A convergence theorem for the fuzzy subspace clustering (FSC) algorithm

G. Gan*, J. Wu

Department of Mathematics and Statistics, York University, Toronto, Ont., Canada M3J 1P3

Abstract

We establish the convergence of the fuzzy subspace clustering (FSC) algorithm by applying Zangwill's convergence theorem. We show that the iteration sequence produced by the FSC algorithm terminates at a point in the solution set S or there is a subsequence converging to a point in S. In addition, we present experimental results that illustrate the convergence properties of the FSC algorithm in various scenarios. © 2007 Elsevier Ltd. All rights reserved.

Keywords: Clustering; Subspace clustering; Analysis of algorithms; Convergence; Fuzzy set

1. Introduction

Data clustering is an unsupervised process that divides a given data set into groups or clusters such that the points within the same cluster are more similar than points across different clusters [1–3]. The difficulty that conventional clustering algorithms encounter in dealing with high dimensional data sets gives rise to the invention of subspace clustering algorithms or projected clustering algorithms [4] whose goal is to find clusters embedded in subspaces of the original data space with their own associated dimensions. However, almost all of the subspace clustering algorithms give non-zero weights to cluster dimensions uniformly and zero weights to non-cluster dimensions.

Motivated by fuzzy clustering [5–7] and LAC [8], Gan et al. [9] proposed a fuzzy subspace clustering (FSC) algorithm to cluster high dimensional data sets. FSC finds regular subspace clusters with each dimension of the original data being associated with each cluster with a probability or weight. The higher the density of a cluster in a dimension, the more the weight that will be assigned to that dimension. In other words, all dimensions of the original data are associated with each cluster, but they have different degrees of association with that cluster.

* Corresponding author. Tel.: +1 416 7362100x20195;
fax: +1 416 7365757.

  E-mail addresses: gjgan@mathstat.yorku.ca (G. Gan),
wujh@mathstat.yorku.ca (J. Wu).

Numerical examples of the FSC algorithm given in Ref. [9] suggest empirically that the algorithm is at least locally convergent, but no proof of convergence was provided therein. Motivated by the proof of the convergence of the fuzzy *c*-Means/ISODATA algorithm [10–12], in this paper we develop a convergence theorem, using the Zangwill's convergence theorem, for the FSC algorithm.

2. The FSC algorithm

To describe the FSC algorithm, we start with some notation. Let $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a finite data set in the Euclidean space $\mathfrak{R}^d$; let k be an integer $2 \leqslant k < n$; and let $V_{kn}(B_{kn})$ denote the vector space of all real (binary) $k \times n$ matrices. A hard k-partition of D can be represented by a binary $k \times n$ matrix $U = (u_{ji})$ which satisfies

$$u_{ji} \in \{0, 1\}, \quad 1 \leqslant j \leqslant k, \ 1 \leqslant i \leqslant n, \tag{1a}$$

$$\sum_{j=1}^{k} u_{ji} = 1, \quad 1 \leqslant i \leqslant n, \tag{1b}$$

$$\sum_{i=1}^{n} u_{ji} > 0, \quad 1 \leqslant j \leqslant k. \tag{1c}$$

The set of all hard k-partitions of D is denoted by M_k, i.e.,

$$M_k = \{U \in B_{kn} | U \text{ satisfies constraint (1)}\}. \tag{2}$$

A $k \times d$ matrix $W = (w_{jh})$ is said to be a fuzzy dimension weight matrix if W satisfies the following conditions:

$$0 \leqslant w_{jh} \leqslant 1, \quad 1 \leqslant j \leqslant k, \ 1 \leqslant h \leqslant d, \tag{3a}$$

$$\sum_{h=1}^{d} w_{jh} = 1, \quad 1 \leqslant j \leqslant k. \tag{3b}$$

The set of all such fuzzy dimension weight matrices is denoted by M_{fk}, i.e.

$$M_{fk} = \{W \in V_{kd} | W \text{ satisfies constraint (3)}\}. \tag{4}$$

The component w_{jh} specifies the probability of the dimension h belonging to the set of cluster dimensions of the cluster j.

Now let $Z = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k\} \subset \Re^d$ be a set of k centers. Then the objective function of the FSC algorithm is defined as [9]

$$E_{\alpha,\varepsilon}(W, Z, U) = \varepsilon \sum_{j=1}^{k} \sum_{h=1}^{d} w_{jh}^{\alpha} + \sum_{j=1}^{k} \sum_{i=1}^{n} u_{ji}$$
$$\times \sum_{h=1}^{d} w_{jh}^{\alpha} (x_{ih} - z_{jh})^2, \tag{5}$$

where W, Z and U are the fuzzy dimension weight matrix, the center and the hard k-partition of D, respectively, $\alpha \in (1, \infty)$ is a weight component or fuzzifier, and ε is a very small positive real number. Note that any one of W, Z and U can be determined from the other two.

It will be shown later that (W^*, Z^*, U^*) is a local minimum of $E_{\alpha,\varepsilon}$ if and only if, for any $\alpha > 1$ and $\varepsilon > 0$, there holds

$$w_{jh}^* = \frac{1}{\sum_{l=1}^{d} \left[\dfrac{\sum_{i=1}^{n} u_{ji}^* (x_{ih} - z_{jh}^*)^2 + \varepsilon}{\sum_{i=1}^{n} u_{ji}^* (x_{il} - z_{jl}^*)^2 + \varepsilon} \right]^{1/(\alpha-1)}} \tag{6a}$$

for $1 \leqslant j \leqslant k$ and $1 \leqslant h \leqslant d$,

$$z_{jh}^* = \frac{\sum_{i=1}^{n} u_{ji}^* x_{ih}}{\sum_{i=1}^{n} u_{ji}^*} \tag{6b}$$

for $1 \leqslant j \leqslant k$ and $1 \leqslant h \leqslant d$ and

$$u_{ji}^* = 1 \quad \text{for some } j \in \left\{ r \in Q | r = \arg \min_{1 \leqslant l \leqslant k} d_{li} \right\} \tag{6c}$$

for $1 \leqslant j \leqslant k$ and $1 \leqslant i \leqslant n$, where $Q = \{1, 2, \ldots, k\}$ and $d_{li} = \sum_{h=1}^{d} (w_{lh}^*)^{\alpha} (x_{ih} - z_{lh}^*)^2$. The FSC algorithm is a Picard iteration through the loop defined by Eq. (6).

Algorithm 1. The pseudo-code of the FSC algorithm.
Require: D—the data set, k—the number of clusters and α—the fuzzifier
1: Initialize Z by choosing k points from D randomly;
2: Initialize W with $w_{jh} = \frac{1}{d} (1 \leqslant j \leqslant k, 1 \leqslant h \leqslant d)$;
3:     Estimate U from initial values of W and Z according to Eq. (6c);
4:     Let $error = 1$ and $Obj = E_{\alpha,\varepsilon}(W, Z)$;
5:     **while** $error > 0$ **do**
6:         Update Z according to Eq. (6b);
7:         Update W according to Eq. (6a);
8:         Update U according to Eq. (6c);
9:         Calculate $NewObj = E_{\alpha,\varepsilon}(W, Z)$;
10:        Let $error = |NewObj - Obj|$, and then $Obj \Leftarrow NewObj$;
11: **endwhile**
12: Output W, Z and U.

