

A Procedure of Global Optimization and its Application to Estimate Parameters in Interaction Spatial Models

Edson Tadeu Bez¹, Mirian Buss Gonçalves², José Eduardo Souza de Cursi^{3,a}

¹ Universidade do Vale do Itajaí – UNIVALI, Rodovia SC 407, Km 04, CEP 88122-000, São José/SC, Brazil

² Universidade Federal de Santa Catarina – UFSC, Depto de Engenharia de Produção e Sistemas, 88010-970, Florianópolis/SC, Brazil

³ LMR – INSA – Rouen, Avenue de l'Université BP 8, Saint-Etienne du Rouvray, FR-76801, France

Received 15 November 2009, Accepted 30 March 2010

Abstract – This work concerns the calibration of trip distribution models issued from Transportation Planning. The optimal estimators are the global minima of a non-convex function. A mathematical result establishes that the solution is the limit of a sequence of means of convenient random variables. A population-based global optimization involving this representation and random perturbations of a local descent method is applied. Experiments are performed on classical test functions. In the application to the models of trip distribution, 4 sets of data, including real data are used. The results show that the use of the representation of the solution leads to a significant improvement and increases the robustness.

Key words: Global Optimization; Parameter Estimation; Calibration; Spatial Interaction Models

1 Introduction

Nowadays many studies focus on the development of robust mathematical methods that allow one to solve important problems in engineering. One carefully analyzes strategies for general solutions of the problems, instead of looking for particular solutions that would solve specific situations.

In Transport Planning, one of the problems that require the application of robust methods of optimization is the estimation of parameters of spatial interaction models. Beyond its utilization in transport planning, these models are also used in urban planning, in particular, in the planning of services as education and health. (Lowe[23], Soot[31], Almeida[1] and Almeida[2]).

Since its origin, in the middle of the past century, a good number of models of spatial interaction was developed and tested, mainly in the United States and the United Kingdom. Between them, the most popular and widely used is the gravitational model, conceived empirically from an analogy with the gravity law of Newton and which received, later, a more solid theoretical justification, based on the principle of the maximization of the entropy (Wilson[36] and Wilson[37]).

The gravitational models constitute the most common form of the models of trip distribution currently in use (Zhao[39]). They are aggregate models that are based on the hypothesis that the number of trips between two zones of traffic is directly proportional to the relative attractiveness of each zone and inversely proportional to some function of spatial separation between the zones.

Another important class of aggregate models is constituted by the models of opportunities, which are derived

from postulates associated to the choice of destination. In this class of models, the models of intervening opportunities (Schneider[27]) and the models of competitive destinations can be cited (Fotheringham[12]).

From the amalgamation of the gravitational models and intervening opportunities models, the hybrid models appeared, also called of gravity-opportunity models (Baxter[3], Wills[35], Gonçalves[16] and Gonçalves[17]).

To incorporate, in a disaggregated manner, the behavior of the users in spatial interaction models, it has appeared, more recently, the models of choice of destination (Ben-Akiva[5] and Ortuzar[24]). These models have been accepted as a possible alternative to the aggregated models, especially, to the gravitational models. However, comparing these models with the gravitational ones, Zhao[39] affirm: “Aside from the disaggregate-versus-aggregate distinction, the two models actually share more similarities than differences. In terms of functional forms, the *logit* model, into which destination choice models are frequently formulated, actually belong to the gravity model family. A production-constrained gravity model with an exponential impedance function is identical, up to a constant multiplier, to a *logit* destination choice model with a utility function. The only difference is that disaggregate *logit* models of destination choice include attributes of the individual or household, and aggregate gravity models include zonal demographic and socioeconomic measures.”

Another form to incorporate behavioral aspects in spatial interaction models was introduced by Almeida[2]. In their work, these authors present a methodology to incorporate behavioral aspects in the intervening opportunities models and/or hybrid models. For this, they use in the definition of intervening opportunities a function of utility, based on the Economic Theory of the Consumer and calibrated through

^a Corresponding author: eduardo.souza@insa-rouen.fr

techniques of state preference. An advantage of the presented methodology is the incorporation of the disaggregated behavioral modeling in the aggregate models of trip distribution, which are more treatable in diverse practical situations.

We remark, in this context, that the aggregate models continue to have an important role in the planning of urban services and transports, both in the developed world (Zhao[39] and Horowitz[21]), and in developing countries (Wirasinghe[38]). So, the evolution of the aggregate models and/or the methodologies of calibration, aiming at making the operational procedures simple for the practitioners, is of great importance.

Although the calibration of the gravitational models is simple, this does not occur with the other aggregate models cited above. Until today, the model of intervening opportunities is cited as a model with calibration problems (Zhao[39]). The same occurs with the model of competitive destinations (Fotheringham[12] and Diplock[10]) and with the hybrid model of Gonçalves and Ulysséa-Neto (Gonçalves[17], Diplock[10] and Gonçalves[19]). All these models need robust methods of optimization for its calibration.

In the literature, we can find a great number of optimization methods that are adequate to solve problems characterized by convex functions, but that do not present good performance for more complex problems, involving non convex functions, finding the optimal in narrow and deep cracks or in extensive and plain regions, or local minimums with similar values, etc.

Many studies are centered in the determination of new procedures, more robust, that make possible the determination of the global optimal, independent of the problem to be solved. These methods combine, in the majority of the cases, already existing methods, generating hybrid procedures, with the objective to advantage what each one has of better. One of the ways mentioned in the literature combines evolutionary algorithms with classic descent methods, aiming at joining advantages of the evolutionary methods, as efficiency and flexibility, to the precision and speed of search of traditional descent methods.

In this paper, we present a hybrid procedure of representations of solutions in global optimization. This procedure uses an evolutionary version presented by Gonçalves[19], of an algorithm based on random perturbations of the gradient method and a formula of representation of the global optimal given by Souza de Cursi[32] and Souza de Cursi[33], on the improvement of the initial population. Also we introduce an optimal step in the gradient method, conveniently defined, in each iteration. Tests with classical functions show that the improvements introduced in the original algorithm made it sufficiently robust, with potential to resolve applied problems (Bez[6] and Bez[8]).

We present an evaluation of the potential of the developed algorithm to solve two relevant problems. The first one consists in calibrating the model of competitive destinations of Fotheringham[12], comparing its performance with the one of an evolutionary strategy used by Diplock[10]. The choice of this strategy is due to the empirical evidence of its good performance (Schwefel[29]), compared with eleven conventional optimization methods in the solving of 28 test problems.

The second problem consists in calibrating the hybrid model of Gonçalves[17]. This problem was object of previ-

ous study of two of the authors (Gonçalves[19]), and they used, in the tests, a real data set, corresponding to the inter-city trips of bus passengers, in a region formed by 77 cities in the State of Santa Catarina, Brazil. Although in this work the calibration was successful, tests with other data bases showed a certain fragility of the algorithms presented in the initialization of its parameters. So, one looked for to develop a more robust algorithm, and to get a default set of parameters for its initialization, in order to make it simple of use for the professionals in the area of urban planning and transports.

The paper is organized as follows: In section 2, we present the models which will be studied, as well as the objective functions that must be minimized for the estimation of the model's parameters. In sections 3 and 4, we introduce the developed algorithm and we leave a proof of its convergence to the appendix. In section 5, we present some tests with classical functions. Next, in section 6, we present an application of the algorithm for the calibration of the model of competitive destinations of Fotheringham[12]. We also make a comparison between the results obtained with the use of the developed algorithm and with the use of an evolutionary strategy as done by Diplock[10]. In section 7, we show the results obtained in the calibration of the hybrid model of Gonçalves[17]. The goal of this application is the search of a set of default initialization parameters. Finally, in section 8, we present some conclusions.

2 Setting, calibrating and calibration criteria generalized gravitational model for spatial interactions

The flows observed between locations L_i and L_j are usually referred as *spatial interactions*. Their estimation is a crucial point in the evaluation or planning of transportation systems. An interesting – and popular – framework for this is furnished by the *gravity models*. These models are based on the empirical observation that spatial interactions decrease with the distance separating L_i and L_j . This observation has suggested to some authors that spatial interaction may be considered as being similar to gravitational one and that Newton's gravity model may be used in order to explain the empirical observations concerning flows between origins and destinations. In the pioneer works, a Newtonian-like law depending on the inverse of distance has been used and the expected flow T_{ij} going from L_i to L_j , separated by a distance d_{ij} has been estimated as $T_{ij} = M_i M_j / d_{ij}^\gamma$, where M_i and M_j are “masses” representing the importance of L_i and L_j , respectively, while γ is a parameter to be obtained from experimental evidence. In subsequent works, *generalized gravity models* have brought improvements to this basic model. For instance,

- the physical distance d_{ij} may be replaced by a more general quantity c_{ij} giving a measurement of the spatial separation of locations $i \in j$. The matrix $c = [c_{ij}]_{1 \leq i, j \leq n}$ is referred as *spatial separation matrix*.
- the standard Newtonian-like law may be replaced by a decreasing function $g(\bullet)$. Usually, g is defined by

using a set of *parameters* \mathbf{x} , which characterizes the attractiveness/generation of flows of the locations and the efficiency of the transport system.

- the “masses” M_i and M_j may be replaced by coefficients $M_{o,i}$ and $M_{d,j}$ representing the importance of L_i as origin and of L_j as destination of flows, respectively: a distinction is introduced in the importance as destination and the importance as origin of each location.

With these notations, the expected flows are given by

$$T_{ij} = M_{o,i} M_{d,j} g_{ij}, \quad g_{ij} = g(\mathbf{x}, c_{ij}).$$

A simple way to define the importance of each location as origin or destination of flows consists in using the global flows O_i originated at L_i and D_j destined to L_j :

$$D_j = T_{\cdot,j} = \sum_i T_{ij}, \quad O_i = T_{i\cdot} = \sum_j T_{ij}. \quad (1)$$

O_i and D_j may be considered as measurements of such an importance and the coefficients $M_{o,i}$ and $M_{d,j}$ may be assumed to be proportional to these values:

$$M_{o,i} = A_i O_i \quad \text{and} \quad M_{d,j} = B_j D_j,$$

what leads to the expected flows

$$T_{ij} = A_i O_i B_j D_j g_{ij}. \quad (2)$$

Eq. (1) implies that

$$A_i = \left(\sum_{j=1}^n B_j D_j g_{ij} \right)^{-1}, \quad B_j = \left(\sum_{i=1}^n A_i O_i g_{ij} \right)^{-1} \quad (3)$$

Eq. (3) shows that the coefficients $\mathbf{A} = (A_1, \dots, A_n)$ and $\mathbf{B} = (B_1, \dots, B_n)$ ensure that global coherence between inflows and outflows of the system: they are referred as *balancing factors*. For a given set of parameters \mathbf{x} and a given spatial separation matrix \mathbf{c} , the values of $[g_{ij}]_{1 \leq i, j \leq n}$ are known and Eqs. (1) – (3) form a nonlinear algebraical system for the unknowns \mathbf{A} , \mathbf{B} , $\mathbf{T} = [T_{ij}]_{1 \leq i, j \leq n}$, $\mathbf{O} = (O_1, \dots, O_n)$, $\mathbf{D} = (D_1, \dots, D_n)$. This feature is exploited in numerical calculations: for a given set of parameters \mathbf{x} , the solution of Eqs. (1) – (3) furnishes balancing factors $\mathbf{A}(\mathbf{x})$ and $\mathbf{B}(\mathbf{x})$, estimated fluxes $\mathbf{T}(\mathbf{x}) = [T_{ij}(\mathbf{x})]_{1 \leq i, j \leq n}$, $\mathbf{O}(\mathbf{x}) = (O_1(\mathbf{x}), \dots, O_n(\mathbf{x}))$ and $\mathbf{D}(\mathbf{x}) = (D_1(\mathbf{x}), \dots, D_n(\mathbf{x}))$.

