Supplementary materials for “What to Select: Pursuing Consistent Motion Segmentation from Multiple Geometric Models”

Table 1: Notations and descriptions.

Notation	Description
V	V is the total number of geometric models used for the dataset.
$\boldsymbol{G}^{(<v)}$	$\boldsymbol{G}^{(<v)}$ is all the $\boldsymbol{G}^{(i)}$ such that $i<v$, *i.e.*, $\boldsymbol{G}^{(<v)}=\left\{\boldsymbol{G}^{(1)},\cdots,\boldsymbol{G}^{(v-1)}\right\}$.
$\boldsymbol{G}^{(>v)}$	$\boldsymbol{G}^{(>v)}$ is all the $\boldsymbol{G}^{(i)}$ such that $i>v$, *i.e.*, $\boldsymbol{G}^{(<v)}=\left\{\boldsymbol{G}^{(v+1)},\cdots,\boldsymbol{G}^{(V)}\right\}$.
$\boldsymbol{G}^{(\backslash v)}$	$\boldsymbol{G}^{(\backslash v)}$ is all the $\boldsymbol{G}^{(i)}$ such that $i\neq v$, *i.e.*, $\boldsymbol{G}^{(\backslash v)}=\left\{\boldsymbol{G}^{(1)},\cdots,\boldsymbol{G}^{(v-1)},\boldsymbol{G}^{(v+1)},\cdots,\boldsymbol{G}^{(V)}\right\}$.
$\boldsymbol{G}$	$\boldsymbol{G}$ stands for all the $\boldsymbol{G}^{(v)}$s when they are considered as a whole.
$\boldsymbol{X}_t$	$\boldsymbol{X}_t$ is the variable $\boldsymbol{X}$ updated at the t-th iteration, where $\boldsymbol{X}$ could be $\boldsymbol{S},\boldsymbol{G}^{(v)},\boldsymbol{U}$ or $\boldsymbol{G}$.
$vec(\boldsymbol{X})$	Given a matrix $\boldsymbol{X}$, $vec(\boldsymbol{X})$ is the vectorization of the matrix $\boldsymbol{X}\in\mathbb{R}^{n_1\times n_2}$ such that $vec(\boldsymbol{X})\in\mathbb{R}^{(n_1\times n_2)\times 1}$.
$\mathcal{I}_{\mathcal{C}}(\cdot)$	Given a set $\mathcal{C}$, $\mathcal{I}_{\mathcal{C}}(\cdot)$ is characteristic function such that $\mathcal{I}_{\mathcal{C}}(\boldsymbol{C})=0$, if $\boldsymbol{C}\in\mathcal{C}$, otherwise $\mathcal{I}_{\mathcal{C}}(\boldsymbol{C})=+\infty$.
$\mathcal{I}_{\mathcal{T}}(\Theta)$	$\mathcal{I}_{\mathcal{T}}(\Theta)=\mathcal{I}_{\mathcal{S}}(\boldsymbol{S})+\mathcal{I}_{\mathcal{U}}(\boldsymbol{U})+\sum_v\mathcal{I}_{\mathcal{G}}(\boldsymbol{G}^{(v)})$
$\ell(\boldsymbol{S},\boldsymbol{G}^{(v)},\boldsymbol{U})$	$\ell(\boldsymbol{S},\boldsymbol{G}^{(v)},\boldsymbol{U})=\frac{1}{2}\left\Vert\boldsymbol{A}^{(v)}-\boldsymbol{S}\odot\boldsymbol{G}^{(v)}\right\Vert_F^2+\frac{\alpha_1}{2}\left\Vert\boldsymbol{G}^{(v)}\right\Vert_F^2+\alpha_2\langle\boldsymbol{L}_{\boldsymbol{S}},\boldsymbol{U}\rangle$
$\ell(\boldsymbol{X})$	$\ell(\boldsymbol{X})$ is the loss function of $\boldsymbol{X}$ subproblem, where $\boldsymbol{X}$ could be $\boldsymbol{S},\boldsymbol{G}^{(v)},\boldsymbol{U}$.
$\mathcal{L}(\boldsymbol{S},\boldsymbol{G},\boldsymbol{U})$	$\mathcal{L}(\boldsymbol{S},\boldsymbol{G},\boldsymbol{U})=\frac{1}{2}\sum_v\left\Vert\boldsymbol{A}^{(v)}-\boldsymbol{S}\odot\boldsymbol{G}^{(v)}\right\Vert_F^2+\frac{\alpha_1}{2}\sum_v\left\Vert\boldsymbol{G}^{(v)}\right\Vert_F^2+\alpha_2\langle\boldsymbol{L}_{\boldsymbol{S}},\boldsymbol{U}\rangle+\frac{\alpha_3}{2}\left\Vert\boldsymbol{U}\right\Vert_F^2$
$\mathcal{F}(\boldsymbol{S},\boldsymbol{G},\boldsymbol{U})$	$\mathcal{F}(\boldsymbol{S},\boldsymbol{G},\boldsymbol{U})=\mathcal{L}(\boldsymbol{S},\boldsymbol{G},\boldsymbol{U})+\mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta})$, where $\boldsymbol{\Theta}=\{\boldsymbol{S},\boldsymbol{G},\boldsymbol{U}\}$. It could also be denoted as $\mathcal{F}(\boldsymbol{\Theta})$.
$dist(\cdot,\cdot)$	$dist(\cdot,\cdot)$ is the euclidean distance between a point and a set of points.
$\boldsymbol{q}_{\mathcal{C}}$	$\boldsymbol{q}_{\mathcal{C}}$ is a subgradient of $\mathcal{I}_{\mathcal{C}}(\cdot)$.
$\boldsymbol{g}_{\boldsymbol{C}}$	$\boldsymbol{g}_{\boldsymbol{C}}$ is a subgradient of $\mathcal{F}(\cdot)$ w.r.t. $\boldsymbol{C}$, where $\boldsymbol{C}$ could be $\boldsymbol{S},\boldsymbol{G}^{(v)},\boldsymbol{U},\boldsymbol{\Theta}$.
$\boldsymbol{q}_{\mathcal{C}}^t,\boldsymbol{g}_{\boldsymbol{C}}^t$	$\boldsymbol{q}_{\mathcal{C}}^t,\boldsymbol{g}_{\boldsymbol{C}}^t$ are the values of $\boldsymbol{q}_{\mathcal{C}},\boldsymbol{g}_{\boldsymbol{C}}$ in the t-th iteration.
$\boldsymbol{X}^{\odot 2}$	$\boldsymbol{X}^{\odot 2}=\boldsymbol{X}\odot\boldsymbol{X}$
$\Vert\cdot\Vert$	$\Vert\cdot\Vert$ means $\Vert\cdot\Vert_F$ by default.

