Solving the Minimum Sum of L1 Distances Clustering Problem by Hyperbolic Smoothing and Partition into Boundary and Gravitational Regions

Adilson Elias Xavier; Vinicius Layter Xavier; Sergio B. Villas-Boas
Federal University of Rio de Janeiro, BRAZIL
Email: {adilson;vinicius;sbvb}@cos.ufrj.br

Paper ID 358

Abstract

The article considers the minimum sum of distances clustering problem, where the distances are measured through the L1 or Manhattan metric (MSDC-L1). The mathematical modelling 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.

We propose the AHSC-L1 method to solve this problem, by combining 2 techniques. The first technique is Hyperbolic Smoothing Clustering (HSC), that adopts a smoothing strategy using a special C^∞ completely differentiable class function. The second technique is the partition of the set of observations into two non overlapping groups: "data in frontier" and "data in gravitational regions". We propose a classification of the gravitational observations by each component, which simplifies of the calculation of the objective function and its gradient. The combination of these 2 techniques for MSDC-L1 problem drastically simplify the computational tasks.

Keywords: Cluster Analysis, Min-Sum-Min Problems, Manhattan Metric, Non differentiable Programming, Smoothing

1 Introduction

Cluster analysis deals with the problems of classification of a set of patterns or observations. In general the observations are represented as points in a multidimensional space. The purpose of cluster analysis is to define the clusters to that each observation belongs, following two basic and simultaneous objectives: patterns in the same clusters must be similar to each other (homogeneity objective) and different from patterns in other clusters (separation objective) Hartigan (1975) [2] and Späth (1980) [4].

In this paper, a particular clustering problem formulation is considered. Among many criteria used in cluster analysis, a frequently adopted criterion is the minimum sum of L1 distances clustering (MSDC-L1); see for example Bradley and Mangasarian (1997) [1]. This criterion corresponds to the minimization of the sum of distances of observations to their centroids, where the distances are measured through the L1 or

Manhattan metric. As broadly recorded by the literature, the Manhattan distance is more robust against outliers.

For the sake of completeness, we present first the Hyperbolic Smoothing Clustering Method (HSC), Xavier (2010) [6]. Basically the method performs the smoothing of the non differentiable *min-sum-min* problem engendered by the modelling of a broad class of clustering problems, including the minimum sum of L1 distances clustering (MSDC-L1) formulation. This technique was developed through an adaptation of the hyperbolic penalty method originally introduced by Xavier (1982) [5]. By smoothing, we fundamentally mean the substitution of an intrinsically non differentiable two-level problem by a C^∞ unconstrained differentiable single-level alternative.

Additionally, the paper presents an accelerated methodology applied to the specific considered problem. The basic idea is the partition of the set of observations into two non overlapping parts. By using a conceptual presentation, the first set corresponds to the observation points relatively close to two or more centroids. The second set corresponds to observation points significantly closer to a single centroid in comparison with others. The same partition scheme was presented first by Xavier and Xavier (2011) [7] in order to solve the specific minimum sum of squares clustering (MSSC) formulation. In this paper, specific features of the minimum sum of L1 distances clustering (MSDC-L1) formulation are explored in order to take additional advantages of the partition scheme.

2 The Minimum Sum of L1 Distances Clustering Problem

Let $S = \{s_1, \dots, s_m\}$ denote a set of m patterns or observations from an Euclidean n-space, to be clustered into a given number q of disjoint clusters. 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 $x_i \in \mathbb{R}^n$. The set of these centroid coordinates will be represented by $X \in \mathbb{R}^{nq}$.

