Supplementary Materials for the paper "Preconditioning Matters: Fast Global Convergence of Non-convex Matrix Factorization via Scaled Gradient Descent"

We first rewrite the main update formulation of both ScaledGD and AltScaledGD

$$\begin{cases} \boldsymbol{U}_{k+1} = \boldsymbol{U}_k - \eta \nabla_{\boldsymbol{U}_k} f(\boldsymbol{U}_k, \boldsymbol{V}_k) \left(\boldsymbol{V}_k^\top \boldsymbol{V}_k\right)^{-1} \\ \boldsymbol{V}_{k+1} = \boldsymbol{V}_k - \eta \nabla_{\boldsymbol{V}_k} f(\boldsymbol{U}_k, \boldsymbol{V}_k) \left(\boldsymbol{U}_k^\top \boldsymbol{U}_k\right)^{-1} \end{cases} \tag{1}$$

$$\begin{cases} \boldsymbol{U}_{k+1} = \boldsymbol{U}_k - \eta \nabla_{\boldsymbol{U}_k} f(\boldsymbol{U}_k, \boldsymbol{V}_k)(\boldsymbol{V}_k^\top \boldsymbol{V}_k)^{-1} \\ \boldsymbol{V}_{k+1} = \boldsymbol{V}_k - \eta \nabla_{\boldsymbol{V}_k} f(\boldsymbol{U}_{k+1}, \boldsymbol{V}_k)(\boldsymbol{U}_{k+1}^\top \boldsymbol{U}_{k+1})^{-1} \end{cases} \tag{2}$$

where $\boldsymbol{U}_k \in \mathbb{R}^{m \times d}$ and $\boldsymbol{V}_k \in \mathbb{R}^{n \times d}$.

1 Proofs of the lemmas and theorem in Section 4 (rank-1 case).

If $d = 1$, then $\boldsymbol{U}_k, \boldsymbol{V}_k$ are all vectors.

Lemma 1. *If $\left\langle \boldsymbol{M}, \boldsymbol{U}_k \boldsymbol{V}_k^\top \right\rangle \geq \|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F^2 \neq 0$, then there is constant $C_u \geq 0$ such that*

$$\left| 1 - \frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F} \right| \leq C_u \left(\|\boldsymbol{M}\|_F - \|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F \right) \tag{3}$$

Proof. Since $\left\langle \boldsymbol{M}, \boldsymbol{U}_k \boldsymbol{V}_k^\top \right\rangle \geq \|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F^2$, it follows directly that

$$\frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F} \geq 1. \tag{4}$$

In consequence,

$$\begin{aligned} \left| 1 - \frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F} \right| &= \frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F} - 1 \\ &\leq \frac{\|\boldsymbol{M}\|_F - \|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F}. \end{aligned} \tag{5}$$

According to Lemma 4 and together with $\frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F} \geq 1$, it is thus suffice to prove that $\|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F$ is monotonically increasing, meanwhile as $\|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|$ is bounded below, which guarantees that there exists constant C_u such that the result in Eq. (3) is true. □

Lemma 2. (Convergence of the distance between subspaces) *For the ScaledGD (1) with $d = 1$, if $\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v \geq \|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F$, then the following holds*

$$\|\mathcal{U}_{*\perp}^\top \mathcal{U}_{k+1}\|_2 \leq (1 - \eta) \|\mathcal{U}_{*\perp}^\top \mathcal{U}_k\|_2, \quad \|\mathcal{V}_{*\perp}^\top \mathcal{V}_{k+1}\|_2 \leq (1 - \eta) \|\mathcal{V}_{*\perp}^\top \mathcal{V}_k\|_2 \tag{6}$$

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

Proof. Since $\boldsymbol{U}_{k+1} \in \mathbb{R}^{m\times 1}$, we have

$$\|\mathcal{U}_{*\perp}^\top \mathcal{U}_{k+1}\|_2 = \frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+1}\|_2}{\|\boldsymbol{U}_{k+1}\|_2}. \tag{7}$$

Meanwhile, $\boldsymbol{U}_{k+1} = (1-\eta)\boldsymbol{U}_k + \eta \boldsymbol{M}\boldsymbol{V}_k(\boldsymbol{V}_k^\top \boldsymbol{V}_k)^{-1}$, thus

$$\begin{aligned}
\|\mathcal{U}_{*\perp}^\top \mathcal{U}_{k+1}\|_2 &= (1-\eta)\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_{k+1}\|_F} \\
&= (1-\eta)\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_2}{\|\boldsymbol{U}_k\|_2}\frac{\|\boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_{k+1}\|_F} \\
&= (1-\eta)\|\mathcal{U}_{*\perp}^\top \mathcal{U}_k\|_2 \frac{\|\boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_{k+1}\|_F}.
\end{aligned} \tag{8}$$

According to Lemma 4 and the condition that $\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$, then

$$\frac{\|\boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_{k+1}\|_F} \leq 1 \text{ and } \frac{\|\boldsymbol{V}_k\|_F}{\|\boldsymbol{V}_{k+1}\|_F} \leq 1, \tag{9}$$

it is thus suffice to guarantee that

$$\|\mathcal{U}_{*\perp}^\top \mathcal{U}_{k+1}\|_2 \leq (1-\eta)\|\mathcal{U}_{*\perp}^\top \mathcal{U}_k\|_2. \tag{10}$$

Similarly, we can guarantee

$$\|\mathcal{V}_{*\perp}^\top \mathcal{V}_{k+1}\|_2 \leq (1-\eta)\|\mathcal{V}_{*\perp}^\top \mathcal{V}_k\|_2. \tag{11}$$

Thus, we finish the proof. □

Theorem 5. (Convergence of the matrix norm) *For the ScaledGD (1) with $d = 1$, if $\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$ for all $k \geq 0$, then we have*

$$\|\boldsymbol{M}\|_F - \|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F \leq (1-\eta)^{2k} k C_\alpha \tag{12}$$

where C_α is a constant and η is the step length $0 \leq \eta < 1$.

Proof. By ScaledGD Eq. (1), we have

$$\|\boldsymbol{U}_{k+1}\|_2^2 = (1-\eta)^2\|\boldsymbol{U}_k\|_2^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v \|\boldsymbol{U}_k\|_F}{\|\boldsymbol{V}_k\|_F} + \eta^2 \frac{\|\boldsymbol{M}\|_F^2 \cos^2\boldsymbol{\theta}_k^v}{\|\boldsymbol{V}_k\|_F^2} \tag{13}$$

and

$$\|\boldsymbol{V}_{k+1}\|_2^2 = (1-\eta)^2\|\boldsymbol{V}_k\|_2^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v \|\boldsymbol{V}_k\|_F}{\|\boldsymbol{U}_k\|_F} + \eta^2 \frac{\|\boldsymbol{M}\|_F^2 \cos^2\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k\|_F^2} \tag{14}$$

which implies

$$\|\boldsymbol{U}_{k+1}\|_F \geq \left(1-\eta+\eta\frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\right)\|\boldsymbol{U}_k\|_F \tag{15}$$

and

$$\|\boldsymbol{V}_{k+1}\|_F \geq \left(1-\eta+\eta\frac{\|\boldsymbol{M}\|_F \cos\boldsymbol{\theta}_k^u \cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\right)\|\boldsymbol{V}_k\|_F \tag{16}$$

In consequence, one obtains

$$\begin{aligned}
\frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F}{\|\boldsymbol{M}\|_F} &\geq (1-\eta)^2\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{M}\|_F} + \eta^2\frac{\|\boldsymbol{M}\|_F\cos^2\boldsymbol{\theta}_k^u\cos^2\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} + 2\eta(1-\eta)\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \\
&\geq (1-\eta)^2\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{M}\|_F} + \eta^2\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k + 2\eta(1-\eta)\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k \\
&= (1-\eta)^2\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{M}\|_F} + \left(2\eta-\eta^2\right)\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k
\end{aligned} \tag{17}$$

then it follows

$$1-\frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F}{\|\boldsymbol{M}\|_F}\leq(1-\eta)^2\left(1-\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{M}\|_F}\right)+(2\eta-\eta^2)(1-\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k), \tag{18}$$

which implies

$$\|\boldsymbol{M}\|_F-\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F\leq(1-\eta)^2\left(\|\boldsymbol{M}\|_F-\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F\right)+(2\eta-\eta^2)\|\boldsymbol{M}\|_F(1-\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k). \tag{19}$$

Meanwhile, according to Lemma 2, we have $\sin\boldsymbol{\theta}_u^{k+1}\leq(1-\eta)\sin\boldsymbol{\theta}_u^k$ and $\sin\boldsymbol{\theta}_v^{k+1}\leq(1-\eta)\sin\boldsymbol{\theta}_v^k$, thus it is suffice to prove that

$$(1-\cos\boldsymbol{\theta}_u^{k+1}\cos\boldsymbol{\theta}_v^{k+1})\leq(1-\eta)^2(1-\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k)\leq(1-\eta)^{2k+1}(1-\cos\boldsymbol{\theta}_u^0\cos\boldsymbol{\theta}_v^0). \tag{20}$$

Together with Eq. (19), the following inequality holds

$$\begin{aligned}\|\boldsymbol{M}\|_F-\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F&\leq(2\eta-\eta^2)\|\boldsymbol{M}\|_F\sum_{i=1}^k(1-\eta)^{2i}(1-\cos\boldsymbol{\theta}_u^{k-i}\cos\boldsymbol{\theta}_v^{k-i})\\&+(1-\eta)^{2k}\left(\|\boldsymbol{M}\|_F-\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F\right)\\&\leq(2\eta-\eta^2)\|\boldsymbol{M}\|_F\sum_{i=1}^k(1-\eta)^{2i}(1-\eta)^{2(k-i)}(1-\cos\boldsymbol{\theta}_u^0\cos\boldsymbol{\theta}_v^0)\\&+(1-\eta)^{2k}\left(\|\boldsymbol{M}\|_F-\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F\right)\\&\leq(1-\eta)^{2k}kC_\alpha,\end{aligned} \tag{21}$$

where $C_\alpha=\max\{(1-\cos\boldsymbol{\theta}_u^0\cos\boldsymbol{\theta}_v^0)(2\eta-\eta^2)\|\boldsymbol{M}\|_F,\frac{\|\boldsymbol{M}\|_F-\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}{k}\}$, which finishes our proof. ☐

Lemma 3. *Let $\eta\leq c_\eta<1$ with c_η a small constant, if $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_0^u\cos\boldsymbol{\theta}_0^v\geq\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F$ then the following is true*

$$\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v\geq\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F,\forall k>0. \tag{22}$$

Proof. We prove this result by induction. Since $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_0^u\cos\boldsymbol{\theta}_0^v\geq\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F$, we assume that $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u\cos\boldsymbol{\theta}_{k-1}^v\geq\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F$, then we need to prove $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v\geq\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$. We first show that

$$\begin{aligned}\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2&=(1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F^2+2\eta(1-\eta)\frac{\|\boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_{k-1}\|_F}\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u\cos\boldsymbol{\theta}_{k-1}^v\\&+\eta^2\frac{\|\boldsymbol{U}_k\|_F^2}{\|\boldsymbol{U}_{k-1}\|_F^2}\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k-1}^u\\&=\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F^2\left((1-\eta)+\eta\frac{\|\boldsymbol{M}\|_F}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F}\cos\boldsymbol{\theta}_{k-1}^u\right)^2\\&=\|\boldsymbol{M}\|_F^2\underbrace{\frac{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F^2}{\|\boldsymbol{M}\|_F^2}\left((1-\eta)+\eta\frac{\|\boldsymbol{M}\|_F}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F}\cos\boldsymbol{\theta}_{k-1}^u\right)^2}_{\mathfrak{M}^2}\end{aligned} \tag{23}$$

To prove the result in Eq. (130), we need to guarantee that $\mathfrak{M}\leq\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v$, which is equivalent to ensure that

$$(1-\eta)\frac{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F}{\|\boldsymbol{M}\|_F}+\eta\frac{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F}\cos\boldsymbol{\theta}_{k-1}^u\leq\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \tag{24}$$