Proof of Theorem 1

The proof could be found in (Yang et al. 2020).

Proof of Theorem 2

Lemma 1. *The following facts hold for the overall objective function of the proposed model:*

(1) Given $\boldsymbol{S},\boldsymbol{G}^{(\backslash v)},\boldsymbol{U}$*, the subproblem for each* $\boldsymbol{G}^{(v)}$ *(Eq.(7)) is* α_1*-strongly convex.*

(2) *Given $\boldsymbol{G}, \boldsymbol{U}$, the subproblem for $\boldsymbol{S}$ (Eq.(9)) is $\alpha_2 V G_{min}^2$-strongly convex.*

(3) *Given $\boldsymbol{S}, \boldsymbol{G}$, the subproblem for $\boldsymbol{U}$ (Eq.(13)) is α_3-strongly convex.*

Proof.

(1) Denote the loss function of $\boldsymbol{G}^{(v)}$ as follows:

$$\ell(\boldsymbol{G}^{(v)}) = \frac{1}{2}\|\boldsymbol{A}^{(v)} - \boldsymbol{S} \odot \boldsymbol{G}^{(v)}\|_F^2 + \frac{\alpha_1}{2}\|\boldsymbol{G}^{(v)}\|_F^2$$

According to (Boyd and Vandenberghe 2004), proving (a) is equivalent to prove that the Hessian matrix of $\mathcal{J}(\boldsymbol{G}^{(v)})$ subjects to

$$\boldsymbol{H} = \frac{\partial \mathcal{J}(\boldsymbol{G}^{(v)})}{\partial vec(\boldsymbol{G}^{(v)}) \partial vec(\boldsymbol{G}^{(v)})^\top} \succeq \alpha_1 \boldsymbol{I}_{N^2}.$$

Let the loss of each element of $\boldsymbol{S}$ be:

$$\ell(\boldsymbol{G}_{ij}^{(v)}) = \frac{1}{2}\left(\boldsymbol{A}_{ij}^{(v)} - \boldsymbol{S}_{ij}\boldsymbol{G}_{ij}^{(v)}\right)^2 + \frac{\alpha_1}{2}\boldsymbol{G}_{ij}^{(v)2}$$

And

$$\ell(\boldsymbol{G}^{(v)}) = \sum_{i,j} \ell(\boldsymbol{G}_{ij}^{(v)})$$

We have

$$\boldsymbol{H} = diag\left(vec\left(\frac{\partial^2 \mathcal{J}(\boldsymbol{G}_{ij}^{(v)})}{\partial^2 \boldsymbol{G}_{ij}^{(v)}}\right)\right) \succeq \alpha_1 \boldsymbol{I}_{N^2}.$$

(2) Similar to (a), we denote the loss function of $\boldsymbol{S}$ as:

$$\ell(\boldsymbol{S}) = \frac{1}{2}\sum_v \|\boldsymbol{A}^{(v)} - \boldsymbol{S} \odot \boldsymbol{G}^{(v)}\|_F^2 + \alpha_2 \langle \boldsymbol{L}_{\boldsymbol{S}}, \boldsymbol{U}\rangle$$

Then the loss of each element of $\boldsymbol{S}$ could be written as:

$$\ell(\boldsymbol{S}_{ij}) = \frac{1}{2}\sum_v \left(\boldsymbol{A}_{ij}^{(v)} - \boldsymbol{S}_{ij}\boldsymbol{G}_{ij}^{(v)}\right)^2 + \alpha_2 \boldsymbol{L}_{\boldsymbol{S}ij}\boldsymbol{U}_{ij}$$

And

$$\ell(\boldsymbol{S}) = \sum_{i,j} \ell(\boldsymbol{S}_{ij})$$

Then we have

$$\boldsymbol{H} = \frac{\partial \ell(\boldsymbol{S})}{\partial vec(\boldsymbol{S}) \partial vec(\boldsymbol{S})^\top} = diag\left(vec\left(\frac{\partial^2 \ell(\boldsymbol{S}_{ij})}{\partial^2 \boldsymbol{S}_{ij}}\right)\right) \succeq \alpha_2 V G_{min}^2 \boldsymbol{I}_{N^2}.$$

Since all the constraints on $\boldsymbol{S}$ are linear, (2) is proved.

