Let $\mathcal{H} \in \mathbb{R}^{m\times n\times m}$ be the ground truth tensor of n different Hessian matrices of the dimension $m \times m$, where the lateral slices represent Hessian matrices generated from the chemical reaction systems. We assume the ground truth tensor $\mathcal{H}$ admits the following Tucker decomposition, for $1 \le i_1, i_3 \le m$ and $1 \le i_2 \le n$,

$$\mathcal{H}(i_1,i_2,i_3) = \sum_{j_1=1}^{r_1}\sum_{j_2=1}^{r_2}\sum_{j_3=1}^{r_3} \mathbf{X}_1(i_1,j_1)\,\mathbf{X}_2(i_2,j_2)\mathbf{X}_3(i_3,j_3)\,\mathcal{G}(j_1,j_2,j_3), \tag{1}$$

namely, $\mathcal{H} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\cdot\mathcal{G}$, where $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times r_3}$ is the core tensor of multilinear rank $\mathbf{r} = (r_1, r_2, r_3)$, $r_k = \mathrm{rank}(\mathbf{H}_{(k)})$ for $k = 1, 2, 3.$, and $\mathbf{X}_1 \in \mathbb{R}^{m\times r_1}$, $\mathbf{X}_2 \in \mathbb{R}^{n\times r_2}$, $\mathbf{X}_3 \in \mathbb{R}^{m\times r_3}$ are factor matrices of each mode.

Several perturbation bounds. We now introduce several perturbation bounds that will be used repeatedly in the proof. Without loss of generality, assume that $\hat{\mathbf{F}} = (\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3, \hat{\mathcal{G}})$ and $\mathbf{F} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathcal{G})$ are aligned, and introduce the following notation that will be used repeatedly:

$$\Delta_{\hat{\mathbf{X}}_1} := \hat{\mathbf{X}}_1 - \mathbf{X}_1, \qquad \Delta_{\hat{\mathbf{X}}_2} := \hat{\mathbf{X}}_2 - \mathbf{X}_2, \qquad \Delta_{\hat{\mathbf{X}}_3} := \hat{\mathbf{X}}_3 - \mathbf{X}_3 \tag{2}$$

$$\breve{\mathbf{X}}_1 := \left(\hat{\mathbf{X}}_2 \otimes \hat{\mathbf{X}}_3\right)\mathcal{M}_1(\hat{\mathcal{G}})^\top, \quad \breve{\mathbf{X}}_2 := \left(\hat{\mathbf{X}}_2 \otimes \hat{\mathbf{X}}_1\right)\mathcal{M}_2(\hat{\mathcal{G}})^\top, \quad \breve{\mathbf{X}}_3 := \left(\hat{\mathbf{X}}_3 \otimes \hat{\mathbf{X}}_1\right)\mathcal{M}_3(\hat{\mathcal{G}})^\top \tag{3}$$

$$\bar{\mathbf{X}}_1 := (\mathbf{X}_2 \otimes \mathbf{X}_3)\,\mathcal{M}_1(\mathcal{G})^\top, \quad \bar{\mathbf{X}}_2 := (\mathbf{X}_2 \otimes \mathbf{X}_1)\,\mathcal{M}_2(\mathcal{G})^\top, \quad \bar{\mathbf{X}}_3 := (\mathbf{X}_3 \otimes \mathbf{X}_1)\,\mathcal{M}_3(\mathcal{G})^\top \tag{4}$$

$$\mathcal{T}_{\hat{\mathbf{X}}_1} := \left(\mathbf{X}_1^\top\Delta_{\hat{\mathbf{X}}_1}, \mathbf{I}_{r_2}, \mathbf{I}_{r_3}\right)\cdot\mathcal{G}, \quad \mathcal{T}_{\hat{\mathbf{X}}_2} := \left(\mathbf{I}_{r_1}, \mathbf{X}_2^\top\Delta_{\hat{\mathbf{X}}_2}, \mathbf{I}_{r_3}\right)\cdot\mathcal{G}, \quad \mathcal{T}_{\hat{\mathbf{X}}_3} := \left(\mathbf{I}_{r_1}, \mathbf{I}_{r_2}, \mathbf{X}_3^\top\Delta_{\hat{\mathbf{X}}_3}\right)\cdot\mathcal{G}. \tag{5}$$

To begin, we define the scaled distance between $\hat{\mathbf{F}} = (\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3, \hat{\mathcal{G}})$ and $\mathbf{F} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathcal{G})$ as:

$$\mathrm{dist}^2(\hat{\mathbf{F}},\mathbf{F}) := \inf_{\mathbf{Q}_k\in\mathrm{GL}(r_k)} \left\|(\hat{\mathbf{X}}_1\mathbf{Q}_1 - \mathbf{X}_1)\mathbf{\Sigma}_1\right\|_F^2 + \left\|(\hat{\mathbf{X}}_2\mathbf{Q}_2 - \mathbf{X}_2)\mathbf{\Sigma}_2\right\|_F^2 + \left\|(\hat{\mathbf{X}}_3\mathbf{Q}_3 - \mathbf{X}_3)\mathbf{\Sigma}_3\right\|_F^2 + \left\|(\mathbf{Q}_1^{-1}, \mathbf{Q}_2^{-1}, \mathbf{Q}_3^{-1})\cdot\hat{\mathcal{G}} - \mathcal{G}\right\|_F^2, \tag{6}$$

where we call the matrices $\{\mathbf{Q}_k\}_{k=1,2,3}$ that attain the infimum the optimal alignment matrices between $\hat{\mathbf{F}}$ and $\mathbf{F}$; in particular, $\hat{\mathbf{F}}$ and $\mathbf{F}$ are said to be aligned if the optimal alignment matrices are identity matrices. The core tensor $\mathcal{G}$ is related to the singular values in each mode as $\mathbf{G}_{(k)}\mathbf{G}_{(k)}^\top = \mathbf{\Sigma}_k^2$, $k = 1, 2, 3$, where $\mathbf{\Sigma}_k := \mathrm{diag}\left[\sigma_1(\mathbf{H}_{(k)}), \ldots, \sigma_{r_k}(\mathbf{H}_{(k)})\right]$ is a diagonal matrix where the diagonal elements are composed of the nonzero singular values of $\mathbf{H}_{(k)}$.

