

A Proof of Main Results for Matrix Sensing (Cont'd)

We continue with the proof of Theorem 3.4 for the alternating gradient and gradient descent algorithms.

A.1 Proof of Theorem 3.4 (Alternating Gradient Descent)

Proof. Throughout the proof for alternating gradient descent, we define a sufficiently large constant ξ . Moreover, we assume that at the t -th iteration, there exists a matrix factorization of M^*

$$M^* = \bar{U}^{*(t)} V^{*(t)\top},$$

where $\bar{U}^{*(t)} \in \mathbb{R}^{m \times k}$ is an orthonormal matrix. We use the following projected oracle divergence

$$\mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) = \left\langle \nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t)}) - \nabla_V \mathcal{F}(\bar{U}^{*(t)}, V^{(t)}), \frac{V^{(t+0.5)} - V^{*(t)}}{\|V^{(t+0.5)} - V^{*(t)}\|_F} \right\rangle.$$

The first lemma is parallel to Lemma 4.3 for alternating exact minimization.

Lemma A.1. Suppose that δ_{2k} , $\bar{U}^{(t)}$, and $V^{(t)}$ satisfy

$$\delta_{2k} \leq \frac{\sqrt{2}(1 - \delta_{2k})\sigma_k}{8\xi k\sigma_1}, \quad \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \leq \frac{\sigma_k^2}{4\xi\sigma_1^2}, \quad \text{and} \quad \|V^{(t)} - V^{*(t)}\|_F \leq \frac{\sigma_1\sqrt{k}}{2}. \quad (\text{A.1})$$

Then we have

$$\mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) \leq \frac{(1 + \delta_{2k})\sigma_k}{\xi} \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F.$$

The proof of Lemma A.1 is provided in Appendix C.1. Lemma A.1 illustrates the projected oracle divergence diminishes with the estimation error of $\bar{U}^{(t)}$, when $\bar{U}^{(t)}$ and $V^{(t)}$ are sufficiently close to $\bar{U}^{*(t)}$ and $V^{*(t)}$.

Lemma A.2. Suppose that the step size parameter η satisfies

$$\eta = \frac{1}{1 + \delta_{2k}}. \quad (\text{A.2})$$

Then we have

$$\|V^{(t+0.5)} - V^*\|_F \leq \sqrt{\delta_{2k}} \|V^{(t)} - V^*\|_F + \frac{2}{1 + \delta_{2k}} \mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}).$$

The proof of Lemma A.2 is in Appendix C.2. Lemma A.2 characterizes the progress of a gradient descent step with a pre-specified fixed step size. A more practical option is adaptively selecting η using the backtracking line search procedure, and similar results can be guaranteed. See [20] for details. The following lemma characterizes the effect of the renormalization step using QR decomposition.

Lemma A.3. Suppose that $V^{(t+0.5)}$ satisfies

$$\|V^{(t+0.5)} - V^{*(t)}\|_F \leq \frac{\sigma_k}{4}. \quad (\text{A.3})$$

Then there exists a factorization of $M^* = U^{*(t+1)} \bar{V}^{*(t+1)}$ such that $\bar{V}^{*(t+1)} \in \mathbb{R}^{n \times k}$ is an orthonormal matrix, and

$$\begin{aligned} \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F &\leq \frac{2}{\sigma_k} \|V^{(t+0.5)} - V^{*(t)}\|_F, \\ \|U^{(t)} - U^{*(t+1)}\|_F &\leq \frac{3\sigma_1}{\sigma_k} \|V^{(t+0.5)} - V^{*(t)}\|_F + \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F, \end{aligned}$$

The proof of Lemma A.3 is provided in Appendix C.3. The next lemma quantifies the accuracy of the initial solutions.

Lemma A.4. Suppose that δ_{2k} satisfies

$$\delta_{2k} \leq \frac{\sigma_k^6}{192\xi^2 k\sigma_1^6}. \quad (\text{A.4})$$

Then we have

$$\|\bar{U}^{(0)} - \bar{U}^{*(0)}\|_F \leq \frac{\sigma_k^2}{4\xi\sigma_1^2} \quad \text{and} \quad \|V^{(0)} - V^{*(0)}\|_F \leq \frac{\sigma_k^2}{2\xi\sigma_1} \leq \frac{\sigma_1\sqrt{k}}{2}.$$

The proof of Lemma A.4 is in Appendix C.4. Lemma A.4 indicates that the initial solutions $\bar{U}^{(0)}$ and $V^{(0)}$ attain sufficiently small estimation errors.

Combining Lemmas A.1, A.2, 4.5, , we obtain the following corollary for a complete iteration of updating V .

Corollary A.5. Suppose that δ_{2k} , $\bar{U}^{(t)}$, and $V^{(t)}$ satisfy

$$\delta_{2k} \leq \frac{\sigma_k^6}{192\xi^2 k \sigma_1^6}, \quad \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \leq \frac{\sigma_k^2}{4\xi\sigma_1^2}, \quad \text{and} \quad \|V^{(t)} - V^{*(t)}\|_F \leq \frac{\sigma_k^2}{2\xi\sigma_1}. \quad (\text{A.5})$$

We then have

$$\|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F \leq \frac{\sigma_k^2}{4\xi\sigma_1^2} \quad \text{and} \quad \|U^{(t)} - U^{*(t+1)}\|_F \leq \frac{\sigma_k^2}{2\xi\sigma_1}.$$

Moreover, we have

$$\|V^{(t+0.5)} - V^{*(t)}\|_F \leq \sqrt{\delta_{2k}} \|V^{(t)} - V^{*(t)}\|_F + \frac{2\sigma_k}{\xi} \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F, \quad (\text{A.6})$$

$$\|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F \leq \frac{2\sqrt{\delta_{2k}}}{\sigma_k} \|V^{(t)} - V^{*(t)}\|_F + \frac{4}{\xi} \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F, \quad (\text{A.7})$$

$$\|U^{(t)} - U^{*(t+1)}\|_F \leq \frac{3\sigma_1\sqrt{\delta_{2k}}}{\sigma_k} \|V^{(t)} - V^{*(t)}\|_F + \left(\frac{6}{\xi} + 1\right) \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F. \quad (\text{A.8})$$

The proof of Corollary A.5 is provided in Appendix C.5. Since the alternating gradient descent algorithm updates U and V in a symmetric manner, we can establish similar results for a complete iteration of updating U in the next corollary.

Corollary A.6. Suppose that δ_{2k} , $\bar{V}^{(t+1)}$, and $U^{(t)}$ satisfy

$$\delta_{2k} \leq \frac{\sigma_k^6}{192\xi^2 k \sigma_1^6}, \quad \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F \leq \frac{\sigma_k^2}{4\xi\sigma_1^2}, \quad \text{and} \quad \|U^{(t)} - U^{*(t+1)}\|_F \leq \frac{\sigma_k^2}{2\xi\sigma_1}. \quad (\text{A.9})$$

We then have

$$\|\bar{U}^{(t+1)} - \bar{U}^{*(t+1)}\|_F \leq \frac{\sigma_k^2}{4\xi\sigma_1^2} \quad \text{and} \quad \|V^{(t+1)} - V^{*(t+1)}\|_F \leq \frac{\sigma_k^2}{2\xi\sigma_1}.$$

Moreover, we have

$$\|U^{(t+0.5)} - U^{*(t+1)}\|_F \leq \sqrt{\delta_{2k}} \|U^{(t)} - U^{*(t+1)}\|_F + \frac{2\sigma_k}{\xi} \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F, \quad (\text{A.10})$$

$$\|\bar{U}^{(t+1)} - \bar{U}^{*(t+1)}\|_F \leq \frac{2\sqrt{\delta_{2k}}}{\sigma_k} \|U^{(t)} - U^{*(t+1)}\|_F + \frac{4}{\xi} \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F, \quad (\text{A.11})$$

$$\|V^{(t+1)} - V^{*(t+1)}\|_F \leq \frac{3\sigma_1\sqrt{\delta_{2k}}}{\sigma_k} \|U^{(t)} - U^{*(t+1)}\|_F + \left(\frac{6}{\xi} + 1\right) \sigma_1 \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F. \quad (\text{A.12})$$

The proof of Corollary A.6 directly follows Appendix C.5, and is therefore omitted..

Now we proceed with the proof of Theorem 3.4 for alternating gradient descent. Recall that Lemma A.4 ensures that (A.5) of Corollary A.5 holds for $\bar{U}^{(0)}$ and $V^{(0)}$. Then Corollary A.5 ensures that (A.9) of Corollary A.6 holds for $U^{(0)}$ and $\bar{V}^{(1)}$. By induction, Corollaries 4.7 and 4.8 can be applied recursively for all T iterations. For notational simplicity, we write (A.6)-(A.12) as

$$\|V^{(t+0.5)} - V^{*(t)}\|_F \leq \alpha_1 \|V^{(t)} - V^{*(t)}\|_F + \gamma_1 \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F, \quad (\text{A.13})$$

$$\sigma_1 \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F \leq \alpha_2 \|V^{(t)} - V^{*(t)}\|_F + \gamma_2 \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F, \quad (\text{A.14})$$

$$\|U^{(t+0.5)} - U^{*(t+1)}\|_F \leq \alpha_3 \|U^{(t)} - U^{*(t+1)}\|_F + \gamma_3 \sigma_1 \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F, \quad (\text{A.15})$$

$$\sigma_1 \|\bar{U}^{(t+1)} - \bar{U}^{*(t+1)}\|_F \leq \alpha_4 \|U^{(t)} - U^{*(t+1)}\|_F + \gamma_4 \sigma_1 \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F, \quad (\text{A.16})$$

$$\|U^{(t)} - U^{*(t+1)}\|_F \leq \alpha_5 \|V^{(t)} - V^{*(t)}\|_F + \gamma_5 \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F, \quad (\text{A.17})$$

$$\|V^{(t+1)} - V^{*(t+1)}\|_F \leq \alpha_6 \|U^{(t)} - U^{*(t+1)}\|_F + \gamma_6 \sigma_1 \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F. \quad (\text{A.18})$$

