

08 a 11 de Outubro de 2018  
Instituto Federal Fluminense  
Búzios - RJ

## COMPARAÇÃO ENTRE MÉTODOS DETERMINÍSTICOS E ESTOCÁSTICOS NA RESOLUÇÃO DE EQUAÇÕES NÃO LINEARES

**Gabriel Pereira de Araujo**<sup>1</sup> - likeahurricane@hotmail.com

**Pedro Gonçalves de Oliveira Neto**<sup>1</sup> - pedrogoncalves001@gmail.com

**Renata Barbosa da Silva**<sup>1</sup> - renatabs97@yahoo.com

<sup>1</sup>Universidade Federal Fluminense, INFES - Santo Antônio de Pádua, RJ, Brazil

**Resumo.** *Este trabalho consiste numa análise comparativa entre alguns métodos estocásticos e determinísticos, empregados na resolução de equações não lineares. Os métodos determinísticos utilizados foram: gradiente, método de Newton, DFP, BFGS e Nelder Mead. Os estocásticos: DE (Evolução Diferencial), PSO (Otimização por Enxame de Partículas) e Luss-Jaakola. A partir dos resultados alcançados, pôde-se verificar que os métodos determinísticos apresentaram uma melhor convergência em relação aos estocásticos por exigirem de menos esforço computacional em sua convergência, entre eles destaca-se o método de Newton, convergindo com menos laços que os demais. Constatou-se que os métodos estocásticos demandam de mais tempo para convergir, entretanto todos os métodos obtiveram resultados aproximados.*

**Palavras-chave:** *Equações não lineares, Comparação, Métodos determinísticos, Métodos estocásticos, Otimização*

### 1. INTRODUÇÃO

Um problema de otimização consiste, basicamente, em determinar os valores ótimos (máximos ou mínimos) de uma certa função, denominada função objetivo. Esta função objetivo depende de um conjunto de variáveis as quais são denominadas de variáveis de decisão ou de controle. A teoria da otimização desenvolve métodos determinísticos e estocásticos, para a obtenção dos valores ótimos das variáveis de decisão dos problemas de otimização (Viana, 2017). Valores estes que vão determinar a imagem ótima (máxima ou mínima) da função objetivo. As técnicas de otimização são amplamente utilizadas na resolução de problemas das mais diversas áreas, tais como: administração, biologia, economia, engenharia, física, logística, entre outras. Este trabalho consiste em apresentar e analisar, alguns métodos de otimização determinísticos e estocásticos para resolução de funções não lineares. Os resultados alcançados permitem realizar uma análise comparativa dos métodos, e desta forma verificar a eficiência, desempenho e convergência dos mesmos.

## 2. MÉTODOS DETERMINÍSTICOS

Os métodos de otimização baseados nos algoritmos determinísticos geram uma seqüência de possíveis soluções requerendo, na maioria das vezes, o uso ao menos da primeira derivada da função objetivo em relação às variáveis de projeto.

Nestes métodos, a função objetivo e as restrições são dadas como funções matemáticas ou relações funcionais. Além disso, a função objetivo deve ser contínua e diferenciável no espaço de busca.

Nos métodos determinísticos Gradiente, Newton, DFP e BFGS utilizou-se o critério de Armijo.

### 2.1 Método do Gradiente

O método de busca por gradiente ou método do declive máximo, tem por objetivo encontrar uma raiz para a equação  $\nabla f(x_k) = 0$  sendo,  $f$  uma função real de  $n$  variáveis reais, continuamente diferenciável. Dada uma aproximação  $x_k$  para um mínimo de  $f$ , o método consiste em estimar  $x_{k+1}$  com uma busca linear na direção oposta ao gradiente da função (Prado & Martinez, 2013). Mais precisamente, dada uma função convexa diferenciável  $z = f(x_1, x_2 \dots, x_n)$  e um ponto  $x^k = (x_1^k, x_2^k \dots, x_n^k)$ , determina-se um novo ponto:

$$x^{k+1} = x^k - \alpha \cdot \nabla f(x^k) \quad (1)$$

onde  $\alpha$  é obtido de tal maneira que:

$$z = \min_{\alpha} f[x^k - \alpha \cdot \nabla f(x^k)] \quad (2)$$

### 2.2 Método de Newton

Utiliza-se o método de Newton na determinação de máximos e mínimos para funções reais de  $n$  variáveis reais (Oliveira & Martinez, 2013). Para isto, procuramos pontos críticos para a função  $f$ , ou seja, estamos interessados em pontos tais que, verifica-se:  $\nabla f(x_k) = 0$ .