Given a point s_j of S, we initially calculate the L1 distance from s_j to a center that is a nearest. This is given by $z_j = \min_{i=1,\dots,q} \|s_j - x_i\|_1$. A frequent measurement of the quality of a clustering associated to a specific position of q centroids is provided by the sum of the L1 distances, which determines the MSDC-L1 problem:

$$\begin{aligned} & minimize \sum_{j=1}^{m} z_j \\ & \text{subject to} \quad z_j = \min_{i=1,\dots,q} \|s_j - x_i\|_1, \qquad j = 1, \dots, m \end{aligned} \tag{1}$$

3 The Hyperbolic Smoothing Clustering Method

Considering its definition, each z_j must necessarily satisfy the following set of inequalities: $z_j - \|s_j - x_i\|_1 \leq 0, i = 1, \dots, q$. Substituting these inequalities for the equality constraints, problem (1) produces the relaxed problem:

$$
\begin{aligned}
&\textit{minimize} \ \sum_{j=1}^{m} z_j && (2)\\
&\text{subject to} \quad z_j - \|s_j - x_i\|_1 \le 0, \quad j = 1,\dots,m, \quad i = 1,\dots,q.
\end{aligned}
$$

Since the variables z_j are not bounded from below, the optimization procedure will determine $z_j \to \infty$, $j = 1,\dots,m$. In order to obtain the desired equivalence, we must, therefore, modify problem (2). We do so by first letting $\varphi(y)$ denote $\max\{0,y\}$ and then observing that, from the set of inequalities in (2), it follows that $\sum_{i=1}^{q} \varphi(z_j - \|s_j - x_i\|_1) = 0, j = 1,\dots,m$. In order to bound the variables z_j, $j = 1,\dots,m$ we include an $\varepsilon > 0$ perturbation.

$$
\begin{aligned}
&\textit{minimize} \ \sum_{j=1}^{m} z_j && (3)\\
&\text{subject to} \quad \sum_{i=1}^{q} \varphi(z_j - \|s_j - x_i\|_1) \ \ge \ \varepsilon, \quad j = 1,\dots,m
\end{aligned}
$$

Since the feasible set of problem (1) is the limit of that of (3) when $\varepsilon \to 0_+$, we can then consider solving (1) by solving a sequence of problems like (3) for a sequence of decreasing values for ε that approaches 0.

Analysing the problem (3), the definition of function φ and the definition of L1 distance 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 approximation functions below:

$$\phi(y,\tau) = \left(y + \sqrt{y^2 + \tau^2}\right)/2 \tag{4}$$

$$\theta_1(s_j, x_i, \gamma) \ = \ \sum_{l=1}^{n} \sqrt{(s_j^l - x_i^l)^2 + \gamma^2} \tag{5}$$

By using the asymptotic approximation properties of the functions θ_1 and ϕ, the following completely differentiable problem is now obtained:

$$
\begin{aligned}
&\textit{minimize} \ \sum_{j=1}^{m} z_j && (6)\\
&\text{subject to} \quad \sum_{i=1}^{q} \phi(z_j - \theta_1(s_j, x_i, \gamma), \tau) \ge \varepsilon, \quad j = 1,\dots,m.
\end{aligned}
$$

So, the properties of functions ϕ and θ_1 allow us to seek a solution to problem (3) by solving a sequence of subproblems like problem (6), produced by the decreasing of the parameters $\gamma \to 0$, $\tau \to 0$ and $\varepsilon \to 0$.

On the other side, the constraints will certainly be active and problem (6) will at last be equivalent to problem:

$$\begin{aligned} & \textit{minimize} \ \sum_{j=1}^{m} z_j \\ & \text{subject to} \quad h_j(z_j, x) = \sum_{i=1}^{q} \phi(z_j - \theta_1(s_j, x_i, \gamma), \tau) - \varepsilon = 0, \qquad j = 1, \dots, m. \end{aligned} \tag{7}$$

Problem (7) has a separable structure, because each variable z_j appears only in one equality constraint. Therefore, as the partial derivative of $h(z_j,x)$ with respect to z_j, $j = 1,\dots,m$ is not equal to zero, it is possible to use the Implicit Function Theorem 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, the unconstrained problem

$$\textit{minimize} \ f(x) = \sum_{j=1}^{m} z_j(x) \tag{8}$$

is obtained, where each $z_j(x)$ results from the calculation of a zero of each equation

$$h_j(z_j, x) = \sum_{i=1}^{q} \phi(z_j - \theta_1(s_j, x_i, \gamma), \tau) - \varepsilon = 0, \qquad j = 1, \dots, m. \tag{9}$$

Again, due to the Implicit Function Theorem, the functions $z_j(x)$ have all derivatives with respect to the variables x_i, $i = 1,\dots,q$, and therefore it is possible to calculate the gradient of the objective function of problem (8),

$$\nabla f(x) = \sum_{j=1}^{m} \nabla z_j(x) \tag{10}$$

where

$$\nabla z_j(x) = - \nabla h_j(z_j, x) \ / \ \frac{\partial h_j(z_j, x)}{\partial z_j}, \tag{11}$$

while $\nabla h_j(z_j,x)$ and $\partial h_j(z_j,x)/\partial z_j$ are obtained from equations (4), (5) and (9).