Note that we have $\gamma_5, \gamma_6 \in (1, 2)$, but $\alpha_1, \dots, \alpha_6, \gamma_1, \dots$, and γ_4 can be sufficiently small as long as ξ is sufficiently large. We then have

$$\begin{aligned}
\|U^{(t+1)} - U^{*(t+2)}\|_F &\stackrel{(i)}{\leq} \alpha_5 \|V^{(t+1)} - V^{*(t+1)}\|_F + \gamma_5 \sigma_1 \|\bar{U}^{(t+1)} - \bar{U}^{*(t+1)}\|_F \\
&\stackrel{(ii)}{\leq} \alpha_5 \alpha_6 \|U^{(t)} - U^{*(t+1)}\|_F + \alpha_5 \gamma_6 \sigma_1 \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F + \gamma_5 \sigma_1 \|\bar{U}^{(t+1)} - \bar{U}^{*(t+1)}\|_F \\
&\stackrel{(iii)}{\leq} (\alpha_5 \alpha_6 + \gamma_5 \alpha_4) \|U^{(t)} - U^{*(t+1)}\|_F + (\gamma_5 \gamma_4 \sigma_1 + \alpha_5 \gamma_6) \sigma_1 \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F \\
&\stackrel{(iv)}{\leq} (\alpha_5 \alpha_6 + \gamma_5 \alpha_4) \|U^{(t)} - U^{*(t+1)}\|_F + (\gamma_5 \gamma_4 \sigma_1 + \alpha_5 \gamma_6) \alpha_2 \|V^{(t)} - V^{*(t)}\|_F \\
&\quad + (\gamma_5 \gamma_4 \sigma_1 + \alpha_5 \gamma_6) \gamma_2 \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F,
\end{aligned} \tag{A.19}$$

where (i) comes from (A.17), (ii) comes from (A.18), (iii) comes from (A.16), and (iv) comes from (A.14). Similarly, we can obtain

$$\begin{aligned}
\|V^{(t+1)} - V^{*(t+1)}\|_F &\leq \alpha_6 \|U^{(t)} - U^{*(t+1)}\|_F + \gamma_6 \alpha_2 \|V^{(t)} - V^{*(t)}\|_F \\
&\quad + \gamma_6 \gamma_2 \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F,
\end{aligned} \tag{A.20}$$

$$\begin{aligned}
\sigma_1 \|\bar{U}^{(t+1)} - \bar{U}^{*(t+1)}\|_F &\leq \alpha_4 \|U^{(t)} - U^{*(t+1)}\|_F + \gamma_4 \alpha_2 \|V^{(t)} - V^{*(t)}\|_F \\
&\quad + \gamma_4 \gamma_2 \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F
\end{aligned} \tag{A.21}$$

$$\begin{aligned}
\|U^{(t+0.5)} - U^{*(t+1)}\|_F &\leq \alpha_3 \|U^{(t)} - U^{*(t+1)}\|_F + \gamma_3 \alpha_2 \|V^{(t)} - V^{*(t)}\|_F \\
&\quad + \gamma_3 \gamma_2 \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F.
\end{aligned} \tag{A.22}$$

For simplicity, we define

$$\begin{aligned}
\phi_{V^{(t+1)}} &= \|V^{(t+1)} - V^{*(t+1)}\|_F, \quad \phi_{V^{(t+0.5)}} = \|V^{(t+0.5)} - V^{*(t)}\|_F, \quad \phi_{\bar{V}^{(t+1)}} = \sigma_1 \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F, \\
\phi_{U^{(t+1)}} &= \|U^{(t+1)} - U^{*(t+2)}\|_F, \quad \phi_{U^{(t+0.5)}} = \|U^{(t+0.5)} - U^{*(t+1)}\|_F, \quad \phi_{\bar{U}^{(t+1)}} = \sigma_1 \|\bar{U}^{(t+1)} - \bar{U}^{*(t+1)}\|_F.
\end{aligned}$$

Then combining (A.13), (A.14) with (A.19)–(A.22), we obtain

$$\max \{ \phi_{V^{(t+1)}}, \phi_{V^{(t+0.5)}}, \phi_{\bar{V}^{(t+1)}}, \phi_{U^{(t+1)}}, \phi_{U^{(t+0.5)}}, \phi_{\bar{U}^{(t+1)}} \} \leq \beta \max \{ \phi_{V^{(t)}}, \phi_{U^{(t)}}, \phi_{\bar{U}^{(t)}} \}, \tag{A.23}$$

where β is a contraction coefficient defined as

$$\begin{aligned}
\beta &= \max \{ \alpha_5 \alpha_6 + \gamma_5 \alpha_4, \alpha_6, \alpha_4, \alpha_3 \} + \max \{ \alpha_1, \alpha_2, (\gamma_5 \gamma_4 \sigma_1 + \alpha_5 \gamma_6), \gamma_6 \alpha_2, \gamma_4 \alpha_2, \gamma_3 \alpha_2 \} \\
&\quad + \max \{ \gamma_1, \gamma_2, (\gamma_5 \gamma_4 \sigma_1 + \alpha_5 \gamma_6) \gamma_2, \gamma_6 \gamma_2, \gamma_4 \gamma_2, \gamma_3 \gamma_2 \}.
\end{aligned}$$

Then we can choose ξ as a sufficiently large constant such that $\beta < 1$. By recursively applying (A.23) for $t = 0, \dots, T$, we obtain

$$\begin{aligned}
\max \{ \phi_{V^{(T)}}, \phi_{V^{(T-0.5)}}, \phi_{\bar{V}^{(T)}}, \phi_{U^{(T)}}, \phi_{U^{(T-0.5)}}, \phi_{\bar{U}^{(T)}} \} &\leq \beta \max \{ \phi_{V^{(T-1)}}, \phi_{U^{(T-1)}}, \phi_{\bar{U}^{(T-1)}} \} \\
&\leq \beta^2 \max \{ \phi_{V^{(T-2)}}, \phi_{U^{(T-2)}}, \phi_{\bar{U}^{(T-2)}} \} \leq \dots \leq \beta^T \max \{ \phi_{V^{(0)}}, \phi_{U^{(0)}}, \phi_{\bar{U}^{(0)}} \}.
\end{aligned}$$

By Corollary A.5, we obtain

$$\begin{aligned}
\|U^{(0)} - U^{*(1)}\|_F &\leq \frac{3\sigma_1 \sqrt{\delta_{2k}}}{\sigma_k} \|V^{(0)} - V^{*(0)}\|_F + \left(\frac{6}{\xi} + 1 \right) \sigma_1 \|\bar{U}^{(0)} - \bar{U}^{*(0)}\|_F \\
&\stackrel{(i)}{\leq} \frac{3\sigma_1}{\sigma_k} \cdot \frac{\sigma_k^3}{12\xi\sigma_1^3} \cdot \frac{\sigma_k^2}{2\xi\sigma_1} + \left(\frac{6}{\xi} + 1 \right) \frac{\sigma_k^2}{4\xi\sigma_1} \\
&\stackrel{(ii)}{=} \frac{\sigma_k^4}{8\xi^2\sigma_1^3} + \frac{3\sigma_k^2}{2\xi^2\sigma_1} + \frac{\sigma_k^2}{4\xi\sigma_1} \stackrel{(iii)}{\leq} \frac{\sigma_k^2}{2\xi\sigma_1},
\end{aligned} \tag{A.24}$$

where (i) and (ii) come from Lemma A.4, and (iii) comes from the definition of ξ and $\sigma_1 \geq \sigma_k$. Combining (A.24) with Lemma A.4, we have

$$\{ \phi_{V^{(0)}}, \phi_{U^{(0)}}, \phi_{\bar{U}^{(0)}} \} \leq \max \left\{ \frac{\sigma_k^2}{2\xi\sigma_1}, \frac{\sigma_k^2}{4\xi\sigma_1^2} \right\}.$$

Then we need at most

$$T = \left\lceil \log \left(\max \left\{ \frac{\sigma_k^2}{\xi\sigma_1}, \frac{\sigma_k^2}{2\xi\sigma_1^2}, \frac{\sigma_k^2}{\xi}, \frac{\sigma_k^2}{2\xi\sigma_1} \right\} \cdot \frac{1}{\epsilon} \right) \log^{-1}(\beta^{-1}) \right\rceil$$

iterations such that

$$\begin{aligned}\|\bar{V}^{(T)} - \bar{V}^*\|_F &\leq \beta^T \max\left\{\frac{\sigma_k^2}{2\xi\sigma_1}, \frac{\sigma_k^2}{4\xi\sigma_1^2}\right\} \leq \frac{\epsilon}{2}, \\ \|U^{(T)} - U^*\|_F &\leq \beta^T \max\left\{\frac{\sigma_k^2}{2\xi\sigma_1}, \frac{\sigma_k^2}{4\xi\sigma_1^2}\right\} \leq \frac{\epsilon}{2\sigma_1}.\end{aligned}$$

We then follow similar lines to (4.9) in Appendix 4.1, and show $\|M^{(T)} - M^*\|_F \leq \epsilon$, which completes the proof. \square

A.2 Proof of Theorem 3.4 (Gradient Descent)

Proof. The convergence analysis of the gradient descent algorithm is similar to that of the alternating gradient descent. The only difference is that for updating U , the gradient descent algorithm employs $V = \bar{V}^{(t)}$ instead of $V = \bar{V}^{(t+1)}$ to calculate the gradient at $U = U^{(t)}$. Then everything else directly follows Appendix A.1, and is therefore omitted. \square

B Lemmas for Theorem 3.4 (Alternating Exact Minimization)

B.1 Proof of Lemma 4.1

Proof. For notational convenience, we omit the index t in $\bar{U}^{*(t)}$ and $V^{*(t)}$, and denote them by \bar{U}^* and V^* respectively. Then we define

$$\begin{aligned}S^{(t)} &= \begin{bmatrix} S_{11}^{(t)} & \cdots & S_{1k}^{(t)} \\ \vdots & \ddots & \vdots \\ S_{k1}^{(t)} & \cdots & S_{kk}^{(t)} \end{bmatrix} \quad \text{with} \quad S_{pq}^{(t)} = \sum_{i=1}^d A_i \bar{U}_{*p}^{(t)} \bar{U}_{*q}^{(t)\top} A_i^\top, \\ G^{(t)} &= \begin{bmatrix} G_{11}^{(t)} & \cdots & G_{1k}^{(t)} \\ \vdots & \ddots & \vdots \\ G_{k1}^{(t)} & \cdots & G_{kk}^{(t)} \end{bmatrix} \quad \text{with} \quad G_{pq}^{(t)} = \sum_{i=1}^d A_i \bar{U}_{*p}^* \bar{U}_{*q}^{*\top} A_i^\top\end{aligned}$$

for $1 \leq p, q \leq k$. Note that $S^{(t)}$ and $G^{(t)}$ are essentially the partial Hessian matrices $\nabla_V^2 \mathcal{F}(\bar{U}^{(t)}, V)$ and $\nabla_V^2 \mathcal{F}(\bar{U}^*, V)$ for a vectorized V , i.e., $\text{vec}(V)$. Before we proceed with the main proof, we first introduce the following lemma.

