

Um Procedimento Numérico para a Otimização Global Baseado em uma Representação da Solução¹

E.T. BEZ², Centro de Educação Superior de São José, UNIVALI, 88122-000 São José, SC, Brasil.

J.E. SOUZA DE CURSI, Laboratoire de Mécanique de Rouen, INSA de Rouen, 76800 Saint Etienne du Rouvray, France.

M.B. GONÇALVES, Departamento de Matemática, Centro de Ciências Físicas e Matemáticas, Departamento de Matemática, UFSC, 88010-970 Florianópolis, SC, Brasil.

Resumo. Na Engenharia nos deparamos com diversas situações, onde a busca de uma solução gera a necessidade de resolução de problemas de otimização global consistindo em determinar o mínimo de uma função contínua f em um conjunto admissível S : $x^* = \text{ArgMin}_S f$. Quando f é unimodal e S convexo, vários métodos estão disponíveis na literatura. Entretanto, no caso multimodal, a não convexidade do problema exige métodos mais robustos. Focamos nosso estudo no desenvolvimento de um método híbrido, composto por um algoritmo de tipo evolucionário, cuja população inicial é gerada por uma Fórmula de Representação, e utilizando um método do gradiente adicionado de perturbações aleatórias, em sua fase de mutação. Supomos dado um operador de projeção *proj* transformando pontos externos a S em pontos de S . Os resultados numéricos mostram que a utilização da Fórmula de Representação acelera significativamente o processo de busca.

1. Introdução

A busca de métodos numéricos eficientes para problemas de otimização não convexos tem interessado pesquisadores de todo o mundo. Na linha de tais trabalhos, consideramos uma função f , contínua a valores reais; um conjunto admissível $S \subset R^n$ fechado e limitado; e o problema seguinte:

$$x^* = \text{ArgMin}_S f. \quad (1.1)$$

Atualmente, um dos caminhos para a resolução numérica da Equação (1.1) passa pela combinação de métodos estocásticos e deterministas. Calcados nesta idéia, desenvolvemos um método híbrido de otimização. A essência do método consiste em

¹Agradecimento a CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

²bolsista de Estágio de doutorado CAPES

combinar uma Fórmula de Representação do ponto de ótimo global x^* e um algoritmo de tipo genético utilizando em sua fase de mutação, um método de descida. A Fórmula de Representação, desenvolvida por [9], é dada por:

$$x^* = \lim_{\lambda \rightarrow +\infty} \frac{E(xg(\lambda, f(x)))}{E(g(\lambda, f(x)))}, \quad (1.2)$$

onde g é uma função convenientemente escolhida, $f(x)$ é a função que se pretende minimizar, $\lambda \in \mathfrak{R}^+$ e $E(xg(\lambda, f(x)))$ é a esperança matemática que determina o valor médio de uma distribuição de probabilidade, considerando o valor e a probabilidade de ocorrência da variável. [9] demonstra que uma possível escolha para g pode ser dada por:

$$g(\lambda, f(x)) = e^{-\lambda f(x)}. \quad (1.3)$$

A utilização prática da equação (1.2) exige a construção de uma distribuição de probabilidades sobre o conjunto viável S , por exemplo, utilizando um operador de projecção *proj* transformando $x \notin S$ em $x_{proj} \in S$. O desenvolvimento de um método mais geral, incluindo o tratamento numérico de restrições faz parte dos trabalhos em andamento.

Na Seção 2 descrevemos o procedimento numérico e, na Seção 3, são apresentadas as funções testes. Na Seção 4 estão os testes numéricos e alguns resultados e, finalmente, as conclusões são apresentadas na Seção 5.