In this work, we solve these equations by the classical iterative method of the equilibrium of the matrices, which is also called method of Furness. By limitation of the room, we do not give here the description of this classical method:

the reader interested in these aspects may refer, for instance, to Gonçalves[19].

One of the main criticisms on the gravity models concerns the fact that other factors than the distance may influence the attractiveness of a destination: for instance, the same facility may be available in several possible destinations, but their attractiveness may be not directly connected to the distance. In such a situation, the destinations may be considered as being in competition in order to generate inflows and the probability of choosing a specific destination depends also of the opportunities offered in the region under analysis – referred as *intervening opportunities*.

This observation has led some authors to consider improvements of the generalized gravity model by the adjunction of supplementary parameters describing other spatial effects. This is generally performed by, on the one hand, the introduction in the model of a supplementary matrix $\mathbf{w} = [w_{ij}]_{1 \leq i, j \leq n}$ involving a measurement the supplementary effects to be considered and, on the other hand, a modification of the function g in order to take into account the supplementary variable \mathbf{w} :

$$g_{ij} = g(c_{ij}, w_{ij}). \quad (4)$$

We consider in the sequel two classical models, which are used to illustrate the procedure proposed.

2.1 A model of competitive destinations

Sheppard[30] has considered that, for an individual from a given origin – for instance, L_i – the probability of a destination – for instance, L_j – depends on its relative accessibility when compared to the alternative destinations. For instance, this probability has not the same value if L_j is the only possible destination for a habitant of L_i or an individual from L_i may choose among several possible destinations, including L_j . These ideas have led to the model of *competitive destinations*, formulated by Fotheringham[12]. We consider here the version of this model proposed by Diplock[10], where \mathbf{w} furnishes a measurement of the accessibility of the potential destinations (Gitlesen[15]). This version reads as:

$$\mathbf{B} = (1, \dots, 1), \quad g(c_{ij}, w_{ij}) = w_{ij}^{-\delta} e^{-\beta c_{ij}}, \quad (5)$$

where w_{ij} is a measurement of the perception by an individual from L_i of the relative accessibility of destination L_j when compared to the other possible destinations. Fotheringham[12], Fotheringham[13] and Fotheringham[14] has defined

$$w_{ij} = \sum_{k \neq i, k \neq j} D_k e^{\sigma c_{jk}} \quad (6)$$

For this model, $\mathbf{x} = (\delta, \beta, \sigma)$ which are impedance parameters connected to the accessibility, the spatial

separation and the importance of the distance in the perception of the accessibility, respectively.

2.2 A model involving intervening opportunities

We consider also a second model, introduced by (Gonçalves[17]). This model takes into account intervening opportunities in the expected flows by using

$$g_{ij} = g(c_{ij}, w_{ij}) = e^{-(\beta c_{ij} + \lambda w_{ij})}. \quad (7)$$

Here, $w = [w_{ij}]_{1 \leq i, j \leq n}$ is a measurement of the intervening opportunities for a flow from location L_i to L_j . In this case, $x = (\beta, \lambda)$, which are impedance parameters for the spatial separation and the intervening opportunities, respectively.

2.3 Calibration of the models

An important step for the practical use of spatial interaction models is their calibration, i. e., the determination of the parameters x of the model and the associated balancing factors $A(x)$ and $B(x)$. These parameters are determined by comparing the estimations $T(x)$ of the model to empirical data T^* : the parameters must be set to values that furnish estimations which are as close as possible to the observed flows. In practice, the empirical data consists in

- The matrix of the measured flows $T^* = [T_{ij}^*]_{1 \leq i, j \leq n}$;
- The spatial separation matrix $c = [c_{ij}]_{1 \leq i, j \leq n}$;
- The supplementary matrix $w = [w_{ij}]_{1 \leq i, j \leq n}$.

The empirical vectors giving the generated and received flows at each location are denoted by $O^* = (O_1^*, O_2^*, \dots, O_n^*)$ and $D^* = (D_1^*, D_2^*, \dots, D_n^*)$, respectively, and are obtained from these data by using the relations (1):

$$D_j^* = T_{\cdot j}^* = \sum_i T_{ij}^*, \quad O_i^* = T_{i \cdot}^* = \sum_j T_{ij}^*.$$

As previously observed, the goal is to determine a set of parameters and balancing factors such that the estimations furnished by the model are, on the one hand, consistent with the balance of inflow and outflow at each traffic zone and, on the other hand, as close as possible of the empirical data available.

Thus, the calibration involves,

- on the one hand, the choice of a *calibration criterion*, i. e., a measurement of the distance between empirical and estimated flows;
- on the other hand, a procedure for the determination of the parameters and balancing factors. This procedure is usually computational and involves optimization methods for the minimization of the calibration criterion.

Thus, the calibration consists in determining

$$x^* = \underset{S}{\text{Arg Min}} F \quad (8)$$

where $S \subset V = R^n$ is the admissible region for the parameters and F is the calibration criterium. The optimal value of F is $F^* = F(x^*)$.

Different choices of F may be considered for a given model, but in general, the calibration criterion is non convex and the determination of x^* is a difficult problem of global optimization. This work is mainly concerned by the resolution of this last difficulty and we shall illustrate the procedure proposed by using two choices of F .

2.4 Calibration criterion for the model of competitive destinations

For this model, we use the mean quadratic error introduced by Diplock[10]:

$$F(x) = \frac{1}{n^2} \sum_{i, j=1}^n (T_{ij}^* - T_{ij}(x))^2. \quad (9)$$

Although the parameters may take any non-negative value, it is more useful to look for the parameters in a box-bounded region:

$$S = \{x = (\delta, \beta, \sigma) \in V : 0 \leq \delta \leq \delta_{\max}, 0 \leq \beta \leq \beta_{\max}, 0 \leq \sigma \leq \sigma_{\max}\}. \quad (10)$$

More general search regions may also be considered, such as, for instance, admissible regions defined by linear or nonlinear constraints (see, for instance, Bouhadi[9] or Souza de Cursi[34]).

2.5 Calibration criterion for the model involving intervening opportunities

By assuming that the flows T_{ij} are random variables multinomially distributed, the maximum likelihood estimations of x , A and B verify, in addition to Eqs. (1) – (4),

$$\frac{1}{T^*} \sum_{i, j=1}^n c_{ij} T_{ij}^* = \frac{1}{T(x)} \sum_{i, j=1}^n c_{ij} T_{ij}(x) \quad (11)$$

$$\frac{1}{T^*} \sum_{i, j=1}^n w_{ij} T_{ij}^* = \frac{1}{T(x)} \sum_{i, j=1}^n w_{ij} T_{ij}(x) \quad (12)$$

where T^* and $T(x)$ are the global flows:

$$T^* = \sum_{i, j=1}^n T_{ij}^* ; \quad T(x) = \sum_{i, j=1}^n T_{ij}(x) \quad (13)$$

In practice, these equations are not exactly satisfied and it is more convenient to look for approximate solutions, such as for instance, least squares ones by

$$E(x, a) = \frac{1}{T^*} \sum_{i, j=1}^n a_{ij} T_{ij}^* - \frac{1}{T(x)} \sum_{i, j=1}^n a_{ij} T_{ij}(x), \quad (14)$$

an approximate solution is determined by minimizing $F : V \rightarrow R$ given by

$$F(\mathbf{x}) = (E(\mathbf{x}, \mathbf{c}))^2 + (E(\mathbf{x}, \mathbf{w}))^2. \quad (15)$$

Here, we look for the parameters in the box-bounded region:

$$S = \{ \mathbf{x} = (\beta, \lambda) \in V : 0 \leq \beta \leq \beta_{\max}, 0 \leq \lambda \leq \delta_{\max} \}. \quad (16)$$

As previously remarked, more general search regions may be considered.

3 A population-based algorithm involving stochastic perturbations

When solving the general problem stated in Eq. (8), one of the main difficulties is connected to the non-convexity of the objective function F .

In order to fix the ideas, let us consider an iterative method generating a sequence $\{\mathbf{x}_k\}_{k \in \mathbb{R}^+}$, recursively defined from a starting point:

$$\mathbf{x}_0 \in S ; \mathbf{x}_{k+1} = \mathbf{Q}_k(\mathbf{x}_k), \quad (17)$$

where \mathbf{Q}_k is an iteration function, which may involve ns substeps of a standard descent method (for instance, steepest descent or Fletcher-Reeves):

$$\mathbf{Q}_k(\mathbf{x}) = \mathbf{Q}^{ns}(\mathbf{x}) = \underbrace{\mathbf{Q}(\mathbf{Q}(\dots \mathbf{Q}(\mathbf{x})))}_{ns \text{ times}}. \quad (18)$$

Here, \mathbf{Q} is the iteration function associated to the descent method used for the substeps. It usually involves a descent direction \mathbf{d} and a step $\alpha \geq 0$ to be determined. For instance, the classical gradient descent with a fixed step uses $\mathbf{d} = -\nabla F(\mathbf{x})$ and a fixed parameter $\alpha_0 > 0$ in order to define $\mathbf{Q}(\mathbf{x}) = \mathbf{x} - \alpha_0 \nabla F(\mathbf{x})$. In a more sophisticated approach, the step may be determined by one dimensional search involving a previously established maximal step α_{\max} (for instance, we may consider the optimal step on the interval $[\theta, \alpha_{\max}]$).

Due to the lack of convexity, the convergence of these iterations to a global minimum \mathbf{x} satisfying (10) is not ensured and may strongly depend on the parameters of the method, such as the initial point \mathbf{x}_0 , the descent method used in the substeps and their number ns , the maximum step α_{\max} and the method used for the unidimensional search.

