Journal of Global Optimization (2005) 33: 597–615
DOI 10.1007/s10898-004-7390-0

A Global Optimization RLT-based Approach for Solving the Fuzzy Clustering Problem

HANIF D. SHERALI and JITAMITRA DESAI
Grado Department of Industrial and Systems Engineering (0118), Virginia Polytechnic Institute and State University, 250 Durham Hall, Blacksburg, VA 24061, USA
(E-mails: hanifs@vt.edu; jidesai@vt.edu)

(Received 11 March 2004; accepted 4 December 2004)

Abstract. The field of cluster analysis is primarily concerned with the partitioning of data points into different clusters so as to optimize a certain criterion. Rapid advances in technology have made it possible to address clustering problems via optimization theory. In this paper, we present a global optimization algorithm to solve the *fuzzy* clustering problem, where each data point is to be assigned to (possibly) several clusters, with a membership grade assigned to each data point that reflects the likelihood of the data point belonging to that cluster. The fuzzy clustering problem is formulated as a nonlinear program, for which a tight linear programming relaxation is constructed via the Reformulation-Linearization Technique (RLT) in concert with additional valid inequalities. This construct is embedded within a specialized branch-and-bound (B&B) algorithm to solve the problem to global optimality. Computational experience is reported using several standard data sets from the literature as well as using synthetically generated larger problem instances. The results validate the robustness of the proposed algorithmic procedure and exhibit its dominance over the popular fuzzy c-means algorithmic technique and the commercial global optimizer BARON.

Key words: clustering problem, fuzzy clustering, fuzzy c-means algorithm, global optimization, Reformulation-Linearization Technique

1. Introduction

Amongst the many areas in which optimization has proven to be an invaluable tool, one notable application is that of cluster analysis. Broadly defined, clustering is the process of partitioning a set of data points, $i = 1, \ldots, n$, into subsets, $j = 1, \ldots, c$, called *clusters*, such that some distance measure is minimized (Sultan et al., 2002). Clustering problems of this type arise in a host of applications related to cellular manufacturing, medicine, archaeology, finance, and marketing (see Hartigan, 1975 for a detailed survey on applications of cluster analysis).

Specifically, there are two different types of clustering problems that have been addressed in the literature: the *hard clustering* problem, wherein a data point is to be assigned to *exactly* one cluster, and the *fuzzy clustering problem* (FCP), which, in contrast, addresses the issue of assigning a data point to one or more clusters along with a designation of a membership grade for each assignment that represents the likelihood of the data point belonging to that cluster. (Here, the word fuzzy is derived from fuzzy programming, and reflects the fact that the specific cluster to which a data point belongs is only fuzzily known, and is not deterministic.) It has been shown in literature that the clustering problem is NP-complete (refer Mangiameli et al., 1996), and indeed, finding a global optimum to this problem is a computationally onerous task.

A first attempt to solve the FCP is credited to Dunn (1973). Subsequently, Bezdek (1981) generalized Dunn's algorithm and developed a more comprehensive iterative procedure, popularly known as the *fuzzy c-means algorithm* (FCMA). Given a set of heuristically prescribed initial cluster centers, the FCMA first computes the membership grade for each data point based on the relative distance measures, and then revises the cluster centers using these resulting membership grades as fixed input quantities. This process is iteratively repeated until no further improvement in the objective function is obtained. However, it has been observed that FCMA often produces local minima and/or suboptimal clustering of the given data. Consequently, several recent algorithms that have appeared in literature to solve the FCP are essentially modifications and improvements of the FCMA (refer Kamel and Selim, 1994). Furthermore, although some of these iterative procedures are guaranteed to converge to optimality under certain assumptions, this convergence can be slow in practice. Other issues such as the validity of the clusters determined by the FCMA (Roubens, 1982; Windham, 1982; Zahid et al., 1999), the geometric shape of the clusters produced (Windham, 1983), and a demonstration that the FCMA can at best produce only local optima (Ismail and Selim, 1986), are also addressed in the literature. As an aside, note that the FCMA produces cluster regions that are always spherical in shape. Gustafson and Kessel (1979) found that replacing the traditional Euclidean distance objective function criterion by another measure formulated from a symmetric, positive semidefinite matrix, yields elliptical clusters when solved via a modified version of the FCMA. Gath and Geva (1989) further generalized this concept by taking into account the size and density of the clusters as well.

A comprehensive survey of fuzzy cluster analysis that is specifically aimed at pattern recognition problems is presented by Baraldi and Blonda (1999a,b), and the use of evolutionary algorithms for fuzzy clustering is discussed by Klawonn and Keller (1998). Also, Mirkin (1996) describes several heuristic procedures such as the *Ideal Type Fuzzy Clustering* method, wherein determining the membership grades for each data point is posed as a problem of computing a vector corresponding to the maximum singular value of the given data matrix. However, despite this notable literature dedicated to solving the FCP, there exist only a limited number of global optimization procedures, such as the algorithms designed via fuzzy set theory as proposed by Ruspini (1973) and Guoyao (1998). This motivates us to consider alternative effective and robust global optimization approaches for solving the FCP.

As a precursor to the present work, we note that Sherali and Desai (2005) have developed a global optimization approach to solve the hard clustering problem based on the Reformulation-Linearization Technique (RLT) (Sherali and Adams, 1990, 1994, 1999; Sherali and Tuncbilek, 1992; Sherali and uncbilek, 1997). Here, the hard clustering problem is formulated as a nonlinear, discrete optimization problem, which is subsequently transformed into an equivalent 0–1 mixed-integer program having a tight linear programming (LP) relaxation as prescribed by the RLT, and a specialized algorithm is designed to derive a global optimum.

In this research effort, we again apply the RLT to develop an effective global optimization algorithm for solving the FCP. However, this approach is completely different from that for the hard clustering case because of the modified structure of the present problem. The remainder of this paper is organized as follows. The FCP is formulated as a nonlinear program in Section 2, and a tight LP relaxation is derived as prescribed via the RLT methodology. Accordingly, the reformulated problem is then embedded in a specialized branch-and-bound (B&B) algorithm along with a branching rule that ensures global convergence, in the spirit of Sherali and Tuncbilek (1992). Section 3 presents computational results using certain standard test problems from the literature as well as using larger synthetically generated data sets, and explores the performance of different formulations. Finally, Section 4 concludes the paper with a summary and a discussion on further avenues for research in this area.

2. Fuzzy clustering problem

The fuzzy clustering problem can be defined as follows. Given a set of n data points, each having some s attributes, we are required to assign each of these points to one or more of some c clusters (where c is given). In this process, we are also required to specify for each assignment a membership grade that represents the likelihood of the data point belonging to that cluster. The objective criterion is to minimize the total weighted squared Euclidean distances of the data points from the centroids of the assigned clusters.

