A Preliminaries

Lemma 1 (Lemma B.1 of [7]). *Suppose $\mathcal{A}(\cdot)$ satisfies $2k$-RIP. For any $X, U \in \mathbb{R}^{m\times k}$ and $Y, V \in \mathbb{R}^{n\times k}$, we have*

$$|\langle \mathcal{A}(XY^\mathsf{T}), \mathcal{A}(UV^\mathsf{T})\rangle - \langle X^\mathsf{T}U, Y^\mathsf{T}V\rangle| \leq 3\delta_{2k}\|XY^\mathsf{T}\|_F \cdot \|UV^\mathsf{T}\|_F$$

Lemma 2 (Lemma 2.1 of [6]). *Let $b = \mathcal{A}(M^*) + \varepsilon$, where M^* is a matrix with the rank of k, $\mathcal{A}$ is the linear mapping operator satisfies $2k$-RIP with constant $\delta_{2k} < 1/3$, and ε is a bounded error vector. Let $M^{(t+1)}$ be the $t+1$-th step iteration of SVP, then we have*

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

Lemma 3 (Lemma 4.5 of [19]). *Suppose that $Y^{(t+0.5)}$ in Alg. 1 satisfies $\|Y^{(0.5)} - V^{(t)}\|_F \leq \sigma_k/4$. Then, there exists a factorization of matrix $M^* = U^{(t+1)}\bar{V}^{(t+1)\mathsf{T}}$ such that $V^{(t+1)} \in \mathbb{R}^{n\times k}$ is an orthonomal matrix, and satisfies*

$$\|\bar{Y}^{(t+1)} - \bar{V}^{(t+1)}\|_F \leq 2/\sigma_k \cdot \|Y^{(0.5)} - V^{(t)}\|_F.$$

B Proofs

B.1 Roadmap

For notation simplicity, we denote a finite positive constant $\xi > 1$ throughout this paper.

Lemma 4. *Suppose the linear operator $\mathcal{A}(\cdot)$ satisfies $2k$-RIP with parameter δ_{2k}. For any orthonormal matrix $\bar{X} \in \mathbb{R}^{m\times k}$, the function $\mathcal{F}(\bar{X}, Y)$ with bounded weights is strongly convex and smooth. In particular, if any weight w_i in $\mathcal{F}(\bar{X}, Y)$ belongs to $[w_-, w_+]$, the value of*

$$\mathcal{F}(\bar{X}, Y') - \mathcal{F}(\bar{X}, Y) - \langle \nabla_Y \mathcal{F}(\bar{X}, Y), Y' - Y\rangle$$

is bounded by

$$[w_-(1-\delta_{2k})\|Y' - Y\|_F^2, w_+(1+\delta_{2k})\|Y' - Y\|_F^2]$$

for all Y, Y'.

The first step is to prove strongly convexity and smoothness of $\mathcal{F}(X,Y)$ if one variable is fixed by a orthonormal matrix as follows:

Lemma 5. *Suppose that δ_{2k} and $\bar{X}^{(t)}$ satisfy*

$$\delta_{2k} \leq \frac{\sqrt{2}w_-^2(1-\delta_{2k})^2\sigma_k}{24\xi w_+ k(1+\delta_{2k})\sigma_1}. \tag{5}$$

and

$$\|\bar{X}^{(t)} - \bar{U}^{(t)}\|_F \leq \frac{w_-(1-\delta_{2k})\sigma_k}{2\xi w_+(1+\delta_{2k})\sigma_1} \tag{6}$$

Then we have:

$$\|Y^{(t+0.5)} - V^{(t)}\|_F \leq \frac{\sigma_k}{2\xi}\|\bar{X}^{(t)} - \bar{U}^{(t)}\|_F.$$

Lemma 4 shows that $\mathcal{F}(X,Y)$ can be block-wise strongly convex and smooth if the weights w_i belongs to $[w_-, w_+]$. In the following, we use U and V to denote the optimal factorization of $M^* = UV^\intercal$. Note that U and V are unique up to orthogonal transformations. The following lemma shows that by taking the block-wise minimum, the distance between the newly updated variable $Y^{(t+0.5)}$ and its "nearby" $V^{(t)}$ is upper bounded by the distance between $X^{(t)}$ and its corresponding neighbor $U^{(t)}$.