Following basic relations, which follow straightforwardly from analogous matrix relations after applying matricizations, will be proven useful:

$$(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\cdot((\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)\cdot\mathcal{G}) = (\mathbf{X}_1\mathbf{Q}_1, \mathbf{X}_2\mathbf{Q}_2, \mathbf{X}_3\mathbf{Q}_3)\cdot\mathcal{G} \tag{7}$$

$$\langle(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\cdot\mathcal{G}, \mathcal{H}\rangle = \left\langle\mathcal{G}, \left(\mathbf{X}_1^\top, \mathbf{X}_2^\top, \mathbf{X}_3^\top\right)\cdot\mathcal{H}\right\rangle, \tag{8}$$

$$\|(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)\cdot\mathcal{G}\|_F \le \|\mathbf{Q}_1\|\,\|\mathbf{Q}_2\|\,\|\mathbf{Q}_3\|\,\|\mathcal{G}\| \tag{9}$$

In [1], it is proven the following bound holds regarding the Frobenius norm when $\epsilon \le 0.2$:

$$\left\|\left(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}} - \mathcal{H}\right\|_F \le 3\,\mathrm{dist}(\hat{\mathbf{F}},\mathbf{F}). \tag{10}$$

Hence, we serve the scaled distance in (6) as a metric to gauge the quality of the tensor recovery in our paper.

0.1. Useful Lemmas

In what follows, we provide several useful lemmas whose proof can be found in [1]. We start with a lemma that ensures the attainability of the infimum in the definition in (6) as long as $\mathrm{dist}(\hat{\mathbf{F}},\mathbf{F})$ is sufficiently small.

Lemma 1. *Fix any factor quadruple $\hat{\mathbf{F}} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathcal{G})$. Suppose that dist$(\hat{\mathbf{F}},\mathbf{F}) < \sigma_{min}(\mathcal{H})$, then the infimum of* (6) *is attained at some $\mathbf{Q}_k \in GL(r_k)$, i.e., the alignment matrices between $\hat{\mathbf{F}}$ and $\mathbf{F}$ exist.*

With the existence of the optimal alignment matrices in place, the following lemma delineates the optimality conditions they need to satisfy.

Lemma 2. *The optimal alignment matrices* $\{\mathbf{Q}_k\}_{k=1,2,3}$ *between* $\hat{\mathbf{F}}$ *and* $\mathbf{F}$*, if exist, must satisfy*

$$(\hat{\mathbf{X}}_1\mathbf{Q}_1)^\top(\hat{\mathbf{X}}_1\mathbf{Q}_1 - \mathbf{X}_1)\mathbf{\Sigma}_1^2 = \mathcal{M}_1\left((\mathbf{Q}_1^{-1},\mathbf{Q}_2^{-1},\mathbf{Q}_3^{-1})\cdot\hat{\mathcal{G}} - \mathcal{G}\right)\mathcal{M}_1\left((\mathbf{Q}_1^{-1},\mathbf{Q}_2^{-1},\mathbf{Q}_3^{-1})\cdot\mathcal{G}\right)^\top, \tag{11}$$

$$(\hat{\mathbf{X}}_2\mathbf{Q}_2)^\top(\hat{\mathbf{X}}_2\mathbf{Q}_2 - \mathbf{X}_2)\mathbf{\Sigma}_2^2 = \mathcal{M}_2\left((\mathbf{Q}_1^{-1},\mathbf{Q}_2^{-1},\mathbf{Q}_3^{-1})\cdot\hat{\mathcal{G}} - \mathcal{G}\right)\mathcal{M}_2\left((\mathbf{Q}_1^{-1},\mathbf{Q}_2^{-1},\mathbf{Q}_3^{-1})\cdot\mathcal{G}\right)^\top, \tag{12}$$

$$(\hat{\mathbf{X}}_3\mathbf{Q}_3)^\top(\hat{\mathbf{X}}_3\mathbf{Q}_3 - \mathbf{X}_3)\mathbf{\Sigma}_3^2 = \mathcal{M}_3\left((\mathbf{Q}_1^{-1},\mathbf{Q}_2^{-1},\mathbf{Q}_3^{-1})\cdot\hat{\mathcal{G}} - \mathcal{G}\right)\mathcal{M}_3\left((\mathbf{Q}_1^{-1},\mathbf{Q}_2^{-1},\mathbf{Q}_3^{-1})\cdot\mathcal{G}\right)^\top. \tag{13}$$

Lemma 3. *Suppose* $\hat{\mathbf{F}} = (\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3,\hat{\mathcal{G}})$ *and* $\mathbf{F} = (\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\mathcal{G})$ *are aligned and satisfy dist*$(\hat{\mathbf{F}},\mathbf{F}) \le \epsilon\sigma_{min}(\mathcal{H})$ *for some* $\epsilon < 1$*. Then, the following bounds hold regarding the spectral norm:*

$$\left\|\Delta_{\hat{\mathbf{X}}_1}\right\| \vee \left\|\Delta_{\hat{\mathbf{X}}_2}\right\| \vee \left\|\Delta_{\hat{\mathbf{X}}_3}\right\| \vee \left\|\mathcal{M}_k\left(\Delta_{\hat{\mathcal{G}}}\right)^\top\mathbf{\Sigma}_k^{-1}\right\| \le \epsilon, \quad k=1,2,3 \tag{14}$$

Now, we prove Theorem 1 via recursion. Recall Theorem 1 in the main paper.

Theorem 1.

