Supplementary Materials: High-order Complementarity Induced Fast Multi-View Clustering with Enhanced Tensor Rank Minimization

Paper ID 188

In this supplementary, we provide the proofs for the theorems proposed in the main manuscript. This document is organized as follows: Section A presents the detailed proofs of Theorem 1. Section B elaborates on the proof of Theorem 2.

A PROOFS OF THEOREM 1

Theorem 1. *Suppose $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with t-SVD $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ and $\beta > 0$. The Enhanced Tensorial Rank Minimization problem (ETRM) can be described as follows:*

$$\arg\min_{\mathcal{G}} \beta \|\mathcal{G}\|_{ETR} + \frac{1}{2} \|\mathcal{G} - \mathcal{A}\|_F^2 . \tag{1}$$

Then, optimal solution $\mathcal{G}^*$ is obtained as:

$$\mathcal{G}^* = \mathcal{U} * ifft(Prox_{f,\beta}(\mathcal{S}_f), [], 3) * \mathcal{V}^T, \tag{2}$$

where $ifft(Prox_{f,\beta}(\mathcal{S}_f), [], 3) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f-diagonal tensor, and $Prox_{f,\beta}(\mathcal{S}_f^k(i,i))$ satisfies the following equation

$$Prox_{f,\beta}(\mathcal{S}_f^k(i,i)) = \arg\min_{x \ge 0} \frac{1}{2}(x - \mathcal{S}_f^k(i,i))^2 + \beta f(x), \tag{3}$$

where $f(x) = \frac{e^{\delta^2} x}{\delta + x}$.

To prove Theorem 1, we first introduce the following lemma.

Lemma 1. *[2] Given $\mathbf{G}, \mathbf{A} \in \mathbb{R}^{m \times n}$, and $\mathbf{A} = \mathbf{U}\mathbf{S}_A\mathbf{V}^T$ is the SVD of $\mathbf{A}$ and $\beta > 0$, then an optimal solution to the following problem*

$$\min_{\mathbf{G}} \beta \|\mathbf{G}\|_{ETR} + \frac{1}{2} \|\mathbf{G} - \mathbf{A}\|_F^2 , \tag{4}$$

is $\mathbf{G}^* = \mathbf{U}\mathbf{S}_G^*\mathbf{V}^T$, where $\mathbf{S}_G^* = diag(\sigma^*)$ and $\sigma^* = prox_{f,\beta}(\sigma_A)$. And $prox_{f,\beta}(\sigma_A)$ is the Moreau-Yosida operator [3] defined as:

$$prox_{f,\beta}(\sigma_A) := \arg\min_{\sigma \ge 0} \beta f(\sigma) + \frac{1}{2} \|\sigma - \sigma_A\|_2^2 , \tag{5}$$

where $f(x) = \frac{e^{\delta^2} x}{\delta + x}$.

Proof In Fourier domain, there is a fact that $\|\mathcal{X}\|_F^2 = \frac{1}{n_3} \left\|\mathcal{X}_f\right\|_F^2$, so the objective function $\frac{1}{2} \|\mathcal{G} - \mathcal{A}\|_F^2 + \beta \|\mathcal{A}\|_{ETR}$ can be rewritten as:

$$\begin{aligned} &\frac{1}{2} \|\mathcal{G} - \mathcal{A}\|_F^2 + \beta \|\mathcal{A}\|_{ETR} \\ =&\frac{1}{2n_3} \left\|\mathcal{G}_f - \mathcal{A}_f\right\|_F^2 + \frac{\beta}{n_3} \sum_{k=1}^{n_3} \left\|\mathcal{A}_f^k\right\|_{ETR} \\ =&\frac{1}{n_3} \sum_{k=1}^{n_3} \left(\frac{1}{2} \left\|\mathcal{G}_f^k - \mathcal{A}_f^k\right\|_F^2 + \beta \left\|\mathcal{A}_f^k\right\|_{ETR} \right) \end{aligned} \tag{6}$$

Thus, the original tensor optimization problem can be transformed into n_3 independent matrix optimization problems as follows:

$$\arg\min_{\mathcal{G}_f^k} \frac{1}{2} \left\|\mathcal{G}_f^k - \mathcal{A}_f^k\right\|_F^2 + \beta \left\|\mathcal{A}_f^k\right\|_{ETR}, \tag{7}$$

for $1 \le k \le n_3$.