Mathematically, this FCP can be stated as follows:

$$\textbf{FCP}: \quad \text{Minimize} \quad \sum_{i=1}^{n}\sum_{j=1}^{c} w_{ij}^2 \left\|a_i - z_j\right\|^2, \tag{1a}$$

$$\text{subject to} \quad \sum_{j=1}^{c} w_{ij} = 1, \quad \forall i = 1, \dots, n, \tag{1b}$$

$$w_{ij} \geqslant 0, \quad \forall (i, j), \tag{1c}$$

where $a_i \equiv (a_{ik}, k=1,\ldots,s)^{\mathrm{T}}$ is the location descriptor for the data point i, $z_j \equiv (z_{jk}, k=1,\ldots,s)^{\mathrm{T}}$ is the centroid of the to-be-determined cluster j, w_{ij} is the membership grade associated with a data point i when assigned to a cluster j, and the norm $\|\ \|$ in (1a) represents the Euclidean distance between the two points in its argument in the s-dimensional space under consideration.

Note that, in general, the objective function for the FCP is sometimes expressed as $\sum_{i=1}^{n}\sum_{j=1}^{c} w_{ij}^{m}\left\|a_i - z_j\right\|^2$, where m represents the *degree* of fuzziness, with the notion that m is increased as the desired extent of fuzziness in the problem increases. Given a data set, the choice of m, also called the *fuzzifier*, is largely dependent on the separation between the clusters. For example, if the data set contains clusters that are far apart, then the data points can be crisply divided into various clusters, thereby leading to the hard clustering problem, with $m=1$, and the associated membership grade for each data point turns out to be either 0 or 1. Conversely, for data sets containing clusters that are indistinguishable, a large value of m ought to be prescribed. Indeed, as $m \to \infty$, it is observed that the membership grade for each data point approaches $1/c$ (refer Höppner et al. (1999) for a general discussion on this subject). In our research, we have adopted the most commonly used value for m, namely $m=2$. Observe that, unlike as in the case of hard clustering, the w-variables can now fractionate, thereby reflecting the fuzziness with which each data point i is assigned to different clusters. Also, consistent with the optimization approach adopted in this paper, we note that for solving problems having a higher degree of fuzziness, some suitable pseudo-global optimization approach coupled with factorable programming techniques might be gainfully employed (see Sherali and Wang, 2001; Sherali and Ganesan, 2003).

Now, note that for a fixed w in Problem FCP, optimality in z requires that

$$\sum_{i=1}^{n} w_{ij}^2 (z_{jk} - a_{ik}) = 0, \quad \forall (j,k). \tag{2}$$

This yields,

$$z_{jk} = \frac{\sum_{i=1}^{n} w_{ij}^2 a_{ik}}{\sum_{i=1}^{n} w_{ij}^2}, \quad \forall (j,k), \quad \text{i.e., } z_j = \sum_i \lambda_i a_i, \quad \text{where } \lambda_i = \frac{w_{ij}^2}{\sum_{i=1}^{n} w_{ij}^2}, \quad \forall i. \tag{3}$$

Hence, each cluster centroid z_j is a convex combination of the vectors a_i for which $w_{ij} > 0$ (since $\sum_{i=1}^{n} \lambda_i = 1, \lambda_i\ 0$, $\forall i$ in (3)). With this motivation, let us define a (conveniently derived) superset approximation to the convex hull of all the data points a_i, $i = 1, \ldots, n$, as given by the inequalities

$$\sum_{k=1}^{s} \gamma_{qk}\xi_k \leqslant \gamma_{q0}, \quad \forall q = 1, \ldots, Q. \tag{4}$$

Accordingly, we can impose the restrictions

$$\sum_{k=1}^{s} \gamma_{qk} z_{jk} \leqslant \gamma_{q0}, \quad \forall q = 1, \ldots, Q \quad \text{for each } j. \tag{5}$$

Remark 1. Let us denote the convex hull of the data points $a_i, i = 1, \ldots, n$, as $\Lambda = conv\{a_i, i = 1, \ldots, n\}$. In the simplest case, Λ might be taken as an enclosing hyper-rectangle as expounded below. Note that Λ is efficiently computable in polynomial time for points in two-dimensions ($s = 2$) using the method described in Manber (1989). (For example, Graham's scan algorithm produces the convex hull in $O(n \log n)$ steps.) However, for higher dimensions, deriving Λ can prove to be an expensive task, although, under some specific assumptions, it has been shown that this set can be obtained using techniques such as neural networks (refer Leung et al., 1997), cutting planes (Chazelle, 1991), and direct convex hull computations for convex polyhedra (refer Klapper, 1987; Balas, 1988). In the context of our problem, we can gainfully employ any such technique to generate suitable valid inequalities for constructing a superset $\bar{\Lambda}$ of Λ. For simplicity, regardless of problem dimension, we will take $\bar{\Lambda}$ to be a hyper-rectangle that bounds the collection of points $a_i, i = 1, \ldots, n$, as defined below.

$$\bar{\Lambda} = \left\{ z_j : \alpha_k^j \leqslant z_{jk} \leqslant \beta_k^j, k = 1, \ldots, s \right\},$$

where, $\alpha_k^j = \min\{a_{ik} : i = 1, \ldots, n\}, \forall k$, and $\beta_k^j = \max\{a_{ik} : i = 1, \ldots, n\}, \forall k$, for each j. Additionally, we could incorporate within the definition of $\bar{\Lambda}$ other valid inequalities that are valid for Λ. In order to maintain generality in presentation of these various viable algorithmic strategies, we will henceforth assume that some such suitable set $\bar{\Lambda}$ as designated by (4) has already been obtained.

Furthermore, given (2), the quartic objective function (1a) can be reduced to a cubic polynomial as follows:

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{c} w_{ij}^2 \left\|a_i - z_j\right\|^2 &= \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{k=1}^{s} w_{ij}^2 (z_{jk}-a_{ik})^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{k=1}^{s} w_{ij}^2 (z_{jk}-a_{ik}) z_{jk} - \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{k=1}^{s} w_{ij}^2 (z_{jk}-a_{ik}) a_{ik} \\
&= \sum_{j=1}^{c}\sum_{k=1}^{s} z_{jk}\left[\sum_{i=1}^{n} w_{ij}^2 (z_{jk}-a_{ik})\right] - \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{k=1}^{s} a_{ik} w_{ij}^2 (z_{jk}-a_{ik}) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{k=1}^{s} a_{ik} w_{ij}^2 (z_{jk}-a_{ik}). \qquad (6)
\end{aligned}$$

In addition, a critical factor that can seriously inhibit the solution of FCP via a B&B approach is the *symmetry* in the problem structure. Note that for any given solution, alternative equivalent solutions could be obtained by simply reindexing each cluster composition, and a B&B algorithm could get mired in sifting through such symmetric reflections. To alleviate the related computational difficulties, we validly impart a somewhat distinctive identity to each cluster set by indexing them in nonincreasing order of their sizes. That is, we impose

$$\sum_{i=1}^{n} w_{ij} \geqslant \sum_{i=1}^{n} w_{i,j+1}, \quad \forall j = 1, \ldots, c-1. \qquad (7)$$

