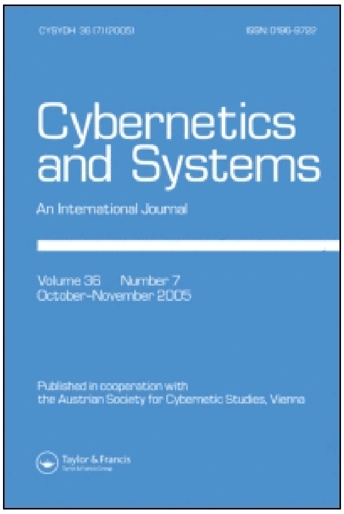

Cybernetics and Systems
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713722751

A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters

J. C. Dunn

To cite this Article Dunn, J. C.(1973) 'A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters', Cybernetics and Systems, 3: 3, 32 — 57
To link to this Article: DOI: 10.1080/01969727308546046
URL: http://dx.doi.org/10.1080/01969727308546046

Journal of Cybernetics, 1974, 3, 3, pp. 32-57

A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters

J. C. Dunn
Department of Theoretical and Applied Mechanics, Cornell University

Abstract

Two fuzzy versions of the k-means optimal, least squared error partitioning problem are formulated for finite subsets X of a general inner product space. In both cases, the extremizing solutions are shown to be fixed points of a certain operator T on the class of fuzzy k-partitions of X, and simple iteration of T provides an algorithm which has the descent property relative to the least squared error criterion function. In the first case, the range of T consists largely of ordinary (i.e. non-fuzzy) partitions of X and the associated iteration scheme is essentially the well known ISODATA process of Ball and Hall. However, in the second case, the range of T consists mainly of fuzzy partitions and the associated algorithm is new; when X consists of k compact well separated (CWS) clusters, X_i, this algorithm generates a limiting partition with membership functions which closely approximate the characteristic functions of the clusters X_i. However, when X is not the union of k CWS clusters, the limiting partition is truly fuzzy in the sense that the values of its component membership functions differ substantially from 0 or 1 over certain regions of X. Thus, unlike ISODATA, the "fuzzy" algorithm signals the presence or absence of CWS clusters in X. Furthermore, the fuzzy algorithm seems significantly less prone to the "cluster-splitting" tendency of ISODATA and may also be less easily diverted to uninteresting locally optimal partitions. Finally, for data sets X consisting of dense CWS clusters embedded in a diffuse background of strays, the structure of X is accurately reflected in the limiting partition generated by the fuzzy algorithm. Mathematical arguments and numerical results are offered in support of the foregoing assertions.

1. Introduction

Computer implemented partitioning algorithms have proved to be useful in many areas of applied science, beginning with Sneath's work on bacteriological taxonomy [1] and extending over a wide range of unrelated fields, including psychology, sociology, geology, medicine, experimental particle physics, operations research, and the technology of automatic reading machines. Extensive bibliographies, good general overviews of partitioning techniques, and applications can be found in [2]–[6].

The function of a partitioning algorithm is to detect natural subgroupings (clusters) within a large finite data set X of multidimensional vectors (patterns), relative to some given quantitative measure of pairwise distance (or similarity) between the elements of X. However, since the specification of a distance measure does not by itself impose a unique interpretation on the ambiguous phrase "natural grouping," it can happen that different algorithms working within the same metrical framework on X will nevertheless partition X

in different ways; in this sense, the organization which an algorithm "sees" in X is dependent upon the structure of the data and the structure of the algorithm. Our particular concern in this paper is with algorithms that are sensitive to the existence of clusters in X which are "compact and well separated" (CWS) relative to a given metric on X (in a sense to be made precise in Sect. 2). Our objective is to devise an algorithm which signals the presence or absence of CWS clusters in X and, in the former case, identifies the characteristic functions of these clusters.

The graph theoretic techniques employed by Sneath [1], Johnson [7], Zahn [4], Wishart [8], and others, afford one possible avenue of approach to the detection of CWS clusters. However, here we follow a different line, which combines Zadeh's fuzzy set concept [9] with the criterion function approach to clustering. Ruspini appears to have been the first to suggest this general scheme and to propose specific fuzzy criterion functions [10] and associated algorithms [11] applicable to a very broad class of distance functions on X. More recently, Gitman and Levine [12] have also applied the theory of fuzzy sets to clustering problems. In the present investigation, we restrict ourselves to the case where X is a finite subset of a general inner product space V, and where the distance function on X is the inner product-induced metric. In this setting, we consider two of many possible differentiable extensions of the k-means squared error criterion function[1] from the class of all "hard" (i.e. ordinary set theoretic) k-partitions of X to the class of all fuzzy k-partitions of X (Sect. 4), and derive corresponding properties of extremal points for these criteria (Sect. 4 and 5): in each case, the extrema are necessarily fixed points of certain operators T on the class of fuzzy k-partitions of X.

The first criterion function is of interest principally because simple iteration of the corresponding operators T is essentially the well-known ISODATA clustering algorithm [13]. Hence our development amounts to a formal derivation of this algorithm. However, in this case, the range of T consists for the most part of hard partitions irrespective of whether CWS clusters are present in X. Consequently, the first criterion function and the associated ISODATA process are unsuitable for our purposes. When CWS clusters are present, experience indicates that the ISODATA process converges rapidly to a partition consisting of these clusters. However, when CWS clusters are not present, the process still converges to some hard partition defined by "unequivocal" characteristic functions, and there is no way to tell from a simple inspection of this limiting partition that its component subsets are in fact not CWS.[2]

The situation is different for the second criterion function, where the range of the corresponding operators T consist essentually of fuzzy partitions. When CWS clusters are present in X, it appears that there is always an extremizing fuzzy partition whose membership functions closely approximate the characteristic functions of the clusters. Furthermore, numerical experiments indicate that simple iteration of T produces a sequence of fuzzy partitions which converge rapidly to the extremizing partition from virtually all starting guesses. On the other hand, when CWS clusters are not present, the iterates of T converge (more slowly) to some limiting partition which is truly fuzzy in the sense that the values of its component membership functions depart significantly from the hard limits 0 and 1. The results of several specific numerical experiments are offered in support of these

[1] Neither extension belongs to Ruspini's scheme.

[2] This objection can be levelled at any algorithm whose output consists exclusively of hard partitions. In such cases it is often necessary to make very extensive secondary computations in order to determine whether or not the sought-after structural property actually is present in the limiting partition.

contentions in Sect. 6. Of special interest are the results obtained for two planar CWS clusters embedded in a diffuse low density background of "strays," forming a connecting bridge and a halo (Fig. 3). The behavior of the algorithm in this case suggests that it may be quite effective when X is a sample drawn from a mixture of unimodal probability distributions. However, a further pursuit of this question and the related parameter estimation problem is not attempted here.

2. Compact Well-Separated Clusters

In this section, we introduce a parameter which provides a simple quantitative index of separation among the subsets of a partition of the data set X. This parameter has explicit theoretical significance for the k-means squared error criterion function approach to clustering, and is also meaningful for the fuzzy algorithm developed in Sect. 5.

For present purposes, we take X to be a non-empty finite subset of an arbitrary real vector space V, and let d denote an arbitrary metric on V. We define set diameters and set distances in the usual way relative to d, namely:

$$\begin{aligned} \operatorname{diam} A &= \sup_{x,y \in A} d(x,y) \\ \operatorname{dist}(A,B) &= \inf_{\substack{x \in A \\ y \in B}} d(x,y) . \end{aligned} \tag{1}$$

Finally $\mathcal{P}(k)$ will denote the class of all partitions $P = \{X_1, \ldots, X_k\}$ of X into k disjoint non-empty subsets X_i, i.e.,

$$\begin{aligned} X_i &\neq \phi \text{ (empty set)} \\ X_i \cap X_j &= \phi \quad i \neq j \\ \bigcup_{i=1}^{k} X_i &= X . \end{aligned} \tag{2}$$

Definition 1. The subsets X_i of a partition P in $\mathcal{P}(k)$ are said to be compact separated (CS) clusters relative to d if and only if they have the following property: for all p,q,r, with $q \neq r$, any pair of points x,y in X_p are closer together (as measured by d) than any pair of points u,v, with u in X_q and v in X_r.

For each fixed k, the existence or non-existence of a k-partition consisting of CS clusters is an intrinsic property of the pair $\{X,d\}$. Furthermore, this property is readily quantified as follows. For each P in $\mathcal{P}(k)$ let,

$$\alpha(k,P) = \frac{\displaystyle\min_{1 \leq q \leq k} \min_{\substack{1 \leq r \leq k \\ r \neq q}} \operatorname{dist}(X_q, X_r)}{\displaystyle\max_{1 \leq p \leq k} \operatorname{diam}(X_p)} . \tag{3}$$

and let

$$\bar{\alpha}(k) = \max_{P \in P(k)} \alpha(k,P) . \tag{4}$$

Then it is easily shown that X can be partitioned into k CS clusters relative to d if and only if $\bar{\alpha}(k) > 1$. The CS property is sometimes regarded as an indispensable feature of any intuitively acceptible concept of "cluster" based on metrics. Nevertheless, other apparently quite different definitions are possible and indeed, some of these appear to have greater operational significance for the prototypical cluster-detecting apparatus of the human visual system [4].

For example, we might agree that clusters should be recognized by the following fundamental "connectivity" property [7]: the subsets X_i of P are connected clusters relative to d if and only if for every x,y in X_p there exists a chain $\xi = \{x = \xi_1, \ldots, \xi_l = y\}$ of elements in X connecting x and y, such that the maximum edge length, $\max_{2 \leq i \leq l} d(\xi_{i-1}, \xi_i)$, of ξ is less than the maximal edge length of any chain connecting u to v, with u in X_q, v in X_r and $r \neq q$. On the face of it, this concept of cluster is quite different from the CS definition; interestingly enough, however, it is not difficult to prove that while subsets X_i which satisfy this connectivity criterion relative to d are in general not CS relative to d, they are always CS relative to a certain (ultra) metric $\hat{d}$ induced by d, namely:

$$\hat{d}(x,y) = \min_{\xi} \left\{ \max_{2 \leq i \leq l} d(\xi_{i-1}, \xi_i) \right\}$$

where the min operation is taken over all chains ξ of arbitrary length l connecting x and y. In short, the intuitive notion of cohesiveness based upon connectivity relative to d has an alternative equivalent expression in terms of the CS concept relative to the induced metric $\hat{d}$. This example argues for the fundamental nature of Definition 1, while at the same time pointing up the fact that we have been talking about structural properties of the pair $\{X,d\}$ and not X alone.