Lemma B.1. Suppose that $\mathcal{A}(\cdot)$ satisfies $2k$ -RIP with parameter δ_{2k} . We then have

$$1 + \delta_{2k} \geq \sigma_{\max}(S^{(t)}) \geq \sigma_{\min}(S^{(t)}) \geq 1 - \delta_{2k}.$$

The proof of Lemma B.1 is provided in Appendix B.7. Note that Lemma B.1 is also applicable $G^{(t)}$, since $G^{(t)}$ shares the same structure with $S^{(t)}$. Given a fixed \bar{U} , $\mathcal{F}(\bar{U}, V)$ is a quadratic function of V . Therefore we have

$$\begin{aligned}\mathcal{F}(\bar{U}, V') &= \mathcal{F}(\bar{U}, V) + \langle \nabla_V \mathcal{F}(\bar{U}, V), V' - V \rangle \\ &\quad + \langle \text{vec}(V') - \text{vec}(V), \nabla_V^2 \mathcal{F}(\bar{U}, V) (\text{vec}(V') - \text{vec}(V)) \rangle,\end{aligned}$$

which further implies

$$\begin{aligned}\mathcal{F}(\bar{U}, V') - \mathcal{F}(\bar{U}, V) - \langle \nabla_V \mathcal{F}(\bar{U}, V), V' - V \rangle &\leq \sigma_{\max}(\nabla_V^2 \mathcal{F}(\bar{U}, V)) \|V' - V\|_F^2 \\ \mathcal{F}(\bar{U}, V') - \mathcal{F}(\bar{U}, V) - \langle \nabla_V \mathcal{F}(\bar{U}, V), V' - V \rangle &\geq \sigma_{\min}(\nabla_V^2 \mathcal{F}(\bar{U}, V)) \|V' - V\|_F^2.\end{aligned}$$

Then we can verify that $\nabla_V^2 \mathcal{F}(U, V)$ also shares the same structure with $S^{(t)}$. Thus applying Lemma B.1 to the above two inequalities, we complete the proof. \square

B.2 Proof of Lemma 4.3

Proof. For notational convenience, we omit the index t in $\bar{U}^{*(t)}$ and $V^{*(t)}$, and denote them by \bar{U}^* and V^* respectively. We define

$$J^{(t)} = \begin{bmatrix} J_{11}^{(t)} & \cdots & J_{1k}^{(t)} \\ \vdots & \ddots & \vdots \\ J_{k1}^{(t)} & \cdots & J_{kk}^{(t)} \end{bmatrix} \quad \text{with} \quad J_{pq}^{(t)} = \sum_{i=1}^d A_i \bar{U}_{*p}^{(t)} \bar{U}_{*q}^{*\top} A_i^\top,$$

$$K^{(t)} = \begin{bmatrix} K_{11}^{(t)} & \cdots & K_{1k}^{(t)} \\ \vdots & \ddots & \vdots \\ K_{k1}^{(t)} & \cdots & K_{kk}^{(t)} \end{bmatrix} \quad \text{with} \quad K_{pq}^{(t)} = \bar{U}_{*p}^{(t)\top} \bar{U}_{*q}^* I_n$$

for $1 \leq p, q \leq k$. Before we proceed with the main proof, we first introduce the following lemmas.

Lemma B.2. Suppose that $\mathcal{A}(\cdot)$ satisfies $2k$ -RIP with parameter δ_{2k} . We then have

$$\|S^{(t)} K^{(t)} - J^{(t)}\|_2 \leq \delta_{2k} \sqrt{2k} \|\bar{U}^{(t)} - \bar{U}^*\|_F.$$

The proof of Lemma B.2 is provided in Appendix B.8. Note that Lemma B.2 is also applicable to $G^{(t)} K^{(t)} - J^{(t)}$, since $G^{(t)}$ and $S^{(t)}$ share the same structure.

Lemma B.3. Given $F \in \mathbb{R}^{k \times k}$, we define

$$\mathbb{F} = \begin{bmatrix} F_{11} I_n & \cdots & F_{1k} I_n \\ \vdots & \ddots & \vdots \\ F_{k1} I_n & \cdots & F_{kk} I_n \end{bmatrix}.$$

For any $V \in \mathbb{R}^{n \times k}$, let $v = \text{vec}(V)$, then we have $\|\mathbb{F}v\|_2 = \|FV^\top\|_F$.

Proof. By linear algebra, we have

$$[FV]_{ij} = F_{i*}^\top V_{j*} = \sum_{\ell=1}^k F_{i\ell} V_{j\ell} = \sum_{\ell=1}^k F_{i\ell} I_{*\ell}^\top V_{*j},$$

which completes the proof. \square

We then proceed with the main proof. Since $b_i = \text{tr}(V^{*\top} A_i U^*)$, then we rewrite $\mathcal{F}(\bar{U}, V)$ as

$$\mathcal{F}(\bar{U}, V) = 1/2 \cdot \sum_{i=1}^d (\text{tr}(V^\top A_i \bar{U}) - b_i)^2 = 1/2 \cdot \sum_{i=1}^d \left(\sum_{j=1}^k V_{j*}^\top A_i \bar{U}_{*j} - \sum_{j=1}^k V_{j*}^{*\top} A_i \bar{U}_{*j}^* \right)^2.$$

For notational simplicity, we define $v = \text{vec}(V)$. Since $V^{(t+0.5)}$ minimizes $\mathcal{F}(\bar{U}^{(t)}, V)$, we have

$$\text{vec}(\nabla_U \mathcal{F}(\bar{U}^{(t)}, V^{(t+0.5)})) = S^{(t)} v^{(t+0.5)} - J^{(t)} v^* = 0.$$

Solving the above system of equations, we obtain

$$v^{(t+0.5)} = (S^{(t)})^{-1} J^{(t)} v^*. \quad (\text{B.1})$$

Meanwhile, we have

$$\begin{aligned} \text{vec}(\nabla_V \mathcal{F}(\bar{U}^*, V^{(t+0.5)})) &= G^{(t)} v^{(t+0.5)} - G^{(t)} v^* \\ &= G^{(t)} (S^{(t)})^{-1} J^{(t)} v^* - G^{(t)} v^* = G^{(t)} ((S^{(t)})^{-1} J^{(t)} - I_{nk}) v^*, \end{aligned} \quad (\text{B.2})$$

where the second equality come from (B.1). By triangle inequality, (B.2) further implies

$$\begin{aligned} \|((S^{(t)})^{-1} J^{(t)} - I_{nk}) v^*\|_2 &\leq \|(K^{(t)} - I_{nk}) v^*\|_2 + \|(S^{(t)})^{-1} (J^{(t)} - S^{(t)} K^{(t)}) v^*\|_2 \\ &\leq \|(\bar{U}^{(t)\top} \bar{U}^* - I_k) V^*\|_F + \|(S^{(t)})^{-1}\|_2 \|(J^{(t)} - S^{(t)} K^{(t)}) v^*\|_2 \\ &\leq \|\bar{U}^{(t)\top} \bar{U}^* - I_k\|_F \|V^*\|_2 + \|(S^{(t)})^{-1}\|_2 \|(J^{(t)} - S^{(t)} K^{(t)}) v^*\|_2, \end{aligned} \quad (\text{B.3})$$

where the second inequality comes from Lemma B.3. Plugging (B.3) into (B.2), we have

$$\begin{aligned}
& \|\text{vec}(\nabla_V \mathcal{F}(\bar{U}^*, V^{(t+0.5)}))\|_2 \leq \|G^{(t)}\|_2 \|((S^{(t)})^{-1} J^{(t)} - I_{nk}) v^*\|_2 \\
& \stackrel{(i)}{\leq} (1 + \delta_{2k}) (\sigma_1 \|\bar{U}^{(t)\top} \bar{U}^* - I_k\|_2 + \|(S^{(t)})^{-1}\|_2 \|S^{(t)} K^{(t)} - J^{(t)}\|_2 \sigma_1 \sqrt{k}) \\
& \stackrel{(ii)}{\leq} (1 + \delta_{2k}) \sigma_1 \left(\|(\bar{U}^{(t)} - \bar{U}^*)^\top (\bar{U}^{(t)} - \bar{U}^*)\|_F + \frac{\sqrt{2} \delta_{2k} k}{1 - \delta_{2k}} \|\bar{U}^{(t)} - \bar{U}^*\|_F \right) \\
& \stackrel{(iii)}{\leq} (1 + \delta_{2k}) \sigma_1 \left(\|\bar{U}^{(t)} - \bar{U}^*\|_F^2 + \frac{\sqrt{2} \delta_{2k} k}{1 - \delta_{2k}} \|\bar{U}^{(t)} - \bar{U}^*\|_F \right) \stackrel{(iv)}{\leq} \frac{(1 - \delta_{2k}) \sigma_k}{2\xi} \|\bar{U}^* - \bar{U}^{(t)}\|_F,
\end{aligned}$$

where (i) comes from Lemma B.1 and $\|V^*\|_2 = \|M^*\| = \sigma_1$ and $\|V^*\|_F = \|v^*\|_2 \leq \sigma_1 \sqrt{k}$, (ii) comes from Lemmas B.1 and B.2, (iii) from Cauchy-Schwartz inequality, and (iv) comes from (4.1). By Cauchy-Schwartz inequality again, we obtain

$$\mathcal{D}(V^{(t+0.5)}, V^{(t+0.5)}, \bar{U}^{(t)}) \leq \|\nabla_V \mathcal{F}(\bar{U}^*, V^{(t+0.5)})\|_F \leq \frac{(1 - \delta_{2k}) \sigma_k}{2\xi} \|\bar{U}^* - \bar{U}^{(t)}\|_F,$$

which completes the proof. \square

B.3 Proof of Lemma 4.4

Proof. For notational convenience, we omit the index t in $\bar{U}^{*(t)}$ and $V^{*(t)}$, and denote them by \bar{U}^* and V^* respectively. By the strong convexity of $\mathcal{F}(\bar{U}^*, \cdot)$, we have

$$\begin{aligned}
& \mathcal{F}(\bar{U}^*, V^*) - \frac{1 - \delta_{2k}}{2} \|V^{(t+0.5)} - V^*\|_F^2 \\
& \geq \mathcal{F}(\bar{U}^*, V^{(t+0.5)}) + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t+0.5)}), V^* - V^{(t+0.5)} \rangle.
\end{aligned} \tag{B.4}$$