(3) See (Yang et al. 2020).

$\square$

Lemma 2. *Let $\mathcal{L}(\boldsymbol{S}, \boldsymbol{G}, \boldsymbol{U}) = \frac{1}{2}\sum_v \|\boldsymbol{A}^{(v)} - \boldsymbol{S} \odot \boldsymbol{G}^{(v)}\|_F^2 + \frac{\alpha_1}{2}\sum_v \|\boldsymbol{G}^{(v)}\|_F^2 + \alpha_2\langle \boldsymbol{L}_{\boldsymbol{S}}, \boldsymbol{U}\rangle + \frac{\alpha_3}{2}\|\boldsymbol{U}\|_F^2$. The objective function could be reformulated as $\mathcal{F}(\boldsymbol{S}, \boldsymbol{G}, \boldsymbol{U}) = \mathcal{L}(\boldsymbol{S}, \boldsymbol{G}, \boldsymbol{U}) + \mathcal{I}_{\mathcal{S}}(\boldsymbol{S}) + \sum_v \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}^{(v)}) + \mathcal{I}_{\mathcal{U}}(\boldsymbol{U})$, and let $\boldsymbol{S}_t, \boldsymbol{G}_t^{(v)}, \boldsymbol{U}_t$ be the parameters obtained at iteration t, the following properties hold for Alg. 1.*

(1) *(Sufficient Decrease Condition): The sequence $\{\mathcal{F}(\boldsymbol{S}_t, \boldsymbol{G}_t, \boldsymbol{U}_t)\}_t$ is non-increasing in the sense that:*

$$\mathcal{F}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_{t+1}) \leq \mathcal{F}(\boldsymbol{S}_t, \boldsymbol{G}_t, \boldsymbol{U}_t) - \min\left\{\frac{\alpha_1}{2}, \frac{\alpha_2 V G_{min}^2}{2}, \frac{\alpha_3}{2}\right\}\|\Delta\boldsymbol{\Theta}_t\|^2.$$

where $\Delta\boldsymbol{\Theta}_t = \left[vec(\Delta\boldsymbol{S}_t);\ vec(\Delta\boldsymbol{G}_t^{(1)});\ \cdots;\ vec(\Delta\boldsymbol{G}_t^{(V)});\ vec(\Delta\boldsymbol{U}_t)\right]$, $\Delta\boldsymbol{S}_t = \boldsymbol{S}_{t+1} - \boldsymbol{S}_t$, $\Delta\boldsymbol{G}_t^{(v)} = \boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{G}_t^{(v)}$, $\Delta\boldsymbol{U}_t = \boldsymbol{U}_{t+1} - \boldsymbol{U}_t$.

(2) *(Square Summable): $\sum_{i=1}^{\infty}\|\Delta\boldsymbol{\Theta}_t\|_F^2 < \infty$. Furthermore, we have $\lim_{t\to\infty}\|\Delta\boldsymbol{S}_t\| = 0, \lim_{t\to\infty}\|\Delta\boldsymbol{G}_t^{(v)}\| = 0, \lim_{t\to\infty}\|\Delta\boldsymbol{U}_t\| = 0$.*

(3) *(Continuity Condition): There exists a subsequence of* $\{\boldsymbol{S}_{t_j}, \boldsymbol{G}_{t_j}, \boldsymbol{U}_{t_j}\}_j$ *with a limit point* $\{\boldsymbol{S}^\star, \boldsymbol{G}^\star, \boldsymbol{U}^\star\}$*, such that*

$$\{\boldsymbol{S}_{t_j}, \boldsymbol{G}_{t_j}, \boldsymbol{U}_{t_j}\} \rightarrow \{\boldsymbol{S}^\star, \boldsymbol{G}^\star, \boldsymbol{U}^\star\}, \ \ and \ \ \mathcal{F}(\boldsymbol{S}_{t_j}, \boldsymbol{G}_{t_j}, \boldsymbol{U}_{t_j}) \rightarrow \mathcal{F}(\boldsymbol{S}^\star, \boldsymbol{G}^\star, \boldsymbol{U}^\star),$$

(4) *(KL Property):* $\mathcal{F}(\cdot,\cdot,\cdot)$ *is a KL function.*

(5) *(Relative Error Condition): There holds that*

$$dist(\mathbf{0}, \partial_{\boldsymbol{\Theta}}\mathcal{F}(\Theta_{t+1})^2 = \min_{\boldsymbol{g}\in\partial_{\Theta}\mathcal{F}(\boldsymbol{\Theta}_{t+1})} \|\boldsymbol{g}\|_F^2 \leq C_{max}^2\|\Theta_{t+1} - \Theta_t\|_F^2, \ \ for\ t \in \mathbb{N}. \tag{16}$$

where:

$$C_{max} = \max\left\{V\left(2S_{max}\cdot\sqrt{\frac{2}{\alpha_1}\mathcal{F}(\boldsymbol{S}_0, \boldsymbol{G}_0, \boldsymbol{U}_0)} + A_{max}\right), \alpha_2(\sqrt{N}+1)\right\}, \tag{17}$$

and $A_{max} = \max_{i,j,v}|A_{i,j}^{(v)}|$.