Lemma 6. *Suppose that δ_{2k} satisfies*

$$\delta_{2k} \leq \frac{w_-^2(1-\delta_{2k})^2\sigma_k^4}{48\xi^2 k w_+^2(1+\delta_{2k})^2\sigma_1^4}.$$

We have $\|\bar{Y}^{(t+1)} - \bar{V}^{(t+1)}\|_F \leq \frac{1}{\xi}\|\bar{X}^{(t)} - \bar{U}^{(t)}\|_F$.

By this lemma, we can prove that the distance between $\bar{Y}^{(t)}$ and $\bar{V}^{(t)}$ decreases with a rate of $1/\xi'$. The same applies to the distance between $\bar{X}^{(t)}$ and $\bar{U}^{(t)}$.

B.2 Proof of Lemma 4

Now we begin to prove these lemmas. Note that a similar technique has also been used by [19]. Since we should fix $X^{(t)}$ or $Y^{(t)}$ as orthonormal matrices, we perform a QR decomposition after getting the minimum. The following lemma shows the distance between $\bar{Y}^{(t+1)}$ and its "nearby" $\bar{V}^{(t+1)}$ is still under control.

Proof. Since $\mathcal{F}(\bar{X}, Y)$ is a quadratic function, we have

$$\begin{aligned}
\mathcal{F}(\bar{X}, Y') &= \mathcal{F}(\bar{X}, Y) + \langle \nabla_Y \mathcal{F}(\bar{X}, Y), Y' - Y \rangle \\
&\quad + \frac{1}{2}(\mathrm{vec}(Y') - \mathrm{vec}(Y))^\top \nabla_Y^2 \mathcal{F}(\bar{X}, Y)(\mathrm{vec}(Y') - \mathrm{vec}(Y)),
\end{aligned}$$

and it suffices to bound the singular values of the Hessian matrix $S_\omega := \nabla_Y^2 \mathcal{F}(\bar{X}, Y)$ so that

$$\begin{aligned}
\mathcal{F}(\bar{X}, Y') - \mathcal{F}(\bar{X}, Y) - \langle \nabla_Y \mathcal{F}(\bar{X}, Y), Y' - Y \rangle &\leq \frac{\sigma_{\max}(S_\omega)}{2}\|Y' - Y\|_F^2 \\
\mathcal{F}(\bar{X}, Y') - \mathcal{F}(\bar{X}, Y) - \langle \nabla_Y \mathcal{F}(\bar{X}, Y), Y' - Y \rangle &\geq \frac{\sigma_{\min}(S_\omega)}{2}\|Y' - Y\|_F^2.
\end{aligned}$$

Now we proceed to derive the Hessian matrix S_ω. Using the fact $\mathrm{vec}(AXB) = (B^\top \otimes A)\mathrm{vec}(X)$, we can write S_ω as follows:

$$\begin{aligned}
S_\omega &= \sum_{i=1}^{p} 2w_i \cdot \mathrm{vec}(A_i^\top \bar{X})\mathrm{vec}^\top(A_i^\top \bar{X}) \\
&= \sum_{i=1}^{p} 2w_i \cdot (I_k \otimes A_i^\top)\mathrm{vec}(\bar{X})\mathrm{vec}^\top(\bar{X})(I_k \otimes A_i).
\end{aligned}$$