2. Um Procedimento Numérico Utilizando a Representação

No desenvolvimento do procedimento, fizemos uso da equação (1.2) na geração da população inicial de um algoritmo evolucionário ([1] e [3]). O principal objetivo do uso desta expressão, foi obter uma melhoria na geração dos pontos iniciais. Como já observado, utiliza-se um operador de projecção *proj* para transformar pontos não viáveis em pontos viáveis. A implementação numérica da equação (1.2) pode ser feita trocando o limite $\lambda \rightarrow +\infty$ por um valor fixo de λ : gera-se uma seqüência (x_1, \dots, x_{ntirm}) de pontos viáveis, de acordo com uma probabilidade P , e usa-se esta seqüência para aproximar:

$$x^* \approx x_e^* = \frac{\sum_{i=1}^{ntirm} x_i g(\lambda, f(x_i))}{\sum_{i=1}^{ntirm} g(\lambda, f(x_i))}. \quad (2.1)$$

A equação (2.1) é aplicada na determinação de cada elemento da população inicial.

Na fase de mutação foi usado o método do gradiente adicionado de perturbações aleatórias ([4] e [5]) que, em sua forma original, tem como processo de busca do ponto de mínimo o seguinte procedimento:

Escolhe-se um ponto inicial x_0 . Em seguida, determina-se uma seqüência de pontos

(x_n) através da relação de recorrência, $x_{n+1} = Q(x_n)$, onde, $Q(x) = x - \mu \nabla f(x)$, com $\mu > 0$ e $\nabla f(x)$ é o gradiente de f .

Para f não convexa, a seqüência pode convergir para um mínimo local e não para um mínimo global, ou seja, não verifica a condição que $f(x_n) \rightarrow \min f$. É possível evitar este fenômeno adicionando um termo estocástico adequado à relação de recorrência:

$$x_{n+1} = Q(x_n) + P_n,$$

onde P_n é uma perturbação aleatória que decresce de forma lenta de modo a evitar a convergência para um mínimo local. A convergência do método, independentemente do ponto inicial foi estabelecida em [7].

A implementação deste método pode ser feita seguindo os passos do seguinte algoritmo ([4] e [5]):

Escolhe-se um número inteiro positivo NR e gera-se uma seqüência $\{x_n\}_{n \geq 1}$ de forma que, no passo n , x_{n+1} é obtido através de

$$x_{n+1} = \text{ArgMin}\{f(Q(x_n) + P_n^r); 0 \leq r \leq NR\},$$

onde $P_n^0 = 0$ e $P_n^1, P_n^2, \dots, P_n^{NR}$ são NR valores aleatórios de P_n .

Para cada r , $1 \leq r \leq NR$, o vetor P_n^r é dado por $P_n^r = \phi_n Z_n^r$, sendo $(\phi_n)_{n \geq 1}$ uma seqüência de números reais convergindo para zero gerada pela expressão:

$$\phi_n = a \frac{c}{\sqrt{\log(n+d)}}, \quad (2.2)$$

em que a , c e d são constantes positivas e (Z_n^r) é um vetor gerado através da distribuição gaussiana padronizada $N(0, Id)$, onde Id é uma matriz identidade.

Este método pode ser implementado com o parâmetro μ fixo, ou seja, o tamanho do passo do gradiente é predefinido no início do processo e permanece o mesmo até o fim. No entanto, se fixamos um valor muito pequeno para μ , o método converge para o mínimo de maneira muito lenta. Caso optemos por um valor alto, temos uma convergência mais rápida porém, o algoritmo pode ficar oscilando próximo do mínimo.

Uma forma mais adequada de trabalhar com o método do gradiente, é fazer com que o tamanho do passo seja convenientemente definido a cada iteração, ou seja, busca-se o passo ótimo. Este, denotado por t_n , é definido de modo que seja válida a relação $f(x_n - t_n \nabla f(x_n)) \leq f(x_n - t \nabla f(x_n))$, $\forall t \in (0, t_{max})$ sendo t_{max} determinado através da razão $\frac{\alpha}{\|\nabla f(x_n)\|}$, onde α delimita o tamanho inicial do passo.