Proof.

(1) In the $t+1$-th iteration, we update the variables sequentially with the values $\boldsymbol{G}_t^{(1)}, \cdots, \boldsymbol{G}_t^{(V)}, \boldsymbol{S}_t, \boldsymbol{U}_t$.
First, we update each $\boldsymbol{G}^{(v)}$. According to Lemma 1 and Eq.(7), $\mathcal{L}(\boldsymbol{S}_t, \boldsymbol{G}_{t+1}^{(<v)}, \boldsymbol{G}^{(v)}, \boldsymbol{G}_t^{(>v)}, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}^{(v)})$ is α_1-strongly convex with respect to $\boldsymbol{G}^{(v)}$. Thus with the solution $\boldsymbol{G}_{t+1}^{(v)}$ we have $\forall v = 1, 2, \cdots, V$,

$$\mathcal{L}(\boldsymbol{S}_t, \boldsymbol{G}_{t+1}^{(<v)}, \boldsymbol{G}_{t+1}^{(v)}, \boldsymbol{G}_t^{(>v)}, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}_{t+1}^{(v)}) \leq \mathcal{L}(\boldsymbol{S}_t, \boldsymbol{G}_{t+1}^{(<v)}, \boldsymbol{G}_t^{(v)}, \boldsymbol{G}_t^{(>v)}, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}_t^{(v)}) - \frac{\alpha_1}{2}\|\boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{G}_t^{(v)}\|_F^2. \tag{18}$$

After updating all $\boldsymbol{G}^{(v)}$s, we start updating $\boldsymbol{S}$. Similarly, we have that $\mathcal{L}(\boldsymbol{S}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{S}}(\boldsymbol{S})$ is $\alpha_2 V G_{min}^2$-strongly convex w.r.t $\boldsymbol{S}$. Hence

$$\mathcal{L}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{S}}(\boldsymbol{S}_{t+1}) \leq \mathcal{L}(\boldsymbol{S}_t, \boldsymbol{G}_{t+1}, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{S}}(\boldsymbol{S}_t) - \frac{\alpha_2 V G_{min}^2}{2}\|\boldsymbol{S}_{t+1} - \boldsymbol{S}_t\|_F^2. \tag{19}$$

Finally, for the update of $\boldsymbol{U}$, we have $\mathcal{L}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}) + \mathcal{I}_{\mathcal{U}}(\boldsymbol{U})$ is α_3-strongly convex. The

$$\mathcal{L}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_{t+1}) + \mathcal{I}_{\mathcal{U}}(\boldsymbol{U}_{t+1}) \leq \mathcal{L}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{U}}(\boldsymbol{U}_t) - \frac{\alpha_3}{2}\|\boldsymbol{U}_{t+1} - \boldsymbol{U}_t\|_F^2. \tag{20}$$

Let $\mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}) = \sum_v \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}^{(v)}) + \mathcal{I}_{\mathcal{S}}(\boldsymbol{S}) + \mathcal{I}_{\mathcal{U}}(\boldsymbol{U})$. Combing Eq. (18)-(20), we have

$$\begin{aligned}\mathcal{L}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_{t+1}) + \mathcal{I}_{\mathcal{T}}(\Theta_{t+1}) &\leq \mathcal{L}(\boldsymbol{S}_t, \boldsymbol{G}_t, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{T}}(\Theta_t) \\ &- \frac{\alpha_1}{2}\sum_v \|\boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{G}_t^{(v)}\|_F^2 - \frac{\alpha_2 V G_{min}^2}{2}\|\boldsymbol{S}_{t+1} - \boldsymbol{S}_t\|_F^2 - \frac{\alpha_3}{2}\|\boldsymbol{U}_{t+1} - \boldsymbol{U}_t\|_F^2\end{aligned} \tag{21}$$

This ends the proof of (1).

(2) Since $\boldsymbol{L}_{\boldsymbol{S}} \succeq 0$, it holds that $\mathcal{L}(\boldsymbol{S}_t, \boldsymbol{G}_t, \boldsymbol{U}_t) + \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}_t) \geq 0$. With feasible initial values $\boldsymbol{S}_0, \boldsymbol{G}_0, \boldsymbol{U}_0$, by adding up Eq. (21) for all $t = 1, 2, \cdots$, we then have

$$\sum_{t=1}^{\infty} \frac{\alpha_1}{2}\|\Delta\boldsymbol{G}_t^{(v)}\|_F^2 + \frac{\alpha_2 V G_{min}^2}{2}\|\Delta\boldsymbol{S}_t\|_F^2 + \frac{\alpha_3}{2}\|\Delta\boldsymbol{U}_t\|_F^2 \leq \mathcal{L}(\boldsymbol{S}_0, \boldsymbol{G}_0, \boldsymbol{U}_0)$$

By the assumption that $\mathcal{L}(\boldsymbol{S}_0, \boldsymbol{G}_0, \boldsymbol{U}_0) < \infty$, this implies that when $t \rightarrow \infty$

$$\boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{G}_t^{(v)} \rightarrow 0, \forall v = 1, 2, \cdots, V, \quad \boldsymbol{S}_{t+1} - \boldsymbol{S}_t \rightarrow 0, \quad \boldsymbol{U}_{t+1} - \boldsymbol{U}_t \rightarrow 0.$$

This ends the proof of (2).