Consider a matrix $Z \in \mathbb{R}^{n \times k}$ with $\|Z\|_F = 1$, and we denote $z = \mathrm{vec}(Z)$. Then we have

$$\begin{aligned}
z^\top S_\omega z &= \sum_{i=1}^{p} 2w_i \cdot z^\top (I_k \otimes A_i^\top)\mathrm{vec}(\bar{X})\mathrm{vec}^\top(\bar{X})(I_k \otimes A_i) \\
&= \sum_{i=1}^{p} 2w_i \cdot \mathrm{vec}^\top(A_i Z)\mathrm{vec}(\bar{X})\mathrm{vec}^\top(\bar{X})\mathrm{vec}(A_i Z) \\
&= \sum_{i=1}^{p} 2w_i \cdot \mathrm{tr}^2(\bar{X}^\top A_i Z) = \sum_{i=1}^{p} 2w_i \cdot \mathrm{tr}^2(A_i^\top \bar{X} Z^\top).
\end{aligned}$$

From the $2k$-RIP property of $\mathcal{A}_\rangle$, we have

$$\begin{aligned}
z^\top S_\omega z &\leq \sum_{i=1}^{p} 2w_+ \mathrm{tr}^2(\bar{X}^\top A_i Z) \\
&\leq 2w_+(1 + \delta_{2k})\|\bar{X} Z^\top\|_F \\
&= 2w_+(1 + \delta_{2k})\|Z^\top\|_F = 2w_+(1 + \delta_{2k}).
\end{aligned}$$

Similiarly, we also have

$$z^\mathsf{T} S_\omega z \geq 2w_-(1-\delta_{2k}).$$

Therefore, the maximum singular value $\sigma_{\max}$ is upper bounded by $2w_+(1+\delta_{2k})$ and the minimum singular value $\sigma_{\min}$ is lower bounded by $2w_-(1-\delta_{2k})$, and the Lemma has been proved. $\square$

B.3 Proof of Lemma 5

We prove this lemma by introducing a divergence function as follows.

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

Lemma 7. *Under the same condition in Lemma 5, we have*

$$\mathcal{D}(Y^{(t+0.5)}, Y^{(t+0.5)}, \bar{X}^{(t)}) \leq \frac{3(1-\delta_{2k})\sigma_k}{2\xi} \cdot \frac{w_+^2}{w_-} \|\bar{X}^{(t)} - \bar{U}^{(t)}\|. \tag{7}$$

Proof of Lemma 7. In this proof we omit the iteration superscriptor, and Y stands particularly for $Y^{(t+0.5)}$. Since b_i is measured by $\langle A_i, \bar{U} V^\mathsf{T} \rangle$, we have

$$\mathcal{F}(\bar{X}, Y) = \sum_{i=1}^{p} w_i (\langle A_i, \bar{X} Y^\mathsf{T} \rangle - \langle A_i, \bar{U} V^\mathsf{T} \rangle)^2.$$

By taking the partial derivatives on Y we have

$$\begin{aligned} \nabla_Y \mathcal{F}(\bar{X}, Y) &= \sum_{i=1}^{p} 2w_i (\langle A_i, \bar{X} Y^\mathsf{T} \rangle - \langle A_i, \bar{U} V^\mathsf{T} \rangle) A_i^\mathsf{T} X \\ &= \sum_{i=1}^{p} 2w_i (\langle A_i^\mathsf{T} \bar{X}, Y \rangle - \langle A_i^\mathsf{T} \bar{U}, V \rangle) A_i^\mathsf{T} X \end{aligned}$$

Let $x := \mathrm{vec}(\bar{X})$, $y := \mathrm{vec}(Y)$, $u := \mathrm{vec}(\bar{U})$, and $v := \mathrm{vec}(V)$. Since Y

minimizes $\mathcal{F}(\bar{X}, \hat{Y})$, we have

$$\begin{aligned}
\mathrm{vec}(\nabla_Y \mathcal{F}(\bar{X}, Y)) &= \sum_{i=1}^{p} 2w_i(\langle A_i^\mathsf{T} \bar{X}, Y\rangle - \langle A_i^\mathsf{T} \bar{U}, V\rangle) A_i^\mathsf{T} x \\
&= \sum_{i=1}^{p} 2w_i((\mathrm{vec}(A_i^\mathsf{T} \bar{X}) \cdot \langle A_i^\mathsf{T} \bar{X}, Y\rangle - \mathrm{vec}(A_i^\mathsf{T} \bar{X}) \cdot \langle A_i^\mathsf{T} \bar{X}, Y\rangle)) \\
&= \sum_{i=1}^{p} 2w_i((I_k \otimes A_i^\mathsf{T}) x x^\mathsf{T} (I_k \otimes A_i) y - (I_k \otimes A_i^\mathsf{T}) x u^\mathsf{T} (I_k \otimes A_i) v)
\end{aligned}$$

