Flying Elephants: Método para Resolver Problemas Não-Diferenciáveis

Adilson Elias Xavier
Vinicius Layter Xavier
Federal University of Rio de Janeiro - BRAZIL

Seminários PESC

Rio de Janeiro, 27 Novembro 2015.
Outline of Presentation

1 - Introduction
2 - Fundamental Smoothing Procedures
3 - Geometry Distance Problem
4 - Covering Problems
5 - Clustering Problems
6 - Fermat-Weber Problem
7 - Hub location Problems
8 - Conclusions
Many do not believe, but the elephants really fly!!
Introduction

The core idea of the Flying Elephants method is the smoothing of a given non-differentiable problem.

In a sense, the process whereby this is achieved is a generalization and a new interpretation of a smoothing scheme, called Hyperbolic Smoothing (HS).
Introduction

The new name of the methodology, Flying Elephants, is definitively not associated to any analogy with the biology area. It is only a metaphor, but this name is fundamentally associated with properties of the method.
Introduction

The Flying feature is directly derived from the complete differentiability property of the method, which has the necessary power to make the flight of the heavy elephant feasible.

Moreover, it permits intergalactic trips into spaces with large number of dimensions, differently from the short local searches associated to traditional heuristic algorithms. On the other side, the convexification feature also associated to the FE method is analogous to the local action of the Elephant landing, eliminating a lot of local minima points.
Fundamental

Smoothing

Procedures
Fundamental Smoothing Procedures

There are only two Fundamental Smoothing Procedures

The main principle is always to perform transformations on the original formulation to make possible to use these two fundamental procedures in order to obtain a succedaneous problem completely differentiable.

This is the idea!!
Smoothing of the absolute value function

To smooth the absolute value function $|u|$, we use the function:

$$\theta(u, \gamma) = \left(u^2 + \gamma^2\right)^{1/2}$$
Fundamental Smoothing Procedures

Function $\theta$ has the following properties:

(a) $\lim_{\gamma \to 0} \theta(u, \gamma) = |u|$

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

(c) $\theta'(u, \gamma) = u / (u^2 + \gamma^2)^{1/2}$

(d) $\theta''(u, \gamma) = \gamma^2 / (u^2 + \gamma^2)^{3/2}$

(e) $\theta'''(0, \gamma) = 1 / \gamma$

(f) $\lim_{\gamma \to 0} \theta''(0, \gamma) \to \infty$
Fundamental Smoothing Procedures

Smoothing of the absolute value function

\[ \gamma_1 > \gamma_2 > \gamma_3 \]
Now, we will present the smoothing procedure of the function $\psi(y, \lambda) = \lambda \max(0, y)$. For this purpose, let us define the function

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

for $y \in \mathbb{R}$ and $\tau > 0$. 
Function $\phi$ has the following properties:

(a) $\phi(y, \lambda, \tau) > \psi(y, \lambda), \quad \forall \tau > 0$

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

(c) $\phi(y, \lambda, \tau)$ is a nondecreasing convex $C^\infty$ function in $y$

(d) $\phi'(y, \lambda, \tau) = \lambda + \lambda^2 y / (\lambda^2 y^2 + \tau^2)^{1/2}$

(e) $\phi''(y, \lambda, \tau) = \lambda^2 \tau^2 / (\lambda^2 y^2 + \tau^2)^{3/2}$

(f) $\phi''(0, \lambda, \tau) = \lambda^2 / \tau$

(g) $\lim_{\tau \to 0} \phi''(0, \lambda, \tau) \to \infty$
Fundamental Smoothing Procedures

Smoothing of the function $\psi$

$\tau_1 > \tau_2 > \tau_3$

$\phi(y, \tau)$

$\phi(y, \tau_1) \rightarrow$

$\phi(y, \tau_2) \rightarrow$

$\phi(y, \tau_3) \rightarrow$

$0 \rightarrow y$
Geometry

Distance

Problem
Publications


Geometry Distance Problem

Let $G = (V, E)$ denote a graph, in which for each $\text{arc}(i, j) \in E$, it is associated a measure $a_{ij} > 0$. The problem consists of associating a vector $x_i \in \mathbb{R}^n$ for each knot $i \in V$, basically addressed to represent the position of this knot into a $n$-dimensional space, so that Euclidean distances between knots, $\|x_i - x_j\|$, corresponds appropriately to the given measures $a_{ij}$.

