https://eventos.utfpr.edu.br//sicite/sicite2020

Métodos de otimização: resolução de problemas de programação não-linear.

Optimization method: solving nonlinear programming problems.

Paulo Roberto Machado Silva Junior
probertomsjr@gmail.com
Universidade Tecnológica Federal do Paraná, Campo Mourão, Paraná, Brasil.

Tatiane Cazarin da Silva
tatianecazarin@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Campo Mourão, Paraná, Brasil.

Recebido: 19 ago. 2020.
Aprovado: 01 out. 2020.

RESUMO

A otimização matemática é um estudo que vem sempre utilizado afim de solucionar problemas, sejam esses relacionados a redução em gastos com processamentos, materiais e recursos em atividades, de natureza fabril ou matemática. Esse estudo, tem como objetivo trabalhar métodos acerca da otimização não-linear irrestrita e de como determinar a solução ótima de uma função. A otimização não-linear, como o próprio nome diz, não possui um padrão a ser seguido em suas direções de minimização e nas escolhas do tamanho de passo. Já a solução ótima, é responsável por gerar valores à função, que serão capazes de maximizar ou minimizar a função objetivo. No presente trabalho, serão discutidos os principais métodos de otimização não-linear irrestritos: Gradiente, Newton, Direções Conjugadas e Quase-Newton. Também é abordado como a escolha do tamanho de passo e do método de descida, influenciam diretamente ao tempo de execução e posteriormente a convergência da função objetivo.

PALAVRAS-CHAVE. Otimização. Não-linear. Métodos de otimização.

ABSTRACT

The math optimization is the study that come used for solve problems, be that linked the reduction of processing costs, stuffs and activity resources, from factory or mathematical nature. This study has goal work meth about unrestricted nonlinear optimization and how discover the optimum solution about the function. The nonlinear optimization, as the name says, doesn't have patter to be followed at its minimization direction and at choose about size step. The optimum solution is responsible for generation values that be able to maximize or minimize the object function. In this work will be talk about the mains methods from unrestricted nonlinear optimization: Gradient, Newton, Conjugated Directions and Quasi-Newton. Also is approached how the step choose and the descent method, directly influence in the execution time and, posteriorly the convergence about objective function.

KEYWORDS: Optimization. Nonlinear. Optimization methods.

INTRODUÇÃO

As técnicas de otimização podem ser utilizadas quando não existe uma solução simples e diretamente calculável para o problema. Isso geralmente ocorre quando a estrutura do problema é complexa, ou existem milhões de possíveis soluções (VIALI, 2016).

Um problema geral de otimização consiste em encontrar um vetor de coordenadas finais que satisfaçam as condições de minimizadores ou maximizadores da função trabalhada (ANDRETTA, 2010).

Neste contexto, considere o problema de otimização definido por minimizar $f(x)$, sujeita a $x \in S$, em que $f: R^n \rightarrow R$ e $S \subset R^n$ e S é o conjunto factível.

Definição 1. Um ponto $x^* \in S$ é um minimizador local de f $\leftrightarrow$ existe $\varepsilon > 0$ / $f(x) \geq f(x^*)$ $\forall\, x \in S$/ $||x - x^*|| < \varepsilon$.

Quando a desigualdade acima é válida para todo $x \in S$, será tratado como um minimizador global ou absoluto. Se as desigualdades sujeitas a $x \in S$ forem estritas para $x \neq x^*$, será dito que que x^*é um minimizador estrito. Se não for mencionado no conjunto S, significa que $S = R^n$ e, portanto, deverá ser tratado como um problema irrestrito (RIBEIRO e KARAS, 2014).

Existem algumas condições que valem ser apresentadas para que ocorra a otimização da função objetivo. O Teorema 1 e o Teorema 2, apresentam essas condições.

Teorema 1 (Condição necessária de 1º ordem). *Seja $f: R^n \rightarrow R$ diferenciável no ponto $x^* \in R^n$. Se x^*é um minimizador local de f, então,*

$$\nabla f(x^*) = 0. \tag{1}$$

Um ponto x^* que cumpre a condição proposta pelo teorema, é dito ponto crítico ou ponto estacionário da função.

Teorema 2 (Condição necessária de 2º ordem). *Seja $f: R^n \rightarrow R$ duas vezes diferenciável no ponto $x^* \in IR^n$. Se x^*é um minimizador local de f, então a matriz Hessiana de f, dada por $\nabla^2 f(x^*)$, semidefinida positiva, isto é,*

$$d^T \nabla^2 f(x^*) d \geq 0 \tag{2}$$

para todo $d \in R^n$.

