On the Convergence of Some Possibilistic Clustering Algorithms

Jian Zhou

School of Management, Shanghai University, Shanghai 200444, China

Longbing Cao

Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

Nan Yang‡

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

Abstract

In this paper, an analysis of the convergence performance is conducted for a class of possibilistic clustering algorithms utilizing the Zangwill convergence theorem. It is shown that under certain conditions the iterative sequence generated by a possibilistic clustering algorithm converges, at least along a subsequence, to either a local minimizer or a saddle point of the objective function of the algorithm. The convergence performance of more general possibilistic clustering algorithms is also discussed.

Keywords: Fuzzy clustering, possibilistic clustering, convergence

1 Introduction

Possibilistic clustering, initiated by Krishnapuram and Keller [7], is an approach of fuzzy clustering based on the possibilistic memberships representing the degrees of typicality, which has been extensively studied and successfully applied in many areas (see, e.g., [3][6][9][13]). The process of fuzzy clustering partitions a data set $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \subset \Re^p$ into c $(1 < c < n)$ clusters, and each datum $\boldsymbol{x}_j$ may belong to some clusters simultaneously with different degrees μ_{ij}. The possibilistic clustering algorithm (PCA) in [7], denoted by PCA93, performs clustering by minimizing the objective function

$$J_{PCA93}(\boldsymbol{\mu}, \boldsymbol{A}) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^m ||\boldsymbol{x}_j - \boldsymbol{a}_i||^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - \mu_{ij})^m \quad (1)$$

subject to

$$\begin{cases} 0 \le \mu_{ij} \le 1, \quad 1 \le i \le c, 1 \le j \le n & (2a) \\ \sum_{i=1}^{c} \mu_{ij} > 0, \quad 1 \le j \le n & (2b) \\ \sum_{j=1}^{n} \mu_{ij} > 0, \quad 1 \le i \le c & (2c) \end{cases} \quad (2)$$

where $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_c) \in \Re^{cp}$ is the cluster center matrix, $m \ge 1$ is a weighting exponent called the fuzzifier, $\|\cdot\|$ is a norm induced by any inner product, and the coefficients η_i $(1 \le i \le c)$

‡Corresponding author. Tel.: +86-13816247965. E-mail address: yangnan@mail.shufe.edu.cn (N. Yang).

are positive. The constraint $(2b)$ guarantees that each feature point should belong to at least one cluster with nonzero membership, and $(2c)$ assures that none of the clusters is empty and thus we really have a partition into no less than c clusters. It should be noted that throughout this paper we take the l_2 norm for $\|\cdot\|$, i.e., $||\boldsymbol{x}_j - \boldsymbol{a}_i|| = \sqrt{\sum_{k=1}^{p}(x_{jk} - a_{ik})^2}$. Let U_X denote the set of all the matrices $\boldsymbol{\mu} = (\mu_{ij})_{c\times n}$ satisfying the constraints $(2a) \sim (2c)$. In order to solve the optimization problem above, Krishnapuram and Keller [7] suggested an iterative algorithm, i.e., PCA93, through the update equations for $\boldsymbol{\mu}$ and $\boldsymbol{A}$, which are both obtained from the necessary conditions for a minimizer of J_{PCA93} with

$$\mu_{ij} = \frac{1}{1 + \left(\frac{||\boldsymbol{x}_j - \boldsymbol{a}_i||^2}{\eta_i}\right)^{1/(m-1)}}, \quad 1 \le i \le c, 1 \le j \le n, \tag{3}$$

and

$$\boldsymbol{a}_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m \boldsymbol{x}_j}{\sum_{j=1}^{n} \mu_{ij}^m}, \quad 1 \le i \le c, \tag{4}$$

respectively. After that, the other three PCAs were presented in [8][5][10], denoted as PCA96, PCA03, and PCA06, respectively, which are listed as follows,

(PCA96, Krishnapuram and Keller [8]) the optimization problem:

$$\begin{cases} J_{PCA96}(\boldsymbol{\mu}, \boldsymbol{A}) = \sum\limits_{i=1}^{c}\sum\limits_{j=1}^{n} \mu_{ij}\, \|\boldsymbol{x}_j - \boldsymbol{a}_i\|^2 + \sum\limits_{i=1}^{c} \eta_i \sum\limits_{j=1}^{n} (\mu_{ij} \ln \mu_{ij} - \mu_{ij}) \\ \text{subject to:} \\ \quad 0 < \mu_{ij} \le 1, \quad 1 \le i \le c, 1 \le j \le n \end{cases} \tag{5}$$

with the update equations for $\boldsymbol{\mu}$

$$\mu_{ij} = \exp\left\{-\frac{1}{\eta_i}\|\boldsymbol{x}_j - \boldsymbol{a}_i\|^2\right\}, \quad 1 \le i \le c, 1 \le j \le n \tag{6}$$

and the update equations for $\boldsymbol{A}$

$$\boldsymbol{a}_i = \frac{\sum_{j=1}^{n} \mu_{ij}\, \boldsymbol{x}_j}{\sum_{j=1}^{n} \mu_{ij}}, \quad 1 \le i \le c; \tag{7}$$

(PCA03, Höppner and Klawonn [5]) the optimization problem:

$$\begin{cases} J_{PCA03}(\boldsymbol{\mu}, \boldsymbol{A}) = \sum\limits_{i=1}^{c}\sum\limits_{j=1}^{n} \mu_{ij}^m ||\boldsymbol{x}_j - \boldsymbol{a}_i||^2 + \sum\limits_{i=1}^{c} \eta_i \sum\limits_{j=1}^{n} (\mu_{ij}^m - m\mu_{ij}) \\ \text{subject to:} \\ \quad \boldsymbol{\mu} \in U_X \end{cases} \tag{8}$$

with the update equations for $\boldsymbol{\mu}$

$$\mu_{ij} = \frac{1}{\left(1 + \frac{||\boldsymbol{x}_j - \boldsymbol{a}_i||^2}{\eta_i}\right)^{1/(m-1)}}, \quad 1 \le i \le c, 1 \le j \le n, \tag{9}$$

and the update equations (4) for $\boldsymbol{A}$;

(PCA06, Yang and Wu [10]) the optimization problem:

$$\begin{cases} J_{PCA06}(\boldsymbol{\mu}, \boldsymbol{A}) = \sum\limits_{i=1}^{c}\sum\limits_{j=1}^{n} \mu_{ij}^m ||\boldsymbol{x}_j - \boldsymbol{a}_i||^2 + \frac{\beta}{m^2\sqrt{c}} \sum\limits_{i=1}^{c}\sum\limits_{j=1}^{n} (\mu_{ij}^m \ln \mu_{ij}^m - \mu_{ij}^m) \\ \text{subject to:} \\ \quad 0 < \mu_{ij} \le 1, \quad 1 \le i \le c, 1 \le j \le n \end{cases} \tag{10}$$

with the update equations for $\boldsymbol{\mu}$

$$\mu_{ij} = \exp\left\{-\frac{m\sqrt{c}}{\beta}||\boldsymbol{x}_j - \boldsymbol{a}_i||^2\right\}, \quad 1 \leq i \leq c, 1 \leq j \leq n \tag{11}$$

and the update equations (4) for $\boldsymbol{A}$, where

$$\beta = \sum_{j=1}^{n} ||\boldsymbol{x}_j - \overline{\boldsymbol{x}}||^2 \;/\; n \text{ with } \overline{\boldsymbol{x}} = \sum_{j=1}^{n} \boldsymbol{x}_j \;/\; n. \tag{12}$$

Furthermore, it was claimed in [2][5] that for different choices of the second term in the objective functions of the PCAs, different algorithms can be obtained with different membership functions. Subsequently, a general framework of the PCAs was provided in [12] by examining the characteristics of these membership functions. However, except the aforementioned four classes of functions, no more objective functions were suggested for possibilistic clustering in the literature.

