



Gabarito PNL - Algoritmos e Métodos de Otimização

Kelvin (Pol)

Exercício 1.

Considere $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2}x_1^2 + x_2$ e a sequência (x^k) , gerada pelo algoritmo de descida com $H(x) = I$ e busca exata, a partir do ponto $x^0 = (1, 1)$. Prove que $x_1^k = (-1)^k$; $x_2^k = 1 - 2k$.

Solução:

Note que

$$\nabla f(x) = \begin{pmatrix} x_1 \\ 1 \end{pmatrix}.$$

Como $H(x) = I$, a direção de descenso é

$$d^k = -H(x^k)\nabla f(x^k) = -\nabla f(x^k) = \begin{pmatrix} -x_1^k \\ -1 \end{pmatrix}.$$

Obtemos t_k por busca exata, isto é, minimizando

$$\varphi(t) = f(x^k + td^k), \quad t > 0.$$

Temos

$$x^k + td^k = \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} + t \begin{pmatrix} -x_1^k \\ -1 \end{pmatrix} = \begin{pmatrix} x_1^k(1-t) \\ x_2^k - t \end{pmatrix}.$$

Assim,

$$\varphi(t) = f(x^k + td^k) = \frac{1}{2}(x_1^k)^2(1-t)^2 + x_2^k - t.$$

Derivando, obtemos

$$\varphi'(t) = -(x_1^k)^2(1-t) - 1.$$

Se $\varphi'(t) = 0$, temos

$$-(x_1^k)^2(1-t) - 1 = 0,$$

o que implica

$$(x_1^k)^2(1-t) = -1 \implies t_k = 1 + \frac{1}{(x_1^k)^2}.$$

Dai atualizamos

$$x^{k+1} = x^k + t_k d^k.$$

Para a primeira coordenada,

$$x_1^{k+1} = x_1^k + t_k(-x_1^k) = x_1^k \left(1 - 1 - \frac{1}{(x_1^k)^2} \right) = -\frac{1}{x_1^k}.$$

Logo,

$$x_1^{k+1} = -\frac{1}{x_1^k}.$$

Avaliando em $k = 0$, e como $x^0 = (1, 1)$, temos

$$x_1^1 = -1.$$

Por indução,

$$x_1^2 = -\frac{1}{x_1^1} = 1, \quad x_1^3 = -1, \quad \dots$$

Assim, concluímos que

$$x_1^k = (-1)^k.$$

Para a segunda coordenada,

$$x_2^{k+1} = x_2^k + t_k(-1) = x_2^k - \left(1 + \frac{1}{(x_1^k)^2} \right).$$

Mas como $(x_1^k)^2 = 1$ para todo k , temos

$$x_2^{k+1} = x_2^k - 2.$$

Avaliando em $k = 0$, como $x_2^0 = 1$, obtemos

$$x_2^1 = -1, \quad x_2^2 = -3, \quad x_2^3 = -5, \quad \dots$$

Logo,

$$x_2^k = 1 - 2k.$$

Portanto,

$$x_1^k = (-1)^k \quad \text{e} \quad x_2^k = 1 - 2k. \odot$$

Exercício 2.

Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2}x^T A x + b^T x + c$, onde $A \in \mathbb{R}^{n \times n}$ é uma matriz definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$.

- (a) Mostre que se $\nabla f(x)^T d = 0$, então a função cresce a partir de x ao longo de d ;
- (b) Suponha que d é uma direção de descida a partir de x . Mostre que a busca exata fornece $t^* = -\frac{\nabla f(x)^T d}{d^T Ad}$;
- (c) Mostre que se t^* satisfaz a condição de Armijo

$$f(x + t^* d) \leq f(x) + \eta t^* \nabla f(x)^T d,$$

então $\eta \leq \frac{1}{2}$.

Solução:

- (a) Fixemos $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$, e consideremos $t \in \mathbb{R}$. Temos

$$f(x + td) = \frac{1}{2}(x + td)^T A(x + td) + b^T(x + td) + c.$$

Desenvolvendo,

$$(x + td)^T A(x + td) = x^T Ax + tx^T Ad + td^T Ax + t^2 d^T Ad.$$

Como A é simétrica, vale $x^T Ad = d^T Ax$, então

$$(x + td)^T A(x + td) = x^T Ax + 2tx^T Ad + t^2 d^T Ad.$$

Assim,

$$f(x + td) = \frac{1}{2}x^T Ax + tx^T Ad + \frac{1}{2}t^2 d^T Ad + b^T x + tb^T d + c.$$

Agrupando os termos, obtemos

$$f(x + td) = f(x) + t(Ax + b)^T d + \frac{1}{2}t^2 d^T Ad.$$

Mas $\nabla f(x) = Ax + b$, logo

$$f(x + td) = f(x) + t\nabla f(x)^T d + \frac{1}{2}t^2 d^T Ad.$$

Por hipótese temos que $\nabla f(x)^T d = 0$, então

$$f(x + td) = f(x) + \frac{1}{2}t^2 d^T Ad.$$

Como A é definida positiva, temos $d^T Ad > 0$ para todo $d \neq 0$, daí

$$f(x + td) > f(x), \quad \forall t \neq 0.$$

Logo, a função cresce a partir de x ao longo da direção d .

- (b) Suponha agora que d é direção de descida em x . A busca exata consiste em minimizar

$$\varphi(t) = f(x + td), \quad t > 0.$$

Da expressão obtida acima,

$$\varphi(t) = f(x) + t\nabla f(x)^T d + \frac{1}{2}t^2 d^T Ad.$$

Derivando em relação a t , temos

$$\varphi'(t) = \nabla f(x)^T d + t d^T Ad.$$

Anulando a derivada, obtemos

$$\nabla f(x)^T d + t d^T Ad = 0,$$

o que implica

$$t^* = -\frac{\nabla f(x)^T d}{d^T Ad}.$$

Como A é definida positiva, temos $d^T Ad > 0$, e como d é direção de descida, $\nabla f(x)^T d < 0$, logo $t^* > 0$, o que conclui.

- (c) Fixemos $x \in \mathbb{R}^n$ e suponhamos que d é direção de descida em x . Do item anterior, temos

$$f(x + td) = f(x) + t\nabla f(x)^T d + \frac{1}{2}t^2 d^T Ad.$$

Da busca exata, obtivemos

$$t^* = -\frac{\nabla f(x)^T d}{d^T Ad}.$$

Substituindo t^* na expressão de $f(x + td)$, temos

$$f(x + t^* d) = f(x) + t^* \nabla f(x)^T d + \frac{1}{2}(t^*)^2 d^T Ad.$$

Agora,

$$t^* \nabla f(x)^T d = -\frac{(\nabla f(x)^T d)^2}{d^T Ad},$$

e

$$(t^*)^2 d^T A d = \frac{(\nabla f(x)^T d)^2}{(d^T A d)^2} d^T A d = \frac{(\nabla f(x)^T d)^2}{d^T A d}.$$

Logo,

$$f(x + t^* d) = f(x) - \frac{(\nabla f(x)^T d)^2}{d^T A d} + \frac{1}{2} \frac{(\nabla f(x)^T d)^2}{d^T A d} = f(x) - \frac{1}{2} \frac{(\nabla f(x)^T d)^2}{d^T A d}.$$

Por outro lado,

$$f(x) + \eta t^* \nabla f(x)^T d = f(x) - \eta \frac{(\nabla f(x)^T d)^2}{d^T A d}.$$

Se t^* satisfaz a condição de Armijo,

$$f(x + t^* d) \leq f(x) + \eta t^* \nabla f(x)^T d,$$

então

$$f(x) - \frac{1}{2} \frac{(\nabla f(x)^T d)^2}{d^T A d} \leq f(x) - \eta \frac{(\nabla f(x)^T d)^2}{d^T A d}.$$

Cancelando $f(x)$ e observando que

$$\frac{(\nabla f(x)^T d)^2}{d^T A d} > 0,$$

pois $d^T A d > 0$ e $(\nabla f(x)^T d)^2 > 0$, obtemos

$$-\frac{1}{2} \leq -\eta,$$

o que implica

$$\eta \leq \frac{1}{2}.$$

Logo, se t^* satisfaz a condição de Armijo, necessariamente $\eta \leq \frac{1}{2}$. ☺

Exercício 3.

Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2} x^T A x + b^T x + c$, onde $A \in \mathbb{R}^{n \times n}$ é uma matriz definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Sejam x^* o minimizador de f e $v \in \mathbb{R}^n$ um autovetor de A . Faça uma busca exata a partir do ponto $x = x^* + v$, na direção $d = -\nabla f(x)$. Que ponto é obtido? Qual é a interpretação geométrica deste exercício?

Solução:

Seja $f(x) = \frac{1}{2} x^T A x + b^T x + c$, com A definida positiva, e seja x^* o minimizador de f . Então, como f é estritamente convexa, x^* é o único ponto crítico, e portanto

$$\nabla f(x^*) = 0,$$

equivalentemente,

$$A x^* + b = 0.$$

Seja agora v um autovetor de A associado ao autovalor $\lambda > 0$, isto é,

$$A v = \lambda v.$$

Consideremos o ponto $x = x^* + v$. Então

$$\nabla f(x) = A(x^* + v) + b = A x^* + A v + b.$$

Como $A x^* + b = 0$, obtemos

$$\nabla f(x) = A v = \lambda v.$$

Por hipótese, $d = -\nabla f(x)$, logo

$$d = -\lambda v.$$

Fazemos agora busca exata a partir de x na direção d . Pelo Exercício 2, sabemos que

$$t^* = -\frac{\nabla f(x)^T d}{d^T A d}.$$

Calculando,

$$\begin{cases} \nabla f(x)^T d = (\lambda v)^T (-\lambda v) = -\lambda^2 v^T v, \\ d^T A d = (-\lambda v)^T A (-\lambda v) = \lambda^2 v^T A v = \lambda^2 v^T (\lambda v) = \lambda^3 v^T v. \end{cases}$$

Logo,

$$t^* = -\frac{-\lambda^2 v^T v}{\lambda^3 v^T v} = \frac{1}{\lambda}.$$

Atualizando,

$$x^{novo} = x + t^* d = x^* + v + \frac{1}{\lambda} (-\lambda v) = x^*.$$

Portanto, o ponto obtido após a busca exata é exatamente o minimizador x^* .

A interpretação geométrica é que, ao nos deslocarmos a partir de x^* na direção de um autovetor de A , o gradiente permanece colinear com essa mesma direção. Assim, o método do gradiente com busca exata escolhe um passo que compensa exatamente o deslocamento inicial, retornando ao minimizador em uma única iteração. ☺

Exercício 4.

Considere $A \in \mathbb{R}^{n \times n}$ uma matriz definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$f(x) = \frac{1}{2}x^T A x + b^T x + c.$$

Pede-se:

(a) Mostre que a busca exata na direção $d \in \mathbb{R}^n$, a partir de $\bar{x} \in \mathbb{R}^n$, fornece

$$t^* = -\frac{\nabla f(\bar{x})^T d}{d^T A d}.$$

(b) Mostre que se $d = -\nabla f(\bar{x})$ é um autovetor de A , então uma iteração do algoritmo do gradiente com busca exata, a partir de \bar{x} , encontra o minimizador global de f .

Solução:

Temos

$$f(x) = \frac{1}{2}x^T A x + b^T x + c,$$

com $A \succ 0$. Logo,

$$\nabla f(x) = Ax + b \quad \text{e} \quad \nabla^2 f(x) = A.$$

Como A é definida positiva, f é estritamente convexa e possui minimizador global único, caracterizado por

$$Ax + b = 0.$$

(a) Fixe $\bar{x} \in \mathbb{R}^n$ e uma direção $d \in \mathbb{R}^n$. Definimos

$$\varphi(t) = f(\bar{x} + td).$$

Então

$$\varphi(t) = \frac{1}{2}(\bar{x} + td)^T A(\bar{x} + td) + b^T(\bar{x} + td) + c.$$

Desenvolvendo,

$$\varphi(t) = \frac{1}{2}\bar{x}^T A \bar{x} + t d^T A \bar{x} + \frac{1}{2}t^2 d^T A d + b^T \bar{x} + t b^T d + c.$$

Agrupando os termos em t ,

$$\varphi(t) = \text{constante} + t d^T (A \bar{x} + b) + \frac{1}{2}t^2 d^T A d.$$

Como $A \bar{x} + b = \nabla f(\bar{x})$, temos

$$\varphi(t) = \text{constante} + t d^T \nabla f(\bar{x}) + \frac{1}{2}t^2 d^T A d.$$

Derivando em relação a t ,

$$\varphi'(t) = d^T \nabla f(\bar{x}) + t d^T A d.$$

Na busca exata impomos $\varphi'(t) = 0$, obtendo

$$d^T \nabla f(\bar{x}) + t^* d^T A d = 0.$$

Logo,

$$t^* = -\frac{\nabla f(\bar{x})^T d}{d^T A d}.$$

(b) Suponha agora que

$$d = -\nabla f(\bar{x})$$

e que d é autovetor de A . Então existe $\lambda > 0$ tal que

$$A d = \lambda d.$$

Como $d = -\nabla f(\bar{x})$, segue que

$$A \nabla f(\bar{x}) = \lambda \nabla f(\bar{x}).$$

Pela parte (a), o passo exato na direção d é

$$t^* = -\frac{\nabla f(\bar{x})^T d}{d^T A d}.$$

Substituindo $d = -\nabla f(\bar{x})$,

$$t^* = -\frac{\nabla f(\bar{x})^T (-\nabla f(\bar{x}))}{(-\nabla f(\bar{x}))^T A (-\nabla f(\bar{x}))}.$$

Então

$$t^* = \frac{\|\nabla f(\bar{x})\|^2}{\nabla f(\bar{x})^T A \nabla f(\bar{x})}.$$

Como $A \nabla f(\bar{x}) = \lambda \nabla f(\bar{x})$, temos

$$\nabla f(\bar{x})^T A \nabla f(\bar{x}) = \lambda \|\nabla f(\bar{x})\|^2.$$

Logo,

$$t^* = \frac{1}{\lambda}.$$

O novo ponto gerado pelo método do gradiente com busca exata é

$$x_1 = \bar{x} + t^*d = \bar{x} - \frac{1}{\lambda}\nabla f(\bar{x}).$$

Agora calculamos o gradiente em x_1 :

$$\nabla f(x_1) = Ax_1 + b = A\left(\bar{x} - \frac{1}{\lambda}\nabla f(\bar{x})\right) + b.$$

Então,

$$\begin{aligned}\nabla f(x_1) &= A\bar{x} + b - \frac{1}{\lambda}A\nabla f(\bar{x}) \\ &= \nabla f(\bar{x}) - \frac{1}{\lambda}A\nabla f(\bar{x}) \\ &= \nabla f(\bar{x}) - \frac{1}{\lambda}\lambda\nabla f(\bar{x}) \\ &= 0.\end{aligned}$$

Portanto x_1 é ponto estacionário.

Como $A \succ 0$, a função f é estritamente convexa. Em funções estritamente convexas, todo ponto estacionário é o único minimizador global.

Concluimos que

$$x_1 = x^*$$

é o minimizador global de f . \odot

Exercício 5.

Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in \mathcal{C}^2$ e $\bar{x} \in \mathbb{R}^n$ tal que $\nabla f(\bar{x}) = 0$ e $\nabla^2 f(\bar{x})$ não é semi-definida positiva. Prove que existe uma direção de descida d em \bar{x} .