(3) To prove (3), we first prove the existence of a limit point of the parameter sequence $\{\boldsymbol{\Theta}_t\}_t$ generated by our Algorithm 1. It could be seen that the sequences $\{\boldsymbol{S}_t\}_t$ and $\{\boldsymbol{U}_t\}_t$ are bounded by the construction of $\mathcal{S}$ and $\mathcal{U}$. According to (1), the loss sequence $\{\mathcal{F}(\boldsymbol{S}_t, \boldsymbol{G}_t, \boldsymbol{U}_t)\}_t$ is non-decreasing. Then we could obtain that $\|\boldsymbol{G}_t^{(v)}\|_F^2 \leq \frac{2}{\alpha_1}\mathcal{F}(\boldsymbol{S}_0, \boldsymbol{G}_0^{(v)}, \boldsymbol{U}_0)$ and $\|\boldsymbol{A}^{(v)} - \boldsymbol{S}_t \odot \boldsymbol{G}_t^{(v)}\|_F^2 \leq 2\mathcal{F}(\boldsymbol{S}_0, \boldsymbol{G}_0^{(v)}, \boldsymbol{U}_0)$. Namely, $\{\boldsymbol{G}_t^{(v)}\}_t$ is also bounded. Based on the Bolzano-Weierstrass theorem (Bartle and Sherbert 2000), any bounded sequence must have convergent subsequence. Hence there exists at least one subsequence $\{\boldsymbol{\Theta}_{t_j}\}_{t_j}$ that converges to a limit point $\boldsymbol{\Theta}^\star$.
Now we prove the continuity condition of $\mathcal{F}(\boldsymbol{\Theta})$ by separately considering the continuity of $\mathcal{F}(\boldsymbol{\Theta}) - \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta})$ and $\mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta})$. Note that since $\mathcal{L}(\boldsymbol{\Theta}) = \mathcal{F}(\boldsymbol{\Theta}) - \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta})$ is continuous, we must have $\lim_{t_j\rightarrow\infty}\mathcal{F}(\boldsymbol{\Theta}_{t_j}) - \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}_{t_j}) = \mathcal{F}(\boldsymbol{\Theta}^\star) - \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}^\star)$. Then we only need to prove that $\lim_{t_j\rightarrow\infty}\mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}_{t_j}) = \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}^\star)$.

Due to the lower semi-continuity of $\mathcal{I}_{\mathcal{G}}, \mathcal{I}_{\mathcal{S}}, \mathcal{I}_{\mathcal{U}}$, we have

$$\liminf_{t_j \to \infty} \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}_{t_j}) \geq \sum_v \left(\liminf_{t_j \to \infty} \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}_{t_j}^{(v)}) \right) + \liminf_{t_j \to \infty} \mathcal{I}_{\mathcal{S}}(\boldsymbol{S}_{t_j}) + \liminf_{t_j \to \infty} \mathcal{I}_{\mathcal{U}}(\boldsymbol{U}_{t_j}) \tag{22}$$

$$\geq \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}^{\star}) \tag{23}$$

It only remains to prove that $\limsup_{t_j \to \infty} \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}_{t_j}) \leq \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}^{\star})$.

For $\mathcal{I}_{\mathcal{G}}$,

$$\boldsymbol{G}_{t_j}^{(v)} = \arg\min_{\boldsymbol{G}^{(v)}} \frac{1}{2}\|\boldsymbol{A}^{(v)} - \boldsymbol{S}_{t_j-1} \odot \boldsymbol{G}^{(v)}\|_F^2 + \frac{\alpha_1}{2}\|\boldsymbol{G}^{(v)}\|_F^2 + \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}^{(v)})$$

Thus

$$\frac{1}{2}\|\boldsymbol{A}^{(v)} - \boldsymbol{S}_{t_j-1} \odot \boldsymbol{G}_{t_j}^{(v)}\|_F^2 + \frac{\alpha_1}{2}\|\boldsymbol{G}_{t_j}^{(v)}\|_F^2 + \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}_{t_j}^{(v)}) \leq \frac{1}{2}\|\boldsymbol{A}^{(v)} - \boldsymbol{S}_{t_j-1} \odot \boldsymbol{G}^{(v)\star}\|_F^2 + \frac{\alpha_1}{2}\|\boldsymbol{G}^{(v)\star}\|_F^2 + \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}^{(v)\star})$$

Taking the limits of both sides as $t_j \to \infty$, we then have $\limsup_{t_j \to \infty} \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}_{t_j}^{(v)}) \leq \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}^{(v)\star})$.

Likewise, we could prove that $\limsup_{t_j \to \infty} \mathcal{I}_{\mathcal{S}}(\boldsymbol{S}_{t_j}) \leq \mathcal{I}_{\mathcal{S}}(\boldsymbol{S}^{\star})$ and $\limsup_{t_j \to \infty} \mathcal{I}_{\mathcal{U}}(\boldsymbol{U}_{t_j}) \leq \mathcal{I}_{\mathcal{U}}(\boldsymbol{U}^{\star})$.

