Letters

Fuzzy Order Statistics and Their Application to Fuzzy Clustering

Paul R. Kersten, *Member, IEEE*

Abstract— **The median and the median absolute deviation (MAD) are robust statistics based on order statistics. Order statistics are extended to fuzzy sets to define a fuzzy median and a fuzzy MAD. The fuzzy c-means (FCM) clustering algorithm is defined for any p-norm (pFCM), including the ℓ_1-norm (1FCM). The 1FCM clustering algorithm is implemented via the alternating optimization (AO) method and the clustering centers are shown to be the fuzzy median. The resulting AO-1FCM clustering algorithm is called the fuzzy c-medians (FCMED) clustering algorithm. An example illustrates the robustness of the FCMED.**

Index Terms—**FCM, FCMED, fuzzy sets, MAD.**

I. Introduction

ROBUST statistics are designed to be resistant to outliers. Two examples are the median for estimating the center of the data and the median of the absolute deviations from the median (MAD) for estimating the dispersion of the data. These statistics do *not* apply directly to fuzzy sets since both are based on order statistics, which implicitly assume the data belongs entirely in one set. These statistics are extended to apply to fuzzy sets and then used to implement an alternating optimization (AO) version of the ℓ_1-norm fuzzy c-means (1FCM) clustering algorithm, where the membership functions (MF's) are given by [1] and the cluster centers are fuzzy medians. This version is called the fuzzy c-medians (FCMED) clustering algorithm since the weighted median plays the same role as the weighted mean in the fuzzy c-means (FCM). The FCMED algorithm improves clustering on outlier-ladent data sets, where the clusters are generated by heavy-tailed distributions.

Fuzzy medians are a special case of weighted medians, where the weights associated with the data points may be interpreted as memberships. According to Bloomfield and Steiger [2], weighted medians were first named by Edgeworth [3] (circa 1887). The 1FCM clustering algorithm requires the minimization of a functional that consists of the weighted sum of absolute differences with respect to the clustering center. Jajuga [4] seems to be the first to have formulated the 1FCM minimization as a regression problem, which then allowed him to apply the solution found in [2] attributed to Laplace (circa 1789). The optimal cluster center is the weighted median, although Jajuga [4], [5] does not seem to mention that his solution is the weighted median. The fuzzy median set forth in this paper was first derived by the author in [6] and [7] and used to independently derive the 1FCM centering statistic [8]. The weighted median appears in numerous applications. For example, it is used in risk management [9] and image processing [10]. In regression, the weighted median provides a robust slope estimate [11]. Another example is in the remedian approximation to the median [12]. Fuzzy clustering using the ℓ_1 norm is *not* new and has been researched by others [1], [13]. In [13], the authors use a reformulated version of the FCM and apply a general search method to find the cluster center and memberships. In [1], the AO-1FCM is used, where the memberships are solved for explicitly as in the FCM and the cluster centers are determined by a linear programming algorithm. The k-medoid method is a *collection* of algorithms that may use the ℓ_1 metric and could include a k-median hard-clustering algorithm [14]. Unfortunately, the k-median is also another name for the k-medoid method, leading to some confusion [14, p. 72].

This paper is organized as follows. Section II contains a definition of fuzzy order statistics as well as the extension of the median and the MAD to fuzzy sets. In Section III, the quantiles are extended to fuzzy sets. The FCMED clustering algorithm is presented in Section IV. The conclusions are contained in Section V.

II. Fuzzy Median and Fuzzy MAD

Robust statistics are resistant to outliers because they are designed assuming variations to the underlying statistical distribution will occur [15]–[18]. Often, a robust statistic is rated by its breakdown point, which is loosely defined as the fraction of outliers that must be present before the statistic no longer provides a meaningful estimate. Just as the median is a robust alternative to the mean, the MAD is the robust alternative to the standard deviation. Both statistics have a high breakdown point. Throughout this section, the data samples are assumed to be one-dimensional.

The median is defined on the data set $X = \{x_1, x_2, \cdots, x_N\}$, where each element is a real number $x_i \in R$. The ordered N sample is denoted by $\{x_{(1)}, x_{(2)}, \cdots, x_{(N)}\}$, where $x_{(1)} \le x_{(2)} \le \cdots \le x_{(N)}$ are collectively defined as the order statistics [19, p. 22]. Here, the median of X is defined to be $x_{(l+1)}$ if $N = 2l+1$ and to be $[x_{(l)} + x_{(l+1)}]/2$ if $N = 2l$. The median represents the halfway point of the samples, having an equal number of samples smaller and larger than itself.

