Supplementary Materials: Improved Weighted Tensor Schatten p-Norm for Fast Multi-view Graph Clustering

Anonymous reviewers

1 SUMMARY

In this supplementary material, we provide additional information. Concretely, we mainly supplement "Complexity analysis of competitors" in Section 2.1, "Proofs of **Theorem 1** and **Theorem 2**" in Section 2.2, and "Parameter Analysis and Convergence" in Section 2.3, respectively. The codes and benchmark dataset for this paper are also provided for repetition.

2 CONTENT

2.1 Complexity Analysis

As shown in Table 1, both space complexity and time complexity are linear to n, which enables handling large-scale datasets with $100,000 \leq n$.

Table 1: Complexity analysis.

Method	Space Complexity ≈	Time Complexity ≈
Our method	$O(n)$	$O(n)$
TBGL [13]	$O(n)$	$O(n)$
MVBGC [5]	$O(n)$	$O(n)$
FastMICE [2]	$O(n)$	$O(n)$
UDBGL [1]	$O(n)$	$O(n)$
SDAFG [9]	$O(n)$	$O(n)$
FPMVS [12]	$O(n)$	$O(n)$
FMCNOF [16]	$O(n)$	$O(n)$
SFMC [7]	$O(n)$	$O(n)$
SMVSC [11]	$O(n)$	$O(n)$
LMVSC [4]	$O(n)$	$O(n)$
BMVC [17]	$O(n)$	$O(n)$
PMSC [3]	$O(n^2)$	$O(n^3)$
FMR [6]	$O(n^2)$	$O(n^3)$
AMGL [10]	$O(n^2)$	$O(n^3)$

2.2 Optimization

$$\min_{\mathbf{B}^r,\mathbf{W}^r,\mathbf{A}^r,\mathcal{J}} \gamma \sum_{r=1}^{v}\sum_{i=1}^{m}\sum_{j=1}^{n} \|\mathbf{b}_i^r - (\mathbf{W}^r)^\top \mathbf{x}_j^r\|_F^2 a_{ij}^r + \alpha\|\mathbf{A}^r\|_F^2 + \|\mathcal{A}\|_{\mathbb{I}\omega,S_p}$$
$$\text{s.t.} (\mathbf{A}^r)^\top \mathbf{1} = \mathbf{1}, \mathbf{A}^r \geq 0, (\mathbf{W}^r)^\top \mathbf{W}^r = \mathbf{I}_m, (\mathbf{B}^r)^\top \mathbf{B}^r = \mathbf{I}_m \tag{1}$$

ACM MM, 2024, Melbourne, Australia

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

To solve Eq. (1), the auxiliary variable $\mathcal{K}$ is introduced to make Eq. (1) separable. Then, Eq. (1) is rewritten as the following augmented Lagrangian function

$$\min_{\substack{\mathbf{B}^r,\mathbf{W}^r,\\ \mathbf{A}^r,\mathcal{Y},\mathcal{K}}} \gamma \sum_{r=1}^{v}\sum_{i=1}^{m}\sum_{j=1}^{n} \|\mathbf{b}_i^r - (\mathbf{W}^r)^\top \mathbf{x}_j^r\|_F^2 a_{ij}^r + \alpha\|\mathbf{A}^r\|_F^2$$
$$+ \|\mathcal{K}\|_{\mathbb{I}\omega,S_p} + \frac{\mu}{2}\|\mathcal{J} - \mathcal{K} + \frac{\mathcal{Y}}{\mu}\|_F^2 \tag{2}$$
$$\text{s.t. } \mathbf{A}^r \geq 0, (\mathbf{A}^r)^\top \mathbf{1} = \mathbf{1}, (\mathbf{W}^r)^\top \mathbf{W}^r = \mathbf{I}_m, (\mathbf{B}^r)^\top \mathbf{B}^r = \mathbf{I}_m,$$
$$\mathcal{J} = \Psi(\mathbf{A}^1, \mathbf{A}^r, \cdots, \mathbf{A}), \mathcal{J} = \mathcal{K}$$

Eq. (4) could be separately solved by developing an alternating iterative algorithm as follows.

