Solving Medium and Large Size Problems of the Literature by the Accelerated Hyperbolic Smoothing Clustering Method

Vinicius Layter Xavier

Adilson Elias Xavier
Dept. of Systems Engineering and Computer Science, Graduate School of Engineering (COPPE)
{vinicius , adilson}@cos.ufrj.br

RESUMO

Este trabalho considera o problema de agrupamento segundo o critério de mínima soma de quadrados de distâncias. Sua formulação é do tipo $min - sum - min,$ que, em adição a sua intrínseca natureza dois-níveis, tem a significante característica de ser fortemente não-diferenciável. O método de resolução proposto adota uma estratégia de suavização que engendra um problema irrestrito completamente diferenciável C^∞. O algoritmo proposto aplica também uma partição do conjunto de observações em dois grupos disjuntos: "dados na fronteira" e "dados nas regiões gravitacionais", que drasticamente simplifica as tarefas computacionais. Resultados de experimentos numéricos em tradicionais problemas teste da literatura mostram um desempenho inaudito da metodologia proposta.

PALAVRAS CHAVE. Análise de Agrupamentos, Problemas Min-Sum-Min, Programação Não-Diferenciável, Suavização.

Área Principal: Análise de Agrupamento ou Clustering

ABSTRACT

This article considers the minimum sum-of-squares clustering (MSSC) problem. 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. The proposed resolution method adopts a smoothing strategy which engenders an unconstrained completely differentiable problem C^∞. The proposed algorithm applies also a partition of the set of observations into two non overlapping groups: "data in frontier" and "data in gravitational regions", which drastically simplify the computational tasks. Results of numerical experiments on traditional test problems of the literature show an unprecedented performance of the proposed methodology.

KEYWORDS. Cluster Analysis, Min-Sum-Min Problems, Nondifferentiable Programming, Smoothing.

Main Area: Clustering

1. Introduction

Cluster analysis deals with the problems of classification of a set of patterns or observations, in general represented as points in a multidimensional space, into clusters, 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), as recorded by Anderberg (1973), Hartingan (1975) and Späth (1980).

In this paper, a particular clustering problem formulation is considered. Among many criteria used in cluster analysis, the most natural, intuitive and frequently adopted criterion is the minimum sum-of-squares clustering (MSSC). The minimum sum-of-squares clustering (MSSC) formulation produces a mathematical problem of global optimization. It is both a non-differentiable and a non-convex mathematical problem, with a large number of local minimizers.

There are two main strategies for solving clustering problems: hierarchical clustering methods and partition clustering methods. Hierarchical methods produce a hierarchy of partitions of a set of observations. Partition methods, in general, assume a given 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 (Mc Queen (1967)) and its modern variations [Likas et al (2003), Bagirov (2008) and Bagirov et al (2011)] are broadly used in practical applications (Wu et al (2008)).

We adopt a novel strategy: the smoothing of the MSSC clustering problem. For the sake of completeness, we present first the Hyperbolic Smoothing Clustering Method (HSCM), Xavier (2010). Basically the method performs the smoothing of the non-differentiable $min-sum-min$ clustering formulation. 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^∞ unconstrained differentiable single-level alternative.

In second place, we present the Accelerated Hyperbolic Smoothing Clustering Method, Xavier and Xavier (2011), a new faster implementation. 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. This set of observations, named boundary band points, can be managed by using the previously presented smoothing approach. The second set corresponds to observation points significantly closer to a single centroid in comparison with others. This set of observations, named gravitational points, is managed in a direct and simple way, offering much faster performance.

2. The Hyperbolic Smoothing Clustering Method

Let $S = \{s_1, \ldots, s_m\}$ 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 $x_i, i = 1, \ldots, q$ 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 Euclidean distance from s_j to the center in X that is nearest. This is given by $z_j = \min_{i=1,\ldots,q} \|s_j - x_i\|_2$. 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, which determines the MSSC problem:

$$minimize \sum_{j=1}^{m} z_j^2 \qquad (1)$$

$$\text{subject to} \quad z_j = \min_{i=1,\ldots,q} \|s_j - x_i\|_2, \quad j = 1, \ldots, m$$