Solução:

Perceba que $\nabla^2 f(\bar{x})$ não é semi-definida positiva, então existe um vetor d tal que $d^T \nabla^2 f(\bar{x}) d < 0$.

Por outro lado, pelo desenvolvimento de Taylor de segunda ordem em torno de \bar{x} na direção td onde $t \neq 0$, temos

$$f(\bar{x} + td) = f(\bar{x}) + t\nabla f(\bar{x})^T d + \frac{t^2}{2}d^T \nabla^2 f(\bar{x})d + r(t), \quad \text{onde } \frac{r(t)}{t^2} \rightarrow 0.$$

Além disso, $\nabla f(\bar{x}) = 0$, assim

$$f(\bar{x} + td) - f(\bar{x}) = \frac{t^2}{2}d^T \nabla^2 f(\bar{x})d + r(t).$$

Dividindo t^2 ,

$$\frac{f(\bar{x} + td) - f(\bar{x})}{t^2} = \frac{1}{2}d^T \nabla^2 f(\bar{x})d + \frac{r(t)}{t^2}.$$

Tomando limite

$$\begin{aligned}\frac{f(\bar{x} + td) - f(\bar{x})}{t^2} &= \frac{1}{2}d^T \nabla^2 f(\bar{x})d + \frac{r(t)}{t^2}. \\ \lim_{t \rightarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t^2} &= \frac{1}{2}d^T \nabla^2 f(\bar{x})d < 0,\end{aligned}$$

pois $\lim_{t \rightarrow 0} \frac{r(t)}{t^2} = 0$. Pela preservação de sinal, existe um $\delta > 0$ tal que

$$f(\bar{x} + td) < f(\bar{x}), \quad \forall t \in (0, \delta).$$

Portanto, existe uma direção de descida d em \bar{x} . \odot

Exercício 6.

Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2}x^T Ax + b^T x + c$, onde $A \in \mathbb{R}^{n \times n}$ é uma matriz definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Mostre que se $x \neq -A^{-1}b$, então $d = -A^{-1}\nabla f(x)$ é uma direção de descida a partir de x , calcule o tamanho do passo da busca exata, t^* , e mostre que o novo ponto $x^+ = x + t^*d$ é justamente o minimizador de f .

Solução:

Temos

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c,$$

com $A \succ 0$. Logo,

$$\nabla f(x) = Ax + b.$$

Sabemos que o minimizador global é caracterizado por

$$Ax + b = 0 \iff x^* = -A^{-1}b.$$

Suponha agora que $x \neq -A^{-1}b$. Então

$$g := \nabla f(x) = Ax + b \neq 0.$$

Definimos

$$d = -A^{-1}g.$$

d é direção de descida.

Como $A \succ 0$, temos que A é simétrica definida positiva, logo A^{-1} também é simétrica definida positiva (para mostrar isto, usar o teorema espectral).

Calculamos

$$\nabla f(x)^T d = g^T (-A^{-1}g) = -g^T A^{-1}g.$$

Como $A^{-1} \succ 0$ e $g \neq 0$, segue que

$$g^T A^{-1}g > 0.$$

Portanto,

$$\nabla f(x)^T d < 0,$$

o que mostra que d é direção de descida a partir de x .

Cálculo do passo ótimo t^* .

Consideramos

$$\varphi(t) = f(x + td).$$

Pela regra da cadeia,

$$\varphi'(t) = \nabla f(x + td)^T d.$$

Temos

$$\nabla f(x + td) = A(x + td) + b = Ax + b + tAd = g + tAd.$$

Logo,

$$\varphi'(t) = (g + tAd)^T d = g^T d + t d^T Ad.$$

Na busca exata impomos $\varphi'(t) = 0$, isto é,

$$g^T d + t^* d^T Ad = 0.$$

Assim,

$$t^* = -\frac{g^T d}{d^T Ad}.$$

Substituímos $d = -A^{-1}g$.

Primeiro,

$$g^T d = g^T (-A^{-1}g) = -g^T A^{-1}g.$$

Agora,

$$d^T Ad = (-A^{-1}g)^T A(-A^{-1}g) = g^T (A^{-1})^T A A^{-1}g.$$

Como A é simétrica, A^{-1} também é simétrica, logo

$$(A^{-1})^T = A^{-1}.$$

Portanto,

$$d^T Ad = g^T A^{-1}g.$$

Concluimos que

$$t^* = -\frac{-g^T A^{-1}g}{g^T A^{-1}g} = 1.$$

Novo ponto.

O novo ponto é

$$x^+ = x + t^* d = x + d.$$

Então

$$x^+ = x - A^{-1}g = x - A^{-1}(Ax + b).$$

Logo,

$$x^+ = x - x - A^{-1}b = -A^{-1}b.$$

Conclusão.

Temos

$$x^+ = -A^{-1}b,$$

e como $A \succ 0$, a função f é estritamente convexa. Portanto, esse ponto é o único minimizador global de f .

Além disso,

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

Assim, uma única iteração produz exatamente o minimizador global. ☺

Exercício 7.

Considere uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 e suponha que existem constantes $M \geq m > 0$ tais que $\nabla^2 f(x) - mI$ e $M I - \nabla^2 f(x)$ são semi-definidas positivas para todo $x \in \mathbb{R}^n$.

- (a) Prove que para todo $x \in \mathbb{R}^n$ e todo autovalor λ de $\nabla^2 f(x)$ temos $m \leq \lambda \leq M$.
 (b) Prove que existem $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$ satisfazendo

$$f(x) \geq \frac{m}{2} \|x\|^2 + b^T x + c$$

e conclua que f tem um único minimizador global. Sugestão: use a fórmula de Taylor com resto de Lagrange em torno de zero.

- (c) Denotando o minimizador por x^* , prove que

$$f(x) - f(x^*) \geq \frac{m}{2} \|x - x^*\|^2$$

para todo $x \in \mathbb{R}^n$. Sugestão: use a fórmula de Taylor com resto de Lagrange em torno de x^* .

- (d) Considere a sequência $(x^k)_{k \in \mathbb{N}}$ dada por

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

com $t = \frac{1}{M}$. Prove que

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2M} \|\nabla f(x^k)\|^2$$

para todo $k \in \mathbb{N}$.

- (e) Prove que a sequência $(f(x^k))_{k \in \mathbb{N}}$ converge e que $\nabla f(x^k) \rightarrow 0$.
 (f) Prove que $(x^k)_{k \in \mathbb{N}}$ é limitada e que $x^k \rightarrow x^*$.

Solução:

- (a) Seja (λ, v) o auto-par da matriz $\nabla^2 f(x)$ e como as matrizes $\nabla^2 f(x) - mI$ e $MI - \nabla^2 f(x)$ são semidefinidas positivas para todo $x \in \mathbb{R}^n$ onde $M \geq m > 0$, então

$$\begin{cases} v^T (\nabla^2 f(x) - mI) v \geq 0 \implies \|v\|^2 \lambda - m \|v\|^2 \geq 0 \implies \lambda \geq m \\ v^T (MI - \nabla^2 f(x)) v \geq 0 \implies M \|v\|^2 - \lambda \|v\|^2 \geq 0 \implies M \geq \lambda \end{cases}$$

Assim, $M \geq \lambda \geq m$.

- (b) Seja $x \in \mathbb{R}^n$, note que $x^T (\nabla^2 f(x) - mI) x \geq 0$, então temos a desigualdade

$$x^T \nabla^2 f(x) x \geq m \|x\|^2, \quad \forall x \in \mathbb{R}^n. \quad (1)$$

Pelo desenvolvimento de Taylor com resto de Lagrange entorno no ponto 0 na direção x , temos

$$f(x) = f(0) + \nabla f(0)^T x + \frac{1}{2} x^T \nabla^2 f(\xi) x \quad ; \text{ onde } \xi = \theta x, \quad \text{para algum } \theta \in (0, 1).$$

Logo, pela desigualdade (1), tem-se

$$f(x) \geq f(0) + \nabla f(0)^T x + \frac{m}{2} \|x\|^2,$$

definindo $f(0) = c$ e $\nabla f(0) = b$ se consegue

$$f(x) \geq \frac{m}{2} \|x\|^2 + b^T x + c.$$

Além disso, fazendo $\|x\| \rightarrow \infty$ temos $f(x) \rightarrow \infty$, logo f é coerciva, então existe algum minimizador global. Por outro lado, da desigualdade (1), implica que $\nabla^2 f(x)$ é positiva definida (pois $m > 0$), daí, f é estritamente convexa. Portanto f tem um único minimizador global.