Using (2), (5)–(7), we can re-write FCP as follows where the bounds on w_{ij} can be initialized at $l_{ij}=0$, and $u_{ij}=1$, $\forall (i,j)$, and will be revised subsequently during the algorithmic process.

$$\textbf{FCP1}: \quad \text{Maximize} \quad \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{k=1}^{s} a_{ik} w_{ij}^2 (z_{jk}-a_{ik}), \qquad (8a)$$

$$\text{subject to} \quad \sum_{i=1}^{n} w_{ij}^2 (z_{jk}-a_{ik}) = 0, \quad \forall (j,k), \qquad (8b)$$

$$\sum_{k=1}^{s} \gamma_{qk} z_{jk} \leqslant \gamma_{q0}, \quad q = 1, \ldots, Q, \ \forall j, \qquad (8c)$$

$$\sum_{i=1}^{n} w_{ij} \geqslant \sum_{i=1}^{n} w_{i,j+1}, \quad \forall j = 1, \ldots, c-1, \qquad (8d)$$

$$\sum_{j=1}^{c} w_{ij} = 1, \quad \forall i = 1, \ldots, n, \tag{8e}$$

$$l_{ij} \leqslant w_{ij} \leqslant u_{ij}, \quad \forall (i, j). \tag{8f}$$

We now apply the RLT to FCP1by generating some special additional valid inequalities. Note that in order to curtail the size of the resulting problem obtained via this process, we will only generate RLT product constraints that contain nonlinear terms of the type that are already present within FCP1. Denoting by $(8c)_{qj}$, the particular constraint expression $\gamma_{q0} - \sum_{k=1}^{s} \gamma_{qk} z_{jk} \geqslant 0$ that appears in (8c), $\forall (q, j)$, and denoting by $[\cdot]_{\mathrm{L}}$ the linearization of an expression $[\cdot]$ under the substitution:

$$W_{ij} = w_{ij}^2, \quad x_{ijk} = w_{ij} z_{jk} \text{ and } y_{ijk} = w_{ij}^2 z_{jk}, \quad \forall (i, j, k), \tag{9}$$

we will generate the following constraints, $\forall q, \forall (i, j)$:

$$[(8c)_{qj} * (w_{ij} - l_{ij})^2]_{\mathrm{L}} \geqslant 0, [(8c)_{qj} * (u_{ij} - w_{ij})^2]_{\mathrm{L}} \geqslant 0 \quad \text{and} \quad [(8c)_{qj} * (u_{ij} - w_{ij}) * (w_{ij} - l_{ij})]_{\mathrm{L}} \geqslant 0. \tag{10}$$

Incorporating (10) within FCP1 yields the following enhanced reformulation **FCP2**, where we have now used the substitution (9) in (8a, b) as well, and where we have re-written (8c) in (10) as (11c) below for the sake of convenience in referencing. Proposition 1 below establishes the validity of this model.

$$\textbf{FCP2}: \quad \text{Maximize} \quad \sum_{i=1}^{n} \sum_{j=1}^{c} \sum_{k=1}^{s} a_{ik} y_{ijk} - \sum_{i=1}^{n} \sum_{j=1}^{c} \sum_{k=1}^{s} a_{ik}^2 W_{ij}, \tag{11a}$$

$$\text{subject to}: \quad \sum_{i=1}^{n} y_{ijk} - \sum_{i=1}^{n} a_{ik} W_{ij} = 0, \quad \forall (j, k), \tag{11b}$$

$$\sum_{k=1}^{s} \gamma_{qk} z_{jk} \leqslant \gamma_{q0}, \quad \forall q = 1, \ldots, Q, \quad \forall j, \tag{11c}$$

$$[(11c)_{qj} * (w_{ij} - l_{ij})^2]_{\mathrm{L}} \geqslant 0, \quad \forall q, i, j, \tag{11d}$$

$$[(11c)_{qj} * (u_{ij} - w_{ij})^2]_{\mathrm{L}} \geqslant 0, \quad \forall q, i, j \tag{11e}$$

$$[(11c)_{qj} * (u_{ij} - w_{ij}) * (w_{ij} - l_{ij})]_{\mathrm{L}} \geqslant 0, \quad \forall q, i, j, \tag{11f}$$

$$\sum_{i=1}^{n} w_{ij} \geqslant \sum_{i=1}^{n} w_{i,j+1}, \quad \forall j = 1, \ldots, c-1, \tag{11g}$$

$$\sum_{j=1}^{c} w_{ij} = 1, \quad \forall i = 1, \ldots, n, \tag{11h}$$

$$l_{ij} \leqslant w_{ij} \leqslant u_{ij}, \quad \forall (i, j). \tag{11i}$$

$$\text{Constraints (9).} \tag{12}$$

Note that the complicating constraints (12) are a part of FCP2; however, upper bounds will be computed by solving Problem (11a–i), without constraint (12). We will refer to this LP relaxation as Problem $\overline{\textbf{FCP2}}$. Conditions under which these relaxed constraints (12) would be satisfied by an optimum to $\overline{\text{FCP2}}$, as well as the implication of other plausible RLT constraints that could have been added to this formulation while creating only the product terms of the type (9), are addressed below.

PROPOSITION 1.
(a) *The constraints* $[(w_{ij} - l_{ij})^2]_{\text{L}} \geqslant 0$, $[(u_{ij} - w_{ij})^2]_{\text{L}} 0$, *and* $[(u_{ij} - w_{ij})(w_{ij} - l_{ij})]_{\text{L}} \geqslant 0$, $\forall (i,j)$ *are implied by* $\overline{\text{FCP2}}$.
(b) *The constraints* $[(11\text{c})_{qj} * (w_{ij} - l_{ij})]_{\text{L}} \geqslant 0$ *and* $[(11\text{c})_{qj} * (u_{ij} - w_{ij})]_{\text{L}} \geqslant 0$, $\forall (i,j)$ *are implied by* $\overline{\text{FCP2}}$.
(c) *For any feasible solution* $(\bar{w}, \bar{z}, \bar{W}, \bar{x}, \bar{y})$ *to* $\overline{\text{FCP2}}$, *if* $\bar{w}_{ij} = l_{ij}$ *or* $\bar{w}_{ij} = u_{ij}$, *then we must have* $\bar{W}_{ij} = \bar{w}_{ij}^2$, $\bar{x}_{ijk} = \bar{w}_{ij}\bar{z}_{jk}$, $\forall k$, *and* $\bar{y}_{ijk} = \bar{w}_{ij}^2 \bar{z}_{jk} \forall k$, *holding true*, *i.e., the related constraints in* (9) *or* (12) *are satisfied.*