Manuscript received November 11, 1996; revised July 10, 1997. This work was supported in part by independent research funds from the Naval Air Warfare Center.

The author is with the Computer Science Department, Midwestern State University, Wichita Falls, TX 76308 USA. He is on leave from the Naval Air Warfare Center, Patuxent River, 20670 MD USA.

Publisher Item Identifier S 1063-6706(99)09887-2.

Accordingly, half of the points to the left of the median must be outliers before the median is pulled toward the left, which explains why the finite breakdown point of the median is one-half [15]. For vector samples $x_k \in R^p, p > 1$, the definition is applied to each dimension of the sample and the median vector is defined to be the vector of individual medians.

To construct the MAD, take the data set X and form another data set $Y = \{|x_1 - \mathrm{med}(X)|, \cdots, |x_N - \mathrm{med}(X)|\}$, find the median of Y, and then scale it. For this paper, the MAD is defined as $\mathrm{mad}(X) = \mathrm{med}(Y)/0.6745$, where the constant 0.6745 adjusts the dispersion measure to be one when the sample is Gaussian with unit variance. Intuitively, one folds the centered data $\{x_i - \mathrm{med}(X)\}_{i=1}^N$ about zero, then finds the median of the set of positive deviations from the median. The breakdown point of the MAD is also one-half [15, pp. 105–107].

The median and the MAD are defined on crisp sets, which implicitly assumes that each data point has membership one in the set. The implicit role of the sample memberships is evident when the median m is defined as the solution of $\min_{m \in R} P_{\mathrm{crisp}}(m) = \min_{m \in R} \sum_{k=1}^N |x_k - m|$ [18, pp. 233–234]. An informal solution is found by taking the derivative of $P_{\mathbf{crisp}}(m)$ with respect to m and setting it equal to zero and multiplying though by -1 giving $\Psi_{\mathrm{crisp}}(m) = \sum_{k=1}^N \mathrm{sgn}(x_k - m) = 0$. If $N = 2l + 1$, the unique solution is $m = x_{(l+1)}$ and if $N = 2l$, the derivative is zero for any $m \in (x_{(l)}, x_{(l+1)})$. In the latter case, the root m is not unique, but is made so by arbitrarily choosing a suitable point within this interval, e.g., the average of $x_{(l+1)}$ and $x_{(l)}$. Strictly speaking, this solution is not proper since the derivative of $|x_k - m|$ at $x_k = m$ does not exist; however, it is easily repaired [18, p. 234]. Following [17], define $m = (m^* + m^{**})/2$ where $m^* = \sup\{m \mid \Psi_{\mathrm{crisp}}(m) > 0\}$ and $m^{**} = \inf\{m \mid \Psi_{\mathrm{crisp}}(m) < 0\}$, so that one avoids the problem of taking the derivative at the jump point. The informal solution is used in other sections because it is shorter and easily formalized.

The definition of fuzzy order statistics requires two sequences of real numbers: the data X and their corresponding memberships $U = \{u_1, u_2, \cdots, u_N\}$. A permutation $\mathrm{per}\{1, 2, \cdots, N\}$ of the integers $\{1, 2, \cdots, N\}$ is needed to order X. The fuzzy order statistics are collectively defined as $x_{\mathrm{per}(1)} \leq x_{\mathrm{per}(2)} \leq \cdots \leq x_{\mathrm{per}(N)}$ along with their corresponding memberships $\{u_{\mathrm{per}(1)}, u_{\mathrm{per}(2)}, \cdots, u_{\mathrm{per}(N)}\}$. Since the same permutation that ordered the data vector X is applied to U, the association of data point to its membership is retained.

