NUWC-NPT Technical Report 10,298
1 December 1993

Fuzzy Robust Statistics for Application to the Fuzzy c-Means Clustering Algorithm

P. R. Kersten
Combat Control Systems Department

Naval Undersea Warfare Center Division
Newport, Rhode Island

PREFACE

This report was funded under NUWC Newport IR/IED Project No. 802424 "Fuzzy Expert Systems." The IR/IED program is funded by the Office of Naval Research; the NUWC Newport program manager is K. M. Lima (Code 102).

The technical reviewer for this report was K. F. Gong (Code 2211).

The author expresses his thanks to Professor Avi Kak of Purdue University for his general encouragement and motivating example.

Reviewed and Approved: 1 December 1993

P. A. La Brecque
Head, Combat Control Systems Department

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1 December 1993	

4. TITLE AND SUBTITLE
Fuzzy Robust Statistics for Application to the Fuzzy c-Means Clustering Algorithm

5. FUNDING NUMBERS

6. AUTHOR(S)
P. R. Kersten

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Undersea Warfare Center Division
1176 Howell Street
Newport, Rhode Island 02841-1708

8. PERFORMING ORGANIZATION REPORT NUMBER
TR 10,298

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report introduces two robust statistics -- the fuzzy median and the fuzzy median absolute deviation from the median -- that have been developed for use with fuzzy data sets. The two statistics were applied to the fuzzy c-Means algorithm, a powerful clustering algorithm that normally employs linear statistics. The modified algorithm showed improved performance, being able to cluster data sets generated by heavy-tailed distributions like the Cauchy and Slash distributions.

14. SUBJECT TERMS
Fuzzy Data Sets
Robust Statistics
Clustering Algorithms
Data Analysis

15. NUMBER OF PAGES
17

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

TABLE OF CONTENTS

Section		Page
1.	INTRODUCTION	1
2.	FUZZY C-MEANS CLUSTERING ALGORITHM	3
3.	FUZZY ROBUST STATISTICS	7
4.	MODIFIED FUZZY C-MEANS ALGORITHM	9
5.	CONCLUSIONS	13
6.	REFERENCES	13

LIST OF ILLUSTRATIONS

Figure		Page
1	Scatter Diagram for the Normal Clusters Centered at (-1.27, 0.00) and (1.27,0.00)	5
2	Fuzzy c-Means Convergence on Two Gaussian Clusters Located at Antipodal Positions	5
3	Scatter Diagram for the Cauchy Clusters Centered at (-1.27, 0.00) and (1.27,0.00)	6
4	Robust Fuzzy c-Means Convergence on Two Gaussian Clusters Located at Antipodal Positions	11
5	Robust Fuzzy c-Means Convergence on Two Cauchy Clusters Located at Antipodal Positions	11
6	Scatter Diagram and Exemplar Trajectories for the Slash Distribution	12

FUZZY ROBUST STATISTICS FOR APPLICATION TO THE FUZZY C-MEANS CLUSTERING ALGORITHM

1. INTRODUCTION

Clustering is an important tool for discovering patterns in exploratory data analysis. In pattern recognition, clustering is one technique used before designing a classifier. Clustering is also a form of unsupervised learning helpful in defining the rules in fuzzy system design. Fuzzy sets are useful in clustering algorithms since each data point may belong to more than one cluster at the same time, permitting smoother convergence of the clustering process. The result of a fuzzy clustering algorithm is a fuzzy partition of the data into c classes. Fuzzy c-Means, a powerful clustering algorithm, illustrates the success of the fuzzy algorithms that have emerged over the last three decades.

Fuzzy c-Means is a generalization of the hard c-Means algorithm. Hard c-Means is an iterative procedure that assigns a class to each point based on the closest class exemplar. The class assignment forms a partition of the data set and thus generates equivalence classes. Fuzzy c-Means generalizes hard c-Means by softening class membership of the data points in the c classes. Instead of a data point belonging to a unique class, a data point may belong to all of the classes, but with varying degrees of membership. So, associated with each data point is a fuzzy unit vector (fit vector) where the i-th element in the vector is the membership value of the point in the i-th class, and the sum of the elements of the vector must be one. In fuzzy c-Means, the distances to the class exemplars are used to modify the fit vector to change the relative memberships of the data point in the classes. The clustering yields a fuzzy partition of the data.

In both of the c-Means clustering algorithms, the class exemplar or center is calculated using linear statistics. For hard c-Means, the exemplar is a strict arithmetic average; for fuzzy c-Means, it is a weighted average in which the weights are the memberships in the classes. But linear statistics are notoriously vulnerable to outliers; for example, only one bad data point or outlier can destroy the sample mean as a measure of centrality and the sample variance as a measure of dispersion. One goal of robust statistics is to develop statistics that are more resistant to outliers. Two examples of these statistics are the median for centering the data and the median absolute deviation from the median (MAD) for estimating the dispersion of the data. However, these statistics have *not* been designed to work with fuzzy sets. Both statistics are based on order statistics that do not inherently take into account membership of the data points in more than one set or class.

