

# Appendix of Paper 574

Anonymous

## A THE DETAILED DERIVATION PROCESS OF ALTERNATE OPTIMIZATION

In this part, we give the detailed derivation process of the three-step alternate optimization. The objective equation is

$$\begin{aligned} \min_{Z_t, B_t, H_t} & \|\tilde{X}_t - Z_t B_t\|_F^2 + \|Z_t - Z_{t-1} H_t\|_F^2 + \|B_t - H_t^T B_{t-1}\|_F^2, \\ \text{s.t. } & Z_t^T Z_t = I_m, B_t^T B_t = I_d, H_t^T H_t = I_m. \end{aligned} \quad (1)$$

**Updating  $Z_t$ .** Fixing  $B_t$  and  $H_t$ , the optimization in Eq.(1) w.r.t  $Z_t$  can be reduced to

$$\begin{aligned} & \min_{Z_t} \|\tilde{X}_t - Z_t B_t\|_F^2 + \|Z_t - Z_{t-1} H_t\|_F^2 + \|B_t - H_t^T B_{t-1}\|_F^2 \\ \Leftrightarrow & \min_{Z_t} \|\tilde{X}_t - Z_t B_t\|_F^2 + \|Z_t - Z_{t-1} H_t\|_F^2 \\ \Leftrightarrow & \min_{Z_t} \text{Tr} \left( \tilde{X}_t^T \tilde{X}_t + (Z_t B_t)^T Z_t B_t - 2 \tilde{X}_t^T Z_t B_t \right) + \\ & \text{Tr} (Z_t^T Z_t + (Z_{t-1} H_t)^T Z_{t-1} H_t - 2 Z_t^T Z_{t-1} H_t) \\ \Leftrightarrow & \min_{Z_t} \text{Tr} \left( -2 \tilde{X}_t^T Z_t B_t \right) + \text{Tr} (-2 Z_t^T Z_{t-1} H_t) \\ \Leftrightarrow & \max_{Z_t} \text{Tr} \left( B_t \tilde{X}_t^T Z_t + H_t^T Z_{t-1}^T Z_t \right) \end{aligned} \quad (2)$$

So the optimization Eq. (2) could be rewritten as follows,

$$\max_{Z_t} \text{Tr} (Z_t^T A), \text{ s.t. } Z_t^T Z_t = I_m, \quad (3)$$

where  $A = \tilde{X}_t B_t^T + Z_{t-1} H_t$ . The problem can be solved with a closed-form solution via [1].

By supposing  $A$  has the rank- $k$  and taking the the singular value decomposition (SVD)  $A = S \Sigma V^T$ , the Eq. (3) could be rewritten as,

$$\text{Tr} (Z_t^T S \Sigma V^T) = \text{Tr} (V^T Z_t^T S \Sigma). \quad (4)$$

Supposing that  $P = V^T Z_t^T S$ , we have  $PP^T = V^T Z_t^T S S^T Z_t V$ . Therefore we can take  $\text{Tr} (V^T Z_t^T S \Sigma) = \text{Tr} (P \Sigma) \leq \sum_{i=1}^k \sigma_i$ . Hence in order to maximize the value of Eq. (3), the solution should be given as follows,

$$Z_t = S V^T, \quad (5)$$

The other two optimization sub-questions of  $B_t$  and  $H_t$  are similar, and omitted.

## B DATASETS

We conduct the experiments on seven widely used datasets. The details of the datasets we use in experiments are listed as follows:

- (1) The Dermatology dataset consists of skin lesions along with associated metadata such as patient age and gender. The feature number of each view are 12, and 22.
- (2) Flower17 consists of images of flowers belonging to 17 different categories. The feature number of each view are 537, 512, 537, 537, 123, 537, 537.

- (3) The Cora contains academic papers from various research areas, along with citation links. Each paper is represented by a bag-of-words feature vector derived from its abstract. The feature number of each view are 270, 143, 270, and 270.
- (4) AWA10 consists of animals along with attribute annotations. Each image is associated with a set of attributes, such as fur color, size, and habitat. The feature number of each view are 268, 200, 252, 200, 200 and 200.
- (5) ALOI is a dataset containing images of objects captured from various viewpoints. The feature number of each view are 77, 13, 64, and 12.
- (6-7) YouTubeFace10 and YouTubeFace100 consist of facial images of 10 and 100 persons extracted from YouTube videos. Each image represents a different individual, and the dataset may include multiple views or facial expressions of each person. The feature number of each view in both are 944, 576, 512, and 640.

## C COMPLEXITY COMPARISON

We count the major memory cost and time use of the compared algorithms in the sequential-view scenarios, concerning the sample number  $n$  and the view number  $v$  in the following Table 1. We assume that the rank of the kernel  $r = n - C$ , where  $C$  is a constant.

It is easy to observe that both the space complexity and time complexity of our proposed method keep the minimum among the state-of-the-art algorithms. Although the computational space complexity of many methods is in the same power level as our method, our incremental algorithm greatly reduces the number of calculation instructions for different view data alignment.

## D VIEW-FUSION PERFORMANCE

We compare the view-fusion performance of our method with the existing three incremental multi-view clustering methods. The results of the three metrics are given in the Tab. 2.

## E PREPROCESSING EFFECT

In the LAIMVC, we perform PCA to unify the dimensions of views in datasets. Each of these datasets has low-quality views, probably because there is redundant information that misleads the clustering results. We choose the worst performance view of every dataset as an example and compare the clustering performance of the PCA-processed view with the original view, as shown in the following Table 3.

The Fig. 1 below shows them intuitively.

In summary, the PCA in our proposed method can improve the clustering performance on low-quality data views in addition to making the dimensionality of different views under the dataset uniform. For instance, on all of our datasets, the clustering effect of the views processed by PCA has been improved, especially for ALOI, whose original dataset performs poorly but is qualitatively improved after PCA.