The functional definition of the median generalizes to fuzzy sets. For the c-class problem, if u_{ik} is the membership of x_k in class i, then m_i solves the minimization of this weighted objective functional $\min_{m_i \in R} P_{\mathrm{fuzzy}}\ (m_i) = \min_{m_i \in R} \sum_{k=1}^N u_{ik} |x_k - m_i|$. The solution m_i is a weighted median applied to the ith fuzzy set where the weights are found in the definition of the fuzzy set $X_i = u_{i1}/x_1 + u_{i2}/x_2 + \cdots + u_{iN}/x_N$. Here, the statistic m_i is called the *fuzzy median* of the ith class. The derivative of $-P_{\mathbf{fuzzy}} m_i$ with respect to m_i is given by $\Psi_{\mathrm{fuzzy}}(m_i) = \sum_{k=1}^N u_{ik}\, \mathrm{sgn}(x_k - m_i)$ and its root m_i is the fuzzy median. When the root is not unique, it is made so by averaging the domain values where the derivative is zero. So the fuzzy median is a weighted median where the weights are the membership of the sample points in the fuzzy set. This statistic reduces to the median when the weights are equally likely.

TABLE I
EXAMPLE DATA SET X WITH CORRESPONDING MEMBERSHIP VALUES

X	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
U	0.2	0.4	0.1	0.6	0.9	0.7	0.3	0.4

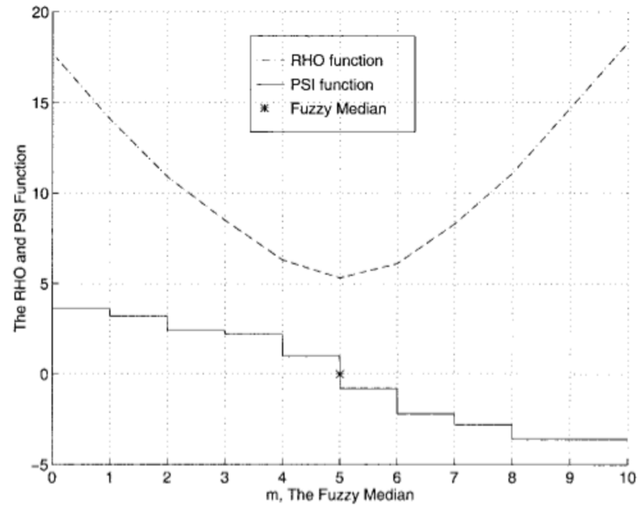

Fig. 1. (ρ) and (Ψ) functionals showing the fuzzy median value.

As an example of the fuzzy median, consider a small one-class data set with its associated membership vector given in Table I and the plots of the ρ and Ψ functions in Fig. 1. The fuzzy median is 5.0, which is *not* equal to the classical median of 4.5. In this example, the fuzzy median is unique because the root is unique.

The MAD can also be reformulated into the functional form $\min_{\eta \in R} \sum_{k=1}^N ||x_k - m| - \eta|$, that is minimized with respect to η and the resulting statistic defined as $\mathrm{mad} = \eta/0.6745$. The mad estimator requires the median m be known beforehand. For a fuzzy data set, the median does not exist; however, the fuzzy median does. For the ith fuzzy data set X_i, one can define the fuzzy MAD in terms of the fuzzy median m_i and the functional $\min_{\eta_i \in R} \sum_{k=1}^N u_{ik} ||x_k - m_i| - \eta_i|$. The fuzzy MAD is given by $\mathrm{fuzmad}_i = \eta_i/0.6745$. From an implementation point of view, 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. For the example in Table I, the MAD is 2.0 whereas the fuzzy MAD is 1.48 since the membership is highest around the central values of the sample. For a p-dimensional data, one applies it on each component separately.

Although defining the fuzzy median and fuzzy MAD is only a simple modification of the crisp statistics, it allows the important application of robust statistics to fuzzy sets. The median is also a Huber M-estimator [16] implicitly defined via functionals of the form $P(m) = \sum_{k=1}^{N} \rho(x_k - m)$, where ρ satisfies certain boundary, symmetry, and nonnegativity conditions. Because these summands can be weighted with the appropriate sample memberships, this whole class of M-estimators applies to fuzzy algorithm development.