$$\text{minimize} \quad f(x) = \sum_{(i, j) \in E} \left(\|x_i - x_j\| - a_{ij}\right)^2$$
Geometry Distance Problem

For solving the previous problem by using the Flying Elephant technique it is only necessary to use the function and \( \theta(u, \gamma) \) to define \( u = \|x_i - x_j\| \):

\[
\text{minimize } \quad f(x) = \sum_{(i,j) \in E} \left( \theta \left( \|x_i - x_j\|, \gamma \right) - a_{ij} \right)^2
\]
Geometry Distance Problem

Moré-Wu Instance
s=4
### Results of FE Technique Applied to Moré-Wu Instance

<table>
<thead>
<tr>
<th>$s$</th>
<th>$m = s^3$</th>
<th>$n = 3s^3$</th>
<th>$p$</th>
<th>Occur.</th>
<th>$f_{\text{Med}}$</th>
<th>Time</th>
<th>Occ.n.s</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>27</td>
<td>81</td>
<td>198</td>
<td>0</td>
<td>-</td>
<td>0.1</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>192</td>
<td>888</td>
<td>6</td>
<td>0.27E-6</td>
<td>0.7</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>125</td>
<td>375</td>
<td>2800</td>
<td>8</td>
<td>0.29E-5</td>
<td>2.8</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>216</td>
<td>648</td>
<td>7110</td>
<td>8</td>
<td>0.19E-4</td>
<td>7.6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>343</td>
<td>1029</td>
<td>15582</td>
<td>5</td>
<td>0.16E-4</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>512</td>
<td>1536</td>
<td>30688</td>
<td>8</td>
<td>0.29E-3</td>
<td>45</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>729</td>
<td>2187</td>
<td>55728</td>
<td>6</td>
<td>0.86E-3</td>
<td>97</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1000</td>
<td>3000</td>
<td>94950</td>
<td>7</td>
<td>0.95E-3</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1331</td>
<td>3993</td>
<td>153670</td>
<td>6</td>
<td>0.17E-2</td>
<td>81</td>
<td>0</td>
</tr>
</tbody>
</table>
Results of FE Technique applied to Moré-Wu Instance

<table>
<thead>
<tr>
<th>s</th>
<th>m = s^3</th>
<th>n = 3s^3</th>
<th>p</th>
<th>Occur.</th>
<th>f_{Med}</th>
<th>Time</th>
<th>Occ.n.s</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1728</td>
<td>5184</td>
<td>238392</td>
<td>8</td>
<td>0.15E-1</td>
<td>143</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2197</td>
<td>6591</td>
<td>356928</td>
<td>7</td>
<td>0.32E-1</td>
<td>222</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>2744</td>
<td>8232</td>
<td>518518</td>
<td>8</td>
<td>0.18E-1</td>
<td>380</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>3375</td>
<td>10125</td>
<td>733950</td>
<td>6</td>
<td>0.65E-1</td>
<td>543</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>4096</td>
<td>12288</td>
<td>1015680</td>
<td>7</td>
<td>0.42E-1</td>
<td>835</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>4913</td>
<td>14739</td>
<td>1377952</td>
<td>6</td>
<td>0.16E0</td>
<td>1270</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>5832</td>
<td>17496</td>
<td>1836918</td>
<td>7</td>
<td>0.21E0</td>
<td>1853</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>6859</td>
<td>20577</td>
<td>2410758</td>
<td>8</td>
<td>0.24E0</td>
<td>2335</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>8000</td>
<td>24000</td>
<td>3119800</td>
<td>8</td>
<td>0.59E0</td>
<td>3187</td>
<td>-</td>
</tr>
</tbody>
</table>
Covering

Problems
Covering Problem Conceptualization

Coverages of Brazil
Covering Problems

We consider the special case of covering a finite plane domain $S$ optimally by a given number $q$ of circles. We first discretize the domain $S$ into a finite set of $m$ points $s_j$, $j = 1, \ldots, m$. Let $x_i$, $i = 1, \ldots, q$ be the centres of the circles that must cover this set of points.

$$X^* = \arg \min_{X \in \mathbb{R}^{2q}} \max_{j=1,\cdots,m} \min_{i=1,\cdots,q} \left\| s_j - x_i \right\|_2$$
Flying Elephants Transformations

Original Problem: Non-differentiable Non-Linear Programming Problem with Constraints

remodel

Parameters \( \tau, \gamma, \varepsilon, \ldots \)