**Table 1: The space and time complexity of the algorithms in the sequential-view scenarios.**

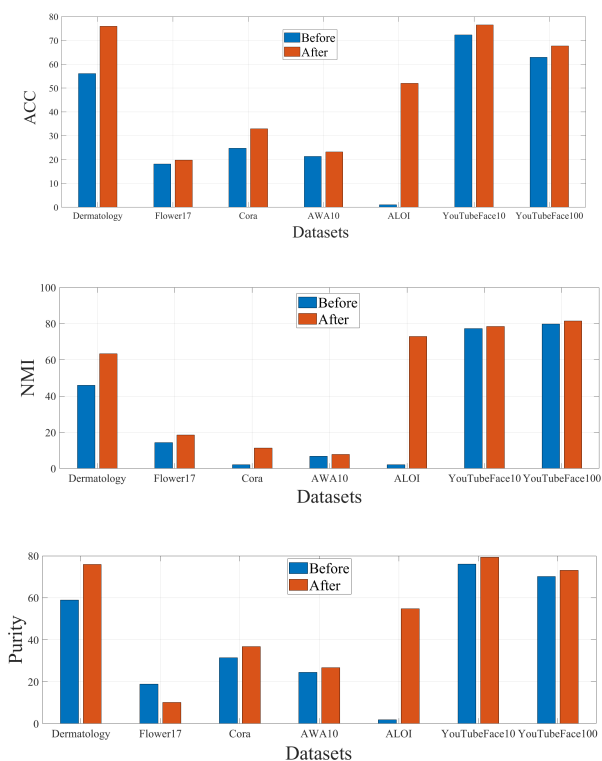
Methods	Time Complexity	Space Complexity
OPLF-MVC	$O(v^3n)$	$O(vn)$
OPMC	$O(v^2n)$	$O(vn)$
LMVSC	$O(v^4 + v^2n)$	$O(vn)$
FPMVS-CAG	$O(v^2 + vn)$	$O(vn)$
SMVSC	$O(v^2n)$	$O(v + n)$
IMSC	$O(vn^2)$	$O(n^2)$
SCGL	$O(vn^3)$	$O(vn)$
CMVC	$O(vn)$	$O(n)$
Proposed	$O(vn)$	$O(n)$

**Table 2: Empirical evaluation and comparison of LAIMVC with three incremental methods on each view of chosen datasets in terms of clustering accuracy (ACC), normalized mutual information (NMI), and Purity, individually.**

View	IMSC			SCGL			CMVC			Proposed		
	ACC(%)	NMI(%)	Purity(%)	ACC(%)	NMI(%)	Purity(%)	ACC(%)	NMI(%)	Purity(%)	ACC(%)	NMI(%)	Purity(%)
Dermatology												
1	57.82	44.23	59.78	67.88	53.48	69.27	51.96	43.36	59.78	57.26	47.38	63.13
2	45.81	62.70	67.32	89.21	80.97	99.31	70.39	61.00	76.54	96.65	92.15	96.65
AWA10												
1	21.17	6.89	24.53	21.52	6.86	24.65	21.65	6.79	25.27	21.36	6.71	25.13
2	22.55	7.34	24.72	21.69	7.38	24.30	19.32	5.42	13.35	23.65	8.01	25.83
3	19.30	4.93	22.74	21.81	7.58	24.82	19.49	5.03	23.74	21.67	6.84	25.18
4	22.08	5.73	23.58	22.57	8.23	25.04	19.83	6.10	23.77	23.75	8.54	26.78
5	22.7	5.99	24.53	21.16	8.13	24.20	21.14	6.80	24.22	24.63	8.28	27.11
6	22.14	5.72	24.01	24.70	9.40	26.63	20.88	6.13	23.82	26.47	10.55	30.75
YouTubeFace10												
1	77.20	80.96	81.57	OM	OM	OM	79.33	82.34	84.42	78.57	81.13	81.44
2	65.21	66.51	68.08	OM	OM	OM	79.19	82.65	82.85	83.35	82.69	85.81
3	64.41	63.62	65.30	OM	OM	OM	85.5	83.41	85.5	85.00	82.84	85.03
4	61.18	65.16	64.65	OM	OM	OM	82.29	83.88	86.37	93.60	88.26	93.60

**Table 3: Comparison of the view performance before PCA and after performing PCA in terms of clustering accuracy (ACC), normalized mutual information (NMI), and Purity.**

Datasets	Dermatology	Flower17	Cora	AWA10	ALOI	YouTubeFace10	YouTubeFace100
ACC(%)							
Before	56.15	18.16	24.78	21.36	1.01	72.33	63.00
After	75.98	19.85	32.98	23.24	52.06	76.55	67.75
NMI(%)							
Before	45.92	14.35	2.07	6.82	2.10	77.32	79.82
After	63.33	18.48	11.34	7.86	72.98	78.51	81.52
Purity(%)							
Before	58.94	18.90	31.46	24.51	1.92	76.13	70.23
After	75.98	10.07	36.78	26.63	54.76	79.39	73.06



**Figure 1: The Comparison between the view before PCA and that after performing PCA in terms of three metrics.**

## F LIMITATION

In our study, the correlation between the selection of the number of anchors and the view data was not explored. Although it can be proved that the clustering effect is less sensitive to the number of anchors, it is still of great significance to effectively select the number of anchors.

## REFERENCES

- [1] Siwei Wang, Xinwang Liu, En Zhu, Chang Tang, Jiyuan Liu, Jingtao Hu, Jingyuan Xia, and Jianping Yin. 2019. Multi-view Clustering via Late Fusion Alignment Maximization.. In *IJCAI*. 3778–3784.