

7. Programação Não-Linear: Problemas Irrestritos

Os problemas de otimização tratados neste capítulo são da forma:

$$\begin{aligned} &\text{Minimizar } f(\mathbf{x}) \\ &\text{sujeita a : } \begin{cases} \mathbf{x} \in \Omega \subset \mathbb{R}^n \\ f : \mathbb{R}^n \rightarrow \mathbb{R} \end{cases} \end{aligned}$$

7.1. Definições Preliminares

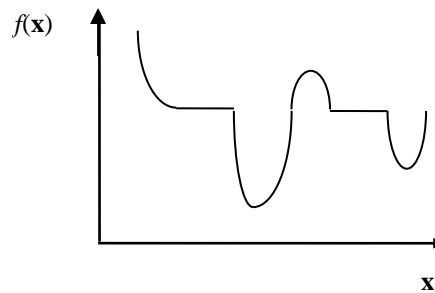
7.1.1. Mínimo Relativo ou Local

$\mathbf{x}^* \in \Omega$ é chamado de ponto mínimo relativo ou local, se $\exists \varepsilon > 0 / f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \Omega, / \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$. Se, nessas condições, tem-se sempre $f(\mathbf{x}^*) < f(\mathbf{x}), \forall \mathbf{x} \in \Omega$, diz-se que \mathbf{x}^* é um ponto de mínimo relativo “estrito”.

7.1.2. Mínimo Absoluto ou Global

O ponto $\mathbf{x}^* \in \Omega$ é chamado de ponto de mínimo absoluto ou global, se $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \Omega$. Caso se tenha sempre $f(\mathbf{x}^*) < f(\mathbf{x}), \forall \mathbf{x} \in \Omega$, diz-se que \mathbf{x}^* é um ponto de mínimo absoluto “estrito”.

Ex.:



7.2. Propriedades Básicas das Soluções

7.2.1. Condição necessária de 1ª. ordem

“Seja f de classe C^1 (contínua, com derivadas de 1ª. ordem contínuas) uma função definida em Ω . Se \mathbf{x}^* é um ponto de mínimo relativo de f em Ω , então:

$$\nabla^T f(\mathbf{x}^*) \cdot \delta \mathbf{x} \geq 0, \forall \delta \mathbf{x} \in \mathbb{R}^n (\mathbf{x}^* + \delta \mathbf{x}) \in \Omega$$

Nessa expressão, define-se o transposto do vetor gradiente de f , calculado em $\mathbf{x} = \mathbf{x}^*$, como:

$$\nabla^T f(\mathbf{x}^*) = \left[\frac{\partial f}{\partial x_1} \quad \dots \quad \frac{\partial f}{\partial x_n} \right]_{\mathbf{x} = \mathbf{x}^*}$$

Demonstração : Expandindo $f(\mathbf{x})$ em série de Taylor, tem-se:

$$f(\mathbf{x}^* + \delta \mathbf{x}) = f(\mathbf{x}^*) + \nabla^T f(\mathbf{x}^*) \cdot \delta \mathbf{x} + \sigma$$

Desprezando σ (i.e., os termos de ordem igual ou superior a 2), tem-se:

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

Pela definição de mínimo relativo:

$$f(\mathbf{x}^* + \delta \mathbf{x}) - f(\mathbf{x}^*) \geq 0$$

Então:

$$\nabla^T f(\mathbf{x}^*) \cdot \delta \mathbf{x} \geq 0 !!!$$

Particularmente, se \mathbf{x}^* é um ponto interior de Ω , $\exists \delta \mathbf{x} > \mathbf{0}$ ou $\delta \mathbf{x} < \mathbf{0}$ / $(\mathbf{x}^* + \delta \mathbf{x}) \in \Omega$.

Logo,

$$\nabla^T f(\mathbf{x}^*) = \mathbf{0} !!!$$

Essa condição é chamada de condição de estacionaridade.