The Hyperbolic Smoothing Clustering Method performs a set of transformations in order to obtain a completely differentiable formulation see Xavier (2010) and Xavier and Xavier (2011). We do so by first letting $\varphi(y)$ denote $max\{0, y\}$ and by including an $\varepsilon > 0$ perturbation in (1). So, the following modified problem is obtained:

$$minimize \sum_{j=1}^{m} z_j^2 \tag{2}$$

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

By using function $\phi(y,\tau) = \left(y + \sqrt{y^2+\tau^2}\,\right)/2$ in the place of function $\varphi(y)$ and by using function $\theta(\,s_j\,,\,x_i\,,\,\gamma\,) \;=\; (\,\sum_{l=1}^{n}(s_j^l - x_i^l)^2 \;+\; \gamma^2\,)^{1/2}$ in the place of the Euclidian distance $\|s_j - x_i\|_2$, the following completely differentiable problem is now obtained:

$$minimize \sum_{j=1}^{m} z_j^2 \tag{3}$$

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

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

$$minimize \;\; f(x) \;=\; \sum_{j=1}^{m} z_j(x)^2 \tag{4}$$

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

$$h_j(z_j, x) \;=\; \sum_{i=1}^{q} \phi(z_j - \theta(s_j, x_i, \gamma), \tau) \;-\; \varepsilon \;=\; 0, \qquad j = 1, \dots, m. \tag{5}$$

Due to property of the hyperbolic smoothing function, see Xavier (2010) and Xavier and Xavier (2011), each term ϕ above is strictly increasing with variable z_j and therefore the equation has a single zero.

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 (4),

$$\nabla\, f(x) \;=\; \sum_{j=1}^{m} 2\, z_j(x)\, \nabla z_j(x) \tag{6}$$

where

$$\nabla z_j(x) \;=\; -\,\nabla\, h_j(z_j,x)\;\;/\;\;\frac{\partial\, h_j(z_j,x)}{\partial\, z_j}, \tag{7}$$

while $\nabla\, h_j(z_j,x)$ and $\partial\, h_j(z_j,x)/\partial\, z_j$ are obtained from equations (5) and from definitions of function $\phi(y,\tau)$ and function $\theta(\,s_j\,,\,x_i\,,\,\gamma\,)$.

In this way, it is easy to solve problem (4) by making use of any method based on first order derivative information. Finally, it must be emphasized that problem (4) is defined on a $(nq)-$dimensional space, so it is a small problem, since the number of clusters, q, is, in general, very small for real applications.

3. The Accelerated Hyperbolic Smoothing Clustering Method

Xavier and Xavier (2011) introduced a faster procedure to the original Hyperbolic Smoothing Clustering Method. The basic idea is the partition of the set of observations into two non overlapping regions. By using a conceptual presentation, the first region J_B, the set of boundary observations, corresponds to the observation points that are relatively close to two or more centroids, within a specified δ tolerance. The second region J_G, the set of gravitational observations, corresponds to the observation points that are significantly close to a unique centroid in comparison with the other ones. Considering this partition, equation (4) can be expressed in the following way:

$$minimize\;\; f(x) \;=\; f_B(x) \,+\, f_G(x) \;=\; \sum_{j\in J_B} z_j(x)^2 \;+\; \sum_{j\in J_G} z_j(x)^2. \tag{8}$$

The first part of expression (8), associated with the boundary observations, can be calculated by using the previous presented smoothing approach, see (4) and (5). The second part of expression (8) can be calculated by using a faster procedure:

$$minimize\;\; f_G(x) \;=\; \sum_{i=1}^{q}\sum_{j\in J_i} \|\, s_j - v_i\,\|^2 \;+\; \sum_{i=1}^{q} |J_i|\;\|\, x_i - v_i\,\|^2 \tag{9}$$

where v_i is the center of the observations in each non-empty subset:

$$v_i \;=\; \frac{1}{|J_i|}\sum_{s_j\in J_i} s_j, \quad \forall\, i = 1,\dots,q. \tag{10}$$

When the position of centroids $x_i\,, i = 1,\dots,q$ moves during the iterative process, the value of the first sum in (9) assumes a constant value, since the vectors s and v are fixed. On the other hand, for the calculation of the second sum, it is only necessary to calculate q distances, $\|\, v_i\; - x_i\,\|\,,\, i = 1,\dots,q$.

The gradient of the second part of the objective function is easily calculated by:

$$\nabla\, f_G(x) \;=\; \sum_{i=1}^{q} 2\,|J_i|\,(\,x_i - v_i\,), \tag{11}$$

where the vector $(v_i - x_i)$ must be in $\mathbb{R}^{nq}$, so it has the first $(i-1)q$ components and the last $l = iq+1,\dots,nq$ components equal zero.

