Generalized Fuzzy c-Means Clustering Strategies Using L_p Norm Distances

Richard J. Hathaway, *Member, IEEE*, James C. Bezdek, *Fellow, IEEE*, and Yingkang Hu

Abstract—Fuzzy c-means (FCM) is a useful clustering technique. Recent modifications of FCM using L_1 norm distances increase robustness to outliers. Object and relational data versions of FCM clustering are defined for the more general case where the L_p norm ($p \geq 1$) or semi-norm ($0 < p < 1$) is used as the measure of dissimilarity. We give simple (though computationally intensive) alternating optimization schemes for all object data cases of $p > 0$ in order to facilitate the empirical examination of the object data models. Both object and relational approaches are included in a numerical study.

Index Terms—Clustering, fuzzy c-means, L_p norm, outlier.

I. INTRODUCTION

THE fuzzy c-means (FCM) algorithm [1] has successfully been applied to a wide variety of clustering problems [2]. This approach partitions a set of object data $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \mathcal{R}^s$ into c-(fuzzy) clusters based on a computed minimizer of the fuzzy within-group least squares functional

$$J_m(U, v) = \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|_2^2 \tag{1}$$

where

$m > 1$	fuzzification parameter;
$\mathbf{v}_i \in R^s$	prototype (or mean) of the ith cluster;
$U_{ik} \in [0, 1]$	degree to which datum $\mathbf{x}_k$ belongs to the ith cluster;
$v = [v_{ji}] = [\mathbf{v}_1, \ldots, \mathbf{v}_c] \in \mathcal{R}^{s \times c}$	matrix of cluster prototypes;
$U = [U_{ik}]$	partition matrix;
$\| * \|_2^2$	Euclidean or 2-norm squared.

For later notational convenience, we will array the object data $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ as columns in the object data matrix $X = [x_{jk}] = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathcal{R}^{s \times n}$. The partition matrix U is a convenient tool for representing cluster structure in the data $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$;

we define the set of all *nondegenerate* fuzzy $c \times n$ partition matrices for partitioning n data into c clusters as

$$M_{fcn} = \left\{ U \in \mathcal{R}^{c \times n} \,|\, \forall i, k : 0 \leq U_{ik} \leq 1; \right.$$
$$\left. \sum_{i=1}^{c} U_{ik} = 1; \; 0 < \sum_{k=1}^{n} U_{ik} \right\}. \tag{2}$$

The most popular and effective method of optimizing (1) is the fuzzy c-means algorithm, which alternates between optimizations of $\tilde{J}_m(U \,|\, v^*)$ over U with v^* fixed and $\hat{J}_m(v \,|\, U^*)$ over v with U^* fixed, producing a sequence $\{(U^{(r)}, v^{(r)})\}$. Specifically, the $r + 1$st value of $v = [\mathbf{v}_1, \ldots, \mathbf{v}_c]$ is computed using the rth value of U in the right-hand side of

$$\mathbf{v}_i = \left(\sum_{k=1}^{n} U_{ik}^m \mathbf{x}_k \right) \Big/ \left(\sum_{k=1}^{n} U_{ik}^m \right), \quad \text{for } i = 1, \ldots, c. \tag{3}$$

Then the updated $r+1$st value of v is used to calculate the $r+1$st value of U via

$$U_{ik} = \left(d_{ik}^{-1/(m-1)} \right) \Big/ \sum_{j=1}^{c} \left(d_{jk}^{-1/(m-1)} \right), \quad \text{where} \tag{4}$$

$$d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|_2^2 > 0$$
$$\text{for } i = 1, \ldots, c \quad \text{and} \quad k = 1, \ldots, n. \tag{5}$$

The FCM iteration is initialized using some $U \in M_{fcn}$ (or possibly $v \in \mathcal{R}^{s \times c}$) and continues by alternating the updates in (3) and (4) until the difference measured in any norm on $\mathcal{R}^{c \times n}$ (or $\mathcal{R}^{s \times c}$) in successive partition matrices (or v matrices) is less than some prescribed tolerance ε.

While FCM has proven itself to be very useful, the quality of the computed cluster centers $\{\mathbf{v}_1, \ldots, \mathbf{v}_c\}$ can sometimes be degraded due to the effects of outliers in the data set. This occurs because $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|_2^2 = \sum_{j=1}^{s} (x_{jk} - v_{ji})^2$, the datum-to-prototype dissimilarity term in (1), can place considerable weight on outlying data points, thus pulling cluster prototypes away from the "center" or main distribution of the (nonoutlying) cluster.