Completely Differentiable Non-Linear Programming Problem

WITHOUT Constraints
Covering Problems

By performing an \( \varepsilon \) perturbation and by using the FE approach, the three-level strongly nondifferentiable \( \min - \max - \min \) problem can be transformed in a one-level completely smooth one:

\[
\begin{align*}
\text{minimize} & \quad z \\
\text{subject to:} & \quad \sum_{i=1}^{q} \phi\left(z - \|s_j - x_i\|_2, \tau\right) \geq \varepsilon, \quad j = 1, \ldots, m
\end{align*}
\]
Covering Problems

Coverages of Netherlands
Covering Problems

Coverages of the state of New York
Covering Problems

Coverages of Dionisio Torres District – Fortaleza - Brazil

64 circles
Covering Problems

Coverages of Dionisio Torres District – Fortaleza - Brazil

64 circles
Clustering Problems
doi:10.1016/j.patcog.2009.06.018

doi:10.1016/j.patcog.2010.07.004


Let $S$ denote a set of $m$ patterns or observations from an Euclidean $n$-space, to be clustered into a given number $q$ of disjoint clusters. Let $x_i, i = 1, \ldots, q$ be the centroids of the clusters, where each $x_i \in \mathbb{R}^n$. Given a point $s_j$ of $S$, we initially calculate the Euclidean distance from $s_j$ to the nearest center. 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 minimum sum of the squares (MSSC) of these distances:
Clustering Problems

The minimum sum of the squares (MSSC) of these distances:

\[
\text{minimize } \sum_{j=1}^{m} z_j^2 \\
\text{subject to: } z_j = \min_{i=1,\ldots,q} \| s_j - x_i \|_2, \quad j = 1, \ldots, m
\]
Clustering Problems

By using FE approach, it is possible to use the Implicit Function Theorem to calculate each component $z_j, j = 1, \cdots, m$ as a function of the centroid variables $x_i, i = 1, \cdots, q$. In this way, the unconstrained problem

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

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

$$h(x, z_j) = \sum_{i=1}^{q} \phi(z_j - \theta(s_j, x_i, \gamma, \tau) - \varepsilon = 0, \quad j = 1, \cdots, m.$$
Clustering 5000000 Synthetic Observations with \( n = 10 \) Dimensions

<table>
<thead>
<tr>
<th>( q )</th>
<th>( f_{\text{AHSC-L2}} )</th>
<th>Occur.</th>
<th>( E_{\text{Mean}} )</th>
<th>( T_{\text{Mean}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.456807E7</td>
<td>3</td>
<td>0.94</td>
<td>16.12</td>
</tr>
<tr>
<td>3</td>
<td>0.373567E7</td>
<td>1</td>
<td>1.21</td>
<td>24.69</td>
</tr>
<tr>
<td>4</td>
<td>0.323058E7</td>
<td>1</td>
<td>0.91</td>
<td>32.90</td>
</tr>
<tr>
<td>5</td>
<td>0.274135E7</td>
<td>1</td>
<td>0.09</td>
<td>26.06</td>
</tr>
<tr>
<td>6</td>
<td>0.248541E7</td>
<td>1</td>
<td>0.04</td>
<td>36.55</td>
</tr>
<tr>
<td>7</td>
<td>0.222897E7</td>
<td>1</td>
<td>0.19</td>
<td>43.24</td>
</tr>
<tr>
<td>8</td>
<td>0.197977E7</td>
<td>2</td>
<td>0.12</td>
<td>45.38</td>
</tr>
<tr>
<td>9</td>
<td>0.173581E7</td>
<td>2</td>
<td>0.10</td>
<td>42.78</td>
</tr>
<tr>
<td>10</td>
<td>0.149703E7</td>
<td>10</td>
<td>0.00</td>
<td>32.98</td>
</tr>
<tr>
<td>( c )</td>
<td>0.150000E7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Fermat-Weber Problem

(Multisource Weber Problem)