A implementação do método proposto, pode ser descrita da seguinte maneira:

1) Dado um número real $H > 0$ e números inteiros não negativos NP (tamanho da população), NF (número de filhos) e NR ; uma população inicial $S_0 = \{x_0^1, \dots, x_0^{NP}\}$, onde cada elemento de S_0 é obtido através da Fórmula de Representação de ótimo global x^* , dada pela equação (2.1).

2) Gera-se a população S_{n+1} , a partir de S_n , fazendo:

• **Reprodução:** Determina-se o conjunto

$$F_n = \{y_n^i = \alpha_n^i x_n^j + \beta_n^i x_n^k + \gamma_n^i : x_n^j, x_n^k \in S_n : i = 1, \dots, NF\},$$

onde α_n^i, β_n^i e γ_n^i são valores aleatórios gerados através da distribuição uniforme sobre $[-H, H]$ e j, k são gerados aleatoriamente.

F_n contém $NF = (2 * NP)$ elementos.

Denotamos $B_n = S_n \cup F_n$. B_n contém $NB = (NP + NF)$ elementos.

• *Mutação*: Determina-se o conjunto

$$M_n = \{xp_n^j = \text{ArgMin } f(Q(x_n^j) + P_n^{j,i}) : x_n^j \in B_n : i = 0, \dots, NR; j = 1, \dots, NB\},$$

M_n tem $NM = (2 * NB)$ elementos. $P_n^{j,0} = 0$ e $P_n^{j,1}, P_n^{j,2}, \dots, P_n^{j,NR}$ são NR valores aleatórios de P_n .

• *Seleção (Elitismo)*:

Ordena-se os elementos de $A_n = B_n \cup M_n = S_n \cup F_n \cup M_n$ em ordem crescente do valor da função. A população S_{n+1} é formada pelos primeiros NP elementos de A_n .

O avanço esperado do algoritmo proposto, em relação aos trabalhos anteriores ([4], [5], [7] e [9]), refere-se à melhoria da população inicial, pelo uso da Fórmula de Representação e pela melhoria na convergência com a introdução do passo ótimo ([2] e [9]) no método do gradiente. É importante, então, investigar como a introdução destes influencia na performance do algoritmo proposto.

3. Funções Testes

A literatura apresenta conjuntos de funções testes, cujo ótimo global é difícil de ser alcançado usando algoritmos determinísticos. Tais funções, caracterizadas pela multimodalidade, são usadas para avaliar a performance de novos algoritmos. Nesse trabalho optamos por escolher as funções, levando em consideração diferenças entre a distância de seus pontos de mínimos locais e seus valores de mínimos locais como observado em funções como Griewank [6], Davis [9] e Rastrigin [9] e, também, pela pouca diferenciação dos valores da função numa região ampla (superfície quase plana) como é o caso da função de Rosenbrock [8].

A Função de Davis [9] testada é dada por:

$$f(x) = 0.5 + \frac{\sin^2(\|x - \bar{x}\|_2) - 0.5}{(1 + \|x - \bar{x}\|_2^2 / 1000)^2}.$$

Esta função possui o valor de mínimo global em $f^* = 0$.

Outra função utilizada foi a Função de Rastrigin [9], descrita pela expressão:

$$f(x) = 3m + \sum_{i=1}^m [(x_i - \bar{x}_i)^2 - 3 \cos(2\pi(x_i - \bar{x}_i))].$$

O valor de mínimo desta função, também, é dado por $f^* = 0$ e o ponto de ótimo global é dado por $x^* = (0, 0, 0, \dots, 0)$, onde x^* é um vetor de dimensão m .

As demais funções testadas foram a Função de Griewank [6], expressa por:

$$f(x) = \frac{1}{200} \|x - \bar{x}\|_2^2 - \prod_{i=1}^m \cos\left(\frac{x_i - \bar{x}_i}{\sqrt{i}}\right),$$

com mínimo global em $f^* = -1$ e, por último, a Função de Rosenbrock [8], dada pela expressão:

$$f(x) = 100 \sum_{i=1}^m (x_i^2 - x_i)^2 + \sum_{i=1}^m (x_i - 1)^2,$$

com valor de mínimo global em $f^* = 0$ e ponto de ótimo global dado por $x^* = (1, 1, 1, \dots, 1)$, onde x^* é um vetor de dimensão m .