There are a number of useful approaches for controlling the harmful effects of outlying data, including the possibilistic clustering approach of Krishnapuram and Keller [3] and the fuzzy noise-clustering approach of Dave [4]. Most important to this note is the work of Kersten [5]–[7] and Miyamoto and Agusta [8], who independently suggested replacing $\|\mathbf{x}_k - \mathbf{v}_i\|_2^2$ with $\|\mathbf{x}_k - \mathbf{v}_i\|_1^1 = \sum_{j=1}^{s} |x_{jk} - v_{ji}|$ in the FCM functional in order
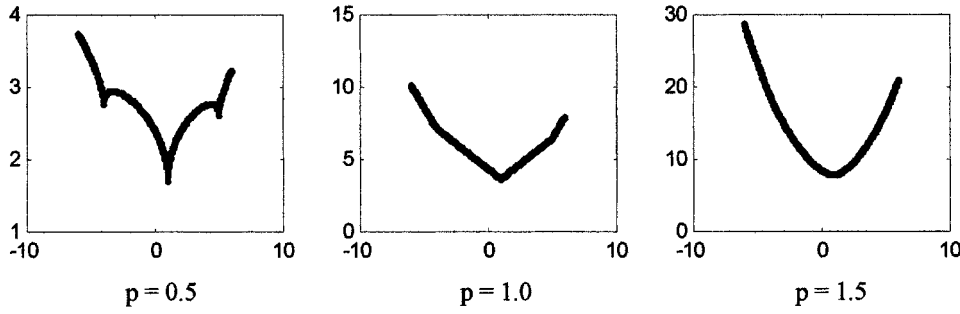
Fig. 1. An illustration of the possible shapes of $f_{ji}(v_{ji}) = \sum_{k=1}^{n} U_{ik}^{m} |x_{jk} - v_{ji}|^{p}$.

to increase robustness against outlying data. Earlier work that uses the L_1 norm or its square in FCM-based clustering appears in Bobrowski and Bezdek [9] and Jajuga [10]. Bobrowski and Bezdek [9] also gave a method for optimizing J_m when the square of the sup norm (L_∞) is used in (1).

In this note, we examine FCM-based clustering using general L_p norm distances, where the L_p norm of the s-dimensional real vector $\mathbf{z}$ is defined as $\|\mathbf{z}\|_p = (\sum_{j=1}^{s} |z_j|^p)^{1/p}$, $p > 1$. In Section II, we present an object-data strategy for using L_p norms due to Miyamoto and Agusta [11], [12] and Overstreet [13]. Additionally, we describe how a relational data approach can be taken using the non-Euclidean fuzzy c-means (NERFCM) algorithm of Hathaway and Bezdek [14]. The object-data approach operates directly and solely, on the object data matrix X. The relational approach clusters $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ indirectly through the use of derived dissimilarity data $R = [R_{jk}]$, where R_{jk} is some measure of the dissimilarity between $\mathbf{x}_j$ and $\mathbf{x}_k$. The two strategies will be compared using numerical examples in Section III. The final section summarizes our findings.

II. L_p EXTENSIONS OF FUZZY c-MEANS

A. Object-Data Strategy

This approach is based on a direct modification of the fuzzy c-means functional $J_m(U, v)$. The generalization of (1) that we consider here was originally proposed by Miyamoto and Agusta [11] and later extended by Overstreet [13]. The objective function is

$$
\begin{aligned}
J_{m,p}(U, v) &= \sum_{i=1}^{c} \sum_{k=1}^{n} U_{ik}^{m} \|\mathbf{x}_k - \mathbf{v}_i\|_p^p \\
&= \sum_{i=1}^{c} \sum_{k=1}^{n} \sum_{j=1}^{s} U_{ik}^{m} |x_{jk} - v_{ji}|^p, \quad m > 1. \quad (6)
\end{aligned}
$$

The optimization of (6) is relatively straightforward and the choice of p has considerable effect on the influence of outliers and other properties of the representation of the clusters. We see that $J_{m,2}$ is the original FCM functional and $J_{m,1}$ is the more robust functional used by Kersten [5]–[7]. Miyamoto and Agusta [11], [12] consider this model in general for $1 \le p < \infty$ and the range of p is extended to include $0 < p < 1$ in Overstreet [13]. The object-data approach considered in this correspondence is based on iterative minimization of $J_{m,p}$.

The optimization of (6) for general p is more complicated and costly than the optimization of the special case $p = 2$ (FCM).

