Lista 1: Otimização II

A. Ramos *

August 13, 2017

Abstract

Lista em constante atualização.

1. Métodos de gradiente;
2. Método de Newton e variantes.

Seja $\mathcal{O}$ um aberto em $\mathbb{R}^n$. Denote por $C^{1,1}_L(\mathcal{O})$ o conjunto das funções deriváveis em $\mathcal{O}$ cuja derivada é Lipschitziana com constante de Lipschitz L em $\mathcal{O}$, isto é, $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$, para todo $x, y \in \mathcal{O}$.

Com essas informações responda:

1. Considere a função $f : \mathbb{R}^n \to \mathbb{R}$ definida como $f(x) = x^T Ax + b^T x + c$. Mostre que
   - $\nabla f(x) = Ax + A^T x + b$. E se A é simetrico ($A = A^T$), $\nabla f(x) = 2Ax + b$
   - $\nabla^2 f(x) = A^T + A$. E quando A é simetrico, $\nabla^2 f(x) = 2A$
   - No caso que A é simétrica. Mostre que f admite solução se, e somente se A é definida positiva.
2. Seja f uma função continuamente derivável em $\mathbb{R}^n$. Suponha que d é uma direção de descida de f em x. Mostre que existe um número $T > 0$ tal que
$$f(x + td) < f(x) \quad \forall t \in (0, T].$$
3. Seja d uma direção de descida para uma função derivável f no ponto $x \in \mathbb{R}^n$. Mostre que se f tem derivada contínua, o ponto $x^+ := x + td$ está bem definido, onde t é escolhido segundo as seguintes condições
   (a) A condição de Armijo, a condição de Goldstein
   (b) A condição de Wolfe e a condição de Wolfe forte.
4. Considere o problema
$$\min\ f(x) := x_1^2 + 2x_2^2 - x_1x_2 - 2x_1 + e^{x_1+x_2}, \text{ s.a. } x = (x_1, x_2) \in \mathbb{R}^2.$$
   (a) Verifique que $(0, 0)$ não é solução do problema
   (b) Minimize a função a partir de $(0, 0)$ ao longo da direção de máxima descida.
5. Seja $f(x) := x_1^4 + x_1^2 + x_2^2$. Considere o ponto $x^k := (1, 1)^T$ e a direção $d^k := (-3, -1)^T$.
   (a) Verifique que d^k é uma direção de descida para f em x^k.
   (b) Use a condição de Wolfe para encontrar o novo ponto $x^{k+1} := x^k + td^k$, com parámetros $\rho := 0.1$ e $\sigma := 0.5$. Isto é:
$$f(x^k + td^k) \leq f(x^k) + t\rho\nabla f(x^k)^T d^k \quad \text{(condição de descrescimo suficiente)}$$
   e
$$\nabla f(x^k + td^k)^T d^k \geq \sigma \nabla f(x^k)^T d^k, \sigma \in (\rho, 1) \quad \text{(condição sobre a curvatura)}.$$
   (c) Considere os valores de $t = 1$, $t = 0.5$ e $t = 0.1$ respectivamente. Quais valores satisfazem a condição de Wolfe?
6. Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função derivável cuja derivada é uma função lipschitziana com constante de Lipschitz L (i.e. $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$, para todo $x, y \in \mathbb{R}^n$). Mostre que
$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|x - y\|^2, \text{ para todo } x, y \in \mathbb{R}^n.$$
   Isto é, a função quadrática $Q(y) := f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|x - y\|^2$ sobre estima a função f em todo $\mathbb{R}^n$
7. Seja f uma funçaõ duas vezes derivável em $\mathbb{R}^n$. As seguintes proposições são equivalentes:

*Department of Mathematics, Federal University of Paraná, PR, Brazil. Email: albertoramos@ufpr.br.

(a) A derivada de f Lipschitziana com constante de Lipschitz L

(b) $\|\nabla^2 f(x)\| \leq L$, para todo $x \in \mathbb{R}^n$.

8. Seja $f \in C_L^{1,1}(\mathbb{R}^n)$ e $\{x^k\}$ uma sequência generada pelo método do gradiente com passo constante $t_k = 1/L$. Suponha que $x^k \to x^*$. Prove que se $\nabla f(x^k) \neq 0$, para $k \in \mathbb{N}$. Então, x^* não é um máximo local.

9. Seja $x^0 \in \mathbb{R}^n$ um ponto inicial e seja $f \in C_L^{1,1}(\mathcal{O})$, onde $\mathcal{O}$ é um aberto que contem o conjunto de nível $\{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$. Adicionalmente suponha que f é limitada inferiormente.

   Se $\{x^{k+1} := x^k + t_k d^k\}$ é uma sequência de iterados, onde d^k é uma direçaõ de descida.