De maneira mais precisa dada uma função  $z = f(x_1, x_2 \dots, x_n)$  convexa e duas vezes diferenciável e um ponto  $x^k = (x_1^k, x_2^k \dots, x_n^k)$ , determina-se um novo ponto  $x^{k+1} = x^k - \alpha \cdot d$  onde  $\alpha$  é obtido de modo que:

$$z = \min_{\alpha} f[x^k - \alpha \cdot d] \quad (3)$$

sendo  $d$  uma direção de busca descendente obtida por:

$$d = -H^{-1}(x^k) \cdot \alpha \cdot \nabla f(x^k) \quad (4)$$

Este método é caracterizado pelo cálculo da matriz hessiana a fim de determinar o ponto crítico da função objetivo, enquanto os métodos quase-Newton possuem a particularidade de estimar a matriz hessiana baseando-se no cálculo do gradiente.

### 2.3 Método DFP (Davidon Fletcher Powell)

O método DFP é um método iterativo conhecido por minimizar uma função irrestrita de  $n$  variáveis ou parâmetros. É um método que utiliza o gradiente para calcular aproximadamente

a matriz Hessiana.

Métodos de minimização iterativos determinam a posição de um  $\bar{x}$  mínimo de uma função como o limite de uma sequência de posições sucessivas  $x_0, x_1, \dots, x_n$ . Essas posições são obtidas minimizando a função ao longo de uma linha através da posição anterior de uma direção especificada de busca  $s_i$  obtida por alguma lógica especificada (Corrêa, 2013). O método DFP é definido pela fórmula abaixo:

$$H_{k+1} = H_k + \frac{p^k (p^k)^t}{(p^k)^t q^k} - \frac{H_k q^k (q^k)^t H_k}{(q^k)^t H_k q^k} \quad (5)$$

Sendo  $H_k$  aproximações da inversa da matriz Hessiana, o  $g^k$  é a diferença entre gradiente atual e o anterior, o  $p^k$  é a diferença entre o chute atual e o anterior,  $q^k$  é a diferença entre o gradiente atual com o anterior.

## 2.4 Método BFGS (Broyden–Fletcher–Goldfarb–Shanno)

Este método é derivado do método DFP, fazendo-se necessário o cálculo de gradientes da função objetivo. A diferença deste método para o DFP é que criamos um  $s$  que é o espaço de busca, este é obtido pela diferença dos chutes iniciais, o cálculo de aproximação da Hessiana é realizado com base nas diferenças de gradientes ao longo de várias iterações.

## 2.5 Método Nelder-Mead

O método de Nelder-Mead é classificado como um método determinístico que não utiliza o cálculo do gradiente e é comumente usado para encontrar o mínimo ou máximo de uma função objetivo em um espaço multidimensional (espaço de busca). É aplicado a problemas de otimização não linear para os quais as derivadas podem não ser conhecidas, ou seja, não requer que a função objetivo seja diferenciável, mas requer que a função seja contínua. O método usa o conceito de um simplex, que é um polítopo especial de  $n + 1$  vértices em  $n$  dimensões. O comportamento do método pode ser visto como a expansão, contração e movimentação desse simplex pelo espaço de busca do problema.

# 3. MÉTODOS ESTOCÁSTICOS

Os métodos de otimização baseados nos algoritmos probabilísticos usam apenas a avaliação da função objetivo e introduzem no processo de otimização dados e parâmetros estocásticos. Além disso, eles não utilizam a derivada da função objetivo.