Here, the SVD of $\mathcal{A}_f^k$ is $\mathcal{A}_f^k = \mathcal{U}_f^k \mathcal{S}_f^k (\mathcal{V}_f^k)^H$. According to Lemmas 1, the optimal solution of Eq. (7) is

$$\mathcal{G}_f^{*k} = \mathcal{U}_f^k Prox_{f,\beta}(\mathcal{S}_f^k)(\mathcal{V}_f^k)^H, \tag{8}$$

where $Prox_{f,\beta}(\mathcal{S}_f^k(i,i))$ is given by solving the following problem:

$$Prox_{f,\beta}(\mathcal{S}_f^k(i,i)) = \arg\min_{x \ge 0} \frac{1}{2}(x - \mathcal{S}_f^k(i,i))^2 + \beta f(x) \tag{9}$$

where $f(x) = \frac{e^{\delta^2} x}{\delta + x}$. □

B PROOFS OF THEOREM 2

Theorem 2. *Let $\{\mathcal{P}_k = (\mathbf{Z}_k^v, \mathbf{C}_k^v, \mathbf{E}_k^v, \mathbf{Y}_k^v, \mathcal{W}_k, \mathcal{G}_k)\}_{k=1}^{\infty}$ be the sequence generated by the Algorithm 1 in the main manuscript, then the sequence $\{\mathcal{P}_k\}_{k=1}^{\infty}$ meets the following two principles:*

1). $\{\mathcal{P}_k\}_{k=1}^{\infty}$ is bounded;
2). Any accumulation point of $\{\mathcal{P}_k\}_{k=1}^{\infty}$ is a KKT point of Algorithm 1.

To prove Theorem 2, we first introduce two important lemmas.

Lemma 2. *[5] Let $\mathcal{H}$ be a real Hilbert space endowed with an inner product $\langle \cdot, \cdot \rangle$, a norm $\|\cdot\|$ with the dual norm $\|\cdot\|^{dual}$, and $y \in \partial\|x\|$, where $\partial f(\cdot)$ is the subgradient of $f(\cdot)$. Then we have $\|y\|^{dual} = 1$ if $x \ne 0$, and $\|y\|^{dual} \le 1$ if $x = 0$.*

Lemma 3. *[4] Suppose that $F : \mathbb{R}^{m \times n} \to \mathbb{R}$ is defined as $F(\mathbf{X}) = f \circ \boldsymbol{\sigma}(\mathbf{X}) = f(\sigma_1(\mathbf{X}), \ldots, \sigma_r(\mathbf{X}))$, where $\mathbf{X} = \mathbf{U}Diag(\boldsymbol{\sigma}(\mathbf{X}))\mathbf{V}^T$ is SVD of matrix $X \in \mathbb{R}^{m \times n}$, $r = \min(\mathrm{m}, \mathrm{n})$, and $f(\cdot) : \mathbb{R}^r \to \mathbb{R}$ be differentiable and absolutely symmetric at $\boldsymbol{\sigma}(\mathbf{X})$. Then, the subdifferential of $F(\mathbf{X})$ at $\mathbf{X}$ is*

$$\frac{\partial F(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{U}Diag(\partial f(\boldsymbol{\sigma}(\mathbf{X})))\mathbf{V}^T, \tag{10}$$

where $\partial f(\boldsymbol{\sigma}(\mathbf{X})) = (\frac{\partial f(\sigma_1(\mathbf{x}))}{\partial \mathbf{X}}), \ldots, \frac{\partial f(\sigma_r(\mathbf{x}))}{\partial \mathbf{X}})$.