We now limit ourselves to the class of metrics d induced by a norm on V, i.e., metrics of the form $d(x,y) = \| x - y \|$ where $\| \cdot \|$ is any positive definite real function on V satisfying $\| \alpha u \| = |\alpha| \, \| u \|$ for α real and u in V, as well as the triangle inequality $\| x + y \| \leq \| x \| + \| y \|$ [15].[3] In this narrower setting we introduce a more demanding separation index based upon the distance between X_i and the convex hull $\mathrm{Co}X_j$ of X_j in V, i.e., the smallest convex subset of V containing X_j. The related concept of cluster is significant for our purposes because it is directly linked to the fundamental CS definition and to the k-means least squared error approximation problems considered later.

Definition 2. The subsets X_i of a partition P in $\mathcal{P}(k)$ are compact well-separated (CWS) clusters relative to d if and only if they have the following property: for all p,q,r, with $q \neq r$, any pair x,y with x in X_p and y in $\mathrm{Co}X_p$ are closer together as measured by d than any pair u,v, with u in X_q and v in $\mathrm{Co}X_r$.

Once again, the existence or non-existence of a k-partition consisting of CWS clusters is an intrinsic property of $\{X,d\}$. Furthermore, when d is norm-induced, this property is readily quantified as follows. For each P in $\mathcal{P}(k)$ let

$$\beta(k,P) = \frac{\min\limits_{1 \leq q \leq k} \min\limits_{\substack{1 \leq r \leq k \\ r \neq q}} \mathrm{dist}(X_q, \mathrm{Co}X_r)}{\max\limits_{1 \leq p \leq k} \mathrm{diam}(X_p)} \tag{5}$$

[3] Ultrametrics are excluded from this class (cf. [14]); in particular, the mini-max chain length metric $\hat{d}$ described above is not induced by any norm. Consequently, the methods described in the following sections are generally not applicable to the detection of "connected" clusters.

and let

$$\overline{\beta}(k) = \max_{P \in \mathcal{P}(k)} \beta(k,P) \,. \tag{6}$$

It is not difficult to prove that X can be partitioned into k CWS clusters relative to d if and only if

$$\overline{\beta}(k) > 1 \,. \tag{7}$$

The proof depends on the fact that norm-induced distances between any x in X_i and any y in CoX_i cannot exceed diam (X_i); details are omitted in the interest of brevity.

In view of Eqs. (3) and (5), and the fact that $X_i \subset CoX_i$, it follows that the subsets X_i of P are CWS clusters only if they are CS clusters; this justifies the terminology of Definitions 1 and 2. Moreover, since the centroid $\overline{x}_i$ of X_i lies in CoX_i, it readily follows that the X_i's in P are CWS clusters only if

$$d(x,\overline{x}_i) < d(x,\overline{x}_j) \tag{8}$$

for all x in X_i and for all i, j, with $i \neq j$. Finally, as a consequence of Eq. (8) it follows that the X_i's in P are CWS clusters only if P is a fixed point of the ISODATA process; this result establishes a link between CS and CWS clusters and the k-means least squared error approximation problems which we now consider.

3. The k-Means Least Square Approximation Problem and ISODATA

The presence of CWS clusters in X relative to some given metric d (or perhaps, any member of some general class of metrics d) will be assumed an "interesting" structural property. We then want algorithms which are capable of indicating the presence or absence of CWS clusters relative to d, and in the former case, of identifying the characteristic functions of the clusters. Since the class $\mathcal{P}(k)$ is finite when X is finite we can, in principal, use brute force exhaustion to find the "most" interesting partition in $\mathcal{P}(k)$, namely, the partition P' which solves

$$\beta(k,P') = \overline{\beta}(k) = \max_{P \in \mathcal{P}(k)} \beta(,P) \,. \tag{9}$$

Once P' and $\overline{\beta}(k)$ are known, the problem is effectively solved. However, even for moderately large sets X, the number of elements in P (k) can be huge (being equal to $\frac{1}{k!}\sum_{i=1}^{k}\binom{k}{i}(-1)^{k-i}i^n$ for $1 \leqslant k \leqslant n$; (cf. [2]), consequently, exhaustion is not feasible in general. In fact, the mere calculation of $\beta(k,P)$ is a decidedly nontrivial task for all but the simplest problems. It follows that the separation parameter $\overline{\beta}(k)$ and the corresponding maximally separated partition P' which solves (9) are effectively inaccessible by a straightforward approach at this level of generality. On the other hand, an indirect heuristic approach based upon criterion functions and associated approximation problems may provide useful algorithms, especially when d is not simply an arbitrary metric but has some geometrical significance in $V \supset X$. Consequently, in the balance of this paper we make the further restriction that d is induced by a norm $\|\cdot\|$ on V which in turn is induced by an inner product $\langle \cdot | \cdot \rangle$ on V, i.e.,

$$d(x,y) \triangleq \|x - y\| \triangleq \langle x - y | x - y \rangle^{1/2} \tag{10}$$

where $\langle \cdot | \cdot \rangle$ is any symmetric positive-definite bi-linear function from the Cartesian product $V \times V$ into R^1 [15]. This abstract formulation includes as a special case,

$$\begin{aligned} V &= R^n \\ \langle x | y \rangle &= x^t M y \\ d(x,y) &= [(x-y)^t M (x-y)]^{1/2} \end{aligned} \tag{11}$$

where M is an arbitrary symmetric positive definite $n \times n$ matrix, and the superscript t signifies the transpose operation. From the standpoint of clustering problems, three sub-cases are worth mentioning here, namely:

a) $M = I$ = identity matrix,
b) $M^{-1} = \text{diag}\{\sigma_1, \ldots, \sigma_n\}$, σ_i = sample variance of the ith component of vectors $x \in X$,
c) M^{-1} = sample covariance matrix for vectors $x \in X$.

In case *a*, the corresponding metric d in (11) is simply the prototypical Euclidean metric which is invariant under the sub-group of orthogonal transformations on $R^n \supset X$; in case *b* the metric in (11) is a weighted Euclidean metric invariant under the sub-group of scale transformations on R^n (i.e., non-singular linear transformations whose matrices are diagonal, relative to the natural basis in R^n); in case *c* the metric in (11) is a weighted Euclidean metric invariant under the full linear group on R^n. For a further discussion of invariance considerations for clustering problems see [2], [16], and [17].[4]

When d is inner product-induced and when $\overline{\beta}(k)$ is sufficiently large, it seems clear that the maximally separated partition P' solving (7) should also solve the following generalized problem.

k-Means Least Square Approximation (LSA) Problem

Let $\mathcal{L}$ denote the linear hull of X in V, i.e., $\mathcal{L}$ is the (finite dimensional) linear subspace spanned by the elements of X. Let $v = \{v_1, \ldots, v_k\}$ denote an ordered k-tuple of vectors in $\mathcal{L}$, i.e., v is a general element in the k-fold Cartesian product $\mathcal{L}^k$ of $\mathcal{L}$ with itself. For each $P = \{X_1, \ldots, X_k\} \in \mathcal{P}(k)$ and each $v \in \mathcal{L}^k$ put

$$J(P,v) = \sum_{i=1}^{k} \sum_{x \in X_i} d(x,v_i)^2 = \sum_{i=1}^{k} \sum_{x \in X_i} \|x - v_i\|^2 = \sum_{i=1}^{k} \sum_{x \in X} \langle x - v_i | x - v_i \rangle . \tag{12}$$

Find $P' \in \mathcal{P}(k)$ and $v' \in \mathcal{L}^k$ such that

$$J(P', v') = \min_{P \in \mathcal{P}(k)} \min_{v \in \mathcal{L}^k} J(P,v) . \tag{13}$$

Since d is inner product-induced, we always have $v_i' = \overline{x}_i$ = the unweighted mean, or centroid of X_i. Furthermore, when $\beta(k)$ is sufficiently large, computational experience suggests that the optimal partition P' in (13) can probably be found, and very efficiently, by the following ISODATA Process [13]:

step 1) Choose $P \in \mathcal{P}(k)$;
step 2) Compute the centroids $\overline{x}_i$ of $X_i \in P$;
step 3) Construct a new partition $\hat{P}$ according to the rule:

[4] The clustering algorithm proposed in [17] is essentially the ISODATA process applied to the metric in (11), corresponding to case *c* above.

$$x \in \hat{X}_i \leftrightarrow d(x, \bar{x}_i) = \min_{1 \leqslant j \leqslant k} d(x, \bar{x}_j);$$

step 4) If $\hat{P} = P$, stop. Otherwise put $P = \hat{P}$ and go to step 2.

We note that step 3 is ambiguous if $d(x, \bar{x}_j)$ does not have a proper minimum over $1 \leqslant j \leqslant k$. Furthermore, it is possible (although not likely) that one or more of the sets $\hat{X}_i$ may be empty at the conclusion of step 3, in which event difficulties will arise after the loop back to step 2. For these reasons, working versions of ISODATA must incorporate "tie-breaking" rules which assign "centroids" to empty subsets X_i in step 2, and resolve ambiguities in step 3 (a common procedure is to take i to be the smallest index satisfying $d(x, \bar{x}_i) = \min_{1 \leqslant j \leqslant k} d(x, \bar{x}_j)$). Regardless of which tie-breaker rule is used, it can be shown that the sequence of partitions $P^{(n)}$ and the corresponding sequence of centroids $\{\bar{x}^{(n)}\} = \{\bar{x}_1^{(n)}, \ldots, \bar{x}_h^{(n)}\}$ has the descent property relatively to J, i.e., $J(P^{(n+1)}, \bar{x}^{(n+1)}) \leqslant J(P^{(n)}, \bar{x}^{(n)})$, for all n (see Theorem 2).

A partition $P \in \mathcal{P}(k)$ is called a fixed point of ISODATA if and only if the derived partition $\hat{P}$ in step 3 is identical to P. In the next section, we explore the relationship between extremal points for the k-means LSA problem and the fixed points of ISODATA. Here we observe that every partition P consisting of CWS clusters is necessarily a fixed point of ISODATA, in view of Eq. (8). As we have noted elsewhere, these considerations motivate Definition 2 and provide explicit links between the k-means LSA problem, the ISODATA process, and the general notion of separated clusters in X.