However, as with FCM, optimization can be done by alternating separate optimizations over the U and v variables. The (U-variable) minimizer of $J_{m,p}(U, v)$ over M_{fcn} (for a fixed v) is given by (4) using the datum-to-prototype dissimilarities

$$
d_{ik} = \sum_{j=1}^{s} |x_{jk} - v_{ji}|^p
$$
$$
\text{for } i = 1, \ldots, c \quad \text{and} \quad k = 1, \ldots, n. \quad (7)
$$

Appropriate methods of computing the v-variable minimizer of $J_{m,p}(U, v)$ over $\mathcal{R}^{s \times c}$ (for a fixed U) depend on the value of p, but in all cases this optimization can be decoupled into $c \times s$ *independent* univariate minimizations of functions of the form

$$
f_{ji}(v_{ji}) = \sum_{k=1}^{n} U_{ik}^{m} |x_{jk} - v_{ji}|^p,
$$
$$
\text{for } j = 1, \ldots, s \quad \text{and} \quad i = 1, \ldots, c. \quad (8)
$$

The geometric form of f_{ji} is nonconvex for $0 < p < 1$, with a cusp at each datum value x_{jk}. For $p = 1$, f_{ji} is convex and piecewise linear, with a corner at each x_{jk}. The function f_{ji} is differentiable and strictly convex if $1 < p < \infty$. The three types of shapes are illustrated in Fig. 1 using the function $f_{ji}(v_{ji}) = 0.4 * |-4 - v_{ji}|^p + 0.7 * |1 - v_{ji}|^p + 0.4 * |5 - v_{ji}|^p$.

We choose to minimize f_{ji} in (8) in the simplest possible way since our emphasis here is on understanding the properties of the clusterings and not on computational efficiency. For $p \le 1$, we note that f_{ji} takes its minimum value over $\mathcal{R}$ for one (or more) elements in the set $\{x_{j1}, \ldots, x_{jn}\}$. For $p < 1$, the minimizing value of v_{ji} is simply taken to be the smallest of the x_{jk} values, which globally minimize f_{ji} over $\{x_{j1}, \ldots, x_{jn}\}$. The mean of the smallest and largest globally minimizing x_{jk} values is used for the special case of $p = 1$. For $p > 1$, the computed value of v_{ji} is taken to be a numerical approximation to the unique zero of $f'_{ji} = -p \sum_{k=1}^{n} U_{ik}^{m} |x_{jk} - v_{ji}|^{p-1} \text{sign}(x_{jk} - v_{ji})$, obtained here using the method of bisection.

We summarize the object-data strategy. It consists of alternating optimizations of $J_{m,p}$ in (6) between the U and v variables. The optimization over the U variable is accomplished using (7) in (4) and the optimization over the v variable is decoupled into $c \times s$ univariate optimizations of functions of the form (8). The univariate optimizations are essentially done using exhaustive search over $\{x_{j1}, \ldots, x_{jn}\}$ for $p \le 1$ and bisection on f'_{ji} for $p > 1$. (Exhaustive search, which is necessary for $p < 1$, is prohibitively expensive for sufficiently large data sets and we, therefore, acknowledge a practical limitation to the usefulness

of this model in some cases. We repeat that the emphasis here is on studying the clustering solutions produced by the various models.) The alternating optimization is continued until successive U partitions are within $\varepsilon = 0.00001$ of each other, as measured by the sup norm of $(U^{\text{new}} - U^{\text{old}})$. It is important to understand that the independence of the components of v (for fixed U) in (6) allows the minimizing v to be calculated using only univariate optimizations; and the result so obtained does optimize $\hat{J}_{m,p}(v \mid U^*)$ over v for U^*.

B. Relational-Data Strategy

The relational-data strategy uses the non-Euclidean relational fuzzy c-means (NERFCM) in [14]. In essence, NERFCM is a safeguarded version of RFCM, which is the relational dual of fuzzy c-means. RFCM produces the FCM clustering of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ indirectly using relational data $R = [R_{jk}] = [\|\mathbf{x}_j - \mathbf{x}_k\|_2^2]$. Given some matrix of relational data R, the NERFCM algorithm iteratively generates a sequence of partition matrices $\{U^{(r)}\}$ according to the following steps. The current U matrix is used to calculate vectors $\mathbf{w}_1, \ldots, \mathbf{w}_c$ according to

$$\mathbf{w}_i = (U_{il}^m, \ldots, U_{in}^m) \bigg/ \sum_{k=1}^{n} U_{ik}^m. \qquad (9)$$

These vectors are then used to calculate new dissimilarities d_{ik} according to

$$d_{ik} = (R\mathbf{w}_i)_k - 0.5(\mathbf{w}_i)^T R(\mathbf{w}_i)$$
$$\text{for } k = 1, \ldots, n \quad \text{and} \quad i = 1, \ldots, c. \qquad (10)$$

If necessary, the dissimilarities are altered to guarantee positivity ([14]) and then they are used in (4) to generate the new U iterate. The iteration is continued until the sup-norm difference in successive U matrices is sufficiently small.

