\sum_{(i,j)\in\mathcal{S}^k\backslash\mathcal{S}^*} \left(u^{-1}\widetilde{\Pi}_{ij}^{k-1}\right)^2} \\
\leq & u^{-1}\left\|\left(\widetilde{\mathbf{\Pi}}^{k-1}\right)_{\mathcal{S}^k\backslash\mathcal{S}^*}\right\|_{\mathsf{F}} \\
\leq & u^{-1}\left\|\left(\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right)_{\mathcal{S}^k\backslash\mathcal{S}^*}\right\|_{\mathsf{F}} \\
\leq & u^{-1}\left\|\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right\|_{\mathsf{F}},
\end{aligned}
\tag{25}
$$

where the third inequality is due to the fact that $\mathbf{\Pi}^*_{\mathcal{S}^k\backslash\mathcal{S}^*} = \mathbf{0}$. Applying Lemma 9 with $\mathcal{E} = \mathcal{E}^{k-1}$, $\mathbf{\Lambda} = \mathbf{\Lambda}^{k-2}$ and $\widetilde{\mathbf{\Pi}} = \widetilde{\mathbf{\Pi}}^{k-1}$, we obtain

$$
\left\|\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right\|_{\mathsf{F}} \leq \frac{2+\sqrt{2}}{\delta}\lambda s. \tag{26}
$$

Combining (25) with (26),

$$
\sqrt{\left|\mathcal{S}^k \backslash \mathcal{S}^*\right|} \leq u^{-1}\frac{2+\sqrt{2}}{\delta}\lambda s = s,
$$

which further implies that

$$
\left|\mathcal{E}^k\right| = \left|\mathcal{S}^*\right| + \left|\mathcal{S}^k\backslash\mathcal{S}^*\right| \leq 2s^2.
$$

By the property (d) in Assumption 1, we have

$$
\left\|\mathbf{\Lambda}^{k-1}_{\overline{\mathcal{E}^k}}\right\|_{\min} \geq \left\|\mathbf{\Lambda}^{k-1}_{\overline{\mathcal{S}^k}}\right\|_{\min} \geq p_\lambda'(u) \geq \frac{\lambda}{2}.
$$

As a consequence, (15) can be proven by directly applying Lemma 9 with $\mathcal{E} = \mathcal{E}^k$, $\mathbf{\Lambda} = \mathbf{\Lambda}^{k-1}$ and $\widetilde{\mathbf{\Pi}} = \widetilde{\mathbf{\Pi}}^k$. ■

Proof of Lemma 2: By Lemma 4, we have

$$
\left\|\widetilde{\mathbf{\Pi}}^{k}-\mathbf{\Pi}^*\right\|_{\mathsf{F}} \leq \frac{2}{\delta}\left(\underbrace{\left\|\mathbf{W}_{\mathcal{E}^k}\right\|_{\mathsf{F}}}_{\mathrm{I}} + \underbrace{\left\|\mathbf{\Lambda}^{k-1}_{\mathcal{S}^*}\right\|_{\mathsf{F}}}_{\mathrm{II}} + \frac{\sqrt{2\delta\varepsilon}}{2}\right). \tag{27}
$$

We then bound I and II respectively.

For I, we divide the set $\mathcal{E}^k$ into $\mathcal{S}^*$ and $\mathcal{E}^k\backslash\mathcal{S}^*$ to get

$$
\begin{aligned}
\mathrm{I} \leq & \left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}} + \left\|\mathbf{W}_{\mathcal{E}^k\backslash\mathcal{S}^*}\right\|_{\mathsf{F}} \\
\leq & \left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}} + \sqrt{\left|\mathcal{E}^k\backslash\mathcal{S}^*\right|}\left\|\mathbf{W}\right\|_{\max} \\
\leq & \left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}} + u^{-1}\left\|\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right\|_{\mathsf{F}}\left\|\mathbf{W}\right\|_{\max} \\
\leq & \left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}} + \frac{\lambda}{2}u^{-1}\left\|\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right\|_{\mathsf{F}},
\end{aligned}
\tag{28}
$$

where the third inequality follows from (25) and the last inequality follows from the condition that $\left\|\mathbf{W}\right\|_{\max} \leq \frac{\lambda}{2}$.