This difficulty makes that global optimization methods, such as stochastic or evolutionary algorithms are considered for the numerical solution of (8): for instance, we may use the random perturbation approach, where the iterations (17) are modified as

$$\mathbf{x}_0 \in S ; \mathbf{x}_{k+1} = \mathbf{Q}_k(\mathbf{x}_k) + \mathbf{P}_k, \quad (19)$$

where \mathbf{P}_k is a convenient random variable which ensures the convergence to a global minimum, such as, for instance,

$$\mathbf{P}_k = \frac{\omega}{\sqrt{\log(k+1)}} \mathbf{Z}, \text{ where } \mathbf{Z} \text{ is a gaussian random vector}$$

following $N(0, \mathbf{Id})$ and $\omega > 0$ is fixed (see, for instance, Pogu[26] or Ellaia[11]). In practice, the implementation of

random perturbations involve a finite sample $(\mathbf{P}_k^1, \dots, \mathbf{P}_k^{nr})$ of nr variates from \mathbf{P}_k : let $\mathbf{P}_k^0 = 0$, $\mathbf{P}_k = (\mathbf{P}_k^0, \mathbf{P}_k^1, \dots, \mathbf{P}_k^{nr})$, $\mathbf{q}_k^0 = \mathbf{Q}_k(\mathbf{x}_k)$; $\mathbf{q}_k^i = \mathbf{q}_k^0 + \mathbf{P}_k^i$, $1 \leq i \leq nr$; $\mathbf{q}_k^{nr+1} = \mathbf{x}_k$; $\mathbf{x}_{k+1} = D(\mathbf{x}_k, \mathbf{P}_k) = \text{Arg Min}\{F(\mathbf{q}_k^i) : 0 \leq i \leq nr+1\}$, i.e., \mathbf{x}_{k+1} is the best point among the $nr+2$ available points \mathbf{q}_k^i , $0 \leq i \leq nr+1$.

The random perturbations may be also used as part of a population-based algorithm. For instance, we may consider the following algorithm

Population-based algorithm involving random perturbations:

Start: let be given: $h > 0$; strictly non negative integers np , nc and nr ; the initial population $\Pi_0 = \{\mathbf{x}_0^1, \dots, \mathbf{x}_0^{np}\} \subset S$; the iteration function \mathbf{Q} and the number of substeps $ns > 0$. k is initialized: $k = 0$.

Step $k > 0$: k is increased and Π_k is obtained from

Π_{k-1} in three substeps:

1) We determine the set

$$C_k = \{ \mathbf{u}_k^i = \rho_k^i \mathbf{x}_k^j + \kappa_k^i \mathbf{x}_k^m + \eta_k^i : \mathbf{x}_k^j, \mathbf{x}_k^m \in \Pi_{k-1} : i = 1, \dots, nc \}$$

where $\rho_k^i, \kappa_k^i, \eta_k^i$ are random values from the uniform distribution on $[-h, h]$ and j, m are randomly chosen. Since C_k contains nc elements,

$B_k = \Pi_{k-1} \cup C_k$ contains $nb = np + nc$ elements.

2) We determine the set

$$M_k = \{ \mathbf{m}_k^i = D(\mathbf{b}_k^i, \mathbf{P}_k^i) : \mathbf{b}_k^i \in B_k : i = 1, \dots, nb \},$$

M_k has nb elements chosen among the $nb(nr+2)$ available points ($nr+2$ points for each element of B_k).

3) The elements of M_k are increasingly ordered according to the associated value of F . The population Π_k is formed by the first np elements of M_k , which correspond to the np best points of M_k .

4) Termination test: a stopping condition is tested and the iterations are terminated if it is verified.

At each iteration, the optimal point \mathbf{x}^* and the optimal value F^* are approximated by

$$\mathbf{x}^* \approx \mathbf{x}_k^* = \mathbf{x}_k^1 = \text{Arg Min}\{F(\mathbf{x}_k^i), 1 \leq i \leq np\}, F^* \approx F_k^* = F(\mathbf{x}_k^*)$$

The stopping conditions usually involve a maximum iteration number k_{\max} , a minimal variation η_{\min} of the optimal point and a minimal improvement ε_F of the objective function: the iterations are stopped when *one* among the following conditions is satisfied:

$$k = k_{\max} \quad \text{OR} \quad \|\mathbf{x}_k^* - \mathbf{x}_{k-1}^*\| \leq \eta_{\min} \quad \text{OR} \quad |F_k^* - F_{k-1}^*| \leq \varepsilon_F$$

We denote by k_{stop} the iteration number where this condition is satisfied. Thus, our final estimation of the optimal

point is $\mathbf{x}_e^* = \mathbf{x}_{k_{stop}}^*$ and the final estimation of the optimal value is $F_e^* = F(\mathbf{x}_{k_{stop}}^*)$.

Each iteration performs $(np+nc) \times ns$ evaluations of \mathbf{Q} and $(nr+2) \times (nc+np)$ evaluations of F . Thus, the computational cost corresponds to $k_{stop} \times ns \times (np+nc)$ evaluations of \mathbf{Q} and $k_{stop} \times (nr+2) \times (nc+np)$ evaluations of F .

In a previous work we have presented a procedure of calibration by an analogous population-based strategy (Gonçalves[18] and Gonçalves[19]). The convergence of this method to the global minimum is ensured by a theorem, independently of the initial population, but the numerical behavior may be widely improved if a convenient initial population Π_0 is chosen: we present here a significant improvement based on the use of a representation of \mathbf{x} . Our main result is a generalization of a pioneer result (Pincus[25]), which has been extended to a more numerically useful representation (Souza de Cursi[32] and Souza de Cursi[34]).

In our numerical experiments, we have considered different methods for the generation of the descent direction: gradient descent (GD), normalized gradient descent with (NGD); Fletcher-Reeves (FR), Polak-Ribière (PR), Davidson-Fletcher-Powell (DFP) and Broyden-Fletcher-Goldfarb-Shanno (BFGS). By reasons of limitation of the room, we do not present here these classical methods: the reader is invited to refer to the literature in continuous optimization, such as, for instance, Bazaraa[4].

4 A representation of the optimal point

In order to properly address the main theorem, we need some notations: let $\varepsilon > 0$ be a real number small enough. We denote by B_ε the open ball having center \mathbf{x} and radius ε , while $S_\varepsilon = S - B_\varepsilon$ denotes its complement in S :

$$B_\varepsilon = \{y \in S : \|y - \mathbf{x}\| < \varepsilon\}; \quad S_\varepsilon = \{y \in S : \|y - \mathbf{x}\| \geq \varepsilon\}.$$

We denote by ψ_ε the characteristic function of S_ε :

$$\psi_\varepsilon(y) = 1, \text{ if } y \in S_\varepsilon; \quad \psi_\varepsilon(y) = 0, \text{ otherwise.}$$

Let us introduce a probability P defined on S . We assume that

$$\forall \eta > 0 : P(B_\eta) > 0.$$

This assumption is verified by any probability defined by a strictly positive density. Let $\tau > 0$ be a real number large enough (in the sequel, we take $\tau \rightarrow +\infty$) and $g: R^2 \rightarrow R$ a continuous function such that $g \geq 0$ and $\zeta \rightarrow g(\tau, \zeta)$ is strictly decreasing for any $\tau > 0$. Let us denote by $E(\bullet)$ the meaning associated to the probability P . We assume that there exists a real number $\varepsilon_0 > 0$ and two functions $h_1, h_2: R^2 \rightarrow R$ such that,

$$\forall \varepsilon \in (0, \varepsilon_0): \\ E(g(\tau, F(y))) \geq h_1(\tau, \varepsilon) > 0; \quad E(\psi_\varepsilon(y)g(\tau, F(y))) \leq h_2(\tau, \varepsilon) \quad (20)$$

and

$$\forall \varepsilon \in (0, \varepsilon_0): \frac{h_2(\tau, \varepsilon)}{h_1(\tau, \varepsilon)} \xrightarrow{\tau \rightarrow +\infty} 0. \quad (21)$$

We have:

Theorem 4.1. Let $S \subset V$ be a non empty, closed, bounded set. Assume that $F: V \rightarrow R$ is continuous and that there is exactly one $\mathbf{x} \in S$ satisfying (10). Assume also that (20) and (21) are satisfied. Then

$$\mathbf{x}^* = \lim_{\tau \rightarrow +\infty} \frac{E(yg(\tau, F(y)))}{E(g(\tau, F(y)))} \quad (22)$$

Proof: Cf. Appendix 1.

The choice of the function g is guided by the following result:

Proposition 4.2. Let $h: R \rightarrow R$ be a continuous function such that $h > 0$, h is strictly decreasing on $(0, +\infty)$ and

$$\forall \xi, \delta > 0 : \lim_{\tau \rightarrow +\infty} \frac{h(\tau(\xi + \delta))}{h(\tau\xi)} = 0 \quad (23)$$

If $\mu: R \rightarrow R$ is continuous, strictly positive and strictly increasing on $(F(\mathbf{x}), +\infty)$ then $g(\tau, \xi) = h(\tau\mu(\xi))$ satisfies (20) and (21). Consequently, we have (22).

Proof: Cf. Appendix 1.

A suitable choice, primarily suggested by (Pincus[25]) is

$$g(\tau, \zeta) = \exp(-\tau\xi), \quad (24)$$

which corresponds to

$$h(\xi) = \exp(-\xi); \quad \mu(\xi) = \xi. \quad (25)$$

The numerical experiments performed show that the representation (Eq. 22) furnishes good starting points for iterative descent methods (Souza de Cursi[32]). This suggests its use for the generation of the initial population S_0 : it is expected an improvement, since the initial population is an approximate numerical representation of the solution. Recent experiments have shown a significant improvement by the use of a representation in the generation of the initial population, when compared to a purely random initial population (Shi[28], Souza de Cursi[33] and Bez[7]).

The representation (22) is numerically used as follows: a large fixed value of τ is selected and a finite sample $(\mathbf{x}_1, \dots, \mathbf{x}_{n_{tir}})$ of admissible points is generated, according to the probability P . Then, the sample is used to generate an initial estimation of the optimal point:

$$\mathbf{x}^* \approx \mathbf{x}_0 = \frac{\sum_{i=1}^{ntirm} \mathbf{x}_i g(\tau, F(\mathbf{x}_i))}{\sum_{i=1}^{ntirm} g(\tau, F(\mathbf{x}_i))} \quad (26)$$

Eq. (26) shows that Eq. (22) may be interpreted as a weighted mean of the values of the points of S . For conveniently chosen g , the weight of a point \mathbf{y} decreases when $F(\mathbf{y})$ increases. The behaviour of g as function of τ must be such that the relative weight of points having values different of $F(\mathbf{x})$ vanishes when $\tau \rightarrow +\infty$.