Meanwhile, since $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u\cos\boldsymbol{\theta}_{k-1}^v\geq\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F$, according to Lemma 2 it is easy to verify

$$\cos^2\boldsymbol{\theta}_k^u\geq(1-\eta)^2\cos^2\boldsymbol{\theta}_{k-1}^u+2\eta-\eta^2\geq(1-\eta^2)\cos^2\boldsymbol{\theta}_{k-1}^u+\eta^2, \tag{25}$$

together with

$$\|\boldsymbol{U}_k\|_F^2 = (1-\eta)^2\|\boldsymbol{U}_{k-1}\|_F^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{U}_{k-1}\|_F}{\|\boldsymbol{V}_{k-1}\|_F}\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u\cos\boldsymbol{\theta}_{k-1}^v + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k-1}^v}{\|\boldsymbol{V}_{k-1}\|_F^2}$$
$$\le (1-\eta)^2\|\boldsymbol{U}_{k-1}\|_F^2 + (2\eta(1-\eta)\cos^2\boldsymbol{\theta}_{k-1}^u + \eta^2)\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k-1}^v}{\|\boldsymbol{V}_{k-1}\|_F^2} \tag{26}$$

and

$$\begin{aligned}\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F^2 &\le (1-\eta)^2\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F^2 + (2\eta(1-\eta)\cos^2\boldsymbol{\theta}_{k-1}^u + \eta^2)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k-1}^v \\ &\le (1-\eta)^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k-1}^u\cos^2\boldsymbol{\theta}_{k-1}^v + (2\eta(1-\eta)\cos^2\boldsymbol{\theta}_{k-1}^u + \eta^2)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k-1}^v \\ &= \left((1-\eta)^2\cos^2\boldsymbol{\theta}_{k-1}^u + (2\eta(1-\eta)\cos^2\boldsymbol{\theta}_{k-1}^u + \eta^2)\right)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k-1}^v \\ &= \left((1-\eta^2)\cos^2\boldsymbol{\theta}_{k-1}^u + \eta^2)\right)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k-1}^v\end{aligned}. \tag{27}$$

We obtain

$$\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_{k-1}^v \ge \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F. \tag{28}$$

By the condition $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u\cos\boldsymbol{\theta}_{k-1}^v \ge \|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F$ and according to Lemma 2, we have

$$\cos\boldsymbol{\theta}_k^v \ge \cos\boldsymbol{\theta}_{k-1}^v, \tag{29}$$

and thus

$$\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \ge \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F, \tag{30}$$

with equality holds if and only if $\cos\boldsymbol{\theta}_k^v = \cos\boldsymbol{\theta}_{k-1}^v$. Denote by

$$\mathfrak{z} = \frac{\frac{\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}\|_F} - 1}{\frac{\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F} - 1}, \tag{31}$$

then it can be easily verified that $\mathfrak{z} > 0$. As a result

$$\mathfrak{z}\left(\frac{\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F} - 1\right)\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}\|_F = \|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v - \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}\|_F, \tag{32}$$

which is equivalent to

$$\mathfrak{z}\left(\frac{\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F} - 1\right)\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}\|_F = \|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v - \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}\|_F, \tag{33}$$

and thus

$$\mathfrak{z}\left(\frac{\|\boldsymbol{U}_k\|_F\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k-1}^u}{\|\boldsymbol{U}_k\|_F} - \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F\right) \le \|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v - \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}\|_F. \tag{34}$$

If $\eta \le c_\eta < \mathfrak{z}$ with c_η sufficiently small, then we can guarantee that

$$\eta\left(\frac{\|\boldsymbol{U}_{k+1}\|_F\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u}{\|\boldsymbol{U}_k\|_F} - \|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F\right) \le \|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k+1}^u\cos\boldsymbol{\theta}_{k+1}^v - \|\boldsymbol{U}_{k+1}\boldsymbol{V}_k\|_F \tag{35}$$

By simple reformulation, one obtains

$$(1-\eta)\frac{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F}{\|\boldsymbol{M}\|_F} + \eta\frac{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F}\cos\boldsymbol{\theta}_{k-1}^u \le \cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \tag{36}$$

which is exactly the inequality we need in Eq. (132), thus we finish the proof. □

Lemma 4. *If the condition $\|\boldsymbol{M}\|_F\max\{\cos\boldsymbol{\theta}_k^u, \cos\boldsymbol{\theta}_k^v\} < \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$ is satisfied then we have*

$$\|\boldsymbol{U}_{k+1}\|_F < \|\boldsymbol{U}_k\|_F \quad \text{and} \quad \|\boldsymbol{V}_{k+1}\|_F < \|\boldsymbol{V}_k\|_F. \tag{37}$$

Furthermore, if the condition $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \ge \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$ is satisfied then we have

$$\|\boldsymbol{U}_{k+1}\|_F \ge \|\boldsymbol{U}_k\|_F \quad \text{and} \quad \|\boldsymbol{V}_{k+1}\|_F \ge \|\boldsymbol{V}_k\|_F. \tag{38}$$

Proof. By Eq. (1), we have

$$\|\boldsymbol{U}_{k+1}\|_F^2 = (1-\eta)^2\|\boldsymbol{U}_k\|_F^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{U}_k\|_F}{\|\boldsymbol{V}_k\|_F}\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_v^k}{\|\boldsymbol{V}_k\|_F^2}. \quad (39)$$

If $\|\boldsymbol{M}\|_F \max\{\cos\boldsymbol{\theta}_k^u, \cos\boldsymbol{\theta}_k^v\} < \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$, then it is easy to verify that

$$\|\boldsymbol{U}_{k+1}\|_F^2 \le \|\boldsymbol{U}_k\|_F^2. \quad (40)$$

Similarly, we have

$$\|\boldsymbol{V}_{k+1}\|_F^2 \le \|\boldsymbol{V}_k\|_F^2, \quad (41)$$

if the condition $\|\boldsymbol{M}\|_F \max\{\cos\boldsymbol{\theta}_k^u, \cos\boldsymbol{\theta}_k^v\} < \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$ is satisfied.

By contrast, if $\|\boldsymbol{M}\|_F\{\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v\} \ge \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$, then we can obtain from Eq. (39) that

$$\|\boldsymbol{U}_{k+1}\|_F^2 \ge \|\boldsymbol{U}_k\|_F^2, \quad (42)$$

and similarly,

$$\|\boldsymbol{V}_{k+1}\|_F^2 \ge \|\boldsymbol{V}_k\|_F^2, \quad (43)$$

which finishes our proof. □

Lemma 5. *After T_1 iterations of ScaledGD Eq.* (1) *with $d = 1$, the following inequalities hold $\forall k \ge T_1$*

$$\|\mathcal{U}_{*\perp}^\top\mathcal{U}_{k+1}\|_2 \le (1-\chi_k)\|\mathcal{U}_{*\perp}^\top\mathcal{U}_k\|_2 \quad (44)$$

$$\|\mathcal{V}_{*\perp}^\top\mathcal{V}_{k+1}\|_2 \le (1-\chi_k)\|\mathcal{V}_{*\perp}^\top\mathcal{V}_k\|_2 \quad (45)$$

$$1 - \cos\boldsymbol{\theta}_{k+1}^u\cos\boldsymbol{\theta}_{k+1}^v \le (1-\chi_k)^2(1-\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v) \quad (46)$$

where $\chi_k = \frac{\eta\tau_k}{1-\eta(1-\tau_k)} < 1$ and $\tau_k = \frac{\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \in [1/2, 1]$.

Proof. According to the analysis in Section 4.1.2 of the main paper, we know that after T_1 iterations, we have

$$\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F \le \|\boldsymbol{M}\|_F,\ \forall k \ge T_1, \quad (47)$$

which implies

$$\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F \le 2\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^u \quad (48)$$

Since the results for $\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F \le \|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^u$ is easy to analyze according to Lemma 2, therefore we mainly focus on

$$\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^u \le \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F \le 2\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^u. \quad (49)$$

By Eq. (1), we have

$$\|\boldsymbol{U}_{k+1}\|_F^2 = (1-\eta)^2\|\boldsymbol{U}_k\|_F^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{U}_k\|_F}{\|\boldsymbol{V}_k\|_F}\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_v^k}{\|\boldsymbol{V}_k\|_F^2}. \quad (50)$$

and consequently

$$\frac{\|\boldsymbol{U}_k\|_F^2}{\|\boldsymbol{U}_{k+1}\|_F^2} = \frac{1}{(1-\eta)^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^u}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2}} \quad (51)$$

thus

$$\frac{\|\boldsymbol{U}_k\|}{\|\boldsymbol{U}_{k+1}\|} \le \frac{1}{\sqrt{(1-\eta)^2 + 2\eta(1-\eta)\tau_k + \eta^2\tau_k^2}} = \frac{1}{1-\eta+\eta\tau_k}. \quad (52)$$

It is easy to verify $\tau_k = \frac{\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \in [1/2, 1], \forall k \ge T_1$ (according to Eq. (49)), then

$$\begin{aligned}\|\mathcal{U}_{*\perp}^\top\mathcal{U}_{k+1}\|_2 &= (1-\eta)\|\mathcal{U}_{*\perp}^\top\mathcal{U}_k\|_2\frac{\|\boldsymbol{U}_k\|_2}{\|\boldsymbol{U}_{k+1}\|_2} \\ &\le \frac{1-\eta}{1-\eta+\eta\tau_k}\|\mathcal{U}_{*\perp}^\top\mathcal{U}_k\|_2 \\ &= (1-\chi_k)\|\mathcal{U}_{*\perp}^\top\mathcal{U}_k\|_2\end{aligned}, \quad (53)$$

and $\chi_k = \frac{\eta\tau_k}{1-\eta(1-\tau_k)} < 1$. Similarly, we can guarantee that

$$\begin{aligned}\|\mathcal{V}_{*\perp}^{\top}\mathcal{V}_{k+1}\|_2 &= (1-\eta)\|\mathcal{V}_{*\perp}^{\top}\mathcal{V}_k\|_2\frac{\|\boldsymbol{V}_k\|_2}{\|\boldsymbol{V}_{k+1}\|_2} \\ &\leq \frac{1-\eta}{1-\eta+\eta\tau_k}\|\mathcal{V}_{*\perp}^{\top}\mathcal{V}_k\|_2 \\ &= (1-\chi_k)\|\mathcal{V}_{*\perp}^{\top}\mathcal{V}_k\|_2\end{aligned}. \tag{54}$$

The above inequalities Eq. (53) and Eq. (54) further imply that

$$\cos^2\boldsymbol{\theta}_{k+1}^u \geq (1-\chi_k)^2\cos^2\boldsymbol{\theta}_k^u + 2\chi_k - \chi_k^2, \tag{55}$$

$$\cos^2\boldsymbol{\theta}_{k+1}^v \geq (1-\chi_k)^2\cos^2\boldsymbol{\theta}_k^v + 2\chi_k - \chi_k^2, \tag{56}$$

thus

$$\begin{aligned}\cos^2\boldsymbol{\theta}_{k+1}^u\cos^2\boldsymbol{\theta}_{k+1}^v &\geq (1-\chi_k)^4\cos^2\boldsymbol{\theta}_k^u\cos^2\boldsymbol{\theta}_k^v + (2\chi_k-\chi_k^2)(1-\chi_k)^2\left(\cos^2\boldsymbol{\theta}_k^u + \cos^2\boldsymbol{\theta}_k^v\right) \\ &\quad + (2\chi_k-\chi_k^2)^2 \\ &\geq \left((1-\chi_k)^2\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v + 2\chi_k - \chi_k^2\right)^2\end{aligned}, \tag{57}$$

which is suffice to guarantee that

$$1-\cos\boldsymbol{\theta}_{k+1}^u\cos\boldsymbol{\theta}_{k+1}^v \leq (1-\chi_k)^2\left(1-\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v\right). \tag{58}$$

Therefore we have finish our proof. □

Lemma 6. (Convergence of the distance between subspaces) *For AltScaledGD (2), if* $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^{\top}\|_F$ *and* $0<\eta\leq 1$*, then the following holds*

$$\|\mathcal{U}_{*\perp}^{\top}\mathcal{U}_{k+1}\|_2 \leq (1-\eta)\|\mathcal{U}_{*\perp}^{\top}\mathcal{U}_k\|_2,\ \ \|\mathcal{V}_{*\perp}^{\top}\mathcal{V}_{k+1}\|_2 \leq (1-\eta)\|\mathcal{V}_{*\perp}^{\top}\mathcal{V}_k\|_2. \tag{59}$$

