8. Métodos de Descida

8.1. Método do Gradiente (descida máxima)

Considere-se a função $f(x)$ desenvolvida em série de Taylor, desprezando-se os termos de ordem ≥ 2 :

$$f(\mathbf{x}+\delta\,\mathbf{x}) = f(\mathbf{x}) + \nabla^T f(\mathbf{x})\delta\,\mathbf{x}$$

Deseja-se estabelecer uma seqüência de valores do vetor **x**, que convirja para que $\mathbf{x}^* / f(\mathbf{x}^*)$ seja mínimo. Denominando-se $\mathbf{x}^{(k)}$ como o k-ésimo valor do vetor **x** e fazendo $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\,\mathbf{x}^{(k)}$, tem-se :

$$f\left(\mathbf{x}^{(k+1)}\right) = f\left(\mathbf{x}^{(k)}\right) + \nabla^T f\left(\mathbf{x}^{(k)}\right)\left(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right)$$

Se a sequência $\{\mathbf{x}^{(k)}\}$ é estabelecida no sentido de minimizar $f(\mathbf{x})$ e, se além disso, deseja-se uma redução máxima possível entre dois valores consecutivos de $f(\mathbf{x})$, deve-se ter :

$$\max\left\{f\left(\mathbf{x}^{(k)}\right) - f\left(\mathbf{x}^{(k+1)}\right)\right\} = \max\left\{-\nabla^T f\left(\mathbf{x}^{(k)}\right)\delta\,\mathbf{x}^{(k)}\right\}$$

i.e., $-\nabla^T f\left(\mathbf{x}^{(k)}\right)$ e $\delta\,\mathbf{x}^{(k)}$ devem ser colineares :

$$\delta\,\mathbf{x}^{(k)} = -\alpha\,\nabla f\left(\mathbf{x}^{(k)}\right)$$

ou ainda :

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha^{(k)}\nabla f\left(\mathbf{x}^{(k)}\right) \text{ (eq. de busca)}$$

onde:

$\alpha^{(k)}$: passo (k-ésima estimativa)

Portanto, a direção de busca que produz a máxima descida é $d^{(k)} = -\nabla f(x^{(k)})$, no ponto $x = x^{(k)}$. O cálculo de $d^{(k)}$ corresponde ao $1^{\underline{0}}$ passo do método.

O $2^{\underline{0}}$ passo consiste em determinar $\alpha^{(k)} \geq 0 / f(\mathbf{x}^{(k)} +^{(k)} \mathbf{d}^{(k)})$ seja mínimo. Esse processo é chamado de “busca linear exata”. O $3^{\underline{0}}$ passo corresponde à aplicação da equação de busca, concluindo assim a k-ésima iteração.

Substituindo $\delta\mathbf{x}_p = -\alpha \nabla f(\mathbf{x}_p)$ na expansão em série de Taylor:

$$f(\mathbf{x}_p) - f(\mathbf{x}_{p+1}) = \alpha \nabla^T f(\mathbf{x}_p).\nabla f(\mathbf{x}_p)$$

Logo:

$\alpha > 0 \Rightarrow$ minimização

$\alpha < 0 \Rightarrow$ maximização

A escolha do passo α é portanto de fundamental importância para a convergência do método.

8.1.1 Exemplo:

Minimizar $f(x) = x_1^2 - 2x_1 + x_2^2$

O vetor $\mathbf{x}$ é dado por:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

O cálculo das derivadas parciais leva a:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 - 2 \\ 2x_2 \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Como $\mathbf{x}^{\mathrm{T}} \nabla^2 f(\mathbf{x}) \mathbf{x} \geq 0 \Rightarrow f$ possui um mínimo e a condição $\nabla f(x) = 0$ é necessária e suficiente para garantir que este é o mínimo global (*f* é convexa). Aplicando então essa condição, obtém-se:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

onde o asterisco indica a situação de otimalidade.

Aplicando o método do gradiente para resolver o problema acima, a fim de estudar o seu desempenho quanto à convergência, tem-se:

$$\mathbf{x}^{(p+1)} = \mathbf{x}^{(p)} - \alpha \begin{bmatrix} 2x_1 - 2 \\ 2x_2 \end{bmatrix}^{(p)}$$

Adotar-se-á como ponto de partida o vetor:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{(0)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Em seguida, alguns valores de α serão testados:

a) $\alpha = 1{,}5$

<u>1ª iteração:</u>

$$x_1^{(1)} = x_1^{(0)} - 1{,}5.\left(\frac{\partial f}{\partial x_1}\right)^{(0)} = -1 - 1{,}5.(-4) = 5$$

$$x_2^{(1)} = x_2^{(0)} - 1{,}5.\left(\frac{\partial f}{\partial x_2}\right)^{(0)} = 1 - 1{,}5.(2) = -2$$