The relational data approach for an L_p extension of FCM consists of applying NERFCM with the L_p-based relational data $R = [R_{jk}] = [\|\mathbf{x}_j - \mathbf{x}_k\|_p^p] = [\sum_{i=1}^{s} |x_{ij} - x_{ik}|^p]$. This approach will produce a terminal partition matrix U^*, which attempts to represent the cluster memberships for $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, but it does not directly provide cluster prototypes $v = [\mathbf{v}_1, \ldots, \mathbf{v}_c]$. We recover meaningful prototypes by using the terminal partition U^* with $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and solving

$$v^* = \arg\min_{v} J_{m,p}(U^*, v). \qquad (11)$$

We remark that the duality theory (in [14] and [15]) that guarantees that the same solution is found using the object and relational versions of FCM holds only for $p = 2$. At other values for p, it is possible that the object data version and its relational derivative yield different (U, v) pairs for the same choices of common algorithmic parameters. One purpose for trying a relational approach for $p \neq 2$ is to discover any general similarities between the object and relational approaches that extend past the limited duality theory. Theoretical convergence of the object data algorithms for $1 \leq p < \infty$ is shown in [12]; existing convergence theory ([14]) for the relational approach only covers the case of $p = 2$.
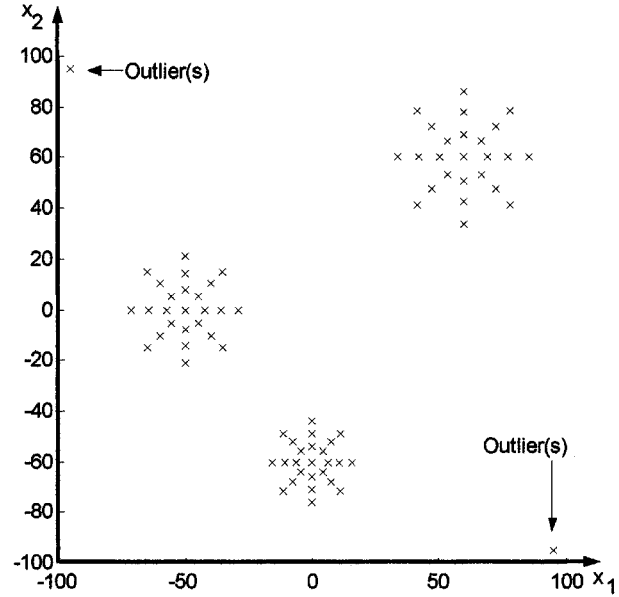

Fig. 2. "Two cluster data" scatterplot and initial prototypes, $c = 2$.

III. NUMERICAL EXPERIMENTS

In all experiments we chose $m = 2$ and stopped iteration as soon as the absolute value of differences of *all* pairs of elements (i.e., $\|U^{\text{new}} - U^{\text{old}}\|_\infty$) in a successive pair of U matrices differed by less than 0.00001. The first experiment uses the (75+ point) data set in Fig. 2, which consists of three (25 point) radial clusters centered at $\mu_1 = (-50, 0)^T$, $\mu_2 = (0, -60)^T$ and $\mu_3 = (60, 60)^T$ and a varying number of outliers located at the indicated positions. The purpose of this experiment is to investigate sensitivity to outliers of the object and relational approaches for various values of p. The leftmost column of the table gives the number of outliers included with the three clusters in the sample. The number of outliers is even and the outliers are evenly divided between the two positions shown in Fig. 2. All iteration sequences are initialized using a hard partition that correctly partitions the three clusters and groups each outlier with its nearest cluster. For a computed set of *terminal prototype vectors* v_1, v_2 and v_3, we measure the sensitivity to the outliers as the Frobenius norm distance between the true centers and terminal prototype vectors: $\|[\mu_1 \ \mu_2 \ \mu_3] - [v_1 \ v_2 \ v_3]\|_F$, where $\|A\|_F$ is defined by $\|A\|_F = (\sum_{j,k} A_{j,k}^2)^{1/2}$.