Although the extensive numerical experiments with these PCAs on different data sets from a wide range of applications have established the applicability and practicality of such techniques, the convergence of the PCAs has not been rigorously established. In [5], the convergence characters of the fuzzy c-means (FCM) algorithm and the PCAs were discussed under a unified view, and the algorithm PCA03 was testified to be convergent through a reformulation of the original objective function J_{PCA03}. It was also stated that the proof can be generalized to other similar algorithms. However, it is not easy for the PCAs because of complexity and diversity of the membership functions of the PCAs, and the convergence issue on the PCAs has not been resolved explicitly.

In this paper, we investigate the convergence performance of the PCAs. Bezdek [1] and Hathaway *et al.* [4] had established the convergence of FCM utilizing the reformulated Zangwill convergence theorem. It is shown in this paper that this approach works as well for the PCAs. We first show by means of the Zangwill's theorem that the iterative sequence generated by PCA93 would converge globally to a minimizer or a saddle point of the objective function J_{PCA93} at worst along a subsequence, where "globally" means the convergence occurs from any initializations. The result is also applicable to PCA96 and PCA03.

The rest of this paper is organized as follows. In Section 2, the problem description of the convergence of the PCAs is stated, and the reformulated Zangwill's convergence theorem to be used is reviewed briefly. Then the convergence of PCA93 is proven in Section 3 utilizing the Zangwill's theorem. In Section 4, we demonstrate that the proof can be extended to PCA96 and PCA03 with some slight modifications. Section 5 contains a short summarization of the proof strategy.

2 Convergence of the PCAs

Numerical experiments with real data have verified the usefulness of the possibilistic clustering algorithms. Our goal below is to prove they are theoretically sound. As a preliminary, this section first describes the problem by defining some new notations, and then expatiate the general proof strategy to be used for this problem.

2.1 Problem description

In general, the procedure of the PCAs can be summarized as follows:

Possibilistic Clustering Algorithms

Step 0 One of the optimization problems (1), (5) and (8), is given. In other words, the objective function and the constraints are predetermined.

Step 1 Initialize $\boldsymbol{\mu}^{(0)} \in \Re^{cp}$, and set a small number $\epsilon > 0$ and iteration counter $l = 0$.

Step 2 Compute $\boldsymbol{A}^{(l+1)}$ using the update equations (4) or (7) for $\boldsymbol{A}$.

Step 3 Compute $\boldsymbol{\mu}^{(l+1)}$ using the evaluation equations (3), (6), or (9) for $\boldsymbol{\mu}$.

Step 4 Increase l until $\max_{i,j} |\mu_{ij}^{(l+1)} - \mu_{ij}^{(l)}| < \epsilon$.

By this procedure, an iterative sequence $\{(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)})\}$ is generated. The problem we seek to resolve is whether or not $\{(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)})\}$ converges. The following notation is given in order to further describe the iteration. Let

$$F : \Re^{cp} \mapsto U_X, F(\boldsymbol{A}) = F(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_c) = \boldsymbol{\mu} \tag{13}$$

where the entries of $\boldsymbol{\mu} = (\mu_{ij})_{c\times n}$ are calculated by (3), (6), (9) or (11). Let

$$G : U_X \mapsto \Re^{cp}, G(\boldsymbol{\mu}) = \boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_c) \tag{14}$$

where the vectors $\boldsymbol{a}_i \in \Re^p$ $(1 \leq i \leq c)$ are calculated via (4) or (7). Using F and G, we define the PCA operator as $T_p : U_X \times \Re^{cp} \mapsto U_X \times \Re^{cp}$ by

$$T_p : T_2 \circ T_1 \tag{15}$$

where

$$T_1 : U_X \times \Re^{cp} \mapsto \Re^{cp}, T_1(\boldsymbol{\mu}, \boldsymbol{A}) = G(\boldsymbol{\mu}), \tag{16}$$

$$T_2 : \Re^{cp} \mapsto U_X \times \Re^{cp}, T_2(\boldsymbol{A}) = (F(\boldsymbol{A}), \boldsymbol{A}). \tag{17}$$

Then we have

$$T_p(\boldsymbol{\mu}, \boldsymbol{A}) = (T_2 \circ T_1)(\boldsymbol{\mu}, \boldsymbol{A}) = (F \circ G(\boldsymbol{\mu}), G(\boldsymbol{\mu})). \tag{18}$$

By (18), the iterative sequence can be rewritten as

$$(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)}) = T_p^l(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)}) = ((F \circ G)^l(\boldsymbol{\mu}^{(0)}), G^l(\boldsymbol{\mu}^{(0)})), \quad l = 1, 2, \cdots \tag{19}$$

One of the most critical issues in the PCAs is to prove whether or not $\{T_p^l(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})\}$ defined in (19) is convergent.

2.2 Proof strategy

The Zangwill's convergence theorem [11] provides a useful approach to analyze the convergence of sequences which has been utilized to establish the convergence of FCM in [1][4]. Motivated by the similarity between the FCM algorithm and the PCAs, our proof strategy will apply the theorem of Zangwill to the PCA operator T_p.

Let $f : \Re^p \mapsto \Re$ be a real function with domain D_f, and S be the solution set of the optimization problem $\min_{D_f} f(\boldsymbol{x})$. Zangwill defined an iterative algorithm for solving the problem as any *point to set* mapping $Z : D_f \mapsto P(D_f)$, where $P(D_f)$ is the power set of D_f. The algorithm of interest here is a *point to point* map $Z = T_p$, so we are interested in the special case $Z : D_f \mapsto D_f$. Consequently, we reduce the closedness constraint on Z in [11] by ordinary continuity, and restate the convergence theorem for our particular case as follows.

Theorem 1 *Let the point-to-point map $Z : D_f \mapsto D_f$ determine an algorithm that generates the sequence $\{z^{(l)}\}_1^\infty$ for a given point $z^{(0)} \in D_f$. Also let a solution set $S \subset D_f$ be given. Suppose*

C1. (Descent Constraint) there is a continuous function $g : D_f \mapsto \Re$ such that:
(a) if z is not a solution, then $g(Z(z)) < g(z)$,
(b) if z is a solution, then $g(Z(z)) \leq g(z)$;

C2. (Continuity Constraint) Z is continuous on $(D_f \setminus S)$;

C3. (Compactness Constraint) all points $z^{(l)}$ are in a compact set in $K \subset D_f$, *then either the algorithm stops at a solution, or the limit of any convergent subsequence is a solution.*

On the convergence issue of the PCAs, we have $Z = T_p$, $z^{(l)} = (\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)})$, and g corresponds to the objective functions of the PCAs. In order to proceed, what we need to do is to verify the objective function (e.g., J_{PCA93}) satisfies the descent constraint for a proper solution set S, T_p satisfies the continuity constraint, and $\{(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)})\}_1^{\infty}$ satisfies the compactness constraint. Following this strategy, Section 3 gives the detailed proof procedure for PCA93.