   Mostre que se t_k satisfaz (i) a condição de Wolfe, ou (ii) a condição de Goldstein ou (iii) a condição de Wolfe-forte. Então, a condição de Zoutendijk é satisfeita i.e. $\sum_{k=1} \cos^2(\theta_k)\|f(x^k)\|^2 < \infty$, onde $\cos(\theta_k) := -{d^k}^T \nabla f(x^k)/\|d^k\|\|\nabla f(x^k)\|$.

10. Considere o problema de minimizar $\min\{x^T Ax, x \in \mathbb{R}^2\}$, onde A é uma matriz 2×2 definida positiva. Considere $D = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1}. \end{pmatrix}$

    Prove que $\mathcal{K}(D^{1/2}AD^{1/2}) \leq \mathcal{K}(A)$. Interprete e analise o método do gradiente para este caso.

11. Seja $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x) = \frac{1}{2}x_1^2 + \frac{1}{4}x_2^4 - \frac{1}{2}x_2^2$. Responda:

    (a) Determine e classifique os pontos estacionários.

    (b) Faça uma iteração do método de gradiente com ponto inicial $x^0 = (1, 0)^T$. Discuta a possível convergência do método de gradiente.

12. Seja $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x) = \frac{1}{2}x_1^2 + \frac{a}{2}x_2^2$, onde $a \geq 1$. Mostre que o método de gradiente, com ponto inicial $x^0 = (a, 1)^T$, gera a seguinte sequência

$$x^k := (x_1^k, x_2^k)^T = \left(\frac{a-1}{a+1}\right)^k (a, (-1)^k)^T, \text{ para todo } k \in \mathbb{N}.$$

13. Seja $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x) = 5x_1^2 + 5x_2^2 - x_1x_2 + 11x_2 - 11x_1 + 11$. Responda:

    (a) Calcule a taxa de convergência de $\|x^k - x^*\|$ e de $f(x^k) - f(x^*)$.

    (b) Considere $x^0 = (0, 0)^T$. Quantas iterações são necessárias para obter uma precisão de 10^{-8} no valor ótimo de f?.

14. Considere $f(x) = \frac{1}{2}(x_1^2 - x_2)^2 + \frac{1}{2}(1 - x_1)^2$. Responda:

    (a) Calcule o minimizador de f

    (b) Calcule uma iteração do método de Newton para minimizar f a partir de $x^0 = (2, 2)^T$. Esse passo é aceitável? *Dica:* Calcule $f(x^0)$ e $f(x^1)$.

15. Em $\mathbb{R}^n$, considere $f(x) := \|x\|^3$. Faça o método de Newton com passo constante $t_k = 1$. Mostre que o método converge linearmente para o mínimo $x^* = 0$. Por quê não temos convergência quadrática?

16. Denote $M(n, \mathbb{R})$ o conjunto de matrices reaias. Seja $GL(n, \mathbb{R})$ o conjunto das matrices não singulares. A função que associa cada matriz com a sua inversa, $Inv : GL(n, \mathbb{R}) \to GL(n, \mathbb{R})$, $Inv(A) = A^{-1}$ é infinitamente derivável. Para isto faça o seguinte:

    (a) Primeiro, calcule as derivadas de Inv em I (onde I é a matriz identidade) e mostre que $D^k Inv(I)[A, \ldots, A] = (-1)^k k! A^k$. Para isto, use a formula de Neumann [1] para escrever a expansão de $(I + tA)^{-1}$, para t suficientemente pequeno.

    (b) No caso geral, para $A \in GL(n, \mathbb{R})$ e $B \in M(n, \mathbb{R})$ mostre que

$$D^k Inv(A)[B, \ldots, B] = (-1)^k k! A^{-1}BA^{-1}BA^{-1} \ldots BA^{-1},$$

    onde temos $k + 1$ matrices A e k matrices B. *Dica:* Escreva $A + tB = A(I + tA^{-1}B)$ e use o item anterior.

    (c) Descreva o método de Newton para calcular a inversa de uma matriz.

17. Considere uma matriz definida positiva $A \in M(n, \mathbb{R})$. (i) Mostre que $\|x\|_A := \sqrt{x^T Ax}$ é uma norma em $\mathbb{R}^n$. (ii) Ainda mais, prove que $\sqrt{\lambda_{min}(A)}\|x\|_2 \leq \|x\|_A \leq \sqrt{\lambda_{max}(A)}\|x\|_2$ para todo $x \in \mathbb{R}^n$, onde $\| \cdot \|_2$ é a norma euclideana.

    (iii) Use o resultado anterior para provar que a sequência x^k generada pelo método de máxima descida (com busca exata) aplicado ao problema $\min f(x) := (1/2)x^T Ax$ converge à solução x^* de dito problema e

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq \sqrt{\mathcal{K}} \left(\frac{\mathcal{K} - 1}{\mathcal{K} + 1}\right), \text{ onde } \mathcal{K} = \frac{\lambda_{min}(A)}{\lambda_{max}(A)}.$$

[1] Formula de Neumann: Para $B \in M(n, \mathbb{R})$ com $\|B\| < 1$, temos que $(I + B)^{-1} = \sum_{k=0}^{\infty}(-1)^k B^k$.

