Appendix of *Hierarchical Multiple Kernel Clustering*

Jiyuan Liu, [1] **Xinwang Liu,** [1*] **Yuexiang Yang,** [1] **Siwei Wang,** [1] **Sihang Zhou** [2]

[1] College of Computer, National University of Defense Technology, Changsha, Hunan, China.
[2] College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan, China.
{liujiyuan13, xinwangliu, yyx, wangsiwei13}@nudt.edu.cn, sihangjoe@gmail.com

Additional details

Due to the space limit, we provide more details in the appendix.

In specific, we compare with the MSC methods in recent literature on *Purity* in Table 1. It can be observed that both *HMKC-1* and *HMKC-2* outperform the comparative ones consistently. Especially, the proposed algorithm exhibits excellent performances on CCV, Flower17 and Flower102, where around 5% - 10% increases are obtained.

Table 1: The performance comparison (Purity) of MSC algorithms in recent literature.

Algorithm	AR10P	BBCSport	CCV	Flower17
A-MKKM	38.46	77.21	23.98	51.99
SB-KKM	49.23	79.60	23.48	44.63
CSRC	37.69	76.65	26.44	53.53
MKKM	39.23	77.57	22.25	46.84
RMSC	36.15	86.03	20.38	55.88
RMKC	44.62	77.39	23.98	53.02
RMKKM	33.85	61.76	20.67	54.34
MKKM-MR	43.08	77.57	25.69	61.03
ONKC	49.23	77.57	26.15	60.15
LFAM	46.92	77.76	28.67	60.66
SPC	40.77	83.09	-	59.78
MVCC	43.85	74.63	24.39	55.44
SPMKC	*56.92*	41.73	18.10	38.38
HMKC-1	*56.92*	*89.52*	*39.33*	*68.16*
HMKC-2	**60.00**	**90.99**	**40.40**	**71.18**
Algorithm	Flower102	Heart	Ionosphere	Plant
A-MKKM	32.28	82.22	64.10	61.70
SB-KKM	38.78	76.30	70.09	56.38
CSRC	42.92	80.37	75.78	59.79
MKKM	27.61	55.56	64.10	56.38
RMSC	40.24	83.33	84.62	60.85
RMKC	39.33	82.22	66.10	61.70
RMKKM	34.33	76.30	65.81	55.32
MKKM-MR	46.48	83.33	64.10	62.87
ONKC	46.65	83.70	64.10	64.89
LFAM	49.87	82.22	68.09	62.77
SPC	-	75.93	71.23	64.47
MVCC	45.30	82.96	64.10	56.17
SPMKC	-	57.04	72.93	53.94
HMKC-1	*53.99*	**86.67**	*86.32*	*65.32*
HMKC-2	**56.64**	*86.30*	**86.89**	**67.02**

Meanwhile, Fig. 1 presents the epoch numbers required by *1-layer* and *2-layer* models over all parameters on *CCV*. It can be observed that the proposed algorithm quickly converge with fewer than 10 epochs.

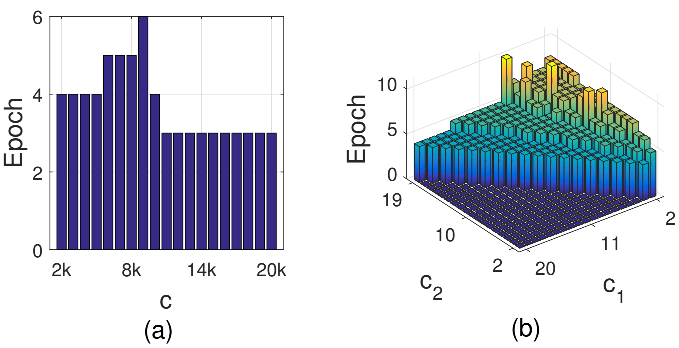

Figure 1: (a). Epoch numbers required by the *1-layer* model to converge at $2k \leq c \leq 20k$ on *CCV*; (b). Epoch numbers required by the *1-layer* model to converge at $2k \leq c_2 \leq c_1 \leq 20k$ on *CCV*.

