

The Hyperbolic Smoothing Clustering Method

Adilson Elias Xavier

Dept. of Systems Engineering and Computer Science

Graduate School of Engineering (COPPE)

Federal University of Rio de Janeiro

P.O. Box 68511

Rio de Janeiro, RJ 21941-972, BRAZIL

e-mail: adilson@cos.ufrj.br

Abstract

It is considered the minimum sum-of-squares clustering problem. The mathematical modeling of this problem leads to a *min – sum – min* formulation which, in addition to its intrinsic bi-level nature, has the significant characteristic of being strongly non-differentiable. In order to overcome these difficulties, the resolution method proposed adopts a smoothing strategy using a special C^∞ differentiable class function. The final solution is obtained by solving a sequence of low dimension differentiable unconstrained optimization subproblems which gradually approach the original problem. The use of this technique, called Hyperbolic Smoothing, allows the main difficulties presented by the original problem to be overcome. A simplified algorithm containing only the essential of the method is presented. For the purpose of illustrating both the reliability and the efficiency of the method, a set of computational experiments was performed making use of traditional test problems described in the literature.

Keywords: Cluster Analysis, Pattern Recognition, Min-Sum-Min Problems, Nondifferentiable Programming, Smoothing

1 Introduction

Cluster analysis deals with the problems of classification of a set patterns or observations, in general represented as a point in a multidimensional space,

into clusters, following two basic and simultaneous objectives: patterns in the same clusters must be similar to another (homogeneity objective) and different from patterns of other clusters (separation objective), see Hartingan (1975) and Späth (1980).

Clustering is an important problem that appears in the broadest spectrum of applications, whose intrinsic characteristics engender many approaches to this problem, see Dubes and Jain (1976), Jain and Dubes (1988) and Hansen and Jaumard (1977).

In this paper, it is considered a particular clustering problem formulation. Among many criteria used in cluster analysis, the most natural, intuitive and frequently adopted criterion is the minimum sum-of-squares clustering (MSSC). This criterion corresponds to the minimization of the sum-of-squares of distances of observations to their cluster means, or equivalently, to the minimization of within-group sum-of-squares. It is a criterion for both homogeneity and separation objective, as, according to the Huygens Theorem, minimizing the within-cluster inertia of a partition (homogeneity within the cluster) is equivalent to maximizing the between-cluster inertia (separation between clusters).

The minimum sum-of-squares clustering (MSSC) formulation produces a global optimization mathematical problem. It is both nondifferentiable and nonconvex mathematical problem, with a large number of local minimizers. It is one of the NP-hard class (Brucker (1978)).

In the cluster analysis scope, algorithms use, traditionally, two main strategies: hierarchical clustering methods and partition clustering methods (Hansen and Jaumard (1997) and Jain et alli (1999)). Hierarchical methods, essentially heuristic procedures, produce a hierarchy of partitions of the set of observations according to an agglomerative strategy or to a divisive one. In the first case, the general algorithm starts from an initial partition, in which each cluster contains one pattern, and merges successively two clusters on the basis of a similarity measure until all patterns are in the same cluster. In the second case, the general algorithm starts from an initial partition with all patterns in the same cluster and, by successive bipartitions, reaches a partition in which each cluster contains one single pattern. In both strategies, the best partition is chosen, by an arbitrary criterion, from the hierarchy of partitions obtained.

Partition methods, in general, assume given the number of clusters and, essentially, seek the optimization of an objective function measuring the homogeneity within the clusters and/or the separation between the clusters. Heuristic algorithms of the exchange type, as the traditional k-means algorithm (McQueen (1967)) and succedaneous variations (Anderberg (1973) and Späth (1980)) are frequently used to find a local minimum of the objective function. However, any mathematical programming technique can be applied to solve the global optimization problem: dynamic programming (Jensen (1969)), branch and bound (Koontz, Narendra and Fukuraga (1975)), interior point algorithm (du Merle et alli (1997)), bilinear programming (Mangasarian (1997)), all kind of metaheuristics (by example: Reeves (1993) or Pacheco and Valencia (2003)) and nonsmooth optimization (Bagirov and Yearwood (2004)).

The core focus of this paper is the smoothing of the *min - sum - min* problem engendered by the modeling of the clustering problem. In a sense, the process whereby this is achieved is an extension of a smoothing scheme, called Hyperbolic Smoothing, presented in Santos (1997) for non-differentiable problems in general, in Chaves (1997) for the *min - max* problem and, more recently, in Xavier and Oliveira (2004) for the covering of plane domains by circles. This technique was developed through an adaptation of the hyperbolic penalty method originally introduced by Xavier (1982).

