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

**Vinicius Layter Xavier**

**Adilson Elias Xavier**

**Daniela Cristina Lubke**

**XLVI SBPO**

Salvador, Bahia, 17 setembro 2014.

# Outline of Presentation

1 - Cluster Analysis

2 - Hyperbolic Smoothing Clustering Method (HSCM)

3 - The HSCM applied for Solving the Minimum Sum-of-Squares Clustering Problem

4 - Partition into Boundary and Gravitational Regions (PBGR)

5 - HSCM + PBGR Applied to General Clustering Problem

6 - HSCM + PBGR Applied to Minimum Sum-of-Squares Clustering Problem (MSSC)

7 – HSCM Applied to MSSC → Large Instances (Bagirov et al – December 2012)

8 - HSCM + PBGR Applied to MSSC → Very Large Instances (March 2013)

9 - HSCM + PBGR Applied to MSSC → Medium and Large Instances (April 2014)

10 - CONCLUSIONS

# Cluster Analysis

In the Cluster Analysis scope, algorithms use traditionally two main strategies:

- Hierarchical clustering methods

- Partition clustering methods...

- Here we introduce a New Approach for solving the problem...

# Hyperbolic Smoothing Clustering Method - HSCM

General Problem Formulation:

$S = \{S_1, \ldots, S_m\}$ Set of m observations

$x_i \; , \; i = 1, \ldots, q$ Centroides of the clusters

$z_j = \min_{i=1,\ldots,q} \| s_j - x_i \|$ Distance of the observation j to nearest centroide

Objective function specification

$$Minimize \quad \sum_{j=1}^{m} f(z_j)$$

where $f(z_j)$ is an arbitrary monotonic increasing function of the distance $z_j$.

# Hyperbolic Smoothing Clustering Method - HSCM

General Problem Formulation:

$$Minimize \quad \sum_{j=1}^{m} f(z_j)$$

$$z_j = \min_{i=1,\dots,q} \left\| s_j - x_i \right\|, \; j = 1,\dots,m$$

Trivial Examples of monotonic increasing functions:

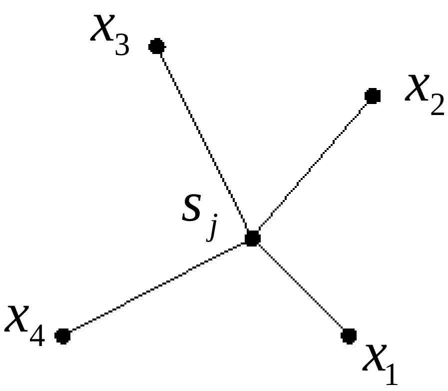
$$\sum_{j=1}^{m} f(z_j) = \sum_{j=1}^{m} Z_j^2 \qquad \sum_{j=1}^{m} f(z_j) = \sum_{j=1}^{m} Z_j$$

Possible Distance Metrics:

$L_2$ (Euclidean) $L_1$ (Manhattan) $L_\infty$ (Chebychev) $L_\mathrm{p}$ (Minkowshi)

# HSCM: Resolution Methodology

The General Clustering Problem



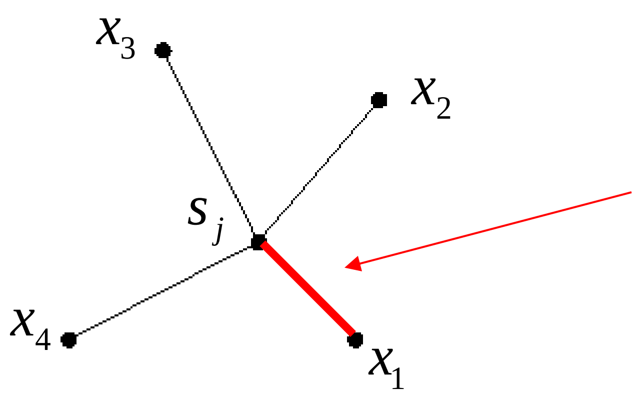

Consider a general observation point $s_j$ .

Let be $x_i, i = 1, \ldots, q$ , the centroids of clustersm where each set of these centroid coordinates will be represented by $X \in \Re^{nq}$.

# HSCM: Resolution Methodology

## The General Clustering Problem

Consider a general observation point $s_j$




$$z_j = \min_{x_i \in X} \left\| s_j - x_i \right\|$$

Let be $x_i, i = 1, \ldots, q$ , the centroids of clusters where each set of $s_j$ these centroid coordinates will be represented by $X \in \Re^{nq}$

This procedure is strongly non differentiable.

# HSCM: Resolution Methodology

## Problem transformation

The general clustering problem is equivalent to the following problem:

$$\text{(P1)}\ \textit{Minimize} \quad \sum_{j=1}^{m} f(z_j)$$

$$\textit{Subject to}: \quad z_j = \min_{i=1,\ldots,q} \left\| s_j - x_i \right\| \ , \quad j = 1,\ldots,m\,.$$

# HSCM: Resolution Methodology

## Problem transformation

From the problem (P1), it is obtained the relaxed problem:

$$(\mathrm{P2})\ \ \mathit{Minimize} \quad \sum_{j=1}^{m} f(z_j)$$

$$\mathit{Subject\ to}: \quad z_j - \left\| s_j - x_i \right\| \leq 0\,, \quad j = 1,\ldots,m, \ \ i = 1,\ldots,q.$$

# HSCM: Resolution Methodology

## Problem transformation

Let us use the auxiliary function

$$\psi(y) = \max\{0, y\}$$

From the inequalities

$$z_j - \|s_j - x_i\| \leq 0\,, \quad j = 1,\ldots,m, \; i = 1,\ldots,q,$$

it follows:

$$\sum_{i=1}^{q} \Psi\left(z_j - \|s_j - x_i\|\right) = 0, \; j = 1,\cdots,m$$

# HSCM: Resolution Methodology

## Problem transformation

Using this set of equality constraints in place of the constraints in (P2), we obtain the following problem

$$(P3) \quad Minimize \quad \sum_{j=1}^{m} f(z_j)$$

$$Subject\ to: \sum_{i=1}^{q} \Psi\left(z_j - \left\|s_j - x_i\right\|\right) = 0,\ j = 1, \cdots, m$$

This problem corresponds also to a relaxation!!