Therefore, we have:

$$\limsup_{t_j \to \infty} \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}_{t_j}) \leq \sum_v \left(\limsup_{t_j \to \infty} \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}_{t_j}^{(v)}) \right) + \limsup_{t_j \to \infty} \mathcal{I}_{\mathcal{S}}(\boldsymbol{S}_{t_j}) + \limsup_{t_j \to \infty} \mathcal{I}_{\mathcal{U}}(\boldsymbol{U}_{t_j}) \leq \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}^{\star}) \tag{24}$$

According to Eq.(22) and Eq.(24), we have:

$$\lim_{t_j \to \infty} \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}_{t_j}) = \mathcal{I}_{\mathcal{T}}(\boldsymbol{\Theta}^{\star})$$

Thus the proof is completed.

(4) The proof follows that of Lemma 1(4) in (Yang et al. 2020).

(5) Due to the optimality of $\boldsymbol{G}_{t+1}^{(v)}$ for the $\boldsymbol{G}^{(v)}$ subproblem in the $t+1$-th iteration, we have

$$\mathbf{0} \in \partial_{\boldsymbol{G}_{t+1}^{(v)}} \mathcal{F}(\boldsymbol{S}_t, \boldsymbol{G}_{t+1}, \boldsymbol{U}_t) = \boldsymbol{S}_t \odot (\boldsymbol{S}_t \odot \boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{A}^{(v)}) + \alpha_1 \boldsymbol{G}_{t+1}^{(v)} + \partial_{\boldsymbol{G}_{t+1}^{(v)}} \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}_{t+1}^{(v)})$$

There exists a $\boldsymbol{q}_{\mathcal{G}}^{t+1} \in \partial_{\boldsymbol{G}_{t+1}^{(v)}} \mathcal{I}_{\mathcal{G}}(\boldsymbol{G}_{t+1}^{(v)})$ such that

$$\begin{aligned} \mathbf{0} =& \boldsymbol{S}_t \odot \left(\boldsymbol{S}_t \odot \boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{A}^{(v)} \right) + \alpha_1 \boldsymbol{G}_{t+1}^{(v)} + \boldsymbol{q}_{\mathcal{G}}^{t+1} \\ =& \boldsymbol{S}_t \odot \left(\boldsymbol{S}_t \odot \boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{A}^{(v)} \right) + \alpha_1 \boldsymbol{G}_{t+1}^{(v)} + \boldsymbol{q}_{\mathcal{G}}^{t+1} \\ &+ \left(\boldsymbol{S}_{t+1}{}^{\odot 2} \odot \boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{S}_{t+1} \odot \boldsymbol{A}^{(v)} \right) - \left(\boldsymbol{S}_{t+1}{}^{\odot 2} \odot \boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{S}_{t+1} \odot \boldsymbol{A}^{(v)} \right) \end{aligned}$$

Let

$$\boldsymbol{g}_{\boldsymbol{G}^{(v)}}^{t+1} = \boldsymbol{S}_{t+1}^{\odot 2} \odot \boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{S}_{t+1} \odot \boldsymbol{A}^{(v)} + \alpha_1 \boldsymbol{G}_{t+1}^{(v)} + \boldsymbol{q}_{\mathcal{G}}^{t+1} \tag{25}$$

then

$$\begin{aligned} \mathbf{0} &= \boldsymbol{g}_{\boldsymbol{G}^{(v)}}^{t+1} + \boldsymbol{S}_t{}^{\odot 2} \odot \boldsymbol{G}_{t+1}^{(v)} - \boldsymbol{S}_t \odot \boldsymbol{A}^{(v)} - \boldsymbol{S}_{t+1}{}^{\odot 2} \odot \boldsymbol{G}_{t+1}^{(v)} + \boldsymbol{S}_{t+1} \odot \boldsymbol{A}^{(v)} \\ &= \boldsymbol{g}_{\boldsymbol{G}^{(v)}}^{t+1} + \boldsymbol{G}_{t+1}^{(v)} \odot \left(\boldsymbol{S}_t{}^{\odot 2} - \boldsymbol{S}_{t+1}{}^{\odot 2} \right) - \boldsymbol{A}^{(v)} \odot (\boldsymbol{S}_t - \boldsymbol{S}_{t+1}) \end{aligned}$$

Thus

$$\|\boldsymbol{g}_{\boldsymbol{G}^{(v)}}^{t+1}\| \leq \underbrace{\|\boldsymbol{G}_{t+1}^{(v)} \odot (\boldsymbol{S}_t{}^{\odot 2} - \boldsymbol{S}_{t+1}{}^{\odot 2})\|}_{(a)} + \underbrace{\|\boldsymbol{A}^{(v)} \odot (\boldsymbol{S}_t - \boldsymbol{S}_{t+1})\|}_{(b)}$$

For (a), we have

$$(a) = \|\boldsymbol{G}_{t+1}^{(v)} \odot (\boldsymbol{S}_t + \boldsymbol{S}_{t+1}) \odot (\boldsymbol{S}_t - \boldsymbol{S}_{t+1})\|$$