## 3.1 Método Luus-Jaakola

O algoritmo de Luus-Jaakola foi proposto em 1973 para a solução de problemas de otimização. Um dos seus objetivos era que pudesse ser aplicado a qualquer região de interesse e para quaisquer funções. Este método gera soluções aleatórias da função objeto em uma ampla região de busca, a cada iteração a região de busca é reduzida e a solução da função é melhorada, o processo é repetido até obter-se um intervalo suficientemente pequeno no qual a solução ótima esteja imersa (Oliveira et al., 2016).

### 3.2 Método PSO (Particle Swarm Optimization)

O método PSO, sigla de Particle Swarm Optimization, cuja tradução é Otimização por Exames de Partículas, basea-se na simulação de um modelo de interação social simplificado. A ideia do algoritmo PSO para a busca de uma solução ótima é calcular a nova posição de cada indivíduo  $x_i(t+1)$  a partir de sua posição atual  $x_i(t)$ , somada a uma nova velocidade  $v_i(t+1)$ , sendo que  $v_i(t+1)$  é calculada utilizando a própria experiência do indivíduo e as experiências dos outros membros da população (Viana, 2017). Desse modo a posição de cada indivíduo  $x_i(t+1)$  é dada por:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (6)$$

A nova velocidade de cada indivíduo é dada por:

$$v_i(t+1) = \omega(t) \cdot v_i(t) + c_1(t) \cdot rand() \cdot (p_i - x_i(t)) + c_2(t) \cdot rand() \cdot (p_g - x_i(t)) \quad (7)$$

onde  $\omega(t)$  é o parâmetro de inércia,  $c_1(t)$  e  $c_2(t)$  são os parâmetros cognitivo e social, respectivamente,  $p_i$  é a melhor posição que o indivíduo  $i$  visitou,  $p_g$  é a melhor posição que todos os indivíduos da população tenham visitado e  $rand()$  são valores aleatórios gerados por distribuição uniforme no intervalo (Corrêa, 2013).

### 3.3 Método DE (Evolução Diferencial)

O algoritmo evolução diferencial surgiu em 1997, foi desenvolvido por Storn e Price e tem como objetivo solucionar problemas de otimização contínua. Este método trabalha com um mecanismo de seleção natural, onde um número de indivíduos é gerado aleatoriamente, cada indivíduo é candidato a solução, uma nova população é gerada passando pelo processo de mutação e cruzamento, no processo de seleção é comparada a geração anterior com a atual, após este processo permanece o melhor indivíduo para a próxima iteração (Corrêa, 2013). Na mutação um indivíduo é perturbado resultando no vetor doador, este processo é dado pela seguinte equação:

$$\mathbf{v}_d^{q+1} = \mathbf{x}_\alpha^q + F_p(\mathbf{x}_\gamma^q - \mathbf{x}_\delta^q) \quad (8)$$

O cruzamento tem como objetivo aumentar a diversidade dos indivíduos que sofrem mutação, o processo acontece através da mistura de componentes do vetor doador  $\mathbf{v}_d^{q+1}$  com as componentes do vetor alvo  $\mathbf{x}_a^q$ , que é escolhido aleatoriamente, resultando num vetor tentativa ou experimental  $\mathbf{v}_t^{q+1}$ . O processo de seleção tem como objetivo encontrar uma nova geração, onde é considerado seu critério de parada, se a função do vetor tentativa for menor que a função do vetor aleatório que foi escolhido, o vetor escolhido recebe o valor do vetor tentativa, se isso não acontecer, o vetor alvo continua para a próxima iteração. É repetido o processo até o critério de parada.

#### 4. RESULTADOS NUMÉRICOS

Nesta seção serão executados os métodos determinísticos e estocásticos, afim de obter-se resultados numéricos que por sua vez permitem realizar uma análise comparativa dos métodos. A primeira função a ser testada, foi retirada de (Ribeiro et al., 2015), que apresentou resultado semelhante.

$$f(x, y) = \frac{x^2}{4000} + \frac{y^2}{4000} - \left(\cos\left(\frac{x}{\sqrt{1}}\right)\right) \cdot \left(\cos\left(\frac{y}{\sqrt{2}}\right)\right) + 1 \quad (9)$$

Figura 1- Gráfico do comportamento da função à esquerda e curva de nível à direita.