By smoothing we fundamentally mean the substitution of an intrinsically non-differentiable two-level problem by a C^∞ differentiable single-level alternative. This is achieved through the solution of a sequence of differentiable subproblems which gradually approaches to the original problem. In the present application, each subproblem, by using the Implicit Function Theorem, can be transformed in an low dimension unconstrained one, which, due the completely differentiability property, can be comfortably solved by using the most powerful and efficient algorithms, such that, conjugate gradient, quasi-Newton or Newton methods.

Although, this paper consider the particular MSSC clustering problem, it must be emphasized, that the proposed methodology, Hyperbolic Smoothing, can even be used for solving other clustering problem formulations. Moreover, it is expected that the smoothing strategy may be established as a new and powerful alternative family of methods for solving clustering problems.

This work is organized in the following way. A step by step definition of the clustering problem, directly connected to the presentation of the proposed hyperbolic smoothing approach, is presented in the next section. The new methodology is described in Section 3. The algorithm and the illustrative computational results are presented in Sections 4 and 5. Brief conclusions are drawn in the Section 5.

2 The Clustering Problem as a Min-Sum-Min Problem

Let $S = \{s_1, \dots, s_m\}$ denote a set of m patterns or observations from a Euclidian n -space to be clustered into a given number q of disjointed clusters.

In order to formulate the original clustering problem as a *min – sum – min* problem, we proceed as follows. Let $x_i, i = 1, \dots, q$ be the centroids of the clusters, where each $x_i \in \mathfrak{R}^n$. The set of these centroid coordinates will be represented by $X \in \mathfrak{R}^{nq}$. Given a point s_j of S , we initially calculate the distance from s_j to the center in X that is nearest. This is given by

$$z_j = \min_{x_i \in X} \|s_j - x_i\|_2. \quad (1)$$

The most frequent measurement of the quality of a clustering associated to a specific position of q centroids is provided by the sum of the squares of these distances

$$D(X) = \sum_{j=1}^m z_j^2. \quad (2)$$

The optimal placing of the centroids must provide the best-quality of this measurement. So, if X^* denotes an optimal placement, then the problem is

$$X^* = \underset{X \in \mathfrak{R}^{nq}}{\operatorname{argmin}} D(X), \quad (3)$$

where X is the set of all placements of the q centroids. Using (1)-(3), we finally arrive at

$$X^* = \underset{X \in \mathfrak{R}^{nq}}{\operatorname{argmin}} \sum_{j=1}^m \min_{x_i \in X} \|s_j - x_i\|_2^2. \quad (4)$$

3 Transforming the Problem

The problem (4) above can be placed under the equivalent one

$$\begin{aligned} & \text{minimize} \quad \sum_{j=1}^m z_j^2 & (5) \\ \text{subject to} \quad & z_j = \min_{i=1, \dots, q} \|s_j - x_i\|_2, \quad j = 1, \dots, m \end{aligned}$$

Considering its definition, z_j must necessarily satisfy the following set of inequalities:

$$z_j - \|s_j - x_i\|_2 \leq 0, \quad i = 1, \dots, q. \quad (6)$$

Putting these inequalities in the place of the equality constraints of problem (5), it is obtained the relaxed problem

$$\begin{aligned} & \text{minimize} \quad \sum_{j=1}^m z_j^2 & (7) \\ \text{subject to} \quad & z_j - \|s_j - x_i\|_2 \leq 0, \quad j = 1, \dots, m, \quad i = 1, \dots, q \end{aligned}$$

Since the variables z_j are not bounded from below, the optimum solution of the relaxed problem will be $z_j = 0, j = 1, \dots, m$. In order to obtain the desired equivalence, we must, therefore, modify the problem (7). We do so by first letting $\varphi(y)$ denote $\max\{0, y\}$ and then observing that from the set of inequalities in (7), it follows that

$$\sum_{i=1}^q \varphi(z_j - \|s_j - x_i\|_2) = 0, \quad j = 1, \dots, m. \quad (8)$$

For fixed j and assuming $d_1 < \dots < d_q$ with $d_i = \|s_j - x_i\|_2$, Figure (1) illustrates the first three summands of (8) as a function of z_j .