In this way, it is easy to solve problem (8) by making use of any method based on first or second order derivative information. At last, it must be emphasized that problem (8) is defined on an $(nq)-$dimensional space, so it is 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.

The Simplified HSC-L1 Algorithm

Initialization Step:

Choose initial values: x^0, γ^1 , τ^1 , ε^1.

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

Main Step: Repeat until a stopping rule is attained

Solve problem (8) 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 $\gamma^{k+1} = \rho_1 \gamma^k$, $\tau^{k+1} = \rho_2 \tau^k$, $\varepsilon^{k+1} = \rho_3 \varepsilon^k$, $k := k+1$. ■

Just as in other smoothing methods, the solution to the clustering problem is obtained, in theory, by solving an infinite sequence of optimization problems. In the HSC-L1 algorithm, each problem to be minimized is unconstrained and of low dimension.

Notice that the algorithm causes τ and γ to approach 0, so the constraints of the subproblems as given in (6) tend to those of (3). In addition, the algorithm causes ε to approach 0, so, in a simultaneous movement, the solved problem (3) gradually approaches the original MSDC-L1 problem (1).

4 The Accelerated Hyperbolic Smoothing Clustering Method

The calculation of the objective function of the problem (8) demands the determination of the zeros of m equations (9), one equation for each observation point. This is a relevant computational task associated to HSC-L1 Algorithm.

In this section, it is presented a faster procedure. The basic idea is the partition of the set of observations into two non overlapping regions. By using a conceptual presentation, the first region corresponds to the observation points that are relatively close to two or more centroids. The second region corresponds to the observation points that are significantly close to a unique centroid in comparison with the other ones.

So, the first part J_B is the set of boundary observations and the second is the set J_G of gravitational observations. Considering this partition, equation (8) can be expressed in the following way:

$$minimize\ f(x) = \sum_{j=1}^{m} z_j(x) = \sum_{j \in J_B} z_j(x) + \sum_{j \in J_G} z_j(x), \tag{12}$$

so that the objective function can be presented in the form:

$$minimize\ f(x) = f_B(x) + f_G(x), \tag{13}$$

where the two components are completely independent.

The first part of expression (13), associated with the boundary observations, can be calculated by using the previously presented smoothing approach, see (8) and (9). The second part of expression (13) can be calculated by using a faster procedure, as we will show right away.

Let us define the two parts in a more rigorous form. Let be $\bar{x}_i, i = 1, \dots, q$ be a referential position of centroids of the clusters taken in the iterative process.

The boundary concept in relation to the referential point $\bar{x}$ can be easily specified by defining a δ band zone between neighbouring centroids. For a generic point $s \in \mathbb{R}^n$, we define the first and second nearest distances from s to the centroids:

$$d_1(s,\bar{x}) = \| s - \bar{x}_{i_1} \| = \min_i \| s - \bar{x}_i \| \tag{14}$$

$$d_2(s,\bar{x}) = \| s - \bar{x}_{i_2} \| = \min_{i \neq i_1} \| s - \bar{x}_i \| \,, \tag{15}$$

where i_1 and i_2 are the labelling indexes of these two nearest centroids.