Proof. The proof of Lemma 6 is similar to that of Lemma 2. It is straightforward that

$$\|\mathcal{U}_{*\perp}^{\top}\mathcal{U}_{k+1}\|_2 = (1-\eta)\|\mathcal{U}_{*\perp}^{\top}\mathcal{U}_k\|_2\frac{\|\boldsymbol{U}_k\|_2}{\|\boldsymbol{U}_{k+1}\|_2}. \tag{60}$$

and

$$\|\mathcal{V}_{*\perp}^{\top}\mathcal{V}_{k+1}\|_2 = (1-\eta)\|\mathcal{V}_{*\perp}^{\top}\mathcal{V}_k\|_2\frac{\|\boldsymbol{V}_k\|_2}{\|\boldsymbol{V}_{k+1}\|_2}. \tag{61}$$

By Eq. (1) and Eq. (2), we have

$$\|\boldsymbol{U}_{k+1}\|_F^2 = (1-\eta)^2\|\boldsymbol{U}_k\|_F^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{U}_k\|_F}{\|\boldsymbol{V}_k\|_F}\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_v^k}{\|\boldsymbol{V}_k\|_F^2}. \tag{62}$$

and

$$\|\boldsymbol{V}_{k+1}\|_F^2 = (1-\eta)^2\|\boldsymbol{V}_k\|_F^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{V}_k\|_F}{\|\boldsymbol{U}_{k+1}\|_F}\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_u^{k+1}\cos\boldsymbol{\theta}_v^k + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_u^{k+1}}{\|\boldsymbol{U}_{k+1}\|_F^2}. \tag{63}$$

respectively. If $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^{\top}\|_F$, then

$$\begin{aligned}\|\boldsymbol{U}_{k+1}\|_F^2 &= (1-\eta)^2\|\boldsymbol{U}_k\|_F^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{U}_k\|_F}{\|\boldsymbol{V}_k\|_F}\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_u^k\cos\boldsymbol{\theta}_v^k + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_v^k}{\|\boldsymbol{V}_k\|_F^2} \\ &\geq (1-\eta)^2\|\boldsymbol{U}_k\|_F^2 + 2\eta(1-\eta)\|\boldsymbol{U}_k\|_F^2 + \eta^2\|\boldsymbol{U}_k\|_F^2 \\ &\geq \|\boldsymbol{U}_k\|_F^2\end{aligned}. \tag{64}$$

thus

$$\|\mathcal{U}_{*\perp}^{\top}\mathcal{U}_{k+1}\|_2 \leq (1-\eta)\|\mathcal{U}_{*\perp}^{\top}\mathcal{U}_k\|_2 \tag{65}$$

which implies

$$\cos^2\boldsymbol{\theta}_{k+1}^u \geq (1-\eta^2)\cos^2\boldsymbol{\theta}_k^u + \eta^2, \tag{66}$$

Meanwhile, we have

$$\begin{aligned}\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 &= (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 + 2\eta(1-\eta)\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v + \eta^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^v \\ &\le (1-\eta)^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^u\cos^2\boldsymbol{\theta}_k^v + 2\eta(1-\eta)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^u\cos^2\boldsymbol{\theta}_k^v + \eta^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^v \\ &\le (1-\eta^2)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^u\cos^2\boldsymbol{\theta}_k^v + \eta^2\cos^2\boldsymbol{\theta}_k^v \\ &\le \|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u\cos^2\boldsymbol{\theta}_k^v\end{aligned} \tag{67}$$

Together with Eq. (63), the following result holds

$$\|\boldsymbol{V}_{k+1}\|_F^2 \ge \|\boldsymbol{V}_k\|_F^2. \tag{68}$$

According to Eq. (61), and Eq. (65), we have the results in Eq. (59), which finishes our proof.

□

Lemma 7. *For AltScaledGD (2), if* $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_0^u\cos\boldsymbol{\theta}_0^v \ge \|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F$ *and* $0 < \eta \le 1$*, then the following is true*

$$\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \ge \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F, \forall k > 0. \tag{69}$$

Proof. We prove the results by induction. Since $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_0^u\cos\boldsymbol{\theta}_0^v \ge \|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F$, we assume that $\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v \ge \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$. holds, we need to prove $\sigma_r(\boldsymbol{M})\cos\boldsymbol{\theta}_{k+1r}^u\cos\boldsymbol{\theta}_{k+1r}^v \ge \sigma_r(\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top)$. According to Lemma 2, we have

$$\cos^2\boldsymbol{\theta}_{k+1}^u \ge (1-\eta)^2\cos^2\boldsymbol{\theta}_{kr}^u + 2\eta - \eta^2 \tag{70}$$

Similarly, we have $\cos^2\boldsymbol{\theta}_{k+1}^v \ge (1-\eta)^2\cos^2\boldsymbol{\theta}_k^v + 2\eta - \eta^2$. Then

$$\cos^2\boldsymbol{\theta}_{k+1}^u\cos^2\boldsymbol{\theta}_{k+1}^v \ge (1-\eta)^2\cos^2\boldsymbol{\theta}_{k+1}^u\cos^2\boldsymbol{\theta}_k^v + (2\eta-\eta^2)\cos^2\boldsymbol{\theta}_{k+1}^u \tag{71}$$

Meanwhile, according to Equation (2), we have

$$\begin{aligned}\|\boldsymbol{U}_{k+1}\|^2 &= \|\boldsymbol{U}_k\|^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{U}_k\|}{\|\boldsymbol{V}_k\|}\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^v}{\|\boldsymbol{V}_k\|^2} + (\eta^2-2\eta)\|\boldsymbol{U}_k\|^2 \\ &\le (1-\eta)^2\|\boldsymbol{U}_k\|^2 + (2\eta(1-\eta)\cos^2\boldsymbol{\theta}_k^u + \eta^2)\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^v}{\|\boldsymbol{V}_k\|^2}\end{aligned} \tag{72}$$

Similarly, we have

$$\|\boldsymbol{V}_{k+1}\|^2 = \|\boldsymbol{V}_k\|^2 + 2\eta(1-\eta)\frac{\|\boldsymbol{V}_k\|}{\|\boldsymbol{U}_{k+1}\|}\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k+1}^u\cos\boldsymbol{\theta}_k^v + \eta^2\frac{\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u}{\|\boldsymbol{U}_{k+1}\|^2} + (\eta^2-2\eta)\|\boldsymbol{V}_k\|^2, \tag{73}$$

thus

$$\begin{aligned}&\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2 \\ &\le (1-\eta)^2\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 + 2\eta(1-\eta)\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F\|\boldsymbol{M}\|_F\cos\boldsymbol{\theta}_{k+1}^u\cos\boldsymbol{\theta}_k^v + \eta^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u.\end{aligned} \tag{74}$$

According to Equation (72), we have

$$\begin{aligned}\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 &\le (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 + (2\eta(1-\eta)\cos^2\boldsymbol{\theta}_k^u + \eta^2)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^v \\ &\le (1-\eta)^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^u\cos^2\boldsymbol{\theta}_k^v + (2\eta(1-\eta)\cos^2\boldsymbol{\theta}_k^u + \eta^2)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^v \\ &= \left((1-\eta^2)\cos^2\boldsymbol{\theta}_k^u + \eta^2)\right)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^v\end{aligned} \tag{75}$$

According to Equation (70), if $0 \le \eta \le 1$, we have

$$\cos^2\boldsymbol{\theta}_{k+1}^u \ge (1-\eta^2)\cos^2\boldsymbol{\theta}_k^u + \eta^2. \tag{76}$$

Thus

$$\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 \le \|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_k^v\cos^2\boldsymbol{\theta}_{k+1}^u, \tag{77}$$

in consequence we have the following result

$$
\begin{aligned}
&\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2 \\
&\le (1-\eta)^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u\cos^2\boldsymbol{\theta}_k^v + 2\eta(1-\eta)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u\cos^2\boldsymbol{\theta}_k^v + \eta^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u \\
&\le (1-\eta)^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u\cos^2\boldsymbol{\theta}_k^v + 2\eta(1-\eta)\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u + \eta^2\|\boldsymbol{M}\|_F^2\cos^2\boldsymbol{\theta}_{k+1}^u \\
&\le \cos^2\boldsymbol{\theta}_{k+1}^u\cos^2\boldsymbol{\theta}_{k+1}^v\|\boldsymbol{M}\|_F^2
\end{aligned} \tag{78}
$$

where the last inequality is due to Equation (71), then we finish the proof. □

2 Proofs of the theorems in Section 3 (rank-d case).

2.1 Proofs for the results of ScaledGD.

Theorem 1 (General random initialization)**.** *Let $\boldsymbol{U}_0 \in \mathbb{R}^{m\times d}$ and $\boldsymbol{V}_0 \in \mathbb{R}^{n\times d}$ be random Gaussian that follow $\mathcal{N}(0,\sigma)$ for $\sigma > c_{\text{init}}$ (c_{init} is a positive constant), and $\boldsymbol{U}_k$, $\boldsymbol{V}_k$ are updated by Eq.* (1)*. If $\eta \le c_\eta < 1$ for small constant c_η, we have that the objective function of the LRMF problem decreases linearly after $T_1 = O\left(\ln\frac{d}{\delta}\right)$ iterations, namely*

$$
\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F \le \alpha_1\left(1-\chi_{k+T_1}\right)^k\|\boldsymbol{M}\|_F \tag{79}
$$

where χ_{k+T_1} is monotonically increasing from $\frac{\eta^2}{(2-\eta)^2}$ to η, δ is a sufficiently small constant, α_1 is a constant.

Proof. Let c_{init} be constant such that $\sigma \ge c_{\text{init}}$ indicates $\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F^2 \ge \left\langle \boldsymbol{U}_0\boldsymbol{V}_0^\top, \boldsymbol{M}\right\rangle$. The proof follows the proof sketch presented in Section 4 of the main paper, specifically, we can upper-bound the objective by four terms as

$$
\begin{aligned}
\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top - \boldsymbol{M}\|_F \le{}& \underbrace{(1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F}_{①} + \underbrace{(1-\eta)\eta\frac{\|\boldsymbol{M}\mathcal{V}_{k\perp}\|_F}{\|\boldsymbol{M}\|_F}\cdot\|\boldsymbol{M}\|_F}_{②} \\
&+ \underbrace{\eta(1-\eta)\frac{\|\mathcal{U}_{k\perp}^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F}\cdot\|\boldsymbol{M}\|_F}_{③} + \underbrace{\eta^2\|\boldsymbol{M}\boldsymbol{V}_k(\boldsymbol{V}_k^\top\boldsymbol{V}_k)^{-1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M} - \boldsymbol{M}\|_F}_{④}
\end{aligned} \tag{80}
$$

According to Lemma 14, we know that

$$
\|\boldsymbol{M}\boldsymbol{V}(\boldsymbol{V}^\top\boldsymbol{V})^{-1}(\boldsymbol{U}^\top\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{M} - \boldsymbol{M}\|_F < C\|\boldsymbol{M}\|_F\left|1 - \frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\right| \tag{81}
$$

It is easy to verify that $\left|1 - \frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\right|$ is bounded as $\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$ is strictly greater than zero, thus with sufficiently small η one can guarantee that all the three terms ②, ③ and ④ are smaller than $\|\boldsymbol{M}\|_F$. Furthermore, it is easy to prove for small η that

$$
\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F < (1-\eta)^{2k}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F + \|\boldsymbol{M}\|_F, \tag{82}
$$

and in consequence if $k \ge T_1 = O(\ln\frac{d}{\delta})$ for sufficiently small δ, the following holds

$$
\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F \le \|\boldsymbol{M}\|_F,\ \forall k \ge 0, \tag{83}
$$

which implies

$$
\frac{1}{2}\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2 \le \left\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\right\rangle, \forall k \ge 0. \tag{84}
$$

Since

$$
\begin{aligned}
\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F \le{}& (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F + \eta(1-\eta)\|\mathcal{U}_k^\top\boldsymbol{M}\|_F + \eta(1-\eta)\|\boldsymbol{M}\mathcal{V}_k\|_F \\
&+ \eta^2\|\boldsymbol{M}\boldsymbol{V}_k(\boldsymbol{V}_k^\top\boldsymbol{V}_k)^{-1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F, \forall k \ge 0.
\end{aligned} \tag{85}
$$

Similar to Eq. (83), we can guarantee that $\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F \le \|\boldsymbol{M}\|_F$ after T_1 iterations, yet still satisfy $\frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \le 1, \forall k \ge T_1$.

