

Flying elephants: a general method for solving non-differentiable problems

Adilson Elias Xavier · Vinicius Layter Xavier

Received: 22 November 2013 / Revised: 23 September 2014 / Accepted: 28 October 2014 /
Published online: 9 November 2014
© Springer Science+Business Media New York 2014

Abstract Flying Elephants (FE) is a generalization and a new interpretation of the Hyperbolic Smoothing approach. The article introduces the fundamental smoothing procedures. It contains a general overview of successful applications of the approach for solving a select set of five important problems, namely: distance geometry, covering, clustering, Fermat–Weber and hub location. For each problem the original non-smooth formulation and the succedaneous completely differentiable one are presented. Computational experiments for all related problems obtained results that exhibited a high level of performance according to all criteria: consistency, robustness and efficiency. For each problem some results to illustrate the performance of FE are also presented.

Keywords Non-differentiable optimization · Smoothing · Distance geometry · Covering · Clustering · Fermat–Weber problem · Hub location problem

1 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), presented in Santos (2003) for non-differentiable problems in general. This technique was developed through an adaptation of the hyperbolic penalty method

A. E. Xavier (✉) · V. L. Xavier
Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: adilson@cos.ufrj.br

V. L. Xavier
e-mail: vinicius@cos.ufrj.br

originally introduced by [Xavier \(1982\)](#) in order to solve the general non-linear programming problem.

By smoothing we fundamentally mean the substitution of an intrinsically non-differentiable two-level problem by a C^∞ differentiable single-level alternative. This is achieved through the solution of a sequence of differentiable sub-problems which gradually approaches the original problem. Each sub-problem, owing to its being indefinitely differentiable, can be comfortably solved by using the most powerful and efficient algorithms, such as conjugate gradient, quasi-Newton or Newton methods.

First, the FE method could incorporate any C^∞ smoothing scheme, for instance the hyperbolic smoothing approach. The HS approach has been applied for solving a set of hard mathematical problems. Despite these problems having a non-differentiable and a non-convex structure with a large number of local minimizers, the HS method produced efficiently and reliably very deep local minima, since it contains an intrinsic convexification power. So, HS can be applied for solving a broad number of non-differentiable optimization problems presented in the literature [Demyanov and Malozemov \(1974\)](#); [Du and Pardalos \(1995\)](#); [Rubinov \(2006\)](#), and [Demyanov et al. \(2014\)](#). The paper presents a survey of successful applications of the HS approach for solving a set of important problems, namely: distance geometry ([Macambira 2003](#); [Xavier 2003](#); [Souza 2010](#)) and ([Souza et al. 2011](#)), covering ([Xavier and Fernandes Oliveira 2005](#)), clustering ([Xavier 2010](#)), ([Xavier and Xavier 2011](#)) and ([Bagirov et al. 2012](#)), Fermat–Weber ([Xavier 2012](#)) and ([Xavier et al. 2014a](#)) and hub location ([Gesteira 2012](#)) and ([Xavier et al. 2014b](#)). There are other successful applications which are not presented in this survey, such as: $\min - \max$ ([Chaves 1997](#)) and ([Bagirov et al. 2013](#)) and packing problems.

The new name of the methodology, Flying Elephants, is definitely 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. The Flying feature is directly derived from the C^∞ 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 HS method is analogous to the local action of the Elephant landing, eliminating a lot of local minima points.

2 The fundamental smoothing procedures

The Flying Elephants method is based on the hyperbolic smoothing of the non-differentiable functions belonging to the optimization problem formulation. We will present the two basic smoothing procedures. First, we will consider the smoothing of the absolute value function $|u|$, where $u \in \Re$. For this purpose, for $\gamma > 0$, let us define the function

$$\theta(u, \gamma) = \sqrt{u^2 + \gamma^2}. \quad (1)$$

Function θ has the following properties:

- (a) $\lim_{\gamma \rightarrow 0} \theta(u, \gamma) = |u|$;