7.2.2. Condição necessária de 2ª. ordem

Seja f uma função de classe C^2 (C^1 + derivadas de 2ª. ordem contínuas), definida em Ω . Se \mathbf{x}^* é um ponto de mínimo relativo de f em Ω , com $\nabla^T f(\mathbf{x}^*) = \mathbf{0}$, então:

$$\delta \mathbf{x}^T \nabla^2 f(\mathbf{x}^*) \delta \mathbf{x} \geq 0, \quad \forall \delta \mathbf{x} \in \mathcal{R}^n \mid (\mathbf{x}^* + \delta \mathbf{x}) \in \Omega$$

Nessa expressão, define-se a Matriz Hessiana, $\nabla^2 f(\mathbf{x}^*)$, como:

$$\nabla^2 f(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Demonstração: Considerando a expansão em série de Taylor da função f e escolhendo $\delta \mathbf{x}$ pequeno o suficiente para que se possam considerar desprezíveis os termos de ordem igual ou superior a 3, dessa expansão, tem-se:

$$f(\mathbf{x}^* + \delta \mathbf{x}) = f(\mathbf{x}^*) + \nabla^T f(\mathbf{x}^*) \cdot \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{x}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \delta \mathbf{x}$$

Como $\nabla^T f(\mathbf{x}^*) = \mathbf{0}$, a expressão acima fica:

$$f(\mathbf{x}^* + \delta \mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2} \delta \mathbf{x}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \delta \mathbf{x}$$

Entretanto, $f(\mathbf{x}^* + \delta \mathbf{x}) - f(\mathbf{x}^*) \geq 0$. Logo:

$$\delta \mathbf{x}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \delta \mathbf{x} \geq 0 !!!$$

Verifica-se, portanto, que $\nabla^2 f(\mathbf{x}^*)$ é semi-definida positiva na vizinhança de \mathbf{x}^* .

Caso \mathbf{x}^* seja um ponto de mínimo local estrito, tem-se:

$$\delta \mathbf{x}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \delta \mathbf{x} > 0$$

i. e., $\nabla^2 f(\mathbf{x}^*)$ é definida positiva.

7.2.3. Condição suficiente de 2ª. ordem

Seja $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$, $f \in C^2$. Se para $\forall \mathbf{x}^* \in \mathfrak{R}^n$ valem as proposições $\nabla f(\mathbf{x}^*) = \mathbf{0}$ e $\mathbf{x}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \mathbf{x} > \mathbf{0}$, então \mathbf{x}^* é um minimizador local estrito de f em \mathfrak{R}^n .

Demonstração:

Seja $\Omega = \{\mathbf{h} \in \mathfrak{R}^n \mid \|\mathbf{h}\| = 1\}$. Seja ainda $g : \Omega \rightarrow \mathfrak{R}$ tal que:

$$g(\mathbf{h}) = \mathbf{h}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \mathbf{h}$$

(Com base na 2ª. hipótese, $g(\mathbf{h}) > \mathbf{0}$)

Uma vez que g é uma função contínua definida no conjunto fechado e limitado Ω , então g possui um máximo e um mínimo. Chamando de a o valor mínimo da função, tem-se:

$$0 < a \leq g(\mathbf{h}), \forall \mathbf{h} \in \Omega \quad (\#)$$

Tomando agora um vetor $\mathbf{d} \in \mathfrak{R}^n$, arbitrário e não-nulo, tem-se que:

$$\frac{\mathbf{d}}{\|\mathbf{d}\|} \in \Omega$$

Então:

$$g\left(\frac{\mathbf{d}}{\|\mathbf{d}\|}\right) = \left[\frac{\mathbf{d}}{\|\mathbf{d}\|}\right]^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \left[\frac{\mathbf{d}}{\|\mathbf{d}\|}\right] = \frac{1}{\|\mathbf{d}\|^2} \mathbf{d}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \mathbf{d} \quad (\#\#)$$

Combinando as eqs. (#) e (\#\#), tem-se:

$$\frac{1}{\|\mathbf{d}\|^2} \mathbf{d}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \mathbf{d} \geq a \quad (\#\#\#)$$

Desenvolvendo agora f em série de Taylor, em torno de \mathbf{x}^* , para uma variação igual a \mathbf{d} , e desprezando os termos de ordem ≥ 3 , tem-se:

$$f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) = \nabla^T f(\mathbf{x}^*) \cdot \mathbf{d} + \frac{1}{2} \mathbf{d}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \mathbf{d}$$

Combinando essa equação com a 1ª. Hipótese, fica:

$$f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) = \frac{1}{2} \mathbf{d}^T \cdot \nabla^2 f(\mathbf{x}^*) \cdot \mathbf{d}$$

$$f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) \geq \frac{a \|\mathbf{d}\|^2}{2}$$

Para \mathbf{d} suficientemente pequeno, o fato de se ter aproximado o desenvolvimento em série de Taylor não afeta o sinal do 2º membro, de forma que:

$$f(\mathbf{x}^* + \mathbf{d}) - f(\mathbf{x}^*) \geq 0$$

Portanto, para qualquer $\mathbf{x} = \mathbf{x}^* + \mathbf{d}$ na vizinhança de \mathbf{x}^* , tem-se que $f(\mathbf{x}^*) \leq f(\mathbf{x})$, o que prova que \mathbf{x}^* é um minimizador local.