<u>2ª iteração:</u>

$$x_1^{(2)} = x_1^{(1)} - 1.,5\left(\frac{\partial f}{\partial x_1}\right)^{(1} = -7$$

$$x_2^{(2)} = x_2^{(1)} - 1.,5\left(\frac{\partial f}{\partial x_2}\right)^{(1)} = 4$$

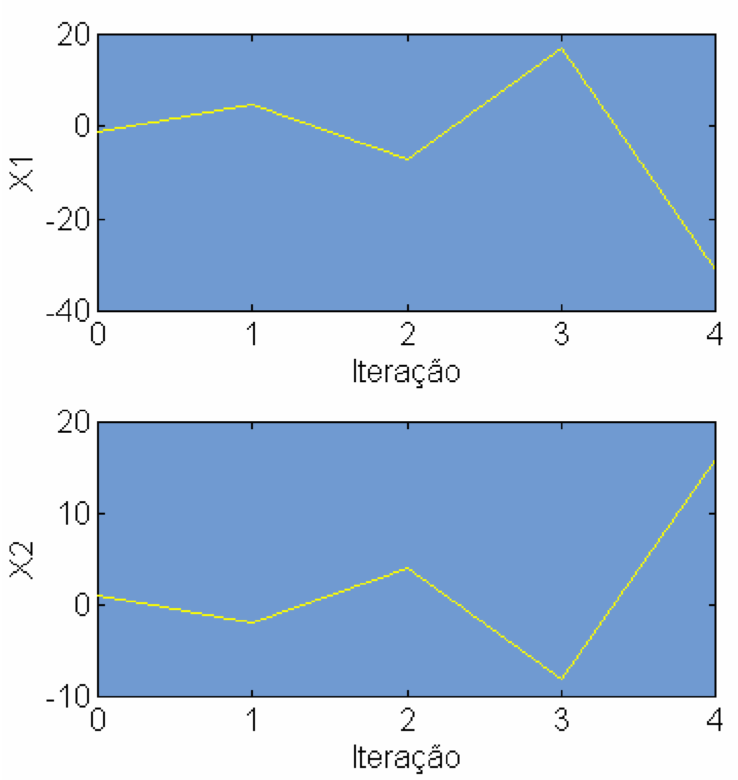
<u>3ª iteração:</u>

$$x_1^{(3)} = -7 - 1{,}5.(-16) = 17$$
$$x_2^{(3)} = 4 - 1{,}5.(8) = -8$$

<u>4ª iteração:</u>

$$x_1^{(4)} = 17 - 1{,}5.(32) = -31$$
$$x_2^{(4)} = -8 - 1{,}5.(-16) = 16$$

Os gráficos de x_1 e x_2, em função do contador de iterações, são mostrados abaixo:



b) $\alpha = 0{,}75$:

Depois de efetuados os cálculos, obtém-se:

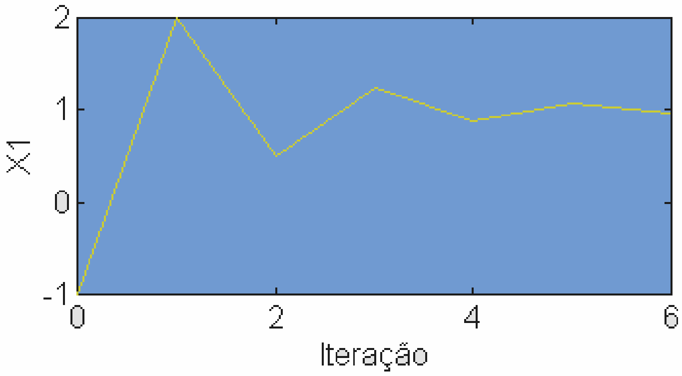
$x_1^{(1)} = 2; \quad x_2^{(1)} = -0{,}5$

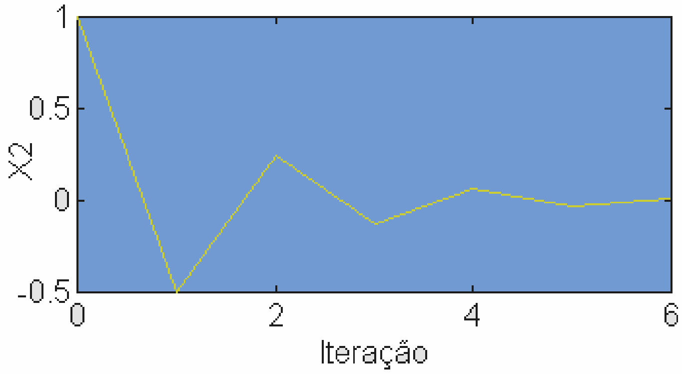
$x_1^{(2)} = 0{,}5; \quad x_2^{(2)} = 0{,}25$