Fizemos uso de um vetor descentrado dado por $\bar{x} = (1, 2, \dots, m)$, nas funções de Davis, Rastrigin e Griewank. O ponto de ótimo global, com o uso deste vetor é dado por $x^* = \bar{x}$.

4. Testes Numéricos

Nesta seção, apresentamos os ensaios numéricos realizados para avaliar a introdução da Fórmula de Representação na performance do algoritmo proposto. Cabe salientar que realizamos uma extensa bateria de testes avaliando a influência dos diversos parâmetros envolvidos. Em virtude da limitação de páginas, mostramos apenas alguns resultados. Antes, porém, apresentaremos algumas considerações referentes a procedimentos adotados na implementação do método.

Na Fórmula de Representação, os pontos aleatórios gerados seguem uma distribuição normal dada por $N(0, \sigma)$.

Na definição das perturbações aleatórias P_n , atribuímos às constantes a e d o valor 1 para a Equação (2.2), e para a distribuição uniforme, atribuímos $H = 1$. A região de busca foi limitada por um disco, de raio r , onde pontos gerados fora da região de busca são projetados para o interior da mesma, através da expressão:

$$x_{proj} = \varphi * r * \frac{x_i}{\|x\|_2},$$

onde $\|x\|_2 = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}}$ e φ é gerado através de uma distribuição uniforme no intervalo $(0,1)$. A supressão do termo φ projeta o ponto para a fronteira.

Foram feitos testes com as 4 funções listadas. Para cada situação o algoritmo foi rodado 100 vezes (número de observações igual a 100), em razão da população inicial ser gerada aleatoriamente.

Na Tabela 1, apresentamos alguns resultados que demonstram o desempenho da Fórmula de Representação. Para estes testes geramos os pontos iniciais, através da função de distribuição normal $N(0,1)$. Os pontos gerados foram projetados para a fronteira e a partir daí duas situações foram criadas. Na primeira, consideramos o método, sem o uso da Fórmula de Representação e variamos o raio do disco responsável pela determinação da região de busca. Desta forma, verificamos a eficiência do método na busca do ótimo global da função. Na segunda situação, fizemos uso da Fórmula de Representação (1.2) para verificar a melhoria da performance do algoritmo.

Uma comparação das duas situações e da influência da Fórmula de Representação é verificada pelo número de iterações (ciclo completo do algoritmo genético) necessárias para determinação do ótimo global.

Considerando que o mínimo global das funções de Rastrigin e Davis é $f^* = 0$, estabelecemos como critério de parada $f \leq 0.001$. Utilizamos $NP = 5$, $m = 5$, Número de pontos gerados para a equação (1.2) ($ntirm$) = 100, $\lambda = 10$, Número de iterações (NIT) = 30, Número de passos do gradiente (NPG) = 10, $NR = 30$, Número de intervalos do passo ótimo ($NIPO$) = 100, $\alpha = 1$ e $c = 0.1$. Fizemos uso de $N(0, 1)$ na geração dos números aleatórios utilizados nas perturbações aleatórias.

Os resultados, demonstraram um grande desempenho da Fórmula de Representação, que pode ser verificado se observarmos a expressão que determina o número total de avaliações da função, dado por:

$$NP + NIT(NF + NB * NPG(NR + NIPO + 1)).$$

	Val. Min.	Nmiter	Nmaval	Val. Min.	Nmiter	Nmaval
($r = 10$)	Rast.	Rast.	Rast.	Davis	Davis	Davis
s/ eq. (1.2)	4,92E-06	8,22	161610.2	7,73E-08	1,72	33820.2
c/ eq. (1.2)	3,12E-07	1	19665	9,49E-08	1	19655
($r = 100$)	Rast.	Rast.	Rast.	Davis	Davis	Davis
s/ eq. (1.2)	5,41E-07	8,7	171047	9,17E-09	2,8	55053
c/ eq. (1.2)	3,75E-08	1	19665	5,60E-10	1,11	21827.6

Tabela 1: Número médio de iterações com ou sem o uso da equação (1.2).