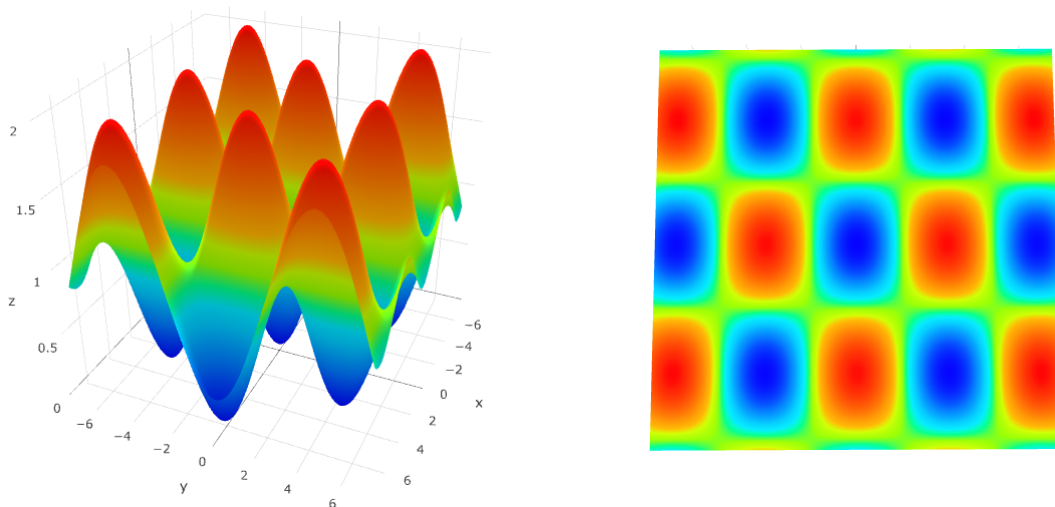


Tabela 1- Resultados numéricos referentes à Eq. (9).

Métodos	Convergência	Tempo (s)	Laços	Chute Inicial	Tolerância
Gradiente	(-5E-06;-6,3E-06)	0,006	18	(1;1)	$10^{-7}$
Newton	(-5E-06;-5E-06)	0,005	4	(0;1)	$10^{-7}$
DFP	(-7E-06;2E-06)	0,009	18	(1;1)	$10^{-7}$
BFGS	(-4,9E-06;-5,1E-06)	0,003	8	(1;1)	$10^{-7}$
Nelder-Mead	(1,2E-08;6,4E-08)	0,011	44	-	$10^{-7}$
Luss-Jaakola	(3E-06;3,6E-06)	0,098	(50;80)	-	-
PSO	(-2,7E-02;4,1E-04)	0,142	100	-	-
DE	(1,8E-08;-1,9E-08)	0,206	63	-	$10^{-7}$

A segunda função analisada foi retirada do recente trabalho (Corrêa Junior & Corrêa, 2016) onde os autores alcançaram resultados próximos.

$$f(x, y) = -\frac{1 + \cos(12\sqrt{x^2 + y^2})}{0.5(x^2 + y^2) + 2} \quad (10)$$

Figura 2- Gráfico do comportamento da função à esquerda e curva de nível à direita.