Unfortunately, it is not difficult to produce simple examples in R^1 and R^2 where ISODATA has many attracting fixed points which are not solutions of (9) and/or of (13) (see Sect. 6). Experience also indicates that the domains of attraction of these spurious fixed points can become quite large as $\bar{\beta}(k)$ decreases, leading to the so called "cluster-splitting" tendency of ISODATA [2]. In short, as $\bar{\beta}(k)$ decreases, the desired relationship between the maximally separated solution of (9) and the attracting fixed points of ISODATA tends to disintegrate. Furthermore, as noted in the introduction, one cannot tell whether this is actually happening from an inspection of the hard partitions generated by ISODATA. Experience indicates that ISODATA converges to some hard partition, irrespective of the value of $\bar{\beta}(k)$. However, since $\bar{\beta}(k)$ is effectively not computable for non-trivial X, we have no way of knowing the true relationship between this limiting partition and the maximally separated solution of (9). For this reason, ISODATA is an effective technique for identifying CWS clusters in X only if one knows in advance that such clusters are actually present; without such *a priori* knowledge, inferences drawn from ISODATA partitions can be very dangerous.

To avoid this difficulty, we shall base our approach to the detection of CWS clusters upon the following mathematical device. We embed the class $\mathcal{P}(k)$ of hard k-partitions in the larger class $\mathcal{P}_f(k)$ of fuzzy k-partitions. Among the class of all possible extensions of the LSA criterion function J from $\mathcal{P}(k)$ to $\mathcal{P}_f(k)$, we then seek one such extension which meets the following conditions:

a) As $\bar{\beta}(k)$ increases beyond 1 the membership functions of the fuzzy partition which minimizes J over $\mathcal{P}_f(k)$ should closely approximate the characteristic functions of the solution of (9).

b) As $\overline{\beta}(k)$ decreases below 1 the extremizing partitions for J should become increasingly fuzzy in the sense that the values of their membership functions depart significantly from 0 or 1, on certain subsets of X.

c) There should exist a simple and efficient algorithm for computing the extremizing fuzzy partitions for J.

In the following sections we consider two of the infinitely many possible fuzzy extensions of J. The first and most obvious extension leads directly back to a formal derivation of the ISODATA process, and consequently does not advance our purposes. However, the second extension apparently does satisfy conditions *a* through *c* above.

4. Fuzzy Embedding I

In carrying out the embedding described in the previous section, we note that by virtue of condition (2) $\mathcal{P}(k)$ is isomorphically represented by the class of all functions $u(\cdot) = \{u_1(\cdot), \ldots, u_k(\cdot)\} : X \to R^k$ such that

$$\begin{aligned} &1)\ \text{for all } i,\ 1 \leqslant i \leqslant k, \text{ there is some } x \in X \text{ such that } u_i(x) \neq 0,\\ &2)\ \text{for all } i,\ 1 \leqslant i \leqslant k, \text{ and for all } x \in X,\ u_i(x) = 0 \text{ or } 1, \\ &3)\ \text{for all } x \in X,\ \sum_{i=1}^{k} u_i(x) = 1. \end{aligned} \tag{14}$$

The isomophic correspondence is obtained by simply identifying $u_i(\cdot)$ with the characteristic function of the ith subset X_i of a given partition $P \in \mathcal{P}(k)$. From now on, we will therefore say that any function $u(\cdot)$: $X \to R^k$ satisfying (14) is a (hard) k-partition of X and we denote the class of all such functions by the same symbol $\mathcal{P}(k)$ used previously with a set theoretic connotation, i.e.,

$$\mathcal{P}(k) = \{u(\cdot) : X \to R^k \mid u(\cdot) \text{ satisfies } (14)\}. \tag{15}$$

The larger class of all fuzzy k-partitions of X is now defined as follows:

$$\begin{aligned} \mathcal{P}_f(k) &= \{u(\cdot) : X \to U_f \subset R^k\} \\ U_f &= \left\{ u \in R^k \mid 0 \le u_i \le 1,\ 1 \le i \le k;\ \sum_{i=1}^{k} u_i = 1 \right\}. \end{aligned} \tag{16}$$

The component functions $u_i(\cdot)$ of a fuzzy partition $u(\cdot)$ are called the membership functions of the partition (see [9] for the origins of fuzzy set theory).

Evidently, $\mathcal{P}_f(k) \supset \mathcal{P}(k)$. In fact, since condition 1) of (14) is not carried out in (16), $\mathcal{P}_f(k)$ includes degenerate k-partitions containing one or more empty subsets X_i; for technical reasons, this is a convenience in the development of Theorems 1 and 2 to follow. If $\overline{\mathcal{P}}(k)$ denotes $P(k)$ + all degenerate hard k-partitions (condition 1 omitted), then $\mathcal{P}_f(k)$ is simply the convex hull of $\overline{\mathcal{P}}(k)$ in the linear space of functions $u(\cdot)$: $X \to R^k$. We now obtain our first extension of the criterion function J from $\mathcal{P}(k)$ to $\mathcal{P}_f(k)$.

Let $u_i(\cdot)$ denote the characteristic functions of $X_i \in P$. Then in view of Eq. (12), we have

$$\begin{aligned} J(P, v) &= J_1(u(\cdot), v) = \sum_{i=1}^{k} \sum_{x \in X} u_i(x) \|x - v_i\|^2 \\ &= \sum_{i=1}^{k} \sum_{x \in X} u_i(x) \langle x - v_i \mid x - v_i \rangle . \end{aligned} \tag{17}$$

If $u(\cdot)$ is now permitted to range over all of $\mathcal{P}_f(k) \supset \mathcal{P}(k)$, formula (17) defines a continuous and differentiable extension J_1 of J from $\mathcal{P}(k) \times \mathcal{L}^k$ to $\mathcal{P}_f(k) \times \mathcal{L}^k$. Corresponding to this extension we have the following problem.

Relaxed k-Means Least Square Approximation Problem I

Find $u'(\cdot) \in \mathcal{P}_f(k)$ and $v' \in \mathcal{L}^k$ such that

$$J_1(u'(\cdot), v') = \min_{u(\cdot) \in \mathcal{P}_f(k)} \min_{v \in \mathcal{L}^k} J_1(u(\cdot), v). \tag{18}$$

The optimal solutions of this problem are characterized in the following theorem.

Theorem 1. If $u'(\cdot) \in \mathcal{P}_f(k)$ and $v' \in \mathcal{L}^k$ solve (18), then the following conditions must hold:

a. For each fixed $x \in X$, let

$$\xi = \min_{1 \leqslant j \leqslant k} \|x - v_j'\|$$

$$I = \{1 \leq i \leq k \mid \|x - v_i'\| = \xi\}$$

$$I^c = \{1 \leq i \leq k \mid \|x - v_i'\| > \xi\} = \text{complement of } I \text{ in } 1 \leq i \leq k.$$

Then
$$i \in I^c \Rightarrow u_i'(x) = 0$$

and
$$\sum_{i \in I} u_i'(x) = 1.$$

In particular, if $\|x - v_j'\|$ has a proper minimum over $1 \leqslant j \leqslant k$, then I consists of a single integer i and we must have

$$u_i'(x) = \begin{cases} 0 & i \neq \mathrm{i} \\ 1 & i = \mathrm{i} \end{cases}.$$

b. For all i, $1 \leqslant i \leqslant k$, there is some $x \in X$ such that $u_i'(x) \neq 0$, i.e., the optimal k-partition $u'(\cdot) \in \mathcal{P}_f(k)$ always consists of k non-empty fuzzy sets.

c.
$$v_i' = \frac{\sum_{x \in X} u_i'(x) \cdot x}{\sum_{x \in X} u_i'(x)}.$$

Proof.

a) If $u'(\cdot)$ and v' satisfy (18) then in particular,

$$J_1(u'(\cdot), v') \leq J_1(u(\cdot), v') \tag{19}$$

for all $u(\cdot) \in \mathcal{P}_f(k)$. In view of (16) and (17), condition (19) holds if and only if

$$\sum_{i=1}^{k} u_i'(x) \|x - v_i'\|^2 \leq \sum_{i=1}^{k} u_i(x) \|x - v_i'\|^2 \tag{20}$$

for all $x \in X$, or equivalently.

$$\sum_{i=1}^{k} u_i'(x)\,||x - v_i'||^2 = \min_{w \in U_f} \sum_{i=1}^{k} w_i\,||x - v_i'||^2 \tag{21}$$

for all $x \in X$. With reference to (16) we have

$$w \in U_f \Rightarrow \sum_{i=1}^{k} w_i\,||x - v_i'||^2 = \left(\sum_{i \in I} w_i\right)\xi^2 + \sum_{i \in I^c} w_i\,||x - v_i'||^2 \geq \left(\sum_{i=1}^{k} w_i\right)\xi^2 = \xi^2 \tag{22}$$

where the strict inequality holds if and only if $w_i > 0$ for some $i \in I^c$. It follows from (22) that $u(\cdot)$ satisfies (20) if and only if $u_i'(x) = 0$, for $i \in I^c$, and $\sum_{i \in I} u_i'(x) = 1$.

b) Suppose that

$$u_l'(x) = 0 \quad \text{all } x \in X\,. \tag{23}$$

Since there are k vectors v_i' and since $1 \leqslant k \leqslant N$ = number of elements in X, there is at least one element of X, say $\hat{x}$, such that

$$||\hat{x} - v_i'|| > 0 \quad \text{all } i \neq l\,. \tag{24}$$

Put $v'' = (v_1'', \ldots, v_k'')$, with

$$v_i'' = \begin{cases} v_i' & i \neq l \\ \hat{x} & i = l \end{cases}. \tag{25}$$

We have $v'' \in \mathfrak{L}^k$ and, in view of (17), (23), and (25)

$$J_1(u'(\cdot), v'') = J_1(u'(\cdot), v')$$

Thus $(u'(\cdot), v'')$ is also an optimal solution of (18) and consequently must satisfy condition a of this theorem. since $\|\hat{x} - v_j''\|$ has a proper minimum over $1 \leqslant j \leqslant k$ at $j = l$, because of (24) and (25), we must have $u_l'(\hat{x}) = 1$, which contradicts (23).