18. **Direções de curvaura negativa**. Considere uma função de classe $C^2(\mathbb{R})$. Então:

(a) Se $\nabla^2 f(x)$ tem um autovalor negativo, dizemos que x é uma ponto indefinido.

(b) Se x é um ponto indefinido e existe uma direção d tal que $d^T \nabla^2 f(x) d < 0$. Dito vetor é chamado de *direção de curvatura negativa.*

(c) Se existe um par de vetores (z, d) tal que

$$\nabla f(x)^T z \leq 0, \quad \nabla f(x)^T d \leq 0, \quad d^T \nabla^2 f(x) d < 0,$$

dizemos que (z, d) é um par de descida no ponto indefinido x. No caso que x não é um ponto indefinido (i.e. $\nabla^f(x) \succeq 0$) se (z, d) satisfaz

$$\nabla f(x)^T z < 0, \quad \nabla f(x)^T d \leq 0, \quad d^T \nabla^2 f(x) d = 0,$$

dizemos que (z, d) é um par de descida no ponto x.

Com um par de descida (z^k, d^k) podemos fazer busca ao longo de uma curva da forma

$$x(t) := x^k + \phi_1(t) z^k + \phi_2(t) d^k$$

para certos ϕ_1 e ϕ_2, em lugar de fazer uma busca linear.

(a) *Condição de Armijo de segunda-ordem.* Nesse caso, a busca é realizada ao longo de curvas da forma:

$$x(t) := x^k + t^2 z^k + t d^k, \text{ i.e. } \phi_1(t) = t^2, \phi_2(t) = t.$$

Considere $\rho, \gamma \in (0, 1)$, e ponha $x^k(i) := x^k + \gamma^{2i} z^k + \gamma^i d^k$. A condição de Armijo de segunda-ordem pede por encontrar $i(k) \in \mathbb{N}$ o menor inteiro não negativo i tal que

$$f(x^k(i)) \leq f(x^k) + \rho \gamma^{2i} (\nabla f(x^k)^T z^k + \frac{1}{2} {d^k}^T \nabla^2 f(x^k) d^k)$$

Atualize $x^{k+1} := x^k(i(k))$. Mostre o seguinte:

i. O passo de Armijo de segunda-ordem está bem definido, se temos que $\nabla f(x^k)^T z^k < 0$ (quando $\nabla f(x^k) \neq 0$) e ${d^k}^T \nabla^2 f(x^k) d^k < 0$ (quando $\nabla f(x^k) = 0$).

ii. Seja $f \in C^2(\mathbb{R}^n)$ tal que $\{x \in \mathbb{R}^n : f(x) \leq f(x^0)\}$ é compacto. Seja $\{x^k\}$ uma sequência que satisfaz a condição de Armijo de segunda-ordem, e suponha que as sequências $\{\|z^k\|\}$ e $\{\|d^k\|\}$ são limitadas. Prove que :

$$\nabla f(x^k)^T z^k \to 0 \text{ e } {d^k}^T \nabla^2 f(x^k) d^k \to 0.$$

iii. Se adicionalmente às hipoteses do item anterior, temos que existem constantes $c_1, c_2, c_3 > 0$ tal que

A. $\|z^k\| \geq c_3 \|\nabla f(x^k)\|$

B. ${d^k}^T \nabla^2 f(x^k) d^k \leq c_2 \lambda_{min}(\nabla^2 f(x^k))$ (lembre que $\lambda_{min}(A)$ denota o mínimo autovalor de A).

C. $-\nabla f(x^k)^T z^k \geq c_1 \|\nabla f(x^k)\| \|z^k\|$.

Então, qualquer ponto de acumulação x^* de x^k é um ponto estacionário de segunda-ordem, isto é, $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*) \succeq 0$.

Obs: Além da condição de segunda-ordem de Armijo, outras condições de segunda-ordem são a condição de Goldfard, a condição de Moré-Sorensen, etc.

19. Seja $F : \mathbb{R}^2 \to \mathbb{R}^2$ uma função cuja componentes são $F_1(x) = x_1^2 + x_2^2 - 9$ e $F_2(x) = x_1 + x_2 - 3$.

(a) Avalie a Jacobiana de F em $x = (1, 0)^T$ e $x = (1, 5)^T$.

(b) Faça duas (três) iterações do método de Newton, para resolver $F(x) = 0$, partindo de $x^0 = (1, 5)^T$.

(c) Escreva o problema $F(x) = 0$, como um problema de otimização com função objetivo $f(x) := \|F(x)\|^2$. Encontre o gradiente e a Hessiana da função objetivo.

20. Seja $f : \mathbb{R} \to \mathbb{R}$ definida $f(x) := x^3 - x$. Construa um modelo linear de f. Nos pontos (1) $x = 0$, (ii) $x = \sqrt{3}/3$ e (iii) $x = 2$. Explique o que acontece em cada uma das situações.