Convém ressaltar que se x^*é minimizador local de f, então, dado uma distância d e um tamanho de passo t, pode-se afirmar que:

$$f(x^*) \leq f(x^* + \mathrm{td}). \tag{3}$$

Existem distintos meios de calcular o tamanho do passo e a direção de descida que será dada para chegar em um ponto que atenda as condições previstas.

TAMANHO DO PASSO

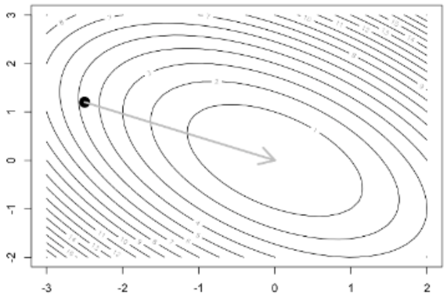
Calcular o tamanho do passo pode ser algo complexo, para alguns tipos de funções, isso influencia diretamente na convergência do algoritmo. Neste trabalho, serão utilizadas a Busca Exata e a Condição de Armijo.

A Busca Exata tem como objetivo minimizar a função $f\left(x^k + t_k d^k\right)$ em t_k, dada a direção de descida. Através da Figura 1(a), é possível observar que é possível chegar ao valor ótimo da função em apenas uma iteração.

A condição de Armijo, por outro lado, utiliza métodos inexatos para obtenção do t_k. Para obtenção do t_k, por Armijo, dados $x^* \in IR^n$, $d \in IR^n$, e $\gamma, \eta \in (0,1)$ o objetivo é testar $t = \gamma t$ até que $f(x^* + td)$> $f(x^*) + \eta t \nabla f(x^*)^T d$. Através da Figura 1(b), é possível observar que foram necessárias algumas iterações para obtenção do tamanho de passo, t.

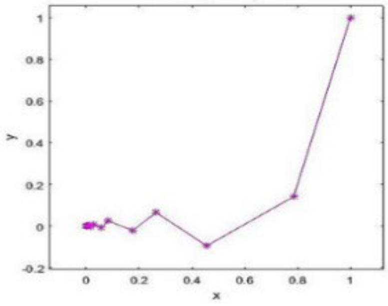
Figura 1 - Tamanho do passo t calculado

(a) por Busca Exata



Fonte: PENG (2020).

(b) pela Condição de Armijo



Fonte: O autor (2020).

Note que, diferente da Busca Exata, a condição de Armijo pode não presar pela exatidão.

DIREÇÃO DE DESCIDA

Fundamental na determinação dos algoritmos, a escolha da direção é o que afeta diretamente na convergência.

Teorema 3. Se $\nabla f(x^*)^T d < 0$, então d é uma direção de descida para f, a partir de x^*.

Satisfazendo a condição do Teorema 3, é possível determinar infinitas direções de descidas. Sendo assim, os diferentes métodos se distinguem pela forma como calcular a direção, d^k. O Algoritmo 1 apresenta o algoritmo geral de descida.

Algoritmo 1- Algoritmo geral de descida.
Dado: $x^0 \in IR^n$,
Repita enquanto $\nabla f\left(x^k\right) \neq 0$
  Defina a direção d^k
  Escolha $t_k > 0$ tal que $f\left(x^k + t_k d^k\right) < f\left(x^k\right)$
  Faça $x^k = x^k + t_k d^k$
$k = k + 1$

Os métodos basicamente se diferenciam pela determinação da direção d^k:

- Quando d^k = - $\nabla f\left(x^k\right)$ temos o método de Cauchy ou método do Gradiente, que consiste em fazer uma busca na direção oposta do gradiente da função objetivo no ponto corrente.

- Quando d^k = - $(\nabla^2 f\left(x^k\right))^{-1}$ $\nabla f\left(x^k\right)$ temos o método de Newton, que visa minimizar uma função f através de uma aproximação local por uma função quadrática e resolver sistemas $\nabla f(x) = 0$, muitas vezes não lineares.

- Quando d^k = - $\nabla f\left(x^{k+1}\right) + \beta_k d^k$ temos o método de Direções Conjugadas. É preciso definir a primeira direção descida por meio do método de Cauchy, e as sucessivas descidas pela inclusão de uma nova combinação do parâmetro β_k, dado por $\beta_k = \frac{(d^k)^T A\, \nabla f(x^{k+1})}{(d^k)^T A d^k}$.

- Quando d^k = - $H_k \nabla f\left(x^k\right)$, onde $H_k \in\ R^{n\,x\,n}$ e $H_k > 0$ temos o método Quase-Newton surgiu com a proposta de não calcular a matriz Hessiana, mas sim, realizar uma aproximação iterativa de sua inversa. Existem variações deste método e, essas variações consistem basicamente, na forma de calcular H_k, a saber Métodos DFP e BDGS.