Suppose that for some $t \ge 0$, one has dist$(\hat{\mathbf{F}}_t,\mathbf{F}) \le \epsilon\sigma_{min}(\mathcal{H})$ for some sufficiently small ϵ whose size will be specified later in the proof. The goal is to bound the scaled distance from the ground truth to the next iterate, i.e., dist$(\hat{\mathbf{F}}_{t+1},\mathbf{F})$. Since dist$(\hat{\mathbf{F}}_t,\mathbf{F}) \le \epsilon\sigma_{min}(\mathcal{H})$, Lemma 1 ensures that the optimal alignment matrices $\{\mathbf{Q}_{t,k}\}_{k=1,2,3}$ between $\hat{\mathbf{F}}_t$ and $\mathbf{F}$ exist. Therefore, in view of the definition of dist$(\hat{\mathbf{F}}_{t+1},\mathbf{F})$, one has

$$\begin{aligned}\text{dist}^2(\hat{\mathbf{F}}_{t+1},\mathbf{F}) \le& \left\|(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1}-\mathbf{X}_1)\mathbf{\Sigma}_1\right\|_F^2 + \left\|(\hat{\mathbf{X}}_{t+1,2}\mathbf{Q}_{t,2}-\mathbf{X}_2)\mathbf{\Sigma}_2\right\|_F^2 + \left\|(\hat{\mathbf{X}}_{t+1,3}\mathbf{Q}_{t,3}-\mathbf{X}_3)\mathbf{\Sigma}_3\right\|_F^2 \\ &+ \left\|(\mathbf{Q}_{t,1}^{-1},\mathbf{Q}_{t,2}^{-1},\mathbf{Q}_{t,3}^{-1})\cdot\hat{\mathcal{G}}_{t+1}-\mathcal{G}\right\|_F^2.\end{aligned} \tag{15}$$

To avoid notational clutter, we denote $\hat{\mathbf{F}} := (\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\mathcal{G})$ with

$$\hat{\mathbf{X}}_1 := \hat{\mathbf{X}}_{t,1}\mathbf{Q}_{t,1}, \quad \hat{\mathbf{X}}_2 := \hat{\mathbf{X}}_{t,2}\mathbf{Q}_{t,2}, \quad \hat{\mathbf{X}}_3 := \hat{\mathbf{X}}_{t,3}\mathbf{Q}_{t,3}, \quad \hat{\mathcal{G}} := \left(\mathbf{Q}_{t,1}^{-1},\mathbf{Q}_{t,2}^{-1},\mathbf{Q}_{t,3}^{-1}\right)\cdot\mathcal{G}_t \tag{16}$$

and adopt the set of notation defined in **??** for the rest of the proof. Clearly, $\hat{\mathbf{F}}$ is aligned with $\mathbf{F}$. With these notation, we can rephrase the consequences of Lemma 2 as:

$$\hat{\mathbf{X}}_1^\top\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1^2 = \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\mathcal{M}_1\left(\hat{\mathcal{G}}\right)^\top, \tag{17}$$

$$\hat{\mathbf{X}}_2^\top\Delta_{\hat{\mathbf{X}}_2}\mathbf{\Sigma}_2^2 = \mathcal{M}_2\left(\Delta_{\hat{\mathcal{G}}}\right)\mathcal{M}_2\left(\hat{\mathcal{G}}\right)^\top, \tag{18}$$

$$\hat{\mathbf{X}}_3^\top\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_3^2 = \mathcal{M}_3\left(\Delta_{\hat{\mathcal{G}}}\right)\mathcal{M}_3\left(\hat{\mathcal{G}}\right)^\top. \tag{19}$$

Proof of Theorem 1. Utilize the gradient descent update rule to write

$$\left(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1}-\mathbf{X}_1\right)\mathbf{\Sigma}_1 \tag{20}$$

$$= \left(\hat{\mathbf{X}}_1 - \eta\mathcal{M}_1\left(2\alpha\left(\left(\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}} - \left(\mathcal{H}+\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)\right) + 2\beta\left(\left(\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}} - \left(\mathcal{H}+\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right)\right)\check{\mathbf{X}}_1 - \mathbf{X}_1\right)\mathbf{\Sigma}_1, \tag{21}$$

where we use the decomposition of the mode-1 matricization

$$\mathcal{M}_1\left(2\alpha\left(\left(\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}} - \left(\mathcal{H}+\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)\right) + 2\beta\left(\left(\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}} - \left(\mathcal{H}+\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right)\right) \tag{22}$$

$$= 2\alpha\mathcal{M}_1\left(\left(\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}} - \left(\mathcal{H}+\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)\right) + 2\beta\mathcal{M}_1\left(\left(\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}} - \left(\mathcal{H}+\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right) \tag{23}$$

$$= 2\,\hat{\mathbf{X}}_1\mathcal{M}_1\left(\hat{\mathcal{G}}\right)\left(\hat{\mathbf{X}}_2\otimes\hat{\mathbf{X}}_3\right)^\top - 2\mathbf{X}_1\mathcal{M}_1\left(\mathcal{G}\right)\left(\mathbf{X}_2\otimes\mathbf{X}_3\right)^\top - 2\alpha\,\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) - 2\beta\,\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right) \tag{24}$$

$$= \Delta_{\hat{\mathbf{X}}_1}\check{\mathbf{X}}_1^\top + \mathbf{X}_1\left(\check{\mathbf{X}}_1 - \bar{\mathbf{X}}_1\right)^\top - 2\alpha\,\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) - 2\beta\,\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right) \tag{25}$$

By plugging (25) into (21), we have

$$
\begin{aligned}
&\left(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1}-\mathbf{X}_1\right)\mathbf{\Sigma}_1 && (26)\\
&=\left(\hat{\mathbf{X}}_1-\eta\mathcal{M}_1\left(2\alpha\left(\left(\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}}-\left(\mathcal{H}+\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)\right)+2\beta\left(\left(\hat{\mathbf{X}}_1,\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_3\right)\cdot\hat{\mathcal{G}}-\left(\mathcal{H}+\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right)\right)\check{\mathbf{X}}_1-\mathbf{X}_1\right)\mathbf{\Sigma}_1 && (27)\\
&=\left(\hat{\mathbf{X}}_1-\eta\left(\Delta_{\hat{\mathbf{X}}_1}\check{\mathbf{X}}_1^{\mathsf{T}}+\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}-2\alpha\,\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)-2\beta\,\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right)\check{\mathbf{X}}_1-\mathbf{X}_1\right)\mathbf{\Sigma}_1 && (28)\\
&=(1-\eta)\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1-\eta\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1+2\alpha\eta\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)\check{\mathbf{X}}_1\mathbf{\Sigma}_1+2\beta\eta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\check{\mathbf{X}}_1\mathbf{\Sigma}_1. && (29)
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\left\|\left(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1}-\mathbf{X}_1\right)\mathbf{\Sigma}_1\right\| && (30)\\
&=\left\|(1-\eta)\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1-\eta\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1+2\alpha\eta\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)\check{\mathbf{X}}_1\mathbf{\Sigma}_1+2\beta\eta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\| && (31)\\
&\leq\underbrace{\left\|(1-\eta)\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1-\eta\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|}_{\odot}+\left\|2\alpha\eta\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)\check{\mathbf{X}}_1\mathbf{\Sigma}_1+2\beta\eta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\| && (32)
\end{aligned}
$$