3 Convergence of PCA93

In this section, we assume that the fuzzifier $m > 1$. In order to establish convergence of PCA93, the three constraints in Theorem 1 are verified in turn.

3.1 Descent constraint

First we show that the descent constraint holds for $J_p = J_{PCA93}$, which is the first requirement of Theorem 1.

Lemma 1 *Let $\varphi : U_X \mapsto \Re$, $\varphi(\boldsymbol{\mu}) = J_p(\boldsymbol{\mu}, \boldsymbol{A})$, where $\boldsymbol{A}$ is fixed. Then $\boldsymbol{\mu}^* \in U_X$ is a global minimum solution of φ if and only if $\boldsymbol{\mu}^* = F(\boldsymbol{A})$, where F is defined by (13) and (3).*

Proof: Minimization of φ over U_X is an optimization problem with $2cn + n + c$ linear inequality constraints $(2a) \sim (2c)$. By letting

$$y_{ij}(\boldsymbol{\mu}) = \mu_{ij} - 1, \quad 1 \le i \le c, 1 \le j \le n, \tag{20}$$

$$z_{ij}(\boldsymbol{\mu}) = -\mu_{ij}, \quad 1 \le i \le c, 1 \le j \le n, \tag{21}$$

$$\zeta_j(\boldsymbol{\mu}) = -\sum_{i=1}^{c} \mu_{ij}, \quad 1 \le j \le n, \tag{22}$$

$$\varsigma_i(\boldsymbol{\mu}) = -\sum_{j=1}^{n} \mu_{ij}, \quad 1 \le i \le c, \tag{23}$$

the original optimization problem is rewritten as

$$\begin{cases} \min \varphi(\boldsymbol{\mu}) \\ \text{subject to:} \\ \quad y_{ij}(\boldsymbol{\mu}) \le 0, \quad 1 \le i \le c, 1 \le j \le n \\ \quad z_{ij}(\boldsymbol{\mu}) \le 0, \quad 1 \le i \le c, 1 \le j \le n. \\ \quad \zeta_j(\boldsymbol{\mu}) < 0, \quad 1 \le j \le n. \\ \quad \varsigma_i(\boldsymbol{\mu}) < 0, \quad 1 \le i \le c. \end{cases} \tag{24}$$

Suppose that $\boldsymbol{\mu}^*$ is a minimizer of (24). Then it must satisfy the following KKT conditions,

(1) $\boldsymbol{\mu}^*$ is feasible, i.e.,

$$\begin{aligned} & y_{ij}(\boldsymbol{\mu}^*) \le 0, \quad 1 \le i \le c, 1 \le j \le n, \\ & z_{ij}(\boldsymbol{\mu}^*) \le 0, \quad 1 \le i \le c, 1 \le j \le n, \\ & \zeta_j(\boldsymbol{\mu}^*) < 0, \quad 1 \le j \le n, \\ & \varsigma_i(\boldsymbol{\mu}^*) < 0, \quad 1 \le i \le c; \end{aligned} \tag{25}$$

(2) There exist $2cn$ nonnegative multiplies $\lambda_{ij} \geq 0$ and $\tau_{ij} \geq 0$ such that

$$\lambda_{ij} y_{ij}(\boldsymbol{\mu}^*) = 0, \quad 1 \leq i \leq c, 1 \leq j \leq n, \tag{26}$$

$$\tau_{ij} z_{ij}(\boldsymbol{\mu}^*) = 0, \quad 1 \leq i \leq c, 1 \leq j \leq n; \tag{27}$$

(3)

$$\frac{\partial \varphi}{\partial \mu_{ij}}(\boldsymbol{\mu}^*) + \sum_{i=1}^{c}\sum_{j=1}^{n} \lambda_{ij} \frac{\partial y_{ij}}{\partial \mu_{ij}}(\boldsymbol{\mu}^*) + \sum_{i=1}^{c}\sum_{j=1}^{n} \tau_{ij} \frac{\partial z_{ij}}{\partial \mu_{ij}}(\boldsymbol{\mu}^*) = 0, \quad 1 \leq i \leq c, 1 \leq j \leq n. \tag{28}$$

Substituting (20) and (21) into (25)∼(28), we have

$$\begin{cases} 0 \leq \mu^*_{ij} \leq 1 \\ \lambda_{ij}(\mu^*_{ij} - 1) = 0 \\ \tau_{ij}\mu^*_{ij} = 0 \\ m(\mu^*_{ij})^{m-1} d^2_{ij} - m\eta_i (1 - \mu^*_{ij})^{m-1} + \lambda_{ij} - \tau_{ij} = 0 \end{cases} \tag{29}$$

for all $1 \leq i \leq c$ and $1 \leq j \leq n$, $\sum_{i=1}^{c} \mu^*_{ij} > 0$ for $1 \leq j \leq n$ and $\sum_{j=1}^{n} \mu^*_{ij} > 0$ for $\leq i \leq c$. Below we will show that all the multipliers λ_{ij} and τ_{ij} are zero by contrapositive. If there exists a multiplier $\lambda_{st} > 0$ for some (s,t), it follows from (29) that

$$\mu^*_{st} = 1, \quad \tau_{st} = 0, \quad m d^2_{st} + \lambda_{st} = 0. \tag{30}$$

Then we have $\lambda_{st} = -m d^2_{st} \leq 0$, which is contradictive to the assumption $\lambda_{st} > 0$. Similarly, if there exists a multiplier $\tau_{st} > 0$ for some (s,t), it follows from (29) that

$$\mu^*_{st} = 0, \quad \lambda_{st} = 0, \quad -m\eta_s - \tau_{st} = 0. \tag{31}$$

Then we have $\tau_{st} = -m\eta_s < 0$, which is contradictive to the assumption $\tau_{st} > 0$. Substituting $\lambda_{ij} = 0$ and $\tau_{ij} = 0$ into (28), we obtain

$$\begin{aligned} &\tfrac{\partial \varphi}{\partial \mu_{ij}}(\boldsymbol{\mu}^*) = m(\mu^*_{ij})^{m-1} d^2_{ij} - m\eta_i(1 - \mu^*_{ij})^{m-1} = 0 \\ &\Leftrightarrow \mu^*_{ij} = \frac{1}{1 + \left(d^2_{ij}/\eta_i\right)^{1/(m-1)}}, \quad 1 \leq i \leq c, 1 \leq j \leq n. \end{aligned} \tag{32}$$

It is clear that $\mu^*_{ij} > 0$ for all (i,j), thus $\boldsymbol{\mu}^*$ is a feasible solution satisfying (25). The necessity is proved.