# HSCM: Resolution Methodology

## Problem transformation

By performing a $\varepsilon$ perturbation, it is obtained the problem:

$$(P4) \quad Minimize \quad \sum_{j=1}^{m} f(z_j)$$

$$Subject\ to: \sum_{i=1}^{q} \Psi\left(z_j - \left\| s_j - x_i \right\|\right) \geq \varepsilon,\ j = 1, \cdots, m$$

# HSCM: Resolution Methodology

## Smoothing the problem

Let us define the auxiliary function

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

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

# HSCM: Resolution Methodology

## Smoothing the problem

Function $\phi$ has the following properties:

(a) $\phi(y,\tau) > \Psi(y), \forall \tau > 0$

(b) $\lim_{\tau \to 0} \phi(y,\tau) = \Psi(y)$

(c) $\phi(y,\tau)$ is an increasing convex $C^{\infty}$ function.

# HSCM: Resolution Methodology

## Smoothing the problem

By using the Hyperbolic Smoothing approach for the problem (P4) it is obtained

$$\text{(P5)}\ \textit{Minimize} \quad \sum_{j=1}^{m} f(z_j)$$

$$\textit{Subject to}: \quad \sum_{i=1}^{q} \phi(z_j - \left\| s_j - x_i \right\| , \tau) \geq \varepsilon\,, \quad j = 1, \ldots, m.$$

# HSCM: Resolution Methodology

## Smoothing the distance calculation

For the Euclidian metric, let us define the auxiliary function

$$\theta(s_j, x_i, \gamma) = \sqrt{\sum_{l=1}^{n} (s_{jl} - x_{il})^2 + \gamma^2}$$

for $\gamma > 0$ .

# HSCM: Resolution Methodology

## Smoothing the Euclidian distance:

Function $\theta$ has the following properties:

(a) $\lim_{\gamma \to 0} \theta(s_j, x_i, \gamma) = \left\| s_j - x_i \right\|_2;$

(b) $\theta$ is a $C^{\infty}$ function.

# HSCM: Resolution Methodology

## Smoothing the problem

Therefore, it is now obtained the completely smooth problem:

$$(\mathrm{P6})\ \ \mathit{Minimize} \quad \sum_{j=1}^{m} f(z_j)$$

$$\mathit{Subject\ to:} \quad \sum_{i=1}^{q} \phi(z_j - \theta(s_j, x_i, \gamma), \tau) \geq \varepsilon\,, \quad j = 1, \ldots, m.$$

# HSCM: Resolution Methodology

## Problem resolution

By considering the KKT conditions for the problem (P6), it is possible to conclude that all inequalities will certainly be active.

$$\text{(P7)}\ \textit{Minimize} \quad \sum_{j=1}^{m} f(z_j)$$

$$\textit{Subject to}: \quad h(x, z_j) = \sum_{i=1}^{q} \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \varepsilon = 0\,, \quad j = 1, \ldots, m.$$

# HSCM: Resolution Methodology

## Problem resolution

Therefore, by using the Implicit Function Theorem, it is obtained the unconstrained problem

$$\text{(P8)}\ Minimize\ \ F(x) = \sum_{j=1}^{m} f(z_j)$$

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

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

# HSCM: Resolution Methodology

## Problem resolution

Again, due to the Implicit Function Theorem, the functions $z_j(x)$ have all derivatives in relation to the variables $x_i, \; i = 1,\ldots,q$ . So, it is possible to calculate the gradient of the objective function of the problem (P8)