- (c) Seja $x \in \mathbb{R}^n$, pelo desenvolvimento de Taylor com resto de Lagrange entorno no ponto x^* na direção $(x - x^*)$, temos

$$f(x^* + (x - x^*)) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2} (x - x^*)^T \nabla^2 f(\xi) (x - x^*);$$

onde $\xi = (1 - \theta)x^* + \theta x$, para algum $\theta \in (0, 1)$.

Suponhamos que x^* é minimizador de f , em particular $\nabla f(x^*) = 0$ e pela desigualdade (1), temos

$$f(x) - f(x^*) \geq \frac{m}{2} \|x - x^*\|^2.$$

- (d) Seja $x \in \mathbb{R}^n$ e como $MI - \nabla^2 f(x)$ é positiva semi-definida, então $x^T (MI - \nabla^2 f(x)) x \geq 0$. Daí,

$$x^T \nabla^2 f(x) x \leq \|x\|^2 M, \quad \forall x \in \mathbb{R}^n. \quad (2)$$

Novamente, pelo desenvolvimento de Taylor com resto de Lagrange entorno no ponto x^k na direção $(x^{k+1} - x^k)$, temos

$$f(x^k + (x^{k+1} - x^k)) = f(x^k) + \nabla f(x^k)^T (x^{k+1} - x^k) + \frac{1}{2} (x^{k+1} - x^k)^T \nabla^2 f(\xi) (x^{k+1} - x^k) \quad , \text{ onde } \xi \text{ fica entre } x^k \text{ e } (x^{k+1} - x^k).$$

Pela desigualdade (2), obtemos

$$f(x^{k+1}) \leq f(x^k) + \nabla f(x^k)^T (x^{k+1} - x^k) + \frac{M}{2} \|x^{k+1} - x^k\|^2,$$

como $x^{k+1} - x^k = -t\nabla f(x^k)$, então

$$f(x^{k+1}) \leq f(x^k) - t\|\nabla f(x^k)\|^2 + \frac{Mt^2}{2}\|\nabla f(x^k)\|^2.$$

Além disso $t = \frac{1}{M}$. Daí,

$$f(x^{k+1}) \leq f(x^k) - \left(\frac{1}{M} - \frac{1}{2M}\right)\|\nabla f(x^k)\|^2.$$

Portanto,

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2M}\|\nabla f(x^k)\|^2$$

para todo $k \in \mathbb{N}$.

- (e) Perceba pelo item (d), f é monótona decrescente e pelo item (b) f é limitada em \mathbb{R} , logo $f(x^k)_{k \in \mathbb{N}}$ é convergente. Por outro lado, do item (d), temos

$$\|\nabla f(x^k)\|^2 \leq 2M \underbrace{[f(x^k) - f(x^{k+1})]}_{\rightarrow 0} \quad ; \forall k \in \mathbb{N}.$$

Pela continuidade da norma e como f é de classe C^2 , fazendo $k \rightarrow \infty$, obtemos $\|\nabla f(x^k)\|^2 \leq 0$, o que implica que $\|\nabla f(x^k)\| \rightarrow 0$. Portanto, $\nabla f(x^k) \rightarrow 0$.

- (f) Pelo item (c), temos que

$$f(x^k) - f(x^*) \geq \frac{m}{2}\|x^k - x^*\|^2, \quad \forall k \in \mathbb{N}.$$

Logo,

$$\|x^k - x^*\|^2 \leq \frac{2}{m}(f(x^k) - f(x^*)).$$

Como pelo item (e) a sequência $(f(x^k))$ é convergente, então é limitada, e portanto $\|x^k - x^*\|$ é limitada. Assim, (x^k) é limitada.

Como (x^k) é limitada em \mathbb{R}^n , existe uma subsequência x^{k_j} tal que $x^{k_j} \rightarrow \bar{x}$.

Além disso, pelo item (e), $\nabla f(x^k) \rightarrow 0$. Como f é de classe C^2 , então ∇f é contínua, logo

$$\nabla f(\bar{x}) = \lim_{j \rightarrow \infty} \nabla f(x^{k_j}) = 0.$$

Portanto, \bar{x} é ponto crítico. Pelo item (b), f é convexa e tem único minimizador global x^* , logo $\bar{x} = x^*$.

Assim, toda subsequência convergente de (x^k) converge para x^* , e portanto $x^k \rightarrow x^*$. ☺

Exercício 8.

Considere uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^2 e suponha que existem constantes $M \geq m > 0$, tais que $\nabla^2 f(x) - mI$ e $M I - \nabla^2 f(x)$ são semi-definidas positivas para todo $x \in \mathbb{R}^n$.

- (a) Mostre que existem $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$ satisfazendo

$$f(x) \geq \frac{m}{2}\|x\|^2 + b^T x + c.$$

Conclua que f possui um minimizador global e que tal ponto é único.

- (b) Denotando por x^* o minimizador global de f , mostre que

$$f(x) - f(x^*) \geq \frac{m}{2}\|x - x^*\|^2$$

para todo $x \in \mathbb{R}^n$.

- (c) Seja (x^k) uma sequência gerada pelo método de Newton com passo constante $t_k = \frac{m}{M}$. Prove que

$$f(x^{k+1}) \leq f(x^k) - \frac{m}{2M} \nabla f(x^k)^T (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

para todo $k \in \mathbb{N}$.

- (d) Mostre que $x^k \rightarrow x^*$.

Solução:

- (a) Seja $x \in \mathbb{R}^n$. Como a matriz $\nabla^2 f(x) - mI$ é semi-definida positiva para todo x , temos a desigualdade:

$$x^T \nabla^2 f(x) v \geq m\|v\|^2, \quad \forall x \in \mathbb{R}^n. \quad (3)$$

Pelo desenvolvimento de Taylor com resto de Lagrange em torno do ponto 0 na direção x , temos

$$f(x) = f(0) + \nabla f(0)^T x + \frac{1}{2} x^T \nabla^2 f(\xi) x, \quad \text{onde } \xi = \theta x \text{ para algum } \theta \in (0, 1).$$

Logo, aplicando a desigualdade (3) no termo quadrático, obtém-se

$$f(x) \geq f(0) + \nabla f(0)^T x + \frac{m}{2}\|x\|^2.$$

Definindo $b = \nabla f(0) \in \mathbb{R}^n$ e $c = f(0) \in \mathbb{R}$, a expressão assume a forma

$$f(x) \geq \frac{m}{2} \|x\|^2 + b^T x + c.$$

Além disso, ao fazer $\|x\| \rightarrow \infty$, observa-se que $f(x) \rightarrow \infty$. Portanto, f é uma função coerciva, o que garante a existência de ao menos um minimizador global. Por outro lado, como $m > 0$, a matriz Hessiana $\nabla^2 f(x)$ é definida positiva para todo x , o que implica que f é estritamente convexa. Consequentemente, o minimizador global é único.

- (b) Denotando x^* como o minimizador global de f , pelo desenvolvimento de Taylor com resto de Lagrange em torno do ponto x^* na direção $(x - x^*)$, temos:

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2} (x - x^*)^T \nabla^2 f(\xi) (x - x^*),$$

onde $\xi = (1 - \theta)x^* + \theta x$ para algum $\theta \in (0, 1)$.