Let $\odot := (1-\eta)\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1-\eta\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1$. Taking squared on each side,

$$
\begin{aligned}
&\left\|\left(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1}-\mathbf{X}_1\right)\mathbf{\Sigma}_1\right\|^2 && (33)\\
&\leq\|\odot\|^2+2\left\langle\odot,2\eta\left(\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)+\beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right)\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle+\left\|2\eta\left(\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)+\beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right)\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2 && (34)\\
&\leq\|\odot\|^2+4\eta\|\odot\|\left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)+\beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right\|\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|+4\eta^2\left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right)+\beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right\|^2\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2. && (35)
\end{aligned}
$$

Now, we bound $\odot=(1-\eta)\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1-\eta\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1$.

$$
\begin{aligned}
\|\odot\|^2&=\left\|(1-\eta)\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1-\eta\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2 && (36)\\
&=(1-\eta)^2\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|^2-\eta(1-\eta)\underbrace{\left\langle\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1,\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle}_{\odot_1}+\eta^2\underbrace{\left\|\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2}_{\odot_2}. && (37)
\end{aligned}
$$

$\odot_1$ is expressed as

$$
\begin{aligned}
\odot_1&=\left\langle\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1,\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle && (38)\\
&=\left\langle\mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1,\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle && (39)\\
&=\underbrace{\left\langle\mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1,\mathcal{M}_1\left(\mathcal{G}\right)\left(\hat{\mathbf{X}}_2\otimes\Delta_{\hat{\mathbf{X}}_3}+\Delta_{\hat{\mathbf{X}}_2}\otimes\mathbf{X}_3\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle}_{\odot_{1,1}}+\underbrace{\left\langle\mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1,\mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\left(\hat{\mathbf{X}}_2\otimes\hat{\mathbf{X}}_3\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle}_{\odot_{1,2}}. && (40)
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
\odot_{1,1}&=\left\langle\mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1,\mathcal{M}_1\left(\mathcal{G}\right)\left(\hat{\mathbf{X}}_2\otimes\Delta_{\hat{\mathbf{X}}_3}+\Delta_{\hat{\mathbf{X}}_2}\otimes\mathbf{X}_3\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle && (41)\\
&=\underbrace{\left\langle\mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1,\mathcal{M}_1\left(\mathcal{G}\right)\left(\mathbf{X}_2\otimes\Delta_{\mathbf{X}_3}+\Delta_{\hat{\mathbf{X}}_2}\otimes\mathbf{X}_3\right)^{\mathsf{T}}\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right\rangle}_{\odot_{1,1,1}}+\underbrace{\left\langle\mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1,\mathcal{M}_1\left(\mathcal{G}\right)\left(\mathbf{X}_2\otimes\Delta_{\hat{\mathbf{X}}_3}\right)^{\mathsf{T}}\left(\check{\mathbf{X}}_1\mathbf{\Sigma}_1-\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right)\right\rangle}_{\odot_{1,1,2}} && (42)
\end{aligned}
$$

$$+\underbrace{\left\langle \mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\mathcal{G}\right)\left(\Delta_{\hat{\mathbf{X}}_2}\otimes\mathbf{X}_3\right)^{\mathsf{T}}\left(\check{\mathbf{X}}_1\mathbf{\Sigma}_1-\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right)\right\rangle}_{\odot_{1,1,3}}+\underbrace{\left\langle \mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\mathcal{G}\right)\left(\Delta_{\hat{\mathbf{X}}_2}\otimes\Delta_{\hat{\mathbf{X}}_3}\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle}_{\odot_{1,1,4}}. \tag{43}$$

Utilizing the definition of $\bar{\mathbf{X}}_1$, $\odot_{1,1,1}$ can be rewritten as an inner product in the tensor space:

$$\odot_{1,1,1}=\left\langle \mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\mathcal{G}\right)\left(\mathbf{X}_2\otimes\Delta_{\mathbf{X}_3}+\Delta_{\hat{\mathbf{X}}_2}\otimes\mathbf{X}_3\right)^{\mathsf{T}}\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right\rangle \tag{44}$$

$$=\left\langle \mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathcal{M}_1\left(\mathcal{G}\right), \mathcal{M}_1\left(\mathcal{G}\right)\left(\mathbf{I}_{r_3}\otimes\Delta_{\hat{\mathbf{X}}_3}^{\mathsf{T}}\mathbf{X}_3+\Delta_{\hat{\mathbf{X}}_2}^{\mathsf{T}}\mathbf{X}_2\otimes\mathbf{I}_{r_2}\right)\right\rangle \tag{45}$$

$$=\left\langle \mathcal{T}_{\hat{\mathbf{X}}_1}, \mathcal{T}_{\hat{\mathbf{X}}_3}+\mathcal{T}_{\hat{\mathbf{X}}_2}\right\rangle \tag{46}$$