III. Fuzzy Quantiles

In this section, the weighted M-estimator functionals are applied to derive fuzzy quantiles by defining $\rho(x)$ asymmetrically. First, the crisp quantiles are redefined using $P(m)$. Define

$$\rho(x) = \begin{cases} px, & \text{if } x > 0 \\ -qx, & \text{otherwise} \end{cases}$$

where for the sake of definiteness it is assumed that $p + q = 1$, then the pth quantile is the value of m that minimizes $P(m) = \sum_{k=1}^{N} \rho(x_k - m)$. The minimum of $P(m)$ is found by taking derivatives. Define $\Psi(m) = -P'(m) = \sum_{k=1}^{N} \rho'(x_k - m) = \sum_{k=1}^{N} \psi(x_k - m)$ where $\rho'(x)$ is defined in terms of indicator functions

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

Then $\psi(x) = \rho'(x) = pI_{\{x>0\}} + \frac{1}{2}(p-q)I_{\{x=0\}} - qI_{\{x<0\}}$, is a step function located at $x = 0$. If $p + q = 1$, the jump size is one, going from $-q$ to $+p$ at $x = 0$. $\Psi(m)$ is a monotone nonincreasing function that starts at pN when $m < x_{(1)}$ and reaches $-qN$ when $m > x_{(N)}$. Intuitively, this is a method of counting since at each data point x_k, as one moves from left to right on the real number line, the functional $\Psi(m)$ decreases by one. So, if $p = q = 1/2$, then the root of $\Psi(m) = 0$ occurs when half of the points are to the left of m and half are to the right. When $p = 1/4$, then $pN = N/4$ and $-qN = -3N/4$, the root m occurs where one-quarter of the samples are to the left and three-quarters are to the right; that is, m is the first sample quartile, ignoring the uniqueness of the root. For large N, it can be shown [19, p. 36] that if X consists of independent and identically distributed random variables (IIDRV'S) with distribution F_X, then $E(x_{(r)}) \approx F_X^{-1}(r/(N+1))$. If $r = pN$, then $E(x_{(r)})$ is approximately equal to the pth quantile or for large N, the sample quantile approaches the population quantile.

The fuzzy quantiles are defined by modifying the functional $P(m)$ to be $P_{\text{fuzzy}}(m_i) = \min_{m \in R} \sum_{k=1}^{N} u_{ik}\rho(x_k - m_i)$ and then with $\Psi_{\text{fuzzy}}(m_i) = -P'_{\text{fuzzy}}(m_i)$ solving $\Psi_{\text{fuzzy}}(m) = \sum_{k=1}^{N} u_{ik}\psi(x_k - m_i) = 0$. Here m_i is the pth quantile of the ith set. When $p = q = 1/2$, then the root of $\Psi_{\text{fuzzy}}(m_i) = 0$ still occurs when the "number of points" to the left of m_i equals the "number of points" to the right. But in this new context, the "number of points" is interpreted to be the "*sum of their fuzzy cardinality*." The fuzzy cardinality of the points in the ith fuzzy set N_i is defined as $N_i = \sum_{k=1}^{N} u_{ik} = \sum_{k=1}^{N} u_i(x_k)$ where the total number of samples is given by $N = \sum_{i=1}^{c} N_i$. So to find the fuzzy median in the set of ordered points $\{x_{\text{per}(i)}\}_1^N$, sum the corresponding memberships $\{u_{\text{per}(i)}\}_1^N$ from left to right until half of the fuzzy cardinality is to the left of median and half is to the right. In like manner, when finding the first fuzzy quartile, one-fourth of the fuzzy cardinality should be to the left of the point and three-fourths should be to the right. Viewed in this manner, fuzzy quantiles possess the same strong intuitive appeal as their crisp counterparts. For the same data given in Table I. Here N_i is 3.6 so $pN_i = 0.9$ and $-qN_i = -2.7$. Again, the fuzzy sample–quartile value of four is *not* the same as the crisp sample–quartile value of two, if the definition of pth sample quantile [19, p. 41] is $X_{(r)}$, $r = \lceil Np \rceil$. To maintain consistency with the fuzzy definition of quantile, the sample quantile convention adopted here is $X_{(r)}$ if pN is not an integer and $[X_{(r)} + X_{(r+1)}]/2$ if pN is an integer. Then if the data memberships are all 1.0, the fuzzy quartile and the crisp quartile will coincide and for the example of Table I, the quartile will be 2.5.

IV. Fuzzy Clustering

The FCMED clustering algorithm is presented after first stating the FCM clustering algorithm. As with the FCM, the FCMED algorithm obtains by first minimizing the objective functional with respect to the MF's and then with respect to the centering statistic. The MF's for the FCMED are stated and the centering statistic shown to be the fuzzy median.

