O MÉTODO CIENTÍFICO
21 a 25 de outubro de 2013
EPCT
VIII Encontro de Produção Científica e Tecnológica

MÉTODOS NEWTON E QUASE-NEWTON PARA OTIMIZAÇÃO IRRESTRITA

Marlon Luiz Dal Pasquale Junior, UNESPAR/FECILCAM,
jr.marlon@hotmail.com
Solange Regina dos Santos (OR), UNESPAR/FECILCAM,
solaregina@fecilcam.br
Gislaine Aparecida Periçaro (CO-OR), (UNESPAR/FECILCAM),
gapericaro@fecilcam.br

1 INTRODUÇÃO

A otimização é uma ferramenta da Matemática Aplicada utilizada na resolução de problemas nos quais é preciso encontrar a solução mais eficiente entre todas as possibilidades existentes. Tais soluções são chamadas de minimizadores ou maximizadores da função objetivo do problema considerado. São diversas as utilidades desse tipo de solução, por exemplo, a construção de uma usina hidrelétrica requer a solução de diversos problemas logísticos em que cada um desses problemas fornece um custo a ser somado ao total da obra. Neste caso saber quais são as soluções mais eficientes é fundamental para determinar a redução dos custos do empreendimento.

Discutimos neste trabalho métodos de otimização empregados na resolução do seguinte problema

$$\begin{array}{ll} \textit{minimizar} & f(x) \\ \text{sujeito a} & x \in \mathbb{R}^n, \end{array} \qquad (1)$$

em que $f: \mathbb{R}^n \to \mathbb{R}$ é uma função continuamente diferenciável. O objetivo na resolução deste problema é determinar os minimizadores locais de f. Dizemos que $\bar{x} \in \mathbb{R}^n$ é um minimizador local de f se $f(\bar{x}) \leq f(x)$ para todo x pertencente a uma vizinhança de $\bar{x}$. O problema (1) foi formulado como um problema de minimização, pois do ponto de vista matemático não existe diferença entre minimização e maximização, uma vez que maximizar $f(x)$ é equivalente a minimizar $-f(x)$ (Izmailov e Solodov, 2007).

Dessa forma, a teoria desenvolvida para problemas de otimização leva em consideração problemas da forma apresentada em (1).

O teorema a seguir estabelece uma condição necessária para que um ponto seja minimizador de f, a qual é conhecida como condição necessária de primeira ordem, cuja demonstração pode ser encontrada em Ribeiro e Karas (2013).

Teorema 1.1: Seja $f: \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Se $\bar{x}$ é um minimizador local de f, então $\nabla f(\bar{x}) = 0$.

PARANÁ GOVERNO DO ESTADO
FUNDAÇÃO ARAUCÁRIA Apoio ao Desenvolvimento Científico e Tecnológico do Paraná
UNIVERSIDADE ESTADUAL DO PARANÁ CAMPUS DE CAMPO MOURÃO
FECILCAM Pública, gratuita, de qualidade
NUPEM Núcleo de Pesquisa Multidisciplinar

De acordo com Martínez e Santos (1998), um algoritmo básico de otimização irrestrita consiste em, a partir de cada ponto obtido, determinar uma direção para dar o próximo passo. Como o objetivo é minimizar *f*, é razoável que a função decresça na direção escolhida. Uma direção que apresenta essa característica é denominada direção de descida, cuja definição formal é apresentada a seguir.

Definição 1.1: Considere uma função $f: \mathbb{R}^n \to \mathbb{R}$, um ponto $\bar{x} \in \mathbb{R}^n$ e uma direção $d \in \mathbb{R}^n \setminus \{0\}$. Dizemos que d é uma direção de descida de f, a partir do ponto $\bar{x}$, quando existe $\delta > 0$ tal que $f(\bar{x} + td) < f(\bar{x})$, para todo $t \in (0, \delta)$.

O teorema a seguir estabelece uma condição para que d seja uma direção de descida, cuja demonstração pode ser encontrada em Ribeiro e Karas (2013).

Teorema 1.2: Se $\nabla f(\bar{x})^T d < 0$, então d é uma direção de descida para f, a partir de $\bar{x}$.