Note the effects of increasing numbers of outliers as we move down the rows of Table I. We see no deterioration in the quality of the terminal prototype vectors for as many as 24 outliers for both the object and relational approaches with $p = 0.5$ and $p = 1$. The last few rows indicate that this resistance to outliers is actually slightly greater for $p = 1$ than for $p = 0.5$. For any value of $p > 1$, the deviation of the computed prototypes from the true cluster centers steadily increases with the number of outliers. The object and relational data results are quite comparable for $p > 1$ and, as predicted by duality theory, they produce identical results for $p = 2$. For $p \leq 1$, the object and relational prototypes eventually vary from the true centers for sufficiently large numbers of outliers, but fewer outliers are required to cause substantial deviation for the relative data approach. Based on the results of this experiment, the object data

TABLE I
DEVIATION OF COMPUTED FROM TRUE CLUSTER CENTERS: $\|[\mu_1 \ \mu_2 \ \mu_3] - [v_1 \ v_2 \ v_3]\|_F$

| Outliers	Object Data Clustering					Relational Data Clustering				
	p = 0.5	p = 1	p = 1.5	p = 2	p = 3	p = 0.5	p = 1	p = 1.5	p = 2	p = 3
0	0	0	0.046	0.374	1.190	0	0	0.058	0.374	0.340
2	0	0	0.340	2.063	10.595	0	0	0.286	2.063	12.348
4	0	0	0.907	4.062	18.323	0	0	0.756	4.062	20.050
6	0	0	1.659	6.155	24.481	0	0	1.402	6.155	25.842
8	0	0	2.560	8.349	29.694	0	0	2.205	8.349	30.657
10	0	0	3.586	10.661	34.313	0	0	3.154	10.661	34.904
12	0	0	4.715	13.109	38.512	0	0	4.245	13.109	38.767
14	0	0	5.920	15.719	42.391	0	0	5.469	15.719	42.347
16	0	0	7.134	18.519	46.012	0	0	6.797	18.519	45.703
18	0	0	8.038	21.547	49.413	0	0	7.953	21.547	48.871
20	0	0	9.258	24.846	52.622	0	0	9.523	24.846	51.881
22	0	0	10.687	28.472	55.659	0	0	11.305	28.472	54.753
24	0	0	11.943	32.496	58.546	0	0	13.365	32.496	57.499
30	0	0	16.171	48.309	66.453	199.749	7.666	23.012	48.309	65.049
40	145.945	9.735	103.461	82.078	77.688	168.819	105.325	99.947	82.078	75.465
50	145.945	145.945	135.110	97.737	87.359	168.819	175.535	124.092	97.737	84.450
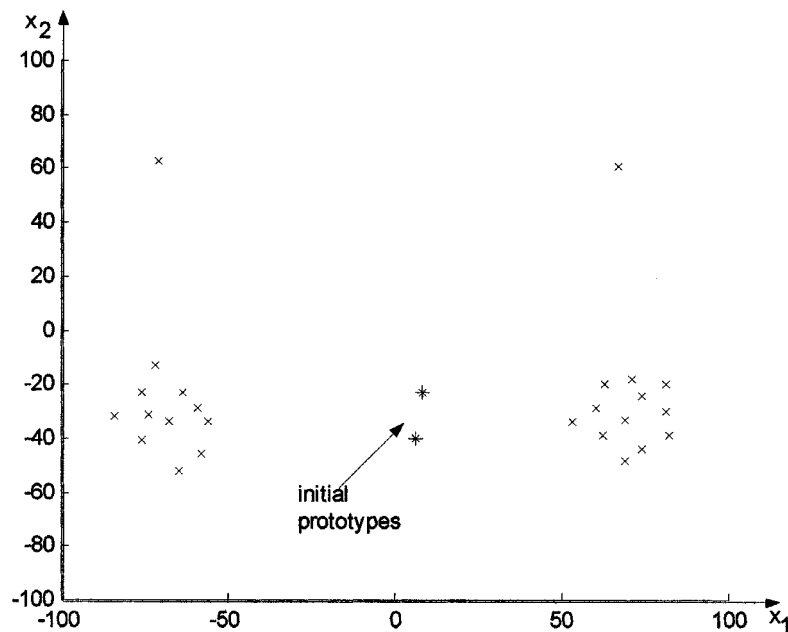
Fig. 3. "Two cluster data" scatterplot and initial prototypes $c = 2$.

approach for $p = 1$ offers the greatest robustness and efficient implementational approaches for it are discussed in [7].

The results of the first experiment show that in at least some cases, the errors produced by the object and relational approaches are of a similar magnitude. Is it also true that the computed prototypes and partition matrices of the two approaches are themselves very similar? The remaining numerical experiments use other artificial two-dimensional data sets that allow us to graphically depict the effect of p on the placement of the terminal prototype vectors. The two data sets (and initial prototype values) are depicted in Figs. 3 and

4 and are, respectively, called the "two cluster data" and "no cluster data." Using identical data p values, initializations, and stopping criteria, we calculate (U^R, v^R) and (U^O, v^O) using the relational and object data approaches, respectively. We calculated the Frobenius norm difference in the terminal partitions and prototypes produced by the two approaches as $U_d = \|U^R - U^O\|_F$ and $v_d = \|v^R - v^O\|_F$. These differences are given for a range of p values using the "two cluster data" and "no cluster data" in Table II.