c) If $(u'(\cdot), v')$ satisfies (18), then, in particular,

$$J_1(u'(\cdot), v') \leq J_1(u'(\cdot), v) \quad \text{all } v \in \mathfrak{L}^k \tag{26}$$

i.e., v' must minimize the quadratic function

$$g(v) \triangleq J_1(u'(x), v) = \sum_{i=1}^{k} \sum_{x \in X} u_i'(x)||x - v_i||^2 = \sum_{i=1}^{k} \sum_{x \in X} u_i'(x)\langle x - v_i \,|\, x - v_i\rangle$$

on the linear space $\mathfrak{L}^k$. Furthermore, since $u_i'(x) \geqslant 0$, g is positive semidefinite. Consequently, v' minimizes g if and only if g is stationary at v', i.e. if and only if the directional derivative $Dg(v', w)$ vanishes for all $w \in \mathfrak{L}^k$. By definition, we have

$$\begin{aligned} Dg(v', w) &\triangleq \left.\frac{d}{dh} g(v' + hw)\right|_{h=0} \\ &= \left.\sum_{i=1}^{k} \sum_{x \in X} u_i'(x) \frac{d}{dh} \langle x - v_i' - hw_i \,|\, x - v_i' - hw_i\rangle\right|_{h=0} \\ &= -2 \sum_{i=1}^{k} \sum_{x \in X} u_i'(x)\,\langle x - v_i' \,|\, w_i\rangle \end{aligned}$$

$$= -2 \sum_{i=1}^{k} \langle \sum_{x \in X} u_i'(x)(x - v_i') \,|\, w_i \rangle \ .$$

Thus v' satisfies (26) if and only if

$$\left\langle \sum_{x \in X} u_i'(x)(x - v_i') \,|\, w_i \right\rangle = 0 \quad \text{for all } w \in \mathfrak{L}^k \ . \tag{27}$$

But (27) holds for arbitrary $w \in \mathfrak{L}^k$ if and only if

$$\sum_{x \in X} u_i'(x)(x - v_i') = 0 \quad 1 \le i \le k$$

i.e. v' satisfies (26) if and only if

$$\left(\sum_{x \in X} u_i'(x) \right) v_i' = \sum_{x \in X} u_i'(x) x \ .$$

Finally, with reference to part b of this theorem, we therefore obtain

$$v_i' = \frac{\sum_{x \in X} u_i'(x) x}{\sum_{x \in X} u_i'(x)} \ . \qquad \text{QED}$$

Corollary 1. Suppose that $(u'(\cdot), v')$ is an optimal solution of the relaxed k-means LSA problem (18) and also satisfies condition (α): for all $x \in X$, $\| x - v_j' \|$ has a proper minimum over $1 \leqslant j \leqslant k$. Then $u'(\cdot)$ is a hard partition (i.e., $u'(\cdot) \in \mathcal{P}(k) \subset \mathcal{P}_f(k)$, v_i' = centroid of the subset $X_i' \subset X$ with characteristic function $u_i'(\cdot)$, and $(u'(\cdot), v')$ is an optimal solution of the original k-means LSA problem (13) over $\mathcal{P}(k) \times \mathfrak{L}^k$.

Corollary 2. The following condition defines a non-empty class $\mathcal{T}_1$ of operators T: $\mathcal{P}_f(k) \to \mathcal{P}_f(k)$.

If $\hat{u}(\cdot)$ denotes $T(u(\cdot))$ = image of $u(\cdot)$ under T, then for some $v \in \mathfrak{L}^k$ satisfying condition (β):

$$\left(\sum_{x \in X} u_i(x) \right) v_i = \sum_{x \in X} u_i(x) x \quad 1 \le i \le k$$

$\hat{u}(\cdot)$ must satisfy

$$\hat{u}_i(x) = 0 \quad ||x - v_i|| \neq \min_{1 \leqslant j \leqslant k} ||x - v_j|| \ .$$

A fuzzy partition $u'(\cdot)$ is part of an optimal solution $(u'(\cdot), v')$ of (18) only if $u'(\cdot)$ is a fixed point of some $T \in \mathcal{T}_1$, i.e.,

$$u'(\cdot) = T(u'(\cdot)) \text{ for some } T \in \mathcal{T}_1 \ . \tag{28}$$

In addition, $u'(\cdot)$ must also satisfy

$$u_i'(\cdot) \neq 0 \quad 1 \le i \le k. \tag{29}$$

Corollary 1 suggests that the characteristic functions of an optimal hard partition for the original k-means LSA problem are likely to differ from the membership functions of an optimal partition for the relaxed problem only on relatively "small" subsets of X, and in many cases, the two solutions will coincide everywhere on X. Furthermore, condition (β) of Corollary 2 suggests that we may be able to approximate optimal solutions of the relaxed problem by iterating some operator in $\mathcal{T}_1$, i.e. by implementing the recursion

$$u^{(m+1)}(\cdot) = T(u)^{(m)}(\cdot) \quad 0 \leq m < \infty \quad T \in \mathcal{T}_1 \tag{30}$$

for some T in $\mathcal{T}_1$ and for some initial guess $u^{\circ}(\cdot)$ in $\mathcal{P}_f(k)$. Indeed, we can now see that the ISODATA process described in Sec. 3 is essentially equivalent to (30). In general, the class $\mathcal{T}_1$ contains infinitely many operators T. However, all of these operators coincide on the subclass of fuzzy partitions $u(\cdot)$ satisfying (29) and condition (α) of Corollary 1. In this subclass, $T(u(\cdot))$ is always a hard partition uniquely prescribed by condition (β) of Corollary 2. Outside this subclass, Theorem 1 does not suggest a unique value for $T(u(\cdot))$,[5] and the tie-breaking rules employed in working versions of ISODATA (cf. Sect. 3) are merely *ad hoc* conventions for keeping $T(u(\cdot))$ in the class $\mathcal{P}$ (k) of hard partitions with k non-empty subsets, and for resolving ambiguities in condition (β). However, regardless of which tie-breaking rule is employed, the following result shows that ISODATA always has the descent property relative to J_1.

Theorem 2. For arbitrary fixed $T \in \mathcal{T}_1$ let $\{u^m(\cdot)\}$ denote the sequence of partitions in $\mathcal{P}_f$ (k) generated by (30), and let $\{v^{(m)}\}$ denote the corresponding sequence in $\mathcal{L}^k$ associated with T and $u^m(\cdot)$ via condition (β). Then for all $m \geqslant 0$,

$$J_1(u^{(m+1)}(\cdot), v^{(m+1)}) \leq J_1(u^{(m)}(\cdot), v^{(m)}) .$$

Proof. As an immediate consequence of the reasoning employed in parts a and c in the proof of Theorem 1, we have

$$J_1(u^{(m+1)}(\cdot), v^m) = \min_{u(\cdot) \in \mathcal{P}_f(k)} J_1(u(\cdot), v^m)$$

and

$$J_1(u^{(m+1)}(\cdot), v^{(m+1)}) = \min_{v \in \mathcal{L}^k} J_1(u^{(m+1)}(\cdot), v).$$

Thus

$$J_1(u^m(\cdot), v^m) \geq J_1(u^{(m+1)}(\cdot), v^m) \geq J_1(u^{(m+1)}(\cdot), v^{(m+1)}). \qquad \text{QED}$$

In conclusion, we observe that the conditions of Theorem 1 are necessary, but in general not sufficient for (global) optimality. Thus, the operators in $\mathcal{T}_1$ may have many non-optimal fixed points and some of these may have substantial zones of attraction for the iterates of (30). This is indeed the case. As noted elsewhere, the ISODATA versions of (30) apparently always converge to some hard partition. However, this limiting partition may vary with the starting guess $u^{\circ}(\cdot)$, and need not be globally optimal. Behavior of this sort is apparently exacerbated by the absence of CWS clusters in X.

5. Fuzzy Embedding II

We now consider a second extension of the criterion function J which apparently does satisfy the condition set forth at the end of Sect. 3.

[5] In analogy with a situation which arises for "singular" extremals in optimal control theory.

Let $u_i(\cdot)$denote the characteristic functions of $X_i \in P$; then in view of Eq. (12) we have

$$J(P, v) = J_2(u(\cdot), v) \triangleq \sum_{i=1}^{k} \sum_{x \in X} u_i^2(x) \|x - v_i\|^2$$

$$= \sum_{i=1}^{k} \sum_{x \in X} u_i^2(x) \langle x - v_i | x - v_i \rangle . \tag{31}$$

If $u(\cdot)$ is once again permitted to range over all of $\mathcal{P}_f(k) \supset \mathcal{P}(k)$, formula (31) defines another continuous and differentiable extension J_2 of J from $\mathcal{P}(k) \times \mathcal{L}^k$ to $\mathcal{P}_f(k) \times \mathcal{L}^k$. Corresponding to this extension we have the following problem.

Relaxed k-Means Least Square Approximation Problem II

Find $u'(\cdot) \in \mathcal{P}_f(k)$ and $v' \in \mathcal{L}^k$ such that

$$J_2(u'(\cdot), v') = \min_{u(\cdot) \in \mathcal{P}_f(k)} \min_{v \in \mathcal{L}^k} J_2(u(\cdot), v) . \tag{32}$$

The optimal solutions of this problem are characterized in the following theorem.