To control the other three terms, $\odot_{1,1,2},\odot_{1,1,3},\odot_{1,1,4}$, Lemma 3 turns out to be useful. For example, $\odot_{1,1,2}$ is bounded by

$$|\odot_{1,1,2}|\le\left\|\mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\mathcal{M}_1\left(\mathcal{G}\right)\left(\mathbf{X}_2\otimes\Delta_{\hat{\mathbf{X}}_3}\right)^{\mathsf{T}}\right\|_F\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1-\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right\| \tag{47}$$

$$\le\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1-\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right\|, \tag{48}$$

we bound,

$$\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1-\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right\|=\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1-\check{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}+\check{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}-\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right\| \tag{49}$$

$$\le\left\|\check{\mathbf{X}}_1\left(\mathbf{\Sigma}_1-\mathbf{\Sigma}_1^{-1}\right)\right\|+\left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)\mathbf{\Sigma}_1^{-1}\right\| \tag{50}$$

$$\le\left(3\epsilon+3\epsilon^2+\epsilon^3\right)+\left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)\mathbf{\Sigma}_1^{-1}\right\|. \tag{51}$$

Therefore, $\odot_{1,1,2}$ is bounded as

$$|\odot_{1,1,2}|\le\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1-\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right\| \tag{52}$$

$$\le\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\left(\left(3\epsilon+3\epsilon^2+\epsilon^3\right)+\left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)\mathbf{\Sigma}_1^{-1}\right\|\right). \tag{53}$$

In a similar way,

$$|\odot_{1,1,3}|\le\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_2}\mathbf{\Sigma}_3\right\|_F\left(\left(3\epsilon+3\epsilon^2+\epsilon^3\right)+\left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)\mathbf{\Sigma}_1^{-1}\right\|\right) \tag{54}$$

$$|\odot_{1,1,4}|\le\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_F. \tag{55}$$

Therefore, combining (46), (53), (54) and (55), $\odot_{1,1}$ is bounded as

$$\odot_{1,1}\le\left\langle \mathcal{T}_{\hat{\mathbf{X}}_1}, \mathcal{T}_{\hat{\mathbf{X}}_3}+\mathcal{T}_{\hat{\mathbf{X}}_2}\right\rangle+\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\left(\left(3\epsilon+3\epsilon^2+\epsilon^3\right)+\left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)\mathbf{\Sigma}_1^{-1}\right\|\right) \tag{56}$$

$$+\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_2}\mathbf{\Sigma}_3\right\|_F\left(\left(3\epsilon+3\epsilon^2+\epsilon^3\right)+\left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)\mathbf{\Sigma}_1^{-1}\right\|\right)+\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_F. \tag{57}$$

Next, $\odot_{1,2}$ can be rewritten as

$$\odot_{1,2}=\left\langle \mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\left(\hat{\mathbf{X}}_2\otimes\hat{\mathbf{X}}_3\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle \tag{58}$$

$$=\left\langle \mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\mathcal{M}_1\left(\mathcal{G}\right)^{\mathsf{T}}\mathbf{\Sigma}_1^{-1}\right\rangle+\underbrace{\left\langle \mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\left(\mathbf{X}_2\otimes\mathbf{X}_3\right)^{\mathsf{T}}\left(\check{\mathbf{X}}_1\mathbf{\Sigma}_1-\bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right)\right\rangle}_{\odot_{1,2,1}} \tag{59}$$

$$+\underbrace{\left\langle \mathbf{X}_1^{\mathsf{T}}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\left(\hat{\mathbf{X}}_2\otimes\hat{\mathbf{X}}_3-\mathbf{X}_2\otimes\mathbf{X}_3\right)^{\mathsf{T}}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle}_{\odot_{1,2,2}} \tag{60}$$

$$= \left\langle \mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\mathcal{M}_1\left(\hat{\mathcal{G}}\right)^\mathsf{T}\mathbf{\Sigma}_1^{-1}\right\rangle \underbrace{-\left\langle \mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)^\mathsf{T}\mathbf{\Sigma}_1^{-1}\right\rangle}_{\odot_{1,2,3}} + \odot_{1,2,1} + \odot_{1,2,2} \tag{61}$$

$$= \left\langle \mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \hat{\mathbf{X}}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\rangle + \odot_{1,2,1} + \odot_{1,2,2} + \odot_{1,2,3} \tag{62}$$

$$= \left\|\mathcal{T}_{\hat{\mathbf{X}}_1}\right\|_F^2 + \odot_{1,2,1} + \odot_{1,2,2} + \odot_{1,2,3} + \underbrace{\left\langle \mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \Delta_{\hat{\mathbf{X}}_1}^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\rangle}_{\odot_{1,2,4}}. \tag{63}$$

Now, we bound $\odot_{1,2,1}$, $\odot_{1,2,2}$, $\odot_{1,2,3}$ and $\odot_{1,2,4}$.

$$\odot_{1,2,1} = \left\langle \mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\left(\mathbf{X}_2\otimes\mathbf{X}_3\right)^\mathsf{T}\left(\check{\mathbf{X}}_1\mathbf{\Sigma}_1 - \bar{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right)\right\rangle \tag{64}$$

$$\leq \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F \left((3\epsilon + 3\epsilon^2 + \epsilon^3) + \left\|\check{\mathbf{X}}_1\left(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^{-1}\right)\right\|\right) \tag{65}$$

$$\odot_{1,2,2} = \left\langle \mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\left(\hat{\mathbf{X}}_2\otimes\hat{\mathbf{X}}_3 - \mathbf{X}_2\otimes\mathbf{X}_3\right)^\mathsf{T}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\rangle \tag{66}$$

$$\leq \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F (2\epsilon + \epsilon^2)\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\| \tag{67}$$

$$\odot_{1,2,3} = \left\langle \mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1, \mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)^\mathsf{T}\mathbf{\Sigma}_1^{-1}\right\rangle \tag{68}$$

$$\leq \left\|\mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\right\|_F \left\|\mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)^\mathsf{T}\mathbf{\Sigma}_1^{-1}\right\|_F \tag{69}$$

$$\leq \epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F \tag{70}$$

$$\odot_{1,2,4} \leq \left\|\mathbf{X}_1^\mathsf{T}\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathbf{X}}_1}\right\|_F \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \leq \epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F^2 \tag{71}$$