$x_1^{(3)} = 1{,}25; \quad x_2^{(3)} = -0{,}125$

$x_1^{(4)} = 0{,}875; \quad x_2^{(4)} = 0{,}0625$

Esses valores correspondem aos gráficos abaixo:





Nesse caso, nota-se que, após algumas iterações, a convergência é obtida para $x_1 = 1$ e $x_2 = 0$. Observa-se que, se o valor de α for escolhido no intervalo $0 < \alpha < 0{,}5$, a convergência é obtida de forma tangencial; se, entretanto, adotar-se $\alpha = 0{,}5$, apenas uma iteração é suficiente para alcançar a solução:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{\text{ot}} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} - 0{,}5.\begin{bmatrix} -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

8.1.2 Função objetivo quadrática

Seja $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{C}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{x} + \mathrm{a}$

Conforme calculado na seção 5.3.4, o vetor gradiente $\nabla f(\mathbf{x})$ é dado por:

$$\nabla f(\mathbf{x}) = 2\mathbf{C}\mathbf{x} + \mathbf{b}$$

A busca linear exata exige que:

$$\nabla f(\mathbf{x}+\alpha.\mathbf{d}) = 2\mathbf{Cx} + 2\alpha.\mathbf{Cd} + \mathbf{b} = \mathbf{0}$$

ou ainda:

$$2\alpha.\mathbf{Cd} = -\nabla f(\mathbf{x})$$

Ou:

$$2\alpha \cdot \begin{bmatrix} c_{11}d_1 + \cdots + c_{1n}d_n \\ \vdots \\ c_{n1}d_1 + \cdots + c_{nn}d_n \end{bmatrix} = -\begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

Multiplicando ambos os membros por $\mathbf{d}^{\mathrm{T}}$ e explicitando α, obtém-se:

$$\alpha = \frac{-1}{2}\frac{\mathbf{d}^{\mathrm{T}}.\nabla f(\mathbf{x})}{\mathbf{d}^{\mathrm{T}}.\mathbf{C}.\mathbf{d}}$$

Como no método do gradiente $\mathbf{d} = -\nabla f(\mathbf{x})$, tem-se:

$$2\alpha.\mathbf{C}\nabla f(\mathbf{x}) = \nabla f(\mathbf{x})$$

$$\alpha = \frac{1}{2}\frac{\nabla^{\mathrm{T}} f(\mathbf{x}).\nabla f(\mathbf{x})}{\nabla^{\mathrm{T}} f(\mathbf{x}).\mathbf{C}.\nabla f(\mathbf{x})}$$

8.2 Método de Newton

8.2.1 Função quadrática

Seja *f* uma função quadrática com matriz hessiana $\mathbf{G} = 2.\mathbf{C}$, definida positiva. Partindo de $\mathbf{x}^{(0)} \in \Re^n$ arbitrário, a direção $\mathbf{d} \in \Re^n$ dada por:

$$\mathbf{d} = -\mathbf{G}^{-1}(\mathbf{G}.\mathbf{x}^{(0)} + \mathbf{b}) = -\mathbf{G}^{-1}\nabla f(\mathbf{x}^{(0)})$$

determina diretamente o minimizador global de *f*, através de:

$\mathbf{x}^{\mathrm{ot}} = \mathbf{x}^{(0)} + \mathbf{d}$

Demonstração:

Seja $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{G}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{x} + \mathrm{a}$. Então $\nabla f(\mathbf{x}) = \mathbf{G}\mathbf{x} + \mathbf{b}$ e, para $\mathbf{x}^{ot} = \mathbf{x}^{(0)} + \mathbf{d}$, tem-se:

$$\nabla f(\mathbf{x}^{\mathrm{ot}}) = \mathbf{G}(\mathbf{x}^{(0)} + \mathbf{d}) + \mathbf{b} = \mathbf{G}[\mathbf{x}^{(0)} - \mathbf{G}^{-1}(\mathbf{G}\mathbf{x}^{(0)} + \mathbf{b})] + \mathbf{b} = \mathbf{G}\mathbf{x}^{(0)} - (\mathbf{G}\mathbf{x}^{(0)} + \mathbf{b}) + \mathbf{b} = \mathbf{0}$$

Portanto, $\mathbf{x}^{ot}$ é um minimizador global. Notar que, para determinar $\mathbf{x}^{ot}$, é suficiente encontrar a direção de busca **d**, o que equivale a resolver o sistema linear dado por:

$$\mathbf{G}\mathbf{d} = -\nabla f(\mathbf{x}^{(0)})$$

partindo de qualquer ponto $\mathbf{x}^{(0)}$.