Let us define

$$\cos\boldsymbol{\theta}_k^v = \frac{\|\boldsymbol{M}\mathcal{V}_k\|_F}{\|\boldsymbol{M}\|_F}, \cos\boldsymbol{\theta}_k^u = \frac{\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F}, \tag{86}$$

then we have $\left|1 - \frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\right| \le (1 - \cos\boldsymbol{\theta}_k^u\cos\boldsymbol{\theta}_k^v), \forall k \ge T_1$. Therefore

$$\begin{aligned}\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top - \boldsymbol{M}\|_F \le &\underbrace{(1-\eta)^2\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F}_{①} + (1-\eta)\eta\underbrace{\frac{\|\boldsymbol{M}\mathcal{V}_{k+T_1\perp}\|_F}{\|\boldsymbol{M}\|_F}}_{②}\|\boldsymbol{M}\|_F \\ &+ \eta(1-\eta)\underbrace{\frac{\|\mathcal{U}_{k+T_1\perp}^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F}}_{③}\|\boldsymbol{M}\|_F + \eta^2 C\|\boldsymbol{M}\|_F\underbrace{(1 - \cos\boldsymbol{\theta}_{k+T_1}^u\cos\boldsymbol{\theta}_{k+T_1}^v)}_{④}\end{aligned} \tag{87}$$

According to Lemma 10, Lemma 15 and Lemma 11, we know that the three terms ②, ③ and ④ are upper-bounded by linearly decreasing functions, therefore we have the upper-bound of the objective as

$$\begin{aligned}&\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top - \boldsymbol{M}\|_F \le (1-\eta)^2\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F + (1-\eta)\eta(1-\chi_{k+T_1})^k\|\boldsymbol{M}\|_F\|\mathcal{V}_{*\perp}^\top\mathcal{V}_0\|_2 \\ &+ \eta(1-\eta)(1-\chi_{k+T_1})^k\|\boldsymbol{M}\|_F\|\mathcal{U}_{*\perp}^\top\mathcal{U}_0\|_2 + \eta^2(1-\chi_{k+T_1})^{2k}C\|\boldsymbol{M}\|_F(1 - C_uC_v) \\ &\le (1-\eta)^2\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F + (2-\eta)\eta(1-\chi_{k+T_1})^k\|\boldsymbol{M}\|_F\alpha \\ &= (2-\eta)\eta\sum_{i=1}^{k}(1-\eta)^{2i}(1-\chi_{k+T_1})^{k-i}\|\boldsymbol{M}\|_F\alpha + (1-\eta)^{2(k+T_1)}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\ &\le (1-\chi_{k+T_1})^k\|\boldsymbol{M}\|_F\alpha + (1-\eta)^{2(k+T_1)}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\ &\le (1-\chi_{k+T_1})^k\|\boldsymbol{M}\|_F\alpha_1\end{aligned} \tag{88}$$

$\forall k \ge 0$ and $\alpha_1 = \max\{\alpha, \|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F/\|\boldsymbol{M}\|_F,\}$ $\alpha = \max\{\|\mathcal{V}_{*\perp}^\top\mathcal{V}_0\|_2, \|\mathcal{U}_{*\perp}^\top\mathcal{U}_0\|_2, C(1 - C_uC_v)\}$. Meanwhile, according to Lemma 10, one can deduce that χ_{k+T_1} is lower-bounded by $\frac{\eta^2}{(2-\eta)^2}$ which indicates the linear convergence of the objective function. With the decrease of the objective function, we know that τ_{k+T_1} is increasing from $1/2$ to 1 and correspondingly χ_{k+T_1} increasing from $\frac{\eta^2}{(2-\eta)^2}$ to η.

Furthermore, if $\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 \le \langle\boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle$ and $\frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \ge 1, \forall k \ge T_1$, then according to Lemma 8 and Theorem 6 we know that

$$\begin{aligned}&\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top - \boldsymbol{M}\|_F \le (1-\eta)^2\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F + (1-\eta)\eta(1-\eta)^k\|\boldsymbol{M}\|_F\|\mathcal{V}_{*\perp}^\top\mathcal{V}_0\|_2 \\ &+ \eta(1-\eta)(1-\eta)^k\|\boldsymbol{M}\|_F\|\mathcal{U}_{*\perp}^\top\mathcal{U}_0\|_2 + \eta^2(1-\eta)^kC_\alpha \\ &\le (1-\eta)^2\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F + (2-\eta)\eta(1-\eta)^k\|\boldsymbol{M}\|_F\beta \\ &= (2-\eta)\eta\sum_{i=1}^{k}(1-\eta)^{2i}(1-\eta)^{k-i}\|\boldsymbol{M}\|_F\beta + (1-\eta)^{2(k+T_1)}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\ &\le (1-\eta)^k\|\boldsymbol{M}\|_F\beta + (1-\eta)^{2(k+T_1)}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\ &\le (1-\eta)^k\|\boldsymbol{M}\|_F\alpha_1\end{aligned} \tag{89}$$

where $\alpha_1 = \max\{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F/\|\boldsymbol{M}\|_F, \beta\}$ and $\beta = \max\{\|\mathcal{V}_{*\perp}^\top\mathcal{V}_0\|_2, \|\mathcal{U}_{*\perp}^\top\mathcal{U}_0\|_2, C_\alpha(2\eta - \eta^2)\}$. Both Eq. (88) and Eq. (91) guarantee the linear convergence of the objective function. Since $\chi_{k+T_1} \le \eta$, thus the result in Eq. (79) is proved. □

Theorem 2 (Small initialization)**.** *Let $\boldsymbol{U}_0 \in \mathbb{R}^{m\times d}$ and $\boldsymbol{V}_0 \in \mathbb{R}^{n\times d}$ be random Gaussian that follow $\mathcal{N}(0, \sigma)$, with $\sigma \le c_{\text{init}}$ and $\boldsymbol{U}_k$, $\boldsymbol{V}_k$ are updated by Eq.* (1)*. If $\eta \le c_\eta < 1$ for small constant c_η, we* *have that the objective function of the LRMF problem decreases linearly, namely*

$$\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F \le \alpha_2 \left(1-\eta\right)^k \|\boldsymbol{M}\|_F \tag{90}$$

where c_{init} is a small constant and α_2 is a constant.

Proof. Let c_{init} be a small constant such that if $\sigma \le c_{\text{init}}$, then we have $\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F^2 \ge \left\langle \boldsymbol{U}_0\boldsymbol{V}_0^\top, \boldsymbol{M}\right\rangle$ [1]. Therefore, according to Lemma 8, Lemma 9 and Theorem 6, we know that the objective function is upper-bounded by

$$\begin{aligned}
&\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top - \boldsymbol{M}\|_F \le (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F + (1-\eta)\eta(1-\eta)^k\|\boldsymbol{M}\|_F\|\mathcal{V}_{*\perp}^\top\mathcal{V}_0\|_2 \\
&+ \eta(1-\eta)(1-\eta)^k\|\boldsymbol{M}\|_F\|\mathcal{U}_{*\perp}^\top\mathcal{U}_0\|_2 + \eta^2(1-\eta)^kC_\alpha \\
&\le (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F + (2-\eta)\eta(1-\eta)^k\|\boldsymbol{M}\|_F\beta \\
&= (2-\eta)\eta\sum_{i=1}^{k}(1-\eta)^{2i}(1-\eta)^{k-i}\|\boldsymbol{M}\|_F\beta + (1-\eta)^{2k}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\
&\le (1-\eta)^k\|\boldsymbol{M}\|_F\beta + (1-\eta)^{2k}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\
&\le (1-\eta)^k\|\boldsymbol{M}\|_F\alpha_2
\end{aligned} \tag{91}$$

where $\alpha_2 = \max\{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F/\|\boldsymbol{M}\|_F, \beta\}$ and $\beta = \max\{\|\mathcal{V}_{*\perp}^\top\mathcal{V}_0\|_2, \|\mathcal{U}_{*\perp}^\top\mathcal{U}_0\|_2, (1-\cos\boldsymbol{\theta}_u^0\cos\boldsymbol{\theta}_v^0)(2\eta-\eta^2)\}$. Thus we finish the proof. □

2.2 Proofs for the results of AltScaledGD.

Theorem 3 (General random initialization)**.** *Let $\boldsymbol{U}_0 \in \mathbb{R}^{m\times d}$ and $\boldsymbol{V}_0 \in \mathbb{R}^{n\times d}$ be random Gaussian that follow $\mathcal{N}(0,\sigma)$ for any $\sigma > c_{\text{init}}$, $\boldsymbol{U}_k$, $\boldsymbol{V}_k$ are updated by Eq.* (2)*, we have that the objective function of LRMF problem decreases linearly after $T_1 = O(\ln\frac{d}{\delta})$ iterations, namely*

$$\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F \le \alpha_1 \left(1-\chi_{k+T_1}\right)^k \|\boldsymbol{M}\|_F, \forall k \ge 0 \tag{92}$$

where χ_{k+T_1} is monotonically increasing from $\frac{\eta^2}{(2-\eta)^2}$ to η, $0 < \eta \le 1$ and α_1 is a constant.

Proof. The proof follows the proof sketch given in Section 4.2 of the main paper, while different to Section 4.2, the Theorem 3 focuses on the case $\operatorname{rank}(\boldsymbol{M}) = d$. We decompose the proof into three phases: initial phase, saddle avoid phase and linear convergence phase. Specifically, the objective function is upper-bounded by three terms as

$$\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top - \boldsymbol{M}\|_F \le \underbrace{(1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F}_{①} + (\eta-\eta^2)\underbrace{\|\mathcal{V}_{k\perp}^\top\boldsymbol{M}\|_F}_{②} + \eta\underbrace{\|\mathcal{U}_{k+1\perp}^\top\boldsymbol{M}\|_F}_{③}, \tag{93}$$

1) Initial phase. It can be easily verified that

$$\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F < (1-\eta)^{2k}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F + \|\boldsymbol{M}\|_F, \tag{94}$$

and in consequence if $k \ge T_1 = O(\ln\frac{d}{\delta})$ for sufficiently small δ, the following holds

$$\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F \le \|\boldsymbol{M}\|_F,\ \forall k \ge 0. \tag{95}$$

The initial phase lasts at most T_1 iterations.

2) Saddle avoid phase. The Eq. (95) implies that

$$\frac{1}{2}\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2 \le \langle \boldsymbol{M}, \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\rangle, \tag{96}$$

and furthermore

$$\frac{\|\mathcal{U}_{k+T_1}^\top\boldsymbol{M}\mathcal{V}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} \ge \frac{\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2} = \tau_{k+T_1} \ge \frac{1}{2}. \tag{97}$$

[1] If $\left\langle \boldsymbol{U}_0\boldsymbol{V}_0^\top, \boldsymbol{M}\right\rangle \le 0$ we can simply reset $\boldsymbol{U}_0 = -\boldsymbol{U}_0$ or $\boldsymbol{V}_0 = -\boldsymbol{V}_0$

If furthermore $\frac{\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2} \leq 1$, and according to Eq. (2), we have $\forall k \geq 0$

$$\begin{aligned} \|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 &= (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 + (2\eta - 2\eta^2)\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle + \eta^2\langle \boldsymbol{M}\mathcal{V}_k\mathcal{V}_k^\top, \boldsymbol{M}\mathcal{V}_k\mathcal{V}_k^\top\rangle \\ &\geq (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 + (2\eta - 2\eta^2)\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle + \eta^2\|\mathcal{U}_k^\top\boldsymbol{M}\mathcal{V}_k\|_F^2 \end{aligned} \tag{98}$$

together with Eq. (97), the following holds

$$\frac{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1}^\top\|_F^2} \leq \frac{1}{(1-\eta)^2 + \tau_{k+T_1}(1-\eta) + \eta^2\tau_{k+T_1}^2}. \tag{99}$$

Similarly, one obtains

$$\frac{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1}^\top\|_F^2}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F^2} \leq \frac{1}{(1-\eta)^2 + \tau_{k+T_1}(1-\eta) + \eta^2\tau_{k+T_1}^2}, \tag{100}$$

thus we can guarantee

$$\frac{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F} \leq \frac{1}{(1-\eta+\eta\tau_{k+T_1})^2} \tag{101}$$