The FSC algorithm is implemented recursively (see Algorithm 1). FSC starts with initial estimates of Z, W and the partition U calculated from Z and W, and then repeats estimating the centers Z given the estimates of W and U, estimating the fuzzy dimension weight matrix W given the estimates of Z and U, and estimating the partition U given the estimates of W and Z until it converges.

FSC requires two parameters: the number of clusters k and the fuzzifier α. Choosing an appropriate k is a challenging problem in data clustering and there are no widely accepted methods [13,14]. In fuzzy clustering, the fuzzifier α is usually specified to be around 1.1 [15]. In our algorithm we desire such an α that the resulting weights of cluster dimensions are large and the resulting weights of non-cluster dimensions are small or close to zero. Suppose we have a cluster with small identical variances in first three dimensions. If we choose α close to 1, then we cannot differentiate between weights $(1, 0, \ldots, 0)$ and $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \ldots, 0)$ for this cluster, since $1 + 0 + \cdots + 0 = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + 0 + \cdots + 0$. Regarding the specification of α, we suggest a value around 2. If we specify $\alpha = 2$, then $1^2 + 0^2 + \cdots + 0^2 = 1 > \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + 0 + \cdots + 0 = \frac{1}{3}$.

The computational cost of the FSC algorithm per iteration (of the while loop) can be decomposed into four parts:

(1) The time required to update Z is $O(nd)$.
(2) The time required to update W is $O(nkd)$.
(3) The time required to update U is $O(nkd)$.
(4) The time required to calculate the objective function is $O((n+1)kd)$.

The number of iterations required for the FSC algorithm to converge depends on the size of data, the number of clusters, initial centers, α and ε. Since the computational cost per iteration is linear to dimension d, the FSC algorithm is well suited for clustering high dimensional data.

3. Zangwill's convergence theorem

Zangwill's results are very useful to establish the convergence properties of iterative algorithms. Before introducing Zangwill's convergence theorem A [16,10], we first recall some concepts.

Definition 1. A point-to-set map Ω from a set X into a set Y is defined as

$$\Omega : X \rightarrow P(Y),$$

which associates a subset of Y with each point of X, where $P(Y)$ denotes the power set of Y.

Definition 2. A point-to-set map $\Omega : X \rightarrow P(Y)$ is said to be open at a point $\bar{\mathbf{x}}$ in X if $\{\mathbf{x}^{(m)}\} \subset X$, $\mathbf{x}^{(m)} \rightarrow \bar{\mathbf{x}}$ and $\bar{\mathbf{y}} \in \Omega(\bar{\mathbf{x}})$ imply the existence of an integer m_0 and a sequence $\{\mathbf{y}^{(m)}\} \subset Y$ such that $\mathbf{y}^{(m)} \in \Omega(\mathbf{x}^{(m)})$ for $m \geqslant m_0$ and $\mathbf{y}^{(m)} \rightarrow \bar{\mathbf{y}}$.

Definition 3. A point-to-set map $\Omega : X \rightarrow P(Y)$ is said to be closed at a point $\bar{\mathbf{x}}$ in X if $\{\mathbf{x}^{(m)}\} \subset X$, $\mathbf{x}^{(m)} \rightarrow \bar{\mathbf{x}}$, $\mathbf{y}^{(m)} \in \Omega(\mathbf{x}^{(m)})$ and $\mathbf{y}^{(m)} \rightarrow \bar{\mathbf{y}}$ imply that $\bar{\mathbf{y}} \in \Omega(\bar{\mathbf{x}})$.

Definition 4. A point-to-set map $\Omega : X \rightarrow P(Y)$ is said to be continuous at a point $\bar{\mathbf{x}}$ in X if it is both open and closed at $\bar{\mathbf{x}}$.

Regarding the composite of a function and a point-to-set map, we have the following property [10].

Corollary 5. Let $C : M \rightarrow V$ be a function and $\Omega : V \rightarrow P(V)$ be a point-to-set map. If C is continuous at $\mathbf{w}_0$ and Ω is closed at $C(\mathbf{w}_0)$, then the point-to-set map $\Omega \circ C : M \rightarrow P(V)$ is closed at $\mathbf{w}_0$.

Now we are ready to introduce Zangwill's convergence theorem A [10,16].

Theorem 6 (*Zangwill's convergence theorem A*). *Given a point $\mathbf{z}^{(0)} \in V$, let the point-to-set map $\Omega : V \rightarrow P(V)$ determine an algorithm that generates the sequence $\{\mathbf{z}^{(m)}\}$. Also a solution set $S \subset V$ is given. Assume that*

(1) *All points $\mathbf{z}^{(m)}$ are in a compact subset of V.*
(2) *There is a continuous function $J : V \rightarrow \Re$ such that:*
   (a) *If $\mathbf{z} \notin S$, then for any $\mathbf{y} \in \Omega(\mathbf{z})$, $J(\mathbf{y}) < J(\mathbf{z})$.*
   (b) *If $\mathbf{z} \in S$, then either the algorithm terminates or for any $\mathbf{y} \in \Omega(\mathbf{z})$, $J(\mathbf{y}) \leqslant J(\mathbf{z})$.*
(3) *The map Ω is closed at $\mathbf{z}$ if $\mathbf{z} \notin S$.*

Then either the algorithm stops at a solution or the limit of any convergent subsequence is a solution.

Obviously, the most difficult part of applying Zangwill's convergence theorem is to find the appropriate solution set S.

4. Convergence of FSC

To establish the convergence properties of FSC by applying Zangwill's convergence theorem, we need some additional notation and propositions.

Let $G_1 : M_{fk} \times M_k \rightarrow \Re^{kd}$ be a function defined as

$$G_1(W, U) = Z = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k)^\mathrm{T}, \tag{7}$$

where the vectors $\mathbf{z}_j = (z_{j1}, z_{j2}, \ldots, z_{jd})^\mathrm{T} \in \Re^d$, $1 \leqslant j \leqslant k$, are computed via Eq. (6b) using W and U. Let $G_2 : \Re^{kd} \times M_k \rightarrow M_{fk}$ be a function defined as

$$G_2(Z, U) = W = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k)^\mathrm{T}, \tag{8}$$

where the vectors $\mathbf{w}_j = (w_{j1}, w_{j2}, \ldots, w_{jd})^\mathrm{T}$, $j = 1, 2, \ldots, k$, are calculated via Eq. (6a). Let $G_3 : M_{fk} \times \Re^{kd} \rightarrow P(M_k)$ be a point-to-set map

$$G_3(W, Z) = \{U \in M_k | U \text{ satisfies Eq. (6c)}\}. \tag{9}$$

Similar to the fuzzy c-means iteration [11], the FSC iteration can be expressed using a point-to-set map $T_{\alpha,\varepsilon} : M_{fk} \times \Re^{kd} \times M_k \rightarrow P(M_{fk} \times \Re^{kd} \times M_k)$ defined by the composition