Theorem 3. If $u'(\cdot) \in \mathcal{P}_f(k)$ and $v' \in \mathcal{L}^k$ solve (32), then the following conditions must hold:

a. For each fixed $x \in X$, let

$$I = \{1 \le i \le k | v_i' = x\}$$

$$I^c = \{1 \le i \le k | v_i' \neq x\} = \text{complement of } I \text{ in } 1 \le i \le k .$$

case 1) Suppose $I = \phi$ = empty set. Then

$$u_i'(x) = \frac{1/\|x - v_i'\|^2}{\sum_{j=1}^{k} (1/\|x - v_j\|^2)} \qquad 1 \le i \le k$$

case 2) Suppose $I \neq \phi$. Then

$$i \in I^c \Rightarrow u_i'(x) = 0$$

and

$$\sum_{i \in I} u_i'(x) = 1 .$$

In particular, if I consists of a single integer i, then

$$u_i'(x) = \begin{cases} 1 & i = \mathrm{i} \\ 0 & i \neq \mathrm{i} \end{cases} .$$

b. For all i, $1 \leqslant i \leqslant k$, there is some $x \in X$ such that $u_i'(x) \neq 0$, i.e., the optimal k-partition $u'(\cdot) \in \mathcal{P}_f(k)$ always consists of k non-empty fuzzy sets.

$$v_i' = \frac{\sum_{x \in X} (u_i'(x))^2 x}{\sum_{x \in X} (u_i'(x))^2} .$$

Proof.

a) If $(u'(\cdot), v')$ satisfies (32), then in particular,

$$J_2(u'(\cdot), v') \leq J_2(u(\cdot), v') \tag{33}$$

for all $u(\cdot) \in \mathcal{P}_f(k)$. In view of (16) and (17), $u'(\cdot)$ satisfies (33) if and only if:

$$\sum_{i=1}^{k} (u_i'(x))^2 \|x - v_i'\|^2 \leq \sum_{i=1}^{k} (u_i(x))^2 \|x - v_i\|^2$$

for all $x \in X$, or equivalently,

$$\sum_{i=1}^{k} (u_i'(x))^2 \|x - v_i'\|^2 = \min_{w \in U_f} \sum_{i=1}^{k} w_i^2 \|x - v_i'\|^2 \tag{34}$$

for all $x \in X$.

case 1 ($I = \phi$): Let $\overline{U}_f$ denote the set $\left\{ u \in R^k \,\middle|\, \sum_{i=1}^{k} u_i = 1 \right\} \supset U_f$, obtained by relaxing the inequality constraints in the definition of U_f (Eq. (16)). Consider the corresponding relaxed version of (34), namely

$$\min_{w \in \overline{U}_f} \sum_{i=1}^{k} w_i^2 \|x - v_i'\|^2 . \tag{35}$$

This quadratic minimization problem always has a solution. Furthermore, since no inequality constraints are present, the classical Lagrange multiplier rule can be applied here and yields the following results: $w' \in \overline{U}_f$ is an optimal solution of (35) only if for some real λ, the augmented function $\sum_{i=1}^{k} w_i^2 \|x - v_i'\|^2 + \lambda \sum_{i=1}^{k} w_i$ is stationary at w'. This condition, together with the constraint $\sum_{i=1}^{k} w_i = 1$ and the fact that $I = \phi$, insure that the solution of (35) is unique and is given by

$$w_i' = \frac{1/\|x - v_i'\|^2}{\sum_{j=1}^{k} (1/\|x - v_j'\|^2)} .$$

Evidently $0 < w_i' < 1$ for $1 \leqslant i \leqslant k$. Consequently, the optimal solution of (35) lies in $U_f \subset \overline{U}_f$ after all, and is therefore also the unique solution of (34). It follows that $u'(\cdot)$ satisfies (33) if and only if $u'(x) = w'$ for all $x \in X$.

Case 2 ($I \neq \phi$): If $I \neq \phi$ then $\| x - v_i' \|^2 = 0$ for some i and it readily follows that

$$\min_{w \in U_f} \sum_{i=1}^{k} w_i^2 \|x - v_i'\|^2 = 0 .$$

Suppose $u_j'(x) \neq 0$ for some $j \in I^c$. Then $(u_j'(x))^2 \| x - v_j \|^2 > 0$ and it follows that

$$\sum_{i=1}^{k} (u'(x))^2 \|x - v_i'\|^2 > 0 = \min_{w \in U_f} \sum_{i=1}^{k} w_i^2 \|x - v_i'\|^2$$

which contradicts (34). Thus $u'(\cdot)$ satisfies (33) only if:

$$u_i'(x) = 0 \quad i \in I^c$$

and

$$\sum_{i \in I} u_i'(x) = 1 . \tag{36}$$

Conversely if $u'(x)$ satisfies (36) then

$$\sum_{i=1}^{k} (u_i'(x)|^2 \|x - v_i'\|^2 = 0 = \min_{w \in U_f} \sum_{i=1}^{k} w_i^2 \|x - v_i'\|^2 .$$

b) Suppose that

$$u_l'(x) = 0 \quad \text{all } x \in X . \tag{37}$$

Since there are k vectors, v_i', and since $1 \leqslant k \leqslant N$ = number of elements in X, there is at least one element of X, say $\hat{x}$, such that

$$v_i' \neq \hat{x} \quad i \neq l . \tag{38}$$

Put $v'' = \{ v_1'' , \ldots , v_k'' \}$ with

$$v_i'' = \begin{cases} v_i' & i \neq l \\ \hat{x} & i = l \end{cases} . \tag{39}$$

We have $v'' \in \mathfrak{L}^k$, and in view of (31), (37), and (39)

$$J_2(u'(\cdot), v'') = J_2(u'(\cdot), v') .$$

Thus $(u'(\cdot), v'')$ is also an optimal solution of (32) and consequently must satisfy condition *a* of Theorem 3. With reference to (38) and (39), we have $\hat{x} = v_i'' \Leftrightarrow i = l$, hence $u_l'(\hat{x})$ must equal 1, which contradicts (37).

c) Proof is obtained by replacing J_1 with J_2 and $u_i'(x)$ with $(u_i'(x))^2$ in part *c* of the proof of Theorem 1. QED

Corollary 1. If $(u'(\cdot), v')$ is an optimal solution of (32), then v_i' is a weighted mean of X and consequently falls in $Co(X)$ for $1 \leqslant i \leqslant k$. Furthermore, for at least $n - k$ elements $x \in X$,

$$0 < u_i'(x) < 1 \quad 1 \leq i \leq k . \tag{40}$$

Corollary 2. The following condition defines a non-empty class $\mathcal{T}_2$ of operators T: $\mathcal{P}_f(k) \to \mathcal{P}_f(k)$. Let $\hat{u}(\cdot) = T(u(\cdot))$ = image of $u(\cdot)$ under T. For some $v \in L^k$ satisfying

$$\left(\sum_{x \in X} (u_i(x))^2\right) v_i = \sum_{x \in X} (u_i(x))^2 x \quad 1 \leq i \leq k, \tag{41}$$

Condition (γ): $\hat{u}(\cdot)$ must satisfy the following constraint: for each fixed $x \in X$, let

$$I = \{1 \leq i \leq k \mid v_i = x\}$$

$$I^c = \{1 \leq i \leq k \mid v_i \neq x\} .$$

If $I = \phi$ then

$$\hat{u}_i(x) = \frac{1/\|x - v_i\|^2}{\sum_{j=1}^{k} (1/\|x - v_j\|^2)} .$$

If $I \neq \phi$ then

$$u_i(x) = 0 \quad i \in I_c$$

and

$$\sum_{i \in I} u_i(x) = 1 .$$

A fuzzy partition $u'(\cdot)$ is part of an optimal solution $(u'(\cdot), v')$ of (32) only if $u'(\cdot)$ is a fixed point of some operator $T \in \mathcal{T}_2$, i.e.,

$$u'(\cdot) = T(u'(\cdot)) \quad \text{for some } T \in \mathcal{T}_2 . \tag{42}$$

In addition, $u'(\cdot)$ must also satisfy

$$u_i'(x) \neq 0 \quad 1 \leq i \leq k. \tag{43}$$

As in the previous sections, Corollary 2 of Theorem 3 suggests an iterative approach to the relaxed problem (32); specifically, it suggests that for some $T \in \mathcal{T}_2$ we implement

$$u^{(m+1)}(\cdot) = T(u^{(m)}(\cdot)) \quad 0 \leq m < \infty \quad T \in \mathcal{T}_2 . \tag{44}$$

Once again, the class $\mathcal{T}_2$ contains infinitely many operators. In general however, all of these operators coincide on the subclass of fuzzy partitions $u(\cdot)$ satisfying (43) and condition (δ): For each $x \in X$, I contains at most one element; i.e., $T(u(\cdot))$ is uniquely prescribed by conditions (43), (γ), and (δ). Furthermore, in contrast to the development of the previous section, we also have the following useful result.[6]

Theorem 4. Let $\mathcal{P}_f(k)$ denote the subclass of non-degenerate fuzzy k-partitions satisfying

$$u_i(\cdot) \neq 0 \quad 1 \leq i \leq k . \tag{45}$$

Then for all $T \in \mathcal{T}_2$, T maps $\mathcal{P}_f(k)$ into $\underline{\mathcal{P}}_f(k)$.

Proof. Since there are k vectors v_i and N elements $x \in X$ ($k \leqslant N$), either (i) there is at least one $x \in X$ such that $I = \phi$, or (ii) $k = N$ and for each $x \in X$, I contains precisely one element. In either case, condition (γ) insures that $\hat{u}(\cdot) = T(u(\cdot)) \in \underline{\mathcal{P}}_f(k)$. QED

In view of this result, the first step of (44) always places $u^{(1)}(\cdot)$ in $\underline{\mathcal{P}}_f(k)$ where condition (41) uniquely prescribes the vectors v_i. Thus, the only significant ambiguity in (γ) occurs when $u(\cdot)$ fails to satisfy (δ); in such cases (which appear to be very rare) a supplementary tie-breaking rule is required to make (44) a completely well-posed algorithm. For example if I contains c elements, we might put $u_i(x) = 1/c$ for $i \in I$; or, we might put $u_i(x) = 1$, where i is the smallest integer in I. The first order analysis of Theorem 3 makes no distinction between such rules. However, regardless of which rule is selected, the following result shows that the algorithm (44) always has the descent property relative to J_2.

Theorem 5. For arbitrary fixed $T \in \mathcal{T}_2$, let $\{u^{(m)}(\cdot)\}$ denote the sequence of partitions in $\mathcal{P}_f(k)$ generated by (44) and let $\{v^{(m)}\}$ denote the corresponding sequence in $\mathcal{L}^k$ associated with T and $u^m(\cdot)$ via condition (β). Then for all $m \geqslant 0$,

$$J_2(u^{(m+1)}(\cdot), v^{(m+1)}) \leq J_2(u^{(m)}(\cdot), v^{(m)}) .$$

[6]This result is generally false for operators $T \in \mathcal{T}_1$. The more restrictive result, $T: \mathcal{P}_f(k) \to \underline{\mathcal{P}}_f(k)$ also fails on $\mathcal{T}_1$.

Proof. As an immediate consequence of the reasoning employed in parts a and c in the proof of Theorem 3, we have

$$J_2(u^{(m+1)}(\cdot), v^{(m)}) = \min_{u(\cdot) \in \mathcal{P}_f(k)} J_2(u(\cdot), v^{(m)})$$

and

$$J_2(u^{(m+1)}(\cdot), v^{(m)}) = \min_{v \in \mathcal{L}^k} J_2(u^{(m+1)}(\cdot), v)$$

thus

$$J_2(u^{(m)}(\cdot), v^{(m)}) \geq J_2(u^{(m+1)}(\cdot), v^{(m)}) \geq J_2(u^{(m+1)}(\cdot), v^{(m+1)})$$ QED

Several important questions remain. First, we would like to know whether the process (44) actually does converge to a fixed point of T, i.e., to a partition $u(\cdot)$ satisfying the necessary conditions of Theorem 3. In view of Corollary 1 we can see that such a limiting partition is typically fuzzy, however, when $\overline{\beta}(k) > 1$ we would like to know whether the membership functions of the limiting partition closely approximate the characteristic functions of the subsets X_i' in the maximally separated partition P' solving (9). Finally, when $\overline{\beta}(k)$ decreases below 1 we would like to know whether the limiting partition(s) produced by (44) is always genuinely fuzzy in the sense that $u_i(x)$ differs from 0 or 1 over a substantial subset of X. At present, these questions have not been decided rigorously. However, the numerical experiments presented in the next section suggest that the process (44) does behave the way we would like it to behave.