Proof **1). Proof of** $1st$ **principle:** On the $k+1$ iteration, from the updating rule of $\mathbf{E}_{k+1}^v$, the first-order optimal condition should be satisfied.

$$\begin{aligned} 0 &= \alpha\partial \left\|\mathbf{E}_{k+1}^v\right\|_{2,1} \\ &+ \mu_k \big(\mathbf{E}_{k+1}^v - (\mathbf{X}^v - (\mathbf{C}_{k+1}^v + \mathbf{Z}_{k+1}^v)\mathbf{H}^v + \mathbf{Y}_k^v/\mu_k)\big) \\ &= \alpha\partial \left\|\mathbf{E}_{k+1}^v\right\|_{2,1} - \mathbf{Y}_{k+1}^v, \end{aligned} \tag{11}$$

Thus, we have

$$\frac{1}{\alpha}[\mathbf{Y}^v_{k+1}]_{:,j} = \partial \left\|[\mathbf{E}^v_{k+1}]_{:,j}\right\|_2, \tag{12}$$

where $[\mathbf{Y}^v_{k+1}]_{:,j}$ and $[\mathbf{E}^v_{k+1}]_{:,j}$ are the j-th columns of $\mathbf{Y}^v_{k+1}$ and $\mathbf{E}^v_{k+1}$. And the ℓ_2 norm is self-dual, so based on the Lemma 2, we have $\left\|\frac{1}{\alpha}[\mathbf{Y}^v_{k+1}]_{:,j}\right\|_2 \leq 1$. So the sequence $\{\mathbf{Y}^v_{k+1}\}$ is bounded.

Then, we prove the sequence $\{\mathcal{W}_{k+1}\}$ is bounded. According to the updating rule of $\mathcal{G}$, the first-order optimality condition holds

$$\mathbf{0} \in \partial \|\mathcal{G}_{k+1}\|_{ETR} + \mu_k(\mathcal{G}_{k+1} - (\mathcal{Z}_{k+1} - \mathcal{W}_{k+1}/\mu_k)). \tag{13}$$

According to rule $\mathcal{W}_{k+1} = \mathcal{W}_k + \mu_k(\mathcal{Z}_{k+1} - \mathcal{G}_{k+1})$, we have

$$\mathcal{W}_{k+1} \in \partial \|\mathcal{G}_{k+1}\|_{ETR}. \tag{14}$$

Let $\mathcal{U} * \mathcal{S} * \mathcal{V}^T$ be the t-SVD of tensor $\mathcal{G}$. According to the Lemma 3 and definition of ETR, we have:

$$\begin{aligned}
&\left\|\partial \|\mathcal{G}_{k+1}\|_{ETR}\right\|_F^2 \\
&= \left\|\frac{1}{n_3}\mathcal{U} * ifft(\partial(\mathcal{S}_f), [], 3) * \mathcal{V}^T\right\|_F^2 \\
&= \frac{1}{n_3^3}\left\|\partial f(\mathcal{S}_f)\right\|_F^2 \\
&\leq \frac{1}{n_3^3}\sum_{i=1}^{n_3}\sum_{j=1}^{min(n_1,n_2)}[\partial f(\mathcal{S}_f^v(j,j))]^2 \\
&\leq \frac{e^{2\delta^2} min(n_1,n_2)}{\delta^2 n_3^2}
\end{aligned} \tag{15}$$

where the second inequality is by the fact $\partial f(x) \leq \frac{e^{\delta^2}}{\delta}$, and $f(x) = \frac{e^{\delta^2} x}{\delta + x}$ is our rank approximation function. So $\partial \|\mathcal{G}_{k+1}\|_{ETR}$ is bounded, meanwhile the sequence $\{\mathcal{W}_{k+1}\}$ is also bounded.