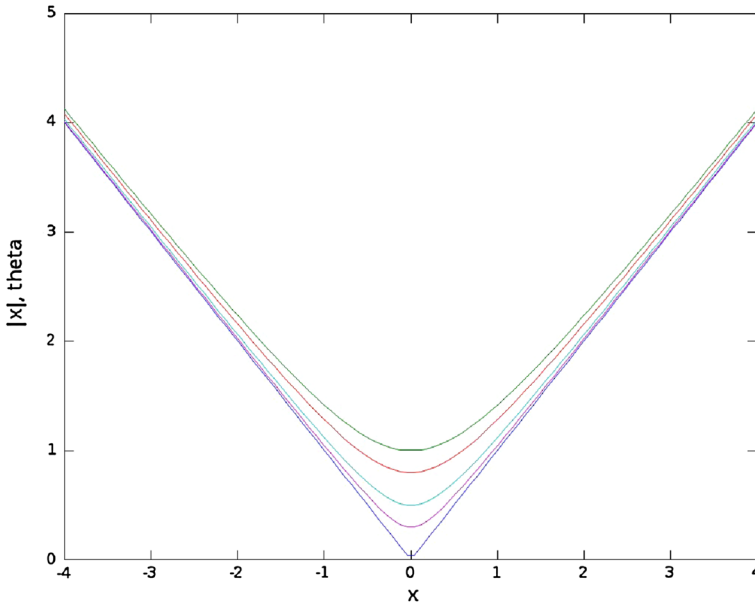


Fig. 1 $|u|$ and $\theta(u, \gamma)$, for $\gamma = [1, 0.8, 0.5, 0.3]$

- (b) θ is a C^∞ 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) $\lim_{\gamma \rightarrow 0} \theta''(0, \gamma) = \infty$.

Figure 1 depicts the gradual approximation of the smooth function $\theta(u, \gamma)$ to the absolute value function $|u|$, when parameter $\gamma \rightarrow 0_+$.

For smoothing the function $\psi(u) = \max(0, u)$ we use:

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

Function ϕ has the following properties:

- (a) $\phi(u, \tau) > \psi(u), \forall \tau > 0$;
- (b) $\lim_{\tau \rightarrow 0} \phi(u, \tau) = \psi(u)$;
- (c) $\phi(u, \tau)$ is an increasing convex C^∞ function in variable u ;
- (d) $\phi'(u, \tau) = 1 + u/(u^2 + \tau^2)^{1/2}$;
- (e) $\phi''(u, \tau) = \tau^2/(u^2 + \tau^2)^{3/2}$;
- (f) $\phi''(0, \tau) = 1/\tau$;
- (g) $\lim_{\tau \rightarrow 0} \phi''(0, \tau) = \infty$.

Figure 2 depicts the gradual approximation of the smooth function $\phi(u, \tau)$ to the value function $\psi(u)$, when parameter $\tau \rightarrow 0_+$. Then, function ϕ can be used as an approximation of function ψ .

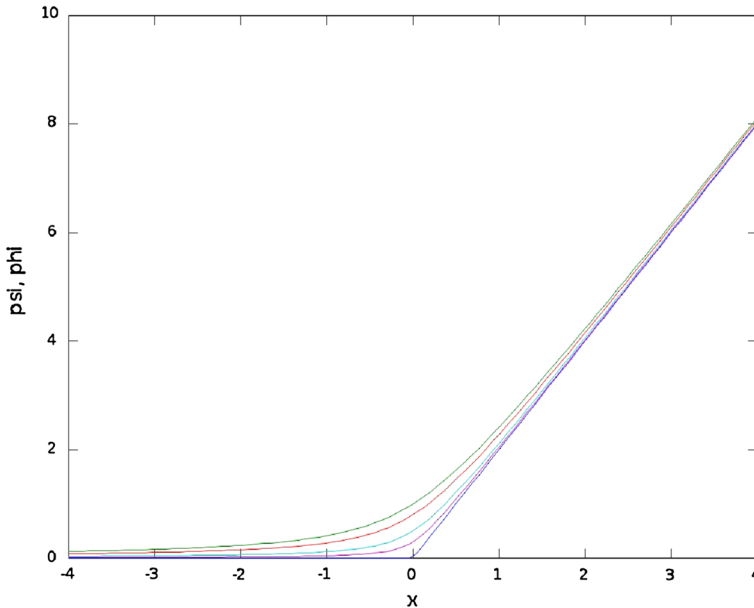


Fig. 2 $\psi(u)$ and $\phi(u, \tau)$, for $\tau = [1, 0.8, 0.5, 0.3]$

3 Distance geometry problem

Let $G = (V, E)$ denote a graph, in which for each arc $(i, j) \in E$, 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} :

$$\text{minimize } f(x) = \sum_{(i,j) \in E} (\|x_i - x_j\| - a_{ij})^2, \tag{3}$$

where x denotes the set of knot positions.

This problem is associated with important practical applications as in the protein folding field [Pardalos et al. \(1996\)](#). Its formulation presents the non-differentiable property due the presence of the Euclidean norm term. Moreover, the objective function is non-convex, so the problem has innumerable local minima. For solving the problem (3) by using the FE technique it is only necessary to use the function $\theta(u, \gamma)$ and to define $u = \|x_i - x_j\|$:

$$\text{minimize } f(x) = \sum_{(i,j) \in E} (\theta(\|x_i - x_j\|, \gamma) - a_{ij})^2. \tag{4}$$

Beside its smoothing properties the function θ also has the important convexification power. [Xavier \(2003\)](#) shows the following theoretical result:

Proposition 1 *There is a value $\bar{\gamma}$, such as, for all values $\gamma > \bar{\gamma}$, the Hessian matrix $\nabla^2 f(x)$ of the smoothed function will be positive definite.*

Souza (2010) and Souza et al. (2011) considers an alternative formulation, where the distances $\|x_i - x_j\|$ must be inside given intervals $[l_{i,j}, u_{i,j}]$:

$$\text{minimize } f(x) = \sum_{(i,j) \in E} \max [(l_{ij} - \|x_i - x_j\|), 0] + \sum_{(i,j) \in E} \max [(\|x_i - x_j\| - u_{ij}), 0] .$$

By using function $\phi(u, \tau) = (u + \sqrt{u^2 + \tau^2}) / 2$ in the place of function $\max(0, u)$, and by using function $\theta(u, \gamma)$ in the place of the Euclidean distance $u = \|x_i - x_j\|$, it is possible to obtain the smooth formulation:

$$\begin{aligned} \text{minimize } f_s(x) = & \sum_{(i,j) \in E} \phi (l_{ij} - \theta(\|x_i - x_j\|, \gamma), \tau) \\ & + \sum_{(i,j) \in E} \phi (\theta (\|x_i - x_j\| - u_{ij}, \gamma), \tau) . \end{aligned}$$

Souza (2010) and Souza et al. (2011) extends the previous theoretical result of Proposition 1 showing the convexification of the above problem for all values $\gamma > \max_{(i,j) \in E} u_{ij}$.

In order to illustrate the computational properties of the Flying Elephants method in this overview , we took the traditional lattice problem originally proposed by Moré and Wu (1995). This instance is a synthetic problem, where the knots are located on the intersection of s planes that cut a cube in the three principal directions in equal intervals.

Following the notation adopted by Moré and Wu , the knot positions are refereed by their coordinates indexes $\{(i_1, i_2, i_3), 0 \leq i_1 \leq s, 0 \leq i_2 \leq s, 0 \leq i_3 \leq s\}$. The relative position i of the knot x_i is given by the rule $i = 1 + x_1 + si_2 + s^2i_3$. The distances a_{ij} associated to the arcs (i, j) are exactly given by $a_{ij} = \|x_i - x_j\|_2$ for each arc $(i, j) \in S$, where $S = \{(i, j) \mid |i - j| < s^2\}$. The set of the $m = s^3$ knots is represented by $x = (x_1, \dots, x_m) \in \mathbb{R}^{3s^3}$. So, the problem has $n = 3s^3$ components and $p = s^5 + s^3 + s$ arcs.

The numerical experiments have been carried out on a Intel Core i7-2620M Windows Notebook with 2.70GHz and 8GB RAM. The columns of Table 1 show the number of splits of the cube (s), the number of variables of the problem ($m = 3s^3$), the occurrences of correct solutions obtained in 10 tentative solutions (*Occur.*), the average value of the correct solutions $f_{FE_{Aver}}$ and the mean CPU time (*Time*), given in seconds, associated to 10 tentative solutions.

This lattice instance considered is a synthetic problem which global solution f^* assumes a value equal zero, $f^* = 0$. So, the small values exhibited in the column $f_{FE_{Aver}}$ of Table 1 indicate the robustness of the FE approach, from cases $s = 4$ to $s = 20$, the last one with 3,119,800 arcs. Hoai and Tao (2000) proposed an approach for solving this problem based on difference of convex functions optimization algorithms and exhibits similar results, but only up to $s = 16$.

Table 1 Distance geometry problem—Moré-Wu Lattice Instance

s	$n = 3s^3$	$Occur.$	$f_{FE_{Aver}}$	T_{Mean}	s	$n = 3s^3$	$Occur.$	$f_{FE_{Aver}}$	T_{Mean}
3	81	0	–	0.1	12	5184	8	0.15E-1	143
4	192	6	0.27E-6	0.7	13	6591	7	0.32E-1	222
5	375	8	0.29E-5	2.8	14	8232	8	0.18E-1	380
6	648	8	0.19E-4	7.6	15	10125	6	0.65E-1	543
7	1029	5	0.16E-4	19	16	12288	7	0.42E-1	835
8	1536	8	0.29E-3	45	17	14739	6	0.16E0	1270
9	2187	6	0.86E-3	97	18	17496	7	0.21E0	1853
10	3000	7	0.95E-3	45	19	20577	8	0.24E0	2335
11	3993	6	0.17E-2	81	20	24000	8	0.59E0	3187

4 Covering problems

Let S be a finite region in \mathbb{R}^2 . A set of q figures constitutes a covering of order 1 of S if each point $s \in S$ belongs to at least one figure. Coverages of a higher order can be defined in a similar manner. Problems inherent to the covering of \mathbb{R}^2 regions by circles, of \mathbb{R}^3 regions by spheres, and even regions in higher dimensional spaces have been the object of research for many decades. 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 $s_j, j = 1, \dots, m$. Let $x_i, i = 1, \dots, q$ be the centres of the circles that must cover this set of points.