Furthermore, by (1), $\mathcal{F}(\boldsymbol{S}_t, \boldsymbol{G}_t^{(v)}, \boldsymbol{U}_t)$ is non-increasing w.r.t. t, we have

$$\max_{i,j} \boldsymbol{G}_{t+1_{i,j}}^{(v)} \leq \|\boldsymbol{G}_{t+1}^{(v)}\|_F \leq \sqrt{\frac{2}{\alpha_1}\mathcal{F}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}^{(v)}, \boldsymbol{U}_{t+1})} \leq \sqrt{\frac{2}{\alpha_1}\mathcal{F}(\boldsymbol{S}_0, \boldsymbol{G}_0^{(v)}, \boldsymbol{U}_0)}$$

$$\max_{i,j} \left(\boldsymbol{S}_{t+1_{i,j}} + \boldsymbol{S}_{t i,j} \right) \leq 2S_{max}$$

Thus

$$
\begin{aligned}
(a) &\leq \left(\max_{i,j}\ \boldsymbol{G}^{(v)}_{t+1_{i,j}}\right) \cdot \left(\max_{i,j}\ \left(\boldsymbol{S}_{t+1_{i,j}} + \boldsymbol{S}_{t_{i,j}}\right)\right) \cdot \|\boldsymbol{S}_{t+1} - \boldsymbol{S}_t\| \\
&\leq 2S_{max} \cdot \sqrt{\frac{2}{\alpha_1}\mathcal{F}(\boldsymbol{S}_0, \boldsymbol{G}_0^{(v)}, \boldsymbol{U}_0)} \cdot \|\boldsymbol{S}_{t+1} - \boldsymbol{S}_t\|
\end{aligned}
$$

By assumption we have

$$
(b) \leq A_{max} \cdot \|\boldsymbol{S}_{t+1} - \boldsymbol{S}_t\|
$$

Above all, we have

$$
\|\boldsymbol{g}^{t+1}_{\boldsymbol{G}^{(v)}}\| \leq \left(2S_{max} \cdot \sqrt{\frac{2}{\alpha_1}\mathcal{F}(\boldsymbol{S}_0, \boldsymbol{G}_0^{(v)}, \boldsymbol{U}_0)} + A_{max}\right) \cdot \|\boldsymbol{S}_{t+1} - \boldsymbol{S}_t\|
$$

From the optimality of $\boldsymbol{S}_{t+1}$ for the $\boldsymbol{S}$ subproblem in the $t+1$-th iteration,

$$
\mathbf{0} \in \sum_v \boldsymbol{G}^{(v)}_{t+1} \odot (\boldsymbol{S}_{t+1} \odot \boldsymbol{G}^{(v)}_{t+1} - \boldsymbol{A}^{(v)}) + \alpha_2(diag(\boldsymbol{U}_t)\mathbf{1}^\top - \boldsymbol{U}_t) + \partial_{\boldsymbol{S}_{t+1}}\mathcal{I}_{\mathcal{S}}(\boldsymbol{S}_{t+1})
$$

There exists a $\boldsymbol{q}^{t+1}_{\mathcal{S}} \in \partial_{\boldsymbol{S}_{t+1}}\mathcal{I}_{\mathcal{S}}(\boldsymbol{S}_{t+1})$ such that

$$
\begin{aligned}
\mathbf{0} &= \sum_v \boldsymbol{G}^{(v)}_{t+1} \odot \left(\boldsymbol{S}_{t+1} \odot \boldsymbol{G}^{(v)}_{t+1} - \boldsymbol{A}^{(v)}\right) + \alpha_2\left(diag(\boldsymbol{U}_t)\mathbf{1}^\top - \boldsymbol{U}_t\right) + \boldsymbol{q}^{t+1}_{\mathcal{S}} \\
&= \sum_v \boldsymbol{G}^{(v)}_{t+1} \odot \left(\boldsymbol{S}_{t+1} \odot \boldsymbol{G}^{(v)}_{t+1} - \boldsymbol{A}^{(v)}\right) + \alpha_2\left(diag(\boldsymbol{U}_t)\mathbf{1}^\top - \boldsymbol{U}_t\right) + \boldsymbol{q}^{t+1}_{\mathcal{S}} \\
&\quad + \alpha_2\left(diag(\boldsymbol{U}_{t+1})\mathbf{1}^\top - \boldsymbol{U}_{t+1}\right) - \alpha_2\left(diag(\boldsymbol{U}_{t+1})\mathbf{1}^\top - \boldsymbol{U}_{t+1}\right)
\end{aligned}
$$

Let

$$
\boldsymbol{g}^{t+1}_{\boldsymbol{S}} = \sum_v \boldsymbol{G}^{(v)}_{t+1} \odot \left(\boldsymbol{S}_{t+1} \odot \boldsymbol{G}^{(v)}_{t+1} - \boldsymbol{A}^{(v)}\right) + \alpha_2\left(diag(\boldsymbol{U}_{t+1})\mathbf{1}^\top - \boldsymbol{U}_{t+1}\right) + \boldsymbol{q}^{t+1}_{\mathcal{S}}. \tag{26}
$$