Se tomarmos, por exemplo, o número médio de iterações (Nmiter) apresentado na Tabela 1 para a função de Davis, com $r = 10$, aparentemente não houve melhoria. No entanto, quando observamos o número médio de avaliações (Nmaval) da função, obtemos 33820.2 avaliações sem a equação (1.2) e 19655 com a equação (1.2), determinando uma diferença significativa.

Testamos a influência da variação do desvio padrão da distribuição normal usada na equação (1.2). Na Figura 1, apresentamos os resultados obtidos para a função de Rastrigin com dimensão $m = 8$. Nesse ensaio, considerando que o mínimo global desta função é $f^* = 0$, estabelecemos como critério de parada $f \leq 0.001$. Utilizamos $NP = 5$, $ntirm = 100$, $\lambda = 10$, $NIT = 50$, $NPG = 10$, $NR = 30$, $NIPO = 100$, $\alpha = 1$ e $c = 0.1$ e a região de busca é um disco com $r = 10$. Fizemos uso de $N(0, 0.5)$ na geração dos números aleatórios utilizados nas perturbações aleatórias.

Observando a Figura 1, percebe-se a queda na probabilidade de sucesso quando aumentamos o valor do desvio padrão. A partir do valor de $\sigma = 2$ esta oscilação continua, porém, com menor intensidade.

Outro parâmetro analisado em nossos testes foi a influência da variação de α utilizado na determinação de t_{max} , extremo do intervalo do passo ótimo. Na Tabela 2 e Figura 2, apresentamos alguns resultados obtidos com relação a probabilidade

de sucesso e ao número de iterações necessárias para obtenção do mínimo global da função de Davis descentrada. Verificou-se que a partir de $\alpha = 3.5$, a probabilidade de sucesso é total.

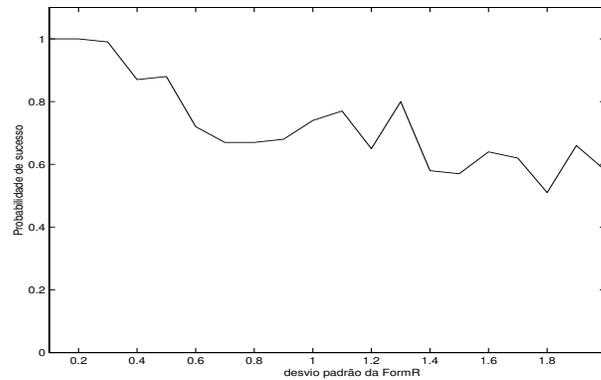


Figura 1: Função Rastrigin – Influência do desvio padrão da equação (1.2) ($m = 8$).

Considerando que o mínimo global da função de Davis é $f^* = 0$, estabelecemos como critério de parada $f \leq 0.001$. Utilizamos $NP = 5$, $ntirm = 50$, $\lambda = 10$, $NIT = 30$, $NPG = 10$, $NR = 30$, $NIPO = 100$ e $c = 0.1$ e a região de busca é um disco com $r = 100$. Fizemos uso de $N(0,1)$ na geração dos números aleatórios utilizados nas perturbações aleatórias e na equação (1.2).

α	Número médio de iterações	Prob. de sucesso
3	30	0,00
3.1	9,09	0,99
3.2	13,73	0,95
3.5	10,86	1,00
25	1,72	1,00
50	1,91	1,00
100	2,94	1,00

Tabela 2: Função Davis Descentrada – Valores da influência do parâmetro α .