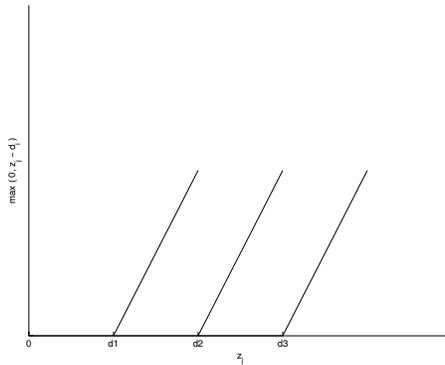


Figure 1: Summands in (8)

Using (8) in place of the set of inequality constraints in (7), we would obtain an equivalent problem maintaining the undesirable property that $z_j, j = 1, \dots, m$ still have no lower bound. Considering, however, that the objective function of problem (7) will force each $z_j, j = 1, \dots, m$, downward, we can think of bounding the latter variables from below by considering “ $>$ ” in place of “ $=$ ” in (8) and considering the resulting “non-canonical” problem

$$\text{minimize } \sum_{j=1}^m z_j^2 \quad (9)$$

$$\text{subject to } \sum_{i=1}^q \varphi(z_j - \|s_j - x_i\|_2) > 0, \quad j = 1, \dots, m.$$

The canonical formulation can be recovered from (9) by perturbing (8) and considering the modified problem:

$$\begin{aligned} & \text{minimize } \sum_{j=1}^m z_j^2 & (10) \\ & \text{subject to } \sum_{i=1}^q \varphi(z_j - \|s_j - x_i\|_2) \geq \varepsilon, \quad j = 1, \dots, m \end{aligned}$$

for $\varepsilon > 0$. Since the feasible set of problem (9) is the limit of that of (10) when $\varepsilon \rightarrow 0_+$, we can then consider solving (9) by solving a sequence of problems like (10) for a sequence of decreasing values for ε that approaches 0.

4 Smoothing the Problem

Analyzing the problem (10), the definition of function φ endows it with an extremely rigid non-differentiable structure, which makes its computational solution very hard. In view of this, the numerical method we adopt for solving problem (1), takes a smoothing approach. From this perspective, let us define the function:

$$\phi(y, \tau) = \left(y + \sqrt{y^2 + \tau^2} \right) / 2 \quad (11)$$

for $y \in \Re$ and $\tau > 0$.

Function ϕ has the following properties:

(a) $\phi(y, \tau) > \varphi(y), \quad \forall \tau > 0;$

(b) $\lim_{\tau \rightarrow 0} \phi(y, \tau) = \varphi(y);$

(c) $\phi(\cdot, \tau)$ is an increasing convex C^∞ function.

So, function ϕ constitutes an approximation of function φ . Adopting the same assumptions used in Figure 1, the first three summands of (8) and their corresponding smoothed approximations, given by (11), are depicted in Figure 2.

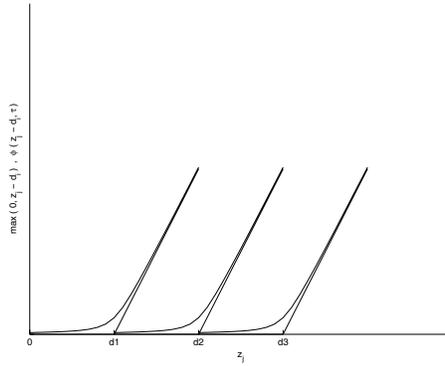


Figure 2: Original and smoothed summands in (8)

By using function ϕ in the place of function φ , it is produced the problem

$$\begin{aligned} & \text{minimize} \quad \sum_{j=1}^m z_j^2 & (12) \\ & \text{subject to} \quad \sum_{i=1}^q \phi(z_j - \|s_j - x_i\|_2, \tau) \geq \varepsilon, \quad j = 1, \dots, m. \end{aligned}$$

In order to obtain a differentiable problem, it is yet necessary to smooth the Euclidian distance $\|s_j - x_i\|_2$. For this purpose, let us define the function

$$\theta(s_j, x_i, \gamma) = \sqrt{\sum_{l=1}^n (s_j^l - x_i^l)^2 + \gamma^2} \quad (13)$$

for $\gamma > 0$.

Function θ has the following properties:

- (a) $\lim_{\gamma \rightarrow 0} \theta(s_j, x_i, \gamma) = \|s_j - x_i\|_2$;
- (b) θ is a C^∞ function.

By using function θ in the place of the distance $\|s_j - x_i\|_2$, it is now obtained the completely differentiable problem

$$\begin{aligned} & \text{minimize} \quad \sum_{j=1}^m z_j^2 \quad (14) \\ & \text{subject to} \quad \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) \geq \varepsilon, \quad j = 1, \dots, m. \end{aligned}$$

So, the properties of functions ϕ and θ allows us to seek a solution to problem (10) by solving a sequence of subproblems above, produced by the decreasing of the parameters $\gamma \rightarrow 0$, $\tau \rightarrow 0$ and $\varepsilon \rightarrow 0$.