Duality theory for NERFCM ([11], [12]) guarantees that the difference is zero when $p = 2$, but note that it is reasonably
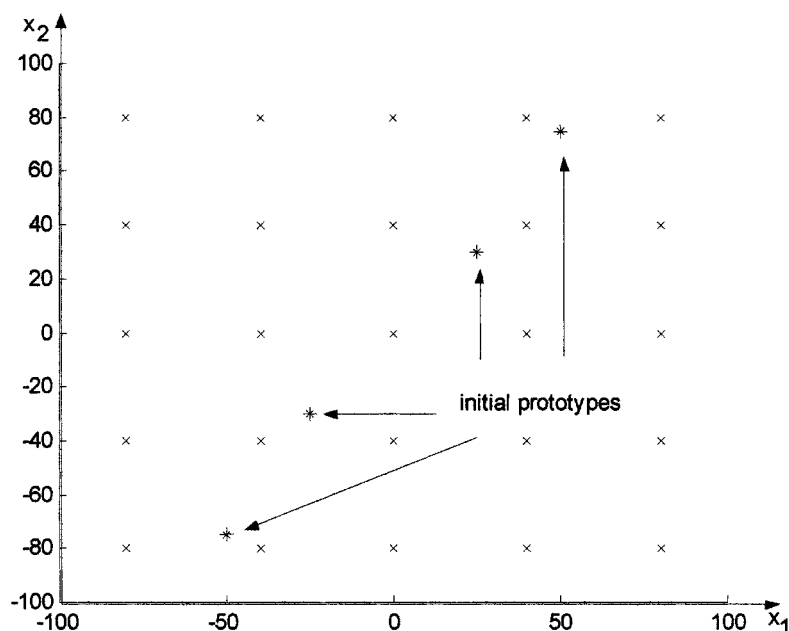
Fig. 4. "No cluster data" scatterplot and initial prototypes $c = 4$.

TABLE II
FROBENIUS NORM DIFFERENCE IN TERMINAL OBJECT AND RELATIONAL
(U, v) VALUES

p	"Two Cluster Data"		"No Cluster Data"	
	v_d	U_d	v_d	U_d
0.1	145.674	3.347	174.356	2.472
0.25	145.086	3.350	174.356	2.472
0.5	0.000	1.437	174.356	2.471
0.75	0.000	0.806	120.000	2.159
1.0	0.000	0.453	97.980	1.837
1.25	0.119	0.251	50.573	1.110
1.5	0.135	0.130	1.923	0.433
1.75	0.070	0.053	2.050	0.205
2.0	0.000	0.000	0.000	0.000
2.25	0.117	0.029	0.704	0.053
2.5	0.361	0.059	0.311	0.173
3.0	2.321	0.139	0.784	0.417
3.5	7.952	0.241	15.803	0.743
4.0	21.181	0.609	114.242	2.484
4.5	69.355	3.185	98.309	2.374

small for p near two. Note also how the difference v_d is sometimes much greater than the difference U_d. We also see from this table that the relational and object-based results can be very different for the important case of $p = 1$.

The remaining figures in this section graphically depict the position of the computed terminal prototypes v_t for a range of p values. Fig. 5 shows the results for the "two cluster data" obtained by the two approaches. Note that the outliers have increasing effect as p increases from $p = 0.5$ to $p = 2$. As p continues to increase above two, the outlying data has an even more powerful draw on the prototypes, which move ever nearer to the approximate center of the figure. Note that for $p = 5$, the terminal prototypes produced by the object-data approach collapse into coincident clusters. Because the "two

cluster data" set has vertical symmetry, so do $\mathbf{v}_1$ and $\mathbf{v}_2$ from either approach.

Fig. 6 shows a similar experiment for the relational-data approach applied to the "no cluster data." While this is not a "clustering" example, we used it to better understand the behavior of the methods. Surprisingly, we observed coincident prototypes at the center of the data for small values of p such as $p = 0.1$ and nearly coincident prototypes for large values such as $p = 100$; the prototypes are most different and farthest from the center when $p = 2$. The behavior of the object-data approach on this example was similar in that $p = 2$ gave the most separated prototype values.