This report addresses the above deficiency by introducing a fuzzy median and a fuzzy MAD estimator. When these statistics are used with the fuzzy c-Means algorithm, the result is a more robust algorithm.

In section 2, the fuzzy c-Means algorithm is described in detail. Examples show how this algorithm works well on light-tailed clusters but not on heavy-tailed clusters, the problem lying with linear statistics used within this algorithm. Section 3 shows how the median and MAD estimators are "fuzzified" so that they can be applied to the fuzzy c-Means algorithm. Section 4 outlines the modified fuzzy c-Means procedure and gives an example of how it works; the modified algorithm is shown to cluster data sets generated by heavy-tailed distributions like the Cauchy and the Slash distributions. Section 5 summarizes the results and indicates the direction of future work on fuzzy robust statistics.

2. FUZZY C-MEANS CLUSTERING ALGORITHM

Fuzzy c-Means is a practical clustering algorithm that generalizes the hard c-Means algorithm. The latter is an iterative procedure that is best described operationally. Given an initial assignment of data points to classes, one calculates class exemplars by averaging the data points in the various classes. These class exemplars are then used to assign new classes by calculating the closest exemplar for each data point. Data points assume the class identity of their closest exemplar. This procedure is repeated until some form of convergence occurs. The algorithm yields a partition of the data points into classes.

Fuzzy c-Means generalizes the hard c-Means algorithm by replacing the class assignment with a membership vector whose elements represent the membership of the data points in each of the classes. The algorithm produces a fuzzy partition of the data into c classes, i.e., each point has a membership vector or fit vector associated it, rather than a single class assignment. The algorithm is basically an unsupervised learning technique, and the following description is based on Bezdek's approach to fuzzy pattern recognition (references 1-2).

Consider N data samples forming the data set denoted by $X = \{x_1, x_2, \ldots, x_N\}$, where each sample is a p-dimensional real vector, $x_i \in R^p$. Assume there are c classes and $u_{ik} = u_i(x_k) \in [0,1]$ is the membership of the k-th sample in the i-th class. This leads to a matrix representation of the membership function associated with the fuzzy c-Partition M_{fc} defined as

$$M_{fc} = \left\{ U \in V_{cn} \mid u_{ik} \in [0,1] \forall i,k; \sum_{i=1}^{c} u_{ik} = 1 \; \forall k; 0 < \sum_{k=1}^{N} u_{ik} < N \; \forall i \right\},$$

where V_{cn} is the set of real c x n matrices and c is an integer such that $2 \le c < N$ (reference 1, p. 26). Each row of U is a class and each column of U is a fit vector in which the i-th vector element represents the membership of the data vector in the i-th class. The sum of the elements in any column must be one. Thus, each column describes the degree that each point belongs to each of the classes. $v = \{v_1, v_2, \ldots, v_c\}$ is the set of cluster exemplars or prototypes for the c clusters. The fuzzy c-Means algorithm minimizes the functional

$$J(U,v) = \sum_{k=1}^{N} \sum_{i=1}^{c} (u_{ik})^{m_c} (d_{ik})^2$$

using the following algorithm (found in reference 1):

1. Fix c, the number of classes such that $c \in \{2, \ldots, N-1\}$.

   Choose an inner product norm metric in R^p and fix the weighting exponent $m_c \in [1, \infty)$. (The c in m_c refers to the c-Means algorithm and m refers to the median.)

   Initialize the membership matrix denoted by $U^{(0)} \in M_{fc}$.

2. Construct the c exemplars by a weighted average $v_i = \sum_{k=1}^{N} w_{ik} x_k$, where the weights w_{ik} are the normalized membership functions given by

$$w_{ik} = (u_{ik})^{m_c} \Big/ \sum_{j=1}^{N} (u_{ij})^{m_c} .$$

3. Update the memberships u_{ik} in the membership matrix with

$$u_{ik} = 1 \Big/ \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m_c - 1)} \right],$$

provided of course that none of the d_{jk} are zero. In the latter case, the u_{ik} are assigned differently (reference 1, p. 66).

4. Compare the last two membership matrices, $U^{(l)}$ and $U^{(l+1)}$, and when they are sufficiently close, terminate the algorithm; otherwise, return to step 2.

Note that in step 2 the exemplars are constructed using a linear combination of the data points where the weights are normalized membership functions.

The distances can be generalized to include the covariance of each cluster; the covariance is a fuzzy version as estimated by Kessel (reference 3) where the i-th class covariance matrix is given by $\Sigma_i = KS_i$, where K is related to a volume constraint on the i-th cluster and S_i is the fuzzy scatter matrix given by

$$S_i = \sum_{k=1}^{N} w_{ik} (x_k - v_i)(x_k - v_i)^t .$$

The constant is given as $K = \left[\rho_i \det(S_i) \right]^{(-1/p)}$, where ρ_i is related to a volume constraint on the i-th cluster and p is the dimension of the vector space. If the dispersion of the cluster is used within the algorithm, then between step 2 where the exemplars are updated and step 3 where the memberships are updated, the fuzzy covariances must be estimated since the distances are now defined by $(d_{ik})^2 = (x_k - v_i)^t \Sigma_i^{-1} (x_k - v_i)$. The fuzzy scatter matrix is also a linear combination of the outer products of the centered data vectors, so it too is vulnerable to the presence of outliers.