Proof. To begin with, let us define $\alpha_k = \min\{\xi_k : \text{constraints}(4)\}$, and $\beta_k = \max\{\xi_k : \text{constraints}(4)\}$, $\forall k$. Note that α_k and β_k exist for all k since (4) defines a nonempty compact set. Moreover, we can compose surrogates of (4) composed by using multipliers equal to the optimal dual solutions to these problems to yield the restrictions $\xi_k \geqslant \alpha_k$, and $\xi_k \leqslant \beta_k$, $\forall k$. Applying this same surrogation process equivalently to (5) or (11c), we get

$$\alpha_k \leqslant z_{jk} \leqslant \beta_k, \quad \forall (j, k). \tag{13}$$

(a) To prove part (a), consider the RLT constraints $[(w_{ij} - l_{ij})^2]_{\text{L}} \geqslant 0$. Pick some k for which $\alpha_k < \beta_k$ (this must exist; else FCP is trivial). By surrogating (11d) using the same Lagrange multipliers with respect to $(11\text{c})_{qj}$ as those that produced (13), the algebra readily yields the constraints

$$[(z_{jk} - \alpha_k)(w_{ij} - l_{ij})^2]_{\text{L}} \geqslant 0 \quad \text{and} \quad [(\beta_k - z_{jk})(w_{ij} - l_{ij})^2]_{\text{L}} \geqslant 0. \tag{14}$$

Summing the constraints in (14) (in the linearized form) produces

$$[(\beta_k - \alpha_k)(w_{ij} - l_{ij})^2]_{\text{L}} \geqslant 0,$$

which implies that $[(w_{ij}-l_{ij})^2]_L \geqslant 0$ because $\alpha_k < \beta_k$. The other constraints in Part (a) are similarly implied by (11e) and (11f), respectively.

(b) If $l_{ij} = u_{ij}$, then the stated constraints are null upon fixing $w_{ij} = l_{ij} = u_{ij}$ in $\overline{\text{FCP2}}$. Hence, suppose that $l_{ij} < u_{ij}$. The constraints of Part (b) can then be obtained by summing the corresponding constraints in (11d, f) and (11e, f), respectively, and are hence implied.

(c) Finally, consider Part (c), and assume that $\bar{w}_{ij} = l_{ij}$. (The case of $\bar{w}_{ij} = u_{ij}$ is similar.) First, let us show that $\bar{W}_{ij} = \bar{w}_{ij}^2$. By Part (a), since the stated constraints are implied by $\overline{\text{FCP2}}$, we have that when $\bar{w}_{ij} = l_{ij}$

$$[(w_{ij}-l_{ij})^2]_L = [w_{ij}(w_{ij}-l_{ij})]_L - l_{ij}(w_{ij}-l_{ij}) \geqslant 0 \Rightarrow \bar{W}_{ij} \geqslant l_{ij}^2$$

and similarly,

$$[(u_{ij}-w_{ij})(w_{ij}-l_{ij})]_L = u_{ij}(w_{ij}-l_{ij}) - [w_{ij}(w_{ij}-l_{ij})]_L \geqslant 0 \Rightarrow \overline{W}_{ij} \leqslant l_{ij}^2.$$

Hence, we have $\bar{W}_{ij} = l_{ij}^2 = \bar{w}_{ij}^2$.

Next, let us show that $\bar{x}_{ijk} = \bar{w}_{ij}\bar{z}_{jk}$, $\forall k$. For any k, noting (13) and Part (b), we have that the constraints of $\overline{\text{FCP2}}$ imply the restrictions $[(z_{jk}-\alpha_k)(w_{ij}-l_{ij})]_L \geqslant 0$ and $[(\beta_k - z_{jk})(w_{ij}-l_{ij})]_L \geqslant 0$. Under the condition $\bar{w}_{ij} = l_{ij}$, these constraints respectively imply that $\bar{x}_{ijk} \geqslant l_{ij}\bar{z}_{jk}$ and $\bar{x}_{ijk} \leqslant l_{ij}\bar{z}_{jk}$, which yields $\bar{x}_{ijk} = l_{ij}\bar{z}_{jk} = \bar{w}_{ij}\bar{z}_{jk}$.

Finally, let us establish that $\bar{y}_{ijk} = \bar{w}_{ij}^2\bar{z}_{jk}$, $\forall\ k$. Again, for any k, noting (11d) and (13), we have that the corresponding surrogates of the former yield

$$\begin{aligned}&[(z_{jk}-\alpha_k)(w_{ij}-l_{ij})^2]_L \geqslant 0 \quad \text{and} \quad [(\beta_k - z_{jk})(w_{ij}-l_{ij})^2]_L \geqslant 0, \quad \text{i.e.,}\\ &[z_{jk}(w_{ij}-l_{ij})^2]_L - \alpha_k[(w_{ij}-l_{ij})^2]_L \geqslant 0 \quad \text{and}\\ &[z_{jk}(w_{ij}-l_{ij})^2]_L - \beta_k[(w_{ij}-l_{ij})^2]_L \geqslant 0.\end{aligned} \tag{15}$$

But when $\bar{w}_{ij} = l_{ij}$, we have $[(w_{ij}-l_{ij})^2]_L = W_{ij} + l_{ij}^2 - 2l_{ij}w_{ij} = 0$ since $\bar{W}_{ij} = l_{ij}^2$ from above. Hence, (15) asserts that when $\bar{w}_{ij} = l_{ij}$, we have $[z_{jk}(w_{ij}-l_{ij})^2]_L = 0$, i.e., $\bar{y}_{ijk} + \bar{z}_{jk}l_{ij}^2 - 2l_{ij}\bar{x}_{ijk} = 0$. Using $\bar{x}_{ijk} = l_{ij}\bar{z}_{jk}$ from above, this implies that $\bar{y}_{ijk} = l_{ij}^2\bar{z}_{jk} = \bar{w}_{ij}^2\bar{z}_{jk}$. This completes the proof. □

We now design a B&B algorithm for solving Problem FCP2, based on partitioning the hyperrectangle (11i) alone. For any node in this B&B tree, we compute an upper bound by solving the LP relaxation $\overline{\text{FCP2}}$ for the corresponding subproblem (i.e., $\overline{\text{FCP2}}$ with modified bounds in (11i), and hence in (11d)–(11f)). If the resulting solution $(\bar{w}, \bar{z}, \bar{W}, \bar{x}, \bar{y})$ satisfies (12), it is optimal to this subproblem. Otherwise, a heuristic solution could be computed by fixing $\bar{w}$, solving for $\bar{z}$ via (3), then fixing the resulting z-variables and solving for the w variables in (1a)–(1c) (see Remark 2 below for the relevant formulae), and so on, alternating in this fashion until the objective function value no longer improves. The node selection strategy in this process picks a node that has the greatest upper bound for further exploration. Finally, to select a branching variable, we compute the index

$$\theta_{ij}=\max\left\{\left|\bar{W}_{ij}-\bar{w}_{ij}^2\right|, \left|\bar{x}_{ijk}-\bar{w}_{ij}\bar{z}_{jk}\right| \text{for all } k, \left|\bar{y}_{ijk}-\bar{w}_{ij}^2\bar{z}_{jk}\right| \text{for all } k\right\}. \quad (16)$$