Verificamos a probabilidade de sucesso na busca do ótimo quando variamos a dimensão e os testes apresentaram um bom desempenho do método. Algumas funções como Rastrigin e Rosenbrock, a partir de uma certa dimensão, começam a diminuir sua probabilidade de sucesso até torná-la nula. Entretanto, para Davis e Griewank, com ou sem a utilização do vetor utilizado para descentrar a função, a probabilidade de sucesso é total, independente da dimensão utilizada.

Na Figura 3, apresentamos a probabilidade de sucesso na busca do mínimo global da função de Rosenbrock. O mínimo global desta função é $f^* = 0$, o critério de parada foi estabelecido para o caso em que o valor de $f \leq 0.01$. Os testes

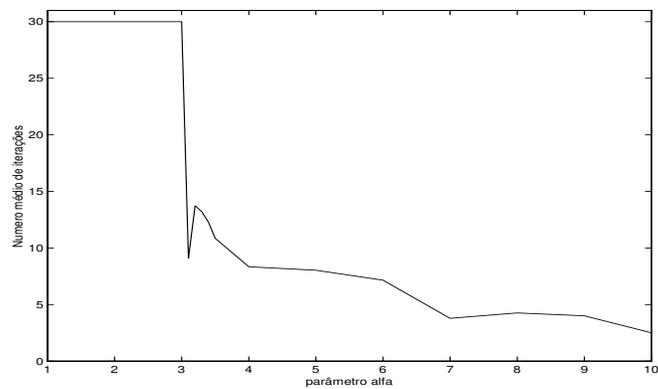


Figura 2: Função Davis Descentrada – Influência de α no número de iterações.

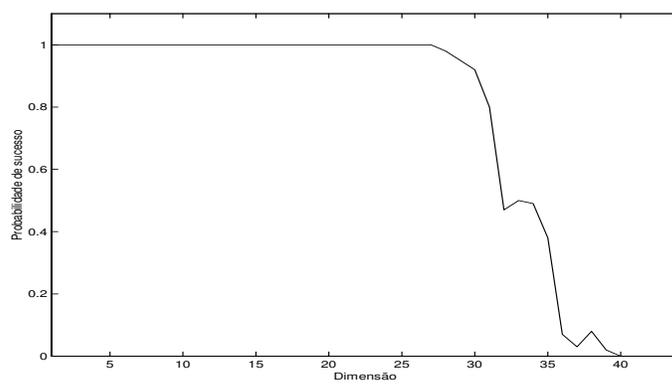


Figura 3: Função Rosenbrock – Influência da Dimensão.

demonstraram que a partir da dimensão 28 a probabilidade de sucesso começa a cair e com dimensão a partir de 40, esta probabilidade torna-se nula. Utilizamos $NP = 5$, $ntirm = 200$, $\lambda = 5$, $NIT = 100$, $NPG = 10$, $NR = 50$, $NIPO = 50$, $\alpha = 0,5$ e $c = 0.1$ e a região de busca é um disco com $r = 10$. Fizemos uso de $N(0, 0.3)$ e $N(0, 0.5)$ na geração dos números aleatórios utilizados nas perturbações aleatórias e na equação (1.2), respectivamente.

5. Conclusões

A proposta deste trabalho, de desenvolver um método híbrido, teve como principal objetivo, avaliar a influência da Fórmula de Representação na melhoria da geração da população inicial do algoritmo evolucionário.

Por intermédio dos inúmeros testes realizados, verificamos o desempenho do método com a variação dos vários parâmetros envolvidos e observamos que o uso da Fórmula de Representação, na melhoria da população inicial, tornou-o mais robusto.

O gradiente com perturbações aleatórias, aplicado à fase de mutação, mostrou-se eficiente, tendo em vista a característica não convexa das funções testadas e levando em consideração que, mesmo em testes com pontos gerados na fronteira, sem o auxílio da Fórmula de Representação, o ótimo era alcançado. Os resultados demonstraram um desempenho superior do método quando fizemos uso da Fórmula de Representação, ficando evidente sua excelente contribuição.