(continuous p-center Problem)
Publications


Fermat-Weber Problem

Let \( S = \{s_1, \cdots, s_m\} \) denote a set of \( m \) cities or locations in an Euclidean planar space \( \mathbb{R}^2 \), with a corresponding set of demands \( W = \{w_1, \cdots, w_m\} \) to be attended by \( q \) a given number of facilities. To formulate the Fermat-Weber problem as a min - sum - min problem, we proceed as follows. Let \( x_i, i = 1, \cdots, q \) be the locations of facilities or centroids, \( x_i \in \mathbb{R}^2 \). Given a point \( s_j \in S \), we initially calculate the Euclidean distance from \( s_j \) to the nearest centroid: 

\[
z_j = \min_{i=1,\cdots,q} \|s_j - x_i\|_2.
\]
Fermat-Weber Problem

The Fermat-Weber problem considers the placing of facilities in order to minimize the total transportation cost:

\[
\begin{align*}
\text{minimize} & \quad \sum_{j=1}^{m} w_j z_j \\
\text{subject to} & \quad z_j = \min_{i=1,\ldots,q} \| s_j - x_i \|_2, \quad j = 1, \ldots, m
\end{align*}
\]
Fermat-Weber Problem

By using FE approach, it is possible to use the Implicit Function Theorem to calculate each component $z_j, j = 1, \ldots, m$ as a function of the centroid variables $x_i, i = 1, \ldots, q$. In this way, the unconstrained problem is obtained

$$\text{minimize} \quad f(x) = \sum_{j=1}^{m} w_j z_j(x)$$
Fermat-Weber Problem

Where each \( z_j(x) \) results from the calculation of the single zero of each equation below, since each term \( \phi \) above strictly increases together with variable \( z_j \):

\[
 h_j(z_j, x) = \sum_{j=1}^{m} \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \epsilon = 0, \quad j = 1, \ldots, m
\]
### Fermat-Weber Problem – Pla85900

<table>
<thead>
<tr>
<th>$q$</th>
<th>$f_{FE_{Best}}$</th>
<th>occur.</th>
<th>$E_{Mean}$</th>
<th>$Time_{Mean}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.163625E11</td>
<td>6</td>
<td>0.27</td>
<td>25.33</td>
</tr>
<tr>
<td>3</td>
<td>0.127835E11</td>
<td>10</td>
<td>0.00</td>
<td>50.91</td>
</tr>
<tr>
<td>4</td>
<td>0.108063E11</td>
<td>10</td>
<td>0.00</td>
<td>74.62</td>
</tr>
<tr>
<td>5</td>
<td>0.984539E10</td>
<td>7</td>
<td>0.11</td>
<td>121.02</td>
</tr>
<tr>
<td>6</td>
<td>0.902515E10</td>
<td>10</td>
<td>0.00</td>
<td>156.63</td>
</tr>
<tr>
<td>7</td>
<td>0.836416E10</td>
<td>3</td>
<td>0.18</td>
<td>206.71</td>
</tr>
<tr>
<td>8</td>
<td>0.778239E10</td>
<td>10</td>
<td>0.00</td>
<td>260.89</td>
</tr>
<tr>
<td>9</td>
<td>0.737264E10</td>
<td>9</td>
<td>0.09</td>
<td>317.09</td>
</tr>
<tr>
<td>10</td>
<td>0.704126E10</td>
<td>1</td>
<td>0.19</td>
<td>381.33</td>
</tr>
<tr>
<td>15</td>
<td>0.576935E10</td>
<td>10</td>
<td>0.00</td>
<td>937.84</td>
</tr>
<tr>
<td>20</td>
<td>0.502191E10</td>
<td>1</td>
<td>0.13</td>
<td>1690.06</td>
</tr>
<tr>
<td>30</td>
<td>0.411982E10</td>
<td>2</td>
<td>0.08</td>
<td>4062.92</td>
</tr>
<tr>
<td>40</td>
<td>0.358238E10</td>
<td>1</td>
<td>0.11</td>
<td>8169.64</td>
</tr>
</tbody>
</table>
Continuous Hub Location Problem

(multiple allocation p-Hub median problem)

Hub location Problems

The continuous p-hub median problem is a location problem which requires finding a set of \( p \) hubs in a planar region, in order to minimize a particular transportation cost function.

The assumption is that each pair of cities is directly connected by the shortest distance route between them.
Hub location Problems

The connections between each pair of cities $j$ and $l$, have always three parts:
1 - from the origin city $j$ to a first hub $a$;
2 - from hub $a$ to a second hub $b$;
3 - from hub $b$ to destination city $l$. 
Hub location Problems

Let $S = \{s_1, \ldots, s_m\}$ denote a set of $m$ cities or consumer points in a planar region. Let $w_{jl}$ be the demand between two points $j$ and $l$. Let $x_i, i = 1, \ldots, p$ be the hubs, where each $x_i \in \mathbb{R}^2$. 
Hub location Problems

The set of possible connections between city \( j \) and city \( l \).