In other hand, given any set of centroids $x_i, i = 1, \dots, q$, due the property (c) of the hyperbolic smoothing function ϕ , the constraints of the problem (14) will certainly be active. So, it will be equivalent to the problem:

$$\begin{aligned} & \text{minimize} \quad \sum_{j=1}^m z_j^2 \quad (15) \\ & \text{subject to} \quad h_j(z_j, x) = \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \varepsilon = 0, \quad j = 1, \dots, m. \end{aligned}$$

The variable domain space of the problem (15) has $(nq + m)$ dimensions. As, in general, the value of the parameter m , the cardinality of the set S of the observations s_j , is very large, the problem (15) has a very large number of variables. However, it has a separable structure, because each variable z_j appears only in one equality constraint. So, as the partial derivative of $h(z_j, x)$ in relation to $z_j, j = 1, \dots, m$ is not equal zero, it is possible to use the Implicit Function Theorem in order to calculate each component $z_j, j = 1, \dots, m$ as a function of the centroid variables $x_i, i = 1, \dots, q$. In this way, it is obtained the unconstrained problem

$$\text{minimize } f(x) = \sum_{j=1}^m z_j(x)^2 \quad (16)$$

where each $z_j(x)$ is obtained by the calculation of a zero of each equation

$$h_j(z_j, x) = \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \varepsilon = 0, \quad j = 1, \dots, m. \quad (17)$$

Due the property (c) of the hyperbolic smoothing function, each term ϕ above is strictly crescent with the variable z_j , so the equation has only one zero.

Again, due the Implicit Function Theorem, the functions $z_j(x)$ have all derivatives in relation to the variables $x_i, i = 1, \dots, q$. So, it is possible to calculate the gradient of the objective function of the problem (16)

$$\nabla f(x) = \sum_{j=1}^m 2 z_j(x) \nabla z_j(x) \quad (18)$$

where

$$\nabla z_j(x) = - \nabla h_j(z_j, x) / \frac{\partial h_j(z_j, x)}{\partial z_j}. \quad (19)$$

The above approach is nothing else the basic idea used by Abadie and Carpentier (1969) for the development of the general reduced gradient algorithm, intended to the solution of the general nonlinear programming problem subject to equality constraints.

In this way, it is very easy to solve the problem (16), by making use of any method based in the first order derivative information. At last, it must be emphasized that problem (16) is defined in a (nq) -dimensional space, so a small problem, since the number of clusters, q , is, in general, very small for real applications.

The solution of the original clustering problem can be obtained by using the Hyperbolic Smoothing Clustering Algorithm, described below in a simplified form.

Simplified HSC Algorithm

Initialization Step: Choose initial values: $x^0, \gamma^1, \tau^1, \varepsilon^1$.

Choose values $0 < \rho_1 < 1, 0 < \rho_2 < 1, 0 < \rho_3 < 1$; let $k = 1$.

Main Step: Repeat indefinitely

Solve problem (16) with $\gamma = \gamma^k, \tau = \tau^k$ and $\varepsilon = \varepsilon^k$, starting at the initial point x^{k-1} and let x^k be the solution obtained.

Let $\gamma^{k+1} = \rho_1 \gamma^k, \tau^{k+1} = \rho_2 \tau^k, \varepsilon^{k+1} = \rho_3 \varepsilon^k, k := k + 1. \blacksquare$

Just as in other smoothing methods, the solution to the clustering problem is obtained by resolving an infinite sequence of optimization problems, in the present case, a sequence of low dimension unconstrained minimization subproblems, $k = 1, 2, \dots$, in the Main Step.

Notice that the algorithm causes τ approach 0, so the constraints of the subproblems it solves, given as in (14), tend to those of (10). In addition, the algorithm causes ε approach 0, so, in a simultaneous movement, the solved problem (10) gradually approaches to problem (9).

5 Computational Results

The computational results presented below were withdrawn from a preliminary implementation developed by Sousa (2005) for your M.Sc. thesis. The numerical experiments have been carried out on a Pentium 4, 2GHz,

PC, with a Digital FORTRAN, Version 6.0A. The unconstrained minimization tasks were carried out by means of a Quasi-Newton algorithm employing the BFGS updating formula from the Harwell Library.

First, in order to illustrate the functioning of the method, we present some computational results on one small test instance ($n = 2, q = 3, m = 41$), originally presented in Demyanov (2004). In the initialization step of the algorithm, it was taken the following choices: $\rho_1 = 1/10, \rho_2 = 1/10, \rho_3 = 1/10, \gamma^1 = 1/10, \tau^1 = 1/10$ and $\varepsilon^1 = 1/10$. The initial point, $x^0 = (2.19093620, 4.62296337, 5.59448532, 4.63150882, 3.88412198, 8.04804420)$, was taken over the inertia ellipse made up by using the center gravity of the observation points and the first and second eigen values and eigen vectors of the matrix $S^T S$, where S is the matrix (m, n) whose row j is formed with the components of the observation s_j .