$$\nabla F(x) = \sum_{j=1}^{m} \frac{df(z_j(x))}{dz_j} \nabla z_j(x)$$

where

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

# HSCM: Resolution Methodology

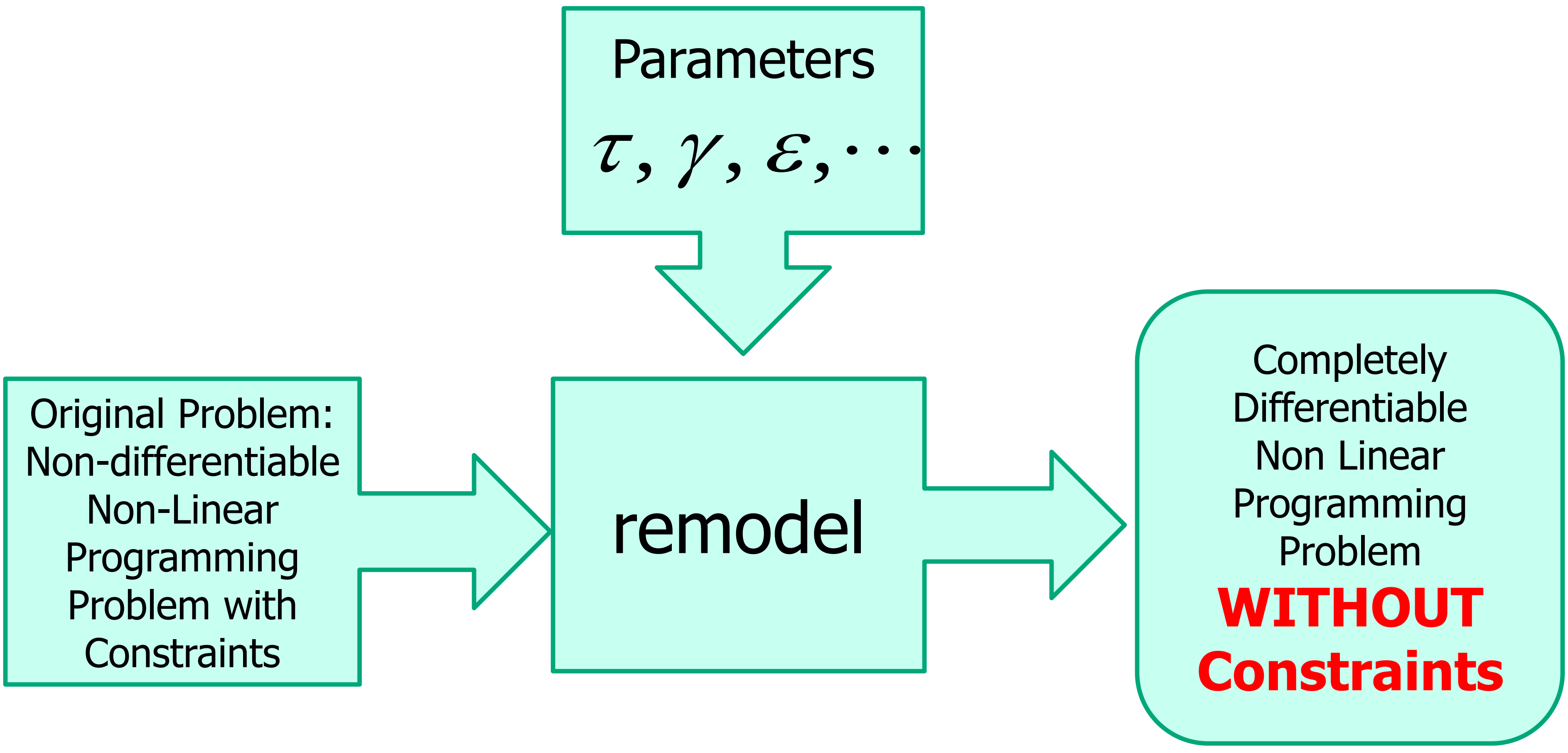
## Problem resolution

Some remarks about the problem (P8):

It can be solved by making use of any method based in the first and second order derivatives information (conjugate gradient, quasi-Newton, Newton methods, etc).

It is defined in a *nq*-dimensional space, therefore a smaller dimension problem in relation to the original problem (P7).

# Hyperbolic Smoothing Clustering Method: The Main Result



# Simplified HSC Algorithm

Initialization Step: Choose values $0 < \rho_1 < 1,\ 0 < \rho_2 < 1,\ 0 < \rho_3 < 1;$

Choose initial values: $x^0,\ \gamma^1,\ \tau^1,\ \varepsilon^1$; Let $k = 1.$

Main Step: Repeat indefinitely

Solve the unconstrained problem (P8)

$$\mathit{minimize}\ \ F(x) = \sum_{j=1}^{m} f(z_j(x))$$

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

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

Hyperbolic Smoothing Clustering Method

applied to:

# Minimum Sum-of-Squares Clustering Problem

# Hyperbolic Smoothing and Partition into Boundary and Gravitational Regions Applied to General Clustering Problem

# The Computational Task of the HSC Algorithm

The most relevant computational task associated to HSC Algorithm is the determination of the zeros of $m$ equations:

$$\sum_{i=1}^{q} \phi\left(z_j - \theta(s_j, x_i, \gamma), \tau\right) - \varepsilon = 0,\ j = 1, \cdots, m$$

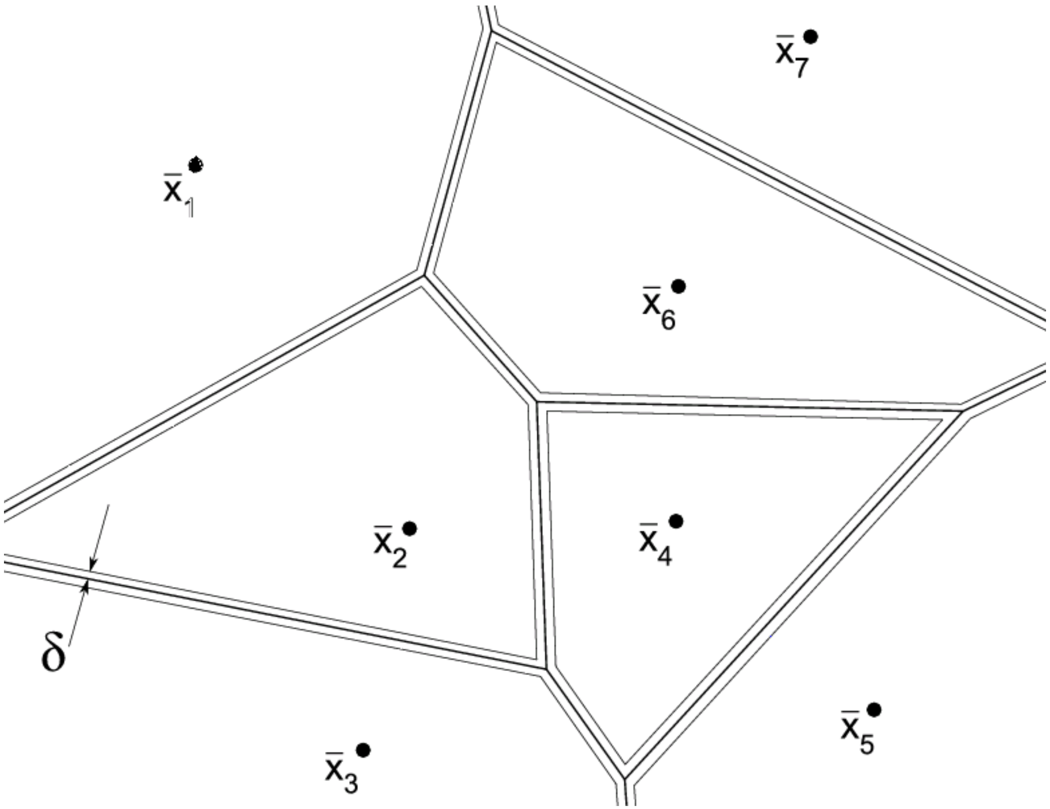
# The Partition into Boundary and Gravitational Regions

The basic idea of the approach is the partition of the set of observations in two non overlapping parts:

- the first set corresponds to the observation points that are relatively close to two or more centroids;

- the second set corresponds to the observation points that are significantly close to a unique centroid in comparison with the other ones.

