Supplementary Materials: Anchor Structure Regularization Induced Multi-view Subspace Clustering via Enhanced Tensor Rank Minimization

Jintian Ji[1,2], Songhe Feng[1,2,*]
[1]Key Laboratory of Big Data & Artificial Intelligence in Transportation,
Ministry of Education, Beijing Jiaotong University, Beijing, 100044, China
[2]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044, China
{22120385,shfeng}@bjtu.edu.cn

In this supplementary, we provide the proofs for the theorems we proposed in the main manuscript. In addition, we also provide more experimental results. This document is organized as follows. Section A presents the detailed proofs of Theorem 1. In Section B, it elaborates on proof of Theorem 2. At last, Section C completes the experimental results.

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

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

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:

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

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

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

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

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 $1 \leq k \leq n_3$.

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 **Lemma 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:

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

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

*Corresponding Author

B. Proofs of Theorem 2

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

1). $\{\mathcal{P}_k\}$ *is bounded;*

2). Any accumulation point of $\{\mathcal{P}_k\}$ *is a KKT point of the 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* $x \neq 0$*, and* $\|y\|^{dual}$ *if* $x = 0$*.*

Lemma 3. *[4] Suppose that* $F : \mathbb{R}^{m\times n} \to \mathbb{R}$ *is defined as* $F(\mathbf{X}) = f \circ \boldsymbol{\sigma}(\mathbf{X}) = f(\sigma_1(\mathbf{X}), \dots, \sigma_r(\mathbf{X}))$*, 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 $\partial f(\boldsymbol{\sigma}(\mathbf{X})) = (\frac{\partial f(\sigma_1(\mathbf{x}))}{\partial \mathbf{X}}), \dots, \frac{\partial f(\sigma_r(\mathbf{x}))}{\partial \mathbf{X}})$*.*

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

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

Thus, we have

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

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

$$\partial \left\|\mathcal{G}_{k+1}\right\|_{\mathrm{ETR}} = \mathcal{W}_{t+1}. \tag{13}$$

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:

$$\begin{aligned} &\left\|\partial \left\|\mathcal{G}_{k+1}\right\|_{\mathrm{ETR}}\right\|_F^2 \\ &= \left\|\frac{1}{n_3}\mathcal{U} * ifft(\partial(\mathcal{S}_f), [], 3) * \mathcal{V}^T\right\|_F^2 \\ &= \frac{1}{n_3^3}\left\|\partial f(\mathcal{S}_f)\right\|_F^2 \\ &\leq \frac{1}{n_3^3}\sum_{i=1}^{n_3}\sum_{j=1}^{min(n_1,n_2)}[\partial f(\mathcal{S}_f^v(j,j))]^2 \\ &\leq \frac{e^{2\delta^2}min(n_1,n_2)}{\delta^2 n_3^2} \end{aligned} \tag{14}$$

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 $\partial\left\|\mathcal{G}_{k+1}\right\|_{\mathrm{ETR}}$ is bounded, meanwhile the sequence $\{\mathcal{W}_{k+1}\}$ is also bounded.

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

$$\begin{aligned} &\mathcal{L}_{\mu_k,\rho_k}(\mathbf{Z}_{k+1}^v, \mathbf{E}_{k+1}^v, \mathbf{A}_{k+1}^v, \mathcal{G}_{k+1}, \mathbf{Y}_k^v, \mathcal{W}_k) \\ &\leq \mathcal{L}_{\mu_k,\rho_k}(\mathbf{Z}_k^v, \mathbf{E}_k^v, \mathbf{A}_k^v, \mathbf{A}_k^v, \mathcal{G}_k, \mathbf{Y}_k^v, \mathcal{W}_k) \\ &= \mathcal{L}_{\mu_{k-1},\rho_{k-1}}(\mathbf{Z}_k^v, \mathbf{E}_k^v, \mathbf{A}_k^v, \mathcal{G}_k, \mathbf{Y}_{k-1}^v, \mathcal{W}_{k-1}) \\ &+ \frac{\rho_k + \rho_{k-1}}{2\rho_{k-1}^2}\left\|\mathcal{W}_k - \mathcal{W}_{k-1}\right\|_F^2 \\ &+ \frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2}\sum_{v=1}^m\left\|\mathbf{Y}_k^v - \mathbf{Y}_{k-1}^v\right\|_F^2, \end{aligned} \tag{15}$$