By the strong convexity of $\mathcal{F}(\bar{U}^*, \cdot)$ again, we have

$$\begin{aligned}
& \mathcal{F}(\bar{U}^*, V^{(t+0.5)}) \geq \mathcal{F}(\bar{U}^*, V^*) + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^*), V^{(t+0.5)} - V^{(t+0.5)} \rangle + \frac{1 - \delta_{2k}}{2} \|V^{(t+0.5)} - V^*\|_F^2 \\
& \geq \mathcal{F}(\bar{U}^*, V^*) + \frac{1 - \delta_{2k}}{2} \|V^{(t+0.5)} - V^*\|_F^2,
\end{aligned} \tag{B.5}$$

where the last inequality comes from the optimality condition of $V^* = \arg\min_V \mathcal{F}(\bar{U}^*, V)$, i.e.

$$\langle \nabla_V \mathcal{F}(\bar{U}^*, V^*), V^{(t+0.5)} - V^* \rangle \geq 0.$$

Meanwhile, since $V^{(t+0.5)}$ minimizes $\mathcal{F}(\bar{U}^{(t)}, \cdot)$, we have the optimality condition

$$\langle \nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t+0.5)}), V^* - V^{(t+0.5)} \rangle \geq 0,$$

which further implies

$$\begin{aligned}
& \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t+0.5)}), V^* - V^{(t+0.5)} \rangle \\
& \geq \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t+0.5)}) - \nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t+0.5)}), V^* - V^{(t+0.5)} \rangle.
\end{aligned} \tag{B.6}$$

Combining (B.4) and (B.5) with (B.6), we obtain

$$\|V^{(t+0.5)} - V^*\|_2 \leq \frac{1}{1 - \delta_{2k}} \mathcal{D}(V^{(t+0.5)}, V^{(t+0.5)}, \bar{U}^{(t)}),$$

which completes the proof. \square

B.4 Proof of Lemma 4.5

Proof. Before we proceed with the proof, we first introduce the following Lemma

Lemma B.4. Suppose that $A^* \in \mathbb{R}^{n \times k}$ is a rank k matrix. Let $E \in \mathbb{R}^{n \times k}$ satisfy $\|E\|_2 \|A^{*\dagger}\|_2 < 1$. Then given a QR decomposition $(A^* + E) = QR$, there exists a factorization of $A^* = Q^* O^*$ such that $Q^* \in \mathbb{R}^{n \times k}$ is an orthonormal matrix, and satisfies

$$\|Q - Q^*\|_F \leq \frac{\sqrt{2} \|A^{*\dagger}\|_2 \|E\|_F}{1 - \|E\|_2 \|A^{*\dagger}\|_2}.$$

The proof of Lemma B.4 is provided in [26], therefore omitted.

We then proceed with the main proof. We consider $A^* = V^{*(t)}$ and $E = V^{(t+0.5)} - V^{*(t)}$ in Lemma B.4 respectively. We can verify that

$$\|V^{(t+0.5)} - V^{*(t)}\|_2 \|V^{*(t)\dagger}\|_2 \leq \frac{\|V^{(t+0.5)} - V^{*(t)}\|_F}{\sigma_k} \leq \frac{1}{4}.$$

Then there exists a $V^{*(t)} = \bar{V}^{*(t+1)} O^*$ such that $\bar{V}^{*(t+1)}$ is an orthonormal matrix, and satisfies

$$\|\bar{V}^{*(t+0.5)} - \bar{V}^{*(t+1)}\|_F \leq 2\|V^{*(t)\dagger}\|_2 \|V^{(t+0.5)} - V^{*(t)}\|_F \leq \frac{2}{\sigma_k} \|V^{(t+0.5)} - V^{*(t)}\|_F.$$

Thus we conclude the proof. \square

B.5 Proof of Lemma 4.6

Proof. We first introduce the following lemma.

Lemma B.5. Let $b = \mathcal{A}(M^*) + \varepsilon$, M is a rank- k matrix, and \mathcal{A} is a linear measurement operator that satisfies $2k$ -RIP with constant $\delta_{2k} < 1/3$. Let $X^{(t+1)}$ be the $(t+1)$ -th step iterate of SVP, then we have

$$\|\mathcal{A}(X^{(t+1)}) - b\|_2^2 \leq \|\mathcal{A}(M^*) - b\|_2^2 + 2\delta_{2k} \|\mathcal{A}(X^{(t)}) - b\|_2^2$$

The proof of Lemma B.5 is provided in [12], therefore omitted. We then explain the implication of Lemma B.5. [12] show that $X^{(t+1)}$ is obtained by taking a projected gradient iteration over $X^{(t)}$ using step size $\frac{1}{1+\delta_{2k}}$. Then taking $X^{(t)} = 0$, we have

$$X^{(t+1)} = \frac{\bar{U}^{(0)} \bar{\Sigma}^{(0)} \bar{V}^{(0)\top}}{1 + \delta_{2k}}.$$

Suppose that M^* has a compact singular value decomposition $M^* = \tilde{U}^* \tilde{D}^* \tilde{V}^{*\top}$. Then Lemma B.5 implies

$$\left\| \mathcal{A} \left(\frac{\bar{U}^{(0)} \bar{\Sigma}^{(0)} \bar{V}^{(0)\top}}{1 + \delta_{2k}} - \tilde{U}^* \tilde{D}^* \tilde{V}^{*\top} \right) \right\|_2^2 \leq 4\delta_{2k} \|\mathcal{A}(\tilde{U}^* \tilde{D}^* \tilde{V}^{*\top})\|_2^2. \quad (\text{B.7})$$

Since $\mathcal{A}(\cdot)$ satisfies $2k$ -RIP, (B.7) further implies

$$\left\| \frac{\bar{U}^{(0)} \bar{\Sigma}^{(0)} \bar{V}^{(0)\top}}{1 + \delta_{2k}} - \tilde{U}^* \tilde{D}^* \tilde{V}^{*\top} \right\|_F^2 \leq 4\delta_{2k} (1 + 3\delta_{2k}) \|\tilde{D}^*\|_F^2. \quad (\text{B.8})$$

We then project each column of $\tilde{U}^* \tilde{D}^* \tilde{V}^{*\top}$ into the subspace spanned by $\{\bar{U}_{*i}^{(0)}\}_{i=1}^k$, and obtain

$$\|\bar{U}^{(0)} \bar{U}^{(0)\top} \tilde{U}^* \tilde{D}^* \tilde{V}^{*\top} - \tilde{U}^* \tilde{D}^* \tilde{V}^{*\top}\|_F^2 \leq 6\delta_{2k} \|\tilde{D}^*\|_F^2.$$

Let $\bar{U}_\perp^{(0)}$ denote the orthonormal complement of $\bar{U}^{(0)}$, i.e.,

$$\bar{U}_\perp^{(0)\top} \bar{U}_\perp^{(0)} = I_{n-k} \quad \text{and} \quad \bar{U}_\perp^{(0)\top} \bar{U}^{(0)} = 0.$$

Then we have

$$\frac{6\delta_{2k} k \sigma_1^2}{\sigma_k^2} \geq \|(\bar{U}^{(0)} \bar{U}^{(0)\top} - I_n) \tilde{U}^*\|_F^2 = \|\bar{U}_\perp^{(0)\top} \tilde{U}^*\|_F^2.$$

Thus there exists a unitary matrix $O \in \mathbb{R}^{k \times k}$ such that $OO^\top = I_k$ and

$$\|\bar{U}^{(0)} - \tilde{U}^* O\|_F \leq \sqrt{2} \|\bar{U}_\perp^{(0)\top} \tilde{U}^*\|_F \leq 2\sqrt{3\delta_{2k} k} \cdot \frac{\sigma_1}{\sigma_k}.$$

We define $\bar{U}^{*(0)} = \tilde{U}^* O$. Then combining the above inequality with (4.3), we have

$$\|\bar{U}^{(0)} - \bar{U}^{*(0)}\|_F \leq \frac{(1 - \delta_{2k})\sigma_k}{4\xi(1 + \delta_{2k})\sigma_1}.$$

Moreover, we define $V^{*(0)} = \tilde{V}^* \tilde{D}^* O$. Then we have $\bar{U}^{*(0)} V^{*(0)\top} = \tilde{U}^* O O^\top \tilde{D}^* \tilde{V}^* = M^*$. \square