A Simplified Version of AHSCM Algorithm

Initialization Step:

Choose the starting point: x^0

Choose smoothing 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

Determine partitions J_B and J_B by using tolerance $\delta = \delta^k$.

Calculate centres $v_i, i = 1, \dots, q$ of gravitational regions by (10)

Solve problem (8) starting at the initial point x^{k-1} and let x^k be the solution obtained:

For solving the first part $\sum_{j \in J_B} z_j(x)^2$, associated to the boundary band zone, calculate each zero of the equation (5) by using the smoothing parameters: $\gamma = \gamma^k$, $\tau = \tau^k$ and $\varepsilon = \varepsilon^k$;

For solving the second part, given by (9), use the above calculated centres of the gravitational regions.

Updating procedure:

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

Redefine the boundary value: δ^{k+1}

Let $k := k + 1$. ■

4. Computational Results

In order to verify the performance of the incremental version of the Accelerated Hyperbolic Smoothing Clustering Method, AHSCM algorithm we perform computational experiments with some traditional test problems of the literature. In order to demonstrate accuracy and efficiency of the proposed algorithm we perform a comparison of our computational results with those presented by Bagirov et al (2013).

This paper presents computational results produced by three different algorithms: GKM (Likas et al (2003)) - the global k-means , MGKM (Bagirov (2008)) - the modified global k-means and HSCM an incremental version the original hyperbolic smoothing clustering method (Xavier (2010)) without the pruning procedure provided by the partition scheme and with initial starting points generated by an incremental procedure. These numerical experiments by Bagirov et al (2013) have been carried on a PC Intel(R) Core(TM) i5-34705 with CPU 2.90 GHz and RAM 8 GB.

The numerical experiments associated to the AHSCM algorithm, a new version with initial starting points generated by an incremental procedure, have been carried out on a PC Intel (R) Core (TM) i7-2620M Windows Notebook with 2.70GHz and 8 GB RAM. The programs are coded with Intel(R) Visual Fortran Composer XE 2011 Update 7 Integration for Microsoft Visual Studio* 2010. 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:

(www.cse.scitech.ac.uk/nag/hsl/).

Table 1 presents a short description of the 8 used test problems, which correspond to same medium size and large size data sets considered in the referential work. Test TSPLIB1060, TSPLIB3038, D15112 and Pla85900 belong to the collection of salesman travel problems organized by Reinelt (1991) and other belongs to UCI repository:
http://www.ics.uci.edu/mlearn/MLRepository.html.

Data sets	Number of instances	Number of attributes
Breast cancer	683	9
TSPLIB1060	1060	2
TSPLIB3038	3038	2
Pendigit	10992	16
D15112	15112	2
Letters	20000	16
Shuttle Control	58000	9
Pla85900	85900	2

Table 1. Short description of test problems

The computational results presented below were obtained from a particular implementation of the incremental AHSCM algorithm. The AHSC-L2 is a general framework that bears a broad numbers 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 smoothing parameter τ^1 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 (2) was specified by $\epsilon^1 = 4\,\tau^1$ and the Euclidean distance smoothing parameter by $\gamma^1 = \tau^1 / 100$. The boundary width parameter δ^k at the beginning of each iteration k was specified by using the average distance between all pairs of the initial position of centroids times $\alpha = 0.05$, a fixed factor.

The adopted stopping criterion was the execution of the main step for the ASHCM algorithm a fixed number of 10 iterations. In this way, the final values of the τ, ϵ, and γ parameters were reduced to $1/(4^9)$ of the initial values. The adopted stopping criteria for the unconstrained minimization procedure was fixed in all iterations, supplying precise solutions with 10 significant digits.

Tables 2 - 3 present a synthesis of the computational results produced by the new version of AHSCM algorithm. For each data set, we vary the number of clusters $q = 2, 5, 10, 15, 20$ and 25. For each data set and for each number of clusters, only one initial starting point was used. The first columns show the number of clusters (q) and the putative global optimum (f_{opt}), that is the best known solution produced at all times. Then, for each algorithm: GKM - the global k-means, MGKM - the modified global k-means, HSCM - the original hyperbolic smoothing clustering method with initial starting points generated by an incremental procedure and the novel version of AHSCM - algorithm, are presented: (E), the relative percentage deviation of the obtained solution f^* in relation to the putative value (f_{opt}), so $E = 100\,(f^* - f_{opt})/f_{opt}$, and (T), the CPU time given in seconds.