Meanwhile, according to Lemma 15, we can upper-bound the terms ② and ③ by

$$\|\mathcal{U}_{k+T_1\perp}^\top\boldsymbol{M}\|_F \leq \frac{\|\mathcal{U}_{*\perp}^\top\boldsymbol{U}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}\|\boldsymbol{M}\|_F\mathfrak{L}_u. \tag{102}$$

and

$$\|\mathcal{V}_{k+T_1\perp}^\top\boldsymbol{M}^\top\|_F \leq \frac{\|\mathcal{V}_{*\perp}^\top\boldsymbol{V}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}\|\boldsymbol{M}\|_F\mathfrak{L}_v \tag{103}$$

respectively. Then we show that the upper-bound Eq. (102) and Eq. (173) decreases linearly. Specifically,

$$\begin{aligned} \frac{\|\mathcal{U}_{*\perp}^\top\boldsymbol{U}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} &= (1-\eta)\frac{\|\mathcal{U}_{*\perp}^\top\boldsymbol{U}_{k+T_1-1}\|_F}{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F}\frac{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} \\ &\leq \frac{1-\eta}{(1-\eta+\eta\tau_{k+T_1})^2}\frac{\|\mathcal{U}_{*\perp}^\top\boldsymbol{U}_{k+T_1-1}\|_F}{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F} \\ &= (1-\chi_{k+T_1})\frac{\|\mathcal{U}_{*\perp}^\top\boldsymbol{U}_{k+T_1-1}\|_F}{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F} \end{aligned}, \tag{104}$$

similarly, we have

$$\begin{aligned} \frac{\|\mathcal{V}_{*\perp}^\top\boldsymbol{V}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} &= (1-\eta)\frac{\|\mathcal{V}_{*\perp}^\top\boldsymbol{V}_{k+T_1-1}\|_F}{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F}\frac{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} \\ &\leq \frac{1-\eta}{(1-\eta+\eta\tau_{k+T_1})^2}\frac{\|\mathcal{V}_{*\perp}^\top\boldsymbol{V}_{k+T_1-1}\|_F}{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F} \\ &= (1-\chi_{k+T_1})\frac{\|\mathcal{V}_{*\perp}^\top\boldsymbol{V}_{k+T_1-1}\|_F}{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F} \end{aligned}, \tag{105}$$

where $\tau_{k+T_1} \in [1/2, 1]$ and correspondingly $1 > \chi_{k+T_1} = \frac{\eta^2(1-\tau_{k+T_1})^2 + 2\eta\tau_{k+T_1} - \eta}{(1-\eta+\eta\tau_{k+T_1})^2} > 0$.

When

$$1 \geq \frac{\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2} = \tau_{k+T_1} \geq \frac{1}{2}. \tag{106}$$

according to Eq. (98) it is easy to verify that $\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F^2 \leq \|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2$. Meanwhile, the objective function is monotonically decreasing according to Eq. (104) and Eq. (105). Together with Eq. (95), we have

$$\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F^2 \leq \|\boldsymbol{M}\|_F^2 - \zeta_{k+T_1}, \tag{107}$$

and ζ_{k+T_1} is monotonically increasing. In consequence, we have

$$\left\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\right\rangle \geq \frac{1}{2}\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2 + \frac{1}{2}\zeta_{k+T_1}, \tag{108}$$

as $\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2$ is monotonically decreasing and ζ_{k+T_1} is monotonically increasing, we can obtain that τ_{k+T_1} is monotonically increasing from $1/2$ to 1 until $\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\rangle \geq \|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2$. Therefore χ_{k+T_1} is monotonically increasing from $\left(\frac{\eta/2}{1-\eta/2}\right)^2$ to η.

3) Linear convergence phase. If $\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\rangle \geq \|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2$ for $k \geq 0$, then according to Lemma 12, we have

$$\langle \boldsymbol{U}_{t+T_1}\boldsymbol{V}_{t+T_1}^\top, \boldsymbol{M}\rangle \geq \|\boldsymbol{U}_{t+T_1}\boldsymbol{V}_{t+T_1}^\top\|_F^2, \forall t > k, \tag{109}$$

together with Eq. (98), we can easily prove that $\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1}^\top\|_F^2 \geq \|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2$ and similarly $\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F^2 \geq \|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1}^\top\|_F^2$, which yields

$$\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F \geq \|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F. \tag{110}$$

According to Eq. (104), Eq. (105) and Eq. (116) we know

$$\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} \leq (1-\eta)\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+T_1-1}\|_F}{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F}, \tag{111}$$

and

$$\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} \leq (1-\eta)\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+T_1-1}\|_F}{\|\boldsymbol{U}_{k+T_1-1}\boldsymbol{V}_{k+T_1-1}^\top\|_F}. \tag{112}$$

Holding all these results, we are guarantee to prove that

$$\begin{aligned}
&\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top - \boldsymbol{M}\|_F \leq (1-\eta)^2\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F \\
&\quad + (\eta-\eta^2)(1-\chi_{k+T_1})^k\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_0\|_F}{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}\|\boldsymbol{M}\|_F\mathfrak{L}_v + \eta(1-\chi_{k+T_1})^k\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_0\|_F}{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}\|\boldsymbol{M}\|_F\mathfrak{L}_u \\
&\quad \leq (2\eta-\eta^2)\sum_{i=1}^{k}(1-\eta)^{2i}(1-\chi_{k+T_1})^{k-i}C_\zeta\|\boldsymbol{M}\|_F + (1-\eta)^{2(k+T_1)}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\
&\quad \leq (1-\chi_{k+T_1})^k C_\zeta\|\boldsymbol{M}\|_F + (1-\chi_{k+T_1})^k\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\
&\quad \leq \alpha_1(1-\chi_{k+T_1})^k\|\boldsymbol{M}\|_F
\end{aligned} \tag{113}$$

where $\chi_{k+T_1} \leq \eta$ is monotonically increasing from $\frac{\eta^2}{(2-\eta)^2}$ to η, $C_\zeta = \max\{\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_0\|_F}{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}, \frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_0\|_F}{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}\}$and $\alpha_1 = \max\{C_\zeta, \frac{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F}\}$, thus we finish the proof. □

Theorem 4 (Small initialization)**.** *Let $\boldsymbol{U}_0 \in \mathbb{R}^{m\times d}$ and $\boldsymbol{V}_0 \in \mathbb{R}^{n\times d}$ be random Gaussian that follow $\mathcal{N}(0,\sigma)$, with $\sigma \leq c_{\text{init}}$, $\boldsymbol{U}_k$, $\boldsymbol{V}_k$ are updated by Eq.* (2) *then we have that the objective function of LRMF problem decreases linearly, namely*

$$\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F \leq \alpha_2\,(1-\eta)^k\,\|\boldsymbol{M}\|_F \tag{114}$$

where $0 < \eta \leq 1$ is the step size, α_2 is a constant and c_{init} is a small constant.

Proof. Let c_{init} be small enough such that $\langle \boldsymbol{U}_0\boldsymbol{V}_0^\top, \boldsymbol{M}\rangle \geq \|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F^2$. According to Lemma 12 the following holds

$$\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2, \forall k. \tag{115}$$

together with Eq. (98), we can prove that $\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$ and similarly $\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2 \geq \|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2$, which yields

$$\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F. \tag{116}$$

According to Eq. (104), Eq. (105) and Eq. (116) we have

$$\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \le (1-\eta)\,\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k-1}\|_F}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F}, \tag{117}$$

and

$$\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \le (1-\eta)\,\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k-1}\|_F}{\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F}. \tag{118}$$

together with Eq. (102) and Eq. (173), we are guarantee to prove

$$\begin{aligned}
\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top - \boldsymbol{M}\|_F &\le (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{M}\|_F + (\eta-\eta^2)(1-\eta)^k\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_0\|_F}{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}\|\boldsymbol{M}\|_F\mathfrak{L}_v \\
&\quad + \eta(1-\eta)^k\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_0\|_F}{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}\|\boldsymbol{M}\|_F\mathfrak{L}_u \\
&\le (2\eta-\eta^2)\sum_{i=1}^{k}(1-\eta)^{2i}(1-\eta)^{k-i}C_\zeta\|\boldsymbol{M}\|_F + (1-\eta)^{2k}\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\
&\le (1-\eta)^k C_\zeta\|\boldsymbol{M}\|_F + (1-\eta)^k\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F \\
&\le \alpha_2(1-\eta)^k\|\boldsymbol{M}\|_F
\end{aligned} \tag{119}$$

where $C_\zeta = \max\{\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_0\|_F}{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}, \frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_0\|_F}{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}\}$and $\alpha_2 = \max\{C_\zeta, \frac{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top - \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F}\}$. We thus finish the proof. □

2.3 Lemmas and some preliminary results.

Lemma 8. *For the ScaledGD (1), if* $\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle \ge \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$*, then the following holds*

$$\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+1}\|_F}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F} \le (1-\eta)\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F},\quad \frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+1}\|_2}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F} \le (1-\eta)\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \tag{120}$$

Proof. Since $\boldsymbol{U}_{k+1} = (1-\eta)\boldsymbol{U}_k + \eta\boldsymbol{M}\boldsymbol{V}_k(\boldsymbol{V}_k^\top\boldsymbol{V}_k)^{-1}$, it is obvious that

$$\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+1}\|_F}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F} = (1-\eta)\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\cdot\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F}. \tag{121}$$

Since $\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle \ge \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$, according to Lemma 17 we know that $\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F} \le 1$, which guarantees that $\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+1}\|_F}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F} \le (1-\eta)\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}$.

Meanwhile, $\boldsymbol{V}_{k+1} = (1-\eta)\boldsymbol{V}_k + \eta\boldsymbol{M}^\top\boldsymbol{U}_k(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}$, we have

$$\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+1}\|_2}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F} \le (1-\eta)\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\cdot\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F}, \tag{122}$$

similarly we can also guarantee that $\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+1}\|_2}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F} \le (1-\eta)\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}$. Thus, we finish the proof. □

Theorem 6. (Convergence of the matrix norm) *For the ScaledGD (1), if* $\langle \boldsymbol{U}_0\boldsymbol{V}_0^\top, \boldsymbol{M}\rangle \ge \|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F^2$*, then we have*

$$\|\boldsymbol{M}\|_F - \|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F \le (1-\eta)^k C_\alpha, \forall k \ge 0 \tag{123}$$

where C_α *is a constant and* η *is the step length* $0 \le \eta < 1$.

Proof. By ScaledGD Eq. (1), we have

$$\begin{aligned}
\frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} &= (1-\eta)^2\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} + 2\eta(1-\eta)\frac{\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{M}\|_F^2} + \eta^2\frac{\|\boldsymbol{M}\mathcal{V}_k\|_F^2}{\|\boldsymbol{M}\|_F^2} \\
&\ge (1-\eta^2)\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} + \eta^2\cos^2\boldsymbol{\theta}_k^v
\end{aligned} \tag{124}$$

the second inequality is due to $\left\langle \boldsymbol{U}_k \boldsymbol{V}_k^\top, \boldsymbol{M} \right\rangle \geq \|\boldsymbol{U}_k \boldsymbol{V}_k^\top\|_F^2$. In consequence,

$$\begin{aligned} \frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} &\geq \frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} \geq (1-\eta^2)\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} + \eta^2\cos^2\boldsymbol{\theta}_k^v \\ &\geq (1-\eta)\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} + \eta^2\cos^2\boldsymbol{\theta}_k^v \end{aligned} \tag{125}$$

where we define

$$\cos\boldsymbol{\theta}_k^v = \frac{\|\boldsymbol{M}\mathcal{V}_k\|_F}{\|\boldsymbol{M}\|_F}, \cos\boldsymbol{\theta}_k^u = \frac{\|\mathcal{U}_k^\top \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F}$$

In consequence, one obtains

$$1 - \frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} \leq (1-\eta)\left(1 - \frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{M}\|_F^2}\right) + \eta(1-\cos^2\boldsymbol{\theta}_v^k), \tag{126}$$

Meanwhile, according to Lemma 8, Lemma 11 and Lemma 15, we have

$$\frac{\|\mathcal{U}_{k\perp}^\top \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F} \leq (1-\eta)^k C_u, \frac{\|\boldsymbol{M}\mathcal{V}_{k\perp}\|_F}{\|\boldsymbol{M}\|_F} \leq (1-\eta)^k C_v, \forall k > 0 \tag{127}$$

