Lista 3: Otimização I

A. Ramos *

October 29, 2017

Abstract

Lista em constante atualização.

1. Métodos de descida, Newton, quase-Newton e gradiente conjugados
2. Para os exercícios que forem convenientes pode ser usado alguma linguagem de programação.

1. lista de exercício

   MÉTODO DE DESCIDA

2. 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 $x^+ := x + \alpha d$ está bem definido, onde α é escolhido segundo as seguintes condições
   (a) A condição de Armijo,
   (b) A condição de Goldstein
   (c) A condição de Wolfe e a condição de Wolfe forte.

3. 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 $\mathbb{R}^n$

   Se ainda supomos que f é convexo, mostre que

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

4. Seja f uma função duas vezes derivável em $\mathbb{R}^n$. As seguintes proposições são equivalentes:
   (a) A derivada de f Lipschitziana com constante de Lipschitz L
   (b) $\|\nabla^2 f(x)\| \leq L$, para todo $x \in \mathbb{R}^n$.

5. Seja $f \in C_L^{1,1}(\mathbb{R}^n)$ e $\{x^k\}$ uma sequência gerada 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.

6. Seja $x^0 \in \mathbb{R}^n$ um ponto inicial e considere uma função $f \in C_L^{1,1}(\mathcal{O})$, 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. Seja $\{x^{k+1} := x^k + \alpha_k d^k\}$ uma sequência de iterados, onde d^k é uma direção de descida e $\alpha_k > 0$.

   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 $\cos(\theta_k) := -\langle d^k, \nabla f(x^k)\rangle / \|d^k\|\|\nabla f(x^k)\|$.

7. Considere uma matriz simétrica definida positiva $A \in M(n, \mathbb{R})$. (i) Mostre que $\langle x, y\rangle_A := \sqrt{x^T A y}$ é um produto interno 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 gerada pelo método de máxima descida (com busca exata) aplicado ao problema min $f(x) := (1/2)x^T A x$ converge à solução x^* de dito problema e

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

   MÉTODOS QUASE-NEWTON

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

8. Verificar a formula de Sherman-Morrison-Woodbury para a inversa dada em aula.

9. Seja $F : \mathbb{R}^n \to \mathbb{R}^n$ uma função derivável tal que $\|DF(x) - DF(x^*)\| \le \|x - x^*\|^p$ para todo $x \in B(x^*, r)$ com $r > 0, p > 0$. Prove que para todo $x, y \in B(x^*, r)$ temos que

$$\|F(x) - F(y) - DF(x^*)(x - y)\| \le L\|x - y\| \max\{\|x - x^*\|^p, \|y - x^*\|^p\}.$$

10. Prove que se x^* é tal que $F(x^*) = 0$ e $DF(x^*)$ é invertível. Existe, uma vizinhança de x^*, tal que para todo x nessa vizinhança temos que $c_1\|x - x^*\| \le \|F(x)\| \le c_2\|x - x^*\|$, para certos c_1, c_2 positivos.

11. Mostre que todas atualização dadas em aula dos $B_{k+1}(H_{k+1})$ (SR1, PSP, BFGS, DFP, etc) satisfazem os problemas de otimização dados.

12. Demonstre o Teorema 5.4.4 do livro de Sun et al.

MÉTODO DE GRADIENTES CONJUGADOS LINEAR E NÃO LINEAR

13. Seja Q uma matriz simétrica definida positiva. Verifique que no método de Gradiente Conjugados linear temos que para $k \ge 1$
$$\text{span}\{r^0, r^1, r^2, \dots, r^k\} = \text{span}\{d^0, d^1, d^2, \dots, d^k\} = \text{span}\{r^0, Qr^0, Q^2r^0, \dots, Q^kr^0\}.$$

14. Encontre os mínimos das quadráticas usando método dos gradientes conjugados

    (a) $q(x, y) = -xy + 1 - y + x^2 + (1/2)y^2$, com $x^0 = (0, 0)^T$

    (b) $q(x, y) = -3x - 4y - 0.5 + 2xy + x^2 + y^2$, com $x^0 = (2, 1)^T$

15. Suponha que o método de gradientes conjugados não linear é implementada de forma que o parâmetro do passo α_k satisfaz a condição forte de Wolfe, com $c_2 \in (0, 1/2)$, e que $|\beta_k| \le \beta_k^{FR}$, $\forall k \in \mathbb{N}$. Então, mostre que

$$-\frac{1}{1 - c_2} \le \frac{\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\|^2} \le \frac{2c_2 - 1}{1 - c_2}, \text{ para } k = 0, 1, 2, \dots,$$

    Conclua que as direções d^k são direções de descida.

16. Seja Q uma matriz simétrica definida positiva e considere $\{v^1, \dots, v^n\}$ uma família de vetores linearmente independentes. Defina

$$d^1 := v^1 \quad \text{e} \quad d^{k+1} := v^{k+1} - \sum_{i=1}^{k} \theta_i^{k+1} d^i, \quad \text{para} \quad k = 1, 2, \dots, n - 1,$$

    onde $\theta_i^{k+1} = (v^{k+1})^T Qd^i/(d^i)^T Qd^i$. Mostre que $\{d^1, \dots, d^n\}$ são Q-conjugados.

17. Prove que para o método de Gradiente conjugados linear, sempre temos que $\langle d^k, Ad^k\rangle = -\langle d^k, Ag^k\rangle$, onde $g^k := \nabla q(x^k)$.