7.3. Funções Convexas

7.3.1. Definição

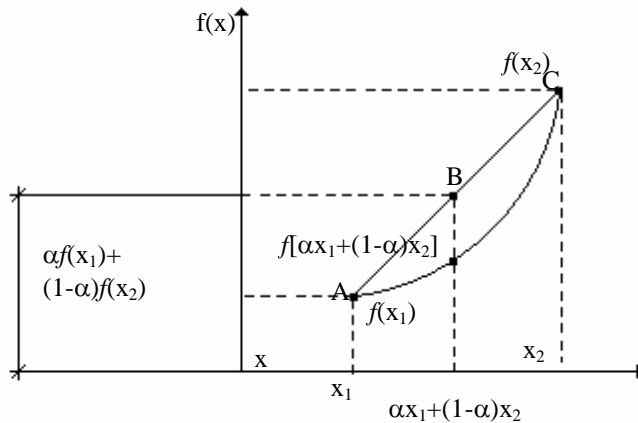
Diz-se que uma função f , definida em um conjunto convexo Ω é convexa, se $\forall \mathbf{x}, \mathbf{y} \in \Omega$ e $\forall \alpha \in \mathbb{R} / 0 \leq \alpha \leq 1$, tem-se:

$$f[\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}] \leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$$

Se $f[\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}] < \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$, diz-se que f é estritamente convexa.

7.3.2. Interpretação gráfica

Tomemos como exemplo $f: \mathbb{R} \rightarrow \mathbb{R}$, como mostrada abaixo :



No estudo de conjuntos convexos, mostrou-se que o ponto

$$\alpha \begin{bmatrix} x_1 \\ f(x_1) \end{bmatrix} + (1 - \alpha) \begin{bmatrix} x_2 \\ f(x_2) \end{bmatrix}$$

pertence ao segmento de reta que une os pontos A e C, conforme figura ao lado (ponto B). Qualquer função f definida em $[x_1, x_2]$, cujo valor em $x_1 \leq x \leq x_2$ seja inferior à

ordenada do ponto correspondente no segmento de reta AC, é dita *convexa*.

7.3.3. Propriedades

Seja $f: \Omega \rightarrow \mathbb{R}$.

a) Se f é convexa de classe $C^1 \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$.

Demonstração : se f é convexa \Rightarrow

$$\begin{aligned} f[\alpha \mathbf{y} + (1 - \alpha)\mathbf{x}] &\leq \alpha f(\mathbf{y}) + (1 - \alpha)f(\mathbf{x}) \\ f[\alpha \mathbf{y} + (1 - \alpha)\mathbf{x}] &\leq f(\mathbf{x}) + \alpha[f(\mathbf{y}) - f(\mathbf{x})] \\ \frac{f[\alpha \mathbf{y} + (1 - \alpha)\mathbf{x}] - f(\mathbf{x})}{\alpha} &\leq [f(\mathbf{y}) - f(\mathbf{x})] \end{aligned}$$

Por Taylor:

$$\begin{aligned} f[\alpha \mathbf{y} + (1 - \alpha)\mathbf{x}] &= f[\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})] = f(\mathbf{x}) + \nabla^T f(\mathbf{x}) \cdot \alpha(\mathbf{y} - \mathbf{x}) + \sigma \\ 0 \leq \alpha \leq 1, \quad \sigma &\rightarrow 0 \end{aligned}$$

Logo, tem-se:

$$\frac{f[\alpha \mathbf{y} + (1 - \alpha)\mathbf{x}] - f(\mathbf{x})}{\alpha} = \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$$

Ou seja:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) !$$

b) Se f é convexa de classe $C^2 \Leftrightarrow \nabla^2 f(\mathbf{x})$ é semi-definida positiva.

Demonstração:

Por Taylor :

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x})(\mathbf{y} - \mathbf{x}), \text{ com } \sigma \rightarrow 0$$

Da propriedade anterior, tem-se:

$$f(\mathbf{y}) - f(\mathbf{x}) - \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geq 0$$

Portanto:

$$\frac{1}{2}(\mathbf{y} - \mathbf{x})^T \nabla^2 f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geq 0$$

c) Como consequência da propriedade anterior, $\nabla f(\mathbf{x}^*)=0$ é condição necessária e suficiente para mínimo local.

d) Se f é convexa definida em um conjunto convexo Ω então um mínimo local \mathbf{x}^* é mínimo global.