The practical use of Eq. (26) (or Eq. (22)) requests the construction of a probability distribution on S . When the geometry of S is complex, this may be a hard point. In the sequel, we assume the existence of a projection operator $proj$ transforming $\mathbf{x} \notin S$ into $\mathbf{x}_{proj} \in S$.

The use of (26) for the generation of the initial population Π_0 needs $np \times ntirm$ supplementary evaluations of F : the computational cost becomes $k_{stop} \times ns \times (np+nc)$ evaluations of \mathbf{Q} and $np \times ntirm + k_{stop} \times (nr+2) \times (nc+np)$ evaluations of F .

5 Numerical experiments

In this section, we present the results of some numerical experiments destined to validate the algorithm.

Extensive experimentation has been performed concerning the effects of the parameters: choice of the descent method used in the mutation phase and its internal parameters, size of population, standard deviation of the distributions used for generation, maximum step size etc. For each set of parameters, one hundred runs have been performed in order to estimate the probability of success of the method: a run is considered as successful when the relative error in the final evaluation \mathbf{x}_e^* of the optimum \mathbf{x}^* is inferior to 0.1%, i.e., the distance between the final estimation \mathbf{x}_e^* and the exact solution \mathbf{x}^* ($\neq \mathbf{0}$ in all experiments) is inferior to 0.1% of the norm of the exact solution: $\|\mathbf{x}_e^* - \mathbf{x}^*\| \leq 10^{-3} \|\mathbf{x}^*\|$ (we have $\mathbf{x}^* \neq \mathbf{0}$ in all the experiments). The probability of success is estimated by the proportion of successful runs on a finite sample of 100 independent runs. We consider also the mean value and the standard deviation of k_{stop} in the sample, denoted by $m(k_{stop})$ and $s(k_{stop})$, respectively.

By reasons of limitation of the room, we present only the results concerning a few experiments, chosen in order to illustrate significant aspects.

We have considered different choices of the function g : (25) has been tested, but we have also considered the situations where $h(\xi) = \exp(-\xi)$ and

$$\mu(\xi) = \varphi \ln(\theta + \xi^\gamma) \Rightarrow g(\tau, \xi) = \left(\frac{1}{\theta + \xi^\gamma} \right)^{\nu\tau}, \quad \nu, \gamma, \theta > 0; \quad (27)$$

$$\mu(\xi) = \varphi \sqrt{\theta + \xi^\gamma} \Rightarrow g(\tau, \xi) = \exp\left(-\nu\tau \sqrt{\theta + \xi^\gamma}\right), \quad \nu, \gamma, \theta > 0; \quad (28)$$

In our experiments, the search region S has had a simple geometry:

- *disk*: $S = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y}\| \leq r\}$, where r is the radius; or
- *rectangular*: S is a product of intervals, $S = [x_{\min}, x_{\max}]^n$.

In order to apply the theorem 4.1, we need a probability P defined on S . In our experiments, we consider the restriction to S of a Gaussian distribution $N(0, \rho Id)$.

In practice, we generate points from $N(0, \rho Id)$. As previously observed, trial points lying outside S are projected by an operator $proj$ in order to get an admissible point. We have considered two kinds of projection:

- The first one is the standard orthogonal projection (SOP) on S . This kind of projection brings the exterior point to the boundary of S . In our experiments, this projection reads as

- *disk*: $proj(\mathbf{y}) = r \frac{\mathbf{y}}{\|\mathbf{y}\|}$, if $\|\mathbf{y}\| > r$; $proj(\mathbf{y}) = \mathbf{y}$, otherwise.

- *rectangular*: $proj(\mathbf{y}) = (p(y_1), \dots, p(y_n))$,

where $p(y) = x_{\min}$, if $y < x_{\min}$; $p(y) = x_{\max}$, if $y > x_{\max}$; $p(y) = y$, otherwise.

- The second one is a randomly perturbed orthogonal projection (RPOP) where a random term is added to SOP in order to obtain an interior point whenever the trial point \mathbf{y} lies outside from S . In our experiments, this projection is performed as follows:

- *disk*: $proj(\mathbf{y}) = (1-U)r \frac{\mathbf{y}}{\|\mathbf{y}\|}$, if $\|\mathbf{y}\| > r$ (U is a

variable from the uniform distribution on $(0,1)$); $proj(\mathbf{y}) = \mathbf{y}$, otherwise.

- *rectangular*: let $\ell = (x_{\max} - x_{\min})$. $proj(\mathbf{y}) = (rp(y_1), \dots, rp(y_n))$, where $rp(y) = x_{\min} + \ell U$, if $y < x_{\min}$; $rp(y) = x_{\max} - \ell U$, if $y > x_{\max}$; $rp(y) = y$, otherwise (U is a variable from the uniform distribution on $(0,1)$).

In order to investigate the effect of the projection, we have tested initial populations formed by purely boundary points and purely interior points, as well as mixed ones.

We recall that another distribution of probability is needed in order to generate the perturbations \mathbf{P}_k (Eq. (19)).

As previously observed, we use $\mathbf{P}_k = \frac{\omega}{\sqrt{\log(k+1)}} \mathbf{Z}$, where

\mathbf{Z} is a gaussian random vector following $N(0, Id)$ and $\omega > 0$ is fixed.

5.1. Test functions

We have considered classical test functions existing in the literature. These functions are characterized by multimodal behaviour and difficulties in numerically obtaining the points of global optima. They are usually destined to the

evaluation of the performance of global optimization algorithms.

In order to prevent bias introduced by the use of probability distributions having $\theta = (0,0,0,\dots,0) \in \mathcal{R}^n$ as mean, we consider only situations where the theoretical solution is $\mathbf{x}^* \neq \theta$. The functions used are the following:

1) Davis' Function:

$$F(\mathbf{x}) = 0.5 + \frac{\sin^2(\|\mathbf{x} - \bar{\mathbf{x}}\|_2) - 0.5}{(1 + \|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 / 1000)^2} \quad (29)$$

2) Rastrigin's Function :

$$F(\mathbf{x}) = 3n + \sum_{i=1}^n [(x_i - \bar{x}_i)^2 - 3\cos(2\pi(x_i - \bar{x}_i))] \quad (30)$$

3) Ackley's function:

$$F(\mathbf{x}) = 20 \left(1 - e^{-0.2 \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}_i)^2}{n}}} \right) + e - e^{\frac{\sum_{i=1}^n \cos(2\pi(x_i - \bar{x}_i))}{n}} \quad (31)$$

4) Griewank's function (Griewank[20]) :

$$F(\mathbf{x}) = \frac{1}{200} \|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 - \prod_{i=1}^n \cos\left(\frac{x_i - \bar{x}_i}{\sqrt{i}}\right), \quad (32)$$

5) Rosenbrock's function :

$$F(\mathbf{x}) = 100 \sum_{i=1}^n (x_i^2 - x_i)^2 + \sum_{i=1}^n (x_i - 1)^2 \quad (33)$$

6) Schwefel's function :

$$F(\mathbf{x}) = 418.9829 n - \sum_{i=1}^n x_i \sin(\sqrt{|x_i|}) \quad (34)$$

The first four functions attain their global minima at $\mathbf{x}^* = \bar{\mathbf{x}}$ - our experiments use $\bar{\mathbf{x}} = (1, 2, 3, \dots, n)$. We have $F(\mathbf{x}^*) = 0$ for the 3 first ones (Davis, Rastrigin, Ackley), while $F(\mathbf{x}^*) = -1$ for the 4th one (Griewank). The 5th (Rosenbrock) has a global minimum $F(\mathbf{x}^*) = 0$, attained at $\mathbf{x}^* = (1, 1, 1, \dots, 1) \in \mathcal{R}^n$. The last one (Schwefel) has global minimum $F(\mathbf{x}^*) = 0$, attained at $\mathbf{x}^* = (420.96, 420.96, \dots, 420.96) \in \mathcal{R}^n$.

5.2. Numerical effect of the use of the representation

In order to evaluate the improvement furnished by the use of Eq. (26) for the generation of the initial population Π_0 , the results have been compared to those furnished by a purely random one $\tilde{\Pi}_0$. In all the situations considered, Π_0 has furnished better results. The results have been analogous for all considered test functions: The results obtained for Rastrigin's function are presented in Table 1. They concern the situation where $np = 5$, the search region is rectangular with $x_{\min} = -500$ and $x_{\max} = 500$; $ntirm = 100$; $\rho = 1$; $\omega = 0.5$; $\tau = 10$; $kmax = 300$; $nr = 5$; $a_{\max} = 0.5$; $ns = 10$. When

generating the initial population, we use the random projection RPOP.

Table 1 – Rastrigin's function ($n = 5$, RPOP)

	GD	FR	DFP	BFGS	PR
Prob. of success	0.91	0.92	0.96	0.96	0.96
$m(k_{stop})$	53.98	55.72	47.56	67.25	68.75
$s(k_{stop})$	51.33	52.49	51.60	71.78	75.06
evaluations of \mathcal{Q}	8098	8359	7134	10087	10312
evaluations of F	5668	5851	4994	7061	7218

a) Random Initial Population ($\tilde{\Pi}_0$)

Prob. of success	1	1	0.98	0.99	1
$m(k_{stop})$	23.38	25.56	33.97	35.69	24.75
$s(k_{stop})$	32.07	38.58	37.83	37.92	36.62
evaluations of \mathcal{Q}	3597	3834	5095	5353	3712
evaluations of F	2954	3183	4066	4247	3098

b) Initial population generated by the representation(Π_0)

The method of projection has had no significant influence: standard projection has led to analogous results (see Table 2)

Table 2 – Rastrigin's function ($n = 5$, SOP)

	GD	FR	DFP	BFGS	PR
Prob. of success	0.94	0.91	0.92	0.94	0.97
$m(k_{stop})$	79.02	66.28	65.19	71.24	62.93
$s(k_{stop})$	63.24	54.42	49.56	64.05	54.55
evaluations of \mathcal{Q}	11853	9942	9779	10686	9440
evaluations of F	8297	6960	6845	7480	6608

a) Random Initial Population ($\tilde{\Pi}_0$)

Prob. of success	1	1	1	0.98	1
$m(k_{stop})$	25.70	26.22	34.24	35.43	26.22
$s(k_{stop})$	32.73	36.85	34.96	31.87	37.11
evaluations of \mathcal{Q}	3855	3933	5136	5314	3933
evaluations of F	3198	3253	4095	4220	3253

b) Initial population generated by the representation(Π_0)

The Fig. 1 shows the probability of success for an increasing population size. The search region is a disk of radius $r = 100$; the dimension is $n = 6$; and the parameters are $\rho = 1$; $ntirm = 50$; $\omega = 0.1$; $\tau = 10$; $kmax = 30$; $nr = 30$; $ns = 10$; $a_{\max} = 1$; descent method: FR. We observe the im-

provement due to the use of the representation for the generation of the initial population.