By using the above definitions, let us define precisely the δ boundary band zone:

$$Z_\delta(\bar{x}) = \{ s \in \mathbb{R}^n \mid d_2(s,\bar{x}) - d_1(s,\bar{x}) < 2\,\delta \} \tag{16}$$

and the gravity region, this is the complementary space:

$$G_\delta(\bar{x}) = \{ s \in \mathbb{R}^n - Z_\delta(\bar{x}) \} \,. \tag{17}$$

Figure 1 illustrates in $\mathbb{R}^2$ the $Z_\delta(\bar{x})$ and $G_\delta(\bar{x})$ partitions. The central lines form the Voronoi polygon associated with the referential centroids $\bar{x}_i, i = 1, \ldots, q$. The region between two parallel lines to Voronoi lines constitutes the boundary band zone $Z_\delta(\bar{x})$.

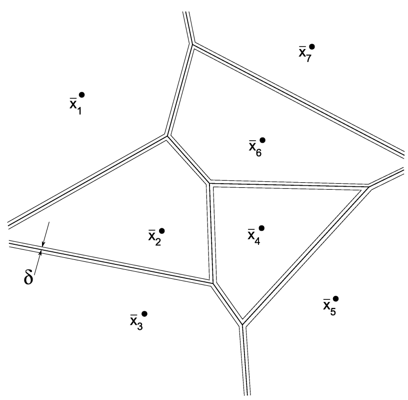

Fig. 1. The $Z_\delta(\bar{x})$ and $G_\delta(\bar{x})$ partitions.

Now, the sets J_B and J_G can be defined in a precise form:

$$J_B(\bar{x}) = \left\{ j = 1, \ldots, m \,|\, s_j \in Z_\delta(\bar{x}) \right\} , \tag{18}$$

$$J_G(\bar{x}) = \left\{ j = 1, \ldots, m \,|\, s_j \in G_\delta(\bar{x}) \right\} . \tag{19}$$

In Xavier & Xavier (2011) [7] it is shown the proof of proposition below.

Proposition 1:

Let s be a generic point belonging to the gravity region $G_\delta(\bar{x})$, with nearest centroid i_1. Let x be the current position of the centroids. Let $\Delta x = \max_i \|x_i - \bar{x}_i\|$

be the maximum displacement of the centroids. If $\Delta x < \delta$ then s will continue to be nearer to centroid x_{i_1} than to any other one. ∎

Since $\delta \geq \Delta x$, Proposition 1 makes it possible to calculate exactly expression (12) in a very fast way. First, let us define the subsets of gravity observations associated with each referential centroid:

$$J_i(\bar{x}) = \left\{ j \in J_G \,|\, \min_{p=1,\dots,q} \left\| s_j - \bar{x}_p \right\| = \left\| s_j - \bar{x}_i \right\| \right\} \tag{20}$$

Let us consider the second sum in expression (12).

$$f_G(x) = \sum_{j \in J_G} z_j(x) = \sum_{i=1}^{q} \sum_{j \in J_i} \left\| s_j - x_i \right\|_1 = \sum_{i=1}^{q} \sum_{j \in J_i} \sum_{l=1}^{n} |\, s_j^l - x_i^l \,| = \sum_{i=1}^{q} \sum_{l=1}^{n} \sum_{j \in J_i} |\, s_j^l - x_i^l \,|.$$

Let us now perform the partition of each set J_i into 3 subsets for each component l in the following form:

$$J_{il}^{+}(\bar{x}) = \left\{ j \in J_i(\bar{x}) \,|\, s_j^l - \bar{x}_i^l \geq \delta \right\} \tag{21}$$

$$J_{il}^{-}(\bar{x}) = \left\{ j \in J_i(\bar{x}) \,|\, s_j^l - \bar{x}_i^l \leq -\delta \right\} \tag{22}$$

$$J_{il}^{0}(\bar{x}) = \left\{ j \in J_i(\bar{x}) \,|\, -\delta < s_j^l - \bar{x}_i^l < \delta \right\} \tag{23}$$

By using the defined subsets, it is obtained:

$$f_G(x) = \sum_{i=1}^{q} \sum_{l=1}^{n} \left[\sum_{j \in J_{il}^{+}} |\, s_j^l - x_i^l \,| + \sum_{j \in J_{il}^{-}} |\, s_j^l - x_i^l \,| + \sum_{j \in J_{il}^{0}} |\, s_j^l - x_i^l \,| \right] =$$

$$\sum_{i=1}^{q} \sum_{l=1}^{n} \left[\sum_{j \in J_{il}^{+}} |\, s_j^l - \bar{x}_i^l + \bar{x}_i^l - x_i^l \,| + \sum_{j \in J_{il}^{-}} |\, s_j^l - \bar{x}_i^l + \bar{x}_i^l - x_i^l \,| + \sum_{j \in J_{il}^{0}} |\, s_j^l - x_i^l \,| \right]$$