The optimal placing of the centres must provide the best-quality covering, that is, it must minimize the most critical covering. If $x \in \mathbb{R}^{2q}$ denotes the set of all placements $x_i, i = 1, \dots, q$ and x^* denotes an optimal placement, the covering problem assumes a *min – max – min* form:

$$x^* \in \operatorname{argmin}_{x \in \mathbb{R}^{2q}} \max_{j=1, \dots, m} \min_{i=1, \dots, q} \|s_j - x_i\|_2. \tag{5}$$

By defining z as the diameter of the circles, the above problem can be written on the equivalent form:

$$\begin{aligned} &\text{minimize } z \\ &\text{subject to } z_j = \min_{i=1, \dots, q} \|v_j - x_i\|_2, \quad j = 1, \dots, m, \\ &\quad z \geq z_j, \quad j = 1, \dots, m. \end{aligned} \tag{6}$$

By performing a relaxation of problem (6) we obtain the problem:

$$\begin{aligned} &\text{minimize } z \\ &\text{subject to } z_j - \|v_j - x_i\|_2 \leq 0, \quad j = 1, \dots, m, \quad i = 1, \dots, q, \\ &\quad z \geq z_j, \quad j = 1, \dots, m. \end{aligned} \tag{7}$$

By using the auxiliary function performing $\psi(u) = \max(0, u)$ and performing $\varepsilon > 0$ perturbation we obtain the problem:

$$\begin{aligned}
 &\text{minimize } z \\
 &\text{subject to } \sum_{i=1}^q \psi(z_j - \|v_j - x_i\|_2) \geq \varepsilon, \quad j = 1, \dots, m, \\
 &z \geq z_j, \quad j = 1, \dots, m.
 \end{aligned} \tag{8}$$

By considering the second set of constraints of problem (8), we can greatly reduce the dimension of the problem:

$$\begin{aligned}
 &\text{minimize } z \\
 &\text{subject to } \sum_{i=1}^q \psi(z - \|v_j - x_i\|_2) \geq \varepsilon, \quad j = 1, \dots, m.
 \end{aligned} \tag{9}$$

The FE approach uses the smooth function ϕ in the place function ψ . So, after the described set of steps, the original three-level strongly non-differentiable *min – max – min* problem can be transformed in a one-level completely smooth one:

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

Just as in other smoothing methods, the solution to the covering problem by the FE approach is obtained by resolving an infinite sequence of constrained minimization subproblems. So, the algorithm causes τ and ε approach 0, so the constraints of the subproblems it solves, given as in (10), tend to those of (9).

In order to show the computational properties of the Flying Elephants method, we reproduce some results originally presented in [Xavier and Fernandes Oliveira \(2005\)](#). Figure 3 depicts the solutions of three covering problems: Brazil (5 circles), The Netherlands (5 circles) and state of New York (7 circles). The number of discretization points were, respectively, 6620, 9220 and 7225. It is impossible to find more than a few works presenting computational results with similar quality to those obtained by FE approach. We quote [Wei \(2008\)](#).

5 Clustering problems

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 $x_i \in \mathbb{R}^n$. Given a point s_j of S , we initially calculate the Euclidean distance from s_j to the nearest center.

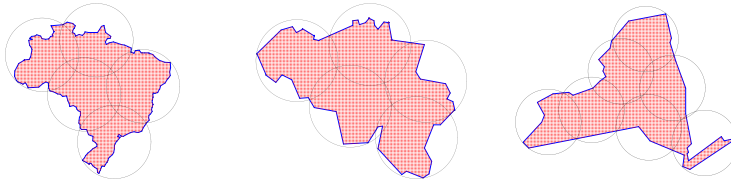


Fig. 3 Coverages of Brazil, Netherlands and the state of New York

This is given by $z_j = \min_{i=1, \dots, q} \|s_j - x_i\|_2$. 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:

$$\begin{aligned} & \text{minimize } \sum_{j=1}^m z_j^2 \\ & \text{subject to } z_j = \min_{i=1, \dots, q} \|s_j - x_i\|_2, \quad j = 1, \dots, m. \end{aligned} \tag{11}$$

By performing a relaxation of problem (11) followed by an $\epsilon > 0$ perturbation, in a similar way to the procedures applied to the covering problem presented in the last section, we obtain the problem:

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

Since $z_j \geq 0, j = 1, \dots, m$, the objective function minimization process will work for reducing these values to the utmost. On the other hand, given any set of centroids $x_i, i = 1, \dots, q$, due to property (c) of the hyperbolic smoothing function ϕ , the constraints of problem (10) are a monotonically increasing function in z_j . So, these constraints will certainly be active and problem (10) will finally be equivalent to problem:

$$\begin{aligned} & \text{minimize } \sum_{j=1}^m z_j^2 \\ & \text{subject to } h_j(z_j, x) = \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \epsilon = 0, \quad j = 1, \dots, m. \end{aligned} \tag{13}$$

The dimension of the variable domain space of problem (13) is $(nq + m)$. As, in general, the value of the parameter m , the cardinality of the set S of the observations s_j , is large, problem (13) has a large number of variables. However, it has a separable structure, because each variable z_j appears only in one equality constraint $h_j(z_j, x)$.