B.6 Proof of Corollary 4.7

Proof. Since (4.4) ensures that (4.1) of Lemma 4.3 holds, then we have

$$\begin{aligned}
\|V^{(t+0.5)} - V^{*(t)}\|_F &\leq \frac{1}{1 - \delta_{2k}} \mathcal{D}(V^{(t+0.5)}, V^{(t+0.5)}, \bar{U}^{(t)}) \\
&\stackrel{(i)}{\leq} \frac{1}{1 - \delta_{2k}} \cdot \frac{(1 - \delta_{2k})\sigma_k}{2\xi} \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \\
&\stackrel{(ii)}{\leq} \frac{1}{1 - \delta_{2k}} \cdot \frac{(1 - \delta_{2k})\sigma_k}{2\xi} \cdot \frac{(1 - \delta_{2k})\sigma_k}{4\xi(1 + \delta_{2k})\sigma_1} \\
&\leq \left(\frac{(1 - \delta_{2k})\sigma_k}{8\xi^2(1 + \delta_{2k})\sigma_1} \right) \sigma_k \stackrel{(iii)}{\leq} \frac{\sigma_k}{4}, \tag{B.9}
\end{aligned}$$

where (i) comes from Lemma 4.4, (ii) comes from (4.4), and (iii) comes from the definition of ξ and $\sigma_k \leq \sigma_1$. Since (B.9) ensures that (4.2) of Lemma 4.5 holds for $V^{(t+0.5)}$, we obtain

$$\|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F \leq \frac{2}{\sigma_k} \|V^{(t+0.5)} - V^{*(t)}\|_F \stackrel{(i)}{\leq} \frac{1}{\xi} \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \stackrel{(ii)}{\leq} \frac{(1 - \delta_{2k})\sigma_k}{4\xi(1 + \delta_{2k})\sigma_1}, \tag{B.10}$$

where (i) comes from (B.9), and (ii) comes from the definition of ξ and (4.4). \square

B.7 Proof of Lemma B.1

Proof. We consider an arbitrary $W \in \mathbb{R}^{n \times k}$ such that $\|W\|_F = 1$. Let $w = \text{vec}(W)$. Then have

$$\begin{aligned}
w^\top B w &= \sum_{p,q=1}^k W_{*p}^\top S_{pq}^{(t)} W_{*p} = \sum_{p,q=1}^k W_{*p}^\top \left(\sum_{i=1}^d A_i \bar{U}_{*p}^{(t)} \bar{U}_{*q}^{(t)\top} A_i^\top \right) W_{*q} \\
&= \sum_{i=1}^d \left(\sum_{p=1}^k W_{*p}^\top A_i \bar{U}_{*p}^{(t)} \right) \left(\sum_{q=1}^k W_{*q}^\top A_i \bar{U}_{*q}^{(t)} \right) = \sum_{i=1}^d \text{tr}(W^\top A_i \bar{U}^{(t)})^2.
\end{aligned}$$

Since $\mathcal{A}(\cdot)$ satisfies $2k$ -RIP, then we have

$$\begin{aligned}
\sum_{i=1}^d \text{tr}(W^\top A_i \bar{U}^{(t)})^2 &\geq (1 - \delta_{2k}) \|\bar{U}^{(t)} W^\top\|_F = (1 - \delta_{2k}) \|W\|_F = 1 - \delta_{2k}, \\
\sum_{i=1}^d \text{tr}(W^\top A_i \bar{U}^{(t)})^2 &\leq (1 + \delta_{2k}) \|\bar{U}^{(t)} W^\top\|_F = (1 + \delta_{2k}) \|W\|_F = 1 + \delta_{2k}.
\end{aligned}$$

Since W is arbitrary, then we have

$$\sigma_{\min}(S^{(t)}) = \min_{\|w\|_2=1} w^\top S^{(t)} w \geq 1 - \delta_{2k} \quad \text{and} \quad \sigma_{\max}(S^{(t)}) = \max_{\|w\|_2=1} w^\top S^{(t)} w \leq 1 + \delta_{2k}.$$

Thus we conclude the proof. \square

B.8 Proof of Lemma B.2

Proof. For notational convenience, we omit the index t in $\bar{U}^{*(t)}$ and $V^{*(t)}$, and denote them by \bar{U}^* and V^* respectively. We first introduce the following lemma.

Lemma B.6. Suppose $\mathcal{A}(\cdot)$ satisfies $2k$ -RIP. For any $U, U' \in \mathbb{R}^{m \times k}$ and $V, V' \in \mathbb{R}^{n \times k}$, we have

$$|\langle \mathcal{A}(UV^\top), \mathcal{A}(U'V'^\top) \rangle - \langle U^\top U', V^\top V' \rangle| \leq 3\delta_{2k} \|UV^\top\|_F \cdot \|U'V'^\top\|_F.$$

The proof of Lemma B.6 is provided in [14], and hence omitted.

We now proceed with the proof. We consider arbitrary $W, Z \in \mathbb{R}^{n \times k}$ such that $\|W\|_F = \|Z\|_F = 1$. Let $w = \text{vec}(W)$ and $z = \text{vec}(Z)$. Then have

$$w^\top (S^{(t)} K^{(t)} - J^{(t)}) z = \sum_{p,q=1}^k W_{*p}^\top [S^{(t)} K^{(t)} - J^{(t)}]_{pq} Z_{*q}.$$

We consider a decomposition

$$\begin{aligned}
[S^{(t)}K^{(t)} - J^{(t)}]_{pq} &= \sum_{\ell=1}^k S_{p\ell}^{(t)} K_{\ell q}^{(t)} - J_{pq}^{(t)} = \sum_{\ell=1}^k S_{p\ell}^{(t)} \bar{U}_{*\ell}^{(t)\top} \bar{U}_{*q}^* I_n - J_{pq}^{(t)} \\
&= \sum_{\ell=1}^k \bar{U}_{*q}^{*\top} \bar{U}_{*\ell}^{(t)} \sum_{i=1}^d A_i \bar{U}_{*p}^{(t)} \bar{U}_{*\ell}^{(t)} A_i^\top - J_{pq}^{(t)} = \sum_{\ell=1}^k A_i \bar{U}_{*q}^{*\top} \bar{U}_{*\ell}^{(t)} \sum_{i=1}^d \bar{U}_{*p}^{(t)} \bar{U}_{*\ell}^{(t)} A_i^\top - \sum_{i=1}^d A_i \bar{U}_{*p}^{(t)} \bar{U}_{*q}^* A_i^\top \\
&= \sum_{i=1}^d A_i \bar{U}_{*p}^{(t)} \bar{U}_{*q}^* (\bar{U}^{(t)} \bar{U}^{(t)\top} - I_n) A_i^\top.
\end{aligned}$$

which further implies

$$\begin{aligned}
w^\top (S^{(t)}K^{(t)} - J^{(t)})z &= \sum_{p,q} W_{*p}^\top \left(\sum_{i=1}^d A_i \bar{U}_{*p}^{(t)} \bar{U}_{*q}^* (\bar{U}^{(t)} \bar{U}^{(t)\top} - I_n) A_i^\top \right) Z_{*q} \\
&= \sum_{i=1}^d \sum_{p,q} W_{*p}^\top A_i \bar{U}_{*p}^{(t)} \bar{U}_{*q}^* (\bar{U}^{(t)} \bar{U}^{(t)\top} - I_n) A_i^\top Z_{*q} \\
&= \sum_{i=1}^d \text{tr}(W^\top A_i \bar{U}^{(t)}) \text{tr}(Z^\top A_i (\bar{U}^{(t)} \bar{U}^{(t)\top} - I_n) \bar{U}^*).
\end{aligned}$$

Therefore by $2k$ -RIP of $\mathcal{A}(\cdot)$ and Lemma B.6, we obtain

$$\begin{aligned}
w^\top (S^{(t)}K^{(t)} - J^{(t)})z &\leq \text{tr}(\bar{U}^* (\bar{U}^{(t)} \bar{U}^{(t)\top} - I_n) \bar{U}^{(t)} W^\top Z) + \delta_{2k} \|\bar{U}^{(t)} W^\top\|_F \|(\bar{U}^{(t)} \bar{U}^{(t)\top} - I_n) \bar{U}^* Z^\top\|_F \\
&\leq \delta_{2k} \|W\|_F \sqrt{\|\bar{U}^{*\top} (\bar{U}^{(t)} \bar{U}^{(t)\top} - I_n) \bar{U}^*\|_F \|Z^\top Z\|_F} \leq \delta_{2k} \sqrt{2k} \|\bar{U}^{(t)} - \bar{U}^*\|_F,
\end{aligned}$$

where the last inequality comes from $(\bar{U}^{(t)} \bar{U}^{(t)\top} - I_n) \bar{U}^{(t)} = 0$. Since W and Z are arbitrary, we have

$$\sigma_{\max}(S^{(t)}K^{(t)} - J^{(t)}) = \max_{\|w\|_2=1, \|z\|_2=1} w^\top (S^{(t)}K^{(t)} - J^{(t)})w \leq \delta_{2k} \sqrt{2k} \|\bar{U}^{(t)} - \bar{U}^*\|_F,$$

which completes the proof. \square

C Lemmas for Theorem 3.4 (Alternating Gradient Descent)

C.1 Proof of Lemma A.1

Proof. For notational convenience, we omit the index t in $\bar{U}^{*(t)}$ and $V^{*(t)}$, and denote them by \bar{U}^* and V^* respectively. We have

$$\text{vec}(\nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t)})) = S^{(t)}v^{(t)} - J^{(t)}v^* \quad \text{and} \quad \text{vec}(\nabla_V \mathcal{F}(\bar{U}^*, V^{(t)})) = G^{(t)}v^{(t)} - G^{(t)}v^*.$$

Therefore, we further obtain

$$\begin{aligned}
&\|\nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t)}) - \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)})\|_F \\
&= \|(S^{(t)} - J^{(t)})(v^{(t)} - v^*) + (S^{(t)} - J^{(t)})v^* + (J^{(t)} - G^{(t)})(v^{(t)} - v^*)\|_2 \\
&\leq \|(S^{(t)} - J^{(t)})(v^{(t)} - v^*)\|_2 + \|(S^{(t)} - J^{(t)})v^*\|_2 + \|(J^{(t)} - G^{(t)})(v^{(t)} - v^*)\|_2 \\
&\leq \|S^{(t)}\|_2 \|((S^{(t)})^{-1} J^{(t)} - I_{nk})(v^{(t)} - v^*)\|_2 \\
&\quad + \|S^{(t)}\|_2 \|((S^{(t)})^{-1} J^{(t)} - I_{nk})v^*\|_2 + \|G\|_2 \|((G^{(t)})^{-1} J^{(t)} - I_{nk})(v^{(t)} - v^*)\|_2. \quad (\text{C.1})
\end{aligned}$$