Conforme apresentado por Ribeiro e Karas (2013), existe uma condição suficiente para que uma direção seja de descida, a saber: se $\nabla f(\bar{x})^T d < 0$ então d é uma direção de descida para f a partir de $\bar{x}$. Assim, as direções que formam um ângulo maior que 90º com o gradiente de f avaliado em $\bar{x}$, são direções de descida para f a partir desse ponto.

Após escolher a direção de descida, é necessário definir o quanto caminhar nessa direção. Pode-se dar um passo completo, ou seja, considerar o tamanho $t_k = 1$, no entanto, esta escolha pode não ser a melhor. Dessa forma, existem métodos denominados métodos de busca que podem ser empregados para esse fim, sendo um deles a busca de Armijo que será discutida na próxima seção.

Com base no que foi discutido anteriormente, apresentamos o algoritmo básico de otimização.

Algoritmo 1.1: ALGORITMO BÁSICO

Dados: $x^0 \in R^n$ (ponto inicial)

$k = 0$

Enquanto $\nabla f(x^k) \neq 0$

Determine uma direção de descida d^k

Determine $t_k > 0$ tal que $f(x^k + t_k d^k) \leq f(x^k)$

Faça $x^{k+1} = x^k + t_k d^k$

$k = k + 1$

Da condição necessária de primeira ordem (Teorema 1.1), sabemos que se x^k for um minimizador da função objetivo, então $\nabla f(x^k) = 0$. Dessa forma, o algoritmo continua a atualizar x^k enquanto o gradiente for diferente de zero.

O que diferencia os métodos de otimização é a escolha da direção de descida e a forma com que é calculado o tamanho do passo. Para determinar a direção de descida em nossa pesquisa estudamos o Método de Newton e os Métodos Quase-Newton e para encontrar o tamanho do passo utilizamos a busca de Armijo, os quais serão discutidos a seguir.

1.1 BUSCA DE ARMIJO

A busca de Armijo consiste em obter uma redução da função objetivo ao longo da direção de descida, sem tentar minimizá-la. A condição (2), conhecida por busca de Armijo, assegura um decréscimo monótono dos valores da função objetivo ao longo do processo iterativo.

Considere então um ponto $\bar{x} \in \mathbb{R}^n$, uma direção de descida $d \in \mathbb{R}^n$ e $\sigma \in (0,1)$. Basicamente, a busca de Armijo encontra $\bar{t} > 0$ tal que

$$f(\bar{x} + \bar{t}d) \leq f(\bar{x}) + \sigma \bar{t} \nabla f(\bar{x})^T d. \tag{2}$$

Para interpretar a busca de Armijo vamos considerar a função $g: [0, \infty) \to \mathbb{R}^n$ dada por $g(t) = f(\bar{x} + td)$ e o seu modelo linear m obtido da aproximação de primeira ordem em torno de $t = 0$, ou seja,

$$m(t) = g(0) + tg'(0) = f(\bar{x}) + t\nabla f(\bar{x})^T d.$$

Da desigualdade (2), temos que

$$f(\bar{x} + \bar{t}d) - f(\bar{x}) \leq \sigma \bar{t} \nabla f(\bar{x})^T d,$$

sendo equivalente a

$$f(\bar{x} + \bar{t}d) - f(\bar{x}) \leq \sigma[m(\bar{t}) - m(0)]. \tag{3}$$

A diferença $m(\bar{t}) - m(0)$ representa a redução do modelo linear para o passo $\bar{t}$ na direção d^k. Portanto, a busca de Armijo estabelece que o decréscimo real de f, dado por $f(\bar{x} + td) - f(\bar{x})$, seja pelo menos uma fração, determinada por $\sigma \in (0,1)$, da redução prevista pelo modelo linear.