6. Numerical Results

The following version of (44) was programmed for the IBM 360 digital computer:

Step 1. Choose a partition $u(\cdot)$ in $\underline{\mathcal{P}}_f(k)$ = class of non-degenerate fuzzy partitions.
Step 2. Compute the weighted mean vectors

$$v_i = \frac{\sum_{x \in X} (u_i(x))^2 x}{\sum_{x \in X} u_i^2(x)} \quad 1 \leq i \leq k. \tag{46}$$

Step 3. Construct a new partition $\hat{u}(\cdot) \in \underline{\mathcal{P}}_f(k)$ according to the following rule: for each $x \in X$, Let $I = \left\{ 1 \leqslant i \leqslant k \;\middle|\; v_i = x \right\}$. If I is empty put

$$\hat{u}_i(x) = \frac{1/\langle x - v_i | x - v_i \rangle}{\sum_{j=1}^{k} (1/\langle x - v_j | x - v_j \rangle)} \quad 1 \leq i \leq k.$$

If I is not empty, let i = smallest integer in I and put

$$\hat{u}_i(x) = \begin{cases} 1 & i = \mathrm{i} \\ 0 & i \neq \mathrm{i} \end{cases}.$$

Step 4. Compute new weighted mean vectors $\hat{v}_i$ corresponding to $\hat{u}(\cdot)$ via (46) and compute the corresponding maximum norm defect

$$\delta = \max_{1 \leqslant i \leqslant k} \max_{1 \leqslant j \leqslant d} |v_{i,j} - \hat{v}_{i,j}|$$

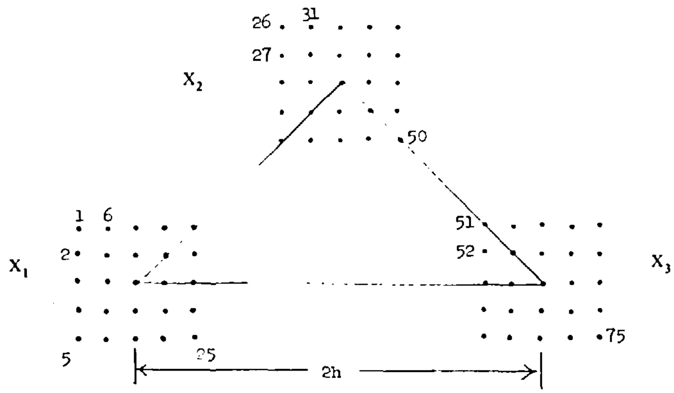

Fig. 1. Three Clusters with centroids on the vertices of an isosceles triangle with altitude h. Clusters are CWS for $h > 8\Delta$, where Δ = bilateral spacing within each lattice.

where d = dimension of $\mathcal{L}(X)$ = linear hull of X and $v_{i,j}$ = jth component of the vector v_i relative to some specified basis for $\mathcal{L}(X)$.

Step 5. If δ is less than some specified threshold ϵ, stop. Otherwise put $v = \hat{v}$ and go to Step 3.

We tried this algorithm on a variety of planar point sets $X \subset R^2$ using the ordinary Euclidean metric as the measure of distance, i.e.,

$$d^2(x_1, x_2) = \langle x_1 - x_2 | x_1 - x_2 \rangle = (x_{1,1} - x_{2,1})^2 + (x_{1,2} - x_{2,2})^2$$

where $x_{i,j}$ = jth entry of the 2-tuple $x_i \in R^2$. In all cases considered, we were able to achieve $\delta < \epsilon^* = 10^{-5}$ by performing a sufficiently large number N^* of iterations. As might be expected, the values of N^* required varied with the structure of X and, to a lesser extent, with the choice of initial partition; these values are quoted for each of the cases reported below, along with descriptions of the limiting fuzzy partition. For purposes of comparison, hard ISODATA partitions were also generated for each example, via the algorithm described in Sect. 3.

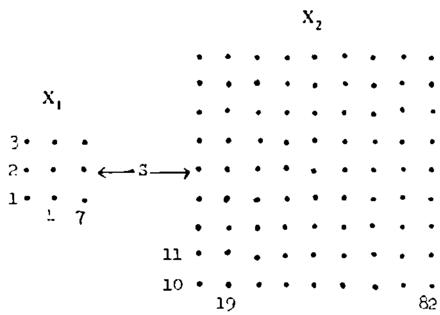

FIG. 2. One large and one small cluster. Clusters are CWS for $s > 8\sqrt{2}\Delta$, where Δ = bilateral spacing within each lattice.

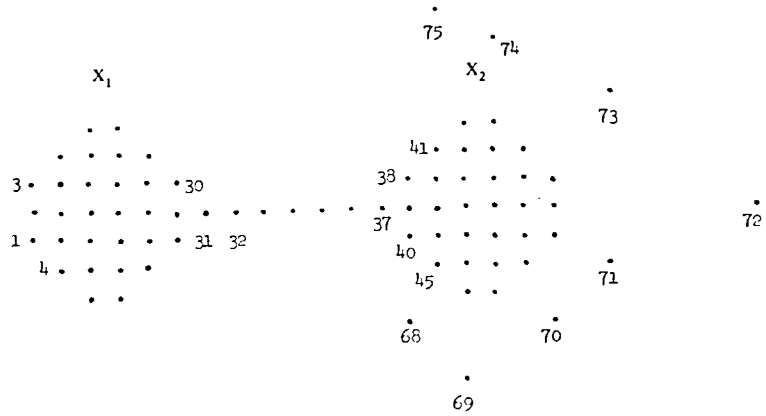

FIG. 3. Two CWS Clusters embedded in a background of strays forming a bridge and a halo.

The point sets considered are shown in Figs. 1, 2, and 3. In Fig. 1, X consists of three identical square lattices X_1, X_2, X_3 each containing 25 points with uniform bilateral internal spacing $\Delta = 1$, with centroids at the vertices of an isosceles triangle with altitude h and base $2h$, and with elements x_i, $1 \leqslant i \leqslant 75$, labeled as indicated. For $h > 5$, the component lattices X_i are "natural" clusters relative to the connectivity criterion discussed in Sect. 2. Furthermore we have $\bar{\alpha}(3) = \bar{\beta}(3) = h/4 - 1$ for $h > 5$. In particular $\{X_1, X_2, X_3\}$ consists of CWS clusters for $h > 8$. The initial partitions employed at the start of all calculations were always hard partitions of the following kind: $P = \{u_1(\cdot), u_2(\cdot), u_3(\cdot)\}$, with

$$
\begin{aligned}
u_1(x_i) &= \begin{cases} 1 & 1 \leq i \leq p \\ 0 & p < i \leq 75 \end{cases} \\
u_2(x_i) &= \begin{cases} 1 & p < i \leq q \\ 0 & 1 \leq i \leq p \text{ or } q < i \leq 75 \end{cases} \\
u_3(x_i) &= \begin{cases} 1 & q < i \leq 75 \\ 0 & 1 \leq i \leq q \end{cases}
\end{aligned}
\tag{47}
$$

where $1 < p < q < 75$.

These partitions are completely characterized by the integer pair (p,q). Runs were made with the fuzzy algorithm and with ISODATA for a variety of different initial partitions (47) and for different values of h (i.e. $\bar{\beta}(3)$). Tables 1–3 give the first two decimal places of the limiting fuzzy partition produced by the fuzzy algorithm for $h = 6, 8$, and 12 ($\bar{\beta}(3) = \frac{1}{2}, 1, 2$) respectively, starting from the initial partition (47) with $(p,q) = (30,55)$; Table 4 gives the corresponding absolute deviations of the components of the limiting weighted mean vector v_i in (46) from the corresponding components of the centroid vector for X_i. The numbers of iterations required to reach the 8-place versions of the limiting partitions displayed in Tables 1–3 were, respectively, $N^* = 14, 10$, and 8. However, the partitions obtained after only $N = 5, 4$, and 3 iterations, respectively, differ from the corresponding limiting partitions only by

TABLE 1. Limiting Fuzzy Partition for the data set in Fig. 1, with $h = 6$ and $k = 3$[a]

i	$u_1(x_i)$	$u_2(x_i)$	i	$u_1(x_i)$	$u_2(x_i)$	i	$u_1(x_i)$	$u_2(x_i)$
1	.88	.09	26	.08	.86	51	.06	.19
2	.93	.05	27	.07	.89	52	.04	.11
3	.95	.04	28	.07	.90	53	.04	.07
4	.94	.04	29	.10	.84	54	.04	.07
5	.91	.06	30	.19	.74	55	.07	.09
6	.90	.07	31	.05	.91	56	.04	.11
7	.96	.03	32	.03	.96	57	.02	.04
8	.98	.01	33	.02	.97	58	.01	.02
9	.97	.02	34	.04	.94	59	.02	.03
10	.93	.04	35	.10	.83	60	.04	.05
11	.90	.07	36	.04	.93	61	.03	.07
12	.98	.02	37	.01	.98	62	.01	.02
13	1.00	.00	38	.00	1.00	63	.00	.00
14	.98	.01	39	.02	.97	64	.01	.01
15	.94	.04	40	.07	.86	65	.03	.04
16	.86	.11	41	.04	.91	66	.03	.07
17	.94	.04	42	.02	.96	67	.01	.03
18	.97	.02	43	.01	.97	68	.01	.01
19	.96	.03	44	.03	.94	69	.01	.02
20	.91	.05	45	.07	.83	70	.03	.04
21	.75	.19	46	.05	.86	71	.04	.09
22	.85	.11	47	.04	.89	72	.02	.05
23	.89	.07	48	.04	.90	73	.02	.04
24	.89	.07	49	.05	.85	74	.02	.04
25	.85	.09	50	.08	.74	75	.04	.06

[a]In Tables 1-3, $u_3(\cdot)$ is not shown, but may be obtained from the relation $\sum_{i=1}^{3} u_i(x) = 1$.

at most one or two units in the second place. Thus, $\epsilon^* = 10^{-5}$ is a conservatively small threshold for the present example.