8.2.2 Função não-quadrática

Nesse caso, considera-se uma aproximação quadrática para a função, entre dois pontos consecutivos da seqüência $\{\mathbf{x}^{(k)}\}$:

$$f(\mathbf{x}^{(\mathrm{k})} + \mathbf{d}^{(k)}) = f(\mathbf{x}^{(\mathrm{k})}) + \nabla^{\mathrm{T}} f(\mathbf{x}^{(\mathrm{k})})\mathbf{d}^{(\mathrm{k})} + \frac{1}{2}\mathbf{d}^{(\mathrm{k}),\mathrm{T}}\nabla^{2} f(\mathbf{x}^{(\mathrm{k})})\mathbf{d}^{(\mathrm{k})}$$

Essa expressão define uma função quadrática $q(\mathbf{d})$, caracterizada por:

$q(\mathbf{d}) = f\left(\mathbf{x}^{(\mathrm{k})} + \mathbf{d}\right)$

t.q.:

$$\begin{cases} q(\mathbf{0}) = f(\mathbf{x}^{(k)}) \\ \nabla q(\mathbf{0}) = \nabla f(\mathbf{x}^{(k)} + \mathbf{d})\Big|_{\mathbf{d}=\mathbf{0}} = \nabla f(\mathbf{x}^{(k)}) \\ \nabla^2 q(\mathbf{0}) = \nabla^2 q(\mathbf{d})\Big|_{\mathbf{d}=\mathbf{0}} = \nabla^2 f(\mathbf{x}^{(k)}) \end{cases}$$

sendo:

$$q(\mathbf{d}) = f(\mathbf{x}^{(k)}) + \nabla^{\mathrm{T}} f(\mathbf{x}^{(k)})\mathbf{d}^{(k)} + \frac{1}{2}\mathbf{d}^{(k),\mathrm{T}}\nabla^2 f(\mathbf{x}^{(k)})\mathbf{d}^{(k)}$$

$$q(\mathbf{d}) = q(\mathbf{0}) + \nabla^{\mathrm{T}} q(\mathbf{0})\mathbf{d}^{(k)} + \frac{1}{2}\mathbf{d}^{(k),\mathrm{T}}\nabla^2 q(\mathbf{0})\mathbf{d}^{(k)}$$

De acordo com o teorema do item anterior, a direção ótima de busca do mínimo de q(**d**) é dada por:

$$\mathbf{d}^{\mathrm{ot}} = -\left[\nabla^2 f\left(\mathbf{x}^{(k)}\right)\right]^{-1}.\left\{\nabla^2 f\left(\mathbf{x}^{(k)}\right).\mathbf{d}^{(0)} + \nabla f\left(\mathbf{x}^{(k)}\right)\right\}$$

Partindo de $\mathbf{d}^{(0)} = \mathbf{0}$, tem-se que a direção ótima de busca para a k-ésima iteração é dada por:

$$\mathbf{d}^{(k)} = -\left[\nabla^2 f\left(\mathbf{x}^{(k)}\right)\right]^{-1}.\nabla f\left(\mathbf{x}^{(k)}\right)$$

8.2.3 Algoritmo

Seja $\mathbf{x}^{(k)}$ tal que $\nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$. O ponto $\mathbf{x}^{(k+1)}$ pode ser obtido através dos seguintes passos:

1 Determinar $\mathbf{d}^{(k)}$ de acordo com:

$$\nabla^2 f\left(\mathbf{x}^{(k)}\right).\mathbf{d}^{(k)} = -\nabla f\left(\mathbf{x}^{(k)}\right)$$

2 Determinar $\alpha^{(k)}$ através de uma busca linear.

3 Calcular $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}.\mathbf{d}^{(k)}$

4 Retornar ao passo 1, caso não se tenha atingido a convergência.

8.3 Métodos quase-Newton

Os métodos do gradiente e de Newton possuem direções de busca definidas por:

$$\mathbf{d}^{(k)} = -\mathbf{H}^{(k)}.\nabla f\left(\mathbf{x}^{(k)}\right)$$

sendo que, no método do gradiente, $\mathbf{H}^{(k)} = \mathbf{E}$ (matriz identidade); no método de Newton, $\mathbf{H}^{(k)} = [\nabla^2 f(\mathbf{x}^{(k)})]^{-1}$.