Demonstração:

Suponha que $\exists \mathbf{y} \in \Omega$ tal que $f(\mathbf{y}) < f(\mathbf{x}^*)$. Considerando um vetor \mathbf{z} como uma combinação convexa de \mathbf{y} e \mathbf{x}^* . tem-se:

$$\begin{aligned} f(\mathbf{z}) &= f[\alpha \mathbf{y} + (1 - \alpha)\mathbf{x}^*] \leq \alpha f(\mathbf{y}) + (1 - \alpha)f(\mathbf{x}^*) \\ f(\mathbf{z}) &\leq f(\mathbf{x}^*) + \alpha[f(\mathbf{y}) - f(\mathbf{x}^*)] \end{aligned}$$

Como $0 \leq \alpha \leq 1$, pode-se supor, particularmente, $\alpha \rightarrow 0$, o que definiria um vetor \mathbf{z} nas imediações de \mathbf{x}^* . Nesse caso, a desigualdade acima só poderia ser verdadeira, para $f(\mathbf{y}) > f(\mathbf{x}^*)$, i.e.:

$$f(\mathbf{z}) \leq f(\mathbf{x}^*) + \epsilon \text{ (com } \epsilon > 0, \text{ o que requer } f(\mathbf{y}) > f(\mathbf{x}^*) \text{).}$$

Portanto, se \mathbf{x}^* é minimizador local, $\nexists \mathbf{z} \in \Omega / f(\mathbf{z}) \leq f(\mathbf{x}^*)$ e $\nexists \mathbf{y} \in \Omega / f(\mathbf{y}) < f(\mathbf{x}^*)$. Como consequência \mathbf{x}^* é **minimizador global**.

e) Seja f **convexa** de classe C^1 definida em Ω (**convexo**). Se existe $\mathbf{x}^* / \forall \mathbf{y} \in \Omega$ a relação de desigualdade $\nabla^T f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geq 0$ seja verdadeira, então \mathbf{x}^* é minimizador global de f em Ω .

Demonstração : do item (a) tem-se:

$$f(\mathbf{y}) - f(\mathbf{x}^*) \geq \nabla^T f(\mathbf{x}^*)(\mathbf{y} - \mathbf{x}^*) \geq 0 \text{ (por hipótese)}$$

Como $f(\mathbf{y}) \geq f(\mathbf{x}^*) \forall \mathbf{y} \in \Omega$, então \mathbf{x}^* é minimizador global.

7.3.4. Funções quadráticas

Caso particular de funções convexas em que:

$$f(\mathbf{x}) = a + \mathbf{b}^T \mathbf{x} + \mathbf{x}^T \mathbf{C} \mathbf{x}, \text{ sendo } \mathbf{C} \text{ simétrica.}$$

A condição necessária e suficiente para existência de mínimo local é:

$$\nabla f(\mathbf{x}^*) = 0 = \frac{\partial}{\partial \mathbf{x}} (\mathbf{b}^T \mathbf{x} + \mathbf{x}^T \mathbf{C} \mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}^*}$$

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{b}^T \mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} (b_1 x_1 + \dots + b_n x_n) = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \mathbf{b}$$

Cálculo de $\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{C} \mathbf{x})$:

$$\mathbf{C} \mathbf{x} = \begin{bmatrix} c_{11}x_1 + \dots + c_{1n}x_n \\ \vdots \\ c_{n1}x_1 + \dots + c_{nn}x_n \end{bmatrix}; \mathbf{x}^T \mathbf{C} \mathbf{x} = \sum_{i=1}^n x_i \sum_{j=1}^n c_{ij} x_j$$

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = y = x_1 (c_{11}x_1 + \dots + c_{1j}x_j + \dots + c_{1n}x_n)$$

$$+ \dots$$

$$\vdots$$

$$x_j (c_{j1}x_1 + \dots + c_{jj}x_j + \dots + c_{jn}x_n)$$

$$+ \dots$$

$$\vdots$$

$$x_n (c_{n1}x_1 + \dots + c_{nj}x_j + \dots + c_{nn}x_n)$$

$$\frac{\partial \mathbf{x}^T \mathbf{C} \mathbf{x}}{\partial x_j} = x_1 c_{1j} + \dots + x_n c_{nj} + c_{j1}x_1 + \dots + 2c_{jj}x_j + \dots + c_{jn}x_n$$

$$\text{com } c_{ij} = c_{ji}$$

$$\frac{\partial \mathbf{x}^T \mathbf{C} \mathbf{x}}{\partial x_j} = 2 \sum_{i=1}^n c_{ij} x_i$$

$$\frac{\partial \mathbf{x}^T \mathbf{C} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{C} \mathbf{x}$$