This algorithm with the covariance modification of Kessel was applied to two Gaussian clusters located at antipodal positions as illustrated in figure 1. These clusters were generated using centers at (b,0) and (-b,0), where b = 1.27, and an identity covariance matrix. The value of b was chosen to give a 10-percent classification error using a linear discriminant function. The fuzzy c-Means algorithm was run for 20 iterations. The results are illustrated in figure 2 where the path of the exemplar solutions produces tracks that converge toward the cluster centers.

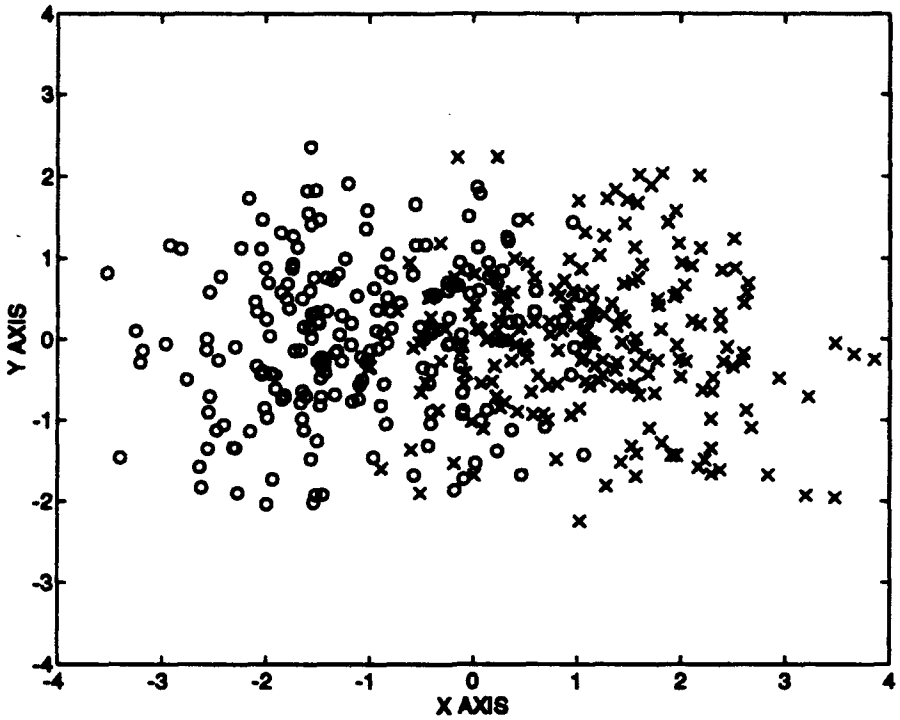

Figure 1. Scatter Diagram for the Normal Clusters Centered at (-1.27, 0.00) and (1.27, 0.00).

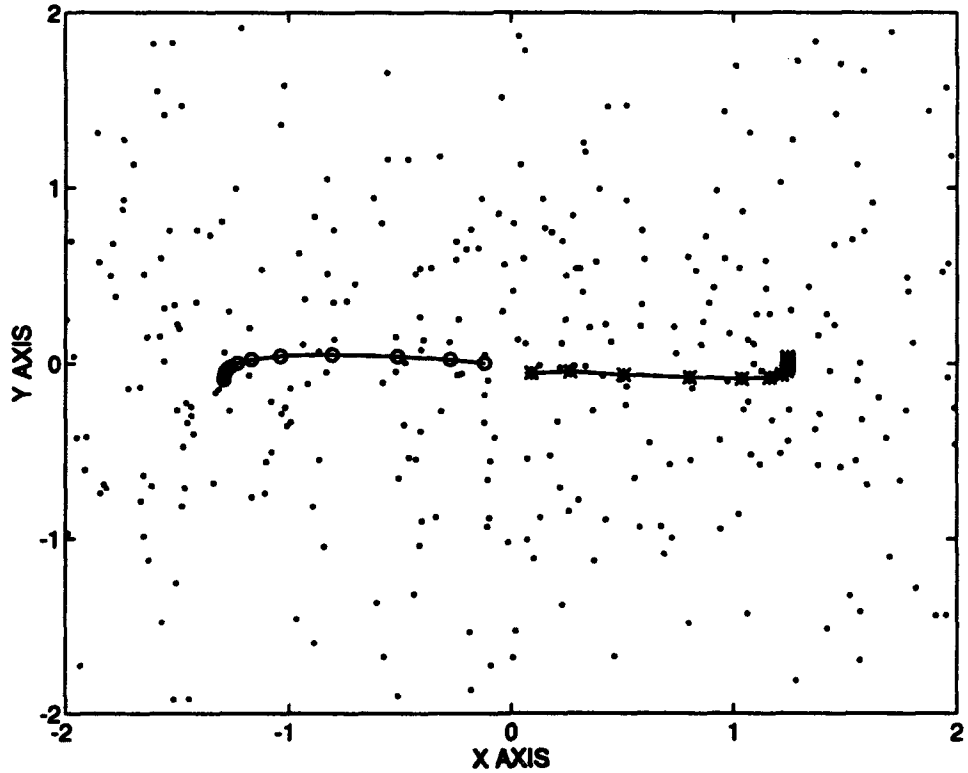

Figure 2. Fuzzy c-Means Convergence on Two Gaussian Clusters Located at Antipodal Positions. (The two distributions used to generate the random deviates are N(-1.27,1) and N(1.27,1).)