Let us define the component displacement of centroid $\Delta x_i^l = x_i^l - \bar{x}_i^l$. Since $|\,\Delta x_i^l\,| < \delta$, from the above definitions of the subsets, it follows that:

$$|\, s_j^l - x_i^l \,| = |\, s_j^l - \bar{x}_i^l \,| - \Delta x_i^l \quad \text{for} \quad j \in J_{il}^{+} \tag{24}$$

$$|\, s_j^l - x_i^l \,| = |\, s_j^l - \bar{x}_i^l \,| + \Delta x_i^l \quad \text{for} \quad j \in J_{il}^{-}$$

So, it follows:

$$f_G(x) = \sum_{i=1}^{q} \sum_{l=1}^{n} \left[\sum_{j \in J_{il}^{+}} \left(|\, s_j^l - \bar{x}_i^l | - \Delta x_i^l \right) + \sum_{j \in J_{il}^{-}} \left(|\, s_j^l - \bar{x}_i^l \,| + \Delta x_i^l \right) + \sum_{j \in J_{il}^{0}} |\, s_j^l - x_i^l \,| \right] =$$

$$\sum_{i=1}^{q}\sum_{l=1}^{n}\left[\sum_{j\in J_{il}^{+}}|s_j^l-\overline{x}_i^l| - |J_{il}^{+}|\,\Delta x_i^l + \sum_{j\in J_{il}^{-}}|s_j^l-\overline{x}_i^l| + |J_{il}^{-}|\,\Delta x_i^l + \sum_{j\in J_{il}^{0}}|s_j^l-x_i^l|\right] \tag{25}$$

where $|J_{il}^{+}|$ and $|J_{il}^{-}|$ are the cardinalities of two first subsets.

When the position of centroids $x_i, i=1,\dots,q$ moves within the iterative process, the value of the first two sums of (25) assumes a constant value, since the values s_j^l and $\overline{x}_i^l$ are fixed. So, to evaluate $f_G(x)$ it is only necessary to calculate the displacements $\Delta x_i^l, i=1,\dots,q, l=1,\dots,n,$ and evaluate the last sum, that normally has only a few number of terms because δ assumes in general a relatively small value.

The function $f_G(x)$ above specified is non differentiable due the last sum, so in order to use gradient information, it is necessary to use a smooth approximation:

$$f_G(x)=\sum_{i=1}^{q}\sum_{l=1}^{n}\left[\sum_{j\in J_{il}^{+}}|s_j^l-\overline{x}_i^l| - |J_{il}^{+}|\,\Delta x_i^l + \sum_{j\in J_{il}^{-}}|s_j^l-\overline{x}_i^l| + |J_{il}^{-}|\,\Delta x_i^l + \sum_{j\in J_{il}^{0}}\sigma(s_j^l,x_i^l,\gamma)\right] \tag{26}$$

where σ is the smoothing function for each unidimensional distance: $\sigma(s_j^l,x_i^l,\gamma) = ((s_j^l-x_i^l)^2+\gamma^2)^{1/2}$.

So, the gradient of the smoothed second part of objective function is easily calculated by:

$$\nabla f_G(x)=\sum_{i=1}^{q}\sum_{l=1}^{n}\left[-|J_{il}^{+}|+|J_{il}^{-}|+\sum_{j\in J_{il}^{0}}-(s_j^l-x_i^l)\,/\,\sigma(s_j^l,x_i^l,\gamma)\right]e_{il} \tag{27}$$

where e_{il} stands for a unitary vector with the component l of centroid i equal to 1.

Therefore, if $\delta \geq \Delta x$ was observed within the iterative process, the calculation of the expression $\sum_{j\in J_G} z_j(x)$ and its gradient can be exactly performed by very fast procedures, equations (26) and (27).