A grande vantagem do método de Newton sobre o método do gradiente reside no fato de sua equação de busca ser derivada de uma aproximação quadrática da função objetivo; no método do gradiente, essa aproximação é linear. Isso faz com que a convergência do método de Newton seja mais rápida. Em contrapartida, o cálculo de sua direção de busca, a cada iteração, requer a avaliação da matriz hessiana, bem como a solução de um sistema de equações lineares.

Os métodos quase-Newton também possuem direção de busca definida pela equação acima, sendo que a matriz **H** não chega a ser tão simples como a matriz identidade, nem tão completa como a matriz hessiana. É evidente que, para se obter uma convergência melhor que a do método do gradiente, alguma informação de segunda ordem precisa ser utilizada. Deve-se ainda atentar para o fato de que, se $\mathbf{H}^{(k)}$ é simétrica e definida positiva, a direção $\mathbf{d}^{(k)}$ é uma direção de descida.

8.3.1 Método de Davidon-Fletcher-Powell

Sejam o vetor $\mathbf{x}^{(0)}$ e a matriz definida positiva $\mathbf{H}^{(0)}$, arbitrários. Se $\nabla f(\mathrm{x}^{(k)}) \neq 0$, uma melhor estimativa de $\mathbf{x}^{ot}$, digamos $\mathbf{x}^{(k+1)}$, pode ser obtida através dos seguintes passos:

1 Calcular $\mathbf{d}^{(k)} = -\mathbf{H}^{(k)}.\nabla f\left(\mathbf{x}^{(k)}\right)$

2 Determinar $\alpha^{(k)}$ através de uma busca linear e calcular $\mathbf{x}^{(k+1)} = \mathbf{x}^{k} + \alpha^{(k)} \cdot \mathbf{d}^{(k)}$.

Sucessivas melhores estimativas podem ser obtidas através de uma atualização da matriz **H**, de acordo com o passo seguinte.

3 Definindo $\mathbf{p}^{(k)} = \alpha^{(k)}.\mathbf{d}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$, $\mathbf{q}^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$, calcular:

$$\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} + \frac{\mathbf{p}^{(k)}.\mathbf{p}^{(k)T}}{\mathbf{p}^{(k)T}.\mathbf{q}^{(k)}} - \frac{\mathbf{H}^{(k)}\mathbf{q}^{(k)}.\mathbf{q}^{(k)T}\mathbf{H}^{(k)}}{\mathbf{q}^{(k)T}\mathbf{H}^{(k)}\mathbf{q}^{(k)}}$$

O número de multiplicações envolvidas em cada atualização da matriz H, pode ser determinado como segue:

$p.p^T : n^2$

$p^T.q : n$

$\mathbf{q}^{(k)T}\mathbf{H}^{(k)}\mathbf{q}^{(k)} : n^2 + n$

$(\mathbf{H}.\mathbf{q}).\mathbf{q}^T.\mathbf{H} : (n^2) + n^2 + n^2 = 2n^2$

Total por iteração : $4n^2 + 2n$

Uma vez que o cálculo de G^{-1}, no método de Newton, envolve n^3 multiplicações, para problemas de grande porte, o método de Davidon-Fletcher-Powell consome muito menos tempo computacional. Pode-se ainda mostrar, para *f* quadrática, que:

- Se $H^{(k)}$ é positiva definida, $H^{(k+1)}$ também o é.
- Os vetores $\mathbf{d}^{(0)},...,\mathbf{d}^{(n-1)}$ são L.I.
- $\mathbf{x}^{(n)} = \mathbf{x}^{ot}$
- $\mathbf{H}^{(n)} = \mathbf{G}^{-1}$

8.3.2 Método de Broyden-Fletcher-Goldfarb-Shanno

$$\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} + \left(1 + \frac{\mathbf{q}^{(k)T}\mathbf{H}^{(k)}\mathbf{q}^{(k)}}{\mathbf{p}^{(k)T}.\mathbf{q}^{(k)}}\right)\frac{\mathbf{p}^{(k)}.\mathbf{p}^{(k)T}}{\mathbf{p}^{(k)T}.\mathbf{q}^{(k)}} - \frac{\mathbf{p}\mathbf{q}^{(k)T}\mathbf{H}^{(k)} + \mathbf{H}^{(k)}\mathbf{q}^{(k)}\mathbf{p}^{T}}{\mathbf{p}^{(k)T}\mathbf{q}^{(k)}}$$