$$T_{\alpha,\varepsilon} = A_3 \circ A_2 \circ A_1, \tag{10}$$

where $A_1 : M_{fk} \times \Re^{kd} \times M_k \rightarrow \Re^{kd} \times M_k$, $A_2 : \Re^{kd} \times M_k \rightarrow M_{fk} \times \Re^{kd}$ and $A_3 : M_{fk} \times \Re^{kd} \rightarrow P(M_{fk} \times \Re^{kd} \times M_k)$ are defined as

$$A_1(W, Z, U) = (G_1(W, U), U), \tag{11a}$$

$$A_2(Z, U) = (G_2(Z, U), Z), \tag{11b}$$

$$A_3(W, Z) = \{(W, Z, U) | U \in G_3(W, Z)\}. \tag{11c}$$

Thus $T_{\alpha,\varepsilon}(W, Z, U) = \{(\hat{W}, \hat{Z}, \hat{U}) | \hat{Z} = G_1(W, U), \hat{W} = G_2(\hat{Z}, U)$, and $\hat{U} \in G_3(\hat{W}, \hat{Z})\}$, i.e.,

$$(\hat{W}, \hat{Z}, \hat{U}) \in T_{\alpha,\varepsilon}(W, Z, U)$$
$$= A_3 \circ A_2 \circ A_1(W, Z, U)$$
$$= A_3 \circ A_2(G_1(W, U), U)$$
$$= A_3 \circ (G_2(\hat{Z}, U), G_1(W, U))$$
$$= \{(W, Z, U) | U \in G_3(G_2(\hat{Z}, U), G_1(W, U))\}$$
$$= \{(W, Z, U) | U \in G_3(\hat{W}, \hat{Z})\}.$$

Now we define an FSC iteration sequence as follows.

Definition 7. A sequence $\{(W^{(m)}, Z^{(m)}, U^{(m)})\}$ is said to be an FSC iteration sequence if $(W^{(1)}, Z^{(1)}, U^{(1)}) \in M_{fk} \times \Re^{kd} \times M_k$ and $(W^{(m)}, Z^{(m)}, U^{(m)}) \in T_{\alpha,\varepsilon}(W^{(m-1)}, Z^{(m-1)}, U^{(m-1)})$ for $m = 2, 3, \ldots$.

Now we present some propositions. The proofs of these propositions are given in Appendix A.

Proposition 8. Let $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and let $\Xi : M_k \rightarrow \Re$ be defined as $\Xi(U) = E_{\alpha,\varepsilon}(W^*, Z^*, U)$, where $W^* \in M_{fk}$, $Z^* \in \Re^{kd}$, $\alpha > 1$ and $\varepsilon > 0$ are fixed. Suppose that there exists k different integers $i_1, i_2, \ldots, i_k$ $(1 \leqslant i_1, i_2, \ldots, i_k \leqslant n)$ such that

$$d_{ji_j} < d_{ji}, \quad \forall i \neq i_j, \tag{12}$$

where

$$d_{ji} = \sum_{h=1}^{d} (w_{jh}^*)^\alpha (x_{ih} - z_{jh}^*)^2,$$

$$i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, k. \tag{13}$$

Then the set of global minimizers of Ξ over M_k is precisely the subset of M_k satisfying Eq. (6c).

Proposition 9. *Let $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and let $\Theta : M_{fk} \to \mathfrak{R}$ be defined as $\Theta(W) = E_{\alpha,\varepsilon}(W, Z^*, U^*)$, where $Z^* \in \mathfrak{R}^{kd}$, $U^* \in M_k$ and $\alpha > 1$ are fixed. Then W is a global minimizer of Θ over M_{fk} if and only if W is calculated via Eq. (6a).*

Proposition 10. *Let $\Psi : \mathfrak{R}^{kd} \to \mathfrak{R}$ be defined as $\Psi(Z) = E_{\alpha,\varepsilon}(W^*, Z, U^*)$, where $W^* \in M_{fk}$, $U^* \in M_k$ and $\alpha > 1$ are fixed. Then Z is a global minimizer of Ψ over $\mathfrak{R}^{kd}$ if and only if Z satisfies Eq. (6b) for W^*.*

Theorem 11. *Let $\alpha > 1$ be fixed and $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ contain at least $k(<n)$ distinct points. Let the solution set S of the optimization problem*

$$\min_{(W,Z,U) \in M_{fk} \times \mathfrak{R}^{kd} \times M_k} E_{\alpha,\varepsilon}(W, Z, U)$$

be defined as

$$S = \{(\bar{W}, \bar{Z}, \bar{U}) \in M_{fk} \times \mathfrak{R}^{kd} \times M_k \mid$$
$$E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) \leqslant E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, U) \; \forall U \in M_k \text{ and}$$
$$E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) < E_{\alpha,\varepsilon}(W, \bar{Z}, \bar{U}) \; \forall W \neq \bar{W} \text{ and}$$
$$E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) < E_{\alpha,\varepsilon}(\bar{W}, Z, \bar{U}) \; \forall Z \neq \bar{Z}\}. \tag{14}$$

Let $(\bar{W}, \bar{Z}, \bar{U}) \in M_{fk} \times \mathfrak{R}^{kd} \times M_k$. Then $E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}) \leqslant E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U})$ for every $(\hat{W}, \hat{Z}, \hat{U}) \in T_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U})$ and the inequality is strict if $(\bar{W}, \bar{Z}, \bar{U}) \notin S$.

Proof. Let $(\hat{W}, \hat{Z}, \hat{U}) \in T_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U})$, then $(\hat{W}, \hat{Z}, \hat{U}) \in A_3 \circ A_2 \circ A_1(\bar{W}, \bar{Z}, \bar{U})$, where A_1, A_2 and A_3 are defined in Eq. (11). From the definitions of A_1, A_2 and A_3, and Propositions 8, 9 and 10, it follows that $(\hat{W}, \hat{Z}, \hat{U}) \in M_{fk} \times \mathfrak{R}^{kd} \times M_k$ and

$$E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) \geqslant E_{\alpha,\varepsilon}(\bar{W}, \hat{Z}, \bar{U}) \geqslant E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \bar{U})$$

$$\geqslant E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}), \tag{15}$$

which implies

$$E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}) \leqslant E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}).$$

We now show that $E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}) < E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U})$ if $(\bar{W}, \bar{Z}, \bar{U}) \notin S$. To do this, we only need to show that $E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}) \neq E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U})$ if $(\bar{W}, \bar{Z}, \bar{U}) \notin S$. Using a proof by contradiction, we assume that $E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}) = E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U})$ for some $(\bar{W}, \bar{Z}, \bar{U}) \notin S$. Then, from Eq. (15) we have

$$E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) = E_{\alpha,\varepsilon}(\bar{W}, \hat{Z}, \bar{U}) = E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \bar{U}) = E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}). \tag{16}$$

From Proposition 10, we know that $\hat{Z} = G_1(\bar{W}, \bar{U})$ is the global minimizer of the function $E_{\alpha,\varepsilon}(\bar{W}, Z, \bar{U})$. From Eq. (16), we have $E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) = E_{\alpha,\varepsilon}(\bar{W}, \hat{Z}, \bar{U})$, which implies $\bar{Z}$ is also the global minimizer of the same function $E_{\alpha,\varepsilon}(\bar{W}, Z, \bar{U})$. Noting that the function has a unique global minimizer, we have $\bar{Z} = \hat{Z}$. Similarly, we have $\bar{W} = \hat{W}$.