Note that by Proposition 1, if $\bar{w}_{ij}=l_{ij}$ or $\bar{w}_{ij}=u_{ij}$, then we have $\theta_{ij}=0$. Also, if (12) is satisfied, then $\theta_{ij}=0$, $\forall(i,j)$. Else, we select $\theta_{pq}\equiv\underset{(i,j)}{\arg\max}\{\theta_{ij}\}>0$, which means that $l_{pq}<\bar{w}_{pq}<u_{pq}$. The node subproblem is then split by imposing the dichotomy that

$$l_{ij}\leqslant w_{ij}\leqslant\bar{w}_{ij} \quad \text{or} \quad \bar{w}_{ij}\leqslant w_{ij}\leqslant u_{ij}. \quad (17)$$

Infinite convergence to a global optimum (in case finite termination does not occur) follows from Sherali and Tuncbilek (1992), noting Proposition 1.

Remark 2. For a fixed $z=\bar{z}$, Problem (1a)–(1c) can be solved for an optimal value of w as follows.

PROPOSITION 2. *Let $z_j=\bar{z}_j$ be fixed for all j in Problem* FCP ((1a)–(1c)). *Then an optimal corresponding solution $\bar{w}$ to* FCP *is obtained as follows:*

For each i, if $\bar{z}_r=a_i$ for some r, then set $\bar{w}_{ir}=1$, and $\bar{w}_{ij}=0$ for all $j\neq r$ (18a)

otherwise,

$$\text{let } \bar{w}_{ij}=\frac{\bar{\pi}_i}{\left\|a_i-\bar{z}_j\right\|^2}, \quad \forall j, \quad \text{where } \bar{\pi}_i=\frac{1}{\sum_{j=1}^{c}\left(1/\left\|a_i-\bar{z}_j\right\|^2\right)}. \quad (18b)$$

Proof. For $z=\bar{z}$ fixed, Problem FCP is a linearly constrained convex program for which the KKT conditions are both necessary and sufficient. Denoting π_i and μ_{ij} as the Lagrange multipliers associated with (1b) and (1c), respectively, $\forall i,j$, these conditions require that (where we have denoted $c_{ij}\equiv\left\|a_i-\bar{z}_j\right\|^2$, $\forall i,j$, and equivalently written the objective function as: Minimize $1/2\sum_{i=1}^{n}\sum_{j=1}^{c}c_{ij}w_{ij}^2$):

$$\sum_{j=1}^{c}w_{ij}=1, \quad \forall i, \quad w\geqslant 0, \quad (19a)$$

$$c_{ij}w_{ij}-\pi_i-\mu_{ij}=0, \quad \mu_{ij}w_{ij}=0, \quad \mu_{ij}\geqslant 0, \ \forall(i,j). \quad (19b)$$

Consider any $i \in \{1, \ldots, n\}$. Let us first show that we must have $\mu_{ij} = 0$, $\forall j$, in (19b). If any $\mu_{ij} > 0$, then (19b) implies that $w_{ij} = 0$, and so $\pi_i = -\mu_{ij} < 0$. But (19a) requires that $w_{ir} > 0$ for some r, and (19b) asserts that μ_{ir} must be zero for this (i, r), which means that we should have $c_{ir} w_{ir} = \pi_i$. Since $c_{ir} \geqslant 0$, this contradicts that $\pi_i < 0$. Hence, $\mu_{ij} = 0$, $\forall j$ in (19b).

Consequently, if $c_{ir} = 0$ (i.e., $\bar{z}_r = a_i$) for any r, then by (19b), we will have $\pi_i = 0$ and the KKT conditions (19a, b) are satisfied by selecting $w_{ij} = \bar{w}_{ij}$ for all j as specified in (18a). On the other hand, if we have $c_{ij} > 0$, $\forall j$, we have from (19b) that $w_{ij} = \pi_i / c_{ij}$, $\forall j$, and using (19a), we obtain $\pi_i = \bar{\pi}_i$ and $w_{ij} = \bar{w}_{ij}$, $\forall j$, as given by (18b). This completes the proof. □

Remark 1. Note that the proof of Proposition 2 asserts that we could impose the constraints

$$w_{ij} \sum_{k=1}^{s} (z_{jk} - a_{ik})^2 = \pi_i, \quad \forall (i, j), \tag{20}$$

within the reformulation of FCP2. However, doing so would produce new nonlinear terms other than those in (9) that would require additional supporting RLT constraints involving the pairwise products of (11c), or the pairwise products of the surrogated implied constraints (13), multiplied by the corresponding bound factors $(w_{ij} - l_{ij})$ and $(u_{ij} - w_{ij})$, $\forall (i, j)$. To avoid this increase in size, we do not include (20) explicitly, and permit FCP2 itself to implicitly attain these conditions ultimately.

3. Computational Results

In this section, we provide computational results using some standard data sets from the literature as well as randomly generated problems to compare the proposed global optimization approach with the popularly implemented FCMA heuristic as well as the commercial global optimizer BARON (see Sahinidis, 1996). Throughout, we will use the following terminology:

- Z_0: Objective function value of FCP corresponding to the heuristic solution found at node zero.
- Z^*: Optimal objective function value of FCP, evaluated at the optimal solution to FCP2 ($\equiv -\nu$[FCP2]).
- Z^*_{FCMA}: Best objective function value of FCP obtained via the FCMA procedure of Bezdek (1981).
- Z^*_{FCP-B}: Best objective function value of FCP obtained via BARON (see Sahinidis, 1996).
- $Z^*_{\mathrm{FCP1-B}}$: Best objective function value of FCP produced by solving FCP1 via BARON.

CPU*: CPU time required to determine a global optimum for FCP2 via the proposed B&B algorithm.
CPU_{FCMA}: CPU time required for the FCMA heuristic procedure.
$\text{CPU}_{\text{FCP-B}}$: CPU time required by BARON to solve Problem FCP.
$\text{CPU}_{\text{FCP1-B}}$: CPU time required by BARON to solve Problem FCP1.
CPU_0: CPU time required to determine a heuristic solution at node zero via the solution to $\overline{\text{FCP2}}$.

To test our proposed methodology, we first used the following standard data sets given in Späth (1980):

1. **Data Set 1**. This is a set of Cartesian coordinates for 22 German towns, which yields a clustering problem having 22 points in a two-dimensional space.
2. **Data Set 2**. This is a set of Cartesian coordinates for 59 German towns, which yields a clustering problem having 59 points in a two-dimensional space.
3. **Data Set 3**. This pertains to 89 postal zones in Germany, where each zone has three attributes, namely, surface area (measured in square kilometers), population, and the density of the population. This yields a clustering problem having 89 points in a three-dimensional space.
4. **Data Set 4**. This is also based on the 89 postal zones of Data Set 3, but considers four attributes, namely, the number of self-employed people, civil servants, clerks, and manual workers. This yields a clustering problem having 89 points in a four-dimensional space.