Moreover, from the iterative method in the algorithm of solving CFMVC-ETR, we can deduce

$$\begin{aligned}
&\mathcal{L}_{\mu_k,\rho_k}(\mathbf{Z}^v_{k+1}, \mathbf{C}^v_{k+1}, \mathbf{E}^v_{k+1}, \mathbf{H}^v_{k+1}, \mathcal{G}_{k+1}, \mathbf{Y}^v_k, \mathcal{W}_k) \\
&\leq \mathcal{L}_{\mu_k,\rho_k}(\mathbf{Z}^v_k, \mathbf{C}^v_k, \mathbf{E}^v_k, \mathbf{H}^v_k, \mathbf{H}^v_k, \mathcal{G}_k, \mathbf{Y}^v_k, \mathcal{W}_k) \\
&= \mathcal{L}_{\mu_{k-1},\rho_{k-1}}(\mathbf{Z}^v_k, \mathbf{C}^v_k, \mathbf{E}^v_k, \mathbf{H}^v_k, \mathcal{G}_k, \mathbf{Y}^v_{k-1}, \mathcal{W}_{k-1}) \\
&+ \frac{\rho_k + \rho_{k-1}}{2\rho_{k-1}^2}\|\mathcal{W}_k - \mathcal{W}_{k-1}\|_F^2 \\
&+ \frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2}\sum_{v=1}^m \left\|\mathbf{Y}^v_k - \mathbf{Y}^v_{k-1}\right\|_F^2,
\end{aligned} \tag{16}$$

Thus, summing two sides of (16) form $k = 1$ to n,

$$\begin{aligned}
&\mathcal{L}_{\mu_k,\rho_k}(\mathbf{Z}^v_{k+1}, \mathbf{C}^v_{k+1}, \mathbf{E}^v_{k+1}, \mathbf{H}^v_{k+1}, \mathcal{G}_{k+1}, \mathbf{Y}^v_k, \mathcal{W}_k) \\
&\leq \mathcal{L}_{\mu_0,\rho_0}(\mathbf{Z}^v_1, \mathbf{C}^v_1, \mathbf{E}^v_1, \mathbf{H}^v_1, \mathcal{G}_1, \mathbf{Y}^v_0, \mathcal{W}_0) \\
&+ \sum_{k=1}^n \frac{\rho_k + \rho_{k-1}}{2\rho_{k-1}^2}\|\mathcal{W}_k - \mathcal{W}_{k-1}\|_F^2 \\
&+ \sum_{k=1}^n \left(\frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2}\sum_{v=1}^m \left\|\mathbf{Y}^v_k - \mathbf{Y}^v_{k-1}\right\|_F^2\right)
\end{aligned} \tag{17}$$

Observe that

$$\sum_{k=1}^n \frac{\mu_k + \mu_{k+1}}{2\mu_{k-1}^2} < \infty, \quad \sum_{k=1}^n \frac{\rho_k + \rho_{k+1}}{2\rho_{k-1}^2} < \infty \tag{18}$$

Note that $\mathcal{L}_{\mu_0,\rho_0}(\mathbf{Z}^v_1, \mathbf{C}^v_1, \mathbf{E}^v_1, \mathbf{H}^v_1, \mathcal{G}_1, \mathbf{Y}^v_0, \mathcal{W}_0)$ is finite, and sequence $\{\mathbf{Y}^v_k\}, \{\mathcal{W}_k\}$, $\sum_{k=1}^n \frac{\mu_k + \mu_{k+1}}{2\mu_{k-1}^2}$ and $\sum_{k=1}^n \frac{\rho_k + \rho_{k+1}}{2\rho_{k-1}^2}$ are all bounded. So $\mathcal{L}_{\mu_k}(\mathbf{Z}^v_{k+1}, \mathbf{E}^v_{k+1}, \mathbf{H}^v_{k+1}, \mathcal{G}_{k+1}, \mathbf{Y}^v_k, \mathcal{W}_k)$ is bounded. Notice