k	a_1^k	b_1^k	a_2^k	b_2^k	a_3^k	b_3^k	$f(x^k)$
1	2.1247416	6.0000027	4.8000021	1.9000180	5.1665170	7.6664565	99.411220
2	2.1249361	6.0000008	4.8000006	1.9000041	5.1666298	7.6666150	99.415309
3	2.1249842	6.0000002	4.8000002	1.9000010	5.1666644	7.6666538	99.416328
4	2.1249961	6.0000001	4.8000000	1.9000003	5.1666661	7.6666635	99.416582
5	2.1249990	6.0000000	4.8000000	1.9000001	5.1666665	7.6666659	99.416646
6	2.1249998	6.0000000	4.8000000	1.9000001	5.1666666	7.6666665	99.416661
7	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666666	99.416666
8	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
-	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667

Table 1: Sequence of points generated in solving the Demyanov instance

Table 1 shows the sequence of points generated by the method in solving the first instance. The exact solution is presented in the last row. Column k represents the iteration numbers, while the pairs (a_i^k, b_i^k) represent the two coordinates of the centroids $x_i^k, i = 1, 2, 3$, and the column $f(x^k)$ shows the objective function values. It is possible to observe a linear convergence of the sequence of the points generated by the algorithm to the optimum solution, with the controlled linear decreasing of the parameters γ^k, ε^k and τ^k .

Let be $S_i, i = 1, \dots, q$ the partition of S generated by the centroids $x_i, i = 1, \dots, q$

$$S = \bigcup_{i=1}^q S_i \quad (20)$$

$$S_{i_1} \cap S_{i_2} = \emptyset, \quad \forall i_1, i_2 = 1, \dots, q, \quad i_1 \neq i_2, \quad (21)$$

$$S_i \neq \emptyset, \quad \forall i = 1, \dots, q. \quad (22)$$

The center of gravity of the clusters is given by

$$v_i = \frac{1}{|S_i|} \sum_{s_j \in S_i} s_j, \quad \forall i = 1, \dots, q. \quad (23)$$

Table 2 shows the resultant clustering generated by the sequence of centroids. The pairs (c_i^k, d_i^k) represent the two coordinates of the $v_i^k, i = 1, 2, 3$, the centers of gravity of the clusters associated to the centroids $x_i^k, i = 1, 2, 3$, and the column $f(v^k)$ shows the correspondent objective function values. It is important to observe that, already at first iteration, the centers of gravity of the formed clusters are equal to the optimum solution.

k	c_1^k	d_1^k	c_2^k	d_2^k	c_3^k	d_3^k	$f(v^k)$
1	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
2	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
3	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
4	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
5	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
6	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
7	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
8	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667
-	2.1250000	6.0000000	4.8000000	1.9000000	5.1666667	7.6666667	99.416667

Table 2: Sequence of gravity centers of clusters generated in solving the Demyanov instance

Figure 3 shows the original observation points of the Demyanov instance and the optimum clustering produced by the algorithm.

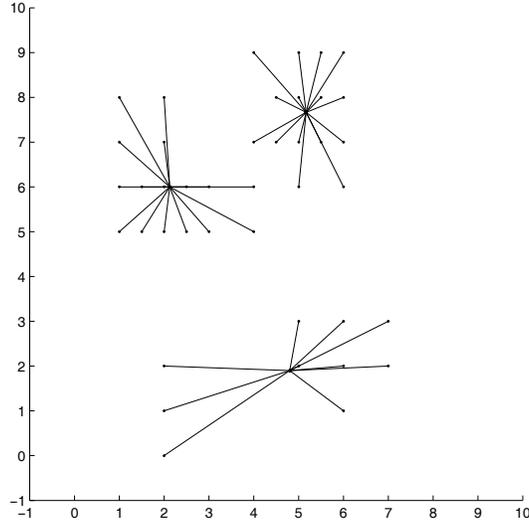


Figure 3: Solution for Demyanov Instance

In order to show some performance of proposed algorithm, it is shown below results obtained by solving four standard test problems from the cluster analysis literature:

1 - German Towns, which uses the two Cartesian coordinates of 59 towns, originally presented by Späth (1980);

2 - Ruspini example, which uses 75 artificial points in the Euclidean plane (Ruspini (1970));

3 - TSPLIB-3038, which uses 3038 points in the plane from a traveling salesman problem of Reinelt (1991);

4 - Iris Fisher example, which uses 4-dimensional data on 150 iris from Gaspé peninsula, published by Anderson(1935) and used by Fisher(1936).