Thus, summing two sides of (15) form $k = 1$ to n,

$$\begin{aligned} &\mathcal{L}_{\mu_k,\rho_k}(\mathbf{Z}_{k+1}^v, \mathbf{E}_{k+1}^v, \mathbf{A}_{k+1}^v, \mathcal{G}_{k+1}, \mathbf{Y}_k^v, \mathcal{W}_k) \\ &\leq \mathcal{L}_{\mu_0,\rho_0}(\mathbf{Z}_1^v, \mathbf{E}_1^v, \mathbf{A}_1^v, \mathcal{G}_1, \mathbf{Y}_0^v, \mathcal{W}_0) \\ &+ \sum_{k=1}^n\frac{\rho_k + \rho_{k-1}}{2\rho_{k-1}^2}\left\|\mathcal{W}_k - \mathcal{W}_{k-1}\right\|_F^2 \\ &+ \sum_{k=1}^n\left(\frac{\mu_k + \mu_{k-1}}{2\mu_{k-1}^2}\sum_{v=1}^m\left\|\mathbf{Y}_k^v - \mathbf{Y}_{k-1}^v\right\|_F^2\right) \end{aligned} \tag{16}$$

Observe that

$$\sum_{k=1}^n\frac{\mu_k + \mu_{k+1}}{2\mu_{k-1}^2} < \infty, \quad \sum_{k=1}^n\frac{\rho_k + \rho_{k+1}}{2\rho_{k-1}^2} < \infty \tag{17}$$

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

Notice

$$
\begin{aligned}
&\mathcal{L}_{\mu_k,\rho_k}(\mathbf{Z}_{k+1}^v, \mathbf{E}_{k+1}^v, \mathbf{A}_{k+1}^v, \mathcal{G}_{k+1}, \mathbf{Y}_k^v, \mathcal{W}_k) \\
&= \|\mathcal{G}_{k+1}\|_{\mathrm{ETR}} + \alpha \|\mathbf{E}_{k+1}\|_{2,1} \\
&+ \gamma \sum_{v=1}^{m} Tr(\mathbf{Z}^v \mathbf{L}^v (\mathbf{Z}^v)^T) \\
&+ \sum_{v=1}^{m} (\langle \mathbf{Y}_k^v, \mathbf{X}^v - \mathbf{Z}_{k+1}^v \mathbf{A}_{k+1}^v - \mathbf{E}_{k+1}^v \rangle \\
&+ \frac{\mu_k}{2} \left\|\mathbf{X}^v - \mathbf{Z}_{k+1}^v \mathbf{A}_{k+1}^v - \mathbf{E}_{k+1}^v\right\|_F^2) \\
&+ \langle \mathcal{W}_k, \mathcal{Z}_{k+1} - \mathcal{G}_{k+1} \rangle + \frac{\rho_k}{2} \|\mathcal{Z}_{k+1} - \mathcal{G}_{k+1}\|_F^2,
\end{aligned} \tag{18}
$$

and each term of (18) is nonnegative, due to the boundedness of $\mathcal{L}_{\mu_k}(\mathbf{Z}_{k+1}^v, \mathbf{E}_{k+1}^v, \mathbf{A}_{k+1}^v, \mathcal{G}_{k+1}, \mathbf{Y}_k^v, \mathcal{W}_k)$, we can deduce each term of (18) is bounded. So the boundedness of $\|\mathcal{G}_{k+1}\|_{\mathrm{ETR}}$ implies that all singular values of $\mathcal{G}_{k+1}$ are bounded. Furthermore, based on the following equation

$$
\|\mathcal{G}_{k+1}\|_F^2 = \frac{1}{n_3} \|\mathcal{G}_{f,k+1}\|_F^2 = \frac{1}{n_3} \sum_{i=1}^{n_3} \sum_{j=1}^{min(n_1,n_2)} (\mathcal{S}_f^i(j,j))^2, \tag{19}
$$

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

Therefore, we can conclude that the sequence $\{\mathcal{P}_k = (\mathbf{Z}_k^v, \mathbf{E}_k^v, \mathbf{A}_k^v, \mathbf{Y}_k^v, \mathcal{W}_k, \mathcal{G}_k)\}_{k=1}^{\infty}$ generated by the Algorithm 1.