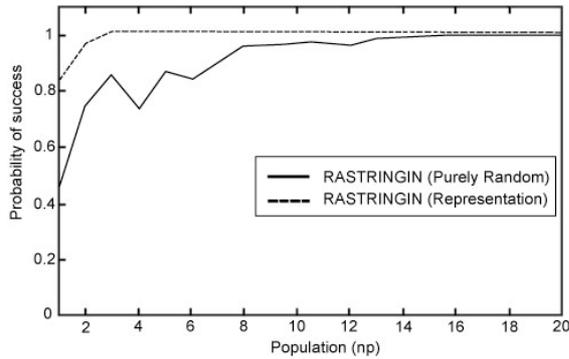


Fig. 1 – Results for an increasing population size (Rastrigin’s function)

In addition, the use of the representation formula may significantly reduce the number of random perturbations to be generated. This confirms that (22) furnishes good starting points. For instance, let us consider the situation above with $nr = 0$ (no random perturbation). The results are shown in Table 3. The projection method is RPOP.

Table 3 – Rastrigin’s function ($n = 5$, RPOP) – no random perturbation ($nr = 0$)

	GD	FR	DFP	BFGS	PR
Prob. of success	0.52	0.47	0.57	0.55	0.46
$m(k_{stop})$	114.67	114.14	123.29	122.6	138.54
$s(k_{stop})$	74.16	68.30	81.05	74.02	78.13
evaluations of \mathcal{Q}	17200	17122	18494	18390	20781
evaluations of F	3440	3424	3698	3678	4156
a) Random Initial Population ($\tilde{\Pi}_0$)					
Prob. of success	0.83	0.68	0.77	0.69	0.70
$m(k_{stop})$	62.33	75.65	64.93	64.04	76.92
$s(k_{stop})$	58.55	84.09	64.15	70.22	72.63
evaluations of \mathcal{Q}	9350	11347	9739	9607	11538
evaluations of F	2370	2769	2447	2421	2807

b) Initial population generated by the representation(Π_0)

The results for the Ackley’s function are given in Tables 4 and 5.

Table 4 – Ackley’s function ($n = 20$, RPOP) – no random perturbation ($nr = 0$)

	GD	FR	DFP	BFGS	PR
Prob. success	1	1	1	1	1
$m(k_{stop})$	41.9	59.93	44.3	44.91	77.54
$s(k_{stop})$	18.88	31.74	19.42	18.66	46.40
evaluations of \mathcal{Q}	3142	4494	3322	3368	5815

evaluations of F	1257	1797	1329	1347	2326
--------------------	------	------	------	------	------

a) Random Initial Population ($\tilde{\Pi}_0$)

Prob. of success	1	1	1	1	1
$m(k_{stop})$	23.44	44.89	29.86	28.15	45.36
$s(k_{stop})$	9.88	20.77	14.63	15.70	22.62
evaluations of \mathcal{Q}	1758	3366	2239	2111	3402
evaluations of F	953	1596	1145	1094	1610

b) Initial population generated by the representation(Π_0)

Table 5 – Ackley’s function ($n = 10$, RPOP) – no random perturbation ($nr = 0$)

	GD	FR	DFP	BFGS	PR
Prob. of success	1	1	1	1	1
$m(k_{stop})$	18.59	29.76	19.42	20.84	24.7
$s(k_{stop})$	6.75	19.23	8.84	8.57	13.23
evaluations of \mathcal{Q}	1394	2232	1456	1563	1852
evaluations of F	557	892	582	625	741
a) Random Initial Population ($\tilde{\Pi}_0$)					
Prob. of success	1	1	1	1	1
$m(k_{stop})$	8.76	16.42	9.74	8.57	14.53
$s(k_{stop})$	7.98	10.27	8.33	7.01	11.30
evaluations of \mathcal{Q}	675	1231	730	642	1089
evaluations of F	512	742	542	507	685

b) Initial population generated by the representation(Π_0)

5.3. Tuning of the main parameters

In order to verify the robustness of the algorithm and the simplicity of tuning, the effects of the variations of the main parameters np , α_{max} , ρ , ω and ns have been studied:

5.3.1. Size of the population

The tuning of the size of the population np was easy: good results have been obtained in all the situations for $np \geq 5$.

For instance, let us consider the situation where the search region is a disk of radius $r = 100$; $\rho = 100$; $ntirm = 150$; $\tau = 10$; $kmax = 50$; $ns = 5$; $nr = 50$; $\omega = 0.01$; $\alpha_{max} = 1$; the descent method is GD. The results obtained are given in Fig. 2 and confirm the observation.

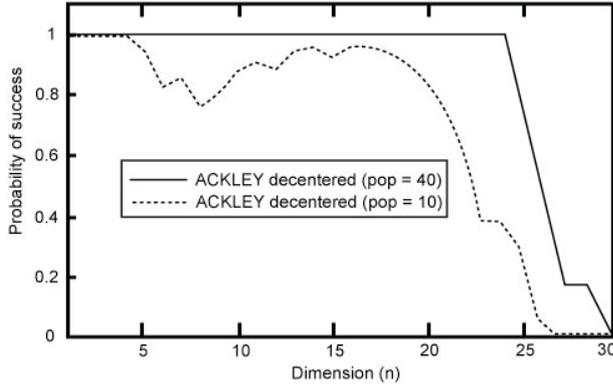


Fig. 2 – Influence of the population size (Ackley’s function)

In Table 6, we present the results for different values of np and ρ . The search region is a disk of radius $r = 10$ and we use $ntirm = 100$; $\tau = 10$; $kmax = 50$; $ns = 10$; $nr = 30$; $\omega = 0.05$; $\alpha_{max} = 1$. The descent method is GD and the projection method is RPOP. As expected, increasing the size of population increases the probability of success.

Table 6 – Probability of succes for different values of np and ρ

P	0,1	0,5	1	5	10	100
$np = 5$	1	0,88	0,74	0,60	0,60	0,59
$np = 30$	1	1	1	1	1	1

5.3.2. Maximum step size α_{max}

The tuning of the step size α_{max} has shown to be easy: good results have been obtained in all the situations for $\alpha_{max} \geq 1$.

For instance, let us consider Davis’s function decentered the situation where the search region is a disk of radius $r = 100$; $\rho = 1$; $np = 5$; $ntirm = 50$; $\tau = 10$; $kmax = 30$; $nr = 30$; $ns = 10$; $\omega = 0.01$; the descent method is GD. The results obtained are exhibited in Fig. 3 and confirm the observation. For ρ superior to 3.5, the probability of success is constant and it equals to 1.

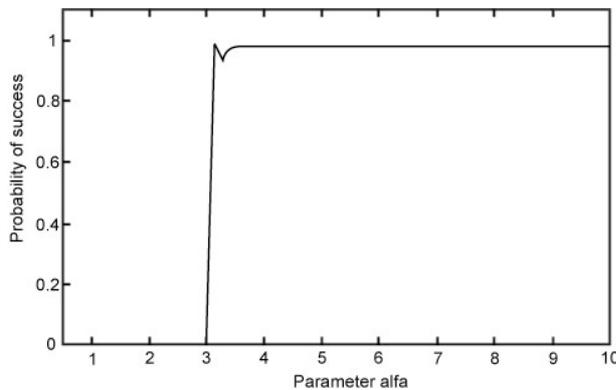


Fig. 3 – Influence of the parameters α_{max} (Davis)

5.3.3. Standard Deviation ω used in the random perturbations

We have analyzed the influence of the standard deviation ω , used in the random perturbations. For instance, Fig. 4 shows the results concerning the Schwefel’s function in

dimension $n = 5$ for an initial population formed of purely random points, without use of the Representation Formula (thus, $ntirm = 0$). The search region is rectangular with $x_{min} = -500$ and $x_{max} = 500$ and we use $np = 5$; $kmax = 30$; $\rho = 20$; $ns = 5$; $nr = 50$; $\alpha_{max} = 1$. The descent method is the GD and the projection method is SOP. We observe that the probability of success increases with the standard deviation.

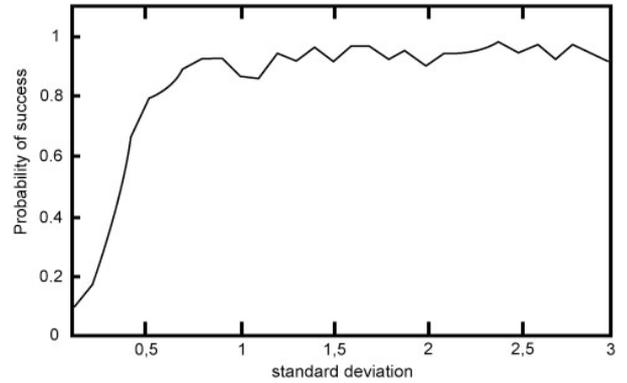


Fig. 4 – Influence of the standard deviation ω with $\tilde{\Pi}_0$ (Purely random)

The use of the Π_0 (Representation) has led to an improvement: the probability of success has been equal to one for all the values of ω . The speed of convergence has also increased with ω .

5.3.4. Number of substeps ns used in mutations

The tuning of ns has shown to be easy: good results have been obtained in all the situations for values superior to 2.

For instance, let us consider the situation where the search region is rectangular with $x_{min} = -500$ and $x_{max} = 500$; $n = 30$; $ntirm = 100$; $\rho = 1$; $\omega = 0.1$; $\tau = 10$; $kmax = 30$; $nr = 10$; $\alpha_{max} = 1$; the descent method is GD. The results obtained are exhibited in Table 7 and confirm the observation (Griewank’s function).

Table 7 – Influence of the number of the substeps ns used in mutations

	$ns=1$	$ns=2$	$ns=3$	$ns=6$	$ns=13$	$ns=17$
Prob. of success	0	0,99	1	1	1	1
$m(k_{stop})$	-----	22,13	15,12	8,94	5,24	4,38
$s(k_{stop})$	-----	2,60	2,04	1,40	0,84	0,52
evaluations of Q		663	680	804	1021	1116
evaluations of F		4483	3221	2109	1443	1288

5.3.5. Standard deviation ρ

The tuning of the standard deviation ρ has shown to be easy: good results have been obtained in all the situations for $0.1 \leq \rho \leq 100$ (see Table 6 for examples of results).

5.4. Choice of g

The different choices of the function g have led to similar results, for a wide range of values of the parameters ν , γ , θ . In our experiments, ν , γ had varied in a interval mesh of 0.5 defined in a rectangle $[0.5, 5] \times [0.5, 5]$ and $\tau = 10$, $\theta = 5$ fixed.