In Tables 3-5, it is presented the number of clusters (q), the best known value for the global optimum (f_{opt}) taken from Merle et alli (2000) and Bagirov and Yearwood (2004), the value produced by the HSC algorithm (f_{HSC}), the error (E) for this solution and the CPU time given in seconds, where the error is calculated as

$$E = \frac{100 (f_{HSC} - f_{opt})}{f_{opt}}. \quad (24)$$

q	f_{opt}	f_{HSC}	E	$Time$
2	0.121426 E6	0.121426 E6	0.00	0.05
3	0.77009 E5	0.77009 E5	0.00	0.08
4	0.49601 E5	0.49601 E5	0.00	0.09
5	0.39453 E5	0.38716 E5	-1.87	0.25

Table 3: Results for the German Towns Instance

The results presented in Table 3 show that for this data set, Algorithm HSC reaches the best known results for $q = 2, 3, 4$, while for $q = 5$ HSC establishes a new best result of cluster literature. The CPU time is small for all q . Figure 4 shows the original observation points of the instance and the optimum clustering produced by the algorithm for $q = 5$.

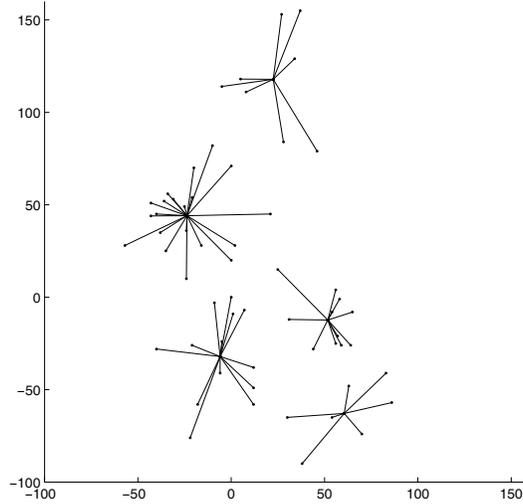


Figure 4: Solution for German Towns Instance, $q = 5$.

From Table 4 it is possible to see that HSC Algorithm again, in small CPU times, gives the best known results for all values of q , except $q = 2$ where the error, $E = 0.00002\%$, is despicable.

The results presented in Table 5 show that for this data set, HSC Algorithm achieves results with small deviations from the best known global minimum, in a significant small computation time than those used by Hansen

q	f_{opt}	f_{HSC}	E	$Time$
2	89337.83	89337.83	0.00	0.07
3	51063.47	51063.48	0.00	0.10
4	12881.05	12881.05	0.00	0.10
5	10126.72	10126.72	0.00	0.20
6	8575.41	8575.41	0.00	0.21
7	7126.20	7126.20	0.00	0.22
8	6149.64	6149.64	0.00	0.28
9	5181.65	5181.65	0.00	0.38
10	4446.28	4446.28	0.00	0.39

Table 4: Results for the Ruspini Instance

q	f_{opt}	f_{HSC}	E	$Time$
2	0.31688 E 10	0.31706 E 10	0.06	1.31
3	0.21763 E 10	0.21777 E 10	0.06	2.57
4	0.14790 E 10	0.14794 E 10	0.03	3.58
5	0.11982 E 10	0.11986 E 10	0.03	7.24
6	0.96918 E 9	0.96936 E 9	0.02	7.56
7	0.83966 E 9	0.83967 E 9	0.00	9.17
8	0.73475 E 9	0.73491 E 9	0.02	20.24
9	0.64477 E 9	0.64607 E 9	0.20	24.71
10	0.56025 E 9	0.56030 E 9	0.01	19.75
20	0.26681 E 9	0.26736 E 9	0.21	49.30
30	0.17557 E 9	0.17593 E 9	0.21	308.53
40	0.12548 E 9	0.12507 E 9	-0.33	737.97
50	0.98400 E 9	0.99142 E 9	0.75	593.34

Table 5: Results for the TSPLIB-3038 Instance

el alli (2002). Moreover, for results for $q = 40$, HSC Algorithm establishes a new best result of the cluster literature.

In Table 6, the results for 10 different initial point restarts are presented. The third column gives the best objective function value produced by HSC Algorithm. The next four columns give the number of occurrences of the best solution, the error of the best solution, the average error and the CPU time.