▷ **Step-1: Solving $\mathbf{A}^r$ with B, W, $\mathbf{A}^b$, and $\mathcal{K}$ fixed.** Then, $\mathbf{A}^r$-subproblem of Eq. (4) changes to

$$\min_{a_{ij}^r \geq 0, (\mathbf{a}^r)^\top \mathbf{1} = 1} \gamma \sum_{r=1}^{v}\sum_{i=1}^{m}\sum_{j=1}^{n} \|\mathbf{b}_i^r - (\mathbf{W}^r)^\top \mathbf{x}_j^r\|_F^2 a_{ij}^r$$
$$+ \alpha\|\mathbf{A}^r\|_F^2 + \frac{\mu}{2}\|\mathbf{A}^r - \mathbf{K}^r + \frac{\mathbf{Y}^r}{\mu}\|_F^2 \tag{3}$$

Denote $\|\mathbf{b}_i^r - (\mathbf{W}^r)^\top \mathbf{x}_j^r\|_F^2 = d_{ij}^r$, with elimination of irrelevant variables for r-th view, Eq. (3) becomes to the following problem

$$\min_{a_{ij}^r \geq 0, (\mathbf{a}_j^r)^\top \mathbf{1} = 1} \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha(a_{ij}^r)^2 + \frac{\mu}{2}(a_{ij}^r)^2 - 2[\frac{1}{2}\mu f_{ij}^r - \frac{1}{2}\gamma d_{ij}^r] a_{ij}^r$$
$$\Leftrightarrow \min_{a_{ij}^r \geq 0, (\mathbf{a}^r)^\top \mathbf{1} = 1} \sum_{i=1}^{m}\sum_{j=1}^{n} (a_{ij}^r)^2 - 2\frac{[\frac{1}{2}\mu f_{ij}^r - \frac{1}{2}\gamma d_{ij}^r]}{(\alpha + \frac{\mu}{2})} a_{ij}^r$$
$$\Leftrightarrow \min_{a_{ij}^r \geq 0, (\mathbf{a}_j^r)^\top \mathbf{1} = 1} \|a_{ij}^r - 2\frac{\frac{1}{2}\mu f_{ij}^r - \frac{1}{2}\gamma d_{ij}^r}{(\alpha + \frac{\mu}{2})}\|_F^2 \tag{4}$$

where f_{ij}^r is element of $\mathbf{F}^r = \mathbf{T}^r - \frac{\mathbf{Y}^r}{\mu}$, respectively. Then, updating $\mathbf{a}^r$ converts to the following column form

$$\min_{\mathbf{a}_j} \left\|\mathbf{a}_j - \hat{\mathbf{a}}_j\right\|_F^2, \text{ s.t. } \forall ij, \mathbf{a}_j \mathbf{1} = 1, a_{ij} \geq 0 \tag{5}$$

where $\hat{\mathbf{a}}_j^r = \frac{\frac{1}{2}\mu \mathbf{f}_j^r - \frac{1}{2}\gamma \mathbf{d}_j^r}{(\alpha+\mu)}$. Each column $\mathbf{a}_j$ could be optimized via the following **Theorem 1**.

Theorem 1. *Given arbitrary v vectors $\{\hat{\mathbf{a}}_j\}_{j=1}^{v}$, we obtain the following closed-form solution $\mathbf{a}_j^*$*

$$\mathbf{a}_j^* = \arg\min_{\mathbf{a}_j} \left\|\mathbf{a}_j - \hat{\mathbf{a}_j}\right\|_F^2, \text{ s.t. } \mathbf{a}_j^\top \mathbf{1} = 1, \mathbf{a}_j \geq 0 \tag{6}$$

Proof. For ease of presentation, we first substitute $\mathbf{a}_j$ and $\hat{\mathbf{a}_j}$ to the vectors $\mathbf{a}$ and $\mathbf{g}$, respectively. Then, by denoting the Lagrange multipliers φ and $\mathbf{e}$ for the constraints $\mathbf{a}^\top \mathbf{1} = 1$ and $\mathbf{a}_j \geq 0$, respectively, the Lagrangian function of Eq. (6) is rewritten as

$$\mathcal{L}(\mathbf{a}, \varphi, \mathbf{e}) = \frac{1}{2}\|\mathbf{a} + \mathbf{g}\|_2^2 - \varphi(\mathbf{a}^\top \mathbf{1} - 1) - \mathbf{e}^\top \mathbf{a}\,. \tag{7}$$

We suppose that the optimal solutions of $\mathbf{a}$, φ, and $\mathbf{e}$ in (27) are $\mathbf{a}^*$, φ^*, and $\mathbf{e}^*$, respectively. According to KKT condition, we have the following equation:

$$\begin{cases} \forall j, \quad a_{ij}^* + g_j - \varphi^* - e_j^* = 0 \\ \forall j, \quad a_{ij}^* e_j^* = 0 \\ \forall j, \quad a_{ij}^* \geqslant 0 \\ \forall j, \quad e_j^* \geqslant 0\,, \end{cases} . \tag{8}$$