Let us consider the situation where the search region is a disk of radius $r = 100$; $np = 5$; $ntirm = 100$; $kmax = 50$; $ns = 10$; $nr = 30$; $\alpha_{max} = 1$; $\omega = 0.1$. The descent method is GD. Table 8 shows the results furnished by the method for various functions g . The results for each choice of g are stable (small standard deviation) and Eq. (27) has been more performing, but its results are very close to those of Eq. (28).

Table 8 – Impact of g – (Griewank, $n = 10$)

	Eq. (25)	Eq. (27)	Eq. (28)
Probability of success	1	1	1
$m(k_{stop})$	2,92	2,05	2,09
$s(k_{stop})$	0,27	0,035	0,039
evaluations of Q	438	307	313
evaluations of F	1901	1484	1503

The results concerning the Rastrigin's function are shown in Table 9. The results are analogous.

Table 9 – Impact of g – (Rastrigin, $n = 5$)

	Eq. (25)	Eq. (27)	Eq. (28)
Probability of success	0,53	0,54	0,60
$m(k_{stop})$	35,33	34,15	34,20
$s(k_{stop})$	3,23	2,30	2,74
evaluations of Q	5299	5122	5130
evaluations of F	17458	16892	16916

6. Calibration of the model of competitive destinations

6.1. Set of initialization parameters and data

In order to evaluate the efficiency of the numeric methods in the calibration of the Fotheringham model, a set of hypothetical data presented by Kühlkamp[22] has been used. These data involve 30 traffic zones and has been often used to analyze the intervening opportunities model.

We compare the proposed numeric procedure ERPG-RF (Evolutionary version of the Random Perturbations of the Gradient – Representation Formula) to the evolutionary strategy of (Diplock[10]), denoted by EDO, where an evolutionary strategy involving a quasi-Newtonian method and a basic generic algorithm is used for the calibration of the Fotheringham model. The work (Diplock[10]) mentions that one of the reasons for the choice of the evolutionary strategy was the work developed by Schwefel[29] that verified, in an empiric way, the good performance of this strategy, when compared to eleven conventional optimization methods out of twenty eight test problems. According to Diplock[10], the same good performance of this method has been observed in

the calibration, which motivates his use to get a comparison with the ERPG-RF approach.

The algorithm EDO reads as follows:

Evolutionary Strategy algorithm (Diplock[10]):

- **Step 1:** Identify initial parameter values and standard deviations for the parent.
- **Step 2:** Generate new parameter values for a descendant by mutation; this involves adding pseudo-random numbers drawn from a normal distribution centered on the parent values with a given standard deviation.
- **Step 3:** If the descendant performs better than the parent, then it becomes the parent for the next generation.
- **Step 4:** Repeat steps 2 and 3 for n mutations and count how many improved values have been found over the preceding $10n$ mutations. If greater than $2n$, then reduce the standard deviations by 0,85; else, increase them divide by 0,85.
- **Step 5:** Return to step 2 until no change occurs over $10n$ mutations on two consecutive occasions.

The tests using the ERPG-RF method involve as search region a disk of radius $r = 10$; $np = 5$; $\rho = 1$; $ntirm = 100$; $\tau = 10$; $kmax = 50$; $nr = 5$; $\alpha_{max} = 0.7$; $\omega = 0.1$; descent method: GD. The iterations have been stopped at the fixed maximal iteration number

For EDO, a random population of 100 elements is generated and $N(0,1)$ is used in the mutation phase.

An extensive set of tests for both methods has been executed, varying parameters, such as population size, standard deviation, among others.

6.2. Analysis of the obtained results

The executed tests verified, among other things, the number of evaluations of the function, the function value evolution, according to the applied numeric method and the influence of the Representation Formula in the performance of the ERPG-RF method.

Table 10 – Parameters estimated by minimizing F defined by Eq. (9)

	β	σ	δ	value of F
ERPG-RF	0,08	0,12	2,48	10679,30
EDO	0,04	0,12	2,29	10896,86

Table 10 presents the parameters obtained by minimizing the objective function given in Eq. (9). It is observed that the optimal values obtained by both methods are very close.

Nevertheless, the behavior of the methods is not the same, as shown in Fig. 5. We can notice that the ERPG-RF method presents a better performance, considering that the function optimal value is reached in a significantly smaller number of iterations. In addition, the ERPG-RF method uses a smaller population. It is known that a good initial population, linked to the method robustness, brings the necessity of a smaller number of evaluations of the function.

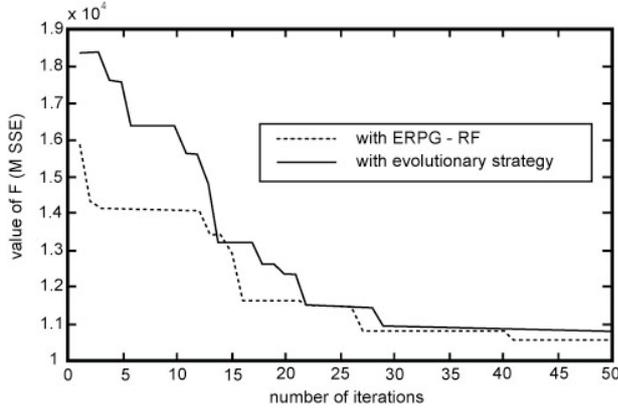


Fig. 5: Minimal value of the function at each interaction (Eq. 9)

The good performance of the ERPG-RF method in the calibration of the competitive destination models, presented in this study, was also verified in an application of the opportunities gravitational model (Bez[8]), according to what we'll see in the next section.

7. Calibration of the gravity-opportunity model

7.1. Set of initialization parameters and data

We have tested the method on 4 data sets: three sets of real data and one set of simulated data. The first set of real data comes from the region of Londrina, Brazil, which is divided into 12 traffic zones (Almeida[1]). The second one corresponds to a region that includes part of the State of Santa Catarina, Brazil, divided into 77 zones (Gonçalves[16]). The third one is a subset of the previous region, corresponding to 44 zones of particular interest (Gonçalves[16]). The set of simulated data is defined on a region divided into 30 zones (Kühlkamp[22]).

All the experiments performed use the following set of parameters: $np = 2$; $ntirm = 100$; $nr = 5$; $\rho = 0.5$; $\omega = 0.02$; $\alpha_{max} = 0.7$. Furness iterations have been stopped when either the corresponding iteration number j has reached a prescribed maximum value $j_{max} = 100$ or the value of the objective function has attained 10^{-3} : $F(\beta, \lambda) < 10^{-3}$.

7.2. Analysis of the obtained results

The objective function under consideration is non-convex, as we can see in the Fig. 6 and 7.

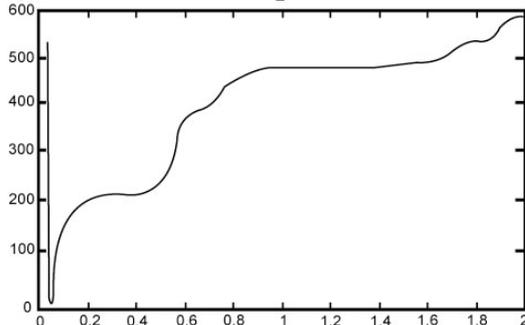


Fig. 6 – Surface section $f(\beta, \lambda)$: straight line $\lambda = 2\beta$ in the interval $[0,009; 2]$ – 77 zones (Gonçalves[18])

In the Fig. 8, we can notice the difficulty in minimizing this function, by using the hypothetical data set generated by Kühlkamp[22], characterized by the large variations of its derivatives in a interval mesh of 0.001 defined in a rectangle $[0, 0.07] \times [0, 0.07]$.

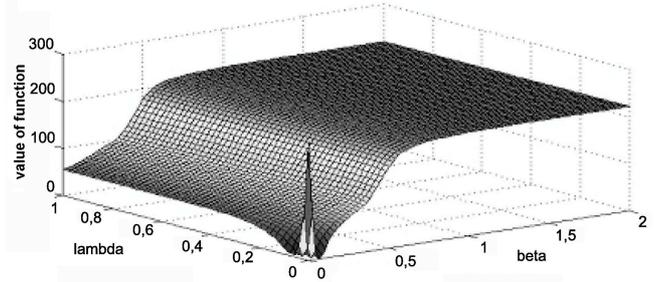


Fig. 7 – Function defined by the Maximum Likelihood criteria (Bez[6])

In the study developed by Gonçalves[19], the ERPG (*Evolutionary version of the random perturbations of the gradient*) algorithm was applied to the hybrid gravitational/intervening opportunities model (Gonçalves[16]). The variation of the parameters, such as the population size, random perturbations added to the gradient method, number of iterations, definition of the error (Gonçalves[19]) to be applied to the Furness method, among others, have been tested and evaluated. In practice, the Furness iterations are stopped at a given finite number of iterations (Gonçalves[19]). In this study, a maximum number of 100 iterations has been considered.

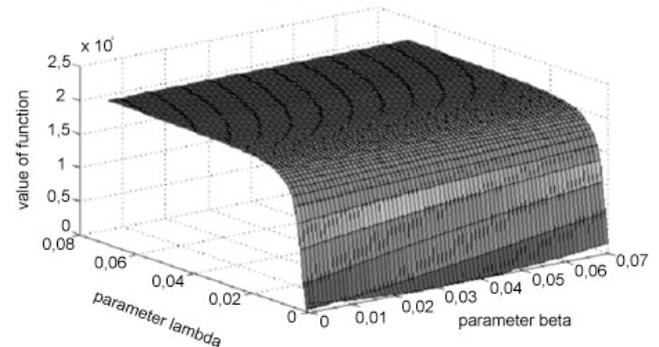


Fig. 8 – Function defined by the maximum verisimilitude criterion – 30 zones

In the application of the proposed method in this study, the same good performance, verified in the tests applied to the classic functions, was confirmed in the calibration of the gravity-opportunity model (Gonçalves[16]). The application of the Representation Formula in the improvement of the initial population has led to a better convergence speed. In the Table 11, we can observe the average value of the function, obtained after the execution of 100 runs, with or without the use of the Representation Formula. The Table 11 uses $\rho = 0.5$ and $ntirm = 100$. The results show the significant drop of these values obtained by the application of this Formula.

Table 11 – Mean optimal value of the objective function given by Eq. (15).

	Gonçalves[16] 77 zones	Gonçalves[16] 44 zones
without RF	70745.27	11178.77
with RF	22.94	7.96
	Almeida[1] 12 zones	Kühlkamp[22] 30 zones
without RF	257.27	52648.36
with RF	1.13	43.32

The use of the Representation Formula in the improvement of the initial population, contributed for the increase in the convergence speed, causing a less number of iterations and, as consequence, the reduction of the number of function evaluations, as we can verify in the results presented in the Table 12.