We denote

$$S_\omega = \sum_{i=1}^{p} 2w_i \cdot (I_k \otimes A_i^\mathsf{T}) x x^\mathsf{T} (I_k \otimes A_i),$$

and

$$J_\omega = \sum_{i=1}^{p} 2w_i \cdot (I_k \otimes A_i^\mathsf{T}) x u^\mathsf{T} (I_k \otimes A_i),$$

So the equation becomes $S_\omega y - J_\omega v = 0$ and since S_ω is invertible we have $y = (S_\omega)^{-1} J_\omega v$. Meanwhile, we denote

$$G_\omega = \sum_{i=1}^{p} 2w_i \cdot (I_k \otimes A_i^\mathsf{T}) u u^\mathsf{T} (I_k \otimes A_i)$$

as the Hessian matrix of $\nabla_Y^2 \mathcal{F}(\bar{U}, Y)$. Then, the partial gradient $\nabla_Y \mathcal{F}(\bar{U}, Y)$ can be written as

$$\begin{aligned}
\mathrm{vec}(\nabla_Y \mathcal{F}(\bar{U}, Y)) &= \sum_{i=1}^{p} 2w_i(\langle A_i^\mathsf{T} \bar{U}, Y\rangle - \langle A_i^\mathsf{T} \bar{U}, V\rangle)(I_k \otimes A_i^\mathsf{T}) u \\
&= \sum_{i=1}^{p} 2w_i((I_k \otimes A_i^\mathsf{T}) u u^\mathsf{T} (I_k \otimes A_i) y - (I_k \otimes A_i^\mathsf{T}) u u^\mathsf{T} (I_k \otimes A_i) v) \\
&= G_\omega (y - v) \\
&= G_\omega (S_\omega^{-1} J_\omega - I_{nk}) v.
\end{aligned}$$

Since we have $\mathrm{vec}(\nabla_Y \mathcal{F}(\bar{X}, Y)) = 0$, the divergence $\mathcal{D} = \langle \nabla_Y(\bar{U}, Y), (Y - V)/\|(\|Y - V)\rangle_F$. So we need to bound $\nabla_Y \mathcal{F}(\bar{U}, Y)$. Let $K := \bar{X}^\mathsf{T} \bar{U} \otimes I_n$. To get the estimate of $S_\omega^{-1} J_\omega - I_{nk}$, we rewrite it as

$$S_\omega^{-1} J_\omega - I_{nk} = K - I_{nk} + S_\omega^{-1}(J_\omega - S_\omega K).$$

We firstly bound the term $(K - I_{nk})v$. Recall $\text{vec}(AXB) = (B^\top \otimes A)\text{vec}(X)$, we have

$$\begin{aligned}(K - I_{nk})v &= ((\bar{X}^\top \bar{U} - I_k) \otimes I_n)v = \text{vec}(V(\bar{U}^\top X - I_k)) \\ \|(K - I_{nk})v\|_2 &= \|V(\bar{U}^\top \bar{X} - I_k)\|_F \leq \sigma_1 \|\bar{U}^\top \bar{X} - I_k\|_F \\ &\leq \sigma_1 \|(\bar{X} - \bar{U})^\top (\bar{X} - \bar{U})\|_F \leq \sigma_1 \|\bar{X} - \bar{U}\|_F^2\end{aligned}$$