To show the sufficiency, we examine $H_\varphi(\boldsymbol{\mu})$, the $(cn \times cn)$ Hessian matrix of φ evaluated at $\boldsymbol{\mu} \in U_X$. It is easy to deduce that

$$\frac{\partial^2 \varphi}{\partial \mu_{ij} \partial \mu_{i'j'}}(\boldsymbol{\mu}) = \begin{cases} \lambda_{ij} & \text{for any } i = i' \text{ and } j = j' \\ 0 & \text{else.} \end{cases} \tag{33}$$

where

$$\lambda_{ij} = m(m-1)(\mu_{ij})^{m-2} d^2_{ij} + m(m-1)\eta_i(1 - \mu_{ij})^{m-2}, \quad 1 \leq i \leq c, 1 \leq j \leq n. \tag{34}$$

Since we assume $m > 1$ in this section, consequently $H_\varphi(\boldsymbol{\mu})$ is a diagonal matrix with all the diagonal element λ_{ij} positive, i.e., a positive definite matrix. Since U_X is a convex set involving a set of linear constraints, minimizing φ subject to $\boldsymbol{\mu} \in U_X$ is a convex program with a strict convex function φ over a convex set U_X. Moreover, it follows from the necessity and (32) that $\boldsymbol{\mu}^* = F(\boldsymbol{A})$ is the one and only KKT point and

$$\frac{\partial \varphi}{\partial \mu_{ij}}(\boldsymbol{\mu}^*) = \frac{\partial \varphi}{\partial \mu_{ij}}(F(\boldsymbol{A})) = 0, \quad 1 \leq i \leq c, 1 \leq j \leq n. \tag{35}$$

As a result, $\boldsymbol{\mu}^* = F(\boldsymbol{A})$ is a strict global minimum solution of φ.

□

Next, we fix $\boldsymbol{\mu} \in U_X$ and consider the minimization of $J_p(\boldsymbol{\mu}, \boldsymbol{A})$ with respect to $\boldsymbol{A}$.

Lemma 2 *Let $\psi : \Re^{cp} \mapsto \Re$, $\psi(\boldsymbol{A}) = J_p(\boldsymbol{\mu}, \boldsymbol{A})$, where $\boldsymbol{\mu} \in U_X$ is fixed. Then $\boldsymbol{A}^*$ is a strict global minimum solution of ψ if and only if $\boldsymbol{A}^* = G(\boldsymbol{\mu})$, where G is defined by (14) and (4).*

Proof: Let us examine the Hessian matrix $H_\psi(\boldsymbol{A})$ of ψ. It is easy to deduce that $H_\psi(\boldsymbol{A})$ is a diagonal matrix defined by

$$\frac{\partial^2 \psi}{\partial a_{ik} \partial a_{i'k'}}(\boldsymbol{A}) = \begin{cases} 2\sum_{j=1}^{n} \mu_{ij}^m & \text{for any } i = i' \text{ and } k = k' \\ 0 & \text{else.} \end{cases} \tag{36}$$

Since $\boldsymbol{\mu} \in U_X$, we have $\sum_{j=1}^{n} \mu_{ij} > 0$ which follows that $\sum_{j=1}^{n} \mu_{ij}^m > 0$ for any $1 \le i \le c$. That implies that $H_\psi(\boldsymbol{A})$ is a positive definite matrix for all $\boldsymbol{A} \in \Re^{cp}$, and hence $\psi(\boldsymbol{A})$ is a strict convex function in $\Re^{cp}$. As a result, $\boldsymbol{A}^*$ is a strict global minimum solution if and only if

$$\begin{aligned} &\frac{\partial \psi}{\partial a_{ik}}(\boldsymbol{A}^*) = 2\sum_{j=1}^{n} \mu_{ij}^m (x_{jk} - a_{ik}^*) = 0 \\ &\Leftrightarrow a_{ik}^* = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_{jk}}{\sum_{j=1}^{n} \mu_{ij}^m}, \quad 1 \le i \le c, 1 \le k \le p \end{aligned} \tag{37}$$

which are equivalent to $\boldsymbol{A}^* = G(\boldsymbol{\mu})$.

□

Based on Lemmas 1 and 2, the first requirement of Theorem 1 – J_p satisfies the descent constraint – can be obtained in the following.

Lemma 3 *Let $S_p = \{(\boldsymbol{\mu}^*, \boldsymbol{A}^*) \in U_X \times \Re^{cp} \big|$*

$$J_p(\boldsymbol{\mu}^*, \boldsymbol{A}^*) < J_p(\boldsymbol{\mu}, \boldsymbol{A}^*) \ \forall \boldsymbol{\mu} \ne \boldsymbol{\mu}^* \tag{38}$$

and

$$J_p(\boldsymbol{\mu}^*, \boldsymbol{A}^*) < J_p(\boldsymbol{\mu}^*, \boldsymbol{A}) \ \forall \boldsymbol{A} \ne \boldsymbol{A}^*\} \tag{39}$$

be the solution set, and let $(\boldsymbol{\mu}, \boldsymbol{A}) \in U_X \times \Re^{cp}$. We have J_p is continuous and $J_p(T_p(\boldsymbol{\mu}, \boldsymbol{A})) \le J_p(\boldsymbol{\mu}, \boldsymbol{A})$ with strictness in the inequality if $(\boldsymbol{\mu}, \boldsymbol{A}) \notin S_p$, where T_p is the algorithm operator of PCA93 in (15).

Proof: First, since $\{y \to \|y\|^2\}, \{y \to 1 - y\}$ and $\{y \to y^m\}$ are continuous, and J_p is the sum of products of such functions, so J_p is continuous on $U_X \times \Re^{cp}$. Next, suppose $(\boldsymbol{\mu}, \boldsymbol{A}) \in U_X \times \Re^{cp}$. Then it follows from (18) that

$$\begin{aligned} J_p(T_p(\boldsymbol{\mu}, \boldsymbol{A})) &= J_p(F \circ G(\boldsymbol{\mu}), G(\boldsymbol{\mu})) \\ &\le J_p(\boldsymbol{\mu}, G(\boldsymbol{\mu})) \text{ by Lemma 1} \\ &\le J_p(\boldsymbol{\mu}, \boldsymbol{A}) \text{ by Lemma 2.} \end{aligned} \tag{40}$$

If the equality prevails throughout in the above argument, then we have

$$\boldsymbol{\mu} = F \circ G(\boldsymbol{\mu}) \text{ and } \boldsymbol{A} = G(\boldsymbol{\mu}). \tag{41}$$

By Lemmas 1 and 2, it follows that

$$\begin{aligned} J_p(\boldsymbol{\mu}, \boldsymbol{A}) &= J(F \circ G(\boldsymbol{\mu}), G(\boldsymbol{\mu})) \\ &< J_p(\boldsymbol{\mu}', G(\boldsymbol{\mu})) \text{ by Lemma 2} \\ &= J_p(\boldsymbol{\mu}', \boldsymbol{A}), \quad \forall \boldsymbol{\mu}' \ne \boldsymbol{\mu} (= F \circ G(\boldsymbol{\mu})) \end{aligned} \tag{42}$$

and

$$\begin{aligned} J_p(\boldsymbol{\mu}, \boldsymbol{A}) &= J(\boldsymbol{\mu}, G(\boldsymbol{\mu})) \\ &< J_p(\boldsymbol{\mu}, \boldsymbol{A}') \text{ by Lemma 1}, \quad \forall \boldsymbol{A}' \neq \boldsymbol{A}(= G(\boldsymbol{\mu})). \end{aligned} \tag{43}$$

(42) and (43) imply that $(\boldsymbol{\mu}, \boldsymbol{A}) \in S_p$.