Writing the first term of Eq.(28) in a vector form as $\mathbf{a}^* + \mathbf{g} - \varphi^* \mathbf{1} - \mathbf{e}^* = 0$. Considering the constraint $\mathbf{a}^\top \mathbf{1} = 1$, we derive $\varphi^* = \frac{\mathbf{1}^\top \mathbf{g} - \mathbf{1}^\top \mathbf{e}^*}{n}$ since $\mathbf{g}^\top \mathbf{1} = \mathbf{1}^\top \mathbf{g}$ and $(\mathbf{e}^*)^\top \mathbf{1} = \mathbf{1}^\top \mathbf{e}^*$. Thus, the optimal solution $\mathbf{a}^*$ can be solved by

$$\mathbf{a}^* = \frac{1}{n}\left(1 - \mathbf{1}^\top \mathbf{e}^*\right)\mathbf{1} - \mathbf{g} - \frac{1}{n}\mathbf{1}^\top \mathbf{g}\mathbf{1} + \mathbf{e}^*\,, \tag{9}$$

and its scalar form becomes

$$a_{ij}^* = \frac{1}{n} - \frac{\mathbf{1}^\top \mathbf{e}^*}{n} - g_j + \frac{1}{n}\mathbf{1}^\top \mathbf{g} + e_j^*\,. \tag{10}$$

Denote $\hat{e^*} = \frac{\mathbf{1}^\top \mathbf{e}^*}{n}$ and $\hat{g}_j = \frac{1}{n} - g_j + \frac{1}{n}\mathbf{1}^\top \mathbf{g}$, then a_{ij}^* becomes

$$a_{ij}^* = \hat{g}_j + e_j^* - \hat{e^*},\ \forall j \tag{11}$$

According to the $2-4$ terms of Eq.(28) and Eq.(31), we have $a_{ij}^* = \lfloor \hat{g}_j - \hat{e^*} \rfloor_+$, where $\lfloor \hat{g}_j - \hat{e^*} \rfloor_+ = \max(0, \hat{g}_j - \hat{e^*})$. Namely, if the optimal solution $\hat{e^*}$ is known, we can gain the corresponding optimal solution a_{ij}^*. We gain $e_j^* = \hat{e^*} + a_{ij}^* - \hat{g}_j$ due to Eq.(31) and $e_j^* = \lfloor \hat{e^*} - \hat{g}_j \rfloor_+$ according to the $2-4$ terms of Eq.(28). Note that we have denoted $\hat{e^*} = \frac{\mathbf{1}^\top \mathbf{e}^*}{n}$, thus we gain the optimal solution of $\hat{e^*}$ as

$$\hat{e^*} = \frac{1}{n}\sum_{j=1}^{n} \lfloor \hat{e^*} - \hat{g}_j \rfloor_+\,. \tag{12}$$

So far, the Newton method can be employed to obtain the optimal solution $\hat{e^*}$ by defining a cost function as

$$\boldsymbol{\Psi}(\hat{e}) = \frac{1}{n}\sum_{j=1}^{n} \lfloor \hat{e} - \hat{g}_j \rfloor_+ - \hat{e}\,. \tag{13}$$

where Ψ denotes the functional symbol. When the root searching problem of $\boldsymbol{\Psi}(\hat{e}) = 0$ is solved, the optimal $\hat{e}^*$ will be obtained. Since $\hat{e} \geq 0$, $\boldsymbol{\Psi}'(\hat{e}^t) \leq 0$, and $\boldsymbol{\Psi}'(\hat{e}^t) \leq 0$ is a piecewise linear and convex function, the Newton method can be used to obtain the root of $\boldsymbol{\Psi}(\hat{e}) = 0$ via

$$\hat{e}^{t+1} = \hat{e}^t - \boldsymbol{\Psi}\left(\hat{e}^t\right) \cdot \left[\frac{\partial \boldsymbol{\Psi}\left(\hat{e}^t\right)}{\partial \hat{e}^t}\right]^{-1} \tag{14}$$

where $\boldsymbol{\Psi}'(\hat{e}^t) = \frac{\partial \boldsymbol{\Psi}(\hat{e}^t)}{\partial \hat{e}^t}$ is first-order derivative of $\boldsymbol{\Psi}(\hat{e}^t)$. □

▷ **Update-2: Solving W with A, B, and $\mathcal{W}$ fixed.** In this case, $\mathbf{W}$-subproblem of Eq. (4) can be written as

$$\max_{\mathbf{W}^r} \mathrm{Tr}((\mathbf{W}^r)^\top \mathbf{E}^r) \text{ s.t.} \mathbf{W}^r(\mathbf{W}^r)^\top = \mathbf{I}_k, \tag{15}$$

where $\mathbf{E}^r = \mathbf{X}^r(\mathbf{A}^r)^\top(\mathbf{B}^r)^\top$. Eq. (15) can be solved via the Singular Value Decomposition (SVD) **Theorem 2** with complexity $\boldsymbol{O}(v\hat{d}(nm + k^2 + km))$ for each iteration, where $\hat{d} = \sum_{p=1}^{v} d^r$.