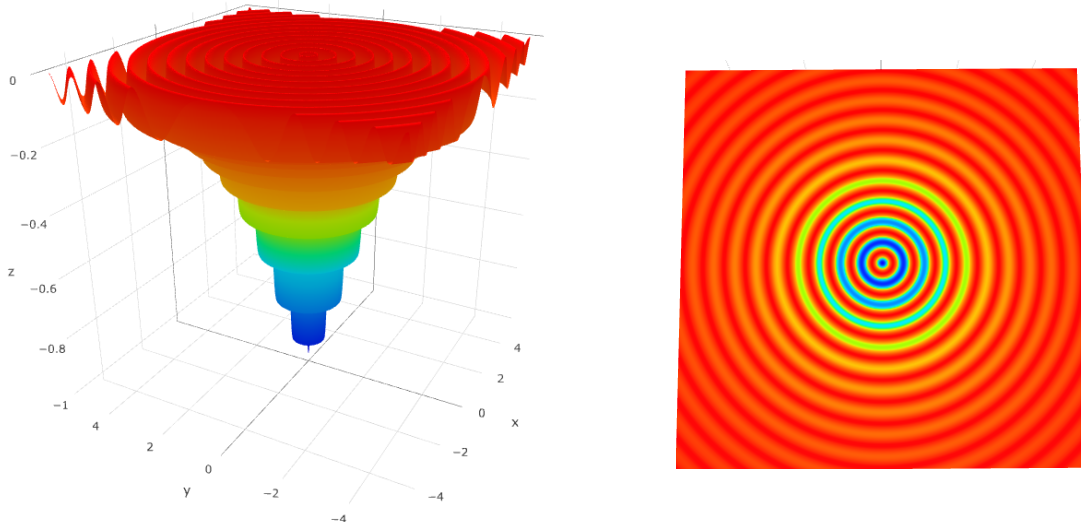


Tabela 2- Resultados numéricos referentes à Eq. (10).

Métodos	Convergência	Tempo (s)	Laços	Chute Inicial	Tolerância
Gradiente	(-2,4E-06;3E-07)	0,007	11	(0,1;0,2)	$10^{-7}$
Newton	(-5E-06;-5E-06)	0,008	5	(0,08;0,1)	$10^{-7}$
DFP	(5,1E-06;5,9E-06)	0,008	12	(0,1;0,2)	$10^{-7}$
BFGS	(-3E-07;4,1E-06)	0,014	36	(0,1;0,2)	$10^{-7}$
Nelder-Mead	(6,3E-04;3,7E-04)	0,006	16	-	$10^{-7}$
Luss-Jaakola	(-2,2E-06;7,5E-06)	0,102	(50;80)	-	-
PSO	(3E-03;4,1E-04)	0,149	100	-	-
DE	(9,3E-08;9,7E-08)	0,461	144	-	$10^{-7}$

A terceira função testada foi obtida do trabalho (Corrêa Junior & Corrêa, 2016) onde os autores alcançaram resultados similares.

$$f(x, y) = 100 \cdot (y - x^2)^2 + (x - 1)^2 \quad (11)$$

Figura 3- Gráfico do comportamento da função à esquerda e curva de nível à direita.

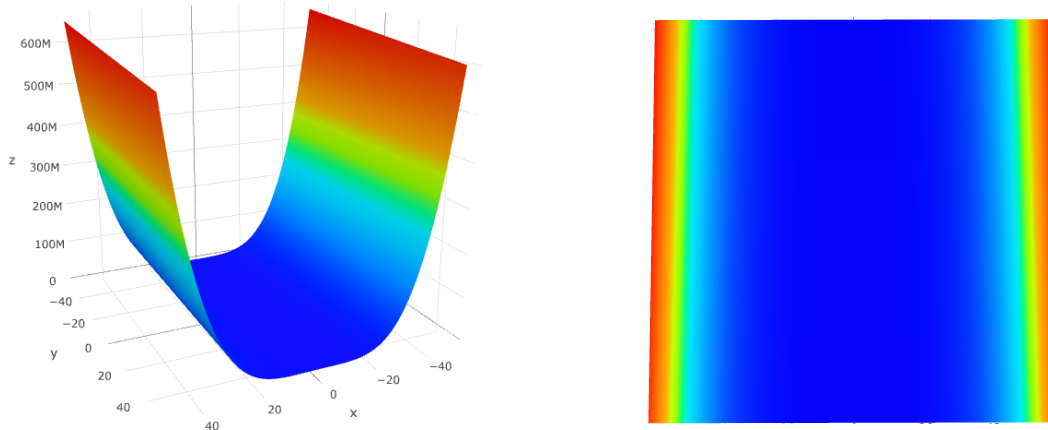


Tabela 3- Resultados numéricos referentes à Eq. (11).