The proposed B&B algorithm for solving FCP2 was implemented in C++, and the commercial solver CPLEX 8.1.0 was invoked for the purpose of solving the LP relaxations at each node. Furthermore, for modeling our problem, the constraints (11c) in FCP2 that represent a superset of the convex hull of the data points were generated by simply constructing a tightest hyperrectangle that encloses the data points. Also, for benchmarking our results, we coded the FCMA procedure in C++, and executed this method with a prescribed termination tolerance of $\varepsilon = 10^{-3}$. Tables 1 and 2 present the relative performance of the proposed algorithm versus the FCMA procedure, measured in terms of various statistics, for three and five cluster centers, respectively.

Note that, on an average, the reformulated problem FCP2 required only 14.05% and 9.87% of the time taken by the FCMA, while producing optimal solutions that further improve the FCMA solutions by 69.32% and 74.88%, for the respective cases of three and five cluster centers. Indeed, from the results in Tables 1 and 2, it can be seen that the (heuristic) solution obtained at node zero itself was uniformly better than that prescribed by the FCMA, yielding an average improvement of 26.77% and 38.93%,

Table 1. Relative performance of the proposed optimization approach versus the FCMA procedure for three cluster centers.

Data sets	Parameters					
	$\frac{z_0}{z^*}$	$\frac{z^*_{FCMA}}{z^*}$	$\frac{z^*_{FCMA}}{z_0}$	CPU*(s)	$\frac{CPU^*}{CPU_{FCMA}}$	$\frac{CPU_0}{CPU_{FCMA}}$
1	2.25	2.48	1.10	0.140	0.254	0.084
2	2.43	2.99	1.23	0.274	0.236	0.078
3	2.19	4.03	1.84	0.288	0.110	0.036
4	2.68	3.55	1.32	0.410	0.114	0.037
Averages	2.387	3.26	1.37	0.278	0.140	0.046

Table 2. Relative performance of the proposed optimization approach versus the FCMA procedure for five cluster centers.

Data sets	Parameters					
	$\frac{z_0}{z^*}$	$\frac{z^*_{FCMA}}{z^*}$	$\frac{z^*_{FCMA}}{z_0}$	CPU* (s)	$\frac{CPU^*}{CPU_{FCMA}}$	$\frac{CPU_0}{CPU_{FCMA}}$
1	2.15	3.22	1.50	0.166	0.152	0.088
2	2.20	4.43	2.01	0.300	0.141	0.082
3	3.27	5.73	1.75	0.414	0.073	0.042
4	3.42	4.68	1.37	0.600	0.099	0.057
Averages	2.76	4.52	1.64	0.37	0.116	0.067

for three and five cluster centers, respectively. Furthermore, from the column of values CPU_0/CPU_{FCMA} in Tables 1 and 2, it can be observed that this node zero heuristic solution process consumed only 4.6% and 6.7% of the CPU time taken by the FCMA at an average, for three and five cluster centers, respectively, while yet producing superior solutions. Moreover, as evident from the results in this table, the global optimum further significantly improved upon the heuristic solution produced at node zero, and was derived within a reasonable computational effort.

Next, to further test the robustness of solving the reformulated problem FCP2 via the proposed approach, a comparative study was conducted by directly solving the nonlinear programs FCP and FCP1, using the commercial software GAMS/BARON software (version 2.50) (see Sahinidis, 1996). The corresponding results obtained are reported in Tables 3 and 4. Assimilating the results obtained in Tables 1–4, note that the proposed approach required only 49.20% and 54.57% of the CPU time consumed by BARON for solving FCP, and 84.24% and 90.24% of the CPU time consumed by BARON for solving FCP1, for the case of three and five cluster centers, respectively. Moreover, examining the objective function values obtained at

Table 3. Relative performance of solving problems FCP and FCP1 via BARON versus the proposed approach, for three cluster centers

		Data sets				Averages
		1	2	3	4	
FCP	$\frac{\text{CPU}^*}{\text{CPU}_{\text{FCP-B}}}$	1.166667	0.668293	0.355556	0.445652	0.492
	$\frac{z^*_{\text{FCP-B}}}{z^*}$	3.00	3.71	3.40	3.91	3.505
FCP1	$\frac{\text{CPU}^*}{\text{CPU}_{\text{FCP1-B}}}$	0.5645	0.9163	0.9350	0.8541	0.8424
	$\frac{z^*_{\text{FCP1-B}}}{z^*}$	1.06	1.83	2.10	2.42	1.852

Table 4. Relative performance of solving problems FCP and FCP1 via BARON versus the proposed approach, for five cluster centers

		Data sets				Averages
		1	2	3	4	
FCP	$\frac{\text{CPU}^*}{\text{CPU}_{\text{FCP-B}}}$	0.922	0.5660	0.4224	0.5870	0.5457
	$\frac{z^*_{\text{FCP-B}}}{z^*}$	2.87	3.55	3.4	4.02	3.46
FCP1	$\frac{\text{CPU}^*}{\text{CPU}_{\text{FCP1-B}}}$	0.6916	0.8219	0.9627	1.00	0.9024
	$\frac{z^*_{\text{FCP1-B}}}{z^*}$	1.108	1.62	2.55	2.87	2.037

termination while solving problems FCP and FCP1 using BARON, it is evident that BARON consistently produced relatively inferior solutions that respectively deviate in value from optimality (as detected by our method) by factors of 3.505 and 1.852 when solving FCP and FCP1 for the case of three cluster centers, and by factors of 3.46 and 2.037 when solving FCP and FCP1, for the case of five cluster centers. The observed robustness of our approach in comparison with BARON stems from the fact that we solve linear, rather than general convex programming relaxations, which yields more reliable bounds for fathoming purposes. Nonetheless, at least in comparison with the FCMA, the solution values obtained by BARON when solving Problem FCP1 dominated the FCMA solution values.