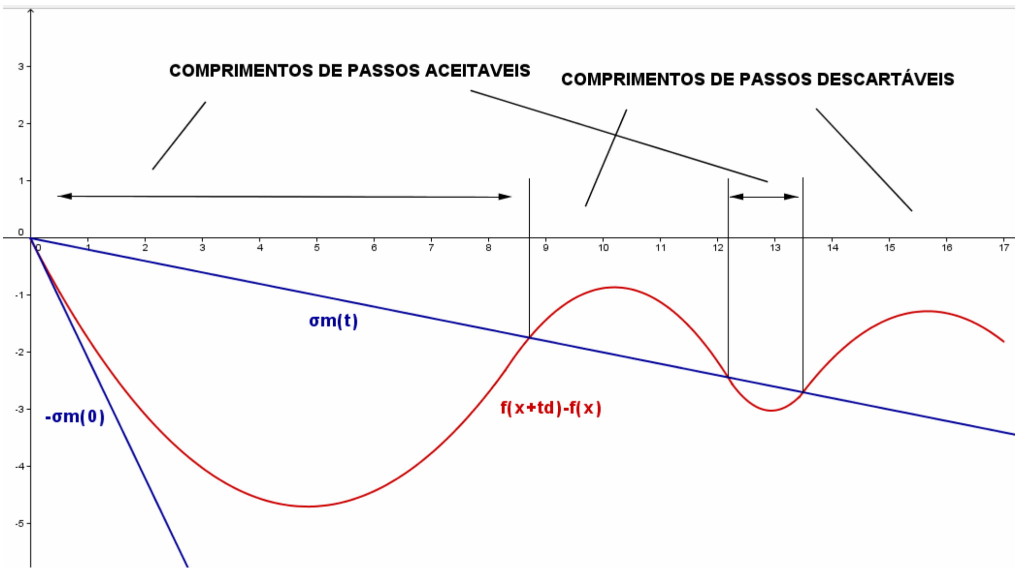

Imagem 01 – Interpretação gráfica da busca de Armijo.

A Imagem 01 representa uma interpretação gráfica da busca de Armijo. Note que a desigualdade (3) é satisfeita quando o gráfico de $f(\bar{x} + td) - f(\bar{x})$ está abaixo da reta $\sigma m(\bar{t})$ Portanto, existem certos intervalos que fornecem tamanhos de passo aceitáveis para a busca. A busca de Armijo está resumida no algoritmo a seguir.

Algoritmo 2.1: BUSCA DE ARMIJO

Dado: $\bar{x} \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ (direção de descida) e $\sigma \in (0,1)$.

$t = 1$

Enquanto $f(\bar{x} + td) - f(\bar{x}) > \sigma t \nabla f(\bar{x})^T d$

$t = 0{,}8t$

Na próxima seção será discutida uma forma de obter a direção de descida para o algoritmo de otimização.

2 MÉTODO DE NEWTON

A resolução do sistema $\nabla f(x) = 0$, muitas vezes não linear, pode ser feita por meio do método de Newton.

Considere $F: \mathbb{R}^n \to \mathbb{R}^n$ de classe C^1 e o seguinte sistema de equações

$$F(x) = 0.$$

Aproximando F por meio do polinômio de Taylor de primeira ordem ao redor de $\bar{x}$, obtemos

$$F(\bar{x}) + J_f(\bar{x})(x - \bar{x}) = 0,$$

em que, J_f é a matriz Jacobiana de F. Caso $J_f(\bar{x})$ seja inversível, a solução do sistema é dada por

$$x^+ = \bar{x} - J_f(\bar{x})^{-1}F(\bar{x}).$$

Retomando o problema de minimizar f e utilizando essa ideia para $F = \nabla f$, temos

$$x^+ = \bar{x} - \nabla^2 f(\bar{x})^{-1}\nabla f(\bar{x}).$$

A seguir apresentamos um algoritmo que resume as ideias discutidas acima.

Algoritmo 2.2: MÉTODO DE NEWTON PARA OTIMIZAÇÃO

Dado: $x^0 \in \mathbb{R}^n$

$k = 0$

REPITA enquanto $\nabla f(x^k) \neq 0$

Define $d^k = -\nabla^2 f(x^k)^{-1}\nabla f(x^k)$

Determine o tamanho do passo $t_k > 0$

Faça $x^{k+1} = x^k + t_k d^k$

$k = k + 1$

A cada iteração o algoritmo caminha na direção de busca d^k com tamanho de passo t_k, o qual pode ser calculado por meio dos métodos de busca exata ou inexata. É possível ainda optar por um comprimento de passo fixo, $t_k = 1$, nesse caso o método é chamado de Newton Puro, e sua convergência é quadrática, se a matriz hessiana é não-singular e o ponto inicial está suficientemente próximo da solução. Portanto, na sua forma básica (com comprimento de passo fixo), o método é considerado um método iterativo local (Martínez e Santos, 1998).