□

3.2 Continuity constraint

The second requirement of Theorem 1 is that T_p should be continuous on the domain of J_p with S_p deleted. T_p is in fact continuous on all of $U_X \times \Re^{cp}$, as we show in the following.

Lemma 4 *T_p is continuous on $U_X \times \Re^{cp}$.*

Proof: Since $T_p = T_2 \circ T_1$, and the composition of continuous functions is again continuous, it suffices to show that T_1 and T_2 are each continuous. Since $T_1(\boldsymbol{\mu}, \boldsymbol{A}) = G(\boldsymbol{\mu})$, T_1 is continuous if G is. To see that G is continuous in the variable $\boldsymbol{\mu}$, note that G is a vector field, with the resolution by (cp) scalar field as

$$G = (G_{11}, G_{12}, \cdots, G_{cp}) : \Re^{cn} \mapsto \Re^{cp} \tag{44}$$

where $G_{ik} : \Re^{cn} \mapsto \Re$ is defined via (4) as

$$G_{ik}(\boldsymbol{\mu}) = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_{jk}}{\sum_{j=1}^{n} \mu_{ij}^m} = a_{ik}, \quad 1 \leq i \leq c, 1 \leq k \leq p. \tag{45}$$

Now $\{\mu_{ij} \to \mu_{ij}^m\}$ is continuous, $\{\mu_{ij}^m \to \mu_{ij}^m x_{jk}\}$ is continuous, and the sum of continuous functions is again continuous, thus G_{ik} is the quotient of two continuous functions. In view of constraint (2c), the denominator $\sum_{j=1}^{n} \mu_{ij}^m$ never vanishes, then G_{ik} are also continuous for all (i, k). Therefore, G, and in turn T_1, are continuous on their entire domains.

Similarly, since $T_2(\boldsymbol{A}) = (F(\boldsymbol{A}), \boldsymbol{A})$, it suffices to show that F is a continuous function in the variable $\boldsymbol{A}$. F is a vector field with the resolution by (cn) scalar fields as

$$F = (F_{11}, F_{12}, \cdots, F_{cn}) : \Re^{cp} \mapsto \Re^{cn} \tag{46}$$

where $F_{ij} : \Re^{cp} \mapsto \Re$ is defined via (3) as

$$F_{ij}(\boldsymbol{A}) = \frac{1}{1 + \left(\frac{||\boldsymbol{x}_j - \boldsymbol{a}_i||^2}{\eta_i}\right)^{1/(m-1)}}. \tag{47}$$

Since $\{\boldsymbol{a}_i \to \|\boldsymbol{x}_j - \boldsymbol{a}_i\|\}$ is continuous, $\{\|\boldsymbol{x}_j - \boldsymbol{a}_k\| \to \|\boldsymbol{x}_j - \boldsymbol{a}_k\|^{2/(m-1)}\}$ is continuous, and the sum of continuous functions is again continuous, thus F_{ij} is the quotient of two continuous functions. It follows from $d_{ij} = \|\boldsymbol{x}_j - \boldsymbol{a}_i\| \geq 0$ that the denominator $1 + (||\boldsymbol{x}_j - \boldsymbol{a}_i||^2/\eta_i)^{1/(m-1)}$ never vanishes, thus F_{ij} are also continuous for all $1 \leq i \leq c$ and $1 \leq j \leq n$. Therefore, F as well as T_2 are continuous on their entire domains.

□

3.3 Compactness constraint

The final condition required for Theorem 1 is compactness of a subset of $(U_X \times \Re^{cp})$ which contains all of the possible iterative sequences generated by T_p. In order to do that, some notations are given first. Let $\text{conv}(X)$ denote the convex hull of data set X, which is the minimal close convex set containing X. Since X is finite, i.e., each $\boldsymbol{x}_k \in X$ has finite components, so the diameter of X is finite, i.e.,

$$d_X = \max_{1 \leq s,t \leq n} \|\boldsymbol{x}_s - \boldsymbol{x}_t\| < \infty. \tag{48}$$

The coefficients η_i $(1 \le i \le c)$ in J_{PCA93} are calculated by

$$\eta_i = K \frac{\sum_{j=1}^n \mu_{ij}^m \; ||\boldsymbol{x}_j - \boldsymbol{a}_i||^2}{\sum_{j=1}^n \mu_{ij}^m}, \quad 1 \le i \le c, \tag{49}$$

where the constant $K > 0$, or alternatively,

$$\eta_i = \frac{\sum_{\mu_{ij} \ge \alpha} ||\boldsymbol{x}_j - \boldsymbol{a}_i||^2}{\sum_{\mu_{ij} \ge \alpha} 1}, \quad 1 \le i \le c, \tag{50}$$

where $0 < \alpha < 1$ is predetermined. In [7], the value of η_i is suggested to be fixed for all iterations for the sake of stabilities. So the parameters η_i, $1 \le i \le c$, are actually positive constants in this case. Let

$$\eta = \min\{\eta_1, \eta_2, \cdots, \eta_c\} \tag{51}$$

and let

$$D = \frac{1}{1 + (d_X^2/\eta)^{1/(m-1)}}, \tag{52}$$

which is a positive constant to be used in the following lemma.

Lemma 5 *Let $[conv(X)]^c$ be the c-fold Cartesian product of the convex hull of X, $[D,1]^{cn}$ be the cn-fold Cartesian product of the closed interval $[D,1]$, and $(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})$ be the starting point of iteration with J_p. Then*

$$(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)}) = T_p^l(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)}) \in [D,1]^{cn} \times [conv(X)]^c, \quad l = 1, 2, \cdots \tag{53}$$

and $[D,1]^{cn} \times [conv(X)]^c$ is compact in $U_X \times \Re^{cp}$.