18. Mostre as seguintes relações para o método de gradientes conjugados linear

    (a)
$$\alpha_k = \frac{\|r^k\|^2}{\langle d^k, Qd^k\rangle} = -\frac{\langle r^k, d^k\rangle}{\langle d^k, Qd^k\rangle} = -\frac{\langle r^0, d^k\rangle}{\langle d^k, Qd^k\rangle}$$

    (b)
$$\beta_{k+1} = \frac{\|r^{k+1}\|^2}{\|r\|^k} = \frac{\langle r^{k+1}, Qd^k\rangle}{\langle d^k, Qd^k\rangle} = -\frac{\langle r^{k+1}, Qr^k\rangle}{\langle d^k, Qd^k\rangle}$$

19. Seja $\{x^k\}$ uma sequência gerada pelo método de gradiente conjugados linear. Mostre que em cada nova iteração, o iterado x^k se aproxima (positivamente) à solução otima x^*, isto é, $\|x^* - x^k\|$ é uma sequência *estritamente* decrescente. Para isso faça o seguinte:

    (a) Mostre que $\langle d^i, d^j\rangle > 0$, para $i \ne j$.

    (b) Compare $\|x^* - x^k\|^2$ com $\|x^* - x^{k-1}\|^2$, escreva $x^k - x^{k-1}$ como combinação linear das direções $\{d^i\}$ e use o item anterior. Conclua.

20. Considere os matrizes e vetores

$$Q = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 2 \\ -1 & 2 & 4 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, v = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, w = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}.$$

    (a) Verifique de v e w são Q-conjugados

(b) Minimize a função quadrática $q(x) := \frac{1}{2}x^TQx - b^Tx$ sobre o plano gerado pelos vetores $\{v, w\}$.
Dica: Use como guia e informação o item anterior

MÉTODO DE REGIÃO DE CONFIANÇA

21. Mostre que d^* é a solução do problema

$$\min\ m(d) := f + g^Td + \frac{1}{2}d^TBd \quad \text{sujeito a} \quad \|d\|_2 \leq \Delta,$$

se, e somente se existe $\lambda \geq 0$ tal que (i) $(B + \lambda I)d^* = -g$, (ii) $\|d^*\| \leq \Delta$, (iii) $\lambda(\|d^*\| - \Delta) = 0$ e (iv) $B + \lambda I$ é uma matriz semi-definida positiva.

Conclua que se a solução ótima d^* está no interior da bola $\{d : \|d\| \leq \Delta\}$ temos que $\nabla m(d^*) = 0$ e $B \succeq 0$.

22. (*Generalização do passo de Cauchy.*) Considere D uma matriz $n \times n$ defina positiva.

(a) Seja d_G^k solução otima de

$$\min\ f(x^k) + d^Tg^k \text{ sujeito a } \|Dd\| \leq \Delta.$$

Prove que $d_G^k = -\frac{\Delta_k}{\|Dg^k\|}D^{-2}g^k$.

(b) O passo de Cauchy Generalizado é definida como

$$m_k(d_{CG}^k) = \min\{m_k(d) : d = \tau d_G^k, \|Dd\| \leq \Delta\}.$$

Assim, o passo de Cauchy Generalizado é

$$d_{CG}^k = \tau_k d_G^k = -\tau_k \frac{\Delta_k}{\|Dg^k\|}D^{-2}g^k,$$

onde $\tau_k :=$ argmin $m_k(\tau d_G^k)$ s.a $\|\tau D d_G^k\| \leq \Delta$. Mostre a seguinte expressão para τ_k

$$\tau_k = \begin{cases} 1 & \text{se } (g^k)^TD^{-2}B_kD^{-2}g^k \leq 0 \\ \min\{\frac{\|D^{-1}g^k\|^3}{\Delta_k(g^k)^TD^{-2}B_kD^{-2}g^k}, 1\} & \text{caso contrário} \end{cases}$$

Observação: Se $D = I$, recuperamos o passo de Cauchy.

23. (a) Descreva o método de região de confiança

Considere as seguintes hipóteses

(H1). A solução aproximada do modelo d^k satisfaz $\text{pred}_k = m_k(0) - m_k(d^k) \geq c_1\|g_k\|\min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}$ para certo $c_1 \in (0, 1)$.

(H2). O passo d^k satisfaz $\|d^k\| \leq \gamma\Delta_k$ para certo $\gamma \geq 1$.

(H3). As hessianas $\{B_k\}$ são uniformemente limitadas por alguma constante β, i.e., $\|B_k\| \leq \beta$, $\forall k$.

(a) Verifique que o passo de Cauchy d_C^k satisfaz a desigualdade descrita em (H1) com $c_1 = 1/2$.

(b) Seja $\{x^k\}$ uma sequência gerada pelo método de região de confiança. Mostre que se $f \in C^1$ e as hipóteses (H1), (H2) e (H3) são satisfeitas. Então:

$$|\rho_k - 1| \leq \frac{\gamma\Delta_k\left(\frac{\beta}{2}\gamma\Delta_k + \sup_{\theta\in[0,1]}\|\nabla f(x^k + \theta_k d^k) - \nabla f(x^k)\|\right)}{c_1\|g_k\|\min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}}$$

(c) Usando as mesma hipóteses do item anterior, conclua que depois de um número finito de passo mal sucedidos, temos um passo bem sucedido.

(d) Seja $\{x^k\}$ uma sequência gerada pelo método de região de confiança. Mostre que se $f \in C^1$ com ∇f uniformemente contínua e as hipóteses (H1), (H2) e (H3) são satisfeitas. Então, $\liminf \nabla f(x^k) = 0$.

(e) Seja $\{x^k\}$ uma sequência gerada pelo método de região de confiança. Mostre que se $f \in C^2$ e as hipóteses (H1), (H2) e (H3) são satisfeitas, com $B_k := \nabla^2 f(x^k)$, $\forall k$. Então, $\liminf \nabla f(x^k) = 0$ e $\liminf \lambda_{min}\nabla^2 f(x^k) \geq 0$.