To this end,

$$\odot_{1,2} \leq \left\|\mathcal{T}_{\hat{\mathbf{X}}_1}\right\|_F^2 + \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F \left((3\epsilon + 3\epsilon^2 + \epsilon^3) + \left\|\check{\mathbf{X}}_1\left(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^{-1}\right)\right\|\right) \tag{72}$$

$$+ \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F (2\epsilon + \epsilon^2)\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\| + \epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F + \epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F^2. \tag{73}$$

To this end, $\odot_1$ is bounded with and as

$$\odot_1 = \odot_{1,1} + \odot_{1,2} \tag{74}$$

$$= \left\langle \mathcal{T}_{\hat{\mathbf{X}}_1}, \mathcal{T}_{\hat{\mathbf{X}}_3} + \mathcal{T}_{\hat{\mathbf{X}}_2}\right\rangle + \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F \left((3\epsilon + 3\epsilon^2 + \epsilon^3) + \left\|\left(\check{\mathbf{X}}_1 - \bar{\mathbf{X}}_1\right)\mathbf{\Sigma}_1^{-1}\right\|\right) \tag{75}$$

$$+ \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathbf{X}}_2}\mathbf{\Sigma}_3\right\|_F \left((3\epsilon + 3\epsilon^2 + \epsilon^3) + \left\|\left(\check{\mathbf{X}}_1 - \bar{\mathbf{X}}_1\right)\mathbf{\Sigma}_1^{-1}\right\|\right) + \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F \left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_F \tag{76}$$

$$+ \left\|\mathcal{T}_{\hat{\mathbf{X}}_1}\right\|_F^2 + \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F \left((3\epsilon + 3\epsilon^2 + \epsilon^3) + \left\|\check{\mathbf{X}}_1\left(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^{-1}\right)\right\|\right) + \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F (2\epsilon + \epsilon^2)\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|$$

$$+ \epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F + \epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F^2 \tag{77}$$

$$= \left\langle \mathcal{T}_{\hat{\mathbf{X}}_1}, \mathcal{T}_{\hat{\mathbf{X}}_1} + \mathcal{T}_{\hat{\mathbf{X}}_2} + \mathcal{T}_{\hat{\mathbf{X}}_3}\right\rangle + (3\epsilon + 3\epsilon^2 + \epsilon^3)\left(1 + \left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2\right)\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F \tag{78}$$

$$+ (3\epsilon + 3\epsilon^2 + \epsilon^3)\left(1 + \left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2\right)\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathbf{X}}_2}\mathbf{\Sigma}_3\right\|_F + \left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F \left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\| \tag{79}$$

$$+ (3\epsilon + 3\epsilon^2 + \epsilon^3)\left(1 + \left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2\right)\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F + (2\epsilon + \epsilon^2)\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F + \epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F \left\|\Delta_{\hat{\mathcal{G}}}\right\|_F + \epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F^2 \tag{80}$$

Now, let us obtain bound for $\odot_2$.

$$\sqrt{\odot_2} = \left\|\mathbf{X}_1\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^\mathsf{T}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_2 \tag{81}$$

$$= \left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)^\mathsf{T}\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_2 \tag{82}$$

$$\leq \left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)\right\|_F\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_2 \tag{83}$$

$\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_2$ is bounded as

$$\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_2 = \frac{1}{\sigma_{min}(\check{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1})} \tag{84}$$

And,

$$\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1}\right\| \leq \left\|\hat{\mathbf{X}}_2\right\|\left\|\hat{\mathbf{X}}_3\right\|\left\|\mathcal{M}_1\left(\hat{\mathcal{G}}\right)\mathbf{\Sigma}_1^{-1}\right\| \tag{85}$$

$$\leq \left\|\hat{\mathbf{X}}_2\right\|\left\|\hat{\mathbf{X}}_3\right\|\left\|\left(\mathcal{M}_1\left(\mathcal{G}\right)-\mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\right)\mathbf{\Sigma}_1^{-1}\right\| \tag{86}$$

$$\leq \left\|\hat{\mathbf{X}}_2\right\|\left\|\hat{\mathbf{X}}_3\right\|\left(\left\|\mathcal{M}_1\left(\mathcal{G}\right)\mathbf{\Sigma}_1^{-1}\right\|+\left\|\mathcal{M}_1\left(\Delta_{\hat{\mathcal{G}}}\right)\mathbf{\Sigma}_1^{-1}\right\|\right) \tag{87}$$

$$= (1+\epsilon)^3 \tag{88}$$

We have

$$\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_2 = \frac{1}{\sigma_{min}(\check{\mathbf{X}}_1\mathbf{\Sigma}_1^{-1})} \leq \frac{1}{(1+\epsilon)^3} \tag{89}$$

Plugging (89) into (83), we have

$$\sqrt{\odot_2} \leq (1+\epsilon)^{-3}\left\|\left(\check{\mathbf{X}}_1-\bar{\mathbf{X}}_1\right)\right\|_F \tag{90}$$

$$\leq (1+\epsilon)^{-3}(1+\epsilon+\frac{\epsilon^2}{3})\left(\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F+\left\|\Delta_{\hat{\mathbf{X}}_2}\mathbf{\Sigma}_3\right\|_F+\left\|\Delta_{\hat{\mathcal{G}}}\right\|_F\right) \tag{91}$$

$$= (1+\epsilon)^{-3}(1+\epsilon+\frac{\epsilon^2}{3})\left(\mathrm{dist}(\hat{\mathbf{F}}_t,\mathbf{F})-\left\|\Delta_{\hat{\mathbf{X}}_1\mathbf{\Sigma}_1}\right\|_F\right) \tag{92}$$

$$\leq (1+\epsilon)^{-3}(1+\epsilon+\frac{\epsilon^2}{3})\mathrm{dist}(\hat{\mathbf{F}}_t,\mathbf{F}). \tag{93}$$

Finally, we have

$$\odot_2 \leq (1+\epsilon)^{-6}(1+\epsilon+\frac{\epsilon^2}{3})^2\mathrm{dist}^2(\hat{\mathbf{F}}_t,\mathbf{F}). \tag{94}$$