Como x^* é o minimizador global de uma função diferenciável em um conjunto aberto, pela condição de otimalidade de primeira ordem, temos que $\nabla f(x^*) = 0$. Assim, o termo linear desaparece, daí

$$f(x) = f(x^*) + \frac{1}{2} (x - x^*)^T \nabla^2 f(\xi) (x - x^*).$$

Utilizando a desigualdade (3), obtemos

$$f(x) \geq f(x^*) + \frac{m}{2} \|x - x^*\|^2,$$

para todo $x \in \mathbb{R}^n$.

- (c) Seja $x^k \in \mathbb{R}^n$. Para simplificar a notação, denotamos $g_k = \nabla f(x^k)$ e $H_k = \nabla^2 f(x^k)$. Pelo desenvolvimento de Taylor com resto de Lagrange em torno de x^k na direção do passo de Newton $(x^{k+1} - x^k)$, temos

$$f(x^{k+1}) = f(x^k) + g_k^T (x^{k+1} - x^k) + \frac{1}{2} (x^{k+1} - x^k)^T \nabla^2 f(\xi) (x^{k+1} - x^k),$$

onde ξ está no segmento entre x^k e x^{k+1} . Como $MI - \nabla^2 f(x)$ é semi-definida positiva, temos que $x^T \nabla^2 f(\xi) x \leq M \|x\|^2$ para todo $x \in \mathbb{R}^n$. Logo

$$f(x^{k+1}) \leq f(x^k) + g_k^T (x^{k+1} - x^k) + \frac{M}{2} \|x^{k+1} - x^k\|^2. \quad (4)$$

Considerando o método de Newton com passo $t_k = \frac{m}{M}$, temos $x^{k+1} - x^k = -\frac{m}{M} H_k^{-1} g_k$. Substituindo na desigualdade (4), tem-se

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - \frac{m}{M} g_k^T H_k^{-1} g_k + \frac{M}{2} \left(\frac{m}{M} \right)^2 \|H_k^{-1} g_k\|^2 \\ f(x^{k+1}) &\leq f(x^k) - \frac{m}{M} g_k^T H_k^{-1} g_k + \frac{m^2}{2M} g_k^T H_k^{-2} g_k. \end{aligned}$$

Note que, do item (a), os autovalores de H_k satisfazem $\lambda \geq m > 0$. Assim, os autovalores de H_k^{-1} satisfazem $\frac{1}{\lambda} \leq \frac{1}{m}$. Disso decorre a desigualdade matricial $H_k^{-2} \preceq \frac{1}{m} H_k^{-1}$, pois $g_k^T H_k^{-2} g_k \leq \frac{1}{m} g_k^T H_k^{-1} g_k$ (prova na afirmação de abaixo). Aplicando este resultado:

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - \frac{m}{M} g_k^T H_k^{-1} g_k + \frac{m^2}{2M} \left(\frac{1}{m} g_k^T H_k^{-1} g_k \right) \\ f(x^{k+1}) &\leq f(x^k) - \frac{m}{M} g_k^T H_k^{-1} g_k + \frac{m}{2M} g_k^T H_k^{-1} g_k. \end{aligned}$$

Portanto, subtraindo os termos, obtemos:

$$f(x^{k+1}) \leq f(x^k) - \frac{m}{2M} g_k^T H_k^{-1} g_k,$$

que é exatamente

$$f(x^{k+1}) \leq f(x^k) - \frac{m}{2M} \nabla f(x^k)^T (\nabla^2 f(x^k))^{-1} \nabla f(x^k).$$

Afirmação 1: Relação entre as formas quadráticas de H_k^{-2} e H_k^{-1}

Considere que a matriz Hessiana H_k é simétrica e definida positiva, cujos autovalores λ_i satisfazem a condição $\lambda_i \geq m > 0$ para todo $i = 1, \dots, n$. Para qualquer vetor g_k , verifica-se a desigualdade

$$g_k^T H_k^{-2} g_k \leq \frac{1}{m} g_k^T H_k^{-1} g_k$$

a qual estabelece a relação matricial $H_k^{-2} \preceq \frac{1}{m} H_k^{-1}$.

Demonstração. Para demonstrar esta afirmação, utilizamos o fato de que H_k é uma matriz simétrica, o que nos permite aplicar o Teorema Espectral para escrever $H_k = Q\Lambda Q^T$. Nesta decomposição, Q é uma matriz ortogonal e Λ é uma matriz diagonal contendo os autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$.

Desta representação, as potências da matriz Hessiana assumem a forma $H_k^{-1} = Q\Lambda^{-1}Q^T$ e $H_k^{-2} = Q\Lambda^{-2}Q^T$. Ao definirmos o vetor $y = Q^T g_k$, as formas quadráticas podem ser expressas em termos dos autovalores como

$$g_k^T H_k^{-1} g_k = y^T \Lambda^{-1} y = \sum_{i=1}^n \frac{y_i^2}{\lambda_i}$$

e analogamente para a potência quadrada temos

$$g_k^T H_k^{-2} g_k = y^T \Lambda^{-2} y = \sum_{i=1}^n \frac{y_i^2}{\lambda_i^2}$$

Considerando a hipótese de que cada autovalor λ_i é limitado inferiormente por m , obtemos a relação $\frac{1}{\lambda_i} \leq \frac{1}{m}$ para todo i . Ao multiplicarmos ambos os lados desta desigualdade pelo termo positivo $\frac{1}{\lambda_i}$, resulta que

$$\frac{1}{\lambda_i^2} \leq \frac{1}{m} \cdot \frac{1}{\lambda_i}$$

Desta forma, ao aplicarmos esta desigualdade em cada componente da soma que define a forma quadrática, observamos que

$$\sum_{i=1}^n \frac{y_i^2}{\lambda_i^2} \leq \frac{1}{m} \sum_{i=1}^n \frac{y_i^2}{\lambda_i}$$

o que implica diretamente que $g_k^T H_k^{-2} g_k \leq \frac{1}{m} g_k^T H_k^{-1} g_k$, como queríamos demonstrar. \square

(d) Pelo item (c), temos

$$f(x^{k+1}) \leq f(x^k) - \frac{m}{2M} g_k^T H_k^{-1} g_k,$$

onde $g_k = \nabla f(x^k)$ e $H_k = \nabla^2 f(x^k)$.

Como $H_k \preceq MI$, então $H_k^{-1} \succeq \frac{1}{M} I$, logo

$$g_k^T H_k^{-1} g_k \geq \frac{1}{M} \|g_k\|^2.$$

Assim,

$$f(x^{k+1}) \leq f(x^k) - \frac{m}{2M^2} \|g_k\|^2.$$

Portanto,

$$\|g_k\|^2 \leq \frac{2M^2}{m} [f(x^k) - f(x^{k+1})], \quad \forall k \in \mathbb{N}.$$

Como $(f(x^k))$ é monótona decrescente e limitada inferiormente por $f(x^*)$, segue que é convergente. Logo,

$$f(x^k) - f(x^{k+1}) \rightarrow 0,$$

e portanto

$$\|\nabla f(x^k)\| \rightarrow 0.$$

Pelo item (b), temos

$$f(x) - f(x^*) \geq \frac{m}{2} \|x - x^*\|^2, \quad \forall x \in \mathbb{R}^n.$$

Em particular,

$$\|x^k - x^*\|^2 \leq \frac{2}{m} (f(x^k) - f(x^*)),$$

logo (x^k) é limitada.

Sendo limitada, existe uma subsequência x^{k_j} tal que $x^{k_j} \rightarrow \bar{x}$. Pela continuidade de ∇f ,

$$\nabla f(\bar{x}) = \lim_{j \rightarrow \infty} \nabla f(x^{k_j}) = 0.$$

Assim, \bar{x} é ponto crítico. Pelo item (a), f é estritamente convexa e possui único minimizador global x^* , logo $\bar{x} = x^*$.