Métodos	Convergência	Tempo (s)	Laços	Chute Inicial	Tolerância
Gradiente	(1,005;1,01)	0,047	173	(1,5;1)	$10^{-7}$
Newton	(1;1)	0,016	2	(1;2)	$10^{-7}$
DFP	(1,002;1,004)	0,016	14	(1,5;2)	$10^{-7}$
BFGS	(1,001;1,003)	0,015	30	(1,5;2)	$10^{-7}$
Nelder-Mead	(1;1)	0,015	77	-	$10^{-7}$
Luss-Jaakola	(1;1)	0,094	(50;80)	-	-
PSO	(1;1)	0,140	100	-	-
DE	(1;1)	0,672	206	-	$10^{-7}$

## 5. ANÁLISE COMPARATIVA

Apresentando os parâmetros utilizados e os resultados alcançados na Eq. (9), o método de Newton foi tomado como referência para comparar do tempo de execução entre os demais métodos:

No método do Gradiente foi utilizado o critério de Armijo, este método convergiu para o ótimo da função  $(-5E-06;-6,3E-06)$  com um total de 18 laços e tempo igual a 6 milissegundos (ms), o critério de parada utilizado neste método foi enquanto o gradiente no ponto for maior que a tolerância ( $10^{-7}$ ), o algoritmo seguia para a próxima iteração, o chute inicial utilizado foi (1, 1). Este método apresentou 120% do tempo de execução em comparação ao método de Newton.

O método de Newton convergiu para o ótimo  $(-5E-06;-5E-06)$ , neste método foi utilizado o critério de Armijo, o tempo para convergencia do método foi 5ms e um total de 4 laços. O critério de parada foi o enquanto o gradiente no ponto for maior que a tolerância ( $10^{-7}$ ), o algoritmo seguia para a próxima iteração, foi utilizado chute inicial (0, 1), o chute inicial foi ajustado para que o método pudesse alcançar o ótimo da função, pois caso o chute fosse igual aos outros métodos determinísticos, não seria possível convergir, este fato mostra uma hipersensibilidade ao chute inicial.

O método DFP também utilizou critério de Armijo, convergindo para o ótimo (-7E-06;2E-06), 18 laços e 9ms foi o tempo necessário para convergência do método, o critério de parada foi o gradiente ser menor ou igual a tolerância ( $10^{-7}$ ), foi utilizado chute inicial (1, 1). Este método apresentou 180% do tempo de execução do método de Newton.

No método BFGS assim como nos outros métodos determinísticos foi utilizado o critério de Armijo, este método convergiu para o ótimo (-4,9E-06;-5,1E-06), sua convergência foi alcançada com tempo recorde de 3ms e 8 laços, o critério de parada utilizado foi enquanto o maior valor do gradiente fosse maior que a tolerância ( $10^{-7}$ ), o algoritmo seguia para próxima iteração, o chute inicial utilizado foi (1, 1). Este método apresentou apenas 60% do tempo de execução comparado ao método de Newton.

O método Nelder-Mead é o único método determinístico apresentado neste trabalho que não utiliza do cálculo do gradiente, este método convergiu para o ótimo (1,2E-08;6,4E-08), sua convergência foi dada com 44 laços e 11ms, este método possui chute inicial aleatório, e o critério de parada utilizado foi enquanto a diferença entre  $x_1$  e  $x_2$  for maior que a tolerância ( $10^{-7}$ ), continuar para a próxima iteração. Este método apresentou 220% do tempo de execução em comparação ao método de Newton.

O método Luus-Jaakola convergiu para o ótimo (3E-06;3,6,3E-06) em 98ms, foi definido nos parâmetros do método 50 laços internos, 80 laços externos e coeficiente de contração igual a 0,1, este método possui chute inicial aleatório, e não possui critério de parada. Este método apresentou 1960% do tempo de execução em comparação ao método de Newton.