As

$$\cos^2\boldsymbol{\theta}_k^u = 1 - \frac{\|\mathcal{U}_{k\perp}^\top \boldsymbol{M}\|_F^2}{\|\boldsymbol{M}\|_F^2}, \cos^2\boldsymbol{\theta}_k^v = 1 - \frac{\|\boldsymbol{M}\mathcal{V}_{k\perp}\|_F^2}{\|\boldsymbol{M}\|_F^2}$$

which guarantees that

$$1 - \cos^2\boldsymbol{\theta}_k^v \leq (1-\eta)^{2k} C_v$$

Together with Eq. (126), the following inequality holds

$$\begin{aligned} 1 - \frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2}{\|\boldsymbol{M}\|_F^2} &\leq \eta^2 \sum_{i=1}^{k} (1-\eta)^i (1-\cos^2\boldsymbol{\theta}_v^{k-i}) + (1-\eta)^k \left(1 - \frac{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F^2}{\|\boldsymbol{M}\|_F^2}\right) \\ &\leq \eta^2 \sum_{i=1}^{k} (1-\eta)^i (1-\eta)^{2(k-i)} C_v + (1-\eta)^k \left(1 - \frac{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F^2}{\|\boldsymbol{M}\|_F^2}\right) \\ &\leq (1-\eta)^k C_{\alpha_1}, \end{aligned} \tag{128}$$

where $C_{\alpha_1} = \max\left\{C_v\eta^2, 1 - \frac{\|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F}{\|\boldsymbol{M}\|_F}\right\}$, which further implies that

$$\|\boldsymbol{M}\|_F - \frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F}{\|\boldsymbol{M}\|_F} \leq (1-\eta)^k C_{\alpha_1} \frac{\|\boldsymbol{M}\|_F^2}{\|\boldsymbol{M}\|_F + \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \leq (1-\eta)^k C_\alpha. \tag{129}$$

thus we finish our proof. □

Lemma 9. *Let $\eta \leq c_\eta < 1$ with c_η a small constant, if $\left\langle \boldsymbol{U}_0\boldsymbol{V}_0^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F^2$ then the following is true*

$$\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2, \forall k > 0. \tag{130}$$

Proof. We prove this result by induction. Since $\left\langle \boldsymbol{U}_0\boldsymbol{V}_0^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_0\boldsymbol{V}_0^\top\|_F^2$, we assume that $\left\langle \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F^2$, then we need to prove $\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$. We first show that

$$\begin{aligned} \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 &= (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F^2 + 2\eta(1-\eta)\left\langle \boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top, \boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\right\rangle \\ &\quad + \eta^2\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\|_F^2 \\ &\leq \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F^2 \left(1 - \eta + \eta\frac{\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F}\right)^2 \\ &= \|\boldsymbol{M}\|_F^2 \underbrace{\frac{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F^2}{\|\boldsymbol{M}\|_F^2}\left(1 - \eta + \eta\frac{\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F}\right)^2}_{\mathfrak{M}} \end{aligned} \tag{131}$$

To prove the result in Eq. (130), we need to guarantee that $\mathfrak{M} \leq \frac{\langle \boldsymbol{M}, \boldsymbol{U}_k \boldsymbol{V}_k^\top \rangle}{\|\boldsymbol{M}\|_F^2}$, which is equivalent to ensure that

$$(1-\eta)\frac{\|\boldsymbol{U}_k \boldsymbol{V}_{k-1}^\top\|_F}{\|\boldsymbol{M}\|_F} + \eta \frac{\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top \boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F} \leq \frac{\sqrt{\langle \boldsymbol{M}, \boldsymbol{U}_k \boldsymbol{V}_k^\top \rangle}}{\|\boldsymbol{M}\|_F} \tag{132}$$

Meanwhile, since $\langle \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top, \boldsymbol{M} \rangle \geq \|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F^2$, and

$$\begin{aligned}
&\|\boldsymbol{U}_k \boldsymbol{V}_{k-1}^\top\|_F^2 = \|(1-\eta)\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top + \eta \boldsymbol{M}\mathcal{V}_{k-1}\mathcal{V}_{k-1}^\top\|_F^2 \\
&= (1-\eta)^2\|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F^2 + 2\eta(1-\eta)\langle \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top, \boldsymbol{M} \rangle + \eta^2\|\boldsymbol{M}\mathcal{V}_{k-1}\|_F^2
\end{aligned} \tag{133}$$

it is easy to verify

$$\begin{aligned}
&\|\boldsymbol{U}_k \boldsymbol{V}_{k-1}^\top\|_F^2 \leq (1-\eta^2)\langle \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top, \boldsymbol{M} \rangle + \eta^2\|\boldsymbol{M}\mathcal{V}_{k-1}\|_F^2 \\
&= (1-\eta)\langle \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top, \boldsymbol{M} \rangle + (\eta-\eta^2)\langle \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top, \boldsymbol{M} \rangle + \eta^2\|\boldsymbol{M}\mathcal{V}_{k-1}\|_F^2 \\
&\leq (1-\eta)\langle \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top, \boldsymbol{M} \rangle + (\eta-\eta^2)\|\boldsymbol{M}\mathcal{V}_{k-1}\|_F^2 + \eta^2\|\boldsymbol{M}\mathcal{V}_{k-1}\|_F^2 \\
&= \langle \boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top, \boldsymbol{M} \rangle,
\end{aligned} \tag{134}$$

where the second inequality is due to Lemma 13. In consequence, one can verify that

$$\langle \boldsymbol{M}, \boldsymbol{U}_k\boldsymbol{V}_k^\top \rangle \geq \langle \boldsymbol{M}, \boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top \rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F^2, \tag{135}$$

as

$$\begin{aligned}
\langle \boldsymbol{M}, \boldsymbol{U}_k\boldsymbol{V}_k^\top - \boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top \rangle &= \langle \boldsymbol{M}, \boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M} \rangle - \langle \boldsymbol{M}, \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top \rangle \\
&\geq \langle \boldsymbol{M}, \boldsymbol{U}_{k-1}(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M} \rangle - \langle \boldsymbol{M}, \boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top \rangle, \\
&\geq 0
\end{aligned} \tag{136}$$

which is due to Lemma 13.

Denote by

$$\mathfrak{Z} = \frac{\sqrt{\langle \boldsymbol{M}, \boldsymbol{U}_k\boldsymbol{V}_k^\top \rangle} - \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F}{\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\|_F - \|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F} \tag{137}$$

then it can be easily verified that $\mathfrak{Z} > 0$. As a result

$$\mathfrak{Z}\left(\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\|_F - \|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F\right) = \sqrt{\langle \boldsymbol{M}, \boldsymbol{U}_k\boldsymbol{V}_k^\top \rangle} - \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F \tag{138}$$

If $\eta \leq c_\eta < \mathfrak{Z}$ with c_η sufficiently small, then we can guarantee that

$$\begin{aligned}
&\eta\left(\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\|_F - \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F\right) \\
&\leq \eta\left(\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\|_F - \|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F\right) \\
&\leq \mathfrak{Z}\left(\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top\boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top\boldsymbol{M}\|_F - \|\boldsymbol{U}_{k-1}\boldsymbol{V}_{k-1}^\top\|_F\right) \\
&= \sqrt{\langle \boldsymbol{M}, \boldsymbol{U}_k\boldsymbol{V}_k^\top \rangle} - \|\boldsymbol{U}_k\boldsymbol{V}_{k-1}^\top\|_F
\end{aligned} \tag{139}$$

By simple reformulation, one obtains

$$(1-\eta)\frac{\|\boldsymbol{U}_k \boldsymbol{V}_{k-1}^\top\|_F}{\|\boldsymbol{M}\|_F} + \eta \frac{\|\boldsymbol{U}_k(\boldsymbol{U}_{k-1}^\top \boldsymbol{U}_{k-1})^{-1}\boldsymbol{U}_{k-1}^\top \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F} \leq \frac{\sqrt{\langle \boldsymbol{M}, \boldsymbol{U}_k \boldsymbol{V}_k^\top \rangle}}{\|\boldsymbol{M}\|_F} \tag{140}$$

which is exactly the inequality we need in Eq. (132), thus we finish the proof. □

Lemma 10. *Let us define*

$$\cos\boldsymbol{\theta}_k^v = \frac{\|\boldsymbol{M}\mathcal{V}_k\|_F}{\|\boldsymbol{M}\|_F}, \cos\boldsymbol{\theta}_k^u = \frac{\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F}, \tag{141}$$

then after $T_1 = O(\ln \frac{d}{\delta})$ iterations of ScaledGD, the following inequalities hold

$$\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+T_1+1}\|_F}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F} \leq (1-\chi_{k+T_1})\,\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}, \forall k > 0. \tag{142}$$

$$\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+T_1+1}\|_F}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F} \leq (1-\chi_{k+T_1})\,\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}, \forall k > 0. \tag{143}$$

$$1-\cos\boldsymbol{\theta}_{k+T_1+1}^u \cos\boldsymbol{\theta}_{k+T_1+1}^v \leq (1-\chi_{k+T_1})^{2k}\,(1-C_u C_v)\,, \forall k > 0. \tag{144}$$

where $\chi_{k+T_1} = \frac{\eta^2(1-\tau_{k+T_1})^2+2\eta\tau_{k+T_1}-\eta}{(1-\eta+\eta\tau_{k+T_1})^2} > 0$, $\tau_{k+T_1} = \frac{\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2} \in [1/2, 1]$.

Proof. According to the analysis in the proof of the Theorem 1 and Eq. (84), we know that after T_1 iterations, we have

$$\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top - \boldsymbol{M}\|_F \leq \|\boldsymbol{M}\|_F,\ \forall k \geq 0, \tag{145}$$

which implies

$$\tau_{k+T_1} = \frac{\langle \boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F^2} \geq 1/2 \tag{146}$$

Since the results for $\tau_{k+T_1} \geq 1$ is easy to analyze according to Lemma 8, therefore we mainly focus on $1 \geq \tau_{k+T_1} \geq 1/2$.

According to the results in Lemma 18, if $1 \geq \tau_{k+T_1} \geq 1/2$, we have

$$\frac{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F} \leq \frac{1}{(1-\eta+\eta\tau_{k+T_1})^2} \tag{147}$$

therefore,

$$\begin{aligned}
\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+T_1+1}\|_F}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F} &= (1-\eta)\,\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} \cdot \frac{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F} \\
&\leq \frac{(1-\eta)}{(1-\eta+\eta\tau_{k+T_1})^2}\,\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F} \\
&= (1-\chi_{k+T_1})\,\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}
\end{aligned} \tag{148}$$

with $\chi_{k+T_1} = \frac{\eta^2(1-\tau_{k+T_1})^2+2\eta\tau_{k+T_1}-\eta}{(1-\eta+\eta\tau_{k+T_1})^2}$ and $0 < \chi_{k+T_1} < 1$ when $1 \geq \tau_{k+T_1} \geq 1/2$.

In the same way, we can prove

$$\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+T_1+1}\|_F}{\|\boldsymbol{U}_{k+T_1+1}\boldsymbol{V}_{k+T_1+1}^\top\|_F} \leq (1-\chi_{k+T_1})\,\frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_{k+T_1}\|_F}{\|\boldsymbol{U}_{k+T_1}\boldsymbol{V}_{k+T_1}^\top\|_F}, \forall k > 0 \tag{149}$$

with $\chi_{k+T_1} = \frac{\eta^2(1-\tau_{k+T_1})^2+2\eta\tau_{k+T_1}-\eta}{(1-\eta+\eta\tau_{k+T_1})^2}$ and $0 < \chi_{k+T_1} < 1$.

According to Lemma 15, we know that there exists bounded value $0 < \mathfrak{b}$ such that

$$\frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \leq \mathfrak{b}\,\frac{\|\mathcal{U}_{k\perp}^\top \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F}, \forall k \geq 0.$$

Together with Lemma 11, Eq. (148) and Eq. (149), we have

$$\frac{\|\mathcal{U}_{k+T_1+1\perp}^\top \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F} \leq (1-\chi_{k+T_1})^k C_u, \forall k > 0 \tag{150}$$

for bounded constant C_u. The same is true for $\boldsymbol{V}$ as

$$\frac{\|\boldsymbol{M}\mathcal{V}_{k+T_1+1\perp}\|_F}{\|\boldsymbol{M}\|_F} \leq (1-\chi_{k+T_1})^k C_v, \forall k > 0 \tag{151}$$

for bounded C_v. One can simply assume that $C_u \leq 1$ and $C_v \leq 1$, since $(1-\chi_{k+T_1})^k C_u = (1-\chi_{k+T_1})^{k_1}(1-\chi_{k+T_1})^{k_2} C_u$ such that $(1-\chi_{k+T_1})^{k_2} C_u \leq 1$.