By using the above results, it is possible to construct a specific method, the Accelerated Hyperbolic Smoothing Method Applied to the Minimum of Sum of L1 Distances Clustering Problem, which has conceptual properties to offer a faster computational performance for solving this specific clustering problem given by formulation (13), since the calculation of the second sum is very simple.

A fundamental question is the proper choice of the boundary parameter δ. Moreover, there are two main options for updating the boundary parameter δ, inside the internal minimization procedure or after it. For simplicity sake, the AHSC-L1 method connected with the partition scheme presented below adopts the second option, which offers a better computational performance, in spite of an eventual violation of the $\delta \geq \Delta x$ condition, which is corrected in the next partition update.

The Simplified AHSC-L1 Algorithm

Initialization Step:

Choose initial start point: x^0;

Choose parameter values: γ^1 , τ^1 , ε^1;

Choose reduction factors:

$0 < \rho_1 < 1,\ \ 0 < \rho_2 < 1,\ \ 0 < \rho_3 < 1;$

Specify the boundary band width: δ^1;

Let $k = 1$.

Main Step: Repeat until an arbitrary stopping rule is attained

For determining the $Z_\delta(\bar{x})$ and $G_\delta(\bar{x})$ partitions, given by (16) and (17), use $\bar{x} = x^{k-1}$ and $\delta = \delta^k$.

Determine the subsets J^+_{il} , J^-_{il} and J^0_{il} and calculate the cardinalities of two first sets: $|J^+_{il}|$ and $|J^-_{il}|$.

Solve problem (13) starting at the initial point x^{k-1} and let x^k be the solution obtained:

For solving the equations associated to the first part given by (9), take the smoothing parameters:

$\gamma = \gamma^k,\ \ \tau = \tau^k$ and $\varepsilon = \varepsilon^k$.

For solving the second part, given by (26), use the above determined subsets and their cardinalities.

Updating procedure:

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

If necessary redefine the boundary value: δ^{k+1} .

Let $k := k+1$. ■

The efficiency of the AHSC-L1 algorithm depends strongly on the parameter δ. A choice of a small value for it will imply an improper definition of the set $G_\delta(\bar{x})$, and frequent violation of the basic condition $\Delta x < \delta$, for the validity of Proposition 1. Otherwise, a choice of a large value will imply a decrease in the number of gravitational observation points and, therefore, the computational advantages given by formulation (26) will be reduced.

As a general strategy, within first iterations, larger δ values must be used, because the centroid displacements are more expressive. The δ values must be gradually decreased in the same proportion of the decrease of these displacements.

5 Computational Results

The numerical experiments have been carried out on a PC Intel Celeron with 2.7GHz CPU and 512MB RAM. The programs are coded with Compac Visual FORTRAN, Version 6.1. The unconstrained minimization tasks were carried out by means of a Quasi-Newton algorithm employing the BFGS updating formula from the Harwell Library, obtained in the site: (http://www.cse.scitech.ac.uk/nag/hsl/).

In order to exhibit the distinct performance of the AHSC-L1 algorithm, tables 1 and 2 present the computational results of AHSC-L1 applied to 4 benchmark problems, all from TSPLIB (http://www.iwr.uni-heidelberg.de/groups/comopt/software). Table 1 represent 2 instances frequently used as benchmark clustering problems. Table 2 left contains data of 15112 German cities. Table 2 right is the largest symmetric problem of TSPLIB.

The AHSC-L1 is a general framework that bears a broad number of implementations. In the initialization steps the following choices were made for the reduction factors: $\rho_1 = 1/4$, $\rho_2 = 1/4$ and $\rho_3 = 1/4$. The specification of initial smoothing and perturbation parameters was automatically tuned to the problem data. So, the initial *max* function smoothing parameter (4) was specified by $\tau^1 = \sigma / 10$ where σ^2 is the variance of set of observation points: $S = \{s_1, \dots, s_m\}$. The initial perturbation parameter (3) was specified by $\varepsilon^1 = 4\tau^1$ and the Euclidian distance smoothing parameter by $\gamma^1 = \tau^1 / 100$.