Theorem 2. Letting the SVD of $\mathbf{E} \in \mathbb{R}^{d\times l}$ be $\mathbf{E} = \mathbf{U}\mathbf{G}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{d\times l}$, $\mathbf{G} \in \mathbb{R}^{l\times l}$ and $\mathbf{V} \in \mathbb{R}^{l\times l}$, the optimal solution of $\max\limits_{\mathbf{W}^\top\mathbf{W}=\mathbf{I}} \mathrm{Tr}(\mathbf{W}^\top \mathbf{E})$ is $\mathbf{W} = \mathbf{U}\mathbf{V}^\top$.

Proof. Letting $\mathbf{E} = \mathbf{U}\mathbf{G}\mathbf{V}^\top \in \mathbb{R}^{d\times l}$ and according to the rotation invariance of trace norm, $\mathrm{Tr}(\mathbf{W}^\top \mathbf{E})$ converts to

$$\mathrm{Tr}(\mathbf{W}^\top \mathbf{E}) = \mathrm{Tr}(\mathbf{W}^\top \mathbf{U}\mathbf{G}\mathbf{V}^\top) = \mathrm{Tr}(\mathbf{G}\mathbf{V}^\top \mathbf{W}^\top \mathbf{U}) \tag{16}$$

Denote $\mathbf{D} = \mathbf{V}^\top \mathbf{W}^\top \mathbf{U}$, we have $\mathrm{Tr}(\mathbf{W}^\top \mathbf{E}) = \mathrm{Tr}(\mathbf{G}\mathbf{D})$, where $\mathbf{G}$ and $\mathbf{D}$ denote their i-th diagonal elements as g_{ii} and d_{ii}, respectively. Considering $\mathbf{D}\mathbf{D}^\top = \mathbf{I}$, we have $|d_{ii}| \leq 1$;. Further, the singular value g_{ii} should enjoy $g_{ii} \geq 0$. So far, we could deduce

$$\mathrm{Tr}(\mathbf{W}^\top \mathbf{E}) = \mathrm{Tr}(\mathbf{G}\mathbf{D}) = \sum_{i=1}^{k} g_{ii} d_{ii} \leq \sum_{i=1}^{b} g_{ii} \tag{17}$$

Based on inequality Eq.(17), the maximization of $\mathrm{Tr}(\mathbf{W}^\top \mathbf{E})$ could be reached when $d_{ii} = 1$. This further deduces $\mathbf{W} = \mathbf{U}[\mathbf{I}_d; 0]\mathbf{V}^\top$, where $\mathbf{D} = [\mathbf{I}_d; 0] \in \mathbb{R}^{l\times d}$ and $\mathbf{D} = \mathbf{V}^\top \mathbf{W}^\top \mathbf{U}$. According to the above analysis, the potimal solution of $\max\limits_{\mathbf{W}^\top\mathbf{W}=\mathbf{I}} \mathrm{Tr}(\mathbf{W}^\top \mathbf{E})$ is $\mathbf{W} = \mathbf{U}\mathbf{V}^\top$ by performing the tiny SVD on $\mathbf{E}$. □

▷ **Step-3 update $\mathbf{B}^r$:** Optimizing $\mathbf{B}^r$ with the irrelevant variables fixed is equivalent to the following optimization problem

$$\max_{\mathbf{B}^r} \mathrm{Tr}((\mathbf{B}^r)^\top \mathbf{C}^r) \text{ s.t.} (\mathbf{B}^r)^\top \mathbf{B}^r = \mathbf{I}_m \tag{18}$$

where $\mathbf{C}^r = \gamma(\mathbf{W}^r)^\top \mathbf{X}^r(\mathbf{A}^r)^\top$. The optimal solution of optimizing $\mathbf{B}^r$ can be effectively obtained via **Theorem 2**.