For II, under Assumption 1, if $\left|\widetilde{\Pi}_{ij}^{k-1} - \Pi_{ij}^*\right| \geq u$ we have

$$
\left|\Lambda_{ij}^{k-1}\right| \leq \lambda \leq \lambda u^{-1}\left|\widetilde{\Pi}_{ij}^{k-1} - \Pi_{ij}^*\right|, \tag{29}
$$

otherwise by Assumption 3 and the triangular inequality we have

$$
\left|\widetilde{\Pi}_{ij}^{k-1}\right| \geq \left|\Pi_{ij}^*\right| - \left|\widetilde{\Pi}_{ij}^{k-1} - \Pi_{ij}^*\right| \geq \alpha\lambda,
$$

which further implies

$$
\left|\Lambda_{ij}^{k-1}\right| = p_\lambda'\left(\left|\widetilde{\Pi}_{ij}^{k-1}\right|\right) = 0. \tag{30}
$$

Combining (29) with (30), we obtain

$$\begin{aligned}\mathrm{II} \leq& \lambda u^{-1}\left\|\left(\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right)_{\mathcal{S}^*}\right\|_{\mathsf{F}}\\ \leq&\lambda u^{-1}\left\|\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right\|_{\mathsf{F}}.\end{aligned}\tag{31}$$

Finally, substituting (28) and (31) back into (27), we have

$$\begin{aligned}\left\|\widetilde{\mathbf{\Pi}}^{k}-\mathbf{\Pi}^*\right\|_{\mathsf{F}} \leq& \frac{2}{\delta}\left(\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}}+\frac{\sqrt{2\delta\varepsilon}}{2}+\frac{3\lambda}{2}u^{-1}\left\|\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right\|_{\mathsf{F}}\right)\\ =&\frac{2}{\delta}\left(\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_{\mathsf{F}}+\frac{\sqrt{2\delta\varepsilon}}{2}\right)+\tau\left\|\widetilde{\mathbf{\Pi}}^{k-1}-\mathbf{\Pi}^*\right\|_{\mathsf{F}},\end{aligned}$$

where $\tau=\frac{3}{2+\sqrt{2}}\in(0,1)$. ■

Proof of Lemma 5: See Corollary 3.3 in [4] for details. ■

Proof of Lemma 6: We first list two useful properties implied from Assumption 2. For $\forall \mathbf{v}\in\mathbb{R}^d\colon \mathbf{v}^\top\mathbf{v}=1$ and $z>0$, we have: 1) $\mathbb{P}\left(\left|\langle\mathbf{x},\mathbf{v}\rangle\right|\geq z\right)\leq 2\exp\left(-\frac{L}{\lambda_1}z^2\right)$, and 2) $\mathbb{E}\left[\exp\left(z\langle\mathbf{v},\mathbf{x}\rangle\right)\right]\leq\exp\left(C_1\frac{\lambda_1}{L}z^2\right)$ for some constant $C_1>0$. The first property is based on the observation that $\left\|\mathbf{\Sigma}^{1/2}\mathbf{v}\right\|_2^2\leq\lambda_1$, and the second property is an equivalent form of definition for sub-Gaussian variables [5].

Define

$$\mathcal{B}\left(d,\mathcal{S}^*\right)\coloneqq\left\{\mathbf{v}\in\mathbb{R}^d\colon \mathbf{v}^\top\mathbf{v}=1,\operatorname{supp}\left(\mathbf{v}\mathbf{v}^\top\right)=\mathcal{S}^*\right\},$$

then we have

$$\begin{aligned}\left\|\mathbf{W}_{\mathcal{S}^*}\right\|_2=&\left\|\left[\frac{1}{n}\sum_{i=1}^n\left\{\mathbf{x}_i\mathbf{x}_i^\top-\mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i^\top\right]\right\}\right]_{\mathcal{S}^*}\right\|_2\\ =&\sup_{\mathbf{v}\in\mathcal{B}(d,\mathcal{S}^*)}\left|\frac{1}{n}\sum_{i=1}^n\left\{\left\langle\mathbf{x}_i,\mathbf{v}\right\rangle^2-\mathbb{E}\left\langle\mathbf{x}_i,\mathbf{v}\right\rangle^2\right\}\right|.\end{aligned}$$

Define $\mu_{\mathbf{v}}\coloneqq\mathbb{E}\left\langle\mathbf{x}_i,\mathbf{v}\right\rangle^2$ for notational simplicity, and fix some $\mathbf{v}\in\mathcal{N}_{\frac{1}{8}}$ where $\mathcal{N}_{\frac{1}{8}}$ is an $\frac{1}{8}$-net of $\mathcal{B}\left(d,\mathcal{S}^*\right)$. The next proof follows from Proposition 1 in [6], and we present the details here for completeness. Specifically, since as $z\to 0$,

$$\begin{aligned}1+\frac{1}{2}\mu_{\mathbf{v}}z^2+o\left(z^2\right)=&\mathbb{E}\left[\exp\left(z\left\langle\mathbf{v},\mathbf{x}_i\right\rangle\right)\right]\\ \leq&\exp\left(C_1\frac{\lambda_1}{L}z^2\right)\\ =&1+\frac{1}{2}C_1\frac{\lambda_1}{L}z^2+o\left(z^2\right),\end{aligned}$$

we obtain $\mu_{\mathbf{v}}\frac{L}{\lambda_1}\leq C_1$. Then for any integer $m\geq 2$,