All experiments where done using 10 initial points. The adopted stopping criterion was the execution of the main step of the AHSC-L1 algorithm in a fixed number of 6 iterations.

The meaning of the columns tables 1 and 2 is as follows. q = the number of clusters. $f_{AHSC-L1-Best}$ = the best results of cost function using points obtained from AHSC-L1 method out of all the 10 random initial points. Occur. = number of times the same best result was obtained from all the 10 random initial points. E_{Mean} = the average error of the 10 solutions in relation to the best solution obtained ($f_{AHSC-L1-Best}$). Finally, T_{cpu} = the average execution time per trial, in seconds.

	TSPLIB 1060				TSPLIB 3038			
q	$f_{AHSC-L1_{Best}}$	Occur.	E_{Mean}	T_{cpu}	$f_{AHSC-L1_{Best}}$	Occur.	E_{Mean}	T_{cpu}
2	0.386500E7	2	0.00	0.10	0.373171E7	2	0.88	0.76
3	0.313377E7	1	0.22	0.14	0.300708E7	1	1.33	0.76
4	0.258205E7	5	0.43	0.21	0.254499E7	1	0.67	0.79
5	0.231098E7	1	0.76	0.29	0.225571E7	1	1.28	0.95
6	0.213567E7	1	0.79	0.35	0.206006E7	1	1.12	0.95
7	0.196685E7	1	1.41	0.48	0.189650E7	1	1.34	1.06
8	0.183280E7	1	2.22	0.53	0.176810E7	1	1.14	1.13
9	0.168634E7	1	3.42	0.60	0.164559E7	1	2.13	1.21
10	0.155220E7	1	2.74	0.68	0.154550E7	1	1.97	1.28

Table 1: Results of AHSC-L1 applied to TSPLIB-1060 and TSPLIB-3038 Instance

The “A” of AHSC-L1 means “accelerated”, that is, the technique that partitions the set of observations into two non overlapping groups: “data in frontier” and “data in gravitational regions”. The sample problems were solved using HSC-L1 and AHSC-L1 methods. Both algorithms obtain the same results with 3 decimal digits of precision. The table 3 shows the speed-up produced by the acceleration technique. The meaning of the columns of table 3 is as follows. q = the number of clusters. Speed-up for TSPLIB-1060 Instance. Speed-up for TSPLIB-3038 Instance. Speed-up for D15112 Instance. Speed-up for Pla85900 Instance.

	D15112				Pla85900			
q	$f_{AHSC-L1_{Best}}$	Occur.	E_{Mean}	T_{cpu}	$f_{AHSC-L1_{Best}}$	Occur.	E_{Mean}	T_{cpu}
2	0.822872E8	6	0.97	10.35	0.883378E10	2	0.00	242.74
3	0.655831E8	1	1.43	7.56	0.667961E10	1	0.21	166.48
4	0.567702E8	1	1.60	6.21	0.551287E10	2	0.06	129.30
5	0.511639E8	1	1.17	5.73	0.482328E10	1	1.43	112.95
6	0.462612E8	1	1.67	5.33	0.432972E10	1	2.47	103.02
7	0.425722E8	1	2.30	4.96	0.401388E10	1	1.87	98.25
8	0.398389E8	1	1.83	5.02	0.373878E10	1	3.47	92.14
9	0.376863E8	1	1.60	5.03	0.355741E10	1	2.40	82.33
10	0.354762E8	1	2.41	5.01	0.341472E10	1	1.77	87.41

Table 2: Results of AHSC-L1 applied to D15112 and Pla85900 Instance

The speed-up was calculated as the ratio between execution times T_{HSC-L1} and $T_{AHSC-L1}$, as shown in equation (28).

$$Speed-up = \frac{T_{HSC-L1}}{T_{AHSC-L1}} \tag{28}$$

The results in the table 3 show that in most cases the "accelerated" technique produces speed-up of the computation effort. In some cases, the speed-up is > 10. In a few cases (e.g. $q = 2$, 85900), the gains produced by the acceleration do not compensate the fixed costs introduced by the calculus of partition. In these cases the speed-up is less than one, that is, AHSC-L1 takes longer to run when compared to HSC-L1.