IV. DISCUSSION

We described relational and object-data approaches for generating L_p norm extensions of FCM. Also, we examined the behavior of the approaches for various p values using artificial data sets. We believe that the two most useful models are object data based and correspond to $p = 1$ and 2. For $p = 2$, the fuzzy c-means algorithm in [1] offers the least expensive clustering technique of all, and it works very well in most cases. For cases where noisy data and outliers may degrade FCM results, we recommend the use of the object data model with $p = 1$, optimized using the fuzzy c-medians algorithm described in [7]. The relational data approach is best saved for cases when object data is unavailable or, in special cases for $p = 2$, when the dimension of the feature space is very high but the number of data is small. (In this case, it may be computationally cheaper to form R and operate on it rather than on the original feature data.) We believe the relational approach for $p = 1$ exhibits some robustness properties, but overall we view it as inferior to the object data approach of Kersten [7]. Our experiments always used $m = 2$, but we believe the importance of $p = 2$ and 1 holds for any choice of the fuzzification constant.
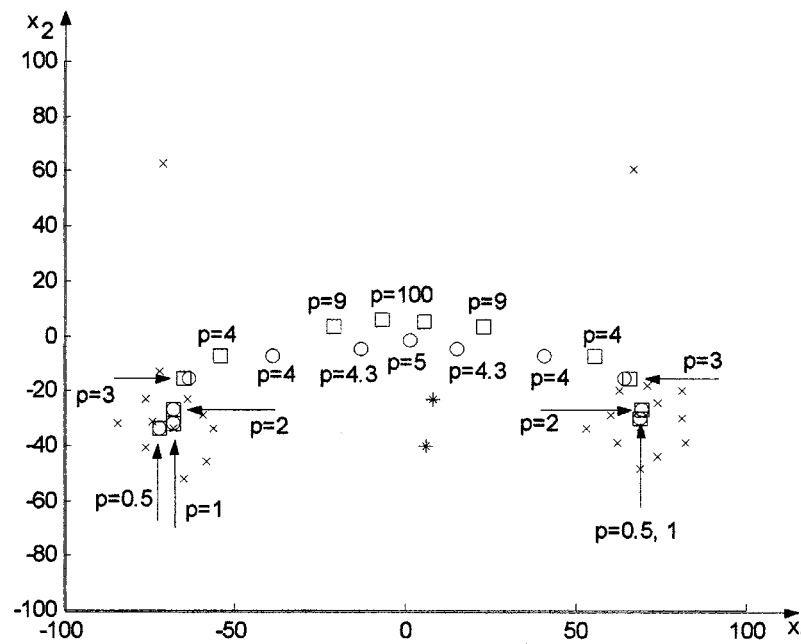
Fig. 5. "Two cluster data" terminal prototypes for the object ($\bigcirc$) and relational ($\square$) approaches $c = 2$.
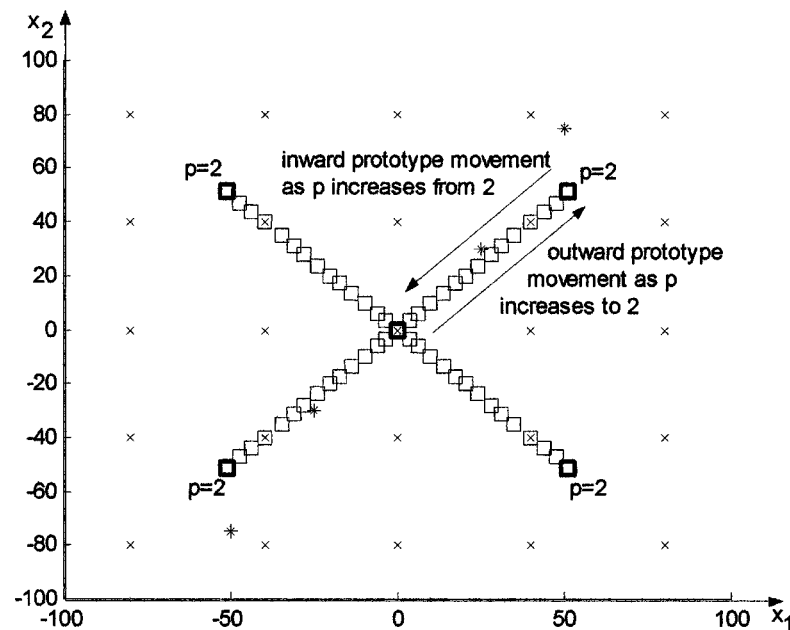

Fig. 6. "No cluster data" terminal prototypes for the relational-data approach $c = 4$.