Runs were also made starting from initial partitions (47) with $(p,q) = (51,61)$ and $(60,70)$; for each fixed value of h, the iterates generated by the fuzzy algorithm led always to the same limiting partition (up to 6 places) obtained for $(p,q) = (30,55)$, with little or no change in N^*. For instance, with $h = 6$, $N^* = 15$ iterations were required to reach the partition displayed in Table 1 from $(p,q) = (51,61)$. As anticipated, the limiting fuzzy partitions become increasingly fuzzy, and N^* becomes increasingly large, as $\bar{\beta}(3)$ decreases.

The behavior exhibited by ISODATA for this example is interesting. As one might expect, ISODATA does very well from the initial partition (47) with $(p,q) = (30,55)$; for all three values of h, just one iteration of ISODATA maps $(p,q) = (30,55)$ into $(p,q) = (25,50)$. However, in marked contrast to the fuzzy algorithm, ISODATA produces some rather surprising and dramatic splittings of the natural clusters X_i when different starting partitions are employed. For example, a simple calculation shows that the partitions (47) with $(p,q) = (50,60)$ and $(50,65)$ are fixed points of the ISODATA algorithm. Furthermore, it turns out that many nearby partitions are either also fixed points or else get mapped quickly into such fixed points. Thus, for $h = 8$, $(p,q) = (60,70)$ maps into $(50,65)$ after four iterations, and $(p,q) = (51,61)$ maps into $(50,61)$ after 3 iterations. Therefore, for this example at least, it appears that the fuzzy algorithm is far superior to ISODATA in the sense that its

TABLE 2. Limiting fuzzy partition for the data set in Fig. 1, with $h = 8$, and $k = 3$

i	$u_1(x_i)$	$u_2(x_i)$	i	$u_1(x_i)$	$u_2(x_i)$	i	$u_1(x_i)$	$u_2(x_i)$
1	.92	.05	26	.05	.91	51	.04	.10
2	.95	.03	27	.04	.93	52	.02	.06
3	.97	.02	28	.04	.94	53	.02	.04
4	.96	.03	29	.05	.92	54	.02	.04
5	.94	.04	30	.10	.85	55	.04	.05
6	.94	.04	31	.03	.94	56	.02	.06
7	.98	.02	32	.02	.97	57	.01	.02
8	.99	.01	33	.01	.98	58	.00	.01
9	.98	.01	34	.02	.96	59	.01	.02
10	.96	.03	35	.05	.91	60	.02	.03
11	.95	.04	36	.02	.95	61	.02	.04
12	.99	.01	37	.01	.99	62	.00	.01
13	1.00	.00	38	.00	1.00	63	.00	.00
14	.99	.01	39	.01	.98	64	.00	.01
15	.96	.02	40	.04	.92	65	.02	.02
16	.92	.06	41	.03	.94	66	.02	.04
17	.97	.02	42	.01	.97	67	.01	.02
18	.99	.01	43	.01	.98	68	.00	.01
19	.98	.02	44	.02	.96	69	.01	.01
20	.95	.03	45	.04	.91	70	.02	.03
21	.87	.10	46	.04	.91	71	.02	.05
22	.92	.06	47	.03	.93	72	.02	.03
23	.94	.04	48	.02	.94	73	.01	.02
24	.94	.04	49	.03	.92	74	.02	.03
25	.91	.05	50	.05	.85	75	.02	.04

TABLE 3. Limiting fuzzy partition for the data set in Fig. 1, with $h = 12$ and $k = 3$

i	$u_1(x_i)$	$u_2(x_i)$	i	$u_1(x_i)$	$u_2(x_i)$	i	$u_1(x_i)$	$u_2(x_i)$
1	.96	.03	26	.03	.95	51	.02	.04
2	.98	.02	27	.02	.97	52	.01	.02
3	.98	.01	28	.02	.97	53	.01	.02
4	.98	.01	29	.02	.96	54	.01	.02
5	.97	.02	30	.04	.94	55	.02	.03
6	.97	.02	31	.02	.97	56	.01	.02
7	.99	.01	32	.01	.99	57	.00	.01
8	1.00	.00	33	.00	.99	58	.00	.00
9	.99	.01	34	.01	.98	59	.00	.01
10	.98	.01	35	.02	.96	60	.01	.02
11	.98	.02	36	.01	.98	61	.01	.02
12	.99	.00	37	.00	.99	62	.00	.00
13	1.00	.00	38	.00	1.00	63	.00	.00
14	1.00	.00	39	.00	.99	64	.00	.00
15	.98	.01	40	.02	.97	65	.01	.01
16	.97	.02	41	.01	.97	66	.01	.02
17	.99	.01	42	.01	.99	67	.00	.01
18	.99	.00	43	.00	.99	68	.00	.00
19	.99	.01	44	.01	.98	69	.00	.01
20	98	.02	45	.02	.96	70	.01	.01
21	.95	.04	46	.02	.95	71	.01	.03
22	.97	.02	47	.01	.97	72	.01	.02
23	.98	.02	48	.01	.97	73	.01	.01
24	.97	.02	49	.02	.96	74	.01	.01
25	.96	.03	50	.03	.94	75	.01	.02

TABLE 4. Absolute deviations of the limiting fuzzy mean vector components from centroid vector components for the data set in Fig. 1

h	$\lvert v_{1,1}-\bar{x}_{1,1}\rvert$	$\lvert v_{1,2}-\bar{x}_{1,2}\rvert$	$\lvert v_{2,1}-\bar{x}_{2,1}\rvert$	$\lvert v_{2,2}-\bar{x}_{2,2}\rvert$	$\lvert v_{3,1}-\bar{x}_{3,1}\rvert$	$\lvert v_{3,2}-\bar{x}_{3,2}\rvert$
6	.032712	.022391	.000111	.035244	.032750	.022422
8	.014845	.010167	.000110	.018010	.014840	.010191
12	.004565	.003089	.000020	.005900	.004570	.003107

convergence to the "correct" limiting partition is not disrupted by a multitude of spurious attracting fixed points corresponding to uninteresting local minimum of J. What seems to be happening here is that hard partitions which provide local minima of the extended payoff J_1 on the continuum $\mathcal{P}_f(3)$ are no longer local minima for the extended payoff J_2; as a consequence, these partitions are not stationary points of J_2, and therefore are not fixed points of the fuzzy algorithm. Furthermore, if spurious fuzzy local minima of J_2 do exist for this data set, their zones of influence were too small to deflect the iterates of the fuzzy algorithm in any of the experiments peformed.

The data set of X in Fig. 2 consists of two square lattices X_1 and X_2 containing 9 and 81 elements respectively, with uniform bilateral internal spacing $\Delta = 1$, and labels as indicated. X_1 and X_2 are natural clusters relative to the connectivity criterion when the distance s between X_1 and X_2 exceeds 1. Furthermore, we have diam $X_2 = 8\sqrt{2}$, and $\bar{\alpha}(2) = \bar{\beta}(2) = s/8\sqrt{2}$; hence, $\{X_1, X_2\}$ consists of CWS clusters when $s > 8\sqrt{2} \approx 11.3$. The starting partitions used for this example were always of the form

$$u_1(x_i) = \begin{cases} 1 & 1 \le i \le p \\ 0 & p < i \le 90 \end{cases} \qquad u_2(x_i) = \begin{cases} 1 & p < i \le 90 \\ 0 & 1 \le i \le p \end{cases} \tag{48}$$

TABLE 5. Limiting fuzzy partition for the data set in Fig. 2, with $s = 5$ and $k = 2$[a]

i	$u_1(x)$	i	$u_1(x)$	i	$u_1(x)$	i	$u_1(x)$	i	$u_1(x)$	i	$u_1(x)$
1	.79	16	.89	31	.63	46	.33	61	.11	76	.06
2	.79	17	.81	32	.65	47	.27	62	.19	77	.04
3	.79	18	.74	33	.63	48	.20	63	.26	78	.04
4	.81	19	.66	34	.60	49	.13	64	.22	79	.09
5	.82	20	.72	35	.57	50	.09	65	.15	80	.15
6	.81	21	.79	36	.55	51	.13	66	.08	81	.20
7	.84	22	.86	37	.43	52	.20	67	.03	82	.20
8	.85	23	.89	38	.41	53	.27	68	.01	83	.16
9	.84	24	.86	39	.38	54	.33	69	.03	84	.12
10	.74	25	.79	40	.34	55	.26	70	.08	85	.09
11	.81	26	.72	41	.32	56	.19	71	.15	86	.08
12	.89	27	.66	42	.34	57	.11	72	.22	87	.09
13	.96	28	.55	43	.38	58	.04	73	.20	88	.12
14	.99	29	.57	44	.41	59	.01	74	.15	89	.16
15	.96	30	.60	45	.43	60	.04	75	.09	90	.20

[a]In Tables 5-8, the value of $u_2(x)$ may be obtained from the relation $\sum_{i=1}^{2} u_i(x) = 1$.

TABLE 6. Limiting fuzzy partition for the data set in Fig. 2, with $s = 7$ and $k = 2$

i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$
1	.95	16	.38	31	.10	46	.12	61	.03	76	.04
2	.95	17	.39	32	.09	47	.08	62	.07	77	.04
3	.95	18	.41	33	.10	48	.04	63	.11	78	.04
4	.97	19	.30	34	.13	49	.01	64	.10	79	.06
5	.97	20	.27	35	.17	50	.00	65	.07	80	.08
6	.97	21	.23	36	.22	51	.01	66	.04	81	.11
7	.98	22	.21	37	.16	52	.04	67	.02	82	.12
8	.99	23	.20	38	.11	53	.08	68	.02	83	.10
9	.98	24	.21	39	.07	54	.12	69	.02	84	.08
10	.41	25	.23	40	.04	55	.11	70	.04	85	.07
11	.39	26	.27	41	.03	56	.07	71	.07	86	.06
12	.38	27	.30	42	.04	57	.03	72	.10	87	.07
13	.36	28	.22	43	.07	58	.01	73	.11	88	.08
14	.36	29	.17	44	.11	59	.00	74	.08	89	.10
15	.36	30	.13	45	.16	60	.01	75	.06	90	.12

with $p = 54$. Runs were made with the fuzzy algorithm and ISODATA for values of s ranging from 5 to 16 ($\bar{\beta}(2)$ ranging from $\approx .44$ to 1.41). Tables 5–8 give the first two decimal places of the limiting fuzzy partitions produced by the fuzzy algorithm for $s = 5, 7, 9$ and 16 ($\bar{\beta}(2) \approx .44, .62, .79$ and 1.4); the numbers of iterations required to obtain these partitions were, respectively, $N^* = 28, 73, 24$, and 12, although in each case substantially fewer iterations would suffice to reach partitions essentially like the limiting partitions. Once again, the limiting fuzzy partitions become increasingly fuzzy, and N^* tends to increase in general ($s = 5$ is the obvious exception) as $\bar{\beta}(2)$ decreases.