The same algorithm applied to two Cauchy clusters placed at the same centers does *not* work as well. In fact, the algorithm does not appear to converge after 20 iterations. Figure 3 shows the two Cauchy clusters that appear to be more dispersed, although benign. Figure 3 is deceiving as not all the data points are shown because they would not fit on this scale. A small percentage of the points are extreme outliers, as is typical of clusters generated using a Cauchy distribution function. In fact, the Cauchy distribution does not technically possess any moments. For the p-th moment of any distribution to be defined, one must have $E|X|^p < \infty$, which is not the case for any $p \geq 1$. The outliers destroy the exemplar estimation in step 2 of the algorithm, as well as the covariance estimation, which is dependent on the results of step 2. This is precisely the reason for developing the fuzzy median and the fuzzy MAD.

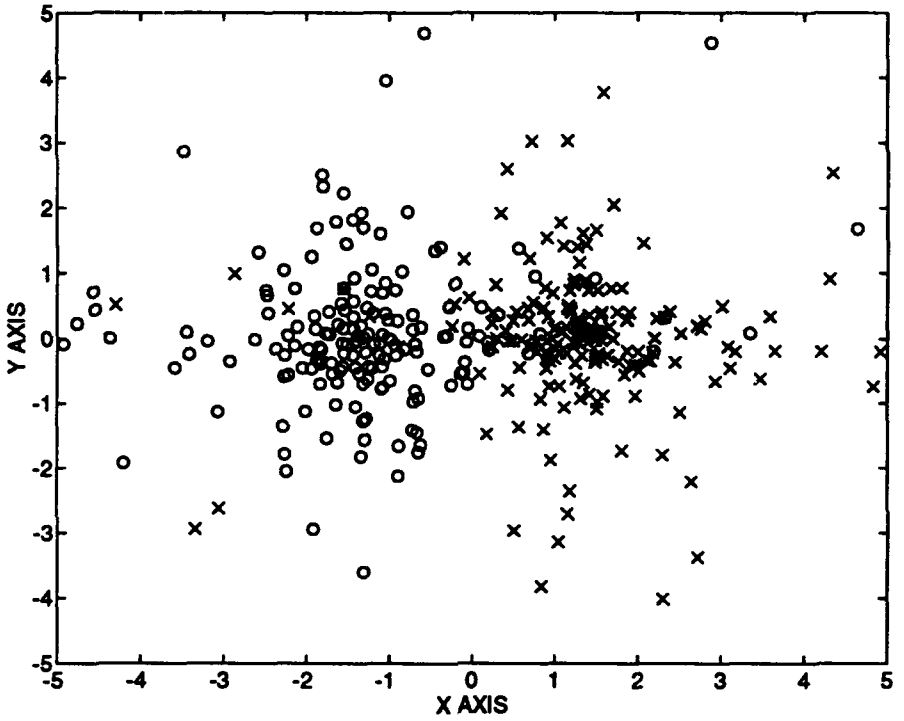

Figure 3. Scatter Diagram for the Cauchy Clusters Centered at (-1.27, 0.00) and (1.27, 0.00) (Not all the points are shown because of the extreme outliers caused by this heavy-tailed distribution.)

It is easy enough to say that one should first remove these outliers so they would be no problem. In the best of all worlds this is a possibility, but it is not easily done, especially in high dimensional data. One would need to run other clustering algorithms or use nonparametric statistics to calculate the cluster centers and then "peel off" the outliers. A better solution is to use a robust method that is resistant to outliers so that the extensive exploratory data analysis could be eliminated. In automated decision systems, this is a necessity. The approach is to replace the centering and dispersion estimates with some robust counterpart resistant to outliers. However, these statistical counterparts have to be compatible with the philosophy of fuzzy clustering, which allows each data point to be a member of all the clusters in differing degrees. Such fuzzy robust statistics are developed in the next section.