Choices for p other than $p = 1$ or $p = 2$ lead to models which can provide good clustering results and possibly classifier designs, but the models are more difficult to optimize in the object data case. For p values near two, the results obtained using the object and relational approaches are quite similar. However, it now appears that the existing duality theory stated in [14] is complete; that is, the object and relational approaches have a strict duality relationship *only* when $p = 2$. As p values increase above one, the attraction of terminal prototype vectors to outliers increases. The empirical migration of the prototypes to the approximate center of the data sets as p increases is interesting to us, but no illuminating result regarding this has been obtained. A partial analysis of the relational case indicates that there is still a strong dependence on the point of initialization, even as p increases without bound. For example, consider clustering $\{(0,0)^T, (0,1)^T, (0,3)^T\}$ into two clusters. It is not hard to show that as $p \to \infty$, the relational based approach is essentially equivalent to that done using NERFCM on the 3×3 matrix R with all zero entries except $R_{13} = R_{31} = 1$. Simple numerical experiments with different crisp initializations show that different solutions are possible. Because of this, we believe there may not be a nice theoretical result regarding the limiting position of the prototypes as $p \to \infty$. We conclude by giving one last question. Why is $p = 2$ special in the sense demon-

strated by Fig. 6 and does this mean that the original FCM is in some sense optimal as a quantization tool?

ACKNOWLEDGMENT

REFERENCES

[1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[2] J. C. Bezdek, J. M. Keller, R. Krishnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Norwell, MA: Kluwer, 1999.

[3] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 98–110, May 1993.

[4] R. N. Dave, "Characterization and detection of noise in clustering," *Pattern Recogn. Lett.*, vol. 12, pp. 657–664, 1991.

[5] P. R. Kersten, "The fuzzy median and the fuzzy MAD," in *Proc. ISUMA-NAFIPS*, College Park, MD, 1995, pp. 85–88.

[6] ——, "Fuzzy order statistics and their application to fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 7, pp. 708–712, Dec. 1999.

[7] ——, "Implementing the fuzzy c-medians clustering algorithm," in *Proc. IEEE Conf. Fuzzy Syst.*, Barcelona, Spain, 1997, pp. 957–962.

[8] S. Miyamoto and Y. Agusta, "An efficient algorithm for L_1 fuzzy c-means and its termination," *Contr. Cybern.*, vol. 25, pp. 421–436, 1995.

[9] L. Bobrowski and J. Bezdek, "C-means clustering with the L_1 and L_∞ norms," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, pp. 545–554, May/June 1991.

[10] K. Jajuga, "L_1 norm-based fuzzy clustering," *Fuzzy Sets Syst.*, vol. 39, pp. 43–50, 1991.

[11] S. Miyamoto and Y. Agusta, "Efficient algorithms for L_p fuzzy c-means and their termination properties," in *Proc. 5th Conf. Int. Federation Classification Soc.*, Kobe, Japan, Mar. 1996, pp. 255–258.

[12] ——, "Algorithms for L_1 and L_p fuzzy c-means and their convergence," in *Studies in Classification, Data Analysis, and Knowledge Organization: Data Science, Classification, and Related Methods*, C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H. H. Bock, and Y. Baba, Eds. Tokyo, Japan: Springer-Verlag, 1998, pp. 295–302.

[13] D. D. Overstreet, "Generalized fuzzy c-means clustering," M.S. thesis in mathematics, Georgia Southern Univ., Statesboro, GA, June 1998.

[14] R. J. Hathaway and J. C. Bezdek, "NERF c-means: Non-Euclidean relational fuzzy clustering," *Pattern Recogn.*, vol. 27, pp. 429–437, 1994.

[15] ——, "Optimization of clustering criteria by reformulation," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 241–245, May 1995.

Richard J. Hathaway (M'93) ("BIG Rivets") received the B.S. degree in applied math from the University of Georgia, Athens, in 1979, and the Ph.D. degree in mathematical sciences from Rice University, Houston, TX, in 1983.

He is currently a Professor in the Mathematics and Computer Science Department, Georgia Southern University, Statesboro. His research interests include pattern recognition and numerical optimization.

James C. Bezdek (M'90–SM'90–F'92) (affectionately known as "Jackhammer" by coworkers in the reviewing construction trades) received the B.S.C.E. degree from the University of Nevada, Reno, in 1969, and the Ph.D. degree in applied math from Cornell University, Ithaca, NY, in 1973.

He is currently a Professor in the Computer Science Department, University of West Florida, Pensacola. His interests include woodworking, fine cigars, optimization, motorcycles, pattern recognition, gardening, fishing, image processing, snow skiing, computational neural networks, blues music, woodworking, and computational medicine.

Dr. Bezdek is the founding editor of the *National Journal of Approximate Reasoning* and the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He is a fellow of IFSA.

Yingkang Hu ("Grinder") received the B.S. degree in mathematics from Beijing University of Chemical Engineering, China, in 1982, and the Ph.D. degree in mathematics from the University of South Carolina, Columbia, in 1989.

He is a Professor in the Mathematics and Computer Science Department, Georgia Southern University, Statesboro. His research interests include approximation theory, numerical computation, and pattern recognition.