Observe que para determinar a direção de busca é necessário o cálculo da derivada segunda da função objetivo e da resolução de um sistema de equações, o que requer alto custo computacional. (Ribeiro e Karas, 2013). Visando reduzir tal custo e manter as boas propriedades do método, foram desenvolvidos métodos que se assemelham tanto quanto possível ao método de Newton, denominados de métodos Quase-Newton.

3 MÉTODO QUASE-NEWTON

O procedimento iterativo utilizado pelos métodos Quase-Newton para minimizar uma função f, considera as direções de busca dadas por

$$d_k = -H_k \nabla f(x^k) \tag{4}$$

onde $H_k \in \mathbb{R}^{n \times n}$ é definida positiva.

Aproximando f por um modelo quadrático em torno de x_k, temos

$$m_k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d, \tag{5}$$

onde $B_k \in \mathbb{R}^{n \times n}$ é uma matriz simétrica qualquer em vez de $\nabla^2 f(x^k)$.

Se B_k for definida positiva, o minimizador do modelo quadrático é dado por $-B_k^{-1} \nabla f(x^k)$.

Deste modo, obtemos (4) escolhendo $B_k = {H_k}^{-1}$.

Os conceitos desenvolvidos anteriormente estão resumidos no algoritmo a seguir.

Algoritmo 3.1: MÉTODO QUASE-NEWTON

Dados $x^0 \in \mathbb{R}^n$, $H_k \in \mathbb{R}^{n \times n}$ definida positiva

$k = 0$

Repita enquanto $\nabla f(x^k) \neq 0$

Defina $d^k = -H_k \nabla f(x^k)$

Obtenha $t_k > 0$ que minimiza $f(x^k + t_k d^k)$

Faça $x^{k+1} = x^k + t_k d^k$

Determine H_{k+1} definida positiva

$k = k + 1$

Observe que se H_k é positiva definida, então a direção $d^k = -H_k \nabla f(x^k)$ é de descida para a função f, a partir de x^k. De fato, se $\nabla f(x^k) \neq 0$, então $\nabla f(\bar{x})^T d = -\nabla f(\bar{x})^T H_k \nabla f(x^k) < 0$. Além disso, se $H_k = \left(\nabla^2 f(x^k)\right)^{-1}$ então a direção de busca coincide com a direção do Algoritmo 2.2. Portanto, se $\nabla^2 f(x^k)$ é definida positiva, então a direção de busca do método de Newton é descida.

Na sequência apresentamos o método DFP usado para calcular uma aproximação para a matriz hessiana.

3.2 O MÉTODO DFP

De acordo com Izmailov e Solodov (2007), a forma que será utilizada para encontrar a matriz H_{k+1} foi proposta por Davidon, Fletcher e Powell. Para tanto, consideremos $H \in \mathbb{R}^{n \times n}$ definida positiva, $p^k, q^k \in \mathbb{R}^n$ tais que $p^T q > 0$, em que $p^k = x^{k+1} - x^k$ e $q^k = \nabla f(x^{k+1}) - \nabla f(x^k)$.

A fórmula do método DFP é obtida por uma correção simétrica de posto dois e seu desenvolvimento pode ser encontrado em Ribeiro e Karas (2013), fornecendo

$$H_{k+1}^{DFP} = H_k + \frac{p^k (p^k)^T}{(p^k)^T q^k} - \frac{H_k q^k (q^k)^T H_k}{(q^k)^T H_k q^k}. \tag{6}$$

A seguir será apresentado mais um desenvolvimento que permite calcular a matriz H_{k+1}.

3.3 O MÉTODO BFGS

Outro modo para atualizar as matrizes no algoritmo do método de Quase-Newton é devido a Broyden, Fletcher, Goldfarb e Shanno (BFGS). A ideia tem uma simetria com a do método DFP e consiste em encontrar uma aproximação para a matriz hessiana, ao invés da sua inversa.