3. FUZZY ROBUST STATISTICS

Robust statistics is an area of study that deals with variations from ideal assumptions (references 4-7). This study area includes the design of statistics that are resistant to outliers. These statistics are sometimes evaluated by the percentage of outliers that must be present before the statistic no longer gives a meaningful estimate of the desired quantity. One example of such a statistic is the median, which can include almost 50-percent outliers before it loses its ability to measure the center of the data sample. A second example is the MAD estimate, which measures the dispersion of the data sample. It too has a high breakdown point. In this section, the median and the MAD are defined and the source of their resistant behavior is explained. Then, alternative definitions are given that can be generalized for application to fuzzy data sets. *Throughout this section, the data samples are assumed to be one dimensional. When the median and MAD statistics are applied to higher dimensional samples, it is on a component-by-component basis.*

Suppose the data set is $X = \{x_1, x_2, \ldots, x_N\}$, where each element x_i is a p-dimensional vector. If p = 1, so that the samples are real numbers, then the median of X is defined in terms of the order statistics. The ordered N-sample is $\{x_{(1)}, x_{(2)}, \ldots, x_{(N)}\}$, where by definition $x_{(1)} \le x_{(2)} \le \ldots \le x_{(N)}$ and the subscript (i) means that the original data have been relabeled or permuted so the sample set is ordered. Then, the median of X is defined to be $x_{(l+1)}$ if $N = 2l + 1$ and $[x_{(l)} + x_{(l+1)}]/2$ if $N = 2l$. If p > 1, the definition is applied to each dimension of the sample and the median vector is constructed from the vector of individual medians. Since the median is defined in terms of the ordered sample, it is an order statistic and there appears to be no way to extend it to a fuzzy set. In one dimension, the median represents the halfway point of the sample set, having an equal number of samples smaller and larger than itself. This interpretation explains why almost half of the data points must be outliers before the median loses its effectiveness as a measure of centrality. Half of the points to the left, say, must be outliers before the median is pulled off to the left. In fact, the finite breakdown point of the median is one-half (reference 4, p. 99).

The robust estimator of dispersion, the MAD, is also an order statistic. This statistic is defined as the median of the absolute deviations from the median. To construct this statistic, one takes the samples $X = \{x_1, x_2, \ldots, x_N\}$ and constructs another data set

$$Y = \{ |x_1 - med(X)|, |x_2 - med(X)|, \ldots, |x_N - med(X)| \},$$

finds the median of this set, and then scales it. For this report, the MAD is defined as $mad(X) = med(Y) / 0.6745$, where the constant 0.6745 adjusts the dispersion measure to be 1 when the sample is Gaussian with unit variance. Intuitively, one folds the data about the $med(X)$ and finds the median of the set of positive deviations about the median. The breakdown point of the MAD is also one-half (reference 5, p. 107).

Although the median and the MAD are resistant to outliers, they are constructed on crisp sets, i.e., the set of data points X and the set of absolute deviations about the median Y are crisp sets. The linear order of the real numbers does not take into account the membership of the points in these sets. However, if these statistics are reformulated, the membership of the samples in the sets can be taken into account. This is accomplished by using another definition of the median that does not depend on the linear ordering of the samples. The median can be defined as the solution m of (reference 8, p. 233)

$$\min_{m \in R} \sum_{k=1}^{N} |x_k - m| .$$

This can be seen most easily by taking the derivative with respect to m, so one wants

$$\sum_{k=1}^{N} sgn(x_k - m) = 0.$$

For $N = 2l+1$, the derivative is zero at the point $m = x_{(l+1)}$, and for $N = 2l$, the derivative is zero for any $m \in (x_{(l)}, x_{(l+1)})$. This definition is amenable to generalization to fuzzy sets. For the c class problem, if u_{ik} is the membership of x_k in class i, then m_i is given by

$$\min_{m_i \in R} \sum_{k=1}^{N} u_{ik} |x_k - m_i|,$$

and the resulting statistic m_i has the characteristics of a median and applies to the fuzzy set $X_i = u_{i1}/x_1 + u_{i2}/x_2 + \cdots + u_{iN}/x_N$. Fuzzy sets are defined using a "+" sign to link the elements u_{ik}/x_i of the set, where u_{ik} is the membership grade and x_i is the element. This is precisely the set that is used to update the exemplars in the fuzzy c-Means algorithm. The derivative also exists and is given by

$$\sum_{k=1}^{N} u_{ik} sgn(x_k - m_i).$$

Numerically solving for the root of this functional using bisection yields both the root and the fuzzy median estimate of the set.

The limiting cases of this statistic are consistent. If $u_{ik} = u_i(x_k) = 1, \forall k$, then the definition reverts to the standard median. For the two-class problem, if $N = 2l$ and the first $N/2$ samples are from class 1 and the second $N/2$ are from class 2, then

$$u_{1k} = u_1(x_k) = \begin{cases} 1, & k \le N/2 \\ 0, & k > N/2 \end{cases} \quad \text{and} \quad u_{2k} = u_2(x_k) = \begin{cases} 0, & k \le N/2 \\ 1, & k > N/2 \end{cases}.$$

The resulting fuzzy medians reduce to the crisp medians on both of the crisp subsets associated with the classes. The fuzzy median is a measure of central tendency that also reflects the membership of the sample points in the fuzzy set and that reduces to the crisp median where appropriate.