q	TSP1060	TSP3038	D15112	Pla85900
2	2.60	1.33	0.94	0.82
3	3.79	1.84	1.43	0.95
4	3.24	3.08	1.95	1.30
5	4.41	3.16	3.17	1.76
6	5.23	4.69	4.60	2.58
7	4.96	5.84	5.94	3.80
8	6.94	7.09	8.44	4.44
9	6.30	7.97	10.34	6.78
10	7.66	12.07	13.26	8.75

Table 3: Speed-up of AHSC-L1 compared to HSC-L1 (the larger the better)

6 Conclusions

In this paper, a new method for the solution of the minimum sum of L1 Manhattan distances clustering problem is proposed, called AHSC-L1 (Accelerated Hyperbolic Smoothing Clustering - L1). It is a natural development of the original HSC method and its descendant AHSC-L2 method, linked to the the minimum sum-of-squares clustering (MSSC) formulation, presented respectively by Xavier (2010) [6] an by Xavier and Xavier (2011) [7].

The special characteristics of L1 distance were taken into account to adapt inside the AHSC-L1 method from the AHSC-L2. The main idea proposed in this paper is the acceleration of the AHSC-L1 method by the partition of the set of observations into two non overlapping parts - gravitational and boundary. The classification of gravitational observations by each component, implemented by equations (21), (22) and (23), simplifies of the calculation of the objective function (26) and its gradient (27). This classification produces a drastic simplification of computational tasks. The computational experiments confirm the speed-up, as shown in Table 3.

The computational experiments presented in this paper were obtained by using a particular and simple set of criteria for all specifications. The AHSC-L1 algorithm is a general framework that can support different implementations.

We could not find in the literature any reference mentioning the solution of cluster L1 problem with instances of sizes similar to those presented in this paper. So, our results represent a challenge for future works.

The most relevant computational task associated with the AHSC-L1 algorithm remains the determination of the zeros of the equations (9), for each observation in the boundary region, with the purpose of calculating the first part of the objective function. However, since these calculations are completely independent, they can be easily implemented using parallel computing techniques.

It must be observed that the AHSC-L1 algorithm, as presented here, is firmly linked to the MSDC-L1 problem formulation. Thus, each different problem formulation requires a specific methodology to be developed, in order to apply the partition into boundary and gravitational regions.

Finally, it must be remembered that the MSDC-L1 problem is a global optimization problem with several local minima, so both HSC-L1 and AHSC-L1 algorithms can only produce local minima. The obtained computational results exhibit a deep local minima property, which is well suited to the requirements of practical applications.

Acknowledgement

The author would like to thank Cláudio Joaquim Martagão Gesteira of Federal University of Rio de Janeiro for the helpful review of the work and constructive comments.

References

1. P. S. Bradley and O. L. Mangasarian, *Clustering via Concave Minimization* Mathematical Programming Technical Report 96-03, May 1996. Advances in Neural Information Processing Systems 9, MIT Press, Cambridge, MA 1997, 368-374.
2. J. A. Hartigan, *Clustering Algorithms*, John Wiley and Sons, Inc., New York, NY, 1975.
3. G. Reinelt, *TSP-LIB - A Traveling Salesman Library*, ORSA J. Comput, 1991. pp. 376-384
4. H. Späth, *Cluster Analysis Algorithms for Data Reduction and Classification*, Ellis Horwood, Upper Saddle River, NJ, 1980.
5. A.E. Xavier, *Penalização Hiperbólica: Um Novo Método para Resolução de Problemas de Otimização*, M.Sc. Thesis - COPPE - UFRJ, Rio de Janeiro, 1982.
6. A.E. Xavier, *The Hyperbolic Smothing Clustering Method*, Pattern Recognition, Vol 43, pp. 731-737, 2010.
7. A.E. Xavier and V.L. Xavier, *Solving the Minimum Sum-of-Squares Clustering Problem by Hyperbolic Smoothing and Partition into Boundary and Gravitational Regions* , Pattern Recognition, Vol. 44, pp 70-77, 2011.