2). Proof of $2nd$ part: 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

$$
\lim_{k\to\infty} (\mathbf{Z}_k^v, \mathbf{E}_k^v, \mathbf{A}_k^v, \mathbf{Y}_k^v, \mathcal{W}_k, \mathcal{G}_k) = (\mathbf{Z}_*^v, \mathbf{E}_*^v, \mathbf{A}_*^v, \mathbf{Y}_*^v, \mathcal{W}_*, \mathcal{G}_*). \tag{20}
$$

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

$$
\begin{aligned}
&\mathbf{X}^v - \mathbf{Z}_{k+1}^v \mathbf{A}_{K+1}^v - \mathbf{E}_{k+1}^v = (\mathbf{Y}_{k+1}^v - \mathbf{Y}_k^v)/\mu_t, \\
&\mathcal{Z}_{k+1} - \mathcal{G}_{k+1} = (\mathcal{W}_{k+1} - \mathcal{W}_k)/\rho_t.
\end{aligned} \tag{21}
$$

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

$$
\begin{aligned}
&\lim_{k\to\infty} \mathbf{X}^v - \mathbf{Z}_{k+1}^v \mathbf{A}_{K+1}^v - \mathbf{E}_{k+1}^v = \lim_{k\to\infty} (\mathbf{Y}_{k+1}^v - \mathbf{Y}_k^v)/\mu_t = 0, \\
&\lim_{k\to\infty} \mathcal{Z}_{k+1} - \mathcal{G}_{k+1} = \lim_{k\to\infty} (\mathcal{W}_{k+1} - \mathcal{W}_k)/\rho_t = 0,
\end{aligned} \tag{22}
$$

then, we can obtain

$$
\mathbf{X}^v - \mathbf{Z}_*^v \mathbf{A}_*^v - \mathbf{E}_*^v = 0, \quad \mathcal{Z}_* - \mathcal{G}_* = 0. \tag{23}
$$

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

$$
\begin{aligned}
&0 = \alpha\partial \left\|\mathbf{E}_{k+1}^v\right\|_{2,1} - \mathbf{Y}_{k+1}^v \Rightarrow \mathbf{Y}_*^v = \alpha\partial \|\mathbf{E}_*^v\|_{2,1} \\
&0 = \partial \|\mathcal{G}_{k+1}\|_{\mathrm{ETR}} - \mathcal{W}_{k+1} \Rightarrow \mathcal{W}_* = \partial \|\mathcal{G}_*\|_{\mathrm{ETR}}
\end{aligned} \tag{24}
$$

Thus, the accumulation point $\mathcal{P}_*$ of sequence $\{\mathcal{P}_k\}_{k=1}^{\infty}$ generated by the algorithm of solving ASR-ETR satisfied the KKT condition. □

C. More Experiment Results

In this part, we complete the results of all seven data sets.

Influence of Enhanced Tensor Rank: We complete the effect of the Enhanced Tensor Rank (ETR) of all seven data sets with Fig. 1, and we can observe that the parameter has a significant effect on the clustering results. The best clustering results of NGs, BBCSport, Caltech101-all, Aloi-100, CIFAR10, and Noisy MNIST are obtained when $\delta = 0.1$, and CCV peaks at $\delta = 10^{-4}$.

Anchor Analysis: We complete the anchor analysis of all seven data sets with Fig. 2, it is clear to observe that the clustering results are stable under different anchors on all seven data sets, which demonstrates that our anchor-representation strategy is robust to the number of anchors, and it is not necessary to use numerous anchors for clustering.

Convergence Analysis: We complete the convergence analysis of all seven data sets with Fig. 3. As shown in Fig. 3, the values of RE and ME rapidly tend to 0 within 10 steps and remain stable, which indicates the excellent convergence property of our ASR-ETR.

References

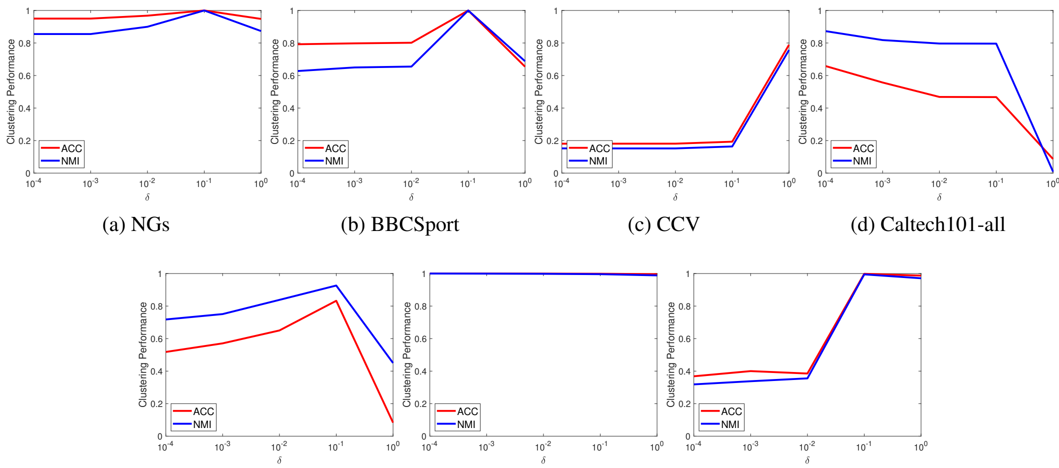
[1] Robert G Bartle and Donald R Sherbert. *Introduction to real analysis*, volume 2. Wiley New York, 2000.