The MAD estimate can also be reformulated in the functional form

$$\min_{\eta \in R} \sum_{k=1}^{N} \left| |x_k - m| - \eta \right|,$$

and the resulting MAD estimate is $mad = \eta / 0.6745$. Note that this requires that the median m be known beforehand. For a fuzzy data set, the median does not exist; however, the fuzzy median does. So one can define the fuzzy MAD recursively for the k-th fuzzy data set by assuming that the fuzzy median m_k exists. This functional definition is given by

$$\min_{\eta_i \in R} \sum_{k=1}^{N} \left| |x_k - m_i| - \eta_i \right|,$$

and the fuzzy MAD is given by $fuzmad_i = \eta_i / 0.6745$. From an implementation point of view, the process is somewhat recursive since one first forms the fuzzy median m_i, uses this to construct a new fuzzy data set

$$Y_i = u_{i1}/|x_1 - m_i| + u_{i2}/|x_2 - m_i| + \cdots + u_{iN}/|x_N - m_i|,$$

finds the fuzzy median on this set, and then scales it. To apply these statistics on a p-dimensional space, one has to apply them on each component separately.

Although this approach is a simple modification of these statistics, it is important to note that it allows the application of robust statistics to fuzzy algorithms. The median is an M-estimator or Huber's generalization of the maximum likelihood estimator. Many times these estimators m are formulated implicitly by specifying functionals of the form

$$\sum_{k=1}^{N} \rho(x_k - m),$$

where ρ satisfies certain boundary, symmetry, and non-negativity conditions. It would seem that all these functionals could be weighted with the appropriate membership functions, thus allowing this whole class of estimators to be applied to fuzzy algorithm development.

4. MODIFIED FUZZY C-MEANS ALGORITHM

At each iteration, the fuzzy c-Means algorithm depends on two estimates made on the fuzzy subsets associated with the c classes. In step 2, a linear combination of data points is used to estimate the exemplars of the fuzzy classes. In step 3, the new membership values are estimated based on the distance to the exemplars, which are normalized with the inverse of the fuzzy covariance matrix of the data sets. These linear statistics are now replaced by the corresponding fuzzy robust statistics. Thus, for each class, the sample mean is replaced by the fuzzy median. The covariance matrix is replaced by a diagonal approximation of $\Sigma = diag(mad^2(X_1),\ldots,mad^2(X_p))$, a simple yet effective approximation.

The robust fuzzy algorithm is then stated as follows:

1. Fix c, the number of classes such that $c \in \{2,\ldots,N-1\}$.

Choose an inner product norm metric in R^p and fix the weighting exponent $m_c \in [1,\infty)$. (The c in m_c refers to the c-Means algorithm and m refers to the median.)

Initialize the membership matrix denoted by $U^{(0)} \in M_{fc}$.

2. For each class $i = 1,\ldots,c$, and for each component of the data vector $j = 1,\ldots,p$, solve for m_{ij}:

$$\sum_{k=1}^{N} u_{ik} sgn(x_{kj} - m_{ij}) = 0,$$

where x_{kj} is the j-th component of the k-th sample vector and m_{ij} is the j-th component of the fuzzy median for the i-th cluster. The new exemplars are $v_i = m_i$.

3. For each class $i = 1,\ldots,c$, form the new fuzzy vector set

$$u_{i1}/|x_1 - m_i| + u_{i2}/|x_2 - m_i| + \cdots + u_{iN}/|x_N - m_i|,$$

and for each component of the data vector $j = 1,\ldots,p$, solve for η_{ij}:

$$\sum_{k=1}^{N} sgn\left(|x_{kj} - m_{ij}| - \eta_{ij}\right) = 0,$$

and then scale $\eta_{ij} \leftarrow \eta_{ij} / 0.6745$ to obtain the fuzzy MAD estimator.

Form the class covariance matrix from $\Sigma_i = diag(\eta_{i1},\ldots,\eta_{ip})$.

4. Update the memberships u_{ik} using $(d_{ik})^2 = (x_k - v_i)^t \Sigma_i^{-1} (x_k - v_i)$ in

$$u_{ik} = 1 \Big/ \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(m_c - 1)} \right],$$

provided of course that none of the d_{jk} are zero. In the latter case, the u_{ik} are assigned differently (reference 1, p. 66).

5. Compare the last two membership matrices, $U^{(l)}$ and $U^{(l+1)}$, and when they are sufficiently close, terminate the algorithm; otherwise, return to step 2.

Figure 4 shows the exemplar tracks for the same Gaussian clusters used to illustrate the fuzzy c-Means algorithm in figure 2. Note that the exemplar tracks again converge on the cluster centers, but by a somewhat more straightlined trajectory. So, for the light-tailed distributions, the convergence is as expected with the fuzzy c-Means algorithm. Figure 5 shows similar exemplar trajectories superimposed on the Cauchy clusters produced by the robust fuzzy c-Means algorithm. The corresponding trajectories for fuzzy c-Means do not exist, since one of the exemplars simply diverged. The divergence was caused by the large outliers destroying the sample mean and covariance estimates.