A partir do estudo de tais métodos, será apresentado algumas discussões numéricas a seguir.

RESULTADO E DISCUSSÕES

O objetivo desse trabalho é fazer com que o leitor tenha uma ideia de como a escolha do método; dos pontos iniciais; da forma de obter o tamanho do passo influenciam diretamente na convergência da função objetivo.

O ideal para que se obtenha o minimizador é fazer com que a condição de parada do algoritmo seja $\nabla f\left(x^k\right) = 0$. Atender essa condição se torna muito custoso, computacionalmente, para certas funções. A condição adotada foi $\nabla f\left(x^k\right) \leq 10^{-4}$. Na obtenção do tamanho do passo por Armijo, foi adotado γ = 0,8 e η = 0,5. Vale destacar que os algoritmos foram projetados em Matlab.

A Tabela 1, apresenta os resultados obtidos para algumas funções.

Tabela 1 - Resultados obtidos a partir da aplicação de funções nos algoritmos estudados

Função objetivo	Direção de descida	Tipo do passo	x^0	k	t(s)
1ª $f(x,y) = 5 + x^2 + 4y^2$	Cauchy	Busca linear exata	[1,1]	11	1,3577
		Armijo	[1,1]	20	2,3527
	Newton	Busca linear exata	[1,1]	1	0,1326
		Armijo	[1,1]	1	0,03727
	Direção Conjugadas_Bk	Busca linear exata	[1,1]	2	0,1509
		Armijo	[1,1]	11	1,3659
	Quase-Newton DFP	Busca linear exata	[1,1]	2	0,2988
		Armijo	[1,1]	5	0,5363
	Quase-Newton BFGS	Busca linear exata	[1,1]	7	14,5616
		Armijo	[1,1]	5	1,3965
2ª $f(x,y) = 2x^2 + 2y^2$	Cauchy	Busca linear exata	[1,1]	1	0,1369
		Armijo	[1,1]	6	0,6270
	Newton	Busca linear exata	[1,1]	1	0,1297
		Armijo	[1,1]	1	0,0375
	Direção Conjugadas_Bkfr	Busca linear exata	[1,1]	1	0,092
		Armijo	[1,1]	4	0,5828
	Quase-Newton DFP	Busca linear exata	[1,1]	1	0,1492
		Armijo	[1,1]	2	0,1715
	Quase-Newton BFGS	Busca linear exata	[1,1]	1	0,1652
		Armijo	[1,1]	5	0,6527
3ª $f(x,y) = \frac{2(xy)^2}{x^2 + y}$	Cauchy	Busca linear exata	[2,2]	-	
		Armijo	[2,2]	6	1,3924
	Newton	Busca linear exata	[2,2]	-	-
		Armijo	[2,2]	3	0,2469
	Dreção Conjugadas_Bfrpr	Busca linear exata	[2,2]	-	-
		Armijo	[2,2]	4	0,4726
	Quase-Newton DFP	Busca linear exata	[2,2]	-	-
		Armijo	[2,2]	4	3,3694
	Quase-Newton BFGS	Busca linear exata	[2,2]	-	-
		Armijo	[2,2]	4	3,577
Legenda:	- Metodo não aplicado para a função.				

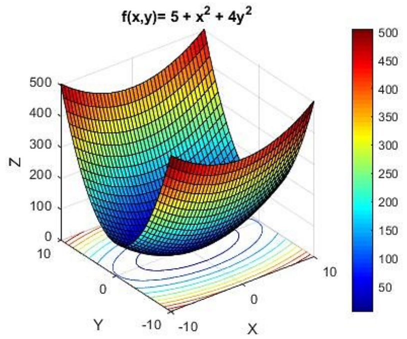
Fonte: O autor (2020).

Como a 3ª função não é quadrática, logo, a escolha do passo por Busca Exata poderia se tornar inviável, computacionalmente. Por esse motivo, o método da Busca Exata, foi aplicado apenas para funções quadráticas.

Analisando os resultados contidos na Tabela 1, é possível afirmar que: (i) Nem sempre o maior número de iterações significa maior tempo de execução; e (ii) Funções quadráticas tem garantia de convergência para todos os métodos estudados, de acordo com o Teorema 2.

Após a apresentação dos resultados numéricos, será apresentada uma discussão gráfica. Para tal, foi a 1ª função, cujo gráfico e curvas de nível podem ser vistas no gráfico da Figura 2.