Multiple allocation is permitted!
Hub location Problems

The p-hub median problem corresponds to minimizing the total cost between all pairs of cities taking the unitary cost value for all connections:

\[
\text{minimize} \quad \sum_{j=1}^{m} \sum_{l=1}^{m} w_{jl} z_{jl}
\]

subject to: \( z_{jl} = \min_{a,b=1,\ldots,p} z_{jabl}, \quad j = 1, \ldots, m \)

Where \( z_{jabl} = \| s_j - x_a \|_2 + \alpha \| x_a - x_b \|_2 + \| x_b - s_l \|_2 \) and \( \alpha \) is the reduction factor: \( 0 \leq \alpha \leq 1 \).
Hub location Problems

By using FE approach, it is possible to use once more the Implicit Function Theorem to calculate each component $z_{jl}$, $j, l = 1, \ldots, m$ as a function of the centroid variables $x_i$, $i = 1, \ldots, q$. So, we obtain the unconstrained problem

$$\text{minimize} \quad f(x) = \sum_{j=1}^{m} \sum_{l=1}^{m} w_{jl} z_{jl}(x)$$
Hub location Problems

German Towns: coordinates of 59 towns (Späth, 1980)
Hub location Problems
Alpha=0. => Fermat Weber problem
Hub location Problems

Alpha=0.25
Hub location Problems
Alpha=0.5
Hub location Problems

Hub Location Problem - dsj1000 TSPLIB instance $\alpha = 0.5$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$f_{FE_{Best}}$</th>
<th>occur.</th>
<th>$E_{Mean}$</th>
<th>$Time_{Mean}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.342083E12</td>
<td>10</td>
<td>0.00</td>
<td>376.66</td>
</tr>
<tr>
<td>3</td>
<td>0.285747E12</td>
<td>10</td>
<td>0.00</td>
<td>1296.32</td>
</tr>
<tr>
<td>4</td>
<td>0.263992E12</td>
<td>9</td>
<td>0.07</td>
<td>3754.33</td>
</tr>
<tr>
<td>5</td>
<td>0.248652E12</td>
<td>4</td>
<td>0.35</td>
<td>8234.88</td>
</tr>
</tbody>
</table>
Conclusions
Performance of the Flying Elephants Method

The performance of Flying Elephants Method can be attributed to the complete differentiability of the approach.

So, the succedaneum formulation can be comfortably solved by using the most powerful and efficient algorithms, such as conjugate gradient or quasi-Newton algorithms.

Computational experiments for all 5 related problems obtained unprecedented results, which exhibits a high level performance according to the different criteria of consistency, robustness and efficiency.
Additional Effect Produced by the Smoothing Procedures: Elimination of Local Minimum Points
Max-Cut Problem

(repto lançado pelo Prof. Manoel Campelo – UFC - Fortaleza)
Max-Cut Problem

The max-cut problem specification:

Temos um grafo $G=(V,E)$, onde $V$ é o conjunto de $n$ nós
$E$ é o conjunto de arcos

O problema é partitionar o conjunto de nós $V$ em duas partes $V_1$ e $V_2$, de maneira que a Soma dos Pesos dos Arcos entre $V_1$ e $V_2$ seja máxima.
Max-Cut Problem

A cada nó $i$ é associada uma variável $x_i$ que é igual a 1 se o nó pertence à partição V1 e igual a 0 em caso contrário.

$$\text{maximize } \sum_{(i,j)} c_{ij} \max(x_i - x_j, x_j - x_i)$$

$$\text{maximize } \sum_{(i,j)} c_{ij} \|x_i - x_j\|$$
Max-Cut Problem

Problema Suavizado

\[
\text{maximize } \sum_{(i,j)} c_{ij} \theta(x_i, x_j, \gamma) \\
1 > x_i > 0
\]
END
Convexification effect by the Hyperbolic Smoothing approach
Geometric distance problem
Smoothing of the objective function terms of a specific classification problem
Smoothing of the objective function terms of a specific classification problem
Smoothing of the objective function terms of a specific classification problem: Global Effect on the Objective Function
Quasi-Convexification effect by the Hyperbolic Smoothing approach associated to the Covering of a Region with Equal Circles Problem