Proof: Let $\boldsymbol{\mu}^{(0)} \in U_X$ be chosen, which is possibly not in $[D,1]^{cn}$. Then $\boldsymbol{A}^{(0)} = G(\boldsymbol{\mu}^{(0)})$ is calculated using (4) so that

$$\boldsymbol{a}_i^{(0)} = \frac{\sum_{j=1}^n (\mu_{ij}^{(0)})^m \, \boldsymbol{x}_j}{\sum_{j=1}^n (\mu_{ij}^{(0)})^m}, \quad 1 \le i \le c. \tag{54}$$

By letting

$$\rho_{ik} = \frac{(\mu_{ik}^{(0)})^m}{\sum_{j=1}^n (\mu_{ij}^{(0)})^m}, \quad 1 \le k \le n, \tag{55}$$

we rewrite (54) as

$$\boldsymbol{a}_i^{(0)} = \sum_{k=1}^n \rho_{ik} \boldsymbol{x}_k, \quad 1 \le i \le c \tag{56}$$

with

$$\sum_{k=1}^n \rho_{ik} = \sum_{k=1}^n \left(\frac{(\mu_{ik}^{(0)})^m}{\sum_{j=1}^n (\mu_{ij}^{(0)})^m} \right) = \frac{\sum_{k=1}^n (\mu_{ik}^{(0)})^m}{\sum_{j=1}^n (\mu_{ij}^{(0)})^m} = 1. \tag{57}$$

Furthermore, it follows from the constraints $(2a)$ and $(2c)$ that $0 \le \rho_{ik} \le 1$ for all $1 \le i \le c$ and $1 \le k \le n$, which implies that $\boldsymbol{a}_i^{(0)}$ is a convex combination of X. Therefore $\boldsymbol{a}_i^{(0)} \in \mathrm{conv}(X)$, and hence $\boldsymbol{A}^{(0)} \in [\mathrm{conv}(X)]^c$. Continuing recursively, $\boldsymbol{\mu}^{(1)}$ is calculated via (3) so that

$$\mu_{ij}^{(1)} = \frac{1}{1 + \left(\frac{||\boldsymbol{x}_j - \boldsymbol{a}_i^{(0)}||^2}{\eta_i} \right)^{1/(m-1)}}, \quad 1 \le i \le c, 1 \le j \le n. \tag{58}$$

It follows from (56) and (57) that for any (i,j),

$$\begin{aligned}\|\boldsymbol{x}_j - \boldsymbol{a}_i^{(0)}\| &= \|\boldsymbol{x}_j - \textstyle\sum_{k=1}^n \rho_{ik}\boldsymbol{x}_k\| \\ &= \|\textstyle\sum_{k=1}^n \rho_{ik}\boldsymbol{x}_j - \sum_{k=1}^n \rho_{ik}\boldsymbol{x}_k\| \\ &= \|\textstyle\sum_{k=1}^n \rho_{ik}(\boldsymbol{x}_j - \boldsymbol{x}_k)\| \\ &\le \textstyle\sum_{k=1}^n \rho_{ik}\|\boldsymbol{x}_j - \boldsymbol{x}_k\| \\ &\le \textstyle\sum_{k=1}^n \rho_{ik} d_X = d_X.\end{aligned} \tag{59}$$

Substituting (59) into (58), we have

$$\mu_{ij}^{(1)} \ge \frac{1}{1 + (d_X^2/\eta_i)^{1/(m-1)}} \ge \frac{1}{1 + (d_X^2/\eta)^{1/(m-1)}} = D, \quad 1 \le i \le c, 1 \le j \le n. \tag{60}$$

Therefore $\mu_{ij}^{(1)} \in [D,1]$, and hence $\boldsymbol{\mu}^{(1)} \in [D,1]^{cn}$. After that $\boldsymbol{a}^{(1)} = G(\boldsymbol{\mu}^{(1)}) \in [\text{conv}(X)]^c$ by the same argument as above. Thus every iterative sequence $(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)})$ of T_p belongs to $[D,1]^{cn} \times [\text{conv}(X)]^c$ for any $l \ge 1$. Furthermore, it is clear that $[D,1]^{cn} \times [\text{conv}(X)]^c$ is a compact set.

□

3.4 Convergence theorem for PCA93

We now assemble the hypotheses and results of the above theorems into a formal statement for the convergence of the algorithm PCA93.

Theorem 2 *(Convergence Theorem for PCA93) Suppose $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \in \Re^p$ are given. Let*

$$J_p(\boldsymbol{\mu}, \boldsymbol{A}) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^m ||\boldsymbol{x}_j - \boldsymbol{a}_i||^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1-\mu_{ij})^m, \quad 1 < m < \infty \tag{61}$$

where $\boldsymbol{\mu} \in U_X$, $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_c)$ with $\boldsymbol{a}_i \in \Re^p$ for all i. If $T_p : U_X \times \Re^{cp} \mapsto U_X \times \Re^{cp}$ is the algorithm operator of PCA93, then for any $(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)}) \in U_X \times \Re^{cp}$, either

(1) $\{T_p^l(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})\}$ terminates at a local minimum solution or saddle point of J_p; or

(2) any convergent subsequence $\{T_p^{l_k}(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})\}$ terminates at a local minimum solution or saddle point of J_p.

Proof: Taking J_p as g in Theorem 1, Lemma 1 shows that J_p satisfies the descent constraint for the solution set S_p. Lemma 2 asserts that the iterative algorithm T_p is continuous on $U_X \times \Re^{cp}$, and by Lemma 3, the iterative sequences of the operator T_p are always in a compact subset of the domain of J_p. The result follows immediately from Theorem 1.

□

4 Extensions to PCA96 and PCA03

It is conceivable that PCA96 and PCA03 can be proved to be convergent through a similar procedure as above by Theorem 1. Below we show that by presenting the results directly and only providing necessary details.

4.1 Descent constraints

Lemma 6 *Let* $\varphi : U'_X \mapsto \Re$, $\varphi(\boldsymbol{\mu}) = J_{PCA96}(\boldsymbol{\mu}, \boldsymbol{A})$, *where* $\boldsymbol{A}$ *is fixed,* U'_X *is the domain of* $\boldsymbol{\mu}$ *with*

$$U'_X = \{\boldsymbol{\mu} \mid 0 < \mu_{ij} \le 1, \quad 1 \le i \le c, 1 \le j \le n\}, \tag{62}$$

and J_{PCA96} *is the objective function of PCA96 defined in (5). Then* $\boldsymbol{\mu}^* \in U'_X$ *is a strict global minimum solution of* φ *if and only if*

$$\mu^*_{ij} = \exp\{-\frac{1}{\eta_i}||\boldsymbol{x}_j - \boldsymbol{a}_i||^2\}, \quad 1 \le i \le c, 1 \le j \le n. \tag{63}$$

Proof: At first examine the Hessian matrix $H_\varphi(\boldsymbol{\mu})$. It is easy to deduce that the entries of H_φ are calculated by

$$\frac{\partial^2 \varphi}{\partial \mu_{ij} \partial \mu_{i'j'}}(\boldsymbol{\mu}) = \begin{cases} \eta_i/\mu_{ij} & \text{for any } i = i' \text{ and } j = j' \\ 0 & \text{else.} \end{cases} \tag{64}$$

Since η_i $(1 \le i \le c)$ are positive constants, the diagonal elements $\eta_i/\mu_{ij} > 0$ for any $\boldsymbol{\mu} \in U'_X$, and the denominators $\mu_{ij} > 0$ for all (i, j). Thus $H_\varphi(\boldsymbol{\mu})$ is positive definite, which implies that φ is a strict convex function of $\boldsymbol{\mu}$. Since U'_X is a convex set, the minimization of $\varphi(\boldsymbol{\mu})$ over U'_X is a convex program. Furthermore, the KKT conditions can be used to show that the point $\boldsymbol{\mu}^*$ calculated by (63) is the one and only KKT point via a similar procedure in the proof of Lemma 1. Hence $\boldsymbol{\mu}^*$ is a strict global minimum solution of φ if and only if $\boldsymbol{\mu}^*$ is calculated via (63).