Since $(\bar{W}, \bar{Z}, \bar{U}) \notin S$, from the definition of S we have the following three cases: (a) $E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) > E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, U^*)$ for some $U^* \in M_k$; or (b) $E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) \geqslant E_{\alpha,\varepsilon}(\bar{W}, Z^*, \bar{U})$ for some $Z^* \neq \bar{Z}$; or (c) $E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) \geqslant E_{\alpha,\varepsilon}(W^*, \bar{Z}, \bar{U})$ for some $W^* \neq \bar{W}$.

In Case (a), noting that $\bar{Z} = \hat{Z}$ and $\bar{W} = \hat{W}$, from Proposition 8 we have

$$E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) > E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, U^*)$$
$$= E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, U^*) \geqslant E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}),$$

which contradicts our assumption.

In Case (b), noting that $\bar{Z} \neq Z^*$ and $\hat{Z} = \bar{Z}$, from Proposition 10 and Eq. (16) we have

$$E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) \geqslant E_{\alpha,\varepsilon}(\bar{W}, Z^*, \bar{U}) > E_{\alpha,\varepsilon}(\bar{W}, \hat{Z}, \bar{U}) = E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}),$$

which contradicts our assumption.

In Case (c), noting that $\bar{Z} = \hat{Z}$, $\bar{W} = \hat{W}$ and $W^* \neq \bar{W}$, from Proposition 9 and Eq. (16) we have

$$E_{\alpha,\varepsilon}(\bar{W}, \bar{Z}, \bar{U}) \geqslant E_{\alpha,\varepsilon}(W^*, \bar{Z}, \bar{U}) = E_{\alpha,\varepsilon}(W^*, \hat{Z}, \bar{U})$$
$$> E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \bar{U}) = E_{\alpha,\varepsilon}(\hat{W}, \hat{Z}, \hat{U}),$$

which contradicts our assumption again. This proves the theorem. $\square$

Theorem 12. *Let $\alpha > 1$ be fixed and $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ contain at least $k(<n)$ distinct points. Then the point-to-set map $T_{\alpha,\varepsilon} : M_{fk} \times \mathfrak{R}^{kd} \times M_k \to P(M_{fk} \times \mathfrak{R}^{kd} \times M_k)$ is closed at every point in $M_{fk} \times \mathfrak{R}^{kd} \times M_k$.*

Theorem 13. *Let $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ contain at least $k(<n)$ distinct points, and let $(W^{(0)}, Z^{(0)}, U^{(0)})$ be the starting point of iteration with $T_{\alpha,\varepsilon}$ with $W^{(0)} \in M_{fk}$, $Z^{(0)} \in \mathfrak{R}^{kd}$ and $U^{(0)} \in G_3(W^{(0)}, Z^{(0)})$. Then the iteration sequence $\{(W^{(r)}, Z^{(r)}, U^{(r)})\}$, $r = 1, 2, \ldots$, is contained in a compact subset of $M_{fk} \times \mathfrak{R}^{kd} \times M_k$.*

Since the function $E_{\alpha,\varepsilon}(W, Z, U)$ defined in Eq. (5) is continuous, the following theorem, which establishes the convergence for the FSC algorithm, follows immediately from Theorems 13, 11, 12 and Zangwill's convergence Theorem 6.

Theorem 14 (*Convergence of FSC*). *Let $D = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ contain at least $k(<n)$ distinct points, and let $E_{\alpha,\varepsilon}(W, Z, U)$ defined as in Eq. (5). Let $(W^{(0)}, Z^{(0)}, U^{(0)})$ be the starting point of iteration with $T_{\alpha,\varepsilon}$, with $W^{(0)} \in M_{fk}$, $Z^{(0)} \in \mathfrak{R}^{kd}$ and $U^{(0)} \in G_3(W^{(0)}, Z^{(0)})$. Then the iteration sequence $\{(W^{(r)}, Z^{(r)}, U^{(r)})\}$, $r = 1, 2, \ldots$, either terminates at a point (W^*, Z^*, U^*) in the solution set S; or there is a subsequence converging to a point in S.*

5. Numerical evaluation of FSC

In this section, we present experimental results that illustrate the convergence properties of the FSC algorithm in various scenarios. For experimental results that illustrate the clustering accuracy of the FSC algorithm, readers are referred to Refs. [9,17]. The FSC algorithm was implemented in $C++$ and the $C++$ code was compiled using the default compiler of Dev-C++. Our experiments were conducted on an Acer Aspire 5502ZWXMi laptop having an Intel Pentium M processor 735 (1.7 GHz, 400 MHz FSB, 2MB L2 cache), with 512 MB DDR2 RAM, using the Windows XP operating system.

5.1. Synthetic data generation

To generate synthetic data, we use the data generation method introduced in Ref. [18] which uses the so-called anchor points to generate clusters embedded in subspaces of a high dimensional space. To generate k clusters embedded in different subspaces of different dimensions, the method proceeds by first generating k uniformly distributed anchor points $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k$ in the d-dimensional space. The method then generates the number of dimensions and the number of points associated with each cluster. Finally, it generates points for each cluster and outliers.

The number of dimensions associated with a cluster is generated by a Poison process with mean μ, with the additional restriction that this number is in $[2, d]$. The dimensions for the first cluster are chosen randomly. Once the dimensions for the $(i-1)$th cluster are chosen, the dimensions for the ith cluster are generated inductively by choosing $\min\{d_{i-1}, d_i/2\}$ dimensions from the $(i-1)$th cluster and generating other dimensions randomly, where d_i is the number of dimensions for the ith cluster. Given the percentage of outliers $F_{outlier}$ and the size of the data set n, the number of points in the ith cluster is $N_c \cdot r_i / \sum_{j=1}^{k} r_j$, where $r_1, r_2, \ldots, r_k$ are generated randomly from an exponential distribution with mean 1, and $N_c = n(1 - F_{outlier})$.

In the final step, points in each cluster and outliers are generated as follows. For the ith cluster, the coordinates of the points in non-cluster dimensions are generated uniformly at random from [0, 100], while the coordinates of the points in a cluster dimension j are generated by a normal distribution with mean at the respective coordinate of the anchor point and variance $(s_{ij}r)^2$, where s_{ij} is a scale factor generated uniformly at random from $[1, s]$ and r is a fixed spread parameter. Outliers are generated uniformly at random from the entire space $[0, 100]^d$. In our experiments, we specify $r = s = 2$ in our synthetic data generation.

Our first high dimensional data set A contains 10,000 100-dimensional points with five clusters embedded in different subspaces of different dimensions. No outliers are introduced in this data set. Table 1 summarizes the clusters and their corresponding cluster dimensions.

Our second high dimensional data set B has 10,000 100-dimensional points with five clusters embedded in different subspaces of different dimensions and contains 500 outliers. Table 2 summarizes the clusters and their corresponding cluster dimensions.