The values of the Table 12 were determined after the execution of 100 tests, performed with every data set.

Table 12 – Mean number of function evaluations given by Eq. (15).

	Gonçalves[16] 77 zones	Gonçalves[16] 44 zones
without RF	8800	6720
with RF	5960	2120
	Almeida[1] 12 zones	Kühlkamp[22] 30 zones
without RF	1960	3720
with RF	1160	2440

In the Fig. 9, we can observe the performance of the numeric procedure, with or without the use of the Representation Formula and, with or without the use of the optimal step. The best performance of the method is evidenced when we use the Representation Formula and the optimal step in the procedure. Then same tests were applied to all the data sets used and the obtained results presented the same tendency.

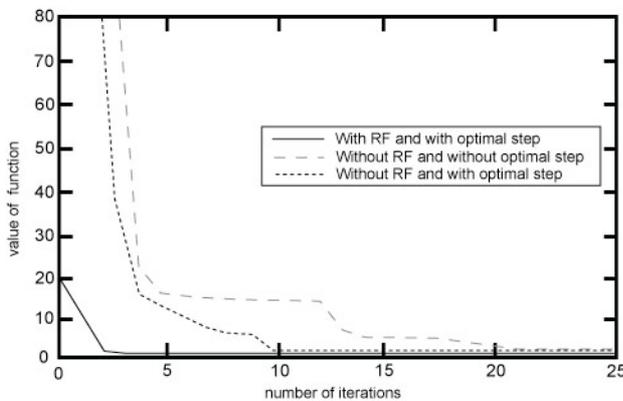


Fig. 9 – Influence of Eq. 22 and of use of the optimum step in the gradient method - Set of data with 44 zones

8. Conclusions

In this study, a stochastic global optimization procedure, of the evolutionary type, was presented, by using descent

methods in the mutation phase and the Representation Formula of the global optimum in the determination of the initial population.

The initial population is one of the significant parameters for the performance of evolutionary methods and the use of the Representation Formula has shown to furnish an improvement. The random perturbations of the deterministic descent methods considered have shown to be effective in the mutation step, enhancing the speed of convergence and preventing from convergence local minima, excepted Newton's method.

The experiments have shown that different choices of g are possible, according to Souza de Cursi[33]: similar results have been obtained for the functions considered.

Another fact to be emphasized was that, in most of the cases, when the Representation Formula was used in the improvement of the initial population, there was no need to use a big number of individuals (points), this way avoiding an excessive number of function evaluations and, as a result, the increase of the processing time.

We verified that the proposed method showed itself as being efficient when compared to the evolutionary strategy presented by Diplock[10]. The use of the Representation Formula in the improvement of the initial population was one of the causes of this success.

In the process of calibration of the Gonçalves model, the use of the Representation Formula in the improvement of the initial population, caused the individuals to be positioned closer to the global optimal point location region and, thus, causing a reduction in the number of iterations and, as a consequence, a small number of functions evaluations.

With the results obtained in the tests executed in the calibration of the Gonçalves model, it was possible to determine, in an empiric way, a default set of parameters for the method, which makes it more attractive to be used by professionals of the transportation and planning areas.

Acknowledgement

This work has been partially financed by the project PRO-NEX "Núcleo de Pesquisa em Logística Integrada", from UFSC with funding from CNPq and FAPESC.

Appendix 1 – Proof of the main results.

Proof of the Theorem 4.1:

1. We have

$$\frac{E(yg(\tau, F(y)))}{E(g(\tau, F(y)))} = \frac{E(xg(\tau, F(y)))}{E(g(\tau, F(y)))} + \frac{E((y-x)g(\tau, F(y)))}{E(g(\tau, F(y)))}.$$

Thus,

$$\frac{E(yg(\tau, F(y)))}{E(g(\tau, F(y)))} = x + \frac{E((y-x)g(\tau, F(y)))}{E(g(\tau, F(y)))}. \quad (A1)$$

2. Let $\varepsilon \in (0, \varepsilon_0)$ and $\chi_\varepsilon(y) = 1 - \psi_\varepsilon(y)$. Then

$$\begin{aligned} E((y-x)g(\tau, F(y))) &= \\ &= E((y-x)\chi_\varepsilon(y) + \psi_\varepsilon(y))g(\tau, F(y)) \end{aligned}$$

and we have

$$E((y-x)g(\tau, F(y))) = E((y-x)\chi_\varepsilon(y)g(\tau, F(y))) + E((y-x)\psi_\varepsilon(y)g(\tau, F(y))) .$$

3. Moreover

$$\begin{aligned} |E((y-x)\chi_\varepsilon(y)g(\tau, F(y)))| &\leq \\ &\leq E(|y-x|\chi_\varepsilon(y)g(\tau, F(y))) \leq \varepsilon M \end{aligned}$$

and, from Eq. (20) :

$$E(\psi_\varepsilon(y)g(\tau, F(y))) \leq h_2(\tau, \varepsilon) .$$

Thus,

$$|E((y-x)g(\tau, F(y)))| \leq \varepsilon E(\chi_\varepsilon(y)g(\tau, F(y))) + h_2(\tau, \varepsilon) .$$

and

$$\begin{aligned} \left| \frac{E((y-x)g(\tau, F(y)))}{E(g(\tau, F(y)))} \right| &\leq \frac{\varepsilon E(\chi_\varepsilon(y)g(\tau, F(y))) + h_2(\tau, \varepsilon)}{E(g(\tau, F(y)))} = \\ &= \frac{\varepsilon E(\chi_\varepsilon(y)g(\tau, F(y)))}{E(g(\tau, F(y)))} + \frac{h_2(\tau, \varepsilon)}{E(g(\tau, F(y)))} . \end{aligned}$$

4. Since

$$\frac{E(\chi_\varepsilon(y)g(\tau, F(y)))}{E(g(\tau, F(y)))} \leq 1$$

and

$$\frac{h_2(\tau, \varepsilon)}{E(g(\tau, F(y)))} \leq \frac{h_2(\tau, \varepsilon)}{h_1(\tau, \varepsilon)} ,$$

we have

$$\left| \frac{E((y-x)g(\tau, F(y)))}{E(g(\tau, F(y)))} \right| \leq \varepsilon + \frac{h_2(\tau, \varepsilon)}{h_1(\tau, \varepsilon)}$$

5. This inequality combined to Eq. (21) yield that

$$\forall \varepsilon \in (0, \varepsilon_0) : \lim_{\tau \rightarrow +\infty} \left| \frac{E((y-x)g(\tau, F(y)))}{E(g(\tau, F(y)))} \right| \leq \varepsilon .$$

Thus

$$\lim_{\tau \rightarrow +\infty} \left| \frac{E((y-x)g(\tau, F(y)))}{E(g(\tau, F(y)))} \right| = 0$$

and we have

$$\lim_{\tau \rightarrow +\infty} \frac{E((y-x)g(\tau, F(y)))}{E(g(\tau, F(y)))} = 0 ,$$

6. By taking the limit for $\tau \rightarrow +\infty$ in Eq. (A1), we have

$$\lim_{\tau \rightarrow +\infty} \frac{E(yg(\tau, F(y)))}{E(g(\tau, F(y)))} = x + \lim_{\tau \rightarrow +\infty} \frac{E((y-x)g(\tau, F(y)))}{E(g(\tau, F(y)))} = x$$

and we have the result.

The proof of the proposition 4.2 needs the following auxiliary result:

Lemma A.1. *Let $\varepsilon, \eta > 0$ be such that $B_\eta \neq \emptyset$ and $S_\varepsilon \neq \emptyset$. Then*

$$\theta(\varepsilon) = \left(\min_{y \in S_\varepsilon} \mu(F(y)) \right) - \mu(F(x)) > 0, \quad (A2)$$

and

$$m(\eta) = \left(\max_{y \in B_\eta} \mu(F(y)) \right) - \mu(F(x)) > 0. \quad (A3)$$

Moreover,

$$m(\eta) \xrightarrow{\eta \rightarrow 0+} 0 ; \quad \theta(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0+} 0 ; \quad (A4)$$

If $g(\tau, \xi) = h(\tau\mu(\xi))$, then $\xi \rightarrow g(\tau, \xi)$ and $\tau \rightarrow g(\tau, \xi)$ are strictly positive, decreasing, continuous. Moreover,

$$\max_{S_\varepsilon} g(\tau, F(y)) \leq h(\tau(\theta(\varepsilon) + \mu(F(x)))) ; \quad (A5)$$

$$\min_{B_\eta} g(\tau, F(y)) \geq h(\tau(m(\eta) + \mu(F(x)))) . \quad (A6)$$

Proof of the Lemma A.1:

1. Since $\xi \rightarrow \mu(\xi)$ is strictly increasing and x verifies (8), we have

$$\forall y \in S : \mu(F(y)) \geq \mu(F(x)). \quad (A7)$$

2. (A5) shows that $\theta(\varepsilon) \geq 0$. Assume that $\theta(\varepsilon) = 0$: then there exists a sequence $\{x_n\}_{n \geq 0} \subset S_\varepsilon$ such that

$$\mu(F(x_n)) \xrightarrow{n \rightarrow +\infty} \mu(F(x)) .$$

Since S_ε is closed bounded, this sequence has a cluster point $\bar{x} \in S_\varepsilon$. We have $\bar{x} \neq x$, since $\|\bar{x} - x\| \geq \varepsilon$. Moreover, the continuity of μ implies that $\mu(F(\bar{x})) = \mu(F(x))$. Since $\xi \rightarrow \mu(\xi)$ is strictly increasing, we have $F(\bar{x}) = F(x)$, what is in contradiction with the uniqueness of x . Thus, $\theta(\varepsilon) \geq 0$ and $\theta(\varepsilon) \neq 0$, what establishes (A2).

3. (A6) implies also that $m(\eta) \geq 0$. Assume that $m(\eta) = 0$: then

$$\forall y \in B_\eta : \mu(F(y)) \leq \mu(F(x)).$$

Since $\xi \rightarrow \mu(\xi)$ is strictly increasing, this implies that

$$\forall y \in B_\eta : F(y) \leq F(x)$$

Thus, (8) implies that: $F(y) = F(x)$, $\forall y \in B_\eta$, what is in contradiction with the uniqueness of x . Thus, we have (A3).

4. (A4) is a consequence of the continuity of μ .

5. $\xi \rightarrow g(\tau, \xi)$ and $\tau \rightarrow g(\tau, \xi)$ are strictly positive, decreasing, continuous, since h is continuous, strictly positive and strictly decreasing on $(0, +\infty)$, while μ is continuous, strictly positive and strictly increasing on $(F(x), +\infty)$.