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

$$\text{minimize } f(x) = \sum_{j=1}^m z_j(x)^2 \tag{14}$$

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

$$h_j(z_j, x) = \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \varepsilon = 0, \quad j = 1, \dots, m. \tag{15}$$

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 (14),

$$\nabla f(x) = \sum_{j=1}^m 2z_j(x) \nabla z_j(x), \tag{16}$$

where

$$\nabla z_j(x) = -\nabla h_j(z_j, x) / \frac{\partial h_j(z_j, x)}{\partial z_j}, \tag{17}$$

while $\nabla h_j(z_j, x)$ and $\partial h_j(z_j, x) / \partial z_j$ are obtained from Eq. (15) and from definitions of function $\phi(y, \tau)$ and function $\theta(s_j, x_i, \gamma)$.

Xavier (2010) introduces the use of the FE approach for solving the MSSC problem. The computational results show a performance with robustness, efficiency and consistency, as well as with the capacity to solve large instances. Xavier and Xavier (2011) introduce a pruning scheme which speeds up the performance of the FE approach up to 500 times maintaining the same robustness. Bagirov et al. (2012) propose an algorithm which is based on the combination of the FE approach, without the pruning scheme, and an incremental approach to get a good starting point. This paper focuses on the solution of the largest clustering instances presented in the literature. A comparison with other two top algorithms, k-means like, demonstrates that the proposed algorithm by using the FE approach is more accurate.

Below we present a new computational experiment in order to demonstrate the performance of the FE method and, in particular, to show its capacity for solving very large clustering problems. We generate a synthetic data set with $m = 5000000$ observations in space with $n = 10$ dimensions. The observations were generated as random perturbations to 10 previously known centres. The numerical experiments have been carried out on a Intel Core i7-2620M Windows Notebook with 2.70 GHz and 8 GB RAM.

Table 2 presents a synthesis of the computational results. We vary the number of clusters $q = 2, \dots, 10$ and for each number of clusters, ten different randomly

Table 2 Clustering 5,000,000 Synthetic Observations with $n = 10$ Dimensions

q	$f_{FE_{Best}}$	$Occur.$	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	–	–	–

chosen starting points were used. The columns show the number of clusters (q), the best solution produced by the FE approach $f_{FE_{Best}}$, the number of occurrences of the best solution ($Occur.$), the average deviation of the 10 solutions in relation to the best solution obtained (E_{Mean}) and the CPU mean time given in seconds (T_{Mean}) associated to 10 tentative solutions. The last row of the table, represented by character c for center, informs the sum of the variances of the 10 synthetic groups, which is greater than the value obtained by the FE approach for the case $q = 10$. This result demonstrates unequivocally the robustness of the new method.

6 The Fermat–Weber problem

Let $S = \{s_1, \dots, s_m\}$ denote a set of m cities or locations in an Euclidean planar space \mathfrak{R}^2 , with a corresponding set of demands $W = \{w_1, \dots, w_m\}$, to be attended by a given number of q facilities. To formulate the Fermat–Weber problem as a *min – sum – min* problem, we proceed as follows. Let $x_i, i = 1, \dots, q$ be the locations of facilities or centroids, $x_i \in \mathfrak{R}^2$. Given a city $s_j \in S$, we initially calculate the Euclidean distance from s_j to the nearest facility or centroid: $z_j = \min_{i=1, \dots, q} \|s_j - x_i\|_2$. The Fermat–Weber problem considers the placing of q facilities in order to minimize the total transportation cost that is one of most important problems in the location sciences field:

$$\begin{aligned}
 & \text{minimize } \sum_{j=1}^m w_j z_j \\
 & \text{subject to } z_j = \min_{i=1, \dots, q} \|s_j - x_i\|_2, \quad j = 1, \dots, m.
 \end{aligned} \tag{18}$$

By performing the same procedures applied to the clustering problems shown in the previous section, we obtain a completely differentiable constrained problem:

$$\begin{aligned}
 & \text{minimize } \sum_{j=1}^m w_j z_j \\
 & \text{subject to } h_j(z_j, x) = \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \varepsilon = 0, \quad j = 1, \dots, m, \quad (19)
 \end{aligned}$$

where, as usual, x denotes the set of facilities or centroids $x_i, i = 1, \dots, q$.