We then bound the term $J_\omega - S_\omega K$. For any two matrices $Z_1, Z_2 \in \mathbb{R}^{n \times k}$, we denote $z_1 := \text{vec}(Z_1)$ and $z_2 := \text{vec}(Z_2)$. Then we have:

$$\begin{aligned}& z_1^\top (S_\omega K - J_\omega) z_2 \\ = \ & \sum_{i=1}^p 2w_i z_1^\top (I_k \otimes A_i^\top) x \{x^\top (I_k \otimes A_i)(\bar{X}^\top \bar{U} \otimes I_n)) - u^\top (I_k \otimes A_i)\} z_2 \\ = \ & \sum_{i=1}^p 2w_i \langle Z_1, A_i^\top \bar{X} \rangle \cdot (x^\top (\bar{X}^\top \bar{U} \otimes A_i) z_2 - \langle \bar{U}, A_i Z \rangle) \\ = \ & \sum_{i=1}^p 2w_i \langle A_i, \bar{X} Z_1^\top \rangle (\langle A_i, \bar{X}\bar{X}^\top - I_m) \bar{U} Z_2^\top \rangle \\ \leq \ & 2w_+ \langle \mathcal{A}(\bar{X} Z_1^\top), \mathcal{A}((\bar{X}\bar{X}^\top - I_m)\bar{U} Z_2^\top) \rangle\end{aligned}$$

Since $\bar{X}^\top (\bar{X}\bar{X}^\top - I_m)\bar{U} = 0$, by Lemma 1 we have

$$\begin{aligned}& z_1^\top (S_\omega K - J_\omega) z_2 \\ \leq \ & 2w_+ \cdot 3\delta_{2k} \|\bar{X} Z_1^\top\|_F \|(\bar{X}\bar{X}^\top - I_m)\bar{U} Z_2^\top\|_F \\ \leq \ & 6w_+ \delta_{2k} \|Z_1\|_F \sqrt{\|\bar{U}^\top (\bar{X}\bar{X}^\top - I_m)\bar{U}\|_F \|Z_2^\top Z_2\|_F} \\ = \ & 6w_+ \delta_{2k} \sqrt{\|\bar{U}^\top (\bar{X}\bar{X}^\top - I_m)\bar{U}\|_F} \\ \leq \ & 6w_+ \delta_{2k} \sqrt{2k} \|\bar{X} - \bar{U}\|_F.\end{aligned}$$

Thus, the spectral norm of this term is upper bounded by $6w_+ \delta_{2k} \sqrt{2k} \|\bar{X} - \bar{U}\|_F$

and finally we have

$$\begin{aligned}
\|\mathrm{vec}(\nabla_Y \mathcal{F}(\bar{U}, Y))\|_2 &= \|G_\omega(S_\omega^{-1} J_\omega - I_{nk})v\|_2 \\
&\leq w_+(1+\delta_{2k})(\sigma_1\|\bar{X} - \bar{U}\|_F^2 + \frac{1}{(1-\delta_{2k})w_-}\|S_\omega K - J_\omega\|_2\|V\|_F) \\
&\leq w_+(1+\delta_{2k})(\sigma_1\|\bar{X} - \bar{U}\|_F^2 + \frac{\sigma_1\sqrt{k}}{(1-\delta_{2k})w_-}\|S_\omega K - J_\omega\|_2) \\
&\leq w_+(1+\delta_{2k})\sigma_1(\|\bar{X} - \bar{U}\|_F^2 + \frac{\sqrt{k}\cdot 6w_+\delta_{2k}\sqrt{2k}}{(1-\delta_{2k})w_-}\|\bar{X} - \bar{U}\|_F) \\
&\leq w_+(1+\delta_{2k})\sigma_1(\|\bar{X} - \bar{U}\|_F^2 + \frac{6\sqrt{2}\cdot w_+\delta_{2k}k}{(1-\delta_{2k})w_-}\|\bar{X} - \bar{U}\|_F).
\end{aligned}$$

Under the given condition, we can upper bound $\|\bar{X} - \bar{U}\|$ and δ_{2k} and we go to the final step as follows:

$$\begin{aligned}
\|\mathrm{vec}(\nabla_Y \mathcal{F}(\bar{U}, Y))\|_2 &\leq \frac{(1-\delta_{2k})\sigma_k w_-}{2\xi} + \frac{(1-\delta_{2k})\sigma_k w_-}{2\xi} \\
&= \frac{(1-\delta_{2k})\sigma_k w_-}{\xi}
\end{aligned}$$