▷ **Step-4 update $\mathcal{K}$:** Ignoring the irrelevant items *w.r.t.* $\mathcal{K}$, updating $\mathcal{K}$ subproblem is

$$\min_{\mathcal{K}} \|\mathcal{K}\|_{\mathbb{I}\omega, S_p}^{p} + \frac{\mu}{2}\|\mathcal{K} - (\mathcal{J} + \frac{\boldsymbol{\mathcal{Y}}}{\mu})\|_F^2 \tag{19}$$

According to [8], Eq. (21) can be solved into two steps as follows: (1) minimizing the core matrix, and (2) minimizing t-TSN.
(1) Updating core matrix as

$$\min_{\mathfrak{P}(\mathcal{T})} \|\mathfrak{P}(\mathcal{T})\|_* + \frac{1}{2\lambda}\|\mathcal{F} - (\mathcal{J} + \frac{\boldsymbol{\mathcal{Y}}}{\mu})\|_F^2 \tag{20}$$

where regularization parameter $\lambda = 1/(max(m,v)n)^{\frac{1}{2}}$. And the tensor $\mathcal{T}$ is obtained from t-SVD on the temporary variable $\mathcal{F}$, *i.e.*, $\mathcal{F} = \boldsymbol{\mathcal{U}} * \mathcal{T} * \boldsymbol{\mathcal{V}}$.
(2) Updating $\mathcal{K}$ as

$$\min_{\mathcal{K}} \|\mathcal{K}\|_{\mathbb{I}\omega, S_p}^{p} + \frac{\mu}{2}\|\mathcal{K} - \mathcal{L}\|_F^2 \tag{21}$$

With the learned low-rank core matrix $\mathfrak{P}(\mathcal{T})$, we can use t-product to reconstruct a tensor as $\mathcal{L} = \boldsymbol{\mathcal{U}} * \mathfrak{P}(\mathcal{T})^{-1} * \boldsymbol{\mathcal{V}}$. The learned $\mathcal{L}$ can further produce a closed-form solution via the following **Theorem 3**.

Theorem 3. *Consider $\mathcal{L} \in \mathbb{R}^{n_1\times n_2\times n_3}$, with $r = \min(n_1, n_2)$. Let $\mathcal{L}_f = \boldsymbol{\mathcal{U}}_f \boldsymbol{\mathcal{M}}_f \boldsymbol{\mathcal{V}}_f^\top$, then the optimization problem for weight tensor Schatten p-norm can be formulated as*

$$\min_{\mathcal{K}} \eta\|\mathcal{K}\|_{\omega, s_p}^{p} + \frac{1}{2}\|\mathcal{K} - \mathcal{L}\|_F^2 \tag{22}$$

and its optimal solution is

$$\mathcal{K}^* = \mathtt{ifft}(\boldsymbol{\mathcal{U}}_f * \mathcal{D}_{\eta,\omega,p}(\mathcal{L}_f) * \boldsymbol{\mathcal{V}}_f^\top) \tag{23}$$

where $\mathbf{M}_f^k = \mathtt{diag}(\boldsymbol{\delta}(\mathbf{M}_f^k))$ and $\mathcal{D}\eta, \omega, p(\mathbf{L}_f^k) = \mathtt{diag}(\boldsymbol{\theta}(\mathbf{L}_f^k))$ are the k-th frontal slices of $\boldsymbol{\mathcal{M}}_f$ and $\mathcal{D}\eta, \omega, p(\mathcal{L}_f)$ in the Fourier domain with respect to $\boldsymbol{\theta}(\mathbf{L}_f^k) = \mathtt{GST}(\boldsymbol{\delta}(\mathbf{M}_f^k), \eta * \boldsymbol{\omega}^k, p)$.

Proof: Eq. (22) becomes to

$$\min_{\mathcal{K}_f} \sum_{l=1}^{n_3} \left(\sum_{j=1}^{r} \eta * \omega_j^k * \theta_j^r(\mathbf{K}_f^k) \right) + \frac{1}{2} \|\mathbf{K}_f^k - \mathbf{L}_f^k\|_F^2 \tag{24}$$

in which $\theta_j(\mathbf{K}_f^k)$ denotes the j-th singular value of $\mathbf{K}_f^k$, and its corresponding weight is $\omega_j^k = \frac{1}{\theta_j^r(\mathbf{K}_f^k)+\varepsilon}$. Initially, each weight ω_j^k is set as $\omega_j^k = \frac{1}{\delta j(\mathbf{M}_f^k)+\varepsilon}$ since $\theta j^r(\mathbf{K}_f^k)$ is unavailable in the first iteration, and updated based on the previous iteration of $\theta j^r(\mathbf{K}_f^k)$.

Eq. (24) can be solved separately for different k as

$$\min_{\mathbf{K}_f^k} \sum_{j=1}^{r} \eta * \omega_j^k * \theta_j^r(\mathbf{K}_f^k) + \frac{1}{2} \|\mathbf{K}_f^k - \mathbf{L}_f^k\|_F^2 \tag{25}$$

The solvers are derived using the following **Theorem 4** and **Lemma 1**.