To reinforce the efficacy of our proposed approach, we also solved several additional randomly generated problems of larger sizes, and compared the results obtained with those produced by the FCMA procedure and by BARON. The number of data points in these test instances was varied from 250 to 1000 in steps of 250, and the dimension of the space was

Table 5. Comparative results for the proposed approach versus FCMA and BARON for randomly generated problem instances having three cluster centers

Data sets	Parameters							
	$\frac{z_0}{z^*}$	$\frac{z^*_{\text{FCMA}}}{z^*}$	$\frac{z^*_{\text{FCMA}}}{z_0}$	$\frac{z^*_{\text{FCP1-B}}}{z^*}$	CPU*(s)	$\frac{\text{CPU}^*}{\text{CPU}_{\text{FCMA}}}$	$\frac{\text{CPU}_0}{\text{CPU}_{\text{FCMA}}}$	$\frac{\text{CPU}^*}{\text{CPU}_{\text{FCP1-B}}}$
(250, 2)	1.23	3.33	2.71	1.21	29.52	4.17	0.55	1.380
(500, 2)	1.60	3.85	2.41	1.40	62.15	5.45	0.43	1.281
(750, 2)	1.82	4.14	2.27	1.51	125.29	5.31	0.84	1.240
(1000, 2)	2.27	4.76	2.10	1.73	150.77	4.01	0.73	1.234
(250, 4)	1.31	1.31	1.00	1.47	116.03	11.85	1.65	1.243
(500, 4)	2.14	2.09	0.98	1.76	205.57	11.22	1.53	1.225
(750, 4)	2.87	5.58	1.94	2.03	314.83	14.42	0.53	1.217
(1000, 4)	3.39	6.30	1.86	2.30	590.27	15.70	0.43	1.210
(250, 6)	1.66	3.92	2.36	1.43	320.38	6.30	0.48	1.215
(500, 6)	2.54	5.13	2.02	1.87	388.86	9.81	0.69	1.215
(750, 6)	4.07	7.23	1.78	2.64	404.66	12.96	0.67	1.214
(1000, 6)	4.92	8.40	1.71	3.06	423.79	10.55	0.55	1.209
(250, 8)	2.12	2.06	0.97	1.75	327.05	8.70	1.80	1.231
(500, 8)	3.37	5.58	1.66	2.03	787.89	18.14	0.51	1.199
(750, 8)	4.85	6.30	1.30	2.30	984.57	14.34	0.94	1.210
(1000, 8)	5.99	3.92	0.65	1.43	1188.13	13.92	0.52	1.217
Averages	2.887	5.13	1.78	1.87	401.24	11.38	0.80	1.214

varied from two to eight, in steps of two, thereby leading to a total of $4 \times 4 = 16$ test problems, with the smallest data set having 250 points in a two-dimensional space, and the largest problem having 1000 points in an eight-dimensional space. The number of clusters (c) for each case was taken to be either three (Table 5) or five (Table 6).

From the results displayed in Tables 5 and 6, note that the FCMA procedure requires a significantly lesser CPU time as compared with the proposed exact approach, but the best solution produced by the FCMA procedure is also substantially inferior. However, the node zero heuristic solution produced by the proposed approach uniformly dominates the FCMA solution with respect to both quality and effort in most of the problem instances, with three exceptions out of the total of 32 problems, all occurring for three centers (Data Sets (250, 4), (500, 4), and (250, 8) in Table 5). On an average, to obtain a feasible solution to Problem FCP2 based on the node zero analysis alone, the CPU time required was 20% lesser than for the FCMA procedure, yet the quality of the solution was 43.2% better in terms of the objective function value for the three cluster center case. Moreover, note that although BARON comsumed about 21.4% lesser effort as compared with the proposed algorithm, it tended to produce suboptimal solutions having an 87% greater objective value for FCP. Thus, for a moderate increase in computational effort, we can obtain significantly improved global optimal solutions via the proposed methodology. A similar performance was observed for the case of five cluster centers. Note that

Table 6. Comparative results for the proposed approach versus FCMA and BARON for randomly generated problem instances having five cluster centers

Data sets	Parameters							
	$\frac{z_0}{z^*}$	$\frac{z^*_{FCMA}}{z^*}$	$\frac{z^*_{FCMA}}{z_0}$	$\frac{z^*_{FCP1}}{z^*}$	CPU*(s)	$\frac{CPU^*}{CPU_{FCMA}}$	$\frac{CPU_0}{CPU_{FCMA}}$	$\frac{CPU^*}{CPU_{FCP1-B}}$
(250, 2)	1.51	5.01	3.32	1.27	36.23	3.82	0.70	1.437
(500, 2)	1.97	5.79	2.94	1.47	72.06	4.53	0.54	1.260
(750, 2)	2.24	6.22	2.78	1.59	141.39	4.16	0.96	1.188
(1000, 2)	2.79	7.16	2.57	1.82	169.36	3.09	0.92	1.176
(250, 4)	1.61	1.97	1.22	1.55	131.22	9.72	1.09	1.193
(500, 4)	2.63	3.14	1.19	1.85	229.53	8.77	1.04	1.161
(750, 4)	3.53	8.39	2.38	2.13	349.50	11.14	0.67	1.147
(1000, 4)	4.17	9.47	2.27	2.42	651.94	11.90	0.54	1.134
(250, 6)	2.04	5.89	2.89	1.50	355.60	4.77	0.61	1.146
(500, 6)	3.13	7.71	2.46	1.97	430.79	7.45	0.87	1.152
(750, 6)	5.01	10.87	2.17	2.77	448.14	9.89	0.85	1.131
(1000, 6)	6.06	12.63	2.08	3.22	469.14	8.00	0.70	1.140
(250, 8)	2.61	3.10	1.19	1.84	362.92	6.63	1.28	1.146
(500, 8)	4.15	8.39	2.02	2.13	868.92	13.69	0.65	1.131
(750, 8)	5.97	9.47	1.59	2.42	1084.88	10.75	0.79	1.129
(1000, 8)	7.37	5.89	0.80	1.50	1308.39	10.41	0.66	1.127
Averages	3.55	7.71	2.17	1.97	444.38	8.66	0.804	1.141

other meta-heuristic procedures such as the genetic algorithm or simulated annealing could also be combined with the node zero analysis to derive enhanced quality feasible solutions via $\overline{\text{FCP2}}$, either as a stand-alone procedure or within the framework of the proposed B&B algorithm. We recommend such investigations for future research.

4. Summary and Extensions for Further Research

In this research effort we have developed an algorithm for determining a global optimum to the FCP, where the objective function seeks to minimize the total degree-two fuzzifier weighted squared Euclidean distance from each data point to the centroids of the clusters to which it is assigned, and requires an accompanying membership grade to be assigned to each data point that reflects the possibility of a data point belonging to each particular cluster. A series of enhanced reformulations of this problem were presented, augmented by optimality-induced, symmetry-defeating, and RLT-based inequalities, and a specialized B&B algorithm was designed for solving the resulting model representation. Several computational experiments were performed using standard data sets as well as synthetically generated test cases to explore the efficacy of the proposed exact solution approach, as well as to study the effectiveness of the heuristic scheme implemented at the root node. This performance was compared with the popular FCMA procedure (see Bezdek, 1981) and the commercial global optimizer BARON (see Sahinidis, 1996). The results revealed the viability and robustness of the proposed approach, and exhibited its superiority over the FCMA procedure, even as a heuristic based on the node zero analysis. The B&B algorithm also consistently dominated BARON in terms of producing significantly improved global optimal solutions with a comparable computational effort.