Thus, the divergence $\mathcal{D}(Y, Y, \bar{X})$ can be upperbounded by

$$\mathcal{D}(Y, Y, \bar{X}) \leq \|\mathrm{vec}(\nabla_Y \mathcal{F}(\bar{U}, Y))\|_2 \leq \frac{(1-\delta_{2k})\sigma_k w_-}{\xi}\|\bar{X}^{(t)} - \bar{U}^{(t)}\|_F. \tag{8}$$

□

Lemma 8.

$$\|Y^{(t+0.5)} - V^{(t)}\|_F \leq \frac{1}{2w_-(1-\delta_{2k})}\mathcal{D}(Y^{(t+0.5)}, Y^{(t+0.5)}, \bar{X}^{(t)}). \tag{9}$$

Proof of Lemma 8. Here we utilize the strongly convexity of $\mathcal{F}(X, Y)$ given a orthonomal matrix X. By Lemma 4, we have

$$\mathcal{F}(\bar{U}, V) \geq \mathcal{F}(\bar{U}, Y) + \langle \nabla_Y \mathcal{F}(\bar{U}, Y), V - Y\rangle + w_-(1-\delta_{2k})\|V - Y\|_F^2. \tag{10}$$

Since V minimizes the function $\mathcal{F}(\bar{U}, \hat{V})$, we have $\langle \nabla_Y \mathcal{F}(\bar{U}, V), Y - V\rangle \geq 0$ and thus

$$\begin{aligned}
\mathcal{F}(\bar{U}, Y) &\geq \mathcal{F}(\bar{U}, V) + \langle \nabla_Y \mathcal{F}(\bar{U}, V), Y - V\rangle + (1-\delta_{2k})w_-\|V - Y\|_F^2 \\
&\geq \mathcal{F}(\bar{U}, V) + w_-(1-\delta_{2k})\|V - Y\|_F^2.
\end{aligned} \tag{11}$$

Add (10) and (11) we have

$$\langle \nabla_Y \mathcal{F}(\bar{U}, Y), Y - V \rangle \geq 2w_-(1-\delta_{2k})\|V - Y\|_F^2. \tag{12}$$

Since Y also minimizes $\mathcal{F}(\bar{X}, \hat{Y})$, we have $\langle \nabla_Y \mathcal{F}(\bar{X}, V), V - Y \rangle \geq 0$ and thus

$$\begin{aligned} \langle \nabla_Y \mathcal{F}(\bar{U}, Y) - \nabla_Y \mathcal{F}(\bar{X}, Y), Y - V \rangle &\geq \langle \nabla_Y \mathcal{F}(\bar{U}, Y), Y - V \rangle \\ &\geq 2w_-(1-\delta_{2k})\|V - Y\|_F^2. \end{aligned} \tag{13}$$

Therefore, we have

$$\|V - Y\|_F \leq \frac{1}{2w_-(1-\delta_{2k})} \mathcal{D}(Y, Y, \bar{X}) \tag{14}$$

□

Given Lemma 7 and Lemma 8, we can now bound $\|Y^{(t+0.5)} - V^{(t)}\|_F$ and thus prove Lemma 5.

Proof of Lemma 5. From Lemma 7, we have

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

and from Lemma 8, we have

$$\|Y^{(t+0.5)} - V^{(t)}\|_F \leq \frac{1}{2w_-(1-\delta_{2k})} \mathcal{D}(Y^{(t+0.5)}, Y^{(t+0.5)}, \hat{X}^{(t)}).$$

Therefore,

$$\begin{aligned} & \|Y^{(t+0.5)} - V^{(t)}\|_F && (15) \\ \leq \; & \frac{(1-\delta_{2k})\sigma_k w_-}{\xi} \cdot \frac{1}{2w_-(1-\delta_{2k})} \|\bar{X}^{(t)} - \bar{U}^{(t)}\|_F && (16) \\ = \; & \frac{\sigma_k}{2\xi} \|\bar{X}^{(t)} - \bar{U}^{(t)}\|_F && (17) \end{aligned}$$