Table 2 presents the results for the medium size data sets: Breast cancer, TSPLIB1060, TSPLIB3038 and Pendigit. As registered by Bagirov et al (2013), the HSCM is able to fins solutions with high accuracy since it locates the putative global solutions, otherwise values near to these solutions. On other hand, HSCM requires significantly more CPU time than GKM and MGKM algorithms. Now, the novel proposed version of AHSCM offers an analogous accuracy than HSCM, but with a expressive gain on the speed. For the instance Pendigit, the CPU time assumes a value smaller than other algorithms.

Table 3 presents the results for the large size data sets: D15112, Letters, Shuttle Control and Pla85900. We can make same comments about the accuracy performance, but now the speed-up offered by AHSCM assumes high values. So, these results demonstrate that the novel proposed AHSCM produces solutions to the clustering problems with high precision in short CPU times. For all instances (D15112, Letters, Shuttle Control and Pla85900) the CPU time assumes a value smaller than other algorithms. Moreover, as larger is the size of the problem larger is the speed-up given by the AHSCM algorithm. For instances Shuttle Control and Pla85900, the speed-up can assume values until 200 times or more yet.

Tables 2 - 3 present the obtained results by the novel proposed version of AHSCM. In the total of 48 cases (8 instances times 6 different number of clusters), the AHSCM algorithm obtained 10 new putative global solutions, best known solution produced at all times, by using only one initial starting point.

q	f_{opt}	GKM		MGKM		HSCM		AHSCM	
		E	T	E	T	E	T	E	T
Breast Cancer									
2	1.9323E4	0	0	0	0.02	0	0.05	0.00	0.04
5	1.3705E4	2.28	0.03	1.86	0.05	0	0.61	0.00	0.14
10	1.0190E4	0.26	0.06	0.28	0.11	0	2.14	0.16	0.37
15	8.6921E3	1.02	0.08	1.07	0.16	0	9.08	**-0.36**	0.62
20	7.6478E3	3.64	0.11	1.80	0.20	0	27.36	**-0.19**	1.12
25	6.9046E3	5.08	0.14	0.93	0.27	0	44.65	0.11	1.53
TSPLIB1060									
2	9.8319E9	0	0	0	0.02	0	0.05	0.31	0.04
5	3.7910E9	0.01	0.03	0.01	0.05	0	0.17	0.54	0.13
10	1.7548E9	0.23	0.06	0.05	0.09	0	0.70	0.42	0.39
15	1.1212E9	0.09	0.09	0.06	0.16	0.02	1.25	1.21	0.87
20	7.9179E8	0.41	0.12	0.41	0.20	0.03	2.70	1.35	1.55
25	6.0670E8	1.80	0.16	1.80	0.25	0.02	4.63	1.81	2.03
TSPLIB3038									
2	3.1688E9	0	0.06	0	0.11	0	0.25	0.06	0.11
5	1.1982E9	0	0.25	0	0.39	0	0.98	0.00	0.21
10	5.6025E8	2.78	0.48	0.58	0.81	0	2.64	0.56	0.69
15	3.5604E8	0.07	0.70	1.06	1.20	0	5.16	0.03	1.40
20	2.6681E8	2.00	0.94	0.48	1.61	0.11	8.86	0.22	2.39
25	2.1450E8	0.78	1.20	0.23	1.98	0.01	14.01	0.18	3.84
Pendigit									
2	1.2812E8	0.39	2.56	0	5.23	0	5.37	0.00	0.38
5	7.5304E7	0	9.73	0	18.72	0	22.60	0.00	1.76
10	4.9302E7	0	20.45	0	39.31	0	61.71	0.00	4.99
15	3.9067E7	0	30.79	0	59.19	0	120.46	0.00	10.33
20	3.4123E7	0	41.25	0.17	78.37	0.16	250.80	**-0.31**	19.21
25	3.0038E7	0.24	51.87	0.24	98.47	0	446.37	**-0.12**	33.06

Table 2. Results for medium size data sets

5. Conclusions

In this paper, computational experiments were performed in order to evaluate the performance of the AHSCM algorithm for solving 8 traditional instances of the literature.