The initial point restarts was produced around the center of the observation points:

$$v = \frac{1}{m} \sum_{j=1}^m s_j, \quad (25)$$

by using very small perturbations

$$x^{0^k} = v^k (1 + RAND \cdot 10^{-6}), \quad (26)$$

where RAND is a random number with uniform distribution on the interval $[-0.5, +0.5]$.

q	f_{opt}	$f_{HSC_{Best}}$	Occur.	E_{Best}	E_{Mean}	$Time_{Mean}$
2	15235.0	15234.8	10	0.00	0.00	0.10
3	7885.1	7885.1	10	0.00	0.00	0.16
4	5722.8	5722.8	10	0.00	0.00	0.29
5	4644.6	4644.6	7	0.00	2.21	0.44
6	3904.0	3904.0	10	0.00	0.00	0.52
7	3429.8	3430.0	7	0.01	2.39	0.97
8	2998.9	2998.9	5	0.00	3.52	1.04
9	2778.6	2778.6	9	0.00	1.20	1.34
10	2583.4	2583.4	9	0.00	0.38	1.70

Table 6: Results for the Fisher Iris Instance

From Table 6, it is possible to see that HSC Algorithm with 10 restarts solves optimally the Fisher Iris problems, in small computational times, except for $q = 7$ where the error, $E = 0.006\%$, is very small.

In the end, from the results presented in Tables 3-6 it is shown that the first implementation of the HSC Algorithm, starting from a unique initial point, reaches the best known solution in the most of cases or calculates a solution which is very close to the best in all other cases (basically from TST3038 instance). Moreover, in this first experiment, it established two new best results of the cluster literature.

6 Conclusions

In this paper, a new method for the solution of the minimum sum-of-squares clustering problem has been proposed. By using the Hyperbolic Smoothing technique, the problem has been reformulated, in an approximation approach, as a completely differentiable constrained optimization problem. By using the Implicit Function Theorem, the last problem has been reformulated as a low dimension unconstrained optimization problem, in the Euclidean space \mathcal{R}^{nq} . Then the basic steps of an algorithm for solving the original clustering problem has been presented.

Although it is considered the particular MSSC formulation, it must be emphasized that this approach can be used for solving other clustering problems, by example, for the norm 1 formulation, it is necessary only a trivial adaptation.

Moreover, it must be observed that the methodology can be applied to any *min – sum – min* problem. Among them, we consider it to be particularly interesting to apply this approach on the general problem considering by Demyanov (2004).

The performance of the HSC Algorithm can be attributed to the complete differentiability of the approach. Inside this context, each centroid get permanently see every observation point. Conversely, each observation point can permanently see every centroid and attract it. The Figure 5 below tries depict this idea.

There are several possibilities for the continuation of this work. One obvious alternative to be explored is to study simple algorithm modifications, such as:

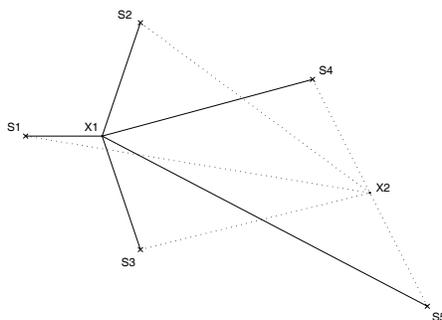


Figure 5: The C^∞ differentiability property effect

- Similarly other clustering algorithms, like Eikelder and Erk (2004), Hansen and Mladenovic (2001) and Merle et alli (2000), heuristics can be used to produce a good initial point through an auxiliary fast algorithm as the k-means or any of its innumerable variations, Hansen et alli (2002).

- In a complementarily way, it is possible to connect it with a heuristic or metaheuristics algorithm in order to do a local search so as to pick up a better point around the initially found local minimum.

- In other hand, it is possible to calculate clusters step-by-step, gradually increasing the number of data clusters until reaching the specified q parameter value, approach successfully adopted, for example, by Hansen et alli (2002) or Bagirov and Yearwood (2004).

Finally, it must be remembered that the MSSC problem is a global optimization problem with a lot of local minima. Although the HSC Algorithm do not offer the guarantee of obtaining a global optimum point, in view of the results obtained, where a preliminary implementation of the proposed algorithm produced efficiently and reliably very deep local minima, perfectly adequate to the necessities and demands of real applications, we believe that the methodology introduced in this article can be used for solving large, practical optimal clustering problems.

Acknowledgments

The authors would like to thank Prof. Valmir Carneiro Barbosa of the Federal University of Rio de Janeiro for the helpful reviewing work and constructive suggestions.