□

B.4 Proof of Lemma 6

From Lemma 5, we have

$$\begin{aligned} \|Y^{(0.5)} - V^{(t)}\|_F &\leq \frac{\sigma_k}{2\xi}\|\bar{X}^{(t)} - \bar{U}_F^{(t)}\| && (18)\\ &\leq \frac{(1-\delta_{2k})\sigma_k w_-}{2\xi^2(1+\delta_{2k})\sigma_1 w_+} \leq \frac{\sigma_k}{4}, && (19)\end{aligned}$$

where (19) is from $\xi > 1$. Thus, we can see from Lemma3 and we obtain that

$$\|\bar{Y}^{(t+1)} - \bar{V}^{(t+1)}\|_F \leq \frac{2}{\sigma_k}\|Y^{(0.5)} - V^{(t)}\|_F \leq \frac{1}{\xi}\|\bar{X}^{(t)} - \bar{U}^{(t)}\| \leq \frac{(1-\delta_{2k})\sigma_k w_-}{2\xi(1+\delta_{2k})\sigma_1 w_+}. \quad (20)$$

B.5 Proof of Theorem 1

Lemma 9. *Suppose that δ_{2k} satisfies*

$$\delta_{2k} \leq \frac{w_-^2(1-\delta_{2k})^2\sigma_k^4}{48\xi^2 k w_+^2(1+\delta_{2k})^2\sigma_1^4}.$$

Then there exists a factorization of $M^* = \bar{U}^0 V^{(0)\mathsf{T}}$ such that $\bar{U}^{(0)} \in \mathbb{R}^{m\times k}$ is an orthonomal matrix, and satisfies

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

Proof of Lemma 9. The initialization step can be regarded as taking a step iterate of singular value projection (SVP) as taking $M^{(t)} = 0$ and the next iterate with the step size $1/(1+\delta_{2k})$ will result $M^{(t+1)} = \bar{X}^{(0)}D^{(0)}\bar{Y}^{(0)}/(1+\delta_{2k})$, where $\bar{X}^{(0)}$,$D^{(0)}$ and $\bar{Y}^{(0)}$ are from the top k singular value decomposition of $\sum_{i=1}^p b_i A_i$.

Then, by Lemma 2 and the fact that $\varepsilon = 0$, we have

$$\left\|\mathcal{A}(\frac{\bar{X}^{(0)}D^{(0)}\bar{Y}^{(0)}}{(1+\delta_{2k})}) - \mathcal{A}(M^*)\right\|_2^2 \leq 4\delta_{2k}\|0 - \mathcal{A}(M^*)\|_2^2. \quad (21)$$

From the $2k$-RIP condition, we have

$$\begin{aligned}
\left\| \frac{\bar{X}^{(0)} D^{(0)} \bar{Y}^{(0)}}{(1+\delta_{2k})} \right\| &\leq \frac{1}{1-\delta_{2k}} \left\| \mathcal{A}\left(\frac{\bar{X}^{(0)} D^{(0)} \bar{Y}^{(0)}}{(1+\delta_{2k})}\right) - \mathcal{A}(M^*) \right\|_2^2 \\
&\leq \frac{4\delta_{2k}}{1-\delta_{2k}} \|\mathcal{A}(M^*)\|_2^2 \\
&\leq \frac{4\delta_{2k}(1+\delta_{2k})}{1-\delta_{2k}} \|M^*\|_F^2 \leq 6\delta_{2k}\|M^*\|_F^2.
\end{aligned}$$

Then, we project each column of M^* into the column subspace of $\bar{X}^{(0)}$ and obtain

$$\|(\bar{X}^{(0)}\bar{X}^{(0)\top} - I)M^*\|_F^2 \leq 6\delta_{2k}\|M^*\|_F^2.$$