We have:

$$
\begin{aligned}
\mathbf{0} &= \boldsymbol{g}^{t+1}_{\boldsymbol{S}} + \alpha_2\left(diag(\boldsymbol{U}_t)\mathbf{1}^\top - \boldsymbol{U}_t\right) - \alpha_2\left(diag(\boldsymbol{U}_{t+1})\mathbf{1}^\top - \boldsymbol{U}_{t+1}\right) \\
&= \boldsymbol{g}^{t+1}_{\boldsymbol{S}} + \alpha_2\left(diag(\boldsymbol{U}_t) - diag(\boldsymbol{U}_{t+1})\right)\mathbf{1}^\top - \alpha_2\left(\boldsymbol{U}_t - \boldsymbol{U}_{t+1}\right)
\end{aligned}
$$

Hence we have

$$
\begin{aligned}
\|\boldsymbol{g}^{t+1}_{\boldsymbol{S}}\| &\leq \alpha_2\left(\|\left(diag(\boldsymbol{U}_{t+1}) - diag(\boldsymbol{U}_t)\right)\mathbf{1}^\top\| + \|\boldsymbol{U}_t - \boldsymbol{U}_{t+1}\|\right) \\
&\leq \alpha_2 \cdot \left(\sqrt{N} + 1\right) \cdot \|\boldsymbol{U}_t - \boldsymbol{U}_{t+1}\|
\end{aligned}
$$

There exists a

$$
\boldsymbol{g}^{t+1}_{\boldsymbol{U}} \in \partial_{\boldsymbol{U}_{t+1}}\mathcal{F}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_{t+1}), \tag{27}
$$

such that

$$
\|\boldsymbol{g}^{t+1}_{\boldsymbol{U}}\| = 0 \leq C_{max} \cdot \|\boldsymbol{G}_t - \boldsymbol{G}_{t+1}\|
$$

where

$$
\begin{aligned}
C_{max} &= \max\left\{V\left(2S_{max} \cdot \sqrt{\frac{2}{\alpha_1}\mathcal{F}(\boldsymbol{S}_0, \boldsymbol{G}_0^{(v)}, \boldsymbol{U}_0)} + A_{max}\right), \alpha_2(\sqrt{N}+1)\right\}, \\
\boldsymbol{G}_t &= \left[\boldsymbol{G}^{(1)}_t; \cdots; \boldsymbol{G}^{(V)}_t\right], \\
\boldsymbol{G}_{t+1} &= \left[\boldsymbol{G}^{(1)}_{t+1}; \cdots; \boldsymbol{G}^{(V)}_{t+1}\right].
\end{aligned}
$$

According to Eq.(25), Eq.(26), and Eq.(27), we have:

$$\boldsymbol{g}_{\boldsymbol{G}^{(v)}}^{t+1} \in \partial_{\boldsymbol{G}_{t+1}^{(v)}} \mathcal{F}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_{t+1})$$
$$\boldsymbol{g}_{\boldsymbol{S}}^{t+1} \in \partial_{\boldsymbol{S}_{t+1}} \mathcal{F}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_{t+1})$$
$$\boldsymbol{g}_{\boldsymbol{U}}^{t+1} \in \partial_{\boldsymbol{U}_{t+1}} \mathcal{F}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_{t+1})$$

Therefore

$$\left[\boldsymbol{g}_{\boldsymbol{G}^{(1)}}^{t+1}; \cdots; \boldsymbol{g}_{\boldsymbol{G}^{(V)}}^{t+1}; \; \boldsymbol{g}_{\boldsymbol{S}}^{t+1}; \; \boldsymbol{g}_{\boldsymbol{U}}^{t+1}\right] \in \partial_{\boldsymbol{\Theta}_{t+1}} \mathcal{F}(\boldsymbol{\Theta}_{t+1})$$

$$\begin{aligned} dist(\boldsymbol{0}, \partial_{\boldsymbol{\Theta}} \mathcal{F}(\boldsymbol{S}_{t+1}, \boldsymbol{G}_{t+1}, \boldsymbol{U}_{t+1}))^2 &= \min_{\boldsymbol{g} \in \partial_{\Theta} \mathcal{F}(\boldsymbol{\Theta}_{t+1})} \|\boldsymbol{g}\|_F^2 \\ &\leq \sum_v \|\boldsymbol{g}_{\boldsymbol{G}^{(v)}}^{t+1}\|_F^2 + \|\boldsymbol{g}_{\boldsymbol{S}}^{t+1}\|_F^2 + \|\boldsymbol{g}_{\boldsymbol{U}}^{t+1}\|_F^2 \\ &\leq C_{max}^2 \|\boldsymbol{\Theta}_{t+1} - \boldsymbol{\Theta}_t\|_F^2. \end{aligned}$$

The proof is then finished.

□

Based on Lemma 2, the proof of Theorem 2 then follows that of Theorem 4 in (Yang et al. 2020).

Proof of Theorem 3

This proof follows that of Theorem 5 in (Yang et al. 2020).

References

Bartle, R. G.; and Sherbert, D. R. 2000. *Introduction to real analysis*, volume 2. Wiley New York.

Boyd, S.; and Vandenberghe, L. 2004. *Convex optimization*. Cambridge university press.

Yang, Z.; Xu, Q.; Cao, X.; and Huang, Q. 2020. Task-Feature Collaborative Learning with Application to Personalized Attribute Prediction. *arXiv preprint arXiv:2004.13930* .