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

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

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

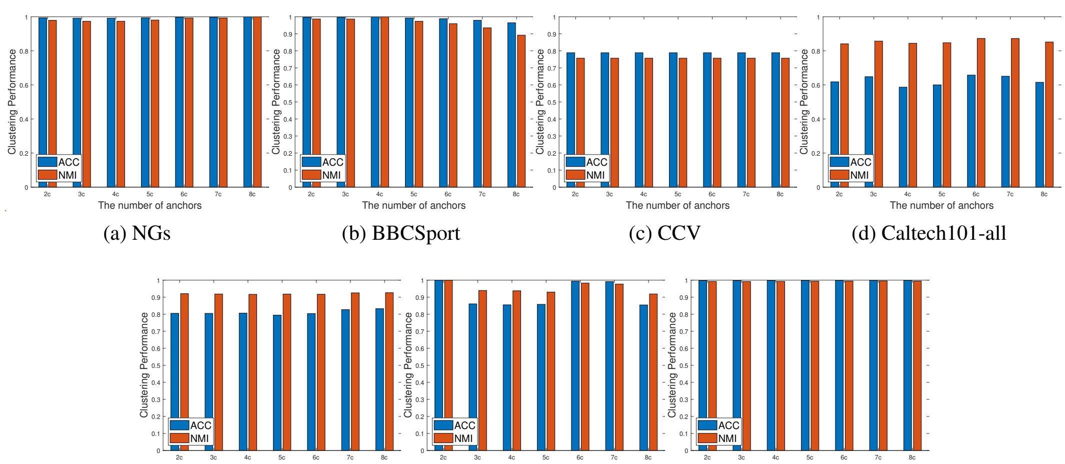
[5] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.



(a) NGs (b) BBCSport (c) CCV (d) Caltech101-all

(e) Aloi-100 (f) CIFAR10 (g) Noisy MNIST

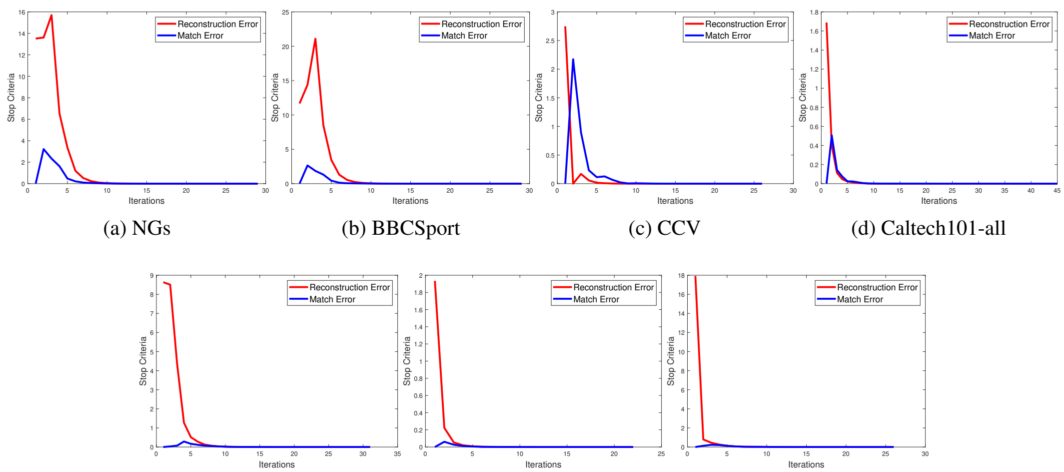
Figure 1: The performance (i.e., ACC and NMI) of ASR-ETR with varying parameter δ on seven data sets.



(a) NGs (b) BBCSport (c) CCV (d) Caltech101-all

(e) Aloi-100 (f) CIFAR10 (g) Noisy MNIST

Figure 2: Anchor Analysis: The performance (i.e., ACC and NMI) of ASR-ETR on seven data sets by varying the number of anchors.



(a) NGs (b) BBCSport (c) CCV (d) Caltech101-all

(e) Aloi-100 (f) CIFAR10 (g) Noisy MNIST

Figure 3: Convergence Analysis: The stop criteria (i.e., RE and ME) variation curves on seven data sets.