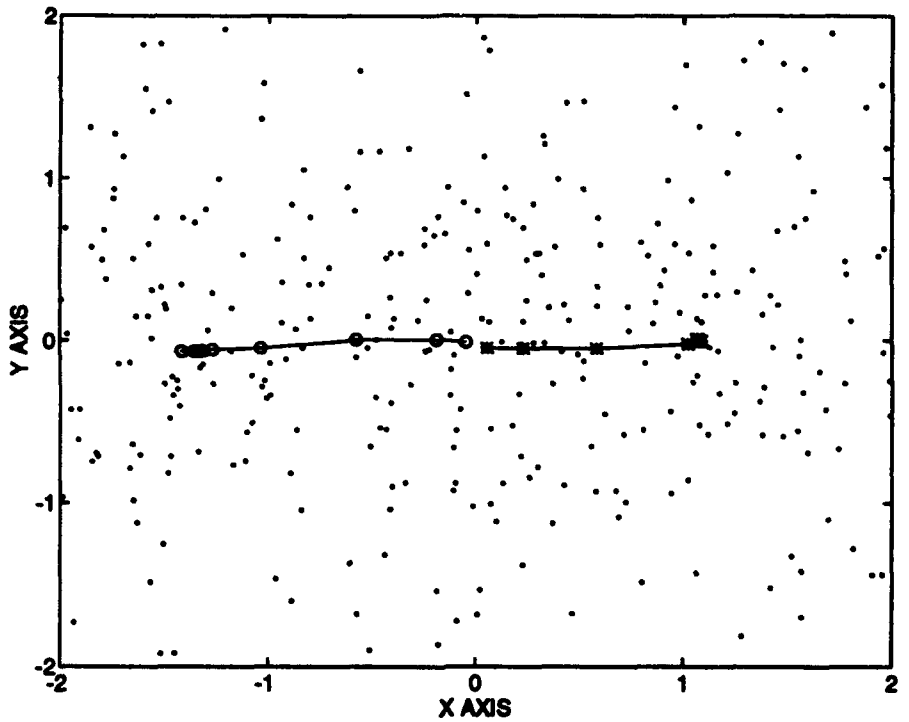

Figure 4. Robust Fuzzy c-Means Convergence on Two Gaussian Clusters Located at Antipodal Positions (The two distributions used to generate the random deviates are N(-1.27,1) and N(1.27,1).)

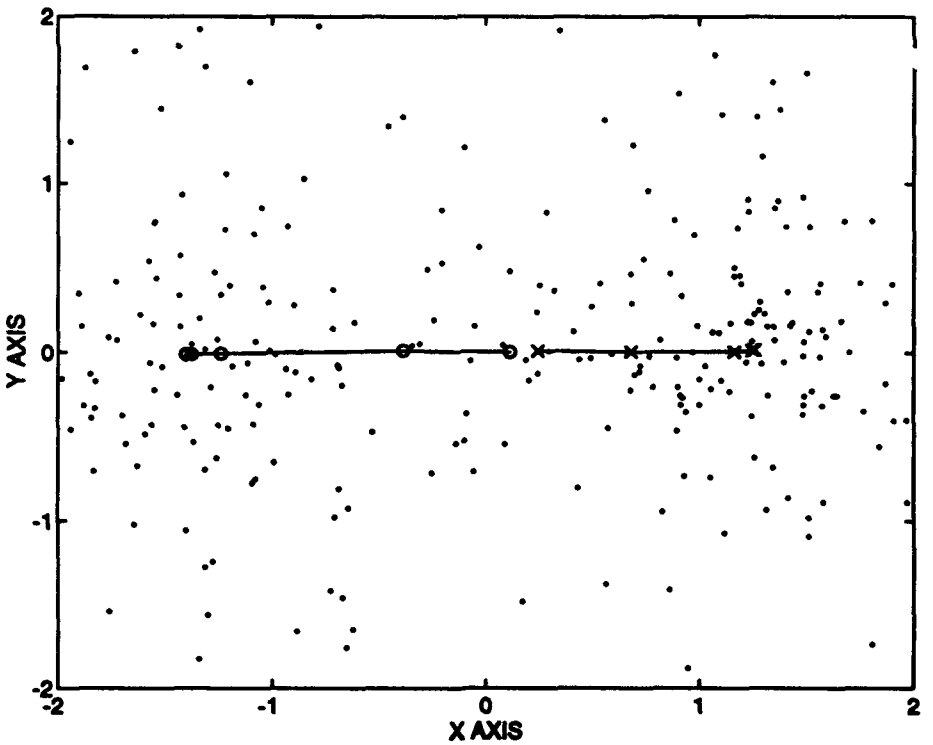

Figure 5. Robust Fuzzy c-Means Convergence on Two Cauchy Clusters Located at Antipodal Positions (The points shown in the figure do not include the outliers caused by the heavy-tailed distribution.)

A second illustration of the same observed behavior is with the Slash distribution, which is defined as the ratio of a Gaussian to a uniform distribution (reference 7). The sample set illustrated in figure 6 was generated using the ratio of N(0.0,0.5) and uniform [0,1] random deviates. Again, the robust fuzzy c-Means algorithm converges nicely, but the standard fuzzy c-Means does not.