5.2. Rate of convergence

The rate of convergence of the FSC algorithm depends on many factors: initial centers, the number of clusters k, the fuzzifier α and ε. Since initial centers were chosen from the original data set randomly, we run the FSC algorithm several times on the data sets A and B with different values of k, α and ε in order to gain insight into the relationship between the rate of convergence and the values of k, α and β. The results are given in Table 3.

In Table 3, the top sections give the average time and the average number of iterations of FSC for different α's with fixed k and ε, the middle sections give the average time and the average number of iterations of FSC for different ε's with fixed

Table 1
The input clusters and their cluster dimensions for the data set A

Input	Dimensions	Points
A	2, 8, 20, 24, 29, 33, 39, 47, 51, 52, 53, 60, 67, 72, 80, 84, 92, 95, 99	2327
B	4, 8, 9, 20, 21, 24, 25, 28, 36, 41, 48, 52, 53, 67, 72, 80, 84, 86, 87, 92, 93, 95	4658
C	2, 9, 11, 18, 25, 41, 42, 48, 52, 57, 67, 68, 72, 92, 95, 97, 98	420
D	2, 14, 18, 33, 41, 44, 47, 52, 53, 72, 89, 93, 95, 98	845
E	2, 7, 11, 12, 14, 18, 23, 27, 33, 41, 44, 47, 53, 65, 67, 69, 72, 86, 89, 98, 99	1750

Table 2
The input clusters and their cluster dimensions for the data set B

Input	Dimensions	Points
A	2, 5, 12, 16, 21, 27, 32, 41, 44, 57, 74, 94	2888
B	1, 2, 7, 9, 12, 13, 21, 27, 33, 40, 41, 44, 58, 74, 94, 95, 100	3147
C	2, 9, 10, 16, 20, 21, 33, 36, 40, 41, 44, 52, 75, 94, 95	1200
D	2, 7, 9, 10, 15, 20, 33, 40, 42, 44, 53, 79, 82, 95, 98, 99, 100	2051
E	2, 7, 19, 28, 33, 37, 40, 42, 53, 69, 70, 80, 83, 97, 100	214
Outlier		500

Table 3
The rate of convergence of FSC on the data sets A and B

k	α	ε	A		B	
			$\bar{T}$	$\bar{I}$	$\bar{T}$	$\bar{I}$
5	1.1	0	179.53	42.0	138.26	32.0
5	1.6	0	142.7	32.8	100.25	22.9
5	2.1	0	88.01	19.8	61.68	13.9
5	2.6	0	89.08	20.4	42.27	9.3
5	3.1	0	120.61	28.0	41.44	9.1
5	3.6	0	127.8	28.2	70.08	15.4
5	2.1	1	85.8	18.6	93.76	16.7
5	2.1	5	50.7	11.2	104.56	17.3
5	2.1	25	55.02	12.1	123.3	20.2
3	2.1	0	30.38	9.3	25.88	6.3
4	2.1	0	27.77	6.9	34.44	6.9
6	2.1	0	123.35	24.9	101.43	16.0
7	2.1	0	197.71	35.1	160.3	22.8

$\bar{T}$ and $\bar{I}$ denote average time (in seconds) and average number of iterations of 10 runs of FSC, respectively.

k and α, and the bottom sections give the average time and the average number of iterations of FSC for different k's with fixed α and ε. From Table 3, we see that, with k and ε fixed, FSC converges faster for α's between 2 and 3 than for other α's. When α and ε are fixed, FSC converges fast for small k's. The effect of ε on the rate of convergence is inconclusive. In addition, the results in Table 3 show that the effect of outliers on the rate of convergence depends on the parameter ε. When $\varepsilon > 0$, FSC converges faster on the data set A, which has no outliers, than on the data set B, which has outliers. When $\varepsilon = 0$, FSC, in general, converges faster on the data set B than on the data set A. In summary, the numerical results show that the FSC algorithm converges in from several to 100 iterations and thus is a scalable algorithm.

6. Conclusions

Subspace clustering algorithms are efficient to deal with high dimensional data sets. The establishment of a subspace clustering algorithm's convergence is very important. In this paper, we showed that the FSC algorithm converges by applying Zangwill's convergence theorem A. Zhangwill's theorem is a very useful tool to establish the convergence of an iterative algorithm. For example, Zhangwill's theorem is also used to show that the generalized alternating minimization (GAM) algorithm, an algorithm based on the EM algorithm [19,20], is convergent [21].

In addition to establishing the convergence theoretically, we presented experimental results that illustrate the convergence properties of the FSC algorithm. We did not give an analytic formula that describes the relation between the rate of convergence of the FSC algorithm and various factors. However, we presented experimental results which show that the FSC algorithm converges in from several to 100 iterations.

Appendix A. Proofs

In this appendix, we give the proofs of the various results presented in the body of this paper.

A.1. Proof of Proposition 8

Proof. To prove this proposition, we first ignore the condition in Eq. (1c), then we will show that the resulting U satisfies Eq. (1c). We rearrange $\Xi(U)$ as

$$\Xi(U) = \sum_{i=1}^{n} \sum_{j=1}^{k} u_{ji} d_{ji} + \varepsilon \sum_{j=1}^{k} \sum_{h=1}^{d} (w_{jh}^{*})^{\alpha}.$$

Note that vectors $\mathbf{u}_i = (u_{1i}, u_{2i}, \ldots, u_{ki})^{\mathrm{T}}$ $(i = 1, 2, \ldots, n)$ are mutually uncorrelated and $\varepsilon \sum_{j=1}^{k} \sum_{h=1}^{d} (w_{jh}^{*})^{\alpha}$ is fixed, therefore $\Xi(U)$ is minimized if and only if

$$\xi_i = \sum_{j=1}^{k} u_{ji} d_{ji}$$

is minimized for $i = 1, 2, \ldots, n$.

On the one hand, if $U \in M_k$ satisfies Eq. (6c), then it is clear that ξ_i is minimized for all $1 \leqslant i \leqslant n$.

On the other hand, if $\xi_i (1 \leqslant i \leqslant n)$ is minimized, we claim that

$$u_{ji} = 1 \text{ implies } j \in \left\{ r \in \{1, 2, \ldots, k\} | r = \arg \min_{1 \leqslant l \leqslant k} d_{li} \right\}$$

for all $1 \leqslant i \leqslant n$. In fact, if this is not true, then there exists an $1 \leqslant i_0 \leqslant n$ and a $1 \leqslant j_0 \leqslant k$ such that $u_{j_0 i_0} = 1$ and

$$d_{j_0 i_0} > \min_{1 \leqslant l \leqslant k} d_{l i_0},$$

which contradicts the assumption that ξ_{i_0} is minimized. From Eq. (12), we have

$$\sum_{i=1}^{n} u_{ji} \geqslant u_{ji_j} = 1 > 0, \quad j = 1, 2, \ldots, k,$$

which shows that U satisfies Eq. (1c). This proves the theorem. $\square$

A.2. Proof of Proposition 9

Lemma 15. *Let M_{f1} be a set defined as*

$$M_{f1} = \left\{ \mathbf{w} \in \Re^d | w_h \in [0, 1], \sum_{h=1}^{d} w_h = 1 \right\}. \tag{A.1}$$

Then M_{f1} is convex.