Although the proposed B&B algorithm produces globally optimal clusters, an important aspect that is often considered in practice is the shape of the clusters produced. Höppner et al. (1999) show that the FCMA tends to produce intersecting clusters that are spherical in shape. Since the termination criterion of the FCMA procedure holds true at the global optimum, the clusters produced by our algorithm would also be represented by intersecting spheres. Nevertheless, for future research, it might be interesting to explore controlling the shapes of the clusters produced as desired, based on imposing suitable additional restrictions on the membership grade variables.

Also, note that in practice, cluster analysis problems can involve very large data sets and, therefore, good heuristic procedures can prove to be critically important for handling such problem instances. In this context, it might be useful to investigate hybrid LP-based and meta-heuristic solution procedures. Our research suggests that designing heuristic methods based on constructs that are borrowed from strong effective exact procedures might be a prudent approach, and offers rich potential for future advances in the domain of cluster analysis.

Acknowledgment

This work has been supported by the *National Science Foundation* under Grant No. DMI - 0094462. The authors also thank Dr. Nick Sahinidis of the $\sum$-Optimization Group at the University of Illinois, Urbana Champaign, for permitting the use of the BARON software.

References

Balas, E. (1988), On the convex hull of the union of certain polyhedra, *Operations Research Letters* 7(6), 279–283.

Baraldi, A. and Blonda, P. (1999a), Survey of fuzzy clustering analysis for pattern recognition – Part I, *IEEE Transactions on Systems, Man, and Cybernetics, Part B:Cybernetics* 29(6), 778–785.

Baraldi, A. and Blonda, P. (1999b), Survey of fuzzy clustering analysis for pattern recognition – Part II, *IEEE Transactions on Systems, Man, and Cybernetics, Part B:Cybernetics* 29(6), 786–801.

Bezdek, J.C. (1981), *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York.

Chazelle, B. (1991), An optimal convex hull algorithm and new results on cuttings. In: *proceedings of the Annual Symposium on Foundations of Computer Science*, IEEE Computer Society, Silver Spring, MD, pp.29–38.

Dunn, J.C. (1973), A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics* 3(3), 32–57.

Gath, I. and Geva, A.B. (1989), Unsupervised optimal fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7), 773–781.

Guoyao, F. (1998), Optimization methods for fuzzy clustering, *Fuzzy Sets and Systems* 93, 301–309.

Gustafson, D.E. and Kessel, W.C. (1979), Fuzzy clustering with a fuzzy covariance matrix. In: *proceedings of the International Proceedings of the IEEE Conference on Decision and Control*, 78, New York, pp. 761–766.

Hartigan, J.A. (1975), *Clustering Algorithms*, Wiley, New York.

Höppner, F., Klawonn, F., Kruse, R. and Runkler, T. (1999), *Fuzzy Cluster Analysis*, Wiley, Chichester, UK.

Ismail, M.A. and Selim, S.Z. (1986), Fuzzy c-means: optimality of solutions and effective termination of the algorithm, *Pattern Recognition* 19, 481–485.

Kamel, M.S. and Selim, S.Z. (1994), New algorithms for solving the fuzzy clustering problem, *Pattern Recognition* 27(3), 421–428.

Klapper, A. (1987), Lower bound on the complexity of the convex hull problem for simply polyhedra, *Information Processing Letters* 25(3), 159–161.

Klawonn, F. and Keller, A. (1998), Fuzzy clustering based on modified distance measures. In: *proceedings of the Advances in Intelligent Data Analysis, Proceedings of the 3rd International Symposium*, *eds.*, Hand, D.J., Kok, J.N. and Berthold, K.R., 291–301, Amsterdam, The Netherlands.

Leung, Y., Zhang, J. and Xu, Z. (1997), Neural networks for convex hull computation, *IEEE Transactions on Neural Networks* 8(3), 601–611.

Manber, U. (1989), *Introduction to Algorithms: A Creative Approach*, Addison-Wesley Publishing Company, Reading, MA.

Mangiameli, P., Chen, K.S. and West, D. (1996), A comparison of SOM neural network and hierarchical clustering methods, *European Journal of Operations Research* 93, 402–417.

Mirkin, B. (1996), *Mathematical Classification and Clustering*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Roubens, M. (1982), Fuzzy clustering algorithms and their cluster validity, *European Journal of Operational Research* 10(3), 294–301.

Ruspini, E.H. (1973), New experimental results in fuzzy clustering, *Information Science* 6, 273–284.

Sahinidis, N.V. (1996), BARON: A general purpose global optimization software package, *Journal of Global Optimization* 8(2), 201–205.

Sherali, H.D. and Adams, W.P. (1990), A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems, *SIAM Journal on Discrete Mathematics* 3(3), 411–430.

Sherali, H.D. and Adams, W.P. (1994), A hierarchy of relaxations and convex hull characterizations for mixed-integer zero-one programming problems, *Discrete Applied Mathematics* 52, 83–106.

Sherali, H.D. and Adams, W.P. (1999), Reformulation-linearization techniques for discrete optimization problems, In: *Handbook of Combinatorial Optimization* 1, Du, D.Z. and Pardalos, P.M. (eds.), Kluwer Academic Publishers, Dordrecht pp. 479–532.

Sherali, H.D. and Desai, J. (2005), An RLT-based global optimization algorithm for solving the hard clustering problem, *Journal of Global Optimization* 32(2), 281–306.

Sherali, H.D. and Ganesan, V. (2003), A pseudo-global optimization approach with application to the design of containerships, *Journal of Global Optimization* 26, 335–360.

Sherali, H.D. and Tuncbilek, C. (1992), A global optimization algorithm for polynomial programming problems using a reformulation-linearization technique, *Journal of Global Optimization* 2, 101–112.

Sherali, H.D. and Tuncbilek, C.H. (1997), Comparison of two Reformulation-Linearization Technique based linear programming relaxations for polynomial programming problems, *Journal of Global Optimization* 10, 381–390.

Sherali, H.D. and Wang, H. (2001), Global optimization of nonconvex factorable programming problems, *Mathematical Programming* 89(3), 459–478.

Späth, H. (1980), *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*, Wiley, New York.

Sultan, M., Wigle, D.A., Cumbaa, C.A., Maziarz, M., Glasgow, J., Tsao, M.S. and Jurisica, I. (2002), Binary tree-structured vector quantization approach to clustering and visualizing microarray data, *Bioinformatics* 18(1), 111–119.

Windham, M.P. (1982), Cluster validity for the fuzzy c-means clustering algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 4, 357–363.

Windham, M.P. (1983), Geometric fuzzy clustering algorithms, *Fuzzy Sets and Systems* 10, 271–279.

Zahid, N., Limouri, M. and Essaid, A. (1999), New cluster-validity for fuzzy clustering, *Pattern Recognition* 32(7), 1089–1097.