O desenvolvimento da fórmula para matriz H_{k+1} do método BGFS também utiliza a mesma correção simétrica de posto dois e a fórmula de Sherman-Morrison para encontrar a inversa da aproximação da matriz hessiana. O desenvolvimento da fórmula do método BFGS tmabém pode ser encontrado em Riberio e Karas (2013) e fornece

$$H_{k+1}^{BFGS} = H_k + \left(1 + \frac{(q^k)^T H_k q^k}{(p^k)^T q^k}\right)\frac{pp^T}{(p^k)^T q^k} - \frac{H_k q^k p^T + p^k (q^k)^T H_k}{(p^k)^T q^k} \tag{7}$$

Pode ser mostrado que, para os métodos Quase-Newton, se o tamanho do passo for obtido por uma minimização local de $f(x^k + td^k)$ e se H_k for definida positiva, então as matrizes H_{k+1}^{DFP} e H_{k+1}^{BFGS} são definidas positivas. Neste caso, as direções dos métodos serão de descida, como discutido anteriormente.

Na próxima seção apresentamos alguns experimentos numéricos para testar o Algoritmo 2.2 e o Algoritmo 3.1. No que segue, usaremos Algoritmo 3.1 DFP e Algoritmo 3.1 BFGS para dizer que testamos o Algoritmo 3.1 com as matrizes H_{k+1}^{DFP} e H_{k+1}^{BFGS}, respectivamente.

4 TESTES NUMÉRICOS

Nesta seção serão abordados os testes numéricos envolvendo a implementação dos algoritmos no software Matlab R2010a e a resolução dos seguintes problemas, sendo que os problemas 5 e 6 foram testados com 2, 10, 100 e 1000 variáveis.

1) $f(x) = 2(x_1 - 2x_2)^2 + \frac{1}{2}(x_2 - x_1^2)^2$;

2) $f(x) = \frac{1}{2}((x_1 - 1)^2 + (x_2 - x_1^2)^2)$;

3) $f(x) = x_1^2 + x_2^2 - 1$;

4) $f(x) = -24x_1x_2 + x_1^2x_2 + x_1x_2^2$;

5) $f(x) = \sum_{i=1}^{n} e^{x_i} - x_i$;

6) $f(x) = \frac{i}{10}\sum_{i=1}^{n} e^{x_i} - x_i$;

7) $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$;

8) $f(x) = x_1^4 + x_1^2 + x_2^2$;

9) $f(x) = (-13 + x_1 + 5x_2^2 - x_2^3 - 2x_2)^2 + (-29 + x_1 + x_2^3 + x_2^2 - 14x_2)^2$.

Para ilustrar os testes realizados, vamos considerar a resolução do problema 8. Nos testes foi usado o ponto inicial $x^0 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ e a precisão considerada (para todos os problemas) foi $\varepsilon = 10^{-4}$. Assim, o algoritmo para quando $\|\nabla f(x^k)\| < 10^{-4}$ ou quando atinge o número máximo de iterações, $k = 1000$.

As figuras a seguir ilustram os resultados da variação nos valores da função objetivo e da norma do gradiente no decorrer das iterações para os métodos de Newton e Quase-Newton (DFP e BFGS).

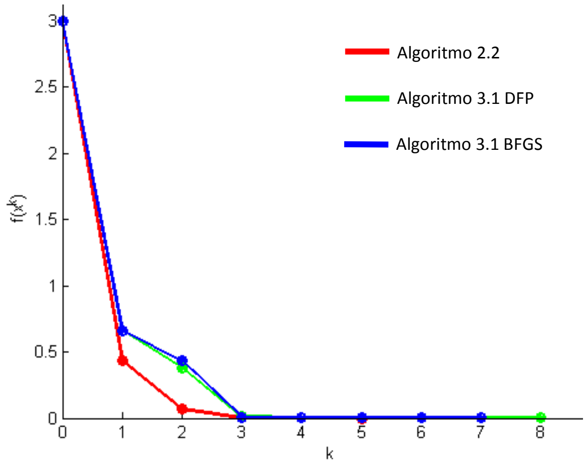

Imagem 02 – Variação na função objetivo.

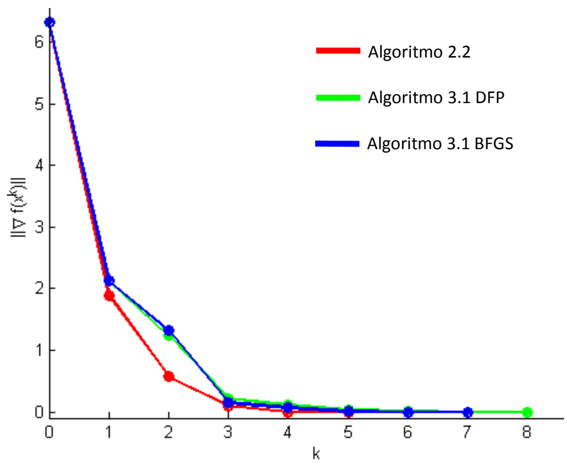

Imagem 03 – Variação na norma do gradiente.