$$\begin{aligned}&\mathbb{E}\left[\left|\left\langle\mathbf{x}_i,\mathbf{v}\right\rangle^2-\mu_{\mathbf{v}}\right|^m\right]\\ \leq&\int_0^\infty\mathbb{P}\left\{\left\langle\mathbf{x}_i,\mathbf{v}\right\rangle^2-\mu_{\mathbf{v}}\geq z^{\frac{1}{m}}\right\}dz+\mu_{\mathbf{v}}^m\\ =&\int_0^\infty\mathbb{P}\left\{\left|\left\langle\mathbf{x}_i,\mathbf{v}\right\rangle\right|\geq\left(z^{\frac{1}{m}}+\mu_{\mathbf{v}}\right)^{\frac{1}{2}}\right\}dz+\mu_{\mathbf{v}}^m\\ \leq&2\int_0^\infty\exp\left(-\frac{L}{\lambda_1}\left(z^{\frac{1}{m}}+\mu_{\mathbf{v}}\right)\right)dz+\mu_{\mathbf{v}}^m\\ =&m!\left(\frac{\lambda_1}{L}\right)^m\left(2\exp\left(-\frac{L}{\lambda_1}\mu_{\mathbf{v}}\right)+\frac{1}{m!}\left(\frac{L\mu_{\mathbf{v}}}{\lambda_1}\right)^m\right)\\ \leq&C_2m!\left(\frac{\lambda_1}{L}\right)^m,\end{aligned}$$

in which C_2 is a constant depending on C_1. The last inequality follows from the fact that $\exp\left(-\frac{L}{\lambda_1}\mu_{\mathbf{v}}\right) \leq 1$ and $\frac{1}{m!}C_1^m$ is bounded above. By the Bernstein's inequality (see Lemma 2.2.11 in [7]),

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\langle\mathbf{x}_i,\mathbf{v}\rangle^2-\mu_{\mathbf{v}}\right\}\right|>z\right)\\
=&\mathbb{P}\left(\left|\sum_{i=1}^{n}\left\{\langle\mathbf{x}_i,\mathbf{v}\rangle^2-\mu_{\mathbf{v}}\right\}\right|>nz\right)\\
\leq&2\exp\left(-\frac{nz^2}{4C_2\left(\frac{\lambda_1}{L}\right)^2+\frac{2\lambda_1}{L}z}\right).
\end{aligned}$$

There exists a constant C_3 satisfying $C_3\lambda_1 \geq 4\left(1+\sqrt{1+C_2}\right)\frac{\lambda_1}{L}$ such that for $0 \leq z \leq C_3\lambda_1$,

$$\exp\left(-\frac{nz^2}{4C_2\left(\frac{\lambda_1}{L}\right)^2+\frac{2\lambda_1}{L}z}\right) \leq \exp\left(-\frac{4nz^2}{\left(C_3\lambda_1\right)^2}\right).$$

Now we unfix $\mathbf{v}$, and consider the event

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\langle\mathbf{x}_i,\mathbf{v}\rangle^2-\mathbb{E}\left[\langle\mathbf{x}_i,\mathbf{v}\rangle^2\right]\right\}\right| \leq z, \quad \forall\mathbf{v}\in\mathcal{N}_{\frac{1}{8}}.$$

It holds at least

$$1-2\left|\mathcal{N}_{\frac{1}{8}}\right|\exp\left(-\frac{4nz^2}{\left(C_3\lambda_1\right)^2}\right) \geq 1-2\exp\left(3s-\frac{4nz^2}{\left(C_3\lambda_1\right)^2}\right),$$

in which we utilize the fact that $\left|\mathcal{N}_{\frac{1}{8}}\right| \leq 17^s$ (see Example 5.8 in [8]). We finish the proof by using Lemma 2.2 in [9] and taking $z = C_3\lambda_1\sqrt{\frac{s}{n}} \leq C_3\lambda_1$. ∎

REFERENCES

[1] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[2] B. He, H. Liu, Z. Wang, and X. Yuan, "A strictly contractive peaceman–rachford splitting method for convex programming," *SIAM Journal on Optimization*, vol. 24, no. 3, pp. 1011–1040, 2014.
[3] J. Lei and V. Q. Vu, "Sparsistency and agnostic inference in sparse pca," 2015.
[4] V. Q. Vu, J. Cho, J. Lei, and K. Rohe, "Fantope projection and selection: A near-optimal convex relaxation of sparse PCA," in *Advances in Neural Information Processing Systems*, vol. 26. Curran Associates, Inc., 2013.
[5] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018, vol. 47.
[6] T. Wang, Q. Berthet, and R. J. Samworth, "Statistical and computational trade-offs in estimation of sparse principal components," 2016.
[7] J. Wellner *et al.*, *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media, 2013.
[8] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge university press, 2019, vol. 48.
[9] R. VERSHYNIN, "How close is the sample covariance matrix to the actual covariance matrix?" *identity (up to normalization)*, vol. 1, p. 1.