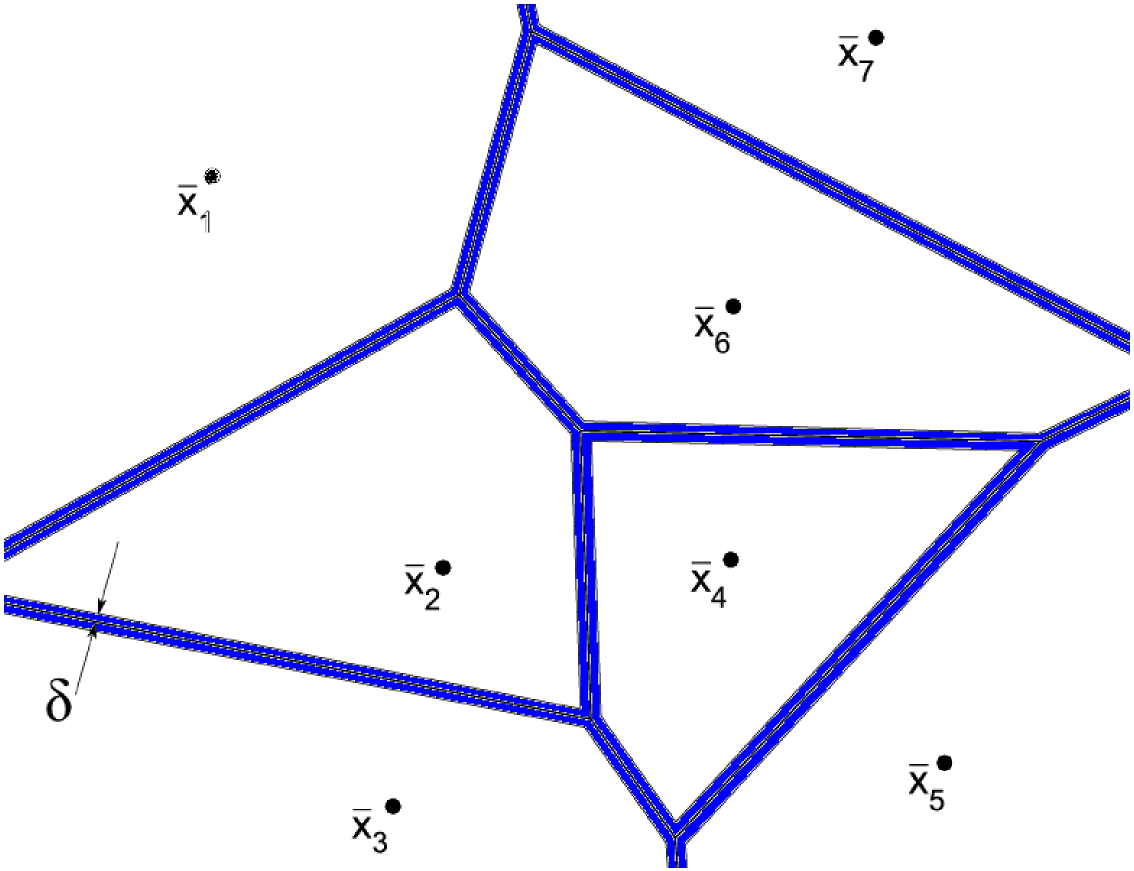
# Partition of the Set of Observations



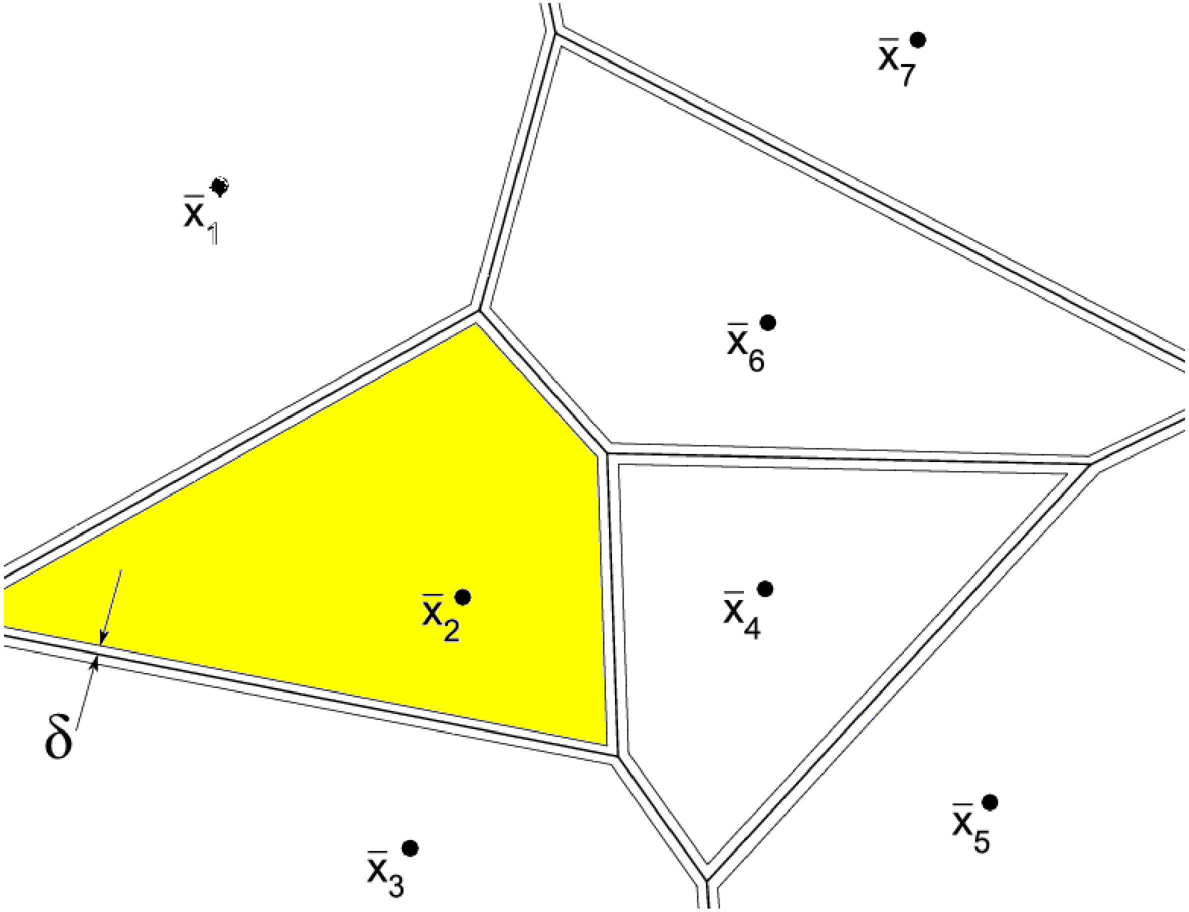

$\bar{x}$ : the referencial point

$\delta$: the band zone width

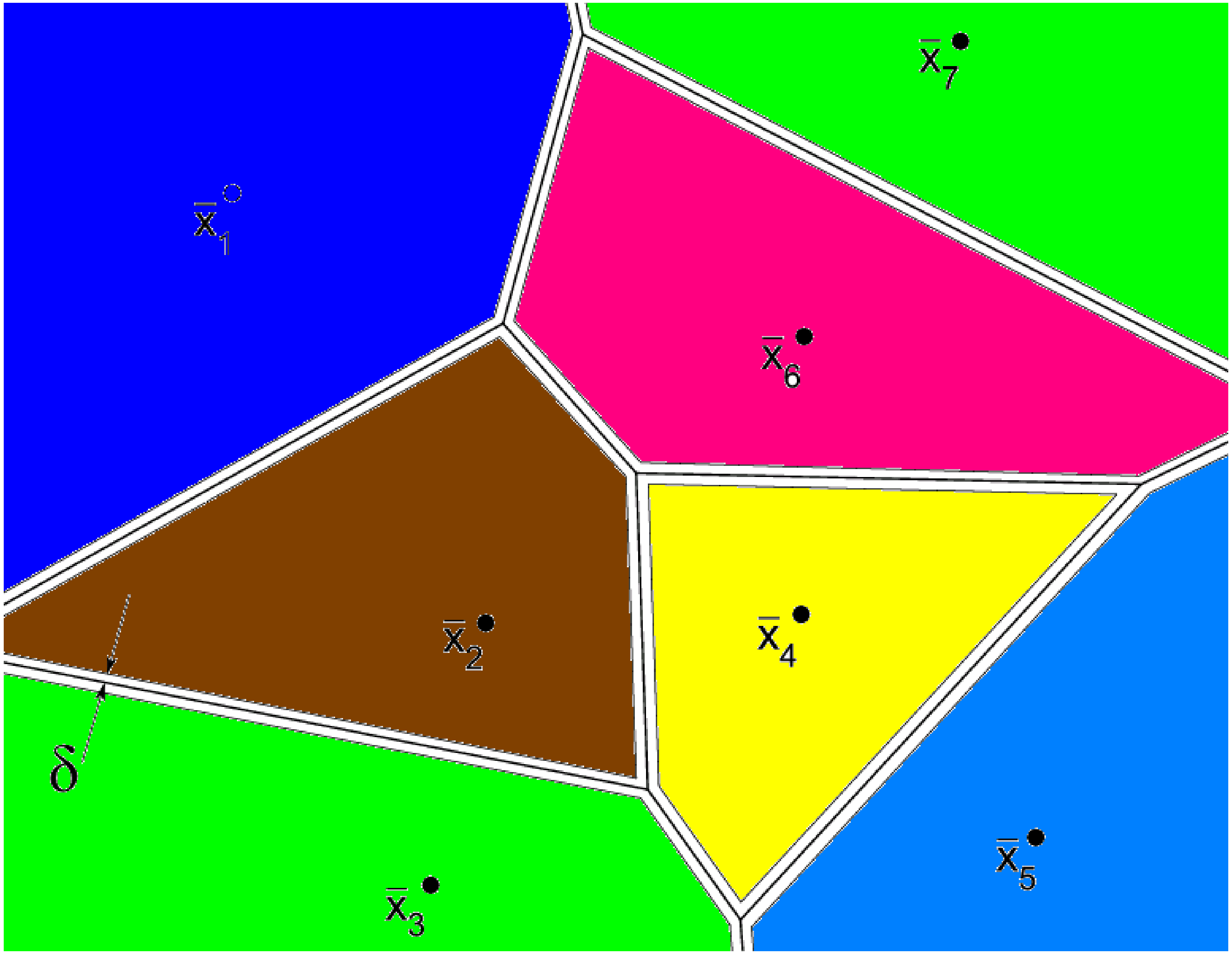
# The Boundary Band Zone



# Gravitational Region => $\bar{x}_2$



# Gravitational Regions



# The Partition of the Set of Observations

Let us define:

- $J_B$ is the set of boundary observations;
- $J_G$ is the set of gravitational observations.

Considering this partition, the objective function can be expressed in the following way:

$$Minimize\ \ F(x) = \sum_{j=1}^{m} f(z_j) = \sum_{j \in J_B} f(z_j) + \sum_{j \in J_G} f(z_j)$$

So, the objective function can be expressed in the following way:

$$Minimize\ F(x) = f_B(x) + f_G(x)$$

# The First Component of the Objective Function

The component associated with the set of boundary observations, can be calculated by using the previous presented smoothing approach:

$$Minimize \;\; F_B(x) = \sum_{j \in J_B} f(z_j)$$

where each $z_j$ results from the calculation of a zero of each equation:

$$h(x, z_j) = \sum_{i=1}^{q} \phi\Big(z_j - \theta(s_j, x_i, \gamma), \tau\Big) - \varepsilon = 0, \; j \in J_B$$

# The Second Component of the Objective Function

The component associated with the gravitational regions, can be calculated in a simpler form:

$$Minimize \;\; F_G(x) = \sum_{i=1}^{q} \sum_{j \in J_i} f(\|s_j - x_i\|)$$

where $J_i$ is the set of observations associated to cluster $i$.