Então;

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

Se \mathbf{C}^{-1} existe, então $\mathbf{x}^* = -\frac{1}{2} \mathbf{C}^{-1} \mathbf{b}$. Além disso, \mathbf{x}^* é um ponto de mínimo global, pois $\mathbf{x}^T \mathbf{C} \mathbf{x} \geq 0$

7.4. Buscas Direcionais

7.4.1. Direções de descida

Se $\forall \nabla f(\mathbf{x}) \neq \mathbf{0}$, sabe-se que \mathbf{x} não é um minimizador local de f . Portanto, $\exists \mathbf{y}$ na vizinhança de \mathbf{x} / $f(\mathbf{y}) < f(\mathbf{x})$. A proposição seguinte ajuda a identificar se uma determinada direção \mathbf{d} / $\mathbf{y} = \mathbf{x} + \alpha \mathbf{d}$, leva a uma diminuição no valor de f .

Teorema :

Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função de classe C^1 , $\mathbf{x} \in \mathbb{R}^n$ / $\forall \nabla f(\mathbf{x}) \neq \mathbf{0}$ e $\mathbf{d} \in \mathbb{R}^n$ / $\nabla^T f(\mathbf{x}) \cdot \mathbf{d} < 0$. Então $\exists \alpha > 0$ / $f(\mathbf{x} + \alpha \mathbf{d}) < f(\mathbf{x})$.

Demonstração :

Seja a função $\phi(\alpha) = f(\mathbf{x} + \alpha \mathbf{d})$. Então $\phi(0) = f(\mathbf{x})$ e:

$$\phi'(0) = \nabla^T f(\mathbf{x} + \alpha \mathbf{d}) \Big|_{\alpha=0} \cdot \mathbf{d} = \nabla^T f(\mathbf{x}) \cdot \mathbf{d} < 0$$

$$\text{Mas } \phi'(0) = \lim_{\alpha \rightarrow 0} \frac{\phi(\alpha) - \phi(0)}{\alpha} < 0$$

Com $\alpha > 0$, o sinal de $\phi'(0)$ é igual ao sinal da diferença:

$$\phi(\alpha) - \phi(0) = f(\mathbf{x} + \alpha \mathbf{d}) - f(\mathbf{x}).$$

Como $\phi'(0) < 0$, então $f(\mathbf{x} + \alpha \mathbf{d}) < f(\mathbf{x})$.

Portanto, qualquer direção definida por $\nabla^T f(\mathbf{x}) \cdot \mathbf{d} < 0$ é uma direção de “descida” (diminuição da função f), a partir de \mathbf{x} .

7.4.2. Modelo de algoritmo de busca direcional

Seja \mathbf{x}^* um minimizador local de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $\mathbf{x}^{(k)}$ a sua k -ésima estimativa.

Modelo do algoritmo :

Passo 1 : Identificar uma direção $\mathbf{d}^{(k)} \in \mathbb{R}^n$ / $\nabla^T f(\mathbf{x}^{(k)}) \cdot \mathbf{d}^{(k)} < 0$.

Passo 2 : Determinar $\alpha^{(k)} > 0$ / $f(\mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)}) < f(\mathbf{x}^{(k)})$. Este problema é conhecido como “busca linear” e define o tamanho do passo $\alpha^{(k)}$.

Passo 3 : Calcular $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)}$ e retornar ao passo 1, enquanto $\nabla f(\mathbf{x}^{(k)}) \neq \mathbf{0}$.

8. Métodos de Descida

8.1. Método do Gradiente (descida máxima)

Considere a função $f(\mathbf{x})$ desenvolvida em série de Taylor, desprezando 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 \mathbf{x} , que convirja para \mathbf{x}^* / $f(\mathbf{x}^*)$ seja mínimo. Denominando $\mathbf{x}^{(k)}$ como o k -ésimo valor do vetor \mathbf{x} e fazendo $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta \mathbf{x}^{(k)}$, tem-se (por Taylor):

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

Se a seqüência $\{\mathbf{x}^{(k)}\}$ 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:

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

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

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

ou ainda :

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

onde:

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

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

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