Na Imagem 02 vemos que o Algoritmo 2.2 convergiu para o minimizador do problema, $\bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, após 5 iterações, enquanto que Algoritmo 3.1 DFP convergiu com 8 iterações e o Algoritmo BFGS convergiu com 7 iterações.

Na Imagem 03 ilustramos a norma do gradiente a cada iteração. Nessa imagem vemos que no Algoritmo 2.2 a norma do gradiente decresce mais rapidamente comparada com os demais, pois com 5 iterações atinge a precisão dada.

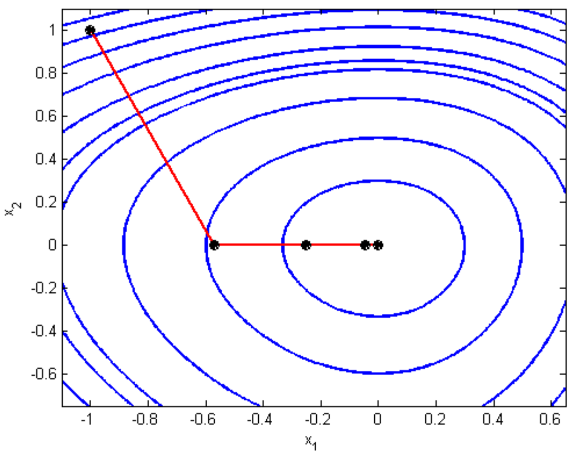

Imagem 04 – Execução do Algoritmo 2.2.

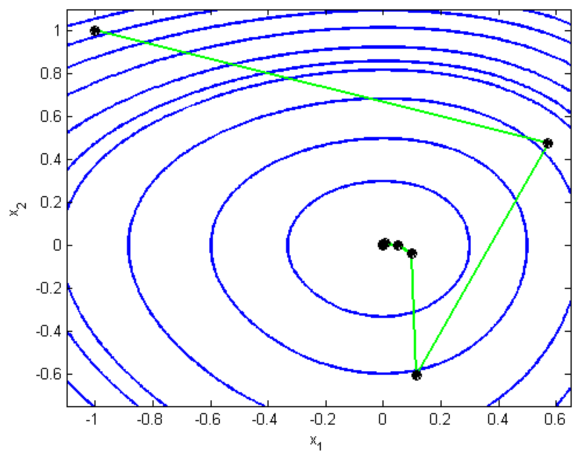

Imagem 05 – Execução do Algoritmo 3.1 DFP.

Nas Imagens 04, 05 e 06 representamos as curvas de nível da função objetivo do problema 8 e também o caminho percorrido pelos iterandos durante a execução dos algoritmos até convergir para o minimizador.

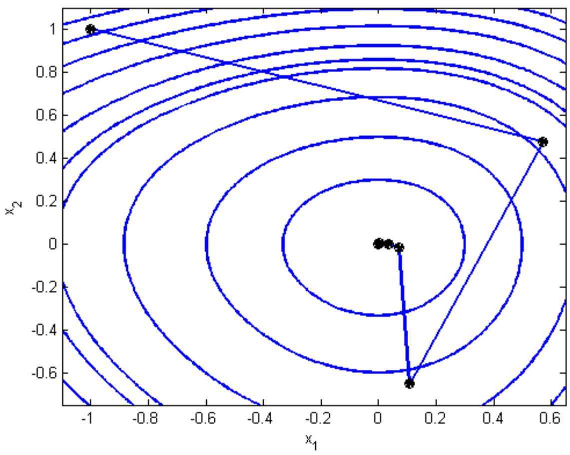

Imagem 06 – Execução do Algoritmo 3.1 BFGS.

Comparando a execução dos três algoritmos, vemos que o Algoritmo 2.2 percorre um caminho mais curto até o minimizador do problema.

Os resultados dos testes para os demais problemas estão apresentados na Tabela 1.