Os testes foram muito encorajadores e nos levam a projetar para trabalhos futuros, a determinação dos tipos de funções que podemos escolher como g , tendo em vista que ela não é única. Cabe-nos também, verificar que outros métodos numéricos podem ser utilizados na fase de mutação e adicionar aos testes outras funções com características diferentes, como por exemplo, a função de Schwefel.

Este método foi aplicado em um problema de distribuição de viagens e observou-se uma ótima performance do mesmo, demonstrado não somente por sua velocidade de convergência, bem como pela pouca sensibilidade dos parâmetros envolvidos [2], o que vem a facilitar sua utilização por profissionais da área de transportes.

Nossos trabalhos em desenvolvimento abordam extensões deste método para problemas de otimização com restrições.

Agradecimentos

Gostaríamos de agradecer ao LMR - Laboratoire de Mécanique de Rouen, INSA de Rouen, pelo espaço e equipamentos disponibilizados para realização deste trabalho.

Abstract. The engineering faces several situations in which the search for solutions creates the need of a resolution of problems of a global optimization consisting on establishing the minimum of a continuous function f in an admissible set S : $x^* = \text{ArgMin}_S f$. Whether f is unimodal and S convex, several methods to solve them are available in the literature. However, in the multimodal case, the non convexity of the problem demands for more robust methods. The study is focused on the development of an hybrid method, composed by an evolutionary algorithm, whose initial population is generated by a Representation Formula, and using a gradient method added of random perturbation on its mutation phase. Considering

given an operator of projection $proj$ transforming points outwards S in points of S . The obtained numerical results showed that the utilization of the Representation Formula accelerate considerably the searching process.

Referências

- [1] H.J.C. Barbosa, C.C. Lavor e F.M.P. Raupp, Computational Experiments with an adaptive genetic algorithm for global minimization of potential energy functions, *Frontiers in Global Optimization*, (C.A. Floudas and P.M. Pardalos Eds.), Kluwer Academic Publishers, Vol. 74, (2003), pp. 71-82.
- [2] E.T. Bez, M.B. Gonçalves e J.E. Souza de Cursi, Uso de um procedimento híbrido de otimização global na determinação dos parâmetros de um modelo gravitacional de oportunidades, por aparecer em “Proceedings do XIII PANAM”, Albany, Nova York, USA, 2004.
- [3] D.E. Goldberg, “Genetic Algorithms in Search, Optimization and Machine Learning”, Addison-Wesley Publishing Company, Inc., 1989.
- [4] M.B. Gonçalves e J.E. Souza de Cursi, Métodos robustos para a calibração de modelos de interação espacial em transportes, em “Anais do XI ANPET”, Vol. 2, pp. 303-313, 1997.
- [5] M.B. Gonçalves e J.E. Souza de Cursi, Parameter estimation in a trip distribution model by random perturbation of a descent method, *Transportation Research B*, **35** (2001), 137-161.
- [6] A.O. Griewank, Generalized descent for global optimization, *Journal of optimization theory and applications*, **34**, No. 1 (1981), 11-39.
- [7] M. Pogu e J.E. Souza de Cursi, Global optimization by random perturbation of the gradient method with a fixed parameter, *Journal of Global Optimization*, **5** (1994), 159-180.
- [8] Y.H. Shi e R.C. Eberhart, Empirical study of particle swarm optimization, *Congress on Evolutionary Computation*, Washington DC, USA (1999) 1945-1950.
- [9] J.E. Souza de Cursi, Representation and numerical determination of the global optimizer of a continuous function on a bounded domain, *Frontiers in Global Optimization*, (C. A. Floudas and P. M. Pardalos, Eds.), Kluwer Academic Publishers, Vol. 74, (2003), pp. 517-540.