A. Fuzzy c-Means (FCM) Clustering Algorithm

The FCM is a practical clustering algorithm that generalizes the crisp c-means algorithm [20], [21]. It generalizes 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 a fuzzy unit vector (fit vector) associated with it, rather than a single class assignment. The algorithm is an unsupervised learning technique. The following description of the FCM is based on [20].

Consider N data samples forming the data set denoted by $X = \{x_1, x_2, \cdots, x_N\}$, where each sample $x_i \in R^p$ is a p-dimensional real vector. Assume there are c classes and $u_{ik} = u_i(x_k) \in [0, 1]$ is the membership of the kth sample in the ith class. Each sample point x_k satisfies the constraint that $\sum_{i=1}^{c} u_{ik} = 1$. The set of exemplars or prototypes for the c clusters is given by $v = (v_1, v_2, \cdots, v_c)$. The FCM algorithm minimizes the functional

$$J(U, v) = \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik}^{m_c} d_{ik}^2 \quad \text{where } d_{ik} = \|v_i - x_k\|_2$$

subject to the above constraint. The AO method is one technique to achieve the minimum. The power m_c of the membership is called the weighting exponent. Using the memberships U, class exemplars are calculated from the data points. The class exemplars are then used to calculate new memberships. This procedure is repeated until some form of convergence occurs. A detailed version of this algorithm is given in [20, p. 66]. The FCM exemplars are linear statistics or weighted averages of the data points where the weights are scaled versions of the memberships. Unfortunately, linear statistics are known to be vulnerable to outliers [22].

B. Fuzzy c-Medians (FCMED) Clustering Algorithm

For the FCMED, the ℓ_1 objective functional is [9]

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

$$d_{ik} = \|v_i - x_k\|_1 = \sum_{j=1}^{p} |x_k(j) - v_i(j)|$$

where $\|\cdot\|_1$ is the ℓ_1 metric that is used throughout this subsection. Following [20, pp. 65–69], the derivation for the weight u_{ik} carries through with d_{ik}^2 replaced by d_{ik}. The optimal memberships are then given by

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

for the samples that do not fall on the exemplars [1, p. 547]. Samples that that fall too close to exemplars are handled in the same way as with the FCM [20]. When the optimum exemplars are sought, one is interested in minimizing $J(U,v)$ with respect to v and, in this case, one minimizes $J(U,v) = \sum_{k=1}^{N}\sum_{i=1}^{c} u_{ik}^{m_c}\sum_{j=1}^{p}|x_k(j)-v_i(j)|$ by first rewriting it as

$$J(U,v) = \sum_{i=1}^{c}\sum_{j=1}^{p} J(U, v_i(j), i, j) \quad \text{where}$$

$$J(U, v_i(j), i, j) = \sum_{k=1}^{N} u_{ik}^{m_c}|x_k(j) - v_i(j)|.$$

The functional is separable in j (the dimension) and i (the class) since each of the functions $J(U, v_i(j), i, j)$ in the objective functional $J(U,v)$ is a function of only one variable $v_i(j)$ [23, p. 8]. Hence, one minimizes $J(U,v)$ by minimizing each component $J(U, v_i(j), i, j)$ separately. For each class i and coordinate j, the minimum of $J(U, v_i(j), i, j)$ with respect to $v_i(j)$ is the fuzzy median. Jajuga [4] also argues that $J(U,v)$ is separable because $J(U, v_i(j), i, j)$ contains only one unknown $v_i(j)$. In Section II, the fuzzy median (weighted median) with memberships (weights) $u_{ik}^{m_c}$ for the i-th class and the j-th coordinate was shown to minimize $J(U, v_i(j), i, j)$. Doing this for each coordinate $j = 1, \cdots, p$ gives the centering vector v_i for class i. Repeating this for each class i, one produces the cluster exemplars v that minimizes $J(U,v)$ with respect to v. The fact that the fuzzy median (weighed median) is the optimal centering statistic for the AO-1FCM does not seem to be widely known. Although the objective functional can be minimized by a general optimization procedure [1], the fuzzy median makes the AO scheme more intuitive. When the cluster distributions are light-tailed, say Gaussian, then the asymptotic relative efficiency of the mean with respect to the median suggest that the FCM should do better than the FCMED [19, p. 283]. Here, the greatest concern is outliers, so estimation efficiency of the centering statistic is not addressed.