As

$$\cos^2\boldsymbol{\theta}_k^u = 1 - \frac{\|\mathcal{U}_{k\perp}^\top \boldsymbol{M}\|_F^2}{\|\boldsymbol{M}\|_F^2}, \cos^2\boldsymbol{\theta}_k^v = 1 - \frac{\|\boldsymbol{M}\mathcal{V}_{k\perp}\|_F^2}{\|\boldsymbol{M}\|_F^2}$$

together with Eq. (150) and Eq. (151) one obtains

$$\cos^2\boldsymbol{\theta}_{k+T_1+1}^u \geq 1 - (1-\chi_{k+T_1})^{2k} C_u^2 = \left(1-(1-\chi_{k+T_1})^{2k}\right) + (1-\chi_{k+T_1})^{2k}(1-C_u^2)$$

$$\cos^2\boldsymbol{\theta}_{k+T_1+1}^v \geq 1 - (1-\chi_{k+T_1})^{2k} C_v^2 = \left(1-(1-\chi_{k+T_1})^{2k}\right) + (1-\chi_{k+T_1})^{2k}(1-C_v^2)$$

consequently, we have

$$1 - \cos\boldsymbol{\theta}_{k+T_1+1}^u \cos\boldsymbol{\theta}_{k+T_1+1}^v \leq (1-\chi_{k+T_1})^{2k}(1-C_uC_v)$$

□

Lemma 11. *Let $\{x_k\}_{k=1}^\infty$ and $\{y_k\}_{k=1}^\infty$ be two sequences such that $x_k \leq (1-\zeta)x_{k-1}$ with $0 \leq \zeta \leq 1$, and*

$$\alpha x_k \leq y_k \leq \beta x_k, \forall k$$

then we have that

$$y_k \leq (1-\zeta)^k c_0, \forall k > 0,$$

where $c_0 = \beta x_0$ and α, β are bounded positive values.

Proof. The proof is trivial since $x_k \leq (1-\zeta)^k x_0$, thus $(1-\zeta)^k \alpha x_0 \leq y_k \leq (1-\zeta)^k \beta x_0$. □

Lemma 12. *If there exits $t \in \mathbb{N}_+$ such that $\boldsymbol{U}_t$ and $\boldsymbol{V}_t \in \mathbb{R}^{n\times r}$ in Eq. (2) satisfy $\left\langle \boldsymbol{U}_t\boldsymbol{V}_t^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_t\boldsymbol{V}_t^\top\|_F^2$, then we have*

$$\left\langle \boldsymbol{U}_{k+t}\boldsymbol{V}_{k+t}^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_{k+t}\boldsymbol{V}_{k+t}^\top\|_F^2 \tag{152}$$

for any $k \in \mathbb{N}_+$.

Proof. Without loss of generality, we assume $t = 0$, thus we prove the result by induction. We assume that the result is true for k, then we prove

$$\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2. \tag{153}$$

We first need to prove that $\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$ implies $\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2$. Note that

$$\begin{aligned}
&\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 = \|(1-\eta)\boldsymbol{U}_k\boldsymbol{V}_k^\top + \eta\boldsymbol{M}\mathcal{V}_k\mathcal{V}_k^\top\|_F^2 \\
&= (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 + 2\eta(1-\eta)\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + \eta^2\|\boldsymbol{M}\mathcal{V}_k\|_F^2
\end{aligned} \tag{154}$$

According to Lemma 13 we know $\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle \leq \|\boldsymbol{M}\mathcal{V}_k\|_F^2$, together with $\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$ therefore

$$\begin{aligned}
&\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 \leq (1-\eta^2)\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + \eta^2\|\boldsymbol{M}\mathcal{V}_k\|_F^2 \\
&= (1-\eta)\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + (\eta-\eta^2)\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + \eta^2\|\boldsymbol{M}\mathcal{V}_k\|_F^2 \\
&\leq (1-\eta)\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + (\eta-\eta^2)\|\boldsymbol{M}\mathcal{V}_k\|_F^2 + \eta^2\|\boldsymbol{M}\mathcal{V}_k\|_F^2 \\
&= \left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle.
\end{aligned} \tag{155}$$

Again, according to Lemma 13, we have $\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle \leq \|\mathcal{U}_{k+1}^\top\boldsymbol{M}\|_F^2$. In the same way, we can prove that

$$\begin{aligned}
&\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2 \leq (1-\eta^2)\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + \eta^2\|\mathcal{U}_{k+1}\boldsymbol{M}\|_F^2 \\
&= (1-\eta)\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + (\eta-\eta^2)\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + \eta^2\|\boldsymbol{M}\mathcal{V}_k\|_F^2 \\
&\leq (1-\eta)\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle + (\eta-\eta^2)\|\mathcal{U}_{k+1}\boldsymbol{M}\|_F^2 + \eta^2\|\mathcal{U}_{k+1}\boldsymbol{M}\|_F^2 \\
&= \left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top, \boldsymbol{M}\right\rangle,
\end{aligned} \tag{156}$$

which finishes our proof. □

Lemma 13. *For any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$, such that $\langle \boldsymbol{A}, \boldsymbol{B}\rangle \geq \|\boldsymbol{A}\|_F^2$, then we have*

$$\langle \boldsymbol{A}, \boldsymbol{B}\rangle \leq \|\mathcal{U}_{\boldsymbol{A}}^\top \boldsymbol{B}\|_F^2, \langle \boldsymbol{A}, \boldsymbol{B}\rangle \leq \|\boldsymbol{B}\mathcal{V}_{\boldsymbol{A}}\|_F^2 \tag{157}$$

where $\mathcal{U}_{\boldsymbol{A}}$ is the left singular vector matrix of $\boldsymbol{A}$.

Proof. Let $\boldsymbol{A} = \mathcal{U}_{\boldsymbol{A}}\Sigma_{\boldsymbol{A}}\mathcal{V}_{\boldsymbol{A}}^\top$ be the SVD of $\boldsymbol{A}$, the following holds

$$\begin{aligned}\langle \boldsymbol{A}, \boldsymbol{B}\rangle &= \left\langle \mathcal{U}_{\boldsymbol{A}}\Sigma_{\boldsymbol{A}}\mathcal{V}_{\boldsymbol{A}}^\top, \boldsymbol{B}\right\rangle \\ &= \left\langle \Sigma_{\boldsymbol{A}}, \mathcal{U}_{\boldsymbol{A}}^\top \boldsymbol{B}\mathcal{V}_{\boldsymbol{A}}\right\rangle \\ &\leq \|\boldsymbol{A}\|_F\|\mathcal{U}_{\boldsymbol{A}}^\top \boldsymbol{B}\mathcal{V}_{\boldsymbol{A}}\|_F \\ &\leq \|\boldsymbol{A}\|_F\|\mathcal{U}_{\boldsymbol{A}}^\top \boldsymbol{B}\|_F\end{aligned} \tag{158}$$

According to $\langle \boldsymbol{A}, \boldsymbol{B}\rangle \geq \|\boldsymbol{A}\|_F^2$, we have

$$\begin{aligned}\langle \boldsymbol{A}, \boldsymbol{B}\rangle^2 &\leq \|\boldsymbol{A}\|_F^2\|\mathcal{U}_{\boldsymbol{A}}^\top \boldsymbol{B}\|_F^2 \\ &\leq \langle \boldsymbol{A}, \boldsymbol{B}\rangle \|\mathcal{U}_{\boldsymbol{A}}^\top \boldsymbol{B}\|_F^2\end{aligned} \tag{159}$$

which implies

$$\langle \boldsymbol{A}, \boldsymbol{B}\rangle \leq \|\mathcal{U}_{\boldsymbol{A}}^\top \boldsymbol{B}\|_F^2. \tag{160}$$

□

Lemma 14. *Let $\boldsymbol{U}_k \in \mathbb{R}^{m\times d}$ and $\boldsymbol{V}_k \in \mathbb{R}^{n\times d}$ be given by Eq.* (1)*, there exists a bounded constant C such that*

$$\|\boldsymbol{M}\boldsymbol{V}(\boldsymbol{V}^\top\boldsymbol{V})^{-1}(\boldsymbol{U}^\top\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{M} - \boldsymbol{M}\|_F < C\|\boldsymbol{M}\|_F\left|1 - \frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\right| \tag{161}$$

when $\operatorname{rank}(\boldsymbol{M}) = 1$ *and $d = 1$, the constant C becomes 1.*

Proof. According the pseudo inverse theorem, we have

$$\boldsymbol{M}\boldsymbol{V}(\boldsymbol{V}^\top\boldsymbol{V})^{-1}(\boldsymbol{U}^\top\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{M} = \boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k\boldsymbol{M},$$

thus

$$\|\boldsymbol{M}\boldsymbol{V}(\boldsymbol{V}^\top\boldsymbol{V})^{-1}(\boldsymbol{U}^\top\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{M} - \boldsymbol{M}\|_F \leq \|\boldsymbol{M}\|_F\|\boldsymbol{M}^\dagger\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k\boldsymbol{M} - \boldsymbol{I}\|_2. \tag{162}$$

Meanwhile,

$$\|\boldsymbol{M}^\dagger\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k\boldsymbol{M} - \boldsymbol{I}\|_2 \leq \max\left\{\|\boldsymbol{M}^\dagger\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k\boldsymbol{M}\|_2 - 1, 1 - \sigma_r(\boldsymbol{M}^\dagger\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k\boldsymbol{M})\right\}. \tag{163}$$

On one hand,

$$\|\boldsymbol{M}^\dagger\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k\boldsymbol{M}\|_2 \leq \frac{\|\boldsymbol{M}\mathcal{V}_k\|_2\|\mathcal{U}_k^\top\boldsymbol{M}\|_2}{\|\boldsymbol{M}\|_d\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_d} \leq \tilde{\kappa}\frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} \tag{164}$$

where $\tilde{\kappa} = \kappa_1\kappa_2\kappa_3\kappa_4$ and $\kappa_1 = \sigma_1(\boldsymbol{M})/\sigma_d(\boldsymbol{M})$...

Since $\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$ is greater than zero, in consequence, we have

$$\|\boldsymbol{M}^\dagger\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k\boldsymbol{M}\|_2 - 1 < C_1\left(\frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F} - 1\right)$$

and similarly, we have

$$1 - \sigma_r(\boldsymbol{M}^\dagger\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k\boldsymbol{M}) < C_2\left(1 - \frac{\|\boldsymbol{M}\mathcal{V}_k\|_F\|\mathcal{U}_k^\top\boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F})\right)$$

together with Eq. (162) and Eq. (163), we finish our proof. □

Lemma 15. *Let $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ are given by Eq.* (1) *or Eq.* (2) *and $\mathcal{U}_k\boldsymbol{\Sigma}_k\mathcal{V}_k^\top = \boldsymbol{U}_k\boldsymbol{V}_k^\top$, $\mathcal{U}_*\boldsymbol{\Sigma}_*\mathcal{V}_*^\top = \boldsymbol{M}$ are the SVD of $\boldsymbol{U}_k\boldsymbol{V}_k^\top$ and $\boldsymbol{M}$ respectively, then we have*

$$\frac{\|\mathcal{U}_{k\perp}^\top \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F} \leq \frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\mathfrak{L}_u. \tag{165}$$

and

$$\frac{\|\mathcal{V}_{k\perp}^\top \boldsymbol{M}^\top\|_F}{\|\boldsymbol{M}\|_F} \leq \frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\mathfrak{L}_v \tag{166}$$

where $\mathfrak{L}_u$ and $\mathfrak{L}_v$ are bounded values.