Now, 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 $x_i, i = 1, \dots, q$. This way, the unconstrained problem

$$\text{minimize } f(x) = \sum_{j=1}^m w_j z_j(x) \quad (20)$$

is obtained, where each $z_j(x)$ results from the calculation of the single zero of each equation below, since each term ϕ above strictly increases together with variable z_j :

$$h_j(z_j, x) = \sum_{i=1}^q \phi(z_j - \theta(s_j, x_i, \gamma), \tau) - \varepsilon = 0, \quad j = 1, \dots, m. \quad (21)$$

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 (20):

$$\nabla f(x) = \sum_{j=1}^m w_j \nabla z_j(x), \quad (22)$$

where

$$\nabla z_j(x) = -\nabla h_j(z_j, x) / \frac{\partial h_j(z_j, x)}{\partial z_j}. \quad (23)$$

This way, it is easy to solve problem (20) by making use of any method based on first or second order derivative information. Finally, it must be emphasized that problem (20) is defined on a $(2q)$ -dimensional space, so it is a small problem, since the number of clusters, q , is, in general, very small in real applications.

Xavier (2012) introduces the use of the FE approach to solve the Fermat–Weber problem. The computational experiments with the new approach show a performance similar to the most efficient algorithms for solving problems up to 1,060 cities, the previous largest instance, see Brimberg et al. (2000) and Plastino et al. (2011). Xavier (2012) and Xavier et al. (2014a) present also results for problems never considered in the literature, with up to 85,900 cities, a new superior bound size about 80 times larger.

We reproduce in Table 3 the results of the experiment for the new largest instance Pla85900. Numerical experiments have been carried out on a PC Intel Pentium T4300, 2.1 GHz CPU with 4GB RAM, Windows 7, 32 bits.

Table 3 Fermat–Weber Problem—Pla85900 Instance

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

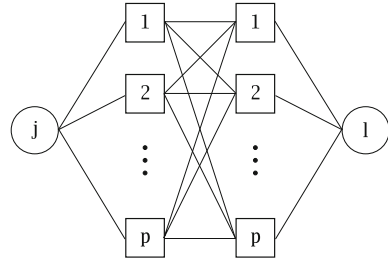
Tables 3 contains: the number of facilities q ; the best objective function value produced by the FE method $f_{FE_{Best}}$ by using a set of random starting points; the number of occurrences of the best solution $Occur.$; the percent deviation value of the set of 10 solutions related to the best value produced by the FE method E_{Mean} and the mean CPU time given in seconds $Time_{Mean}$. The small values in the column E_{Mean} show unequivocally the consistency of the FE algorithm. As there is no recorded result for this instance, the obtained values for the objective function and for the CPU time are a challenge for future research.

7 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. To formulate this problem, we proceed as follows. Let $S = \{s_1, \dots, s_m\}$ denote a set of m cities or consumer points in a planar region. Let w_{jl} be the demand between two cities j and l . Let $x_i, i = 1, \dots, p$ be the locations of the hubs, where each $x_i \in \mathfrak{R}^2$.

Concerning the hub-and-spoke problem under consideration, the connections between each pair of points j and l have always three parts: from the origin point j to a first hub a , from a to a second hub b and from b to destination point l . Multiple allocation is permitted, meaning that any given point can be served by one or more hubs. The first and the second hubs can be coincident (i.e., $a = b$), meaning that a unique hub is used to connect the origin point j and the destination point l . Figure 4 shows the p^2 possible connections between two cities. As shown, the hubs a and b can be the same.

Fig. 4 The set of connections between point j and point l



The unitary cost associated to a general connection (j, a, b, l) is equal to a weighted distance obtained by the sum of three Euclidian distances, with a reduction factor for the second part between hubs:

$$v_{jabl} = \|s_j - x_a\|_2 + \alpha \|x_a - x_b\|_2 + \|x_b - s_l\|_2, \tag{24}$$

where α is the reduction factor: $0 \leq \alpha \leq 1$.

The unitary cost from the origin city j to the destination city l is chosen as the minimum value for all connections:

$$z_{jl} = \min_{a,b=1,\dots,p} v_{jabl} = \min_{a,b=1,\dots,p} \{ \|s_j - x_a\|_2 + \alpha \|x_a - x_b\|_2 + \|x_b - s_l\|_2 \}. \tag{25}$$

For $\alpha = 0$, the connections with minimum values will be those with the closest assignments, which leads to the solutions of the Fermat–Weber problem.