Lemma 16. *Let $c_1, c_2, \ldots, c_d$ be positive real numbers. Then*

$$\sum_{h=1}^{d} c_h w_h^{\alpha}$$

is strictly convex over M_{f1} for $\alpha > 1$, where M_{f1} is defined in Eq. (A.1).

Lemma 17. *Let $c_1, c_2, \ldots, c_d$ be d positive real numbers, $\alpha > 1$, and M_{f1} be defined as in Eq. (A.1). Let the function $\varphi : M_{f1} \rightarrow \Re$ be defined as*

$$\varphi(\mathbf{w}) = \sum_{h=1}^{d} c_h w_h^{\alpha},$$

where w_h is the hth component of $\mathbf{w}$. Then $\mathbf{w}^*$ is a strict local minimum of φ if and only if $\mathbf{w}^*$ is calculated as

$$w_h^* = \frac{1}{\sum_{l=1}^{d} \left[\frac{c_h}{c_l}\right]^{1/(\alpha-1)}}, \quad 1 \leqslant h \leqslant d. \tag{A.2}$$

Proof. We rearrange the objective function in Eq. (5) as follows:

$$\Theta(W) = \sum_{j=1}^{k} \sum_{h=1}^{d} w_{jh}^{\alpha} \left(\sum_{i=1}^{n} u_{ji}^* (x_{ih} - z_{jh}^*)^2 + \varepsilon \right). \tag{A.3}$$

Since $\mathbf{w}_j = (w_{j1}, w_{j2}, \ldots, w_{jd})^{\mathrm{T}}$ are mutually uncorrelated for $j = 1, 2, \ldots, k$, $E_{\alpha,\varepsilon}(W, Z^*)$ is minimized if and only if

$$\varphi_j = \sum_{h=1}^{d} w_{jh}^{\alpha} \left(\sum_{i=1}^{n} u_{ji}^* (x_{ih} - z_{jh}^*)^2 + \varepsilon \right) = \sum_{h=1}^{d} c_{jh} w_{jh}^{\alpha} \tag{A.4}$$

is minimized for every $j = 1, 2, \ldots, k$, where

$$c_{jh} = \sum_{i=1}^{n} u_{ji}^* (x_{ih} - z_{jh}^*)^2 + \varepsilon.$$

Note that the coefficients c_{jh}s of φ_j in Eq. (A.4) are all positive. According to Lemma 16, φ_j is strictly convex for $\alpha > 1$ and, therefore has at most one minimizer, which is necessarily the global minimizer. From Lemma 17 it follows that minimizer exists and satisfies Eq. (6a). $\square$

A.3. Proof of Proposition 10

Lemma 18. *Let $c_1, c_2, \ldots, c_r$ be real numbers, then the function $\psi : \Re \rightarrow \Re$ defined as*

$$\psi(z) = \sum_{i=1}^{r} (x_i - z)^2$$

is strictly convex and z^* is a global minimizer of ψ over $\Re$ if and only if z^* satisfies

$$z^* = \frac{1}{r} \sum_{i=1}^{r} x_i.$$

Proof. We rearrange $\Psi(Z)$ as

$$\Psi(Z) = \sum_{j=1}^{k} \sum_{h=1}^{d} \sum_{i=1}^{n} u_{ji}^* (w_{jh}^*)^{\alpha} (x_{ih} - z_{jh})^2 + \varepsilon \sum_{j=1}^{k} \sum_{h=1}^{d} (w_{jh}^*)^{\alpha}$$

$$= \sum_{j=1}^{k} \sum_{h=1}^{d} (w_{jh}^*)^{\alpha} \psi_{jh} + \varepsilon \sum_{j=1}^{k} \sum_{h=1}^{d} (w_{jh}^*)^{\alpha},$$

where

$$\psi_{jh} = \sum_{i=1}^{n} u_{ji}^* (x_{ih} - z_{jh})^2.$$

Clearly, $\Psi(Z)$ is minimized if and only if each ψ_{jh} is minimized for $j = 1, 2, \ldots, k$ and $h = 1, 2, \ldots, d$. For each ψ_{jh}, from Lemma 18 we know that ψ_{jh} has a global minimizer and z_{jh} is a global minimizer of ψ_{jh} if and only if z_{jh} is calculated as

$$z_{jh} = \frac{\sum_{i=1}^{n} u_{ji}^* x_{ih}}{\sum_{i=1}^{n} u_{ji}^*}.$$

This proves the theorem. $\square$

A.4. Proof of Theorem 12

To prove Theorem 12, we first prove several lemmas.

Lemma 19. *The function $A_1 : M_{fk} \times \Re^{kd} \times M_k \rightarrow \Re^{kd} \times M_k$ defined in Eq. (11a) is continuous at every point in $M_{fk} \times \Re^{kd} \times M_k$.*

Proof. To prove $A_1 = (G_1(W, U), U)$ is continuous, it suffices to show that $G_1(W, U)$ is continuous. Note that G_1 is a vector field; let

$$G_1 = (G_{11}, G_{12}, \ldots, G_{kd}) : M_{fk} \times M_k \rightarrow \Re^{kd},$$

where $G_{jh} : M_{fk} \times M_k \rightarrow \Re$ is defined via Eq. (6b) as

$$G_{jh}(W, U) = \frac{\sum_{i=1}^{n} u_{ji} x_{ih}}{\sum_{i=1}^{n} u_{ji}}, \quad 1 \leqslant j \leqslant k, \quad 1 \leqslant h \leqslant d.$$

Now $(w_{jh}, u_{ji}) \rightarrow u_{ji} x_{ih}$ is continuous and the sum of continuous functions is continuous, thus G_{jh} is the quotient of two continuous scalar fields for all $1 \leqslant j \leqslant k$ and $1 \leqslant h \leqslant d$. According to constraint (1c), G_{jh} never vanishes, so G_{jh} is also continuous for all $1 \leqslant j \leqslant k$ and $1 \leqslant h \leqslant d$. Therefore, G_1 is continuous. This proves the lemma. $\square$

Lemma 20. *The function $A_2 : \Re^{kd} \times M_k \rightarrow M_{fk} \times \Re^{kd}$ defined in Eq. (11b) is continuous at every point in $\Re^{kd} \times M_k$.*