Figura 2 - Representação gráfica da 1ª função

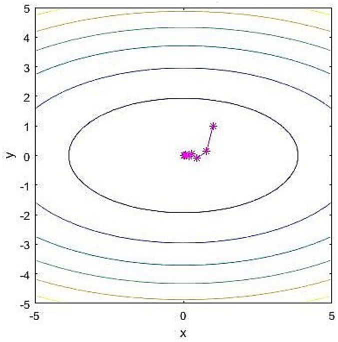


Fonte: O autor (2020).

A Figura 3(a), apresenta o caminho feito pelo algoritmo até chegar ao minimizador, com a escolha de passo dada pela Condição de Armijo. E na Figura 3(b) temos uma aproximação de f, onde, é possível observar melhor o caminho percorrido. Ao todo, foram necessárias 20 iterações para alcançar o mínimo. A escolha do passo por Armijo, resulta em poder caminhar em qualquer direção, formando ângulos distintos entre si, desde que essa seja uma direção minimizadora qualquer.

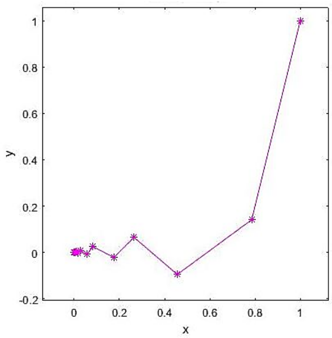
Figura 3 – Convergência do Algoritmo de Cauchy para a função
$f(x, y) = 5 + x^2 + 4y^2$

(a) Definido pela Condição de Armijo



Fonte: O autor (2020).

(b) Aproximação gráfica "zoom"

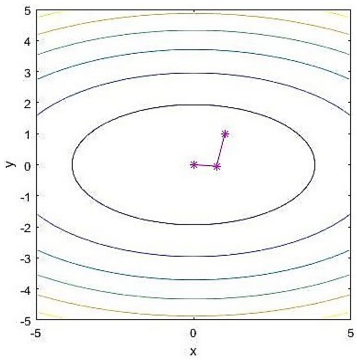


Fonte: O autor (2020).

Para fazer uma comparação gráfica, tem-se a junção do Método de Direções Conjugadas com tamanho de passo dado por Busca Linear Exata, na Figura 4.

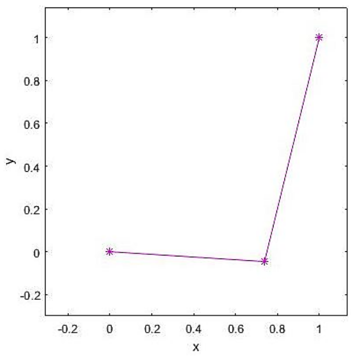
Figura 4 – Convergência do Método Gradientes Conjugados para a função
$f(x, y) = 5 + x^2 + 4y^2$

(a) Definido pela Busca Linear Exata



Fonte: O autor (2020).

(b) Aproximação gráfica "zoom"



Fonte: O autor (2020).

Essa convergência se deu de forma bem sucinta pois, esse método acelera a convergência no método de Cauchy, logo, essa combinação mostrou ser uma das mais melhores a ser feita.

CONCLUSÃO

Nem sempre é possível alcançar os minimizadores, uma vez que, atingi-los pode depender de diversos fatores, dentre eles: a escolha do passo; o método de descida; e um dos mais importantes, os pontos iniciais.

Pontos iniciais mal estabelecidos, podem levar a divergência dos métodos, em oposição, pontos iniciais bem estabelecidos, podem levar a convergência em poucas iterações. Contudo, a obtenção de minimizadores pode se tornar uma tarefa fácil ou muito árdua, o que irá influenciar diretamente é a função objetivo. Portanto, deve-se trabalhar com funções objetivos bem modeladas.

Com o presente estudo, foi concebível o entendimento e compreensão de alguns métodos e meios aplicáveis à otimização de problemas não lineares.

REFERÊNCIAS

ANDRETTA, M. **Método de Busca Linear**. USP. 2010. Disponível em: https://sites.icmc.usp.br/andretta/ensino/aulas/sme5720-2-10/buscalinear.pdf. Acesso em 20 jun. 2020.

PENG, R**. Advanced Statistical Computing**. University of California. 2020. Disponível em: https://bookdown.org/rdpeng/advstatcomp/. Acesso em 10 jun. 2020.

RIBEIRO, A. A.; KARAS, E. W. **Otimização Contínua: Aspectos teóricos e computacionais**. Curitiba: Cengage Learning, 2014.

VIALI, L. **Pesquisa operacional, otimização não linear, conceitos básicos**. PUCRS. 2016.Disponível em: http://www.pucrs.br/ciencias/viali/especializa/po_famat/material/laminaspi/Esp_Famat_3.pdf. Acesso em 21 mai. 2020.