Portanto, toda subsequência convergente converge para x^* , e concluímos que $x^k \rightarrow x^*$. \odot

Exercício 9.

Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função de classe C^1 e considere o método de Newton puro para resolver a equação $f(x) = 0$. Suponha que $f(-x) = -f(x)$ para todo $x \in \mathbb{R}$.

- (a) Prove que se $x_0 \in \mathbb{R}$ é tal que $x_1 = -x_0$, então a sequência gerada pelo método é dada por $x_k = x_0$ para $k \in \mathbb{N}$ par e $x_k = -x_0$ para $k \in \mathbb{N}$ ímpar.
- (b) Considere $f(x) = x^3 - 5x$. Encontre $x_0 \in \mathbb{R} \setminus \{0\}$ satisfazendo as condições do item anterior. Represente geometricamente.
- (c) Considere $f(x) = x^3 - 12x$. Mostre que existe $x_0 \in \mathbb{R}$ tal que o método só consegue dar um passo. Represente geometricamente.

Solução:

O método de Newton puro para resolver $f(x) = 0$ é dado por

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Suponha que f é ímpar, isto é,

$$f(-x) = -f(x), \quad \forall x \in \mathbb{R}.$$

(a) Derivando $f(-x) = -f(x)$,

$$-f'(-x) = -f'(x) \iff f'(-x) = f'(x),$$

logo f' é par.

Suponha que $x_1 = -x_0$. Pelo método de Newton,

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = -x_0 \iff \frac{f(x_0)}{f'(x_0)} = 2x_0.$$

Agora,

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = -x_0 - \frac{f(-x_0)}{f'(-x_0)} = -x_0 - \frac{-f(x_0)}{f'(x_0)} = -x_0 + \frac{f(x_0)}{f'(x_0)}.$$

Usando a condição anterior,

$$x_2 = -x_0 + 2x_0 = x_0.$$

Por indução,

$$x_k = \begin{cases} x_0, & k \text{ par,} \\ -x_0, & k \text{ ímpar.} \end{cases}$$

(b) Considere $f(x) = x^3 - 5x$, então

$$f'(x) = 3x^2 - 5.$$

A condição do item (a) é

$$\frac{f(x_0)}{f'(x_0)} = 2x_0 \iff \frac{x_0^3 - 5x_0}{3x_0^2 - 5} = 2x_0.$$

Multiplicando,

$$x_0^3 - 5x_0 = 2x_0(3x_0^2 - 5) \iff x_0^3 - 5x_0 = 6x_0^3 - 10x_0 \iff 0 = 5x_0^3 - 5x_0 \iff 5x_0(x_0^2 - 1) = 0.$$

Como $x_0 \neq 0$,

$$x_0 = \pm 1.$$

Tomando, por exemplo, $x_0 = 1$, obtém-se $x_1 = -1$ e a sequência oscila entre 1 e -1.

Geometricamente, f é ímpar, possui três zeros reais e a reta tangente em $x_0 = 1$ intercepta o eixo x em -1, produzindo um ciclo de período 2.

(c) Considere $f(x) = x^3 - 12x$, então

$$f'(x) = 3x^2 - 12.$$

Temos

$$f'(2) = 3(4) - 12 = 0.$$

Se x_0 é tal que o método produz $x_1 = 2$, então

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

não está definido, pois $f'(2) = 0$. Logo, o método consegue dar apenas um passo.

Geometricamente, a função possui um máximo local e um mínimo local (obtidos via f' e f'') e intersecta o eixo x em três pontos. Em $x = 2$ a tangente é horizontal; se o método atinge esse ponto, a iteração seguinte não pode ser realizada. ☹

Exercício 10.

Considere a função $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2}x_1^2 + x_2$ e a sequência (x^k) , gerada pelo Algoritmo do Gradiente com busca exata.

(a) Prove que se $x = (a, b)$, com $a \neq 0$, então o tamanho de passo ótimo é dado por $t^* = 1 + \frac{1}{a^2}$.

(b) Prove que se $x^0 = (1, 1)$, então $x_1^k = (-1)^k$ e $x_2^k = 1 - 2k$.

Solução:

Considere

$$f(x_1, x_2) = \frac{1}{2}x_1^2 + x_2.$$

Temos

$$\nabla f(x) = \begin{pmatrix} x_1 \\ 1 \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} =: A.$$

O método do gradiente com busca exata é dado por

$$x^{k+1} = x^k + t^k d^k, \quad d^k = -\nabla f(x^k),$$

onde

$$t^k = \frac{-\nabla f(x^k)^T d^k}{(d^k)^T A d^k}.$$

(a) Se $x = (a, b)$, então

$$\nabla f(x) = \begin{pmatrix} a \\ 1 \end{pmatrix}, \quad d = -\begin{pmatrix} a \\ 1 \end{pmatrix}.$$

Logo,

$$-\nabla f(x)^T d = (a \quad 1) \begin{pmatrix} a \\ 1 \end{pmatrix} = a^2 + 1.$$

Além disso,

$$Ad = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -a \\ -1 \end{pmatrix} = \begin{pmatrix} -a \\ 0 \end{pmatrix},$$

$$d^T Ad = (-a \quad -1) \begin{pmatrix} -a \\ 0 \end{pmatrix} = a^2.$$

Assim,

$$t^* = \frac{a^2 + 1}{a^2} = 1 + \frac{1}{a^2}, \quad a \neq 0.$$

(b) Se $x^0 = (1, 1)$, então

$$\nabla f(x^0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad d^0 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}.$$

Pelo item (a),

$$t^0 = 1 + \frac{1}{1^2} = 2.$$

Logo,

$$x^1 = x^0 + t^0 d^0 = (1, 1) + 2(-1, -1) = (-1, -1).$$

Agora,

$$\nabla f(x^1) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad d^1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

$$t^1 = 1 + \frac{1}{(-1)^2} = 2,$$

$$x^2 = x^1 + 2(1, -1) = (1, -3).$$

Observa-se o padrão

$$x_1^{k+1} = -x_1^k, \quad x_2^{k+1} = x_2^k - 2.$$

Por indução, temos

$$x_1^k = (-1)^k, \quad x_2^k = 1 - 2k. \odot$$

Exercício 11.

Considere $\lambda > 1$, $A = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$ e $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2}x^T Ax$. Dado um ponto arbitrário $x = \begin{pmatrix} r \\ s \end{pmatrix}$, aplique um passo de Cauchy com tamanho de passo $t = \frac{2}{\lambda+1}$.

(a) Mostre que o ponto obtido é $x^+ = \frac{\lambda-1}{\lambda+1} \begin{pmatrix} r \\ -s \end{pmatrix}$.

(b) Qual o valor da razão $\frac{\|x^+\|}{\|x\|}$?

Solução:

(a) Considere $f(x) = \frac{1}{2}x^T Ax$, com

$$A = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \quad \lambda > 1.$$

Note que $\nabla f(x) = Ax$. Assim, dado

$$x = \begin{pmatrix} r \\ s \end{pmatrix},$$

temos

$$\nabla f(x) = \begin{pmatrix} r \\ \lambda s \end{pmatrix}.$$

Como o passo de Cauchy utiliza a direção de maior descida, temos

$$d = -\nabla f(x) = \begin{pmatrix} -r \\ -\lambda s \end{pmatrix}.$$

Aplicando um passo com tamanho

$$t = \frac{2}{\lambda+1},$$

obtemos

$$x^+ = x + td = \begin{pmatrix} r \\ s \end{pmatrix} + \frac{2}{\lambda + 1} \begin{pmatrix} -r \\ -\lambda s \end{pmatrix}.$$

Calculando componente a componente,

$$\begin{aligned} x_1^+ &= r - \frac{2}{\lambda + 1}r = \frac{\lambda - 1}{\lambda + 1}r, \\ x_2^+ &= s - \frac{2\lambda}{\lambda + 1}s = \frac{1 - \lambda}{\lambda + 1}s = -\frac{\lambda - 1}{\lambda + 1}s. \end{aligned}$$

Consequentemente,

$$x^+ = \frac{\lambda - 1}{\lambda + 1} \begin{pmatrix} r \\ -s \end{pmatrix}.$$

(b) Perceba que

$$\|x^+\| = \left\| \frac{\lambda - 1}{\lambda + 1} \begin{pmatrix} r \\ -s \end{pmatrix} \right\| = \frac{\lambda - 1}{\lambda + 1} \left\| \begin{pmatrix} r \\ -s \end{pmatrix} \right\|.$$

Como a norma Euclidiana é invariante por mudança de sinal em uma componente,

$$\left\| \begin{pmatrix} r \\ -s \end{pmatrix} \right\| = \left\| \begin{pmatrix} r \\ s \end{pmatrix} \right\| = \|x\|.$$

Assim,

$$\frac{\|x^+\|}{\|x\|} = \frac{\lambda - 1}{\lambda + 1}. \odot$$

Exercício 12.

Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função de classe C^1 com gradiente Lipschitz com constante $L > 0$. Consequentemente, para todos $x, y \in \mathbb{R}^n$,

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \frac{L}{2} \|y - x\|^2.$$

Suponha que f é limitada inferiormente. Seja $(x^k) \subset \mathbb{R}^n$ uma sequência gerada pelo método de Cauchy com passo constante $t_k = \frac{1}{L}$.

(a) Prove que, para todo $k \in \mathbb{N}$,

$$f(x^k) - f(x^{k+1}) \geq \frac{1}{2L} \|\nabla f(x^k)\|^2;$$

(b) Conclua que $\lim_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0$.

Solução:

Suponha que ∇f é Lipschitz com constante $L > 0$, isto é,

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \frac{L}{2} \|y - x\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

Note que, em particular,

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{L}{2} \|y - x\|^2.$$

A sequência é gerada pelo método de Cauchy com passo constante

$$x^{k+1} = x^k - \frac{1}{L} \nabla f(x^k).$$

(a) Aplicando a desigualdade anterior com $x = x^k$ e $y = x^{k+1}$, temos

$$f(x^{k+1}) \leq f(x^k) + \nabla f(x^k)^T(x^{k+1} - x^k) + \frac{L}{2} \|x^{k+1} - x^k\|^2.$$

Logo,

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2L} \|\nabla f(x^k)\|^2.$$

(b) Pelo item (a), $(f(x^k))$ é monótona decrescente. Como f é limitada inferiormente, segue que $(f(x^k))$ é convergente. Além disso, da desigualdade do item (a),

$$\frac{1}{2L} \|\nabla f(x^k)\|^2 \leq \underbrace{f(x^k) - f(x^{k+1})}_{\rightarrow 0}.$$

Tomando limite quando $k \rightarrow \infty$, o lado direito tende a 0 porque $(f(x^k))$ é convergente. Consequentemente,

$$\|\nabla f(x^k)\|^2 \rightarrow 0,$$

o que implica

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0. \odot$$

Exercício 13.

Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 . Suponha que $L > 0$ é tal que $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$ para todos $x, y \in \mathbb{R}^n$. Considere $t < \frac{1}{L}$ fixado e $x^0 \in \mathbb{R}^n$. Seja (x^k) a sequência gerada pelo método de Cauchy com tamanho de passo $t_k = t$, isto é,

$$x^{k+1} = x^k - t\nabla f(x^k) \quad (5)$$

- (a) Mostre que $|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq L\|y - x\|^2$ para todos $x, y \in \mathbb{R}^n$. (Sugestão: Teorema do valor meio)
- (b) Mostre que existe $M > 0$ tal que $f(x^{k+1}) \leq f(x^k) - M\|\nabla f(x^k)\|^2$ para todo $k \in \mathbb{N}$.
- (c) Conclua que o algoritmo definido por (5) é globalmente convergente, isto é, se $\bar{x} \in \mathbb{R}^n$ é um ponto de acumulação da sequência (x^k) , então $\|\nabla f(\bar{x})\| = 0$. (fato que você pode usar, caso necessite: se uma sequência monótona possui uma subsequência convergente, então a sequência é convergente)

Solução:

Suponha que $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$ para todos $x, y \in \mathbb{R}^n$ e considere

$$x^{k+1} = x^k - t\nabla f(x^k), \quad 0 < t < \frac{1}{L}.$$

- (a) Fixemos $x, y \in \mathbb{R}^n$ e considere $z \in [x, y]$. Definimos

$$h : \mathbb{R}^n \rightarrow \mathbb{R}, \quad h(z) = f(z) - \nabla f(x)^T z.$$

Note que h é diferenciável pois $f \in \mathcal{C}^1$. Além disso,

$$\nabla h(z) = \nabla f(z) - \nabla f(x).$$

Pela hipótese de Lipschitz,

$$\|\nabla h(z)\| = \|\nabla f(z) - \nabla f(x)\| \leq L\|z - x\| \leq L\|y - x\| =: \beta.$$

Logo $\|\nabla h(z)\| \leq \beta$ em $[x, y]$. Pelo Teorema do Valor Médio,

$$|h(y) - h(x)| \leq \beta\|y - x\|.$$

Substituindo a definição de h ,

$$|f(y) - \nabla f(x)^T y - (f(x) - \nabla f(x)^T x)| \leq L\|y - x\|^2.$$

Portanto,

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq L\|y - x\|^2.$$

- (b) Pelo item anterior, em particular,

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + L\|y - x\|^2.$$

Pelo método do gradiente,

$$x^{k+1} - x^k = -t\nabla f(x^k).$$

Tomando $x = x^k$ e $y = x^{k+1}$, obtemos

$$f(x^{k+1}) \leq f(x^k) - t\|\nabla f(x^k)\|^2 + Lt^2\|\nabla f(x^k)\|^2.$$

Assim,

$$f(x^{k+1}) \leq f(x^k) - (t - Lt^2)\|\nabla f(x^k)\|^2.$$

Como $0 < t < \frac{1}{L}$, temos $t - Lt^2 > 0$. Definindo

$$M := t - Lt^2 > 0,$$

segue que

$$f(x^{k+1}) \leq f(x^k) - M\|\nabla f(x^k)\|^2.$$

- (c) Suponha que \bar{x} é ponto de acumulação de (x^k) , isto é, existe uma subsequência $x^{k_j} \rightarrow \bar{x}$.

Pelo item (b),

$$f(x^{k+1}) \leq f(x^k) - M\|\nabla f(x^k)\|^2, \quad M > 0.$$

Logo $(f(x^k))$ é monótona decrescente. Como f é limitada inferiormente, segue que $(f(x^k))$ é convergente.

Da desigualdade acima,

$$M\|\nabla f(x^k)\|^2 \leq f(x^k) - f(x^{k+1}).$$

Fazendo $k \rightarrow \infty$,

$$\lim_{k \rightarrow \infty} M\|\nabla f(x^k)\|^2 \leq \lim_{k \rightarrow \infty} (f(x^k) - f(x^{k+1})) = 0.$$

Logo,

$$\lim_{k \rightarrow \infty} \|\nabla f(x^k)\| = 0.$$

Como ∇f é contínuo e $x^{k_j} \rightarrow \bar{x}$, obtemos

$$\nabla f(\bar{x}) = \lim_{j \rightarrow \infty} \nabla f(x^{k_j}) = 0.$$

Portanto,

$$\|\nabla f(\bar{x})\| = 0. \odot$$

Exercício 14.

Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1 . Suponha que existe $L > 0$ tal que

$$\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$$

para todos $x, y \in \mathbb{R}^n$. Dado $x^0 \in \mathbb{R}^n$, seja $(x_k) \subset \mathbb{R}^n$ a sequência gerada pelo método de Cauchy com tamanho de passo constante $t_k = \frac{1}{L}$, isto é,

$$x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k) \quad (6)$$

- (a) Mostre que $|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \frac{L}{2}\|y - x\|^2$ para todos $x, y \in \mathbb{R}^n$;
- (b) Mostre que $f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2$ para todo $k \in \mathbb{N}$;
- (c) Conclua que o algoritmo definido por (6) é globalmente convergente;
- (d) Assumindo que f é limitada inferiormente, prove que $\|\nabla f(x_k)\| \rightarrow 0$.

Solução:

Suponha que $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$ para todos $x, y \in \mathbb{R}^n$ e considere

$$x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k).$$

- (a) Sejam $x, y \in \mathbb{R}^n$ fixos e considere

$$\varphi(t) = f(x + t(y - x)), \quad t \in [0, 1].$$

Note que $\varphi(0) = f(x)$, $\varphi(1) = f(y)$ e

$$\varphi'(t) = \nabla f(x + t(y - x))^T(y - x).$$

Pelo Teorema Fundamental do Cálculo,

$$f(y) - f(x) = \int_0^1 \varphi'(t) dt = \int_0^1 \nabla f(x + t(y - x))^T(y - x) dt.$$

Subtraindo $\nabla f(x)^T(y - x)$, obtemos

$$f(y) - f(x) - \nabla f(x)^T(y - x) = \int_0^1 (\nabla f(x + t(y - x)) - \nabla f(x))^T(y - x) dt.$$

Tomando módulo e usando Cauchy-Schwarz,

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \int_0^1 \|\nabla f(x + t(y - x)) - \nabla f(x)\| \|y - x\| dt.$$

Pela hipótese de Lipschitz,

$$\|\nabla f(x + t(y - x)) - \nabla f(x)\| \leq Lt\|y - x\|.$$

Logo,

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \int_0^1 Lt\|y - x\|^2 dt = \frac{L}{2}\|y - x\|^2.$$

(b) Da desigualdade anterior, em particular,

$$f(y) \leq f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|^2.$$

Tomando $x = x_k$ e $y = x_{k+1}$ e usando (1),

$$x_{k+1} - x_k = -\frac{1}{L}\nabla f(x_k).$$

Substituindo,

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\frac{1}{L^2}\|\nabla f(x_k)\|^2.$$

Assim,

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2.$$

(c) Queremos mostrar que todo ponto de acumulação é estacionário.

Suponha que exista uma subsequência $x_{k_j} \rightarrow \bar{x}$.

Note que f é contínua, logo

$$\lim_{j \rightarrow \infty} f(x_{k_j}) = f(\bar{x}).$$

Por outro lado, pelo item (b),

$$f(x_{k+1}) \leq f(x_k),$$

pois $\frac{1}{2L} > 0$. Assim $(f(x_k))$ é monótona decrescente.

Como possui uma subsequência convergente, segue que $(f(x_k))$ é convergente e

$$\lim_{k \rightarrow \infty} f(x_k) = f(\bar{x}).$$

Da desigualdade do item (b),

$$\frac{1}{2L}\|\nabla f(x_k)\|^2 \leq f(x_k) - f(x_{k+1}).$$

Fazendo $k \rightarrow \infty$, o lado direito tende a zero, logo

$$\lim_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0.$$

Como ∇f é contínuo e $x_{k_j} \rightarrow \bar{x}$, obtemos

$$\nabla f(\bar{x}) = \lim_{j \rightarrow \infty} \nabla f(x_{k_j}) = 0.$$

Portanto \bar{x} é ponto estacionário.

(d) Assumindo que f é limitada inferiormente, pelo item (b) temos que $(f(x_k))$ é monótona decrescente e limitada inferiormente, logo é convergente.

Da desigualdade

$$\frac{1}{2L}\|\nabla f(x_k)\|^2 \leq f(x_k) - f(x_{k+1}),$$

e passando ao limite quando $k \rightarrow \infty$, obtemos novamente

$$\|\nabla f(x_k)\| \rightarrow 0.$$

Exercício 15.

(Regularidade do Gradiente) Seja $U \subset \mathbb{R}^n$ um conjunto aberto e convexo, e seja $f : U \rightarrow \mathbb{R}$ uma função de classe $\mathcal{C}^2(U)$. Suponha que existam constantes $\alpha \geq 0$ e $L > 0$ tais que

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|^\alpha$$

para todos $x, y \in U$. Mostre que vale a seguinte estimativa:

$$\|\nabla f(x) - \nabla f(y) - \nabla^2 f(y)(x - y)\| \leq \frac{L}{\alpha + 1}\|x - y\|^{\alpha+1}$$

para todos $x, y \in U$.

Solução:

Fixemos $x, y \in U$. Como o conjunto U é convexo, o segmento de reta $[y, x] = \{y + t(x - y) : t \in [0, 1]\}$ está contido inteiramente em U . Definimos a função vetorial auxiliar $\phi : [0, 1] \rightarrow \mathbb{R}^n$ como:

$$\phi(t) = \nabla f(y + t(x - y)).$$

Visto que $f \in \mathcal{C}^2(U)$, a função ϕ é de classe \mathcal{C}^1 no intervalo $[0, 1]$. Seus valores nos extremos são $\phi(0) = \nabla f(y)$ e $\phi(1) = \nabla f(x)$. Pela regra da cadeia, a derivada de ϕ para cada $t \in [0, 1]$ é dada por:

$$\phi'(t) = \nabla^2 f(y + t(x - y))(x - y).$$

Pelo Teorema Fundamental do Cálculo para funções vetoriais, temos:

$$\nabla f(x) - \nabla f(y) = \int_0^1 \phi'(t) dt = \int_0^1 \nabla^2 f(y + t(x - y))(x - y) dt.$$

Subtraindo $\nabla^2 f(y)(x - y)$ em ambos lados da igualdade, temos:

$$\nabla f(x) - \nabla f(y) - \nabla^2 f(y)(x - y) = \int_0^1 [\nabla^2 f(y + t(x - y)) - \nabla^2 f(y)](x - y) dt.$$

Tomando a norma e aplicando a desigualdade triangular para integrais:

$$\begin{aligned} \|\nabla f(x) - \nabla f(y) - \nabla^2 f(y)(x - y)\| &\leq \int_0^1 \|\nabla^2 f(y + t(x - y)) - \nabla^2 f(y)\| \cdot \|x - y\| dt \\ &\leq \int_0^1 L \|t(x - y)\|^\alpha \cdot \|x - y\| dt = L \|x - y\|^{\alpha+1} \int_0^1 t^\alpha dt. \end{aligned}$$

Como $\alpha \geq 0$, a integral resulta em $\int_0^1 t^\alpha dt = \frac{1}{\alpha+1}$. Substituindo este valor, chegamos à estimativa desejada:

$$\|\nabla f(x) - \nabla f(y) - \nabla^2 f(y)(x - y)\| \leq \frac{L}{\alpha + 1} \|x - y\|^{\alpha+1}. \odot$$

Observação

Parece que esta estimativa de regularidade, ao considerar o parâmetro α , oferece uma perspectiva interessante sobre o comportamento local de algoritmos em Otimização Numérica. Nota-se que, para o método de Newton, essa desigualdade poderia sugerir uma transição na velocidade de convergência: enquanto o caso clássico assume $\alpha = 1$ para garantir convergência quadrática, a presença de um $\alpha \in (0, 1)$ indicaria uma convergência superlinear de ordem $1 + \alpha$.

É curioso observar que, se considerarmos o limite inferior $\alpha = 0$, a estimativa assemelha-se formalmente ao limite de erro da aproximação linear utilizado em provas de convergência global (como o Lema de Descida para métodos de gradiente), sugerindo que esta formulação pode atuar como uma ponte teórica entre diferentes regimes de suavidade. Embora estas interpretações ainda careçam de uma análise mais profunda para especialistas, elas parecem indicar que a "folga" na regularidade da Hessiana afeta diretamente a eficiência computacional do método.