To compare the FCMED algorithm to the FCM, the FCMED was tested on both Gaussian and Cauchy samples. As one expects, the FCMED exemplar trajectories for the Gaussian clusters are quite similar to the FCM trajectories. However,

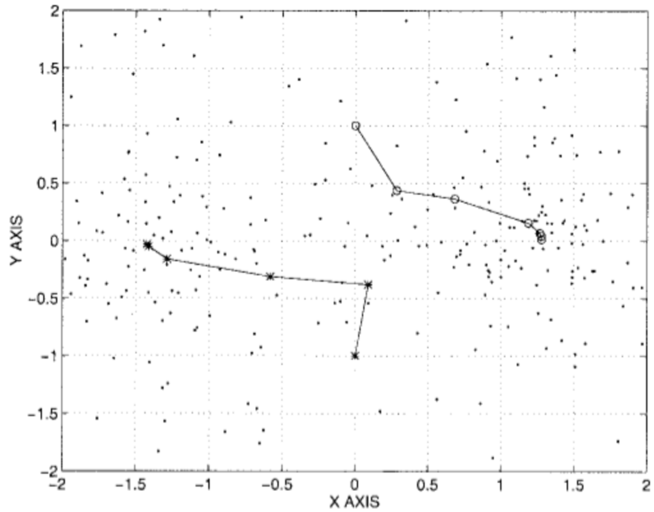

Fig. 2. FCMED exemplar traces for two Cauchy clusters located at antipodal positions (outliers not shown due to scale.)

for the Cauchy sample, the FCM does not converge to the cluster centers, while the FCMED does. Fig. 2 illustrates the FCMED applied to the two-dimensional Cauchy antipodal clusters located at $[\pm 1.27, 0]$, $m_c = 1.25$. The exemplars were initialized as $[0, \pm 1]$.

The FCMED algorithm has the same algorithmic structure as the FCM with the AO method. The FCMED algorithm follows.

1) Fix c, the number of classes such that $c \in \{2, \cdots, N-1\}$. Choose the ℓ_1 metric in R^p and fix the weighting exponent $m_c \in (1, \infty]$. Initialize the membership matrix denoted by $U^{(0)}$.
2) Construct the c exemplars v_i for $i = \{1, \cdots, c\}$ by finding the fuzzy median with memberships $u_{ik}^{m_c}$ for each class. Each class exemplar v_i is p-dimensional so $v_i(j)$ must be found for each $j = \{1, \cdots, p\}$, using just the j-th component of x_k.
3) Update the memberships u_{ik} in the membership matrix with
$$u_{ik} = 1 \Bigg/ \left[\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{1/(m_c-1)}\right]$$
provided of course that none of the d_{jk} are zero. In the latter case, the u_{ik} are assigned as they are in the FCM algorithm [20, p. 66].
4) Compare the last two membership matrices, $U^{(l)}$ and $U^{(l+1)}$. When they are sufficiently close, terminate the algorithm; otherwise, return to step 2.

Note the strong structural similarity of the FCMED and FCM algorithms. The fuzzy median may be calculated by sorting the sample values. In this case, the time complexity for each exemplar v_i is easily shown to be $O(pN \log N)$, since for each of the p-dimensions of the sample vectors it takes $O(N \log N)$ operations to sort the data. There are c classes so the time complexity for Step 2 is $O(cpN \log N)$. The space complexity is $O(N)$, which for large data sets like images can be quite onerous. More refined algorithms for calculating the weighted median can reduce the time complexity [24, p. 193] and approximations to the fuzzy median can reduce the space complexity [25]. A heavy computational price is paid to replace the FCM with the FCMED.

V. Conclusion

The fuzzy median was defined and shown to be weighted median where the weights may be interpreted as memberships. Functional definitions of the median and the MAD provided the formulation to extend these statistics to fuzzy sets. By weighting the functionals with the memberships, both statistics naturally extend to fuzzy data sets. The quantiles were extended to the fuzzy data sets using the same approach of explicitly weighting the defining functionals. The intuitive appeal of the fuzzy quantiles is retained by interpreting counting as summing memberships. The fuzzy median and the fuzzy quartile were illustrated in separate examples. The AO-1FCM is a special case of the FCM clustering algorithm that uses an alternating optimization method and the ℓ_1 norm. In this case, the cluster exemplars are shown to be the fuzzy medians and the resulting algorithm called the fuzzy c-medians (FCMED) clustering algorithm because of its strong similarity to the FCM. This fuzzy median was linked to Jajuga's solution formulation for the cluster exemplars as a regression problem, which yielded the weighted median as the cluster center via the work of Laplace. The FCM and the FCMED clustering algorithms have similar performance for light-tailed clusters, but quite dissimilar performance on heavy-tailed clusters. Both algorithms quickly converge when the data is light-tailed and the number of clusters is fixed. Outliers or heavy-tailed clusters that cause convergence problems for the FCM, are better handled by the FCMED. When the data is unknown or not well behaved, the FCMED is a robust alternative to the FCM with a heavy computational penalty.