Proof. Let $\mathcal{U}_k^\top\mathcal{U}_* = \boldsymbol{A}_{uk}\cos\boldsymbol{\Theta}_{uk}\boldsymbol{B}_{uk}^\top$ be the SVD decomposition. Then $\|\mathcal{U}_{*\perp}^\top\boldsymbol{U}_k\|_F^2$ becomes

$$\mathrm{Trace}\left(\boldsymbol{\Sigma}_{uk}\mathcal{U}_k^\top\mathcal{U}_{*\perp}\mathcal{U}_{*\perp}^\top\mathcal{U}_k\boldsymbol{\Sigma}_{uk}\right) = \mathrm{Trace}\left(\boldsymbol{\Sigma}_{uk}\boldsymbol{A}_{uk}\sin^2\boldsymbol{\Theta}_{uk}\boldsymbol{A}_{uk}^\top\boldsymbol{\Sigma}_{uk}\right) = \|\sin\boldsymbol{\Theta}_{uk}\boldsymbol{A}_{uk}^\top\boldsymbol{\Sigma}_{uk}\|_F^2 \tag{167}$$

Similarly, we have

$$\begin{aligned} \|\mathcal{U}_{k\perp}^\top\boldsymbol{M}\|_F^2 &= \mathrm{Trace}\left(\boldsymbol{\Sigma}_*\mathcal{U}_*^\top\mathcal{U}_{k\perp}\mathcal{U}_{k\perp}^\top\mathcal{U}_*\boldsymbol{\Sigma}_{1*}\right) \\ &= \mathrm{Trace}\left(\boldsymbol{\Sigma}_*\boldsymbol{B}_{uk}\sin^2\boldsymbol{\Theta}_{uk}\boldsymbol{B}_{uk}^\top\boldsymbol{\Sigma}_*\right) = \|\sin\boldsymbol{\Theta}_{uk}\boldsymbol{B}_{uk}^\top\boldsymbol{\Sigma}_*\|_F^2 \end{aligned} \tag{168}$$

The Eq. (167) and Eq. (168) indicate that

$$\|\mathcal{U}_{k\perp}^\top\boldsymbol{M}\|_F = \|\sin\boldsymbol{\Theta}_{uk}\boldsymbol{B}_{uk}^\top\boldsymbol{\Sigma}_*\|_F \text{ and } \|\mathcal{U}_{*\perp}^\top\mathcal{U}_k\boldsymbol{\Sigma}_{uk}\|_F = \|\sin\boldsymbol{\Theta}_{uk}\boldsymbol{A}_{uk}^\top\boldsymbol{\Sigma}_{uk}\|_F, \tag{169}$$

which implies

$$\|\mathcal{U}_{k\perp}^\top\boldsymbol{M}\|_F = \|\sin\boldsymbol{\Theta}_{uk}\boldsymbol{A}_{uk}^\top\boldsymbol{\Sigma}_{uk}\boldsymbol{\Sigma}_{uk}^{-1}\boldsymbol{A}_{uk}\boldsymbol{B}_{uk}^\top\boldsymbol{\Sigma}_*\|_F \leq \|\mathcal{U}_{*\perp}^\top\boldsymbol{U}_k\|_F\|\boldsymbol{\Sigma}_{uk}^{-1}\boldsymbol{A}_{uk}\boldsymbol{B}_{uk}^\top\boldsymbol{\Sigma}_*\|_F. \tag{170}$$

In consequence

$$\frac{\|\mathcal{U}_{k\perp}^\top\mathcal{U}_*\boldsymbol{\Sigma}_*\|_F}{\|\boldsymbol{\Sigma}_*\|_F} \leq \frac{\|\mathcal{U}_{*\perp}^\top\boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\underbrace{\|\boldsymbol{\Sigma}_{uk}^{-1}\boldsymbol{A}_{uk}\boldsymbol{B}_{uk}^\top\boldsymbol{\Sigma}_*\|_F\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{\Sigma}_*\|_F}}_{\mathfrak{L}_u}, \tag{171}$$

thus we obtain

$$\frac{\|\mathcal{U}_{k\perp}^\top \boldsymbol{M}\|_F}{\|\boldsymbol{M}\|_F} \leq \frac{\|\mathcal{U}_{*\perp}^\top \boldsymbol{U}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\mathfrak{L}_u, \tag{172}$$

where $\mathfrak{L}_u$ is bounded value. Let $\mathcal{V}_k^\top\mathcal{V}_* = \boldsymbol{A}_{vk}\cos\boldsymbol{\Theta}_{vk}\boldsymbol{B}_{vk}^\top$, in the same way, we have

$$\frac{\|\mathcal{V}_{k\perp}^\top \boldsymbol{M}^\top\|_F}{\|\boldsymbol{M}\|_F} \leq \frac{\|\mathcal{V}_{*\perp}^\top \boldsymbol{V}_k\|_F}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}\mathfrak{L}_v \tag{173}$$

where $\mathfrak{L}_v = \|\boldsymbol{\Sigma}_{vk}^{-1}\boldsymbol{A}_{vk}\boldsymbol{B}_{vk}^\top\boldsymbol{\Sigma}_*\|_F\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{\Sigma}_*\|_F}$ is bounded value.

□

Lemma 16. *Let $\boldsymbol{U}_k, \boldsymbol{V}_k$ be updated by Eq.* (1)*, then with the learning rate $\eta \leq c_\eta$, we have*

$$\kappa(\boldsymbol{U}_k\boldsymbol{V}_k^\top) \leq \mathfrak{m}, \kappa(\boldsymbol{U}_k) \leq \mathfrak{u}, \kappa(\boldsymbol{V}_k) \leq \mathfrak{v}, \forall k \tag{174}$$

where $\kappa(\boldsymbol{X})$ outputs the condition number of a matrix $\boldsymbol{X}$, $\mathfrak{u}, \mathfrak{v}, \mathfrak{m}$ are bounded positive constant.

Proof. With Eq. (1), the product of the matrix $\boldsymbol{U}$ and $\boldsymbol{V}$ are updated by

$$\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top = (1-\eta)^2\boldsymbol{U}_k\boldsymbol{V}_k^\top + \eta(1-\eta)\mathcal{U}_k\mathcal{U}_k^\top\boldsymbol{M} + \eta(1-\eta)\boldsymbol{M}\mathcal{V}_k\mathcal{V}_k^\top + \eta^2\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k^\top\boldsymbol{M} \tag{175}$$

To guarantee that the condition number $\kappa(\boldsymbol{U}_k\boldsymbol{V}_k^\top)$ of the matrix $\boldsymbol{U}_k\boldsymbol{V}_k^\top, \forall k$ are bounded, we only need to guarantee that $\sigma_r(\boldsymbol{U}_k\boldsymbol{V}_k^\top)$ is strictly greater than 0 according to Eq. (175). Next, we prove by contradiction. If we assume that there exists k such that $\sigma_r(\boldsymbol{U}_k\boldsymbol{V}_k^\top) \leq \epsilon, \forall\epsilon \geq 0$, then we have that $\forall\mathfrak{M} \geq 0, \sigma_1(\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top) \geq \mathfrak{M}$ according to Eq. (175), which indicates that $\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F \geq \mathfrak{M}, \forall\mathfrak{M} \geq 0$.

Meanwhile, we have

$$\begin{aligned}\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F &= \|(1-\eta)^2\boldsymbol{U}_k\boldsymbol{V}_k^\top + \eta(1-\eta)\mathcal{U}_k\mathcal{U}_k^\top\boldsymbol{M} + \eta(1-\eta)\boldsymbol{M}\mathcal{V}_k\mathcal{V}_k^\top + \eta^2\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k^\top\boldsymbol{M}\|_F \\ &\leq (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F + 2\eta(1-\eta)\|\boldsymbol{M}\|_F + \eta^2\|\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k^\top\boldsymbol{M}\|_F\end{aligned} \tag{176}$$

which indicates that the upper-bound of $\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$ will decrease with the increase of k if $\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F$ is greater than some constant related to $\|\boldsymbol{M}\mathcal{V}_k\boldsymbol{\Sigma}_k^{-1}\mathcal{U}_k^\top\boldsymbol{M}\|_F$. The result is in contradict to that $\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F \geq \mathfrak{M}, \forall \mathfrak{M} \geq 0$, which proves that $\sigma_r(\boldsymbol{U}_k\boldsymbol{V}_k^\top)$ is strictly greater than zero. Therefore, we can guarantee that there exist bounded constant value $\mathfrak{m}$ such that $\kappa(\boldsymbol{U}_k\boldsymbol{V}_k^\top) \leq \mathfrak{m}$.

Meanwhile, since

$$\sigma_1(\boldsymbol{U}_k\boldsymbol{V}_k^\top) \leq \sigma_1(\boldsymbol{U}_k)\sigma_1(\boldsymbol{V}_k), \sigma_r(\boldsymbol{U}_k\boldsymbol{V}_k^\top) \geq \sigma_r(\boldsymbol{U}_k)\sigma_r(\boldsymbol{V}_k), \tag{177}$$

we can guarantee that there exist bounded constant $\mathfrak{u}$ and $\mathfrak{v}$ such that

$$\kappa(\boldsymbol{U}_k) \leq \mathfrak{u}, \kappa(\boldsymbol{V}_k) \leq \mathfrak{v}, \forall k \tag{178}$$

thus we finish the proof. □

Lemma 17. *Let $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ be updated by Eq.* (1)*, if $\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$, then we have that*

$$\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F. \tag{179}$$

Proof. Based on the update Eq. (1), we have

$$\begin{aligned}\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 &= (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 + 2\eta(1-\eta)\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{U}_k(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\right\rangle \\ &\quad + \eta^2\|\boldsymbol{U}_k(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2\end{aligned} \tag{180}$$

Meanwhile, according to Lemma 13, we have that $\|\boldsymbol{U}_k(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2 \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$, together with $\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$ the following holds

$$\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$$

On the other hand,

$$\begin{aligned}\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2 &= (1-\eta)^2\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 + 2\eta(1-\eta)\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\right\rangle \\ &\quad + \eta^2\|\boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2\end{aligned} \tag{181}$$

and $\|\boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2 \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$, $\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\right\rangle \geq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$ we have the result in Eq. (179). □

Lemma 18. *Let $\tau_k = \frac{\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2}$, if $1 \geq \tau_k \geq 1/2$, then we have that*

$$\frac{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F} \leq \frac{1}{(1-\eta+\eta\tau_k)^2} \tag{182}$$

Proof. Note that

$$\begin{aligned}\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 &= (1-\eta)^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2 + 2\eta(1-\eta)\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{U}_k(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\right\rangle \\ &\quad + \eta^2\|\boldsymbol{U}_k(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2\end{aligned} \tag{183}$$

thus

$$\begin{aligned}\frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2} &= (1-\eta)^2 + 2\eta(1-\eta)\frac{\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2} + \eta^2\frac{\|\boldsymbol{U}_k(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2} \\ &= (1-\eta)^2 + 2\eta(1-\eta)\tau_k + \eta^2\frac{\|\mathcal{U}_k\mathcal{U}_k^\top\boldsymbol{M}\|_F^2\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^4} \\ &\geq (1-\eta)^2 + 2\eta(1-\eta)\tau_k + \eta^2\frac{\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\rangle}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2} \\ &\geq (1-\eta)^2 + 2\eta(1-\eta)\tau_k + \eta^2\tau_k^2 \\ &= (1-\eta+\eta\tau_k)^2\end{aligned} \tag{184}$$

In the same vein

$$\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2 = (1-\eta)^2\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2 + 2\eta(1-\eta)\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\right\rangle \\ + \eta^2\|\boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2 \tag{185}$$

If $\langle \boldsymbol{U}_k\boldsymbol{V}_k, \boldsymbol{M}\rangle \leq \|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2$ we know that the norm of $\boldsymbol{U}_k\boldsymbol{V}_k^\top$ decreases and $\|\boldsymbol{U}_{k+t}\boldsymbol{V}_k^\top\|_F^2 \geq \|\boldsymbol{U}_{k+t}\boldsymbol{V}_{k+1}^\top\|_F^2$, therefore

$$\begin{aligned}\frac{\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\right\rangle}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2} &\geq \frac{\left\langle \boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top, \boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\right\rangle}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2} \\ &\geq \frac{\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \mathcal{U}_k\mathcal{U}_k^\top\boldsymbol{M}\right\rangle}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2} = \tau_k\end{aligned} \tag{186}$$

and similarly, we have

$$\begin{aligned}\frac{\|\boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2} &\geq \frac{\|\boldsymbol{U}_{k+1}(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2} \\ &\geq \frac{\|\boldsymbol{U}_k(\boldsymbol{U}_k^\top\boldsymbol{U}_k)^{-1}\boldsymbol{U}_k^\top\boldsymbol{M}\|_F^2}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^2} \\ &\geq \frac{\left\langle \boldsymbol{U}_k\boldsymbol{V}_k^\top, \boldsymbol{M}\right\rangle^2}{\|\boldsymbol{U}_k\boldsymbol{V}_k^\top\|_F^4} \geq \tau_k^2\end{aligned} \tag{187}$$

In consequence, we have the result

$$\frac{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_{k+1}^\top\|_F^2}{\|\boldsymbol{U}_{k+1}\boldsymbol{V}_k^\top\|_F^2} \geq (1-\eta)^2 + 2\eta(1-\eta)\tau_k + \eta^2\tau_k^2 = (1-\eta+\eta\tau_k)^2. \tag{188}$$

Together with Eq. (184) and Eq. (188), we have the result in Eq. (182). □