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

$$\begin{aligned} &\text{minimize } \sum_{j=1}^m \sum_{l=1}^m w_{jl} z_{jl} \\ &\text{subject to } z_{jl} = \min_{a,b=1,\dots,p} v_{jabl}, \quad j, l = 1, \dots, m. \end{aligned} \tag{26}$$

By performing the same procedures applied to the clustering and Fermat–Weber problems shown in previous sections, we obtain a completely differentiable constrained problem:

$$\begin{aligned} &\text{minimize } \sum_{j=1}^m \sum_{l=1}^m w_{jl} z_{jl} \\ &\text{subject to } h_{jl}(z_{jl}, x) = \sum_{a=1}^p \sum_{b=1}^p \phi(z_{jl} - (\theta(s_j, x_a, \gamma) \\ &\quad + \alpha \theta(x_a, x_b, \gamma) + \theta(x_b, s_l, \gamma)), \tau) - \varepsilon = 0, \quad j, l = 1, \dots, m. \end{aligned} \tag{27}$$

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

$$\text{minimize } f(x) = \sum_{j=1}^m \sum_{l=1}^m w_{jl} z_{jl}(x) \quad (28)$$

is obtained, where each $z_{jl}(x)$ results from the calculation of a zero of each equation

$$h_{jl}(z_{jl}, x) = \sum_{a=1}^p \sum_{b=1}^p \phi(z_{jl} - (\theta(s_j, x_a, \gamma) + \alpha\theta(x_a, x_b, \gamma) + \theta(x_b, s_l, \gamma)), \tau) - \varepsilon = 0, \quad j, l = 1, \dots, m. \quad (29)$$

Again, due to the Implicit Function Theorem, the functions $z_{jl}(x)$ have all derivatives with respect to the variables $x_i, i = 1, \dots, p$, and therefore it is possible to calculate the gradient of the objective function (28):

$$\nabla f(x) = \sum_{j=1}^m \sum_{l=1}^m w_{jl} \nabla z_{jl}(x), \quad (30)$$

where

$$\nabla z_{jl}(x) = -\nabla h_{jl}(z_{jl}, x) \left/ \frac{\partial h_{jl}(z_{jl}, x)}{\partial z_{jl}} \right., \quad (31)$$

while $\nabla h_{jl}(z_{jl}, x)$ and $\partial h_{jl}(z_{jl}, x) / \partial z_{jl}$ are directly obtained from Eqs. (1), (2) and (29).

This way, it is easy to solve problem (28) by making use of any method based on first order derivative information. Finally, it must be emphasized that problem (28) is defined on a $(2p)$ -dimensional space, so it is a small problem, since the number of hubs, p , is small, in general, for real world applications.

Gesteira (2012) and Xavier et al. (2014b) introduce the use of the FE approach to solve the continuous hub location problem. They present computational results for instances up to 1,000 cities or about 500,000 different origin-destination pairs. The number of hubs reaches the value $p = 5$, which implies 12.5 million different path connections $jabl$, see Fig. 2. Contreras et al. (2011) consider a discrete hub location problem and solve problems up to 500 cities. To the best knowledge of these authors: *the new instances are by far the largest and most difficult ever solved for any type of hub location problem.*

Below, in Table 4, we reproduce the computational results obtained for solving the largest instance. The numerical experiments have been carried out on a PC Intel Celeron with a 2.7 GHz CPU and 512 MB RAM. The first column presents the specified number of hubs (p). The second column presents the best objective function value produced by the FE method in 10 attempts $f_{FE_{Best}}$. The next three columns present the number of occurrences of the best solution (*Occur.*), the percent average deviation of the T solutions in relation to the best solution obtained (E_{Mean}) and the CPU mean time given in seconds (T_{Mean}). The small values in the column E_{Mean} show unequivocally the consistency of the FE algorithm. As there is no recorded result for this instance the obtained values for objective function and CPU time are a challenge for future research.

Table 4 Hub location problem—dsj1000 TSPLIB instance ($\alpha = 0.5$)

P	$f_{FE_{Best}}$	$Occur.$	E_{Mean}	T_{Mean}
2	0.342083E12	10	0.00	376.66
3	0.285747E12	10	0.00	1296.32
4	0.263992E12	9	0.07	3754.33
5	0.248652E12	4	0.35	8234.88

8 Conclusions

In short, computational experiments for all related problems obtained results that exhibited a high level of performance of the FE approach according to the different criteria of consistency, robustness and efficiency. The robustness, consistency and efficiency performances can be attributed to the complete differentiability of the approach. Based on the success of these previous experiences, we believe that the FE methodology can also be used for solving a broad class of non-smooth problems with similar characteristics, like those contemplated in the seminal survey written by [Rubinov \(2006\)](#).

Acknowledgments The authors would like to thank the two referees for their comments.