Tabela 1

Testes Numéricos				
		Quantidade de Iterações		
Problema	Ponto Inicial	Algoritmo 2.2	Algoritmo 3.1 DFP	Algoritmo 3.1 BFGS
1	(2,6)	7	13	14
2	(-1,0)	6	11	9
3	(-2,1)	1	2	2
4	(7,9)	3	6	6
5 (n=2)	(1,1)	4	6	6
5 (n=10)	(1,...,1)	4	6	6
5 (n=100)	(1,...,1)	4	7	7
5 (n=1000)	(1,...,1)	4	7	7
6 (n=2)	(1,1)	4	15	10
6 (n=10)	(1,...,1)	4	21	16
6 (n=100)	(1,...,1)	4	57	63
6 (n=1000)	(1,...,1)	4	253	263

7	(-7,8)	41	> 1000	43
8	(-1,1)	5	8	7
9	(3,-1)	5	7	8

Como visto na Tabela 1, o Algoritmo 2.2 obteve desempenho melhor em todos os problemas quando comparado aos demais algoritmos. Isso se deve ao fato de que o método de Newton converge rapidamente quando o ponto inicial está suficientemente próximo do minimizador (Martínez e Santos, 1998).

Para o problema 7, o Algoritmo 3.1 DFP atingiu o número máximo de iterações e por isso sua execução foi interrompida.

CONSIDERAÇÕES FINAIS

No presente trabalho foram estudados os aspectos teóricos e computacionais dos métodos de Newton e Quase-Newton empregados na resolução de problemas de otimização. Os problemas considerados não possuem nenhum tipo de restrição, o que caracteriza um estudo de otimização irrestrita.

Os métodos estudados são caracterizados como métodos de descida, uma vez que a partir de um ponto inicial o algoritmo caminha na direção em que a função objetivo decresce, denominada direção de descida. Sendo assim, o que diferencia os métodos de otimização é a escolha da direção de descida e a forma com que é calculado o tamanho do passo a ser dado.

O método de Newton é uma ferramenta simples de ser implementada, pois requer, a cada iteração, o cálculo do gradiente e da hessiana da função objetivo. Porém, o cáculo da matriz hessiana apresenta um alto custo computacional.

Visando reduzir o esforço computacional e manter as simplicidade do método de Newton, foram desenvolvidos métodos que aproximam a matriz hessiana usando apenas informações de primeira ordem, esses métodos são denominados Quase-Newton.

Para compreensão do desempenho dos métodos estudados, foram realizados testes computacionais com nove problemas, sendo dois deles (problema 5 e 6) testados com quantidades diferentes de variáveis. Os algoritmos foram implementados e testados via *software* Matlab. Para o problema 8, cuja execução foi detalhada no inicio da seção 4, o método de Newton convergiu para a solução em apenas cinco iterações, enquanto que os demais métodos convergiram com mais de cinco iterações. Para este caso ilustramos o decréscimo do valor da função a cada iteração (Imagem 02), da norma do gradiente (Imagem 03) e, por fim, o caminho percorrido pelo algoritmo ao longo das iterações até a obtenção do minimizador do problema (Imagens 04, 05 e 06).

De modo geral, para os problemas testados, os métodos Quase-Newton apresenteram desempenho inferior uma vez que apresentaram maior número de iterações na resolução dos problemas. Porém, é preciso mencionar que cada método de otimização possui propriedades particulares, de modo que, seu desempenho está associado às características do problema em estudo.

Sendo assim, de acordo com os conceitos pesquisados, os métodos de otimização abordados nesse trabalho são uma alternativa para a resolução de problemas de otimização irrestrita, pois fornecem boas aproximações para a solução de tais problemas, que dificilmente seriam calculados de maneira analítica.

REFERÊNCIAS

IZMAILOV, A.; SOLODOV, M. **Otimização: métodos computacionais**. Volume 2. Rio de Janeito: IMPA, 2007.

MARTINEZ, J. M.; SANTOS, S. **A. Métodos computacionais de otimização**. [S.l.]: Departamento de Matemática Aplicada, IMECC - UNICAMP, 1998.

RIBEIRO, A. A.; KARAS, E. W. **Otimização Contínua: Aspectos teóricos e computacionais**. São Paulo: Cengage Learning, 2013.