Acknowledgment

The author would like to thank the editor and all the reviewers for their helpful suggestions and recommended references.

References

[1] 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.
[2] P. Bloomfield and W. Steiger, *Least Absolute Deviations, Theory, Applications and Algorithms*. Boston, MA: Birkhauser Boston, 1983.
[3] F. Edgeworth, "A new method of reducing observations relating to several quantities," *Phil. Mag., 5th Ser.*, vol. 24, pp. 222–223, 1887.
[4] K. Jajuga, "L_1 norm based fuzzy clustering," *Fuzzy Sets Syst.*, vol. 39, pp. 43–50, 1991.
[5] K. Jajuga, "Optimization in fuzzy clustering," *Contr. Cybern.*, vol. 24, no. 4, pp. 409–419, 1995.
[6] P. Kersten, "The fuzzy median and the fuzzy MAD," in *Proc. ISUMA-NAFIPS'95*, College Park, MD, Sept. 1995, pp. 85–88.
[7] P. Kersten, "Fuzzy robust statistics for application to the fuzzy c-means clustering algorithm," NUWC-NPT Tech. Rep. 10 298, Dec. 1993.
[8] P. Kersten, "Fuzzy order statistics and its application to fuzzy clustering," NAWCWPNS Tech. Paper 8274, Aug. 1995.
[9] K. Engemann, H. Miller, and R. Yager, "Decision making using the weighted median applied to risk management," in *Proc. ISUMA-NAFIPS'95*, College Park, MD, Sept. 1995, pp. 9–13.
[10] E. Dougherty and J. Astola, *An Introduction to Nonlinear Image Processing*. Bellingham, WA: SPIE Opt. Engrg., 1994, vol. TT16.
[11] D. Birkes and Y. Dodge, *Alternative Methods of Regression*. New York: Wiley, 1993.
[12] P. Rousseeuw and G. Bassett Jr., "The remedian: A robust averaging method for large data sets," *J. Amer. Assoc.*, vol. 85, no. 409, pp. 97–104, Mar. 1990.
[13] R. Hathaway and J. Bezdek, "Optimization of clustering criteria by reformulation," *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 241–245, May 1995.
[14] L. Kaufman and P. Rousseeuw, *Finding Groups in Data, an Introduction to Cluster Analysis*. New York: Wiley, 1990.
[15] F. Hampel, P. Rousseeuw, E. Ronchetti, and W. Stahel, *Robust Statistics, The Approach Based on Influence Functions*. New York: Wiley, 1986.
[16] P. Huber, *Robust Statistics*. New York: Wiley, 1981.
[17] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.
[18] R. Randles and D. Wolfe, *Introduction to the Theory of Nonparametric Statistics*. New York: Wiley, 1979.
[19] J. Gibbons, *Nonparametric Statistical Inference*. New York: McGraw-Hill, 1971.
[20] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.
[21] J. Bezdek and S. Pal, *Fuzzy Models for Pattern Recognition, Methods That Search for Structures in Data*. New York: IEEE Press, 1992.
[22] D. Hoaglin, F. Mosteller, and J. Tukey, *Understanding Robust and Exploratory Data Analysis*. New York: Wiley, 1983.
[23] S. Jacoby, J. Kowalik, and J. Pizzo, *Iterative Methods for Nonlinear Optimization Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1972, p. 8.
[24] T. Corman, C. Leiserson, and R. Rivest, *Introduction to Algorithms*. New York: McGraw-Hill, 1989.
[25] P. Kersten, "Implementing the fuzzy c-medians clustering algorithm," in *Proc. IEEE Conf. Fuzzy Syst.*, Barcelona, Spain, July 1997, pp. 957–962.