To this end, we have

$$\|\odot\|^2 \leq (1-\eta)^2\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|^2-\eta(1-\eta)\left(\left\langle\mathcal{T}_{\hat{\mathbf{X}}_1},\mathcal{T}_{\hat{\mathbf{X}}_1}+\mathcal{T}_{\hat{\mathbf{X}}_2}+\mathcal{T}_{\hat{\mathbf{X}}_3}\right\rangle+(3\epsilon+3\epsilon^2+\epsilon^3)\left(1+\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2\right)\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\right) \tag{95}$$

$$-\eta(1-\eta)\left((3\epsilon+3\epsilon^2+\epsilon^3)\left(1+\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2\right)\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_2}\mathbf{\Sigma}_3\right\|_F+\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|\right) \tag{96}$$

$$-\eta(1-\eta)\left((3\epsilon+3\epsilon^2+\epsilon^3)\left(1+\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2\right)\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathcal{G}}}\right\|_F+(2\epsilon+\epsilon^2)\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathcal{G}}}\right\|_F\right) \tag{97}$$

$$-\eta(1-\eta)\epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathcal{G}}}\right\|_F-\eta(1-\eta)\epsilon\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F^2+\eta^2\,(1+\epsilon)^{-6}(1+\epsilon+\frac{\epsilon^2}{3})^2\mathrm{dist}^2(\hat{\mathbf{F}}_t,\mathbf{F}) \tag{98}$$

$$\leq (1-\eta)^2\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|^2-\eta(1-\eta)\left(\left\langle\mathcal{T}_{\hat{\mathbf{X}}_1},\mathcal{T}_{\hat{\mathbf{X}}_1}+\mathcal{T}_{\hat{\mathbf{X}}_2}+\mathcal{T}_{\hat{\mathbf{X}}_3}\right\rangle+C_1\epsilon\mathrm{dist}^2\left(\hat{\mathbf{F}}_t,\mathbf{F}\right)+\left\|\Delta_{\hat{\mathbf{X}}_1}\mathbf{\Sigma}_1\right\|_F\left\|\Delta_{\hat{\mathbf{X}}_3}\mathbf{\Sigma}_2\right\|_F\left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|\right) \tag{99}$$

$$+ \eta^2 \, (1+\epsilon)^{-6} \left(1+\epsilon+\frac{\epsilon^2}{3}\right)^2 \mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) \qquad \leq \tag{100}$$

$$\leq (1-\eta)^2 \mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + 4\eta^2 \, (1+\epsilon)^{-6} \, (1+2\epsilon)^2 \, \mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) \tag{101}$$

$$= \mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) \left(1 - 2\eta + \eta^2 + 4\eta^2(1+\epsilon)^{-6}(1+2\epsilon)^2\right) \tag{102}$$

If $\eta < 1/\left(1 + 4(1+\epsilon)^{-6}(1+2\epsilon)^2\right)$, we have $\|\odot\|^2 \leq (1-\eta)\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F})$.

Recalling the bound of $\left\|(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1} - \mathbf{X}_1)\mathbf{\Sigma}_1\right\|_F^2$ in (83), given $\eta < 1/\left(1 + 4(1+\epsilon)^{-6}(1+2\epsilon)^2\right)$, it can be finally bounded as

$$\left\|(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1} - \mathbf{X}_1)\mathbf{\Sigma}_1\right\|_F^2 \tag{103}$$

$$\leq \|\odot\|^2 + 4\eta \|\odot\| \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\| \left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\| + 4\eta^2 \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\|^2 \left\|\check{\mathbf{X}}_1\mathbf{\Sigma}_1\right\|^2 \tag{104}$$

$$\leq (1-\eta)\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + \frac{4\eta\sqrt{(1-\eta)}}{(1+\epsilon)^3}\mathrm{dist}(\hat{\mathbf{F}}_t, \mathbf{F}) \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\| + \frac{4\eta^2}{(1+\epsilon)^6} \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\|^2 \tag{105}$$

$$\leq (1-\eta)\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\| \left(\frac{4\eta\sqrt{(1-\eta)}}{(1+\epsilon)^3}\mathrm{dist}(\hat{\mathbf{F}}_t, \mathbf{F}) + \frac{4\eta^2}{(1+\epsilon)^6} \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\|\right) \tag{106}$$

In a similar way, we can also obtain the same bound for $\left\|(\hat{\mathbf{X}}_{t+1,2}\mathbf{Q}_{t,2} - \mathbf{X}_2)\mathbf{\Sigma}_2\right\|_F^2$ and $\left\|(\hat{\mathbf{X}}_{t+1,3}\mathbf{Q}_{t,3} - \mathbf{X}_3)\mathbf{\Sigma}_3\right\|_F^2$. For example,

$$\left\|(\hat{\mathbf{X}}_{t+1,2}\mathbf{Q}_{t,2} - \mathbf{X}_2)\mathbf{\Sigma}_2\right\|_F^2$$

$$\leq (1-\eta)\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\| \left(\frac{4\eta\sqrt{(1-\eta)}}{(1+\epsilon)^3}\mathrm{dist}(\hat{\mathbf{F}}_t, \mathbf{F}) + \frac{4\eta^2}{(1+\epsilon)^6} \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\|\right). \tag{107}$$

For $\left\|(\mathbf{Q}_1^{-1}, \mathbf{Q}_2^{-1}, \mathbf{Q}_3^{-1}) \cdot \hat{\mathcal{G}} - \mathcal{G}\right\|_F^2$, we follow the bound in [], such that as long as $\eta^2 \leq 2/5$ and $\epsilon \leq 0.2/C$, for $C > 1$, one has

$$\left\|(\mathbf{Q}_{t,1}^{-1}, \mathbf{Q}_{t,2}^{-1}, \mathbf{Q}_{t,3}^{-1}) \cdot \hat{\mathcal{G}}_{t+1} - \mathcal{G}\right\|_F^2 \leq (1-\eta)^2\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + 2\eta(1-\eta)C_2\epsilon\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + \eta^2 C_3\epsilon\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}). \tag{108}$$

$$\leq \left((1-\eta)^2 + 2\eta(1-\eta)C\epsilon + \eta^2 C\epsilon\right)\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) \tag{109}$$

$$\leq (1-0.7\eta)^2\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}). \tag{110}$$