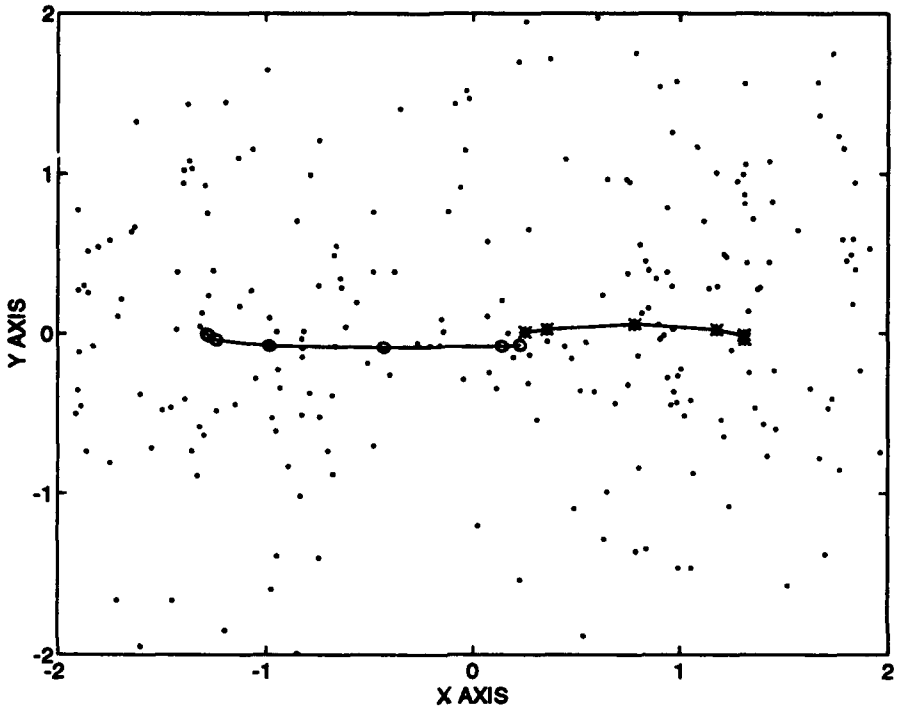

Figure 6. Scatter Diagram and Exemplar Trajectories for the Slash Distribution

5. CONCLUSIONS

Before order statistics can be applied to fuzzy data sets, they must be generalized to take into account the membership of the data points in the fuzzy set. This was done by properly weighting the functionals that generate the statistics with the membership values of the data points. Two such statistics were developed, the fuzzy median and the fuzzy MAD. The fuzzy median can replace the weighted average and the fuzzy MAD can replace the standard deviation. The fuzzy MAD can also be used to generate an approximate replacement for the covariance matrix. Both statistics were applied within the fuzzy c-Means clustering algorithm and shown for two examples to stabilize the convergence.

The approach used to "fuzzify" robust statistics is currently being applied to more sophisticated robust estimates of location and scale and the results will be reported when available. The replacement of linear estimates by more resistant fuzzy estimators in any fuzzy algorithm should help to make that algorithm more robust.

REFERENCES

1. James C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

2. James Bezdek, "Pattern Recognition with Fuzzy Sets and Neural Nets," Tutorial Notes, *Second IEEE International Conference on Fuzzy Systems*, IEEE, 1993.

3. Donald E. Gustafson and William C. Kessel, "Fuzzy Clustering with a Fuzzy Covariance Matrix," *Proceedings of the IEEE CDC*, San Diego, CA, pp. 761-766, Jan., 10-12, 1979.

4. Frank R. Hampel, Peter J. Rousseeuw, Elvezio M. Ronchetti, and Werner A. Stahel, *Robust Statistics, The Approach Based on Influence Functions*, John Wiley, New York, 1986.

5. Peter J. Huber, *Robust Statistics*, John Wiley, New York, 1981.

6. Peter J. Rousseeuw and Annick M. Leroy, *Robust Regression and Outlier Detection*, John Wiley, New York, 1987.

7. David C. Hoaglin, Frederick Mosteller, and John W. Tukey, *Understanding Robust and Exploratory Data Analysis*, John Wiley, New York, 1983.

8. Ronald H. Randles and Douglas A. Wolfe, *Introduction to The Theory of Nonparametric Statistics*, John Wiley, New York, 1979.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
Program Executive Officer for Undersea Warfare (ASTO-B--W. Chen, ASTO-G--G. Kamilakis, ASTO-G3--LCDR Traweek)	3
Chief of Naval Operations (N872E, N872E2)	2
Chief of Naval Research (ONR-4520, ONR-4525)	2
Naval Air Warfare Center Weapons Division (Code 2158--O. McNiel, J. Hodge)	2
Naval Surface Warfare Center Carderock Division (Code N04W--M. Stripling)	1
Commander Submarine Force Atlantic Fleet	1
Commander Submarine Force Pacific Fleet	1
Commander Submarine Development Squadron 12	1
Advanced Research Projects Agency	2
Defense Technical Information Center	12
Center for Naval Analyses	1