Recall that Lemma B.2 is also applicable to $G^{(t)}K^{(t)} - J^{(t)}$. Since we have

$$\|V^{(t)} - V^*\|_2 \leq \|V^{(t)} - V^*\|_F = \|v^{(t)} - v^*\|_2 \leq \sigma_1,$$

following similar lines to Appendix B.2, we can show

$$\begin{aligned}\|((S^{(t)})^{-1}J^{(t)} - I_{mn})v^*\|_2 &\leq \sigma_1 \left(\|\bar{U}^{(t)} - \bar{U}^*\|_F^2 + \frac{\sqrt{2}\delta_{2k}k}{1-\delta_{2k}} \|\bar{U}^{(t)} - \bar{U}^*\|_F \right), \\ \|((G^{(t)})^{-1}J^{(t)} - I_{mn})(v^{(t)} - v^*)\|_2 &\leq \sigma_1 \left(\|\bar{U}^{(t)} - \bar{U}^*\|_F^2 + \frac{\sqrt{2}\delta_{2k}k}{1-\delta_{2k}} \|\bar{U}^{(t)} - \bar{U}^*\|_F \right), \\ \|((S^{(t)})^{-1}J^{(t)} - I_{mn})(v^{(t)} - v^*)\|_2 &\leq \sigma_1 \left(\|\bar{U}^{(t)} - \bar{U}^*\|_F^2 + \frac{\sqrt{2}\delta_{2k}k}{1-\delta_{2k}} \|\bar{U}^{(t)} - \bar{U}^*\|_F \right).\end{aligned}$$

Combining the above three inequalities with (C.1), we have

$$\begin{aligned}\|\nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t)}) - \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)})\|_F \\ \leq 2(1 + \delta_{2k})\sigma_1 \left(\|\bar{U}^{(t)} - \bar{U}^*\|_F^2 + \frac{\sqrt{2}\delta_{2k}k}{1-\delta_{2k}} \|\bar{U}^{(t)} - \bar{U}^*\|_F \right).\end{aligned}\quad (\text{C.2})$$

Since $\bar{U}^{(t)}$, δ_{2k} , and ξ satisfy (A.1), then (C.2) further implies

$$\|\nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t)}) - \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)})\|_F \leq \frac{(1 + \delta_{2k})\sigma_k}{\xi} \|\bar{U}^{(t)} - \bar{U}^*\|_F. \quad (\text{C.3})$$

Therefore by Cauchy-Schwartz inequality, (C.3) implies

$$\mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) \leq \|\nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t)}) - \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)})\|_F \leq \frac{(1 + \delta_{2k})\sigma_k}{\xi} \|\bar{U}^{(t)} - \bar{U}^*\|_F,$$

which completes the proof. \square

C.2 Proof of Lemma A.2

Proof. For notational convenience, we omit the index t in $\bar{U}^{*(t)}$ and $V^{*(t)}$, and denote them by \bar{U}^* and V^* respectively. By the strong convexity of $\mathcal{F}(\bar{U}^*, \cdot)$, we have

$$\begin{aligned}\mathcal{F}(\bar{U}^*, V^*) - \frac{1 - \delta_{2k}}{2} \|V^{(t)} - V^*\|_F^2 &\geq \mathcal{F}(\bar{U}^*, V^{(t)}) + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)}), V^* - V^{(t)} \rangle \\ &= \mathcal{F}(\bar{U}^*, V^{(t)}) + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)}), V^{(t+0.5)} - V^{(t)} \rangle + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)}), V^* - V^{(t+0.5)} \rangle.\end{aligned}\quad (\text{C.4})$$

Meanwhile, we define

$$\mathcal{Q}(V; \bar{U}^*, V^{(t)}) = \mathcal{F}(\bar{U}^*, V^{(t)}) + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)}), V - V^{(t)} \rangle + \frac{1}{2\eta} \|V - V^{(t)}\|_F^2.$$

Since η satisfies (A.2) and $\mathcal{F}(\bar{U}^*, V)$ is strongly smooth in V for a fixed orthonormal \bar{U}^* , we have

$$\mathcal{Q}(V; \bar{U}^*, V^{(t)}) \geq \mathcal{F}(\bar{U}^*, V^{(t)}).$$

Combining the above two inequalities, we obtain

$$\begin{aligned}\mathcal{F}(\bar{U}^*, V^{(t)}) + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)}), V^{(t+0.5)} - V^{(t)} \rangle \\ = \mathcal{Q}(V^{(t+0.5)}; \bar{U}^*, V^{(t)}) - \frac{1}{2\eta} \|V^{(t+0.5)} - V^{(t)}\|_F^2 \geq \mathcal{F}(\bar{U}^*, V^{(t+0.5)}) - \frac{1}{2\eta} \|V^{(t+0.5)} - V^{(t)}\|_F^2.\end{aligned}\quad (\text{C.5})$$

Moreover, by the strong convexity of $\mathcal{F}(\bar{U}^*, \cdot)$ again, we have

$$\begin{aligned}\mathcal{F}(\bar{U}^*, V^{(t+0.5)}) &\geq \mathcal{F}(\bar{U}^*, V^*) + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^*), V^{(t+0.5)} - V^* \rangle + \frac{1 - \delta_{2k}}{2} \|V^{(t+0.5)} - V^*\|_F^2 \\ &\geq \mathcal{F}(\bar{U}^*, V^*) + \frac{1 - \delta_{2k}}{2} \|V^{(t+0.5)} - V^*\|_F^2,\end{aligned}\quad (\text{C.6})$$

where the second equalities comes from the optimality condition of $V^* = \operatorname{argmin}_V \mathcal{F}(\bar{U}^*, V)$, i.e.

$$\langle \nabla_V \mathcal{F}(\bar{U}^*, V^*), V^{(t+0.5)} - V^* \rangle \geq 0.$$

Combining (C.4) and (C.5) with (C.6), we obtain

$$\begin{aligned}\mathcal{F}(\bar{U}^*, V^{(t)}) + \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)}), V^{(t+0.5)} - V^{(t)} \rangle \\ \geq \mathcal{F}(\bar{U}^*, V^*) + \frac{1 - \delta_{2k}}{2} \|V^{(t+0.5)} - V^*\|_F^2 - \frac{1}{2\eta} \|V^{(t+0.5)} - V^{(t)}\|_F^2.\end{aligned}\quad (\text{C.7})$$

On the other hand, since $V^{(t+0.5)}$ minimizes $\mathcal{Q}(V; \bar{U}^*, V^{(t)})$, we have

$$\begin{aligned} 0 &\leq \langle \nabla \mathcal{Q}(V^{(t+0.5)}; \bar{U}^*, V^{(t)}), V^* - V^{(t+0.5)} \rangle \\ &\leq \langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)}), V^* - V^{(t+0.5)} \rangle + (1 + \delta_{2k}) \langle V^{(t+0.5)} - V^{(t)}, V^* - V^{(t+0.5)} \rangle. \end{aligned} \quad (\text{C.8})$$

Meanwhile, we have

$$\begin{aligned} &\langle \nabla_V \mathcal{F}(\bar{U}^*, V^{(t)}), V^* - V^{(t+0.5)} \rangle \\ &= \langle \nabla_V \mathcal{F}(\bar{U}^{(t)}, V^{(t)}), V^* - V^{(t+0.5)} \rangle - \mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) \|V^* - V^{(t+0.5)}\|_2 \\ &\geq (1 + \delta_{2k}) \langle V^{(t)} - V^{(t+0.5)}, V^* - V^{(t+0.5)} \rangle - \mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) \|V^* - V^{(t+0.5)}\|_2 \\ &= (1 + \delta_{2k}) \langle V^{(t)} - V^{(t+0.5)}, V^* - V^{(t)} \rangle + \frac{1}{2\eta} \|V^{(t)} - V^{(t+0.5)}\|_F^2 \\ &\quad - \mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) \|V^* - V^{(t+0.5)}\|_2. \end{aligned} \quad (\text{C.9})$$

Combining (C.8) with (C.9), we obtain

$$\begin{aligned} &2 \langle V^{(t)} - V^{(t+0.5)}, V^* - V^{(t)} \rangle \\ &\leq -\eta(1 - \delta_{2k}) \|V^{(t)} - V^*\|_2^2 - \eta(1 - \delta_{2k}) \|V^{(t+0.5)} - V^*\|_2^2 \\ &\quad - \|V^{(t+0.5)} - V^{(t)}\|_2^2 + \mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) \|V^* - V^{(t+0.5)}\|_2. \end{aligned} \quad (\text{C.10})$$

Therefore, combining (C.7) with (C.10), we obtain

$$\begin{aligned} \|V^{(t+0.5)} - V^*\|_F^2 &\leq \|V^{(t+0.5)} - V^{(t)} + V^{(t)} - V^*\|_F^2 \\ &= \|V^{(t+0.5)} - V^{(t)}\|_F^2 + \|V^{(t)} - V^*\|_F^2 + 2 \langle V^{(t+0.5)} - V^{(t)}, V^{(t)} - V^* \rangle \\ &\leq 2\eta \|V^{(t)} - V^*\|_F^2 - \eta(1 - \delta_{2k}) \|V^{(t+0.5)} - V^*\|_F^2 \\ &\quad - \mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) \|V^* - V^{(t+0.5)}\|_2. \end{aligned}$$

Rearranging the above inequality, we obtain

$$\|V^{(t+0.5)} - V^*\|_F \leq \sqrt{\delta_{2k}} \|V^{(t)} - V^*\|_F + \frac{2}{1 + \delta_{2k}} \mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}),$$

which completes the proof. \square

C.3 Proof of Lemma A.3

Proof. Before we proceed with the proof, we first introduce the following lemma.