Theorem 4. *Consider the singular value decomposition (SVD) of matrix* $\mathbf{T} \in \mathbb{R}^{n_1 \times n_2}$ *as* $\mathbf{T} = \mathbf{U}_A * \mathbf{D}_A * \mathbf{V}_A^\top$, *where* $\eta > 0, r = \min(n_1, n_2)$, *and* $0 \leq \omega 1 \leq \omega_2 \leq \ldots \leq \omega_r$. *The global optimal solution for the following weighted Schatten p-norm minimization problem, adapted from [14], is as follows:*

$$\min_{\mathbf{K}} \eta \|\mathbf{K}\|_{\boldsymbol{\omega}, s_p}^p + \frac{1}{2} \|\mathbf{K} - \mathbf{T}\|_F^2 \tag{26}$$

As shown in [15], the optimal solution of Eq. (26) is given by

$$\mathbf{K}^* = \mathbf{U}_A \mathbf{D}_{\eta, \boldsymbol{\omega}, p}(\mathbf{T}) \mathbf{V}_A^\top \tag{27}$$

where $\mathbf{D}_A = \mathrm{diag}(\boldsymbol{\delta})$, $\mathbf{D}\eta, \boldsymbol{\omega}, p(\mathbf{T}) = \mathrm{diag}(\boldsymbol{\theta})$. *The vector* $\boldsymbol{\delta} = \delta_j(\mathbf{T}) j = 1^r$ *represents the singular values of* $\mathbf{T}$, *each of which can be obtained using **Lemma 1** [18].*

Lemma 1. *Consider the k-th subproblem of Eq. (26), expressed as*

$$\min_{\theta(\mathbf{K}_f^k) \geq 0} f(\theta_j(\mathbf{K}_f^k)) = \frac{1}{2} (\theta_j(\mathbf{K}_f^k) - \delta_j(\mathbf{L}_f^k))^2 + \eta \omega_j \theta_j(\mathbf{K}_f^k)^p \tag{28}$$

Within ω and p, the soft-thresholding function $\eta p^{GST}(\omega_j)$ *is defined as*

$$\eta_p^{GST}(\omega_j) = (2\omega_j(1-p))^{\frac{1}{2-p}} + \omega_j p (2\omega_j(1-p))^{\frac{p-1}{2-p}} \tag{29}$$

The minimum $Sv^{GST}(\delta_j, \omega_j)$ *of Eq. (29) is determined by*

$$T_p^{GST}(\delta_j, \omega_j) = \begin{cases} 0, \ \delta_j < \eta_p^{GST}(\omega_j) \\ \mathrm{sgn}(\delta_j) S_p^{GST}(\delta_j, \omega_j), \delta_j \geq \eta_p^{GST}(\omega_j) \end{cases} \tag{30}$$

in which $S_p^{GST}(\delta_j, \omega_j)$ *satisfies*

$$S_p^{GST}(\delta_j, \omega_j) - \delta_j + \omega_j p \left(S_p^{GST}(\delta_j, \omega_j) \right)^{p-1} = 0 \tag{31}$$

Arranging $\boldsymbol{\omega}$ $(0 \leq \omega_1 \leq \omega_2 \leq \ldots \leq \omega_r)$ *in non-ascending order and* $\boldsymbol{\delta}$ $(\delta_1 \geq \delta_2 \geq \ldots \geq \delta_r \geq 0)$ *in non-descending order aids in determining a global minimizer* $\boldsymbol{\theta}$ $(\theta_1 \geq \theta_2 \geq \ldots \geq \theta_r)$ *using von Neumann's trace inequality, where* $r = min(n_1, n_2)$.

Updating ADMM variables are written as

$$\begin{aligned} \boldsymbol{\mathcal{Y}} &= \boldsymbol{\mathcal{Y}} + \mu(\boldsymbol{\mathcal{J}} - \boldsymbol{\mathcal{K}}) \\ \mu &= min(\rho\mu, \mu_{max}) \end{aligned} \tag{32}$$

In the optimization process, we set $\mu = 1e^{-4}$ and $\mu_{max} = 10^{10}$, with a computational complexity of $O(n)$. Algorithm 1 delineates the entire optimization procedure of Eq. (4), wherein convergence is assessed by evaluating the objective value obj^t after the t-th iteration.