Proof. Since $A_2 = (G_2(Z, U), Z)$, it suffices to prove that G_2 is a continuous function. Note that G_2 is a vector field with the resolution by $k \times d$ scalar fields,

$$G_2 = (F_{11}, F_{12}, \ldots, F_{kd}) : \Re^{kd} \times M_k \rightarrow \Re^{kd},$$

where $F_{jh} : \Re^{kd} \times M_k \to \Re$ is defined via Eq. (6a) as

$$F_{jh}(Z, U) = \frac{1}{\sum_{l=1}^{d} \left[\dfrac{\sum_{i=1}^{n} u_{ji}(x_{ih} - z_{jh})^2 + \varepsilon}{\sum_{i=1}^{n} u_{ji}(x_{il} - z_{jl})^2 + \varepsilon} \right]^{1/(\alpha-1)}},$$

for $1 \leqslant h \leqslant d$. Since $(z_{jh}, u_{ji}) \to u_{ji}(x_{ih} - z_{jh})^2$ is continuous for all $1 \leqslant j \leqslant k$ and $1 \leqslant h \leqslant d$ and the sum of two continuous functions is continuous, $\sum_{i=1}^{n} u_{ji}(x_{ih} - z_{jh})^2$ is continuous. Since $\sum_{i=1}^{n} u_{ji}(x_{ih} - z_{jh})^2 + \varepsilon > 0$ for all $1 \leqslant j \leqslant k$ and $1 \leqslant h \leqslant d$, we know that the quotient

$$\frac{\sum_{i=1}^{n} u_{ji}(x_{ih} - z_{jh})^2 + \varepsilon}{\sum_{i=1}^{n} u_{ji}(x_{il} - z_{jl})^2 + \varepsilon}$$

is continuous. Since $\alpha > 1$,

$$\frac{\sum_{i=1}^{n} u_{ji}(x_{ih} - z_{jh})^2 + \varepsilon}{\sum_{i=1}^{n} u_{ji}(x_{il} - z_{jl})^2 + \varepsilon} \to \left[\frac{\sum_{i=1}^{n} u_{ji}(x_{ih} - z_{jh})^2 + \varepsilon}{\sum_{i=1}^{n} u_{ji}(x_{il} - z_{jl})^2 + \varepsilon} \right]^{1/(\alpha-1)}$$

is continuous. Thus G_2 is continuous. This proves the lemma. $\square$

Lemma 21. *The function $A_3 : M_{fk} \times \Re^{kd} \to P(M_{fk} \times \Re^{kd} \times M_k)$ defined in Eq. (11c) is closed on $M_{fk} \times \Re^{kd} \times M_k$.*

Proof. Since $A_3 = \{(W, Z, U) \in M_{fk} \times \Re^{kd} \times M_k | U \in G_3(W, Z)\}$, it suffices to show that G_3 is a closed point-to-set mapping. To do this, let

$$(W^{(r)}, Z^{(r)}) \to (\bar{W}, \bar{Z}) \quad \text{as } r \to \infty, \tag{A.5a}$$

$$U^{(r)} \in G_3(W^{(r)}, Z^{(r)}), \quad r = 1, 2, \ldots, \tag{A.5b}$$

$$U^{(r)} \to \bar{U} \quad \text{as } r \to \infty. \tag{A.5c}$$

Then it is necessary to show that $\bar{U} \in G_3(\bar{W}, \bar{Z})$.

In fact, let

$$I_{ji}(W, Z) = \left\{ s \in Q \middle| s = \arg \min_{1 \leqslant l \leqslant k} \sum_{h=1}^{d} w_{lh}^{\alpha}(x_{ih} - z_{lh})^2 \right\}$$

for $1 \leqslant j \leqslant k, 1 \leqslant i \leqslant n$. From Eq. (A.5a) it follows that $I_{ji}(W^{(r)}, Z^{(r)}) \to I_{ji}(\bar{W}, \bar{Z})$ as $r \to \infty$. Note that $I_{ji}(W, Z)$ is a finite set for all i, j, thus there exists an r_1 such that

$$I_{ji}(W^{(r)}, Z^{(r)}) = I_{ji}(\bar{W}, \bar{Z}) \quad \text{for all } r > r_1. \tag{A.6}$$

From Eq. (A.5c) it follows that $u_{ji}^{(r)} \to \bar{u}_{ji}$ as $r \to \infty$. Since $u_{ji} = 1$ or 0 for all k-partition U, there exists an r_2 such that

$$u_{ji}^{(r)} = \bar{u}_{ji} \quad \text{for all } r > r_2. \tag{A.7}$$

From Eq. (A.5b) we have

$$u_{ji}^{(r)} = 1 \quad \text{for some } j \in I_{ji}(W^{(r)}, Z^{(r)}) \; \forall r. \tag{A.8}$$

Thus from Eqs. (A.6)–(A.8) it follows that

$$\bar{u}_{ji} = 1 \quad \text{for some } j \in I_{ji}(\bar{W}, \bar{Z}),$$

which gives $\bar{U} \in G_3(\bar{W}, \bar{Z})$. This proves the theorem. $\square$

Theorem 12 follows immediately from Lemmas 19–21 and Corollary 5.

A.5. Proof of Theorem 13

Theorem 13 follows immediately from the following two lemmas, i.e., Lemmas 22 and 23.

Lemma 22. *Let $[\text{conv}(D)]^k$ be the k-fold Cartesian product of the convex hull of D, and let $(W^{(0)}, Z^{(0)}, U^{(0)})$ be the starting point of iteration with $T_{\alpha,\varepsilon}$ with $W^{(0)} \in M_{fk}$, $Z^{(0)} \in \Re^{kd}$ and $U^{(0)} \in G_3(W^{(0)}, Z^{(0)})$. Then*

$$T_{\alpha,\varepsilon}^r(W^{(0)}, Z^{(0)}, U^{(0)}) \in M_{fk} \times [\text{conv}(D)]^k \times M_k,$$

for $r = 1, 2, \ldots,$ where $T_{\alpha,\varepsilon}^r = T_{\alpha,\varepsilon} \circ T_{\alpha,\varepsilon} \circ \cdots \circ T_{\alpha,\varepsilon}$ (r times).

Proof. Let $W^{(0)} \in M_{fk}$ and $Z^{(0)} \in \Re^{kd}$ be chosen. Then $U^{(0)}$ is calculated to satisfy Eq. (6c), hence $U^{(0)} \in M_k$. Continuing recursively, we have $Z^{(1)} = G_1(W^{(0)}, U^{(0)})$ which is calculated via Eq. (6b) as

$$z_{jh}^{(1)} = \frac{\sum_{i=1}^{n} u_{ji}^{(0)} x_{ih}}{\sum_{i=1}^{n} u_{ji}^{(0)}}, \quad 1 \leqslant j \leqslant k, \; 1 \leqslant h \leqslant d,$$

or

$$\mathbf{z}_j^{(1)} = \frac{\sum_{i=1}^{n} u_{ji}^{(0)} \mathbf{x}_i}{\sum_{i=1}^{n} u_{ji}^{(0)}}, \quad 1 \leqslant j \leqslant k.$$

Let

$$\rho_{ji} = \frac{u_{ji}^{(0)}}{\sum_{i=1}^{n} u_{ji}^{(0)}}, \quad 1 \leqslant i \leqslant n.$$

From Eq. (1), it must be that $0 \leqslant \rho_{ji} \leqslant 1$ for all j, i and therefore

$$\mathbf{z}_j^{(1)} = \sum_{i=1}^{n} \rho_{ji} \mathbf{x}_i,$$

with

$$\sum_{i=1}^{n} \rho_{ji} = \sum_{i=1}^{n} \left(\frac{u_{ji}^{(0)}}{\sum_{i=1}^{n} u_{ji}^{(0)}} \right) = \frac{\sum_{i=1}^{n} u_{ji}^{(0)}}{\sum_{i=1}^{n} u_{ji}^{(0)}} = 1.$$

Thus $\mathbf{z}_j^{(1)} \in \text{conv}(D)$ for all $1 \leqslant j \leqslant k$, and therefore $Z^{(1)} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k)^{\mathrm{T}} \in [\text{conv}(D)]^k$.