# Hyperbolic Smoothing and Partition into Boundary and Gravitational Regions Applied to Minimum Sum-of-Squares Clustering Problem

# The Second Component for the MSSC

For The Minimum Sum-of-Squares Clustering formulation,
the second component, associated with the gravitational regions,
can be calculated in a direct form:

$$Minimize \;\; F_G(x) = \sum_{i=1}^{q} \sum_{j \in J_i} \left\| s_j - v_i \right\|_2^2 + \sum_{i=1}^{q} \left| J_i \right| \left\| x_i - v_i \right\|_2^2$$

where $J_i$ is the set of observations associated to centroid $i$

$v_i$ is the gravity center of cluster $i$.

# The Gradient of the Second Component for the MSSC

The gradient of the second component, associated with the gravitational regions, can be calculated in an easy way:

$$\nabla F_G(x) = \sum_{\iota=1}^{q} 2\left|J_i\right| \left(x_i - v_i\right)$$

where $J_i$ is the number the of observations associated to centroid $i$

# Simplified Accelerated Algorithm: AHSCM

Initialization Step: Choose values $0<\rho_1<1,\ 0<\rho_2<1,\ 0<\rho_3<1;$

Choose initial values: $x^0,\ \gamma^1,\ \tau^1,\ \varepsilon^1$; Let $k=1$.

Specify the initial boundary band width: $\delta^1$.

Main Step: Repeat indefinitely

By using $\bar{x}=x^{k-1}$ and $\delta=\delta^k$ determine partitions $J_B$ and $J_G$.

Solve the problem

$$minimize\ \ f(x)=\sum_{j\in J_B} z_j(x)^2+\sum_{j\in J_G} z_j(x)^2$$

with $\gamma=\gamma^k,\ \ \tau=\tau^k,\ \ \varepsilon=\varepsilon^k$, starting at the initial point $x^{k-1}$,

and let $x^k$ be the solution obtained.

Redefine the boundary band width value $\delta^{k+1}$.

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

# Comparison of Results: TSPLIB-3038

## Hyperbolic Smoothing Clustering & Accelerated HSC Methods

Results published in Pattern Recognion 44 (2011) 70-77

| q | $f_{opt}$ | Algorithm HSC | | Algorithm AHSC-L2 | | Speed up |
|---|---|---|---|---|---|---|
| | | Error | Time (s) | Error | Time (s) | |
| 2 | 0.31688E10 | 0.05 | 0.60 | 0.05 | 0.07 | **8.6** |
| 10 | 0.56025E09 | 0.01 | 16.16 | 0.01 | 0.28 | **57.7** |
| 20 | 0.26681E09 | **-0.02** | 62.90 | 0.05 | 0.59 | **107** |
| 30 | 0.17557E09 | **-0.08** | 169.17 | 0.31 | 0.86 | **197** |
| 40 | 0.12548E09 | **-0.06** | 333.92 | **-0.11** | 1.09 | **306** |
| 50 | 0.98400E08 | 0.50 | 662.33 | 0.44 | 1.36 | **487** |
| 60 | 0.82006E08 | **-1.09** | 1049.97 | **-0.80** | 1.91 | **550** |
| 80 | 0.61217E08 | **-0.31** | 2038.31 | **-0.73** | 6.72 | **303** |
| 100 | 0.48912E08 | **-0.90** | 3499.76 | **-0.60** | 9.79 | **357** |

# Comparison of Results: Pla85900

## Hyperbolic Smoothing Clustering & Accelerated HSC Methods

Results published in Pattern Recognion 44 (2011) 70-77

| q | $f_{opt}$ | Algorithm HSC | | Algorithm AHSC-L2 | | Speed up |
|---|---|---|---|---|---|---|
| | | Error | Time (s) | Error | Time (s) | |
| 2 | 0.37491E16 | 0.86 | 23.07 | 0.58 | 3.65 | **6.3** |
| 3 | 0.22806E16 | 0.00 | 47.41 | 0.04 | 4.92 | **9.6** |
| 4 | 0.15931E16 | 0.00 | 76.34 | 0.00 | 5.76 | **13.3** |
| 5 | 0.13397E16 | 0.80 | 124.32 | 1.35 | 7.78 | **16.0** |
| 6 | 0.11366E16 | 0.12 | 173.44 | 1.25 | 7.87 | **22.0** |
| 7 | 0.97110E15 | 0.42 | 254.37 | 0.87 | 9.33 | **27.3** |
| 8 | 0.83774E15 | 0.55 | 353.61 | 0.37 | 12.96 | **27.3** |
| 9 | 0.74660E15 | 0.68 | 438.71 | 0.25 | 13.00 | **33.8** |
| 10 | 0.68294E15 | 0.29 | 551.98 | 0.46 | 14.75 | **37.4** |

# Hyperbolic Smoothing Applied to Minimum Sum-of-Squares Clustering Problem

Bagirov et al - December 2012

Computational Optimization and Applications

# An incremental clustering algorithm based on hyperbolic smoothing

A.M. Bagirov[1], B. Ordin[2], G. Ozturk[3] and A.E. Xavier[4]

[1] *School of Science, Information Technology and Engineering, University of Ballarat, Victoria, Australia,*

[2] *Department of Mathematics, Faculty of Science, Ege University, Izmir, Turkey,*

[2] *Department of Industrial Engineering, Faculty of Engineering, Anadolu University, Eskisehir, Turkey,*

[4] *Department of Systems Engineering and Computer Science, Graduate School of Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.*