Algorithm 1 IWTSN-FMGC

Input: Multi-view data $\{\mathbf{X}^r\}_{r=1}^{v}$, cluster number c, latent space dimension k, and parameters α, γ.
Initialize $\mathbf{Q}^r = \mathbf{I}_k$, and the others matrices as $\mathbf{0}$.

```
repeat
    Update A, W, B, and 𝒦 via Eq (3), Eq. (15), Eq. (18), and Eq. (19), respectively;
    Update ADMM variables via Eq. (32);
until Satisfy (obj^(t) − obj^(t−1))/obj^(t) ≤ 1e − 4.
Perform SVD on Â = ∑_{r=1}^{v} A^r / v.
```

Output: Clustering metrics.

2.3 Parameter and Convergence

More parameters and convergence analyses are provide in Fig. 1 and Fig. 2.

REFERENCES

[1] Si-Guo Fang, Dong Huang, Xiao-Sha Cai, Chang-Dong Wang, Chaobo He, and Yong Tang. 2023. Efficient multi-view clustering via unified and discrete bipartite graph learning. *IEEE Transactions on Neural Networks and Learning Systems* (2023), 1–12.

[2] Dong Huang, Chang-Dong Wang, and Jian-Huang Lai. 2023. Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity. *IEEE Transactions on Knowledge and Data Engineering* (2023).

[3] Zhao Kang, Xinjia Zhao, Chong Peng, Hongyuan Zhu, Joey Tianyi Zhou, Xi Peng, Wenyu Chen, and Zenglin Xu. 2020. Partition level multiview subspace clustering. *Neural Networks* 122 (2020), 279–288.

[4] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. 2020. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 4412–4419.

[5] Liang Li, Junpu Zhang, Siwei Wang, Xinwang Liu, Kenli Li, and Keqin Li. 2023. Multi-view bipartite graph clustering with coupled noisy feature filter. *IEEE Transactions on Knowledge and Data Engineering* (2023), 1–12.

[6] Ruihuang Li, Changqing Zhang, Qinghua Hu, Pengfei Zhu, and Zheng Wang. 2019. Flexible Multi-View Representation Learning for Subspace Clustering.. In *IJCAI*. 2916–2922.

[7] Xuelong Li, Han Zhang, Rong Wang, and Feiping Nie. 2022. Multiview clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 1 (2022), 330–344.

[8] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. 2019. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE transactions on pattern analysis and machine intelligence* 42, 4 (2019), 925–938.

[9] Xun Lu and Songhe Feng. 2023. Structure diversity-induced anchor graph fusion for multi-view clustering. *ACM Transactions on Knowledge Discovery from Data* 17, 2 (2023), 1–18.

[10] Feiping Nie, Xiaoqian Wang, Michael I Jordan, and Heng Huang. 2016. The constrained laplacian rank algorithm for graph-based clustering. In *Thirtieth AAAI Conference on Artificial Intelligence*. 1969–1976.

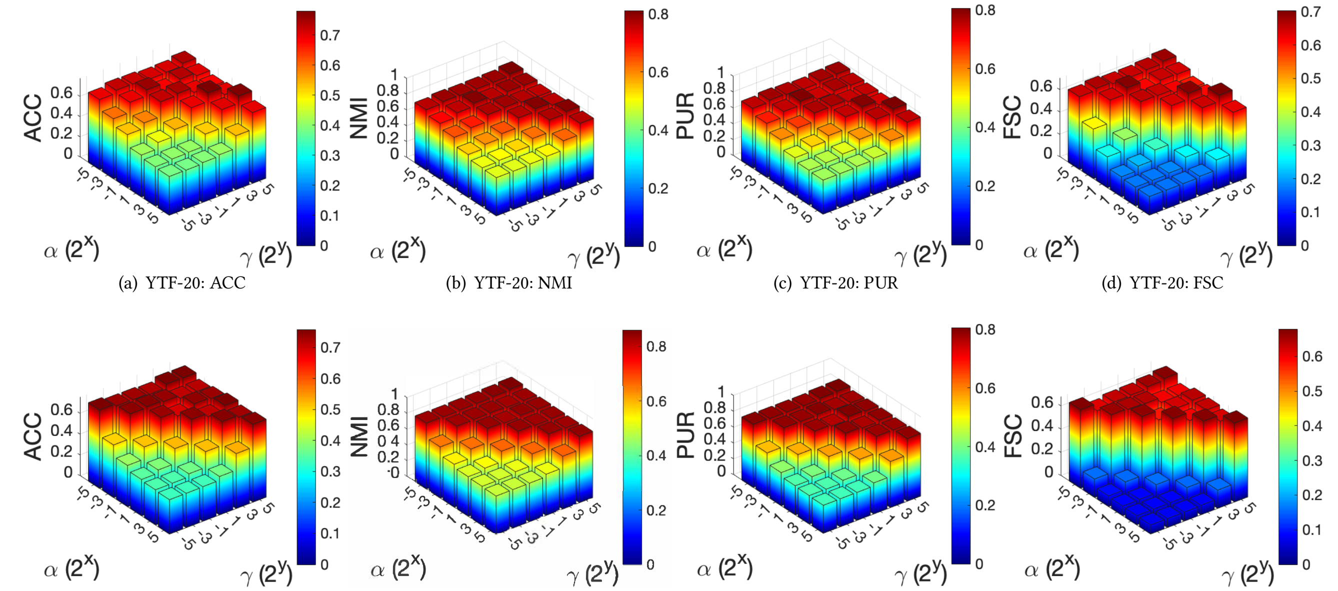
[11] Mengjing Sun, Pei Zhang, Siwei Wang, Sihang Zhou, Wenxuan Tu, Xinwang Liu, En Zhu, and Changjian Wang. 2021. Scalable multi-view subspace clustering with unified anchors. In *Proceedings of the 29th ACM International Conference on Multimedia*. 3528–3536.