□

Lemma 7 *Let* $\varphi : U_X \mapsto \Re$, $\varphi(\boldsymbol{\mu}) = J_{PCA03}(\boldsymbol{\mu}, \boldsymbol{A})$, *where* $\boldsymbol{A}$ *is fixed, and* J_{PCA03} *is the objective function of PCA03 defined in (8). Also suppose that* $m > 1$. *Then* $\boldsymbol{\mu}^* \in U_X$ *is a strict global minimum solution of* φ *if and only if*

$$\mu^*_{ij} = \frac{1}{\left(1 + \frac{||\boldsymbol{x}_j - \boldsymbol{a}_i||^2}{\eta_i}\right)^{1/(m-1)}}, \quad 1 \le i \le c, 1 \le j \le n. \tag{65}$$

Proof: At first examine the Hessian matrix $H_\varphi(\boldsymbol{\mu})$. It is easy to deduce that the entries of H_φ are calculated by

$$\frac{\partial^2 \varphi}{\partial \mu_{ij} \partial \mu_{i'j'}}(\boldsymbol{\mu}) = \begin{cases} \tau_{ij} & \text{for any } i = i' \text{ and } j = j' \\ 0 & \text{else,} \end{cases} \tag{66}$$

where $\tau_{ij} = m(m-1)(d^2_{ij} + \eta_i \mu^{m-2}_{ij})$. Since $m > 1$ and $\eta_i > 0$, the diagonal elements $\tau_{ij} > 0$ for any $\boldsymbol{\mu} \in U_X$. Thus $H_\varphi(\boldsymbol{\mu})$ is positive definite, which implies that φ is a strict convex function of $\boldsymbol{\mu}$. Hence minimizing φ over U_X is a convex program. Furthermore, the KKT conditions can be used to show that the point $\boldsymbol{\mu}^*$ calculated by (65) is a KKT point via a similar procedure in the proof of Lemma 1. Hence $\boldsymbol{\mu}^*$ is a strict global minimum solution of φ if and only if $\boldsymbol{\mu}^*$ is calculated via (65).

□

4.2 Compactness constraints

It is easy to deduce that Lemmas 2∼4 will also hold for PCA96 and PCA03 according to the similar derivation procedures. Now we investigate the compactness of a subset which contains all of the possible iterative sequences generated in PCA96 and PCA03. The results are showed as follows.

Lemma 8 *Let* $[conv(X)]^c$ *be the c-fold Cartesian product of the convex hull of* X, $[D_1, 1]^{cn}$ *be the cn-fold Cartesian product of the closed interval* $[D_1, 1]$ *with* $D_1 = \exp\{-d^2_X/\eta\}$, $(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})$ *be the starting point of iteration with* J_{PCA96}, *and* T_p *is the algorithm operator of PCA96. Then*

$$(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)}) = T^l_p(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)}) \in [D_1, 1]^{cn} \times [conv(X)]^c, \quad l = 1, 2, \cdots \tag{67}$$

and $[D_1, 1]^{cn} \times [conv(X)]^c$ *is compact in* $U'_X \times \Re^{cp}$.

Proof: It follows from the proof of Lemma 5 that for any $\boldsymbol{\mu}^{(0)} \in U'_X$, we have $\boldsymbol{A}^{(0)} = G(\boldsymbol{\mu}^{(0)}) \in [\text{conv}(X)]^c$. Subsequently, $\boldsymbol{\mu}^{(1)}$ is calculated via (63) so that

$$\mu_{ij}^{(1)} = \exp\{-\frac{1}{\eta_i}||\boldsymbol{x}_j - \boldsymbol{a}_i^{(0)}||^2\}, \quad 1 \le i \le c, 1 \le j \le n. \tag{68}$$

Substituting (59) into (68), we have

$$\mu_{ij}^{(1)} \ge \exp\{-d_X^2/\eta_i\} \ge \exp\{-d_X^2/\eta\} = D_1, \quad 1 \le i \le c, 1 \le j \le n. \tag{69}$$

Therefore $\mu_{ij}^{(1)} \in [D_1, 1]$, and hence $\boldsymbol{\mu}^{(1)} \in [D_1, 1]^{cn}$. Consequently, it follows from Lemma 5 that every iterative sequence $(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)})$ of T_p belongs to $[D_1, 1]^{cn} \times [\text{conv}(X)]^c$ for any $l \ge 1$. Furthermore, it is clear that $[D_1, 1]^{cn} \times [\text{conv}(X)]^c$ is a compact set.

□

Lemma 9 *Let $[conv(X)]^c$ be the c-fold Cartesian product of the convex hull of X, $[D_2, 1]^{cn}$ be the cn-fold Cartesian product of the closed interval $[D_2, 1]$ with $D_2 = (1 + d_X^2/\eta)^{-\frac{1}{m-1}}$, $(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})$ be the starting point of iteration with J_{PCA03}, and T_p is the algorithm operator of PCA03. Then*

$$(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)}) = T_p^l(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)}) \in [D_2, 1]^{cn} \times [conv(X)]^c, \quad l = 1, 2, \cdots \tag{70}$$

and $[D_2, 1]^{cn} \times [conv(X)]^c$ is compact in $U_X \times \Re^{cp}$.

Proof: It follows from the proof of Lemma 5 that for any $\boldsymbol{\mu}^{(0)} \in U_X$, we have $\boldsymbol{A}^{(0)} = G(\boldsymbol{\mu}^{(0)}) \in [\text{conv}(X)]^c$. Subsequently, $\boldsymbol{\mu}^{(1)}$ is calculated via (65) so that

$$\mu_{ij}^{(1)} = \frac{1}{\left(1 + \frac{||\boldsymbol{x}_j - \boldsymbol{a}_i^{(0)}||^2}{\eta_i}\right)^{1/(m-1)}}, \quad 1 \le i \le c, 1 \le j \le n. \tag{71}$$

Substituting (59) into (71), we have

$$\mu_{ij}^{(1)} \ge (1 + d_X^2/\eta_i)^{-\frac{1}{m-1}} \ge (1 + d_X^2/\eta)^{-\frac{1}{m-1}} = D_2, \quad 1 \le i \le c, 1 \le j \le n. \tag{72}$$

Therefore $\mu_{ij}^{(1)} \in [D_2, 1]$, and hence $\boldsymbol{\mu}^{(1)} \in [D_2, 1]^{cn}$. Consequently, it follows from Lemma 5 that every iterative sequence $(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)})$ of T_p belongs to $[D_2, 1]^{cn} \times [\text{conv}(X)]^c$ for any $l \ge 1$. Furthermore, it is clear that $[D_2, 1]^{cn} \times [\text{conv}(X)]^c$ is a compact set. □

4.3 Convergence theorems for PCA96 and PCA03

Finally we conclude the convergence theorems for the two PCAs by assembling the hypotheses and results of the above theorems.