We denote the orthonomal complement of $\bar{X}^{(0)}$ as $\bar{X}_{\perp}^{(0)}$. Then, we have

$$\frac{6\delta_{2k} k \sigma_1^2}{\sigma_k^2} \geq \|\bar{X}_{\perp}^{(0)\top} \bar{U}^*\|_F^2,$$

where $\bar{U}^*$ is from the singular value decomposition of $M^* = \bar{U} D \bar{V}^{\top}$. Then, there exists a unitary matrix $O \in \mathbb{R}^{k \times k}$ such that $O^\top O = I_k$ and

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

By taking the condition of δ_{2k}, we have

$$\|\bar{X}^0 - \bar{U}^*\|_F \leq \frac{(1-\delta_{2k})\sigma_k w_-}{2\xi(1+\delta_{2k})\sigma_1 w_+}. \tag{22}$$

□

Proof of Theorem 1. The proof of Theore 1 can be done by induction. Firstly, we note that Lemma 9 ensures that the initial $\bar{X}^{(0)}$ is close to a $\bar{U}^{(0)}$. Then, by Lemma 3 we have the following sequence of inequalities for all T iterations:

$$\|\bar{Y}^{(T)} - \bar{V}^{(T)}\|_F \leq \frac{1}{\xi}\|\bar{X}^{(T-1)} - \bar{U}^{(T-1)}\|_F \leq \cdots \leq \frac{1}{\xi^{2T-1}}\|\bar{X}^{(0)} - \bar{U}^{(0)}\|_F \leq \frac{(1-\delta_{2k})\sigma_k w_-}{2\xi^{2T}(1+\delta_{2k})\sigma_1 w_+}. \tag{23}$$

Therefore, we can bound the right most term by $\varepsilon/2$ for any given precision ε. By algebra, we can derive the required number of iterations T as:

$$T \geq \frac{1}{2}\log\left(\frac{(1-\delta_{2k})\sigma_k w_-}{2\varepsilon(1+\delta_{2k})\sigma_1 w_+}\right)\log^{-1}\xi.$$

Similarly, we can also bound $\|X^{(T-0.5)} - U^{(T)}\|_F$,

$$\|X^{(T-0.5)} - U^{(T)}\|_F \leq \frac{\sigma_k}{2\xi}\|\bar{Y}^{(T)} - \bar{V}^{(T)}\|_F \leq \frac{(1-\delta_{2k})\sigma_k^2 w_-}{4\xi(1+\delta_{2k})\sigma_1 w_+}. \tag{24}$$

To make it smaller than $\varepsilon\sigma_1/2$, we need the number of iterations as

$$T \geq \frac{1}{2}\log\left(\frac{(1-\delta_{2k})\sigma_k^2 w_-}{4\varepsilon(1+\delta_{2k})\sigma_1 w_+}\right)\log^{-1}\xi.$$

Combining all results we have

$$\begin{aligned}
\|M^{(T)} - M^*\|_F &= \|X^{(T-0.5)}\bar{Y}^{(T)\mathsf{T}} - U^{(T)}\bar{V}^{(T)\mathsf{T}}\|_F \\
&= \|X^{(T-0.5)}\bar{Y}^{(T)\mathsf{T}} - U^{(T)}\bar{Y}^{(T)\mathsf{T}} + U^{(T)}\bar{Y}^{(T)\mathsf{T}} - U^{(T)}\bar{V}^{(T)\mathsf{T}}\|_F \\
&\leq \|\bar{Y}^{(T)\mathsf{T}}\|_2\|X^{(T-0.5)} - U^{(T)}\|_F + \|U^{(T)}\|_2\|\bar{Y}^{(T)} - \bar{V}^{(T)}\|_F \leq \varepsilon, \qquad (25)
\end{aligned}$$

Here we use the fact that the orthonomal matrix $\bar{V}^{(T)}$ leads to $\|\bar{V}^{(T)}\|_2 = 1$, and $\|M^*\|_2 = \|U^{(T)}\bar{V}^{(T)\mathsf{T}}\|_2 = \|U^{(T)}\|_2 = \sigma_1$. Now we complete the proof of Theorem 1. $\square$