6. Since, for $y \in S_\varepsilon$,

$$\tau\mu(F(y)) = \tau\mu(F(x)) + \tau(\mu(F(y)) - \mu(F(x))) \geq \tau\mu(F(x)) + \tau\theta(\varepsilon)$$

and h is strictly decreasing, we have

$$g(\tau, F(y)) = h(\tau\mu(F(y))) \leq h(\tau(\theta(\varepsilon) + \mu(F(x)))).$$

7. Analogously, for $y \in B_\eta$,

$$\begin{aligned} \tau\mu(F(y)) &= \mu(F(x)) + \tau(\mu(F(y)) - \mu(F(x))) \leq \\ &\leq \tau\mu(F(x)) + \tau m(\eta) \end{aligned}$$

Thus, using that h is strictly decreasing, we have

$$g(\tau, F(y)) = h(\tau\mu(F(y))) \geq h(\tau(m(\eta) + \mu(F(x)))).$$

Proof of the Proposition 4.2:

1. Let $\varepsilon > 0$ be such that $S_\varepsilon \neq \emptyset$. From (A5):

$$E(\psi_\varepsilon(y)g(\tau, F(y))) \leq h(\tau(\theta(\varepsilon) + \mu(F(x))))E(\psi_\varepsilon(y)).$$

Since $E(\psi_\varepsilon(y)) = P(S_\varepsilon)$, we have

$$E(\psi_\varepsilon(y)g(\tau, F(y))) \leq h_2(\tau, \varepsilon) = h(\tau(\theta(\varepsilon) + \mu(F(x))))P(S_\varepsilon).$$

2. Let $\eta > 0$ be such that $B_\eta \neq \emptyset$. Since $g > 0$, we have

$$E(g(\tau, F(y))) \geq E(\chi_\eta(y)g(\tau, F(y))).$$

Thus, (A6) implies that

$$E(g(\tau, F(y))) \geq h(\tau(m(\eta) + \mu(F(x))))E(\chi_\eta(y)).$$

Since $E(\chi_\eta(y)) = P(B_\eta)$, we have

$$E(g(\tau, F(y))) \geq h_1(\tau, \varepsilon) = h(\tau(m(\eta) + \mu(F(x))))P(B_\eta) > 0$$

3. From (A2) and (A4), there exists $\eta > 0$ such that $0 < m(\eta) < \theta(\varepsilon)$. Thus, (23) applied to $\xi = m(\eta) + \mu(F(x))$ and $\delta = \theta(\varepsilon) - m(\eta)$ yields that

$$\frac{h_2(\tau, \varepsilon)}{h_1(\tau, \varepsilon)} = \frac{P(S_\eta) h(\tau(\theta(\varepsilon) + \mu(F(x))))}{P(B_\eta) h(\tau(m(\eta) + \mu(F(x))))} \xrightarrow{\tau \rightarrow +\infty} 0.$$

4. Thus, we have (20) and (21) and all the requirements of the theorem 4.1 are fulfilled, what establishes the result.

References

1. L. M. W. Almeida, *Desenvolvimento de uma Metodologia para Análise Locacional de Sistemas Educacionais Usando Modelos de Interação Espacial e Indicadores de Acessibilidade*. Ph. D. Thesis (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, Brazil, (1999).
2. L.M.W. Almeida, M.B. Gonçalves, *A methodology to incorporate behavioral aspects in trip distribution models with and application to estimate student flow*, Environment and Planning A, **33**(6), 1125-1138, (2001).
3. M. J. Baxter, G. O. Ewing, *Calibration of Production – Constrained Trip Distribution Models and The Effect of Intervening Opportunities*. Journal Regional Science; **19**(3), 319-330, (1979).
4. M. S. Bazaraa, H. D. Sherali, C. M. Shetty, *Nonlinear Programming – theory and algorithms*. 2. ed. New York: John Wiley & Sons (1993).
5. M. Ben-Akiva, S.R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*, the MIT Press, Cambridge, Massachusetts (1985).
6. E. T. Bez, *Procedimento de representação de soluções em otimização global: aplicação em modelos de interação espacial*. Ph. D. Thesis (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, Brazil (2005).
7. E. T. Bez, J. E. Souza de Cursi, M. B. Gonçalves, *Procedimento de Representação em Otimização Global*. In: XXVI Congresso Nacional de Matemática Aplicada e Computacional, 2003, São José do Rio Preto. Resumos das Comunicações do XXVI CNMAC. Rio de Janeiro: SBMAC **1**, 545-545, (2003).
8. E. T. Bez, J. E. Souza de Cursi, M.B. Gonçalves, *A Hybrid Method for Continuous Global Optimization Involving the Representation of the Solution*. 6th World Congress on Structural and Multidisciplinary Optimization – WCSMO6. Rio de Janeiro, RJ, Brazil (2005).
9. M. Bouhadi, R. Ellaia, J.E. Souza de Cursi, *Stochastic perturbation methods for affine restrictions*, Hadjisavvas, Nicolas (ed.) et al., Advances in convex analysis and global optimization. Honoring the memory of C. Caratheodory (1873-1950). Dordrecht: Kluwer Academic Publishers. Nonconvex Optim. Appl., **54**, 487-499, (2001).
10. G. Diplock, S. Openshaw, *Using Simple Genetic Algorithms to Calibrate Spatial Interaction Models*. Geographical Analysis, **28**(3), 262-279, (1996).
11. R. Ellaia, E. Elmouatasim, J.E. Souza de Cursi, *Random Perturbations of variable metric descent in unconstrained nonsmooth nonconvex perturbation*, International Journal of Applied Mathematics and Computer Science, **16**(4), (2006).
12. A.S. Fotheringham, *A new set of Spatial Interaction Models: The theory of competing destinations*. Environment and Planning A, **15**, 15-36, (1983).
13. A.S. Fotheringham, *Spatial Flows and Spatial Patterns*. Environment and Planning A, **16**, 529-543, (1984).
14. A.S. Fotheringham, *Spatial Competition and Agglomeration in Urban Modelling*. Environment and Planning A, **17**, 213-230, (1985).
15. J. P. Gitlesen, I. Thorsen, J. Ubøe, *Misspecifications due to aggregation of data in models for journeys-to-work*. Discussion Paper 13. NHH: Department of Finance and Management Science (2004).

16. M.B. Gonçalves, *Desenvolvimento e Teste de um Novo Modelo Gravitacional – de Oportunidades de Distribuição de Viagens*. Ph. D. Thesis (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, Brazil (1992).
17. M.B. Gonçalves, I. Ulysséa Neto, *The Development of a new Gravity – Opportunity Model for Trip Distribution*. *Environment and Planning*, **25**, 817-826, (1993).
18. M.B. Gonçalves, J.E. Souza de Cursi, *Métodos Robustos para a Calibração de Modelos de Interação Espacial em Transportes*. Associação Nacional de Pesquisa e Ensino em Transportes, 11., 1997. *Anais...* 1997; **2**, 303-313, (1997).
19. M.B. Gonçalves, J.E. Souza de Cursi, *Parameter Estimation in a Trip Distribution Model by Random Perturbation of a Descent Method*. *Transportation Research*, **35B**, 137-161, (2001).
20. A.O. Griewank, *Generalized descent for global optimization*, *Journal of optimization theory and applications*, **34**(1), 11-39, (1981).
21. A.J. Horowitz, D.D. Farmer, *A Critical Review of Statewide Travel Forecasting Practice*, University of Wisconsin (<http://my.execpc.com/~ajh/Statewid.pdf>) (1998).
22. N. Kühlkamp, *Modelo de oportunidades intervenientes, de distribuição de viagens, com ponderação das posições espaciais relativas das oportunidades*. Ph. D. Thesis (Doutorado em Engenharia Civil) - Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, Brazil (2003).
23. J.M. Lowe, A. Sen, *Gravity model applications in health planning: Analysis of an urban hospital market*. *Journal of Regional Science*, **36**(3), 437-461, (1996).
24. J.D. Ortuzar, L.G. Willumsen, *Modelling Transport*. John Wiley & Sons, New York (1994).
25. M. Pincus, *A closed formula solution of certain programming problems*, *Operations Research*, **16**(3), 690-694, (1968).
26. M. Pogu, J.E. Souza de Cursi, *Global Optimization by Random Perturbation of the Gradient Method with a Fixed Parameter*. *Journal of Global Optimization*, **5**, 159-180, (1994).
27. M. Schneider, *Gravity models and trip distribution theory*. *Papers and Proceedings of the Regional Science Association*, **5**, 51-56, (1959).
28. Y.H. Shi, R.C. Eberhart, *Empirical study of particle swarm optimization*, *Congress on Evolutionary Computation*. Washington DC, USA, 1945-1950 (1999)
29. H.P. Schwefel, *Evolution and Optimization Seeking*. New York: Wiley (1995).
30. E. Sheppard, *Theoretical underpinnings of the gravity hypothesis*, *Geographical Analysis*, **10**(4), 386-402, (1978).
31. S. Soot, A. Sen, *A spatial employment and economic development model*. *Papers in Regional Science*, **70**(2), 149-166, (1991).
32. J.E. Souza de Cursi, *Une Formule de Représentation pour le point d'optimum global d'une fonctionnelle régulière en dimension finie*. *Note de Recherche 061/02, LMR, Rouen, France* (2002).
33. J.E. Souza de Cursi, *Representation and numerical determination of the global optimizer of a continuous function on a bounded domain*, Floudas, Christodoulos A. (ed.) et al., *Frontiers in global optimization*. Boston, MA: Kluwer Academic Publishers. *Nonconvex Optim. Appl.* **74**, 517-539, (2004).
34. J.E. Souza de Cursi, R. Ellaia, M. Bouhadi, *Global Optimization under nonlinear restrictions by using stochastic perturbations of the projected gradient*, Floudas, Christodoulos A. (ed.) et al., *Frontiers in global optimization*. Boston, MA: Kluwer Academic Publishers. *Nonconvex Optim. Appl.* **74**, 541-561, (2004).
35. M.J. Wills, *A Flexible Gravity-Opportunities Model for Trip Distribution*. *Transportation Research*, **20B**, 89-111, (1986).
36. A.G. Wilson, *A Statistical Theory of Spatial Distribution Models*. *Transportation Research*, **1**, 253-269, (1967).
37. A.G. Wilson *Entropy in Urban and Regional Modelling*. London: Pion (1970).
38. S.C. Wirasinghe, A.S. Kumarage, *An aggregate demand model for intercity passenger travel in Sri Lanka*, *Transportation*, **25**, 77-98, (1998).
39. F. Zhao, L.F. Chow, M.T. Li, A. Gan, *Refinement of Fsutms Trip Distribution Methodology*. Final Report, Prepared by Lehman Center for Transportation Research, Department of Civil & Environmental Engineering, Florida International University (2004).