Lemma C.1. For any matrix $U, \tilde{U} \in \mathbb{R}^{m \times k}$ and $V, \tilde{V} \in \mathbb{R}^{n \times k}$, we have

$$\|UV^\top - \tilde{U}\tilde{V}^\top\|_F \leq \|U\|_2 \|V - \tilde{V}\| + \|\tilde{V}\|_2 \|U - \tilde{U}\|_F.$$

Proof. By linear algebra, we have

$$\begin{aligned} \|UV^\top - \tilde{U}\tilde{V}^\top\|_F &= \|UV^\top - U\tilde{V}^\top + U\tilde{V}^\top - \tilde{U}\tilde{V}^\top\|_F \\ &\leq \|UV^\top - U\tilde{V}^\top\|_F + \|U\tilde{V}^\top - \tilde{U}\tilde{V}^\top\|_F \leq \|U\|_2 \|V - \tilde{V}\|_F + \|\tilde{V}\|_2 \|U - \tilde{U}\|_F. \end{aligned} \quad (\text{C.11})$$

Thus, we conclude the proof. \square

By Lemma C.1, we have

$$\begin{aligned} \|R_{\bar{V}}^{(t+0.5)} - \bar{V}^{*(t+1)\top} V^{*(t)}\|_F &= \|\bar{V}^{(t+0.5)\top} V^{(t+0.5)} - \bar{V}^{*(t+1)\top} V^{*(t)}\|_F \\ &\leq \|\bar{V}^{(t+0.5)}\|_2 \|V^{(t+0.5)} - V^{*(t)}\|_F + \|V^{*(t)}\|_2 \|\bar{V}^{(t+0.5)} - \bar{V}^{*(t+1)}\|_F \\ &\leq \|V^{(t+0.5)} - V^{*(t)}\|_F + \frac{2\sigma_1}{\sigma_k} \|V^{(t+0.5)} - V^{*(t)}\|_F, \end{aligned} \quad (\text{C.12})$$

where the last inequality comes from Lemma 4.5. Moreover, let $U^{*(t+1)} = \bar{U}^{*(t)} (\bar{V}^{*(t+1)\top} V^{*(t)})^\top$. Then we can verify

$$U^{*(t+1)} \bar{V}^{*(t+1)} = \bar{U}^{*(t)} V^{*(t)\top} \bar{V}^{*(t+1)} \bar{V}^{*(t+1)\top} = M^* \bar{V}^{*(t+1)} \bar{V}^{*(t+1)\top} = M^*,$$

where the last equality holds, since $\bar{V}^{*(t+1)}\bar{V}^{*(t+1)\top}$ is exactly the projection matrix for the row space of M^* . Thus by Lemma C.1, we have

$$\begin{aligned}\|U^{(t+1)} - U^{*(t+1)}\|_F &= \|\bar{U}^{(t)} R_{\bar{V}}^{(t+0.5)\top} - \bar{U}^{*(t)} (\bar{V}^{*(t+1)\top} V^{*(t)})^\top\|_F \\ &\leq \|\bar{U}^{(t)}\|_2 \|R_{\bar{V}}^{(t+0.5)} - \bar{V}^{*(t+1)\top} V^{*(t)}\|_F + \|\bar{V}^{*(t+1)\top} V^{*(t)}\|_2 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \\ &\leq \left(1 + \frac{2\sigma_1}{\sigma_k}\right) \|V^{(t+0.5)} - V^{*(t)}\|_F + \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F,\end{aligned}$$

where the last inequality comes from (C.12), $\|\bar{V}^{*(t+1)}\|_2 = \sigma_1$, and $\|\bar{U}^{(t)}\|_2 = 1$. \square

C.4 Proof of Lemma A.4

Proof. Following similar lines to Appendix B.5, we have

$$\|\bar{U}^{(0)} - \bar{U}^{*(0)}\|_F \leq \frac{\sigma_k^2}{4\xi\sigma_1^2}.$$

In Appendix B.5, we have already shown

$$\left\| \frac{\bar{U}^{(0)} \bar{\Sigma}^{(0)} \bar{V}^{(0)\top}}{1 + \delta_{2k}} - M^* \right\|_F \leq 2\sqrt{\delta_{2k}(1 + 3\delta_{2k})} \|\bar{\Sigma}^*\|_F.$$

Then by Lemma C.1 we have

$$\begin{aligned}\left\| \frac{\bar{U}^{(0)} \bar{\Sigma}^{(0)}}{1 + \delta_{2k}} - V^{*(0)} \right\|_F &= \left\| \frac{\bar{U}^{(0)\top} \bar{U}^{(0)} \bar{\Sigma}^{(0)} \bar{V}^{(0)\top}}{1 + \delta_{2k}} - \bar{U}^{*(0)\top} M^* \right\|_F \\ &\leq \|\bar{U}^{(0)}\|_2 \left\| \frac{\bar{U}^{(0)} \bar{\Sigma}^{(0)} \bar{V}^{(0)\top}}{1 + \delta_{2k}} - M^* \right\|_F + \|M^*\|_2 \|\bar{U}^{(0)} - \bar{U}^{*(0)}\|_F \\ &\leq 2\sqrt{\delta_{2k}k(1 + 3\delta_{2k})} \sigma_1 + \frac{\sigma_k^2}{4\xi\sigma_1^2}.\end{aligned}\tag{C.13}$$

By triangle inequality, we further have

$$\begin{aligned}\|\bar{U}^{(0)} \bar{\Sigma}^{(0)} - V^{*(0)}\|_F &\leq (1 + \delta_{2k}) \left\| \frac{\bar{U}^{(0)} \bar{\Sigma}^{(0)}}{1 + \delta_{2k}} - V^{*(0)} \right\|_F + \delta_{2k} \|V^{*(0)}\|_F \\ &\stackrel{(i)}{\leq} (1 + \delta_{2k}) \left(2\sqrt{\delta_{2k}k(1 + 3\delta_{2k})} \sigma_1 + \frac{\sigma_k^2}{4\xi\sigma_1^2} \right) + \delta_{2k} \sigma_1 \sqrt{k} \\ &\stackrel{(ii)}{\leq} \left(\frac{\sigma_k^3}{9\sigma_1^3\xi} + \frac{\sigma_k^2}{3\sigma_1^3\xi^2} + \frac{\sigma_k^3}{192\xi^3\sigma_1^2} \right) \sigma_1 \stackrel{(iii)}{\leq} \frac{\sigma_k^2}{2\xi\sigma_1},\end{aligned}$$

where (i) comes from (C.13) and $\|V^{*(0)}\|_F = \|M^*\|_F \leq \sigma_1 \sqrt{k}$, (ii) comes from (A.4), and (iii) comes from the definition of ξ and $\sigma_1 \geq \sigma_k$. \square

C.5 Proof of Corollary A.5

Proof. Since (A.5) ensures that (A.1) of Lemma A.1 holds, we have

$$\begin{aligned}\|V^{(t+0.5)} - V^{*(t)}\|_F &\leq \sqrt{\delta_{2k}} \|V^{(t)} - V^{*(t)}\|_F + \frac{2}{1 + \delta_{2k}} \mathcal{D}(V^{(t+0.5)}, V^{(t)}, \bar{U}^{(t)}) \\ &\stackrel{(i)}{\leq} \sqrt{\delta_{2k}} \|V^{(t)} - V^{*(t)}\|_F + \frac{2}{1 + \delta_{2k}} \cdot \frac{(1 + \delta_{2k})\sigma_k}{\xi} \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \\ &\stackrel{(ii)}{\leq} \frac{\sigma_k^2}{12\xi\sigma_1^2} \|V^{(t)} - V^{*(t)}\|_F + \frac{2\sigma_k}{\xi} \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \\ &\stackrel{(iii)}{\leq} \frac{\sigma_k^2}{12\xi\sigma_1^2} \cdot \frac{\sigma_k^2}{2\xi\sigma_1} + \frac{2\sigma_k}{\xi} \cdot \frac{\sigma_k^2}{4\xi\sigma_1^2} \stackrel{(iv)}{\leq} \frac{13\sigma_k^3}{24\xi^2\sigma_1^2} \stackrel{(v)}{\leq} \frac{\sigma_k}{4},\end{aligned}\tag{C.14}$$

where (i) comes from Lemma A.2, (ii) and (iii) come from (A.5), and (iii) and (iv) come from the definition of ξ and $\sigma_k \leq \sigma_1$. Since (C.14) ensures that (4.2) of Lemma 4.5, we obtain

$$\begin{aligned} \|\bar{V}^{(t+1)} - \bar{V}^{*(t+1)}\|_F &\leq \frac{2}{\sigma_k} \|V^{(t+0.5)} - V^{*(t)}\|_F \stackrel{(i)}{\leq} \frac{2\sqrt{\delta_{2k}}}{\sigma_k} \|V^{(t)} - V^{*(t)}\|_F + \frac{4}{\xi} \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \\ &\stackrel{(ii)}{\leq} \left(\frac{\sigma_k}{3\xi\sigma_1} + \frac{4}{\xi} \right) \cdot \frac{\sigma_k^2}{4\xi\sigma_1^2} \stackrel{(iii)}{\leq} \frac{\sigma_k^2}{4\xi\sigma_1^2}, \end{aligned} \quad (\text{C.15})$$

where (i) and (ii) come from (C.14), and (iii) comes from the definition of ξ and $\sigma_1 > \sigma_k$. Moreover, since (C.14) ensures that (A.3) of Lemma A.3 holds, we have

$$\begin{aligned} \|U^{(t)} - U^{*(t+1)}\|_F &\leq \frac{3\sigma_1}{\sigma_k} \|V^{(t+0.5)} - V^{*(t)}\|_F + \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \\ &\stackrel{(i)}{\leq} \frac{3\sigma_1\sqrt{\delta_{2k}}}{\sigma_k} \|V^{(t)} - V^{*(t)}\|_F + \left(\frac{6}{\xi} + 1 \right) \sigma_1 \|\bar{U}^{(t)} - \bar{U}^{*(t)}\|_F \\ &\stackrel{(ii)}{\leq} \frac{3\sigma_1}{\sigma_k} \cdot \frac{\sigma_k^3}{12\xi\sigma_1^3} \cdot \frac{\sigma_k^2}{2\xi\sigma_1} + \left(\frac{6}{\xi} + 1 \right) \cdot \frac{\sigma_k^2}{4\xi\sigma_1} \\ &= \left(\frac{\sigma_k^2}{4\xi^2\sigma_1^2} + \frac{3}{\xi} + 1/2 \right) \frac{\sigma_k^2}{2\xi\sigma_1} \stackrel{(iii)}{\leq} \frac{\sigma_k^2}{2\xi\sigma_1}, \end{aligned}$$

where (i) comes from (C.14), (ii) comes from (A.5), and (iii) comes from the definition of ξ and $\sigma_1 \geq \sigma_k$. \square

D Algorithms for Matrix Completion

Algorithm 2 A family of nonconvex optimization algorithms for matrix completion. The incoherence factorization algorithm IF(\cdot) is illustrated in Algorithm 3, and the partition algorithm Partition(\cdot), which is proposed by [10], is provided in Algorithm 4. The initialization procedures $\text{INT}_{\bar{U}}(\cdot)$ and $\text{INT}_{\bar{V}}(\cdot)$ are provided in Algorithm 5 and Algorithm 6. Here $\mathcal{F}_{\mathcal{W}}(\cdot)$ is defined in (5.2).