# Comparison Results

## General k-means, Modified General, Hyperbolic Smoothing

## D15112

| | | GKM | | | MGKM | | | SMOOTH | | |
|---|---|---|---|---|---|---|---|---|---|---|
| q | $f_{opt}$ | E | α | t | E | α | t | E | α | t |
| 2 | $0.36840 \cdot 10^{12}$ | 0.00 | 0.22 | 2.75 | 0.00 | 0.46 | 4.51 | 0.00 | 0.42 | 5.15 |
| 5 | $0.13271 \cdot 10^{12}$ | 0.00 | 0.92 | 8.50 | 0.00 | 1.83 | 13.85 | 0.00 | 1.52 | 18.41 |
| 10 | $0.64892 \cdot 10^{11}$ | 0.78 | 2.07 | 14.87 | 0.78 | 4.12 | 24.98 | 0.00 | 3.44 | 41.50 |
| 15 | $0.43136 \cdot 10^{11}$ | 0.26 | 3.24 | 21.00 | 0.26 | 6.44 | 35.51 | 0.00 | 5.74 | 68.23 |
| 20 | $0.32177 \cdot 10^{11}$ | 0.25 | 4.43 | 26.99 | 0.25 | 8.77 | 45.68 | 0.00 | 8.65 | 101.82 |
| 25 | $0.25423 \cdot 10^{11}$ | 0.03 | 5.62 | 32.71 | 0.03 | 11.11 | 55.47 | 0.00 | 11.96 | 139.53 |

# Comparison Results

## General k-means, Modified General k-means, Hyperbolic Smoothing

## Pla85900

| | | GKM | | | MGKM | | | SMOOTH | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **q** | $\mathbf{f_{opt}}$ | **E** | **α** | **t** | **E** | **α** | **t** | **E** | **α** | **t** |
| 2 | $0.37491 \cdot 10^{16}$ | 0.00 | 0.07 | 89.54 | 0.00 | 0.15 | 115.99 | 0.00 | 0.12 | 123.96 |
| 5 | $0.13397 \cdot 10^{16}$ | 0.00 | 0.30 | 408.60 | 0.00 | 0.59 | 512.31 | 0.00 | 0.37 | 452.96 |
| 10 | $0.68294 \cdot 10^{15}$ | 0.00 | 0.67 | 754.67 | 0.00 | 1.33 | 994.29 | 0.00 | 0.79 | 1011.17 |
| 15 | $0.46029 \cdot 10^{15}$ | 0.51 | 1.04 | 1083.04 | 0.51 | 2.07 | 1448.94 | 0.00 | 1.24 | 1596.67 |
| 20 | $0.35087 \cdot 10^{15}$ | 0.01 | 1.42 | 1355.85 | 0.00 | 2.82 | 1844.34 | 0.00 | 1.71 | 2210.91 |
| 25 | $0.28323 \cdot 10^{15}$ | 0.74 | 1.80 | 1570.37 | 0.73 | 3.57 | 2186.20 | 0.00 | 2.21 | 2863.18 |

# Very Large Clustering Instances

# Hyperbolic Smoothing and Partition into Boundary and Gravitational Regions Applied to Minimum Sum-of-Squares Clustering Problem

## Synthetic Problems – 03/2013

# Synthetic Problem 5.000.000 Observations Euclidian Space with n = 10 dimensions

| q | $f_{AHSC-L2}$ | Occur. | $E_{Mean}$ | $T_{Mean}$ |
|---|---|---|---|---|
| 2 | 0.456807E7 | 3 | 0.94 | 16.12 |
| 3 | 0.373567E7 | 1 | 1.21 | 24.69 |
| 4 | 0.323058E7 | 1 | 0.91 | 32.90 |
| 5 | 0.274135E7 | 1 | 0.09 | 26.06 |
| 6 | 0.248541E7 | 1 | 0.04 | 36.55 |
| 7 | 0.222897E7 | 1 | 0.19 | 43.24 |
| 8 | 0.197977E7 | 2 | 0.12 | 45.38 |
| 9 | 0.173581E7 | 2 | 0.10 | 42.78 |
| 10 | 0.149703E7 | 10 | 0.00 | 32.98 |
| **c** | 0.150000E7 | - | - | - |

## Medium and Large Clustering Instances

# Hyperbolic Smoothing and Partition into Boundary and Gravitational Regions Applied to Minimum Sum-of-Squares Clustering Problem

Comparison with Bagirov et al

January-April 2014

# Medium Size Instances
# Starting Point - Tight Way

**Breast Cancer (m=683 , n=9)**

| | | GKM | | MGKM | | HSCM | | AHSCM | |
|---|---|---|---|---|---|---|---|---|---|
| *k* | $F_{opt}$ | *E* | *t* | *E* | *t* | *E* | *t* | *E* | *t* |
| 2 | 1,93E08 | 0 | 0 | 0 | 0.02 | 0 | 0.05 | 0.00 | 0.04 |
| 5 | 1,37E08 | 2.28 | 0.03 | 1.86 | 0.05 | 0 | 0.61 | 0.00 | 0.14 |
| 10 | 1,02E08 | 0.26 | 0.06 | 0.28 | 0.11 | 0 | 2.14 | 0.16 | 0.37 |
| 15 | 8,69E07 | 1.02 | 0.08 | 1.07 | 0.16 | 0 | 9.08 | -0.36 | 0.62 |
| 20 | 7,65E07 | 3.64 | 0.11 | 1.80 | 0.20 | 0 | 27.36 | -0.19 | 1.12 |
| 25 | 6,90E07 | 5.08 | 0.14 | 0.93 | 0.27 | 0 | 44.65 | 0.11 | 1.53 |