$$\begin{aligned}
&\mathcal{L}_{\mu_k,\rho_k}(\mathbf{Z}^v_{k+1}, \mathbf{C}^v_{k+1}, \mathbf{E}^v_{k+1}, \mathbf{H}^v_{k+1}, \mathcal{G}_{k+1}, \mathbf{Y}^v_k, \mathcal{W}_k) \\
&= \|\mathcal{G}_{k+1}\|_{ETR} + \alpha\|\mathbf{E}_{k+1}\|_{2,1} + \gamma\|\mathcal{C}_{k+1}\|_{TER} \\
&+ \sum_{v=1}^m (\langle \mathbf{Y}^v_k, \mathbf{X}^v - (\mathbf{Z}^v_{k+1} + \mathbf{C}^v_{k+1})\mathbf{H}^v_{k+1} - \mathbf{E}^v_{k+1}\rangle \\
&+ \frac{\mu_k}{2}\left\|\mathbf{X}^v - (\mathbf{Z}^v_{k+1} + \mathbf{C}^v_{k+1})\mathbf{H}^v_{k+1} - \mathbf{E}^v_{k+1}\right\|_F^2) \\
&+ \langle \mathcal{W}_k, \mathcal{Z}_{k+1} - \mathcal{G}_{k+1}\rangle + \frac{\rho_k}{2}\|\mathcal{Z}_{k+1} - \mathcal{G}_{k+1}\|_F^2,
\end{aligned} \tag{19}$$

and each term of Eq. (19) is nonnegative, due to the boundedness of $\mathcal{L}_{\mu_k}(\mathbf{Z}^v_{k+1}, \mathbf{C}^v_{k+1}, \mathbf{E}^v_{k+1}, \mathbf{H}^v_{k+1}, \mathcal{G}_{k+1}, \mathbf{Y}^v_k, \mathcal{W}_k)$, we can deduce each term of Eq. (19) is bounded. So the boundedness of $\|\mathcal{G}_{k+1}\|_{ETR}$ implies that all singular values of $\mathcal{G}_{k+1}$ are bounded. Furthermore, based on the following equation

$$\|\mathcal{G}_{k+1}\|_F^2 = \frac{1}{n_3}\left\|\mathcal{G}_{f,k+1}\right\|_F^2 = \frac{1}{n_3}\sum_{i=1}^{n_3}\sum_{j=1}^{min(n_1,n_2)}(\mathcal{S}_f^i(j,j))^2, \tag{20}$$

we can derive the sequence $\{\mathcal{G}_{k+1}\}$ is bounded, then, it is easy to prove the boundedness of $\{\mathbf{Z}_{k+1}\}$, $\{\mathbf{C}_{k+1}\}$ and $\{\mathbf{H}_{k+1}\}$.

Therefore, from the above proof, we can conclude that the sequence $\{\mathcal{P}_k = (\mathbf{Z}^v_k, \mathbf{C}^v_k, \mathbf{E}^v_k, \mathbf{H}^v_k, \mathbf{Y}^v_k, \mathcal{W}_k, \mathcal{G}_k)\}_{k=1}^\infty$ generated by the Algorithm 1 is bounded.