Again, the behavior of ISODATA is interesting. When $s > 3$, a simple calculation reveals that the "natural" partition (48) corresponding to $p = 9$ is indeed a fixed point of ISODATA, however the partitions corresponding to $p = 18, 27, 36$, and 45 are also fixed points when $8 \geqslant s > 4$, $9 > s > 3$, $8 \geqslant s > 1$, and $5 \geqslant s > 1$ respectively. Furthermore calculations for $s = 9, 8, 7, 6$, and 5 show that ISODATA quickly maps the initial partition corresponding to $p = 54$ into $p = 27, 36, 36, 36$, and 45 respectively. For $s = 7, 8$, and 9, the

TABLE 7. Limiting fuzzy partition for the data set in Fig. 2, with $s = 9$ and $k = 2$

i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$
1	.98	16	.20	31	.05	46	.08	61	.02	76	.03
2	.98	17	.23	32	.04	47	.05	62	.04	77	.03
3	.98	18	.26	33	.05	48	.02	63	.07	78	.03
4	.99	19	.18	34	.07	49	.01	64	.07	79	.04
5	1.00	20	.15	35	.10	50	.00	65	.05	80	.06
6	.99	21	.12	36	.13	51	.01	66	.03	81	.08
7	.99	22	.10	37	.10	52	.02	67	.02	82	.09
8	1.00	23	.09	38	.06	53	.05	68	.01	83	.07
9	.99	24	.10	39	.04	54	.08	69	.02	84	.06
10	.26	25	.12	40	.02	55	.07	70	.03	85	.05
11	.23	26	.15	41	.01	56	.04	71	.05	86	.05
12	.20	27	.18	42	.02	57	.02	72	.07	87	.05
13	.19	28	.13	43	.04	58	.01	73	.08	88	.06
14	.18	29	.10	44	.06	59	.00	74	.06	89	.07
15	.19	30	.07	45	.10	60	.01	75	.04	90	.09

TABLE 8. Limiting fuzzy partition for the data set in Fig. 2, with $s = 16$ and $k = 2$

i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$
1	.99	16	.07	31	.02	46	.04	61	.01	76	.02
2	1.00	17	.08	32	.01	47	.02	62	.02	77	.02
3	.99	18	.10	33	.02	48	.01	63	.03	78	.02
4	1.00	19	.07	34	.02	49	.00	64	.04	79	.02
5	1.00	20	.05	35	.04	50	.00	65	.02	80	.03
6	1.00	21	.04	36	.05	51	.00	66	.02	81	.04
7	1.00	22	.03	37	.04	52	.01	67	.01	82	.05
8	1.00	23	.03	38	.03	53	.02	68	.01	83	.04
9	1.00	24	.03	39	.01	54	.04	69	.01	84	.03
10	.10	25	.04	40	.01	55	.03	70	.02	85	.03
11	.08	26	.05	41	.00	56	.02	71	.02	86	.03
12	.07	27	.07	42	.01	57	.01	72	.04	87	.03
13	.06	28	.05	43	.01	58	.00	73	.04	88	.03
14	.06	29	.04	44	.03	59	.00	74	.03	89	.04
15	.06	30	.02	45	.04	60	.00	75	.02	90	.05

TABLE 9. Limiting fuzzy partition for the data set in Fig. 3, with $k = 2$[a]

i	$u_1(x_i)$	i	$u_1(x_i)$	i	$u_1(x_i)$
1	.96	26	.99	51	.03
2	.96	27	.97	52	.06
3	.96	28	.96	53	.05
4	.96	29	.97	54	.02
5	.98	30	.96	55	.01
6	.98	31	.91	56	.00
7	.98	32	.82	57	.01
8	.96	33	.70	58	.02
9	.95	34	.55	59	.05
10	.97	35	.39	60	.03
11	.99	36	.25	61	.01
12	.99	37	.14	62	.01
13	.99	38	.07	63	.01
14	.97	39	.06	64	.03
15	.95	40	.07	65	.03
16	.95	41	.05	66	.02
17	.98	42	.03	67	.03
18	.99	43	.02	68	.16
19	1.00	44	.03	69	.17
20	.99	45	.05	70	.08
21	.98	46	.06	71	.07
22	.95	47	.03	72	.15
23	.97	48	.01	73	.10
24	.99	49	.00	74	.15
25	.99	50	.01	75	.23

[a] $u_2(\cdot)$ may be obtained from the relation $\sum_{i=1}^{2} u_i(x) = 1$.

fuzzy algorithm generates substantially better approximations to the "natural" partition (48) with $p = 9$; for $s = 16$, ISODATA converges to the natural partition, and the fuzzy algorithm also generates a very good approximation to this partition.

Finally, the data set X in Fig. 3 contains two octagonal lattices X_1 and X_2, each comprised of 30 elements with uniform bilateral internal spacing $\Delta = 1$. The remaining elements of X form a six element bridge joining X_1 and X_2 (points 31–36) and a halo of eight elements about X_2 (points 68–75).[7] This data set differs from the previous examples in a significant way: Although $\bar{\beta}(2) << 1$ for X, we have $X = Y \cup Z$ where $\bar{\beta}(2) > 1$ for Y, and where the "average density" of points in Z is much less than in Y (e.g., take Y = union of X_1 and X_2, and Z as the remaining bridge and halo elements.) In this sense X differs negligibly from a set consisting of two CWS clusters. Although we have not attempted a precise quantitative characterization of this kind of structure, it is clearly important in applications, especially where X is a sample drawn from a mixture of unimodal probability distributions. Zahn [4] and Wishart [8] describe pruning techniques for dealing with precisely this problem of extracting dense nuclear clusters from a noisy background. The general idea behind their methods is to first isolate and remove the "negligible" subset of X and then submit the remainder set to an algorithm capable to identifying clusters relative to the metric of interest. In the present investigation, we were curious to see how the fuzzy algorithm would behave on the unaltered set X of Fig. 3, for $k = 2$. Accordingly, calculations were made for initial partitions of the form

$$
\begin{aligned}
u_1(x_i) &= \begin{cases} 1 & 1 \le i \le p \\ 0 & p < i \le 75 \end{cases} \\
u_2(x_i) &= \begin{cases} 1 & p < i \le 75 \\ 0 & 1 \le i \le p \end{cases}
\end{aligned}
\tag{49}
$$

In all cases considered, the fuzzy algorithm converged rapidly to the same limiting partition, which is given to 2 decimal places in Table 9 (the number of iterations required being $N^* = 10$). Since $\bar{\beta}(2) << 1$ for X, we expect fuzziness in this partition. However, it can be seen that divided membership is most pronounced on the bridge and halo elements. For x in the lattice X_1, $u_1(\cdot)$ is essentially equal to the characteristic function of X_1; similarly, for x in X_2, $u_2(\cdot)$ closely approximates the characteristic function of X_2. Thus, the limiting partition generated by the fuzzy algorithm provides insight into the structure of X. In contrast, ISODATA converges rapidly ($N^* \approx 3$) to the limiting hard partition (49) with $p = 34$ for all initial partitions considered; while this limiting partition does separate the nuclear clusters X_1 and X_2, it gives no clue to the structure of X.

References

[1] P. H. A. Sneath, "The application of computers to taxonomy," *J. Gen. Microbiology*, vol. 17, p. 201–226, 1957.

[2] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* New York: Academic Press, 1972.

[3] G. Nagy, "State of the art in pattern recognition," *IEEE Proc.*, vol. 56, no. 5, pp. 836–862, 1968.

[4] C. T. Zahn, "Graph theoretical methods for detecting and describing gestalt clusters," *IEEE Trans. on Comp.*, vol. C-20, no. 1, pp. 68–86, 1971.

[5] R. G. Casey and G. Nagy, "An autonomous reading machine," *IEEE Trans. on Comp.*, vol. C-17, no. 5, pp. 492–503, 1968.

[6] R. G. Casey and G. Nagy, "Recognition of printed chinese characters," *IEEE Trans. on Comp.*, vol. C-15, pp. 91–101, 1966.

[7] Figure 3 is drawn to scale.

[7] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika,* vol. 32, pp. 241–254, 1967.
[8] D. Wishart, "Mode analysis: A generalization of nearest neighbor which reduces chaining effects," in *Numerical Taxonomy*, A. J. Cole, Ed. New York: Academic Press, 1969.
[9] L. Zadeh, "Fuzzy sets," *Inform. and Control,* vol. 8, no. 3, pp. 338–353, 1965.
[10] E. H. Ruspini, "A new approach to clustering," *Inform. and Control,* vol. 15, pp. 22–32, 1969.
[11] E. H. Ruspini, "Numerical methods for fuzzy clustering," *Inf. Science,* vol. 2, p. 319, 1970.
[12] I. Gitman and M. D. Levine, "An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique," *IEEE Trans. on Comp.*, vol. C-19, no. 7, pp. 583–593, 1970.
[13] G. H. Ball and D. J. Hall, "ISODATA, an iterative method of multivariate analysis and pattern classification," *Behavioral Science,* vol. 12, pp. 153–155, 1967.
[14] J. Dieudonne, *Foundations of Modern Analysis.* New York: Academic Press, 1960.
[15] C. W. Curtis, *Linear Algebra.* Boston: Allyn & Bacon, 1963.
[16] H. P. Friedman and J. Rubin, "On some invariant criteria for grouping data," *Amer. Stat. Assoc. J.*, vol. 62, no. 320, pp. 1159–1178, 1967.
[17] K. Fukunaga and W. L. G. Koontz, "A criterion and an algorithm for grouping data," *IEEE Trans. on Comp.*, vol. C-19, no. 10, pp. 917–923, 1970.

Received September 17, 1973