O método PSO convergiu para o ótimo (-2,7E-02;4,1E-04) em um tempo de 142ms, foi definido como parâmetro 100 laços, número de indivíduos igual a 50, coeficiente de inércia igual a 0,7 e coeficiente cognitivo e social é igual a 0,7. Neste método a posição e velocidade inicial são geradas aleatoriamente e não possui critério de parada. Este método apresentou 2840% do tempo de execução em comparação ao método de Newton.

O método DE convergiu para o ótimo (1,8E-08;-1,9E-08) com 63 laços em 206ms, o critério de parada foi enquanto a diferença entre os melhores pontos de  $x_1$  e  $x_2$  for maior que a tolerância ( $10^{-7}$ ) o algoritmo segue para próxima iteração. Foram usados como parâmetro número de indivíduos igual a 50, coeficiente de cruzamento igual a 0,7 e coeficiente de mutação igual a 0,3, a população é gerada aleatoriamente. Este método apresentou 4120% do tempo de execução em comparação ao método de Newton.

É apresentada uma maior complexidade na Eq. (10), porém não houve alteração nos parâmetros utilizados nos métodos estocásticos para gerar os resultados numéricos desta função em relação a Eq. (9), já os métodos determinísticos necessitaram de ajustes no chute inicial, destaque para o método de Newton que precisou de um refinamento maior para alcançar o ótimo da função. Sem estes ajustes os métodos determinísticos convergiam para ótimos locais, este teste reforça a ideia de hipersensibilidade aos parâmetros iniciais que os métodos determinísticos possuem, entretanto o desempenho apresentado é superior em relação aos métodos estocásticos, fazendo-se uso de menos laços para a convergência. Será comparado agora o tempo de execução entre os métodos utilizados na Eq. (10), tomando como referência o método de Newton.

O método do Gradiente apresentou apenas 87,5% do tempo de execução comparado ao método de Newton. O método DFP apresentou o mesmo tempo de execução que o método de Newton. O método BFGS apresentou 175% do tempo de execução comparado ao método de Newton.

O método Nelder-Mead apresentou apenas 75% do tempo de execução comparado ao método de Newton. O método Luus-Jaakola apresentou 1275% do tempo de execução comparado ao método de Newton. O método PSO apresentou 1862,5% do tempo de execução comparado ao



método de Newton. O método DE apresentou 5762,5% do tempo de execução comparado ao método de Newton.

A Eq. (11) diferente das Eq. (9) e Eq. (10) não possui múltiplos mínimos locais, esta característica auxilia a convergência dos métodos, evitando que estes por sua vez fiquem presos em pontos de mínimo ou máximo locais. Apesar do vale onde está contido o mínimo global ser encontrado facilmente, uma boa convergência para o mínimo global é difícil de ser obtida (Picheny et al., 2012). É possível verificar que na Eq. (11) os métodos estocásticos ainda que máis demorados apresentaram maior precisão na convergência. Foram mantidos todos os parâmetros com os métodos estocásticos nos três testes, entretanto os métodos determinísticos necessitaram de ajustes nos parâmetros iniciais para que fosse alcançado o ótimo da função. O método de Newton foi tomado como referência na comparação do tempo de execução entre os demais métodos na Eq. (11).

O método do Gradiente apresentou 293,75% do tempo de execução comparado ao método de Newton. O método DFP apresentou o mesmo tempo de execução que o método de Newton. O método BFGS apresentou apenas 93,75% do tempo de execução comparado ao método de Newton.

O método Nelder-Mead apresentou apenas 93,75% do tempo de execução comparado ao método de Newton. O método Luss-Jaakola apresentou 587,5% do tempo de execução comparado ao método de Newton. O método PSO apresentou 875% do tempo de execução comparado ao método de Newton. O método DE apresentou 4200% do tempo de execução comparado ao método de Newton.

Mesmo que os métodos apresentados nesse trabalho possuam características distintas, é evidente que há proximidade em seus resultados.