References

- Bagirov, A.M., Al-Nuaimat, A., Sultanova, N.: Hyperbolic smoothing function method for minimax problems. *Optimization* **62**(6), 759–782 (2013)
- Bagirov, A.M., Ordin, B., Ozturk, G., Xavier, A.E.: An incremental clustering algorithm based on smoothing techniques. Submitted to *Computational Optimization and Applications* (2012)
- Brimberg, J., Hansen, P., Mladenovic, N., Taillard, E.D.: Improvements and comparison of heuristics for solving the multisource weber problem. *Oper. Res.* **48**, 129–135 (2000)
- Chaves, A.M.V.: Resolução do problema minimax via suavização. M.Sc. Thesis—COPPE-UFRJ, Rio de Janeiro (1997)
- Contreras, I., Cordeau, J.-F., Laporte, G.: Benders decomposition for large-scale uncapacitated hub location. *Oper. Res.* **59**, 1477–1490 (2011)
- Demyanov, V.F., Malozemov, V.N.: *Introduction to Minimax*. Dover Publications Inc., New York (1974)
- Demyanov, V.F., Pardalos, P.M., Batsyn, M.: *Constructive Nonsmooth Analysis and Related Topics*. Springer, New York (2014)
- Du, D.Z., Pardalos, P.M.: *Minimax and Applications*. Kluwer Academic Press, Dordrecht (1995)
- Gesteira, C.: Resolvendo Problemas de Localização de Hubs com Alocação Múltipla numa Modelagem Contínua Tipo p-Mediana Usando a Abordagem de Suavização Hiperbólica. M.Sc. Thesis—COPPE-UFRJ, Rio de Janeiro (2012)
- Hoai An, L.T., Tao, P.D.: Large-Scale Molecular Conformation Via the Exact Distance Geometry Problem. *Lecture Notes in Economics and Mathematical Systems*, vol. 481, pp. 260–277 (2000)
- Macambira, A.F.U.S.: Determinação de estruturas de proteínas via suavização e penalização hiperbólica. M.Sc. Thesis—COPPE—UFRJ, Rio de Janeiro (2003)
- Moré, J.J., Wu, Z.: ϵ -optimal solutions to distance geometry problems via global continuation. Mathematics and Computer Science Division, Argonne Lab., Preprint MCS-P520-0595 (1995)
- Pardalos, P.M., Shalloway, D., Xue, G.: *Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding DIMACS Series*, vol. 23. American Mathematical Society, Providence (1996)

- Plastino, A., Fuchshuber, R., Martins, S.L., Freitas, A.A., Salhi, S.: A hybrid data mining metaheuristic for the p -median problem. *Statist. Anal. Data Mining* **4**, 313–335 (2011)
- Rubinov, A.M.: Methods for global optimization of nonsmooth functions with applications. *Appl. Comput. Math.* **5**, 3–15 (2006)
- Santos, A.B.A.: Problemas de programação não-diferenciável: Uma metodologia de suavização. M.Sc. Thesis—COPPE—UFRJ, Rio de Janeiro (2003)
- Souza, M.F.: Suavização hiperbólica aplicada à otimização de geometria molecular. D.Sc. Thesis—COPPE—UFRJ, Rio de Janeiro (2010)
- Souza, M.F., Xavier, A.E., Lavor, C., Maculan, N.: Hyperbolic smoothing and penalty techniques applied to molecular structure determination. *Oper. Res. Lett.* **39**, 461–465 (2011)
- Wei, H.: Solving continuous space location problems. Ph.D. Thesis—The Ohio State University (2008)
- Xavier, A.E.: Penalização hiperbólica: Um novo método para resolução de problemas de otimização. M.Sc. Thesis—COPPE—UFRJ, Rio de Janeiro (1982)
- Xavier, A.E.: Convexificação do problema de distância geométrica através da técnica de suavização hiperbólica. Workshop em Biociências COPPE UFRJ (2003)
- Xavier, A.E.: The hyperbolic smoothing clustering method. *Pattern Recogn.* **43**, 731–737 (2010)
- Xavier, V.L.: Resolução do Problema de Agrupamento segundo o Critério de Minimização da Soma das Distâncias. M.Sc. Thesis—COPPE—UFRJ, Rio de Janeiro (2012)
- Xavier, A.E., Fernandes Oliveira, A.A.: Optimal covering of plane domains by circles via hyperbolic smoothing. *J. Global Optim.* **31**, 493–504 (2005)
- Xavier, V.L., França, F.M.G., Xavier, A.E., Lima, P.M.V.: A hyperbolic smoothing approach to the Multi-source Weber problem. *J. Global Optim.* **60**(1), 49–58 (2014a)
- Xavier, A.E., Gesteira, C., Xavier, V.L.: The continuous multiple allocation p -hub median problem solving by the hyperbolic smoothing approach: computational performance. *Optimization* (2014b). doi:[10.1080/02331934.2014.929677](https://doi.org/10.1080/02331934.2014.929677)
- Xavier, A.E., Xavier, V.L.: Solving the minimum sum-of-squares clustering problem by hyperbolic smoothing and partition into boundary and gravitational regions. *Pattern Recogn.* **44**, 70–77 (2011)