# Medium Size Instances Starting Point - Tight Way

**TSPLIB3038 (m=3038 , n=2)**

| | | GKM | | MGKM | | HSCM | | AHSCM | |
|---|---|---|---|---|---|---|---|---|---|
| *k* | $F_{opt}$ | *E* | *t* | *E* | *t* | *E* | *t* | *E* | *t* |
| 2 | 3,17E13 | 0 | 0.06 | 0 | 0.11 | 0 | 0.25 | 0.06 | 0.11 |
| 5 | 1,20E13 | 0 | 0.25 | 0 | 0.39 | 0 | 0.98 | 0.00 | 0.21 |
| 10 | 5,60E12 | 2.78 | 0.48 | 0.58 | 0.81 | 0 | 2.64 | 0.56 | 0.69 |
| 15 | 3,56E12 | 0.07 | 0.70 | 1.06 | 1.20 | 0 | 5.16 | 0.03 | 1.40 |
| 20 | 2,67E12 | 2.00 | 0.94 | 0.48 | 1.61 | 0.11 | 8.86 | 0.22 | 2.39 |
| 25 | 2,15E12 | 0.78 | 1.20 | 0.23 | 1.98 | 0.01 | 14.01 | 0.18 | 3.84 |

# Medium Size Instances
# Starting Point - Tight Way

**Pendigit (m=10992 , n=16)**

| | | GKM | | MGKM | | HSCM | | AHSCM | |
|---|---|---|---|---|---|---|---|---|---|
| *k* | $F_{opt}$ | *E* | *t* | *E* | *t* | *E* | *t* | *E* | *t* |
| 2 | 1,28E12 | 0.39 | 2.56 | 0 | 5.23 | 0 | 5.37 | 0.00 | 0.38 |
| 5 | 7,53E11 | 0 | 9.73 | 0 | 18.72 | 0 | 22.60 | 0.00 | 1.76 |
| 10 | 4,93E11 | 0 | 20.45 | 0 | 39.31 | 0 | 61.71 | 0.00 | 4.99 |
| 15 | 3,91E11 | 0 | 30.79 | 0 | 59.19 | 0 | 120.46 | 0.00 | 10.33 |
| 20 | 3,41E11 | 0 | 41.25 | 0.17 | 78.37 | 0.16 | 250.80 | -0.31 | 19.21 |
| 25 | 3,00E11 | 0.24 | 51.87 | 0.24 | 98.47 | 0 | 446.37 | -0.12 | 33.06 |

# Large Size Instances
# Starting Point - Tight Way

**D15112** **(m=15112 , n=2)**

| | | GKM | | MGKM | | HSCM | | AHSCM | |
|---|---|---|---|---|---|---|---|---|---|
| ***k*** | $F_{opt}$ | ***E*** | ***t*** | ***E*** | ***t*** | ***E*** | ***t*** | ***E*** | ***t*** |
| **2** | **3,68E15** | **0** | **2.75** | **0** | **4.51** | **0** | **5.15** | **0.00** | **0.31** |
| **5** | **1,33E15** | **0** | **8.50** | **0** | **13.85** | **0** | **18.41** | **0.00** | **1.36** |
| **10** | **6,49E14** | **0.78** | **14.87** | **0.78** | **24.98** | **0** | **41.50** | **-0.62** | **4.21** |
| **15** | **4,31E14** | **0.26** | **21.00** | **0.26** | **35.51** | **0** | **68.23** | **0.15** | **9.31** |
| **20** | **3,22E14** | **0.25** | **26.99** | **0.25** | **45.68** | **0** | **101.82** | **0.62** | **17.43** |
| **25** | **2,54E14** | **0.03** | **32.71** | **0.03** | **55.47** | **0** | **139.53** | **-0.49** | **30.22** |

# Large Size Instances
# Starting Point - Tight Way

**Shuttle (m=58000 , n=9)**

| | | GKM | | MGKM | | HSCM | | AHSCM | |
|---|---|---|---|---|---|---|---|---|---|
| *k* | $F_{opt}$ | *E* | *t* | *E* | *t* | *E* | *t* | *E* | *t* |
| 2 | 2,13E13 | 0 | 63.20 | 0 | 123.82 | 0 | 86.74 | 0.00 | 0.20 |
| 5 | 7,24E12 | 0 | 251.65 | 0 | 521.81 | 0 | 351.50 | 0.00 | 1.35 |
| 10 | 2,83E12 | 0.02 | 581.28 | 0 | 1207.20 | 0 | 813.73 | 0.00 | 4.87 |
| 15 | 1,53E12 | 0 | 888.24 | 0 | 1776.88 | 0 | 1272.75 | 0.00 | 11.09 |
| 20 | 1,06E12 | 0 | 1195.84 | 0 | 2338.99 | 0 | 1798.88 | -3.60 | 26.80 |
| 25 | 7,98E11 | 0.17 | 1512.55 | 0.17 | 2917.41 | 0 | 2396.82 | -3.13 | 51.97 |

# Large Size Instances
# Starting Point - Tight Way

**Pla85900 (m=85900 , n=2)**

| | | GKM | | MGKM | | HSCM | | AHSCM | |
|---|---|---|---|---|---|---|---|---|---|
| *k* | $F_{opt}$ | *E* | *t* | *E* | *t* | *E* | *t* | *E* | *t* |
| 2 | 3,75E19 | 0 | 89.54 | 0 | 115.99 | 0 | 123.96 | 0.00 | 1.17 |
| 5 | 1,34E19 | 0 | 408.60 | 0 | 512.31 | 0 | 452.96 | 0.00 | 5.04 |
| 10 | 6,83E18 | 0 | 754.67 | 0 | 994.29 | 0 | 1011.17 | 0.00 | 15.66 |
| 15 | 4,60E18 | 0.51 | 1083.04 | 0.51 | 1448.94 | 0 | 1596.67 | 0.00 | 38.01 |
| 20 | 3,51E18 | 0.01 | 1355.85 | 0 | 1844.34 | 0 | 2210.91 | 0.00 | 73.09 |
| 25 | 2,83E18 | 0.74 | 1570.37 | 0.73 | 2186.20 | 0 | 2863.18 | 0.13 | 135.62 |

# Conclusions

# Conclusions

In view of the preliminary results obtained, where the proposed methodology performed efficiently and robustly, these algorithms can represent a possible approach for dealing with the solving of clustering of problems of real applications.

END