In addition, we provide formal convergence proof of the proposed alternate optimization algorithm in the following. For ease of expression, we formulate the objective in Eq. (9) of manuscript as

$$\max_{\{\mathbf{H}_p^{(t)}\}_{t=1}^s, \mathbf{H}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^s, \boldsymbol{\beta}} \mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^s, \mathbf{H}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^s, \boldsymbol{\beta}) \tag{1}$$

Let superscript r represent the optimization at round r.

i) **Optimizing $\{\mathbf{H}_p^{(t)}\}_{t=1}^s$ with fixing the others**. Given $\mathbf{H}^{(r)}$, $\{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r)}$ and $\boldsymbol{\beta}^{(r)}$, the resultant optimization problem can be efficiently solved in closed-form by singular value decomposition. Assuming the obtained solution as $\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}$, we can get

$$\begin{aligned} &\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r)}, \mathbf{H}^{(r)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r)}, \boldsymbol{\beta}^{(r)}) \\ \leq &\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}, \mathbf{H}^{(r)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r)}, \boldsymbol{\beta}^{(r)}). \end{aligned} \tag{2}$$

*Corresponding author

ii) **Optimizing H with fixing the others**.
Given $\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}$, $\{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r)}$ and $\boldsymbol{\beta}^{(r)}$, the resultant optimization problem can be efficiently solved in closed-form by singular value decomposition. Assuming the obtained solution as $\mathbf{H}^{(r+1)}$, we can get

$$\begin{aligned}&\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}, \mathbf{H}^{(r)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r)}, \boldsymbol{\beta}^{(r)})\\ \leq&\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}, \mathbf{H}^{(r+1)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r)}, \boldsymbol{\beta}^{(r)}).\end{aligned} \tag{3}$$

iii) **Optimizing $\{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s}$ with fixing the others**.
Given $\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}$, $\mathbf{H}^{(r+1)}$ and $\boldsymbol{\beta}^{(r)}$, we can get the closed-form solution $\{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r+1)}$, resulting in

$$\begin{aligned}&\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}, \mathbf{H}^{(r+1)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r)}, \boldsymbol{\beta}^{(r)})\\ \leq&\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}, \mathbf{H}^{(r+1)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r+1)}, \boldsymbol{\beta}^{(r)}).\end{aligned} \tag{4}$$

iv) **Optimizing $\boldsymbol{\beta}$ with fixing the others**.
Given $\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}$, $\mathbf{H}^{(r+1)}$ and $\{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r+1)}$, we can get the closed-form solution $\boldsymbol{\beta}^{(r+1)}$, resulting in

$$\begin{aligned}&\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}, \mathbf{H}^{(r+1)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r+1)}, \boldsymbol{\beta}^{(r)})\\ \leq&\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}, \mathbf{H}^{(r+1)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r+1)}, \boldsymbol{\beta}^{(r+1)}).\end{aligned} \tag{5}$$

Together with Eq. (2), (3), (4) and (5), we can get

$$\begin{aligned}&\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r)}, \mathbf{H}^{(r)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r)}, \boldsymbol{\beta}^{(r)})\\ \leq&\mathcal{J}(\{\mathbf{H}_p^{(t)}\}_{t=1}^{s\ (r+1)}, \mathbf{H}^{(r+1)}, \{\boldsymbol{\gamma}^{(t)}\}_{t=1}^{s\ (r+1)}, \boldsymbol{\beta}^{(r+1)}),\end{aligned} \tag{6}$$

which indicates that the objective monotonically increases at each round. Also, it is obvious that the whole optimization objective is upper-bounded. As a result, the alternative algorithm is theoretically guaranteed to converge to a local maximum.