Input: $\mathcal{P}_{\mathcal{W}}(M^*)$

Parameter: Step size η , Total number of iterations T

$(\{\mathcal{W}_t\}_{t=0}^{2T}, \tilde{\rho}) \leftarrow \text{Partition}(\mathcal{W})$, $\mathcal{P}_{\mathcal{W}_0}(\tilde{M}) \leftarrow \mathcal{P}_{\mathcal{W}_0}(M^*)$, and $\tilde{M}_{ij} \leftarrow 0$ for all $(i, j) \notin \mathcal{W}_0$

$(\bar{U}^{(0)}, V^{(0)}) \leftarrow \text{INT}_{\bar{U}}(\tilde{M})$, $(\bar{V}^{(0)}, U^{(0)}) \leftarrow \text{INT}_{\bar{V}}(\tilde{M})$

For: $t = 0, \dots, T-1$

$$\left. \begin{aligned} &\text{Alternating Exact Minimization : } V^{(t+0.5)} \leftarrow \arg\min_V \mathcal{F}_{\mathcal{W}_{2t+1}}(\bar{U}^{(t)}, V) \\ &(\bar{V}^{(t+1)}, R_{\bar{V}}^{(t+0.5)}) \leftarrow \text{IF}(V^{(t+0.5)}) \\ &\text{Alternating Gradient Descent : } V^{(t+0.5)} \leftarrow V^{(t)} - \eta \nabla_V \mathcal{F}_{\mathcal{W}_{2t+1}}(\bar{U}^{(t)}, V^{(t)}) \\ &(\bar{V}^{(t+1)}, R_{\bar{V}}^{(t+0.5)}) \leftarrow \text{IF}(V^{(t+0.5)}), U^{(t)} \leftarrow \bar{U}^{(t)} R_{\bar{V}}^{(t+0.5)\top} \\ &\text{Gradient Descent : } V^{(t+0.5)} \leftarrow V^{(t)} - \eta \nabla_V \mathcal{F}_{\mathcal{W}_{2t+1}}(\bar{U}^{(t)}, V^{(t)}) \\ &(\bar{V}^{(t+1)}, R_{\bar{V}}^{(t+0.5)}) \leftarrow \text{IF}(V^{(t+0.5)}), U^{(t+1)} \leftarrow \bar{U}^{(t)} R_{\bar{V}}^{(t+0.5)\top} \end{aligned} \right\} \text{Updating } V.$$

$$\left. \begin{aligned} &\text{Alternating Exact Minimization : } U^{(t+0.5)} \leftarrow \arg\min_U \mathcal{F}_{\mathcal{W}_{2t+2}}(U, \bar{V}^{(t+1)}) \\ &(\bar{U}^{(t+1)}, R_{\bar{U}}^{(t+0.5)}) \leftarrow \text{IF}(U^{(t+0.5)}) \\ &\text{Alternating Gradient Descent : } U^{(t+0.5)} \leftarrow U^{(t)} - \eta \nabla_U \mathcal{F}_{\mathcal{W}_{2t+2}}(U^{(t)}, \bar{V}^{(t+1)}) \\ &(\bar{U}^{(t+1)}, R_{\bar{U}}^{(t+0.5)}) \leftarrow \text{IF}(U^{(t+0.5)}), V^{(t+1)} \leftarrow \bar{V}^{(t+1)} R_{\bar{U}}^{(t+0.5)\top} \\ &\text{Gradient Descent : } U^{(t+0.5)} \leftarrow U^{(t)} - \eta \nabla_U \mathcal{F}_{\mathcal{W}_{2t+2}}(U^{(t)}, \bar{V}^{(t)}) \\ &(\bar{U}^{(t+1)}, R_{\bar{U}}^{(t+0.5)}) \leftarrow \text{IF}(U^{(t+0.5)}), V^{(t+1)} \leftarrow \bar{V}^{(t)} R_{\bar{U}}^{(t+0.5)\top} \end{aligned} \right\} \text{Updating } U.$$

End for

Output: $M^{(T)} \leftarrow U^{(T-0.5)} \bar{V}^{(T)\top}$ or $\bar{U}^{(T)} V^{(T)\top}$ (Gradient Descent Only)

Algorithm 3 The incoherence factorization algorithm for matrix completion. It guarantees that the solutions satisfy the incoherence condition throughout all iterations.

Input: W^{in}
 $r \leftarrow$ Number of rows of W^{in}
Parameter: Incoherence parameter μ
 $(\bar{W}^{\text{in}}, R_{\bar{W}}^{\text{in}}) \leftarrow \text{QR}(W^{\text{in}})$
 $\tilde{W} \leftarrow \underset{W}{\text{argmin}} \|W - \bar{W}^{\text{in}}\|_{\text{F}}^2$ subject to $\max_j \|W_{j*}\|_2 \leq \mu\sqrt{k/r}$
 $(\bar{W}^{\text{out}}, R_{\bar{W}}^{\text{tmp}}) \leftarrow \text{QR}(W^{\text{out}})$
 $R_{\bar{W}}^{\text{out}} = \bar{W}^{\text{out}\top} W^{\text{in}}$
Output: $\bar{W}^{\text{out}}, R_{\bar{W}}^{\text{out}}$

Algorithm 4 The observation set partition algorithm for matrix completion. It guarantees the independence among all $2T + 1$ output observation sets.

Input: $\mathcal{W}, \bar{\rho}$
 $\tilde{\rho} = 1 - (1 - \bar{\rho})^{\frac{1}{2T+1}}$
For: $t = 0, \dots, 2T$
 $\tilde{\rho}_t = \frac{(mn)!\tilde{\rho}^{t+1}(1 - \bar{\rho})^{mn-t-1}}{\bar{\rho}(mn - t - 1)!(t + 1)!}$
End for
 $\mathcal{W}_0 = \emptyset, \dots, \mathcal{W}_{2T} = \emptyset$
For every $(i, j) \in \mathcal{W}$
Sample t from $\{0, \dots, 2T\}$ with probability $\{\tilde{\rho}_0, \dots, \tilde{\rho}_{2T}\}$
Sample (w/o replacement) a set \mathcal{B} such that $|\mathcal{B}| = t$ from $\{0, \dots, 2T\}$ with equal probability
Add (i, j) to \mathcal{W}_ℓ for all $\ell \in \mathcal{B}$
End for
Output: $\{\mathcal{W}_t\}_{t=0}^{2T}, \tilde{\rho}$

Algorithm 5 The initialization procedure $\text{INT}_{\bar{U}}(\cdot)$ for matrix completion. It guarantees that the initial solutions satisfy the incoherence condition throughout all iterations.

Input: \tilde{M}
Parameter: Incoherence parameter μ
 $(\tilde{U}, \tilde{D}, \tilde{V}) \leftarrow \text{KSVD}(\tilde{M})$
 $\tilde{U}^{\text{tmp}} \leftarrow \underset{U}{\text{argmin}} \|U - \tilde{U}\|_{\text{F}}^2$ subject to $\max_i \|U_{i*}\|_2 \leq \mu\sqrt{k/m}$
 $(\bar{U}^{\text{out}}, R_{\bar{U}}^{\text{out}}) \leftarrow \text{QR}(\tilde{U}^{\text{tmp}})$
 $\tilde{V}^{\text{tmp}} \leftarrow \underset{V}{\text{argmin}} \|V - \tilde{V}^{\text{tmp}}\|_{\text{F}}^2$ subject to $\max_j \|V_{j*}\|_2 \leq \mu\sqrt{k/n}$
 $(\bar{V}^{\text{out}}, R_{\bar{V}}^{\text{out}}) \leftarrow \text{QR}(\tilde{V}^{\text{tmp}})$
 $V^{\text{out}} = \bar{V}^{\text{out}}(\bar{U}^{\text{out}\top} \tilde{M} \bar{V}^{\text{out}})^\top$
Output: $\bar{U}^{\text{out}}, V^{\text{out}}$

Algorithm 6 The initialization procedure $\text{INT}_{\bar{V}}(\cdot)$ for matrix completion. It guarantees that the initial solutions satisfy the incoherence condition throughout all iterations.

Input: \widetilde{M}

Parameter: Incoherence parameter μ

$(\widetilde{U}, \widetilde{D}, \widetilde{V}) \leftarrow \text{KSVD}(\widetilde{M})$

$\widetilde{V}^{\text{tmp}} \leftarrow \underset{V}{\text{argmin}} \|V - \widetilde{V}\|_{\text{F}}^2 \text{ subject to } \max_j \|V_{j*}\|_2 \leq \mu\sqrt{k/n}$

$(\bar{V}^{\text{out}}, R_{\bar{V}}^{\text{out}}) \leftarrow \text{QR}(\widetilde{V}^{\text{tmp}})$

$\widetilde{U}^{\text{tmp}} \leftarrow \underset{U}{\text{argmin}} \|U - \widetilde{U}^{\text{tmp}}\|_{\text{F}}^2 \text{ subject to } \max_i \|U_{i*}\|_2 \leq \mu\sqrt{k/m}$

$(\bar{U}^{\text{out}}, R_{\bar{U}}^{\text{out}}) \leftarrow \text{QR}(\widetilde{U}^{\text{tmp}})$

$U^{\text{out}} = \bar{U}^{\text{out}} (\bar{U}^{\text{out}}{}^{\text{T}} \widetilde{M} \bar{V}^{\text{out}})$

Output: $\bar{V}^{\text{out}}, U^{\text{out}}$