## 6. CONCLUSÃO

Analisando os resultados obtidos pela simulação, é possível concluir que os métodos determinísticos e os estocásticos alcançaram resultados numéricos próximos. Em todos os métodos foram escolhidos parâmetros para que a comparação fosse a mais justa possível.

Observando o tempo necessário para convergência do método e o número de laços, pode-se notar que os métodos determinísticos tendem a convergir mais facilmente para o ótimo da função, isto é, utilizam de menos recursos computacionais para convergir comparado aos estocásticos, entretanto, estes métodos dependem que a função objetiva seja diferenciável, caso contrário será criado um cenário onde o método não convergirá. Outro quadro onde os métodos determinísticos podem divergir é quando o chute inicial é escolhido inadequadamente, durante a execução dos métodos foi necessário ajustar o chute inicial no método de Newton para que este convergisse para o ótimo da função, essa característica é devida as estruturas dos métodos determinísticos apresentarem sensibilidade ao chute inicial.

Os métodos estocásticos, por sua vez, não dependem que a função seja diferenciável nem do chute inicial, estas características permitem solucionar uma ampla gama de funções das quais os métodos determinísticos não são factíveis, com o preço de terem maior custo computacional, houveram cenários durante os testes onde o tempo de execução do método estocástico Evolução Diferencial chegou a ser maior que 40 vezes o valor do tempo de execução do método de Newton.

## Referências

- Corrêa, Rosilene Abreu Portela (2013), “*Identificação de Danos em Estruturas Bi-Dimensionais via Matriz de Flexibilidade Baseada em um Modelo de Dano Contínuo*”, Tese de Doutorado, IPRJ/UERJ, Nova Friburgo.
- Corrêa Junior, C. A.; Corrêa, R. A. P. (2016), “*Alcateia - Um Novo Algoritmo Para Otimização*”, Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, CNMAC, Gramado, vol 5.
- de Oliveira, G. L.; Martinez, A. L. M. (2013), “Um estudo introdutório do método de Newton para minimização de funções de uma ou várias variáveis”, *Congresso de Matemática Aplicada e Computacional*, 288-290.
- Oliveira, Matheus Braga; de Souza, Joviana Sartori (2016), “Determinação de minimizadores globais de funções por meio de métodos híbridos com base no método LuusJaakola”, *XLVIII Simpósio Brasileiro de Pesquisa Operacional*, Vitória, 2607-2617.
- Picheny, V.; Wagner, T.; Ginsbourger, D. (2012), “*A benchmark of kriging-based infill criteria for noisy optimization*”, hal-00658212.
- Prado, J. C. L.; Martinez, A. L. M. (2013), “Um estudo introdutório do método do Gradiente para minimização de funções de uma ou várias variáveis”, *Congresso de Matemática Aplicada e Computacional*, 234-236.
- Ribeiro, M. A. C.; Câmara, L. D. T.; Neto, A. J. S. (2015), “*Validação do Método Estocástico R2W na Obtenção de Ótimos Globais de Funções Não Lineares*”, REUCP, Petrópolis, vol 9, 42-53.
- Viana, Paulo Emanuel Soares. (2017), “*Métodos de Simulação Estocástica Aplicados em Problemas Não Lineares*”, Dissertação de Mestrado, UFF, Volta Redonda.

## COMPARISON BETWEEN DETERMINISTIC AND STOCHASTIC METHODS IN THE RESOLUTION OF NON-LINEAR EQUATIONS

**Abstract.** *This work consists of a comparative analysis between some stochastic and deterministic methods used in solving nonlinear equations. The deterministic methods used were: gradient, Newton method, DFP, BFGS and Nelder Mead. The Stochastics: DE (Differential Evolution), PSO (Optimization by Swarm of Particles) and Luss-Jaakola. From the results obtained, it was possible to verify that the deterministic methods presented a better convergence in relation to the stochastics because they require less computational effort in their convergence, among them is Newton’s method, converging with less loops than the others. It was noted that stochastic methods require more time to converge, however all methods have obtained approximate results.*

**Keywords:** *Nonlinear equations, Comparison, Deterministic methods, Stochastic methods, Optimization*