q	f_{opt}	GKM		MGKM		HSCM		AHSCM	
		E	T	E	T	E	T	E	T
D15112									
2	3.6840E11	0	2.75	0	4.51	0	5.15	0.00	0.31
5	1.3271E11	0	8.50	0	13.85	0	18.41	0.00	1.36
10	6.4892E10	0.78	14.87	0.78	24.98	0	41.50	**-0.62**	4.21
15	4.3136E10	0.26	21.00	0.26	35.51	0	68.23	0.15	9.31
20	3.2177E10	0.25	26.99	0.25	45.68	0	101.82	0.62	17.43
25	2.5423E10	0.03	32.71	0.03	55.47	0	139.53	**-0.49**	30.22
Letters									
2	1.3819E6	0	9.63	0	17.35	0	14.40	0.00	0.43
5	1.0771E6	1.94	38.28	0.87	61.78	0	63.10	0.87	2.03
10	8.5750E5	0	79.76	0	131.77	0.21	163.71	0.00	7.61
15	7.4457E5	0.48	116.70	0	200.07	0.40	312.25	**-0.08**	19.89
20	6.7394E5	0.53	153.21	0.34	265.97	0	509.45	0.30	26.06
25	6.2287E5	1.47	187.79	0.58	331.07	0	749.96	**-0.53**	53.17
Shuttle control									
2	2.1343E9	0	63.20	0	123.82	0	86.74	0.00	0.20
5	7.2448E8	0	251.65	0	521.81	0	351.50	0.00	1.35
10	2.8317E8	0.02	581.28	0	1207.20	0	813.73	0.00	4.87
15	1.5315E8	0	888.24	0	1776.88	0	1272.75	0.00	11.09
20	1.0601E8	0	1195.84	0	2338.99	0	1798.88	**-3.60**	26.80
25	7.9776E7	0.17	1512.55	0.17	2917.41	0	2396.82	**-3.13**	51.97
Pla85900									
2	3.7491E15	0	89.54	0	115.99	0	123.96	0.00	1.17
5	1.3397E15	0	408.60	0	512.31	0	452.96	0.00	5.04
10	6.8294E14	0	754.67	0	994.29	0	1011.17	0.00	15.66
15	4.6029E14	0.51	1083.04	0.51	1448.94	0	1596.67	0.00	38.01
20	3.5087E14	0.01	1355.85	0	1844.34	0	2210.91	0.00	73.09
25	2.8323E14	0.74	1570.37	0.73	2186.20	0	2863.18	0.13	135.62

Table 3. Results for large size data sets

In short, this computational experiments show a high level of performance of the algorithm according to the different criteria of consistency, robustness and efficiency. The robustness and consistency performances can be attributed to the complete differentiability of the approach. The high speed of the algorithm can be attributed to the partition of the set of observations into two non overlapping parts, which simplifies drastically the computational tasks.

References

ANDERBERG, M. R. (1973), *Cluster Analysis for Applications*, Academic Press Inc., New York.
BAGIROV, A. M. (2008), *Modified Global k-means Algorithm for Minimum Sum-of-Squares Clustering Problems*, Pattern Recognition, Vol 41 Issue 10 pp. 3192-3199.
BAGIROV, A.M., UGON, J. and WEBB (2011), *Fast modified global k-means algorithm for incremental cluster construction*, Pattern Recognition, Vol 44, Issue 4, pp. 866-876.
BAGIROV, A.M. et al (2013), *An incremental clustering algorithm based on hyperbolic smoothing*, submitted to Computational Optimization and Applications.
HARTINGAN, J. A. (1975) , *Clustering Algorithms*, John Wiley and Sons, Inc., New York, NY.
LIKAS, A., VLASSIS, M., and VERBEEK, J. (2003), *The Global k-means Clustering Algorithm*, Pattern Recognition, 36, 451-461.
MC QUEEN, J. (1967), *Some Methods for Classification and Analysis of Multivariate Observations*, in Proceedings of the Fifth Berkeley Symposiums on Mathematical Statistics and Probability, 281-297.
REINELT, G (1991), *TSP-LIB: A Traveling Salesman Library* , ORSA Journal on Computing, 3, 376-384.
SPÄTH, H. (1980), *Cluster Analysis Algorithms for Data Reduction and Classification*, Ellis Horwood, Upper Saddle River, NJ.
WU, X. et alli, *Top 10 algorithms in data mining*, Knowledge and Information Systems, Springer, 14, 1-37, 2008.
XAVIER, A.E. (1982), *Hyperbolic penality: A new method for nonlinear programming with inequalities*, International Transactions in Operational Research, 8, 659-672.
XAVIER, A.E. (2010), *The Hyperbolic Smoothing Clustering Method*, Pattern Recognition, 43, 731-737.
XAVIER, A.E. and XAVIER, V.L. (2011), *Solving the Minimum Sum-of-Squares Clustering Problem by Hyperbolic Smoothing and Partition into Boundary and Gravitational Regions*, Pattern Recognition, 44, 70-77.