References

- ABADIE, J. and CARPENTIER, J. (1969) "Generalization of the Wolfe Reduced Gradient Method to the Case of Nonlinear Constraints" in Optimization, R. Fletcher (editor), Academic Press, London, pages 37-47
- ANDERBERGER, M. R. (1973) "Cluster Analysis for Applications", Academic Press Inc., New York.
- ANDERSON, E. (1935) "The Irises of Gaspe Peninsula", Bull. American Iris Society, 55 pp. 2-5
- BAGIROV, A. M. and YEARWOOD, J. (2004) "A New Nonsmooth Optimization Algorithm for Minimum Sum-of-Squares Clustering Problems", to appear in European Journal of Operational Research, Elsevier.
- CHAVES, A.M.V., 1997, Resolução do Problema Minimax Via Suavização, M.Sc. Thesis - COPPE - UFRJ, Rio de Janeiro.
- CONWAY, J. H. and SLOANE, N. J. A., 1988, Sphere Packings, Lattices and Groups, Springer, New York.
- DEMYANOV, A. (2004) "On the Solution of Min-Sum-Min Problems", to appear in Journal of Global Optimization, Kluwer Acad. Publ.
- DUBES, R. C. and JAIN, A. K. (1976) "Cluster Techniques: The User's Dilemma", Pattern Recognition No. 8 pp. 247-260.
- FISHER, R.A. (1936) "The Use of Multiple Measurements in Taxonomic Problems", Ann. Eugenics, VII part II pp. 179-188. Reprinted in R. A. Fisher (1950) "Contributions to Mathematical Statistics" , Wiley
- HANSEN, P. and JAUMARD, B. (1997) "Cluster Analysis and Mathematical Programming", Mathematical Programming No. 79 pp. 191-215.
- HANSEN, P. and MLADENOVIC, N. (2001) "J-Means: A New Heuristic for Minimum Sum-of-Squares Clustering", Pattern Recognition, Vol. 4 pp 405-413.
- HANSEN, P., NGAI, E., CHEUNG, B. K., MLADENOVIC, N. (2002) "Anal-

ysis of Global k-Means, an Incremental Heuristic for Minimum Sum-of-Squares Clustering”, Research Paper G-2002-43, GERAD, Montreal, Canada and M”, Mathematical Programming No. 79 pp. 191-215.

HARTINGAN, J. A. (1975) Clustering Algorithms, John Wiley and Sons, Inc., New York, NY.

JAIN, A. K. and DUBES, R. C. (1988) ”Algorithms for Clustering Data”, Prentice-Hall Inc., Upper Saddle River, NJ.

JAIN, A. K.; MURTY, M. N. and FLYNN, P. J. (1999) ”Data Clustering: A Review”, ACM Computing Surveys, Vol. 31, No. 3, Sept 1999.

JENSEN, R. E. (1969) ”A Dynamic Programming Algorithm for Clustering Analysis”, Operations Research No. 17 pp. 1043-1057.

KOONTZ, W. L. G.; NARENDRA, P. M. and FUKUNAGA (1975) ”A Branch and Bound Clustering Algorithm”, IEEE Transactions on Computers No. 24 pp. 908-915.

MANGASARIAN, O. L. (1996) ”Mathematical Programming in Data Mining”, Technical Report 96-05 - Computer Sciences Department, University of Wisconsin.

MCQUEEN, J. (1967) ”Some Methods for Classification and Analysis of Multivariate Observations” In ”Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability”, pp. 281-297.

MERLE, O. du; HANSEN, P.; JAUMARD, B.; MLADENOVIC, V. (2000) ”An Interior Point Algorithm for Minimum Sum-of-Squares Clustering”, SIAM Journal on Scientific Computing No. 21 pp. 1485-1505.

PACHECO, J. and VALENCIA, O. (2003) ”Design of Hybrids for Minimum Sum-of-Squares Clustering Problem”, Computational Statistics and Data Analysis No. 43 pp. 235-248.

REEVES, C. R. (Ed.) (1993), Modern Heuristics Techniques for Combinatorial Problems, Blackwell, London.

REINELT, G. (1991), ”TSP-LIB A Traveling Salesman Library”, ORSA J.

Comput. pp. 376-384

ROGERS, C. A., (1964), Packing and Clustering, Cambridge University Press, Cambridge.

RUSPINI, E. (1970), "Numerical Methods for Fuzzy Clustering", Information Sci. pp. 319-350

SANTOS, A.B.A. (1997), Problemas de Programação Não- Diferenciável: Uma Metodologia de Suavização, M.Sc. thesis - COPPE - UFRJ, Rio de Janeiro.

SOUSA, L.C.F. (2005), Desempenho Computacional do Método de Agrupamento Via Suavização Hiperbólica, M.Sc. thesis - COPPE - UFRJ, Rio de Janeiro.

SPÄTH, H. (1980), "Cluster Analysis Algorithms for Data Reduction and Classification", Ellis Horwood, Upper Saddle River, NJ.

XAVIER, A.E., 1982, Penalização Hiperbólica: Um Novo Método para Resolução de Problemas de Otimização, M.Sc. Thesis - COPPE - UFRJ, Rio de Janeiro.

XAVIER, A. E. and OLIVEIRA, A. A. F. (2004) "Optimal Covering of Plane Domains by Circles Via Hyperbolic Smoothing", to appear in Journal of Global Optimization, Kluwer.