Since $W^{(1)} = G_2(Z^{(1)}, U^{(0)})$ is calculated using Eqs. (6a) we know that $W^{(1)} \in M_{fk}$. Then $U^{(1)} \in G_3(W^{(1)}, Z^{(1)})$ is calculated to satisfy Eq. (6c), we know that $U^{(1)} \in M_k$.

Thus every iteration of $T_{\alpha,\varepsilon}$ belongs to $M_{fk} \times [\text{conv}(D)]^k \times M_k$. This proves the lemma. $\square$

Lemma 23. *Let $[\text{conv}(D)]^k$ be the k-fold Cartesian product of the convex hull of D. Then $M_{fk} \times [\text{conv}(D)]^k \times M_k$ is compact in $M_{fk} \times \Re^{kd} \times M_k$.*

Proof. From Eq. (3) it follows that M_{fk} is closed and bounded, and therefore compact. $[\text{conv}(D)]^k$ is also compact [22, Theorem 3].

To see that M_k is compact, we note that D is a finite discrete set. The number of elements in M_k is the number of k-partitions of D. Thus M_k is closed and bounded, and therefore compact. Thus $M_{fk} \times [\text{conv}(D)]^k \times M_k$ is compact in $M_{fk} \times \Re^{kd} \times M_k$. $\quad\square$

References

[1] B. Mirkin, Clustering for data mining: A data recovery approach, Computer Science and Data Analysis Series, Chapman & Hall, CRC, Boca Raton, FL, 2005.

[2] A. Jain, M. Murty, P. Flynn, Data clustering: a review, ACM Comput. Surv. 31 (3) (1999) 264–323.

[3] G. Gan, C. Ma, J. Wu, Data Clustering: Theory, Algorithms, and Applications, in: ASA-SIAM Series on Statistics and Applied Probability, vol. 20, SIAM Press, SIAM, Philadelphia, ASA, Alexandria, VA, USA, 2007.

[4] R. Agrawal, J. Gehrke, D. Gunopulos, P. Raghavan, Automatic subspace clustering of high dimensional data for data mining applications, in: SIGMOD Record ACM Special Interest Group on Management of Data, ACM Press, New York, NY, USA, 1998, pp. 94–105.

[5] H. Xu, H. Wang, C. Li, Fuzzy tabu search method for the clustering problem, in: Proceedings of the 2002 International Conference on Machine Learning and Cybernetics, vol. 2, IEEE, Beijing, 2002, pp. 876–880.

[6] A. Baraldi, P. Blonda, A survey of fuzzy clustering algorithms for pattern recognition, IEEE Trans. Syst. Man Cybern. Part B 29 (6) (1999) 778–785.

[7] A. Baraldi, P. Blonda, A survey of fuzzy clustering algorithms for pattern recognition, IEEE Trans. Syst. Man Cybern. Part B 29 (6) (1999) 786–801.

[8] C. Domeniconi, D. Papadopoulos, D. Gunopulos, S. Ma, Subspace clustering of high dimensional data, in: Proceedings of the SIAM International Conference on Data Mining, Lake Buena Vista, Florida, 2004.

[9] G. Gan, J. Wu, Z. Yang, A fuzzy subspace algorithm for clustering high dimensional data, in: X. Li, O. Zaiane, Z. Li (Eds.), Lecture Notes in Artificial Intelligence, vol. 4093, Springer, Berlin, 2006, pp. 271–278.

[10] R. Hathaway, J. Bezdek, W. Tucker, An improved covergence theorem for the fuzzy c-means clustering algorithms, in: J. Bezdek (Ed.), Analysis of Fuzzy Information, vol. III, CRC Press, Boca Raton, 1987, pp. 123–131.

[11] J. Bezdek, R. Hathaway, M. Sabin, W. Tucker, Convergence theory for fuzzy c-means: counterexamples and repairs, in: J. Bezdek, S. Pal (Eds.), Fuzzy Models for Pattern Recognition: Methods that Search for Approximate Structures in Data, IEEE Press, New York, 1992, pp. 138–142.

[12] M. Sabin, Convergence and consistency of fuzzy c-means/ISODATA algorithms, in: J. Bezdek, S. Pal (Eds.), Fuzzy Models for Pattern Recognition: Methods that Search for Approximate Structures in Data, IEEE Press, New York, 1992, pp. 143–150.

[13] C. Fraley, A. Raftery, How many clusters? Which clustering method? Answers via model-based cluster analysis, Comput. J. 41 (8) (1998) 578–588.

[14] W. Krzanowski, Y. Lai, A criterion for determining the number of groups in a data set using sum-of-squares clustering, Biometrics 44 (1) (1988) 23–34.

[15] Z. Huang, M. Ng, A fuzzy k-modes algorithm for clustering categorical data, IEEE Trans. Fuzzy Syst. 7 (4) (1999) 446–452.

[16] W. Zangwill, Non-linear Programming: A Unified Approach, Prentice-Hall, Englewood Cliffs, NJ, 1969.

[17] G. Gan, Subspace clustering based on fuzzy models and mean shifts, Ph.D. Thesis, Department of Mathematics and Statistics, York University, Toronto, Ont., Canada, March 2007.

[18] C. Aggarwal, J. Wolf, P. Yu, C. Procopiuc, J. Park, Fast algorithms for projected clustering, in: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, ACM, New York, 1999, pp. 61–72.

[19] C. Ordonez, E. Omiecinski, Frem: fast and robust em clustering for large data sets, in: Proceedings of the 11th International Conference on Information and Knowledge Management, ACM, McLean, Virginia, USA, 2002, pp. 590–599.

[20] G. Xuan, W. Zhang, P. Chai, EM algorithms of Gaussian mixture model and hidden Markov model, in: Proceedings of the International Conference on Image Processing, vol. 1, 2001, pp. 145–148.

[21] A. Gunawardana, W. Byrne, Convergence theorems for generalized alternating minimization procedures, J. Mach. Learn. Res. 6 (2005) 2049–2073.

[22] J. Bezdek, Ann convergence theorem for the fuzzy ISODATA clustering algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 2 (1) (1980) 1–8.

About the Author—GUOJUN GAN obtained the B.Sc. degree in computational mathematics and applied software in 2001 from Jilin University, Changchun, PR China. He received the M.Sc. degree and Ph.D. in applied mathematics from York University, Toronto, Canada, in 2003 and 2007, respectively. He is a coauthor for one book and several papers in the area of cluster analysis.

About the Author—JIANHONG WU is a Canada Research Chair in Applied Mathematics (Tier I) and a Professor of Mathematics at York University, Toronto, Ont., Canada. He is the author or coauthor for over 140 papers and four monographs in a number of scientific areas including delay differential equations, dynamical systems, mathematical biology, neural networks and data analysis and he is serving on the editorial boards of seven international journals. He has also been leading a research team and the development of the newly founded Laboratory for Industrial and Applied Mathematics (LIAM).