2). Proof of $2nd$ **principle:** According to Weierstrass-Bolzano theorem [1], there is at least one accumulation point of the sequence $\{\mathcal{P}_k\}_{k=1}^\infty$, we denote one of the points as $\mathcal{P}_*$. Then we have

$$\lim_{k\to\infty}(\mathbf{Z}^v_k, \mathbf{C}^v_k, \mathbf{E}^v_k, \mathbf{H}^v_k, \mathbf{Y}^v_k, \mathcal{W}_k, \mathcal{G}_k) = (\mathbf{Z}^v_*, \mathbf{C}^v_*, \mathbf{E}^v_*, \mathbf{H}^v_*, \mathbf{Y}^v_*, \mathcal{W}_*, \mathcal{G}_*). \tag{21}$$

Form the updating rule of $\mathcal{W}$ and $\mathbf{Y}^v$, we have the following equations:

$$\begin{aligned}
&\mathbf{X}^v - (\mathbf{Z}^v_{k+1} + \mathbf{C}^v_{k+1})\mathbf{H}^v_{K+1} - \mathbf{E}^v_{k+1} = (\mathbf{Y}^v_{k+1} - \mathbf{Y}^v_k)/\mu_t, \\
&\mathcal{Z}_{k+1} - \mathcal{G}_{k+1} = (\mathcal{W}_{k+1} - \mathcal{W}_k)/\rho_t.
\end{aligned} \tag{22}$$

According the boundedness of sequences $\{\mathcal{W}_k\}$ and $\{\mathbf{Y}^v_k\}$, and the fact $\lim_{k\to\infty}\mu_k, \rho_k = \infty$, we have:

$$\begin{aligned}
&\lim_{k\to\infty}\mathbf{X}^v - (\mathbf{Z}^v_{k+1} + \mathbf{C}^v_{k+1})\mathbf{H}^v_{K+1} - \mathbf{E}^v_{k+1} = \lim_{k\to\infty}(\mathbf{Y}^v_{k+1} - \mathbf{Y}^v_k)/\mu_t = 0, \\
&\lim_{k\to\infty}\mathcal{Z}_{k+1} - \mathcal{G}_{k+1} = \lim_{k\to\infty}(\mathcal{W}_{k+1} - \mathcal{W}_k)/\rho_t = 0,
\end{aligned} \tag{23}$$

then, we can obtain

$$\mathbf{X}^v - (\mathbf{Z}^v_* + \mathbf{C}^v_*)\mathbf{H}^v_* - \mathbf{E}^v_* = 0, \quad \mathcal{Z}_* - \mathcal{G}_* = 0. \tag{24}$$

Furthermore, due to the first-order optimality conditions of $\mathbf{E}^v_{k+1}$ and $\mathcal{G}_{k+1}$, we can deduce:

$$\begin{aligned}
&0 \in \alpha\partial\left\|\mathbf{E}^v_{k+1}\right\|_{2,1} - \mathbf{Y}^v_{k+1} \Rightarrow \mathbf{Y}^v_* = \alpha\partial\left\|\mathbf{E}^v_*\right\|_{2,1} \\
&0 \in \partial\|\mathcal{G}_{k+1}\|_{ETR} - \mathcal{W}_{k+1} \Rightarrow \mathcal{W}_* = \partial\|\mathcal{G}_*\|_{ETR}
\end{aligned} \tag{25}$$

Thus, the accumulation point $\mathcal{P}_*$ of sequence $\{\mathcal{P}_k\}_{k=1}^\infty$ generated by the algorithm 1 in the main manuscript satisfies the KKT condition. □

REFERENCES

[1] Robert G Bartle and Donald R Sherbert. 2000. *Introduction to real analysis*. Vol. 2. Wiley New York.

[2] Zhao Kang, Chong Peng, and Qiang Cheng. 2015. Robust PCA via nonconvex rank approximation. In *2015 IEEE International Conference on Data Mining*. IEEE, 211–220.

[3] Claude Lemaréchal and Claudia Sagastizábal. 1997. Practical aspects of the Moreau–Yosida regularization: Theoretical preliminaries. *SIAM journal on optimization* 7, 2 (1997), 367–385.

[4] Adrian S Lewis and Hristo S Sendov. 2005. Nonsmooth analysis of singular values. Part I: Theory. *Set-Valued Analysis* 13, 3 (2005), 213–241.

[5] Zhouchen Lin, Minming Chen, and Yi Ma. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055* (2010).