Theorem 3 *(Convergence Theorem for PCA96) Suppose $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \in \Re^p$ are given. Let*

$$J_{PCA96}(\boldsymbol{\mu}, \boldsymbol{A}) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij} \|\boldsymbol{x}_j - \boldsymbol{a}_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (\mu_{ij} \ln \mu_{ij} - \mu_{ij}) \tag{73}$$

where $\boldsymbol{\mu} \in U'_X$, $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_c)$ with $\boldsymbol{a}_i \in \Re^p$ for all i. If $T_p : (U'_X \times \Re^{cp}) \mapsto (U'_X \times \Re^{cp})$ is the algorithm operator of PCA96, then for any $(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)}) \in U'_X \times \Re^{cp}$, either

(1) $\{T_p^l(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})\}$ terminates at a local minimum solution or saddle point of J_{PCA96}; or

(2) any convergent subsequence $\{T_p^{l_k}(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})\}$ terminates at a local minimum solution or saddle point of J_{PCA96}.

Theorem 4 *(Convergence Theorem for PCA03) Suppose $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \in \Re^p$ are given. Let*

$$J_{PCA03}(\boldsymbol{\mu}, \boldsymbol{A}) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^m \|\boldsymbol{x}_j - \boldsymbol{a}_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n}(\mu_{ij}^m - m\mu_{ij}), \quad 1 < m < \infty \tag{74}$$

where $\boldsymbol{\mu} \in U_X$, $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_c)$ with $\boldsymbol{a}_i \in \Re^p$ for all i. If $T_p : (U_X \times \Re^{cp}) \mapsto (U_X \times \Re^{cp})$ is the algorithm operator of PCA03, then for any $(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)}) \in U_X \times \Re^{cp}$, either

(1) $\{T_p^l(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})\}$ terminates at a local minimum solution or saddle point of J_{PCA03}; or

(2) any convergent subsequence $\{T_p^{l_k}(\boldsymbol{\mu}^{(0)}, \boldsymbol{A}^{(0)})\}$ terminates at a local minimum solution or saddle point of J_{PCA03}.

5 Conclusion

Different from the FCM algorithm, the possibilistic clustering algorithms include a family of PCAs with different objective functions and different membership functions. This fact makes the theoretical convergence of the PCAs more complex. Due to the similarity between the FCM and the PCAs, this paper considers to establish convergence of the PCAs by the specific case of the Zangwill's convergence theorem. The proof procedure can be summarized as the following four critical steps.

S1. (Strict Convexity of $\varphi(\boldsymbol{\mu})$) For any fixed $\boldsymbol{A} \in \Re^{cp}$, the function $\varphi(\boldsymbol{\mu}) = J_p(\boldsymbol{\mu}, \boldsymbol{A})$ is a strict convex function of $\boldsymbol{\mu}$ and the domain of $\boldsymbol{\mu}$ is convex, which is attained by examining the Hessian matrix of φ. This step depends on the objective function and the membership function used.

S2. (Strict Convexity of $\psi(\boldsymbol{A})$) For any fixed $\boldsymbol{\mu}$ in the domain, the function $\psi(\boldsymbol{A}) = J_p(\boldsymbol{\mu}, \boldsymbol{A})$ is a strict convex function of $\boldsymbol{A}$, which holds for all the PCAs since $\boldsymbol{\mu}$ is considered as a constant in this step.

S3. (Continuity of Objective Function) The objective function $J_p(\boldsymbol{\mu}, \boldsymbol{A})$ is continuous in the domain, which follows from continuity of the membership function used directly.

S4. (Compactness of Iterative Sequence) The iterative sequence $\{(\boldsymbol{\mu}^{(l)}, \boldsymbol{A}^{(l)})\}$ generated by the PCAs is contained in a compact set. In this step we only need to show that $\boldsymbol{\mu}^{(l)}$ has a positive lower bound.

The above proof strategy can be applied to establish the convergence in more general situations. However, it is not applicable to PCA06 since the objective function J_{PCA06} is not strictly convex on $\boldsymbol{\mu}$, which does not imply that the algorithm PCA06 does not converge. The performance of PCA06 requires further investigation.

Acknowledgments

This work was supported in part by the Shanghai Philosophy and Social Science Planning Project grant (2012XAL022), Australian Research Council Discovery grants (DP1096218 and DP130102691) and Linkage grants (LP100200774 and LP120100566).

References

[1] Bezdek, J.C., A Convergence theorem for the fuzzy ISODATA clustering algorithms, *IEEE Transactions on Pattern Analysis and Machine Intellegence*, Vol. PAMI-2, No. 1, 1-8, 1980.

[2] Dave, R.N., and Krishnapuram, R., Robust clustering methods: a unified view, *IEEE Transactions on Fuzzy Systems*, Vol. 5, No. 2, 270-293, 1997.

[3] Dey, V., Pratihar, D.K., and Datta, G.L., Genetic algorithm-tuned entropy-based fuzzy C-means algorithm for obtaining distinct and compact clusters, *Fuzzy Optimization and Decision Making*, Vol. 10, No. 2, 153-166, 2011.

[4] Hathaway, R.J., Bezdek, J.C., and Tucker, W.T., An improved convergence theory for the fuzzy ISODATA clustering algorithms, *The Analysis of Fuzzy Information*, Vol. 3, Boca Raton: CRC Press, 123-132, 1987.

[5] Höppner, F., and Klawonn, F., A contribution to convergence theory of fuzzy c-means and derivatives, *IEEE Transactions on Fuzzy Systems*, Vol. 11, No. 5, 682-694, 2003.

[6] Krishnapuram, R., Frigui, H., and Nasraoui, O., Fuzzy and possibilistic shell clustering algorihm and their application to boundary detection and surface approximation, *IEEE Transactions on Fuzzy Systems*, Vol. 3, 29-60, 1995.

[7] Krishnapuram, R., and Keller, J.M., A possibilistic approach to clustering, *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 2, 98-110, 1993.

[8] Krishnapuram, R., and Keller, J.M., The possibilistic c-means algorithm: insights and recommendations, *IEEE Transactions on Fuzzy Systems*, Vol. 4, No. 3, 385-393, 1996.

[9] Oussalah, M., and Nefti, S., On the use of divergence distance in fuzzy clustering, *Fuzzy Optimization and Decision Making*, Vol. 7, No. 2, 147-167, 2008.

[10] Yang, M.-S., and Wu, K.-L., Unsupervised possibilistic clustering, *Pattern Recognition*, Vol. 39, No. 1, 5-21, 2006.

[11] Zangwill, W., *Nolinear Programming: A Unified Approach*, Englewood Cliffs, NJ: Prentice-Hall, 1969.

[12] Zhou, J., and Hung, C.C., A generalized approach to possibilistic clustering algorithms, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 15, No. 2 Suppl., 117-138, 2007.

[13] Zhang, Y., Chi, Z.-X., A fuzzy support vector classifier based on Bayesian optimization, *Fuzzy Optimization and Decision Making*, Vol. 7, No. 1, 75-86, 2008.