[12] Siwei Wang, Xinwang Liu, Xinzhong Zhu, Pei Zhang, Yi Zhang, Feng Gao, and En Zhu. 2021. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing* 31 (2021), 556–568.

[13] Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. 2023. Tensorized Bipartite Graph Learning for Multi-View Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 4 (2023), 5187–5202.

[14] Yuan Xie, Shuhang Gu, Yan Liu, Wangmeng Zuo, Wensheng Zhang, and Lei Zhang. 2016. Weighted Schatten p-norm minimization for image denoising and background subtraction. *IEEE transactions on image processing* 25, 10 (2016), 4842–4857.

[15] Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu. 2018. On unifying multi-view self-representations for clustering by tensor



(a) YTF-20: ACC (b) YTF-20: NMI (c) YTF-20: PUR (d) YTF-20: FSC

(e) YTF-50: ACC (f) YTF-50: NMI (g) YTF-20: PUR (h) YTF-50: FSC

Figure 1: The parameter settings (α and γ) on the large-scale datasets, *i.e.*, YTF50 and YTF-20.

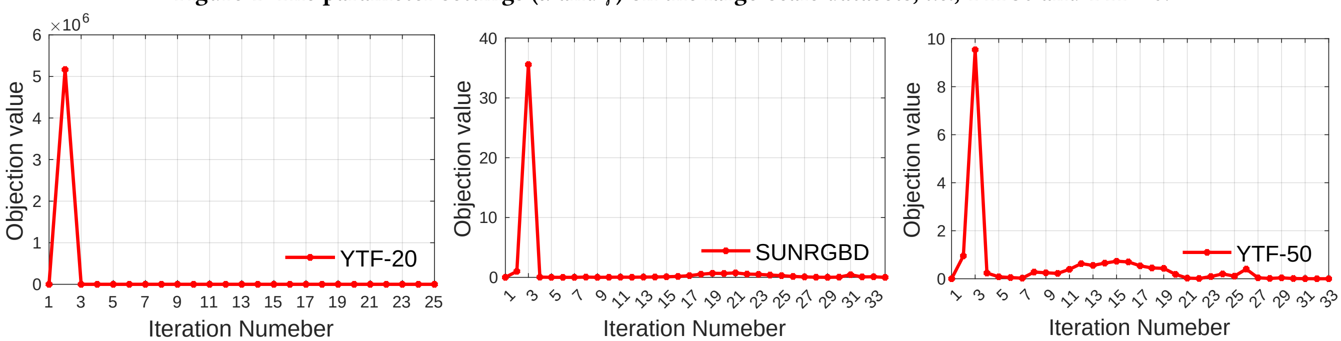


Figure 2: The losses on the YTF-50, YTF-20, and SUNRGBD datasets.

multi-rank minimization. *International Journal of Computer Vision* 126, 11 (2018), 1157–1179.

[16] Ben Yang, Xuetao Zhang, Feiping Nie, Fei Wang, Weizhong Yu, and Rong Wang. 2020. Fast multi-view clustering via nonnegative and orthogonal factorization. *IEEE Transactions on Image Processing* 30 (2020), 2575–2586.

[17] Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. 2018. Binary multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence* 41, 7 (2018), 1774–1782.

[18] Wangmeng Zuo, Deyu Meng, Lei Zhang, Xiangchu Feng, and David Zhang. 2013. A generalized iterated shrinkage algorithm for non-convex sparse coding. In *Proceedings of the IEEE international conference on computer vision*. 217–224.