For the proof, refer to []. Combining the bounds for $\left\|(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1} - \mathbf{X}_1)\mathbf{\Sigma}_1\right\|_F^2$, $\left\|(\hat{\mathbf{X}}_{t+1,2}\mathbf{Q}_{t,2} - \mathbf{X}_2)\mathbf{\Sigma}_2\right\|_F^2$, $\left\|(\hat{\mathbf{X}}_{t+1,3}\mathbf{Q}_{t,3} - \mathbf{X}_3)\mathbf{\Sigma}_3\right\|_F^2$ and $\left\|(\mathbf{Q}_1^{-1}, \mathbf{Q}_2^{-1}, \mathbf{Q}_3^{-1}) \cdot \hat{\mathcal{G}} - \mathcal{G}\right\|_F^2$ in (106), (107) and (108), we can bound (15) as

$$\mathrm{dist}^2(\hat{\mathbf{F}}_{t+1}, \mathbf{F}) \tag{111}$$

$$\leq \left\|(\hat{\mathbf{X}}_{t+1,1}\mathbf{Q}_{t,1} - \mathbf{X}_1)\mathbf{\Sigma}_1\right\|_F^2 + \left\|(\hat{\mathbf{X}}_{t+1,2}\mathbf{Q}_{t,2} - \mathbf{X}_2)\mathbf{\Sigma}_2\right\|_F^2 + \left\|(\hat{\mathbf{X}}_{t+1,3}\mathbf{Q}_{t,3} - \mathbf{X}_3)\mathbf{\Sigma}_3\right\|_F^2 + \left\|(\mathbf{Q}_{t,1}^{-1}, \mathbf{Q}_{t,2}^{-1}, \mathbf{Q}_{t,3}^{-1}) \cdot \hat{\mathcal{G}}_{t+1} - \mathcal{G}\right\|_F^2 \tag{112}$$

$$\leq 3(1-\eta)\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\| \left(\frac{12\eta\sqrt{(1-\eta)}}{(1+\epsilon)^3}\mathrm{dist}(\hat{\mathbf{F}}_t, \mathbf{F}) + \frac{12\eta^2}{(1+\epsilon)^6} \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_\mathcal{I}}\right)\right\|\right) \tag{113}$$

$$+ (1-0.7\eta)^2\mathrm{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}), \tag{114}$$

if $\eta \leq 4/49$,

$$
\begin{aligned}
&\text{dist}^2(\hat{\mathbf{F}}_{t+1}, \mathbf{F}) && (115)\\
&\leq 4(1-\eta)\text{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + \left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right\| \left(\frac{12\eta\sqrt{(1-\eta)}}{(1+\epsilon)^3}\text{dist}(\hat{\mathbf{F}}_t, \mathbf{F}) + \frac{12\eta^2}{(1+\epsilon)^6}\left\|\alpha\mathcal{M}_1\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\mathcal{M}_1\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right\|\right) && (116)\\
&= 4(1-\eta)\text{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + \frac{12\eta\sqrt{(1-\eta)}}{(1+\epsilon)^3}\left(\alpha\sigma_{max}\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\sigma_{max}\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right)\text{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) && (117)\\
&\qquad + \frac{12\eta^2}{(1+\epsilon)^6}\left(\alpha\sigma_{max}\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\sigma_{max}\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right)^2 && (118)\\
&\leq (1-\eta)\left(4 + \frac{12\eta}{(1+\epsilon)^3}\left(\alpha\sigma_{max}\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\sigma_{max}\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right)\right)\text{dist}^2(\hat{\mathbf{F}}_t, \mathbf{F}) + \frac{12\eta^2}{(1+\epsilon)^6}\left(\alpha\sigma_{max}\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\sigma_{max}\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)\right)^2. && (119)
\end{aligned}
$$

Note that we suppose $\text{dist}(\hat{\mathbf{F}}_t, \mathbf{F}) \leq \epsilon\sigma_{min}(\mathcal{H})$ for some sufficiently small ϵ. Then, according to (10), when $\epsilon \leq 0.2$ and $\eta \leq \min(4/49, 1/(1+4(1+\epsilon)^{-6}(1+2\epsilon)^2))$, we have

$$
\begin{aligned}
\|(\hat{\mathbf{X}}_1^t, \hat{\mathbf{X}}_2^t, \hat{\mathbf{X}}_3^t)\cdot\hat{\mathcal{G}}^t - \mathcal{H}\|_F &\leq (1-\eta)^t\left(4 + \frac{12\eta\delta}{(1+\epsilon)^3}\right)^t \text{dist}^2\left(\hat{\mathbf{F}}_t, \mathbf{F}\right) + \frac{12\eta^2\delta^2}{(1+\epsilon)^6}\left(\sum_{k=0}^{t-1}(1-\eta)^k\left(4 + \frac{12\eta\delta}{(1+\epsilon)^3}\right)^k\right) && (120)\\
&\leq \epsilon(1-\eta)^t\left(4 + \frac{12\eta\delta}{(1+\epsilon)^3}\right)^t \sigma_{min}(\mathcal{H}) + \frac{12\eta^2\delta^2}{(1+\epsilon)^6}\left(\sum_{k=0}^{t-1}(1-\eta)^k\left(4 + \frac{12\eta\delta}{(1+\epsilon)^3}\right)^k\right) && (121)
\end{aligned}
$$

where $\delta := \alpha\ \sigma_{max}\left(\mathcal{E}_{\hat{\mathcal{Q}}\mathbf{S}}\right) + \beta\ \sigma_{max}\left(\mathcal{E}_{\mathcal{C}\hat{\mathbf{\Lambda}}_{\mathcal{I}}}\right)$ and $\alpha + \beta = 1$. This concludes the proof of Theorem 1.

1. REFERENCES

[1] Tian Tong, Cong Ma, Ashley Prater-Bennette, Erin Tripp, and Yuejie Chi, "Scaling and scalability: Provable nonconvex low-rank tensor estimation from incomplete measurements," *Journal of Machine Learning Research*, vol. 23, no. 163, pp. 1–77, 2022.