Lista 2: Otimização II

A. Ramos *

September 5, 2017

Abstract

Lista em constante atualização.

1. Gradiente conjugados
2. Métodos de Região de Confiança.
3. Para os exercícios que forem convenientes pode ser usado alguma linguagem de programação.

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

2. 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 $x^0 = (2, 1)^T$

3. 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 $|\beta_k| \leq \beta_k^{FR}$, $\forall k \in \mathbb{N}$. Então, mostre que
$$-\frac{1}{1-c_2} \leq \frac{\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\|^2} \leq \frac{2c_2 - 1}{1 - c_2}, \text{ para } k = 0, 1, 2, \ldots,$$
Conclua que as direções d^k são direções de descida.

4. Considere Q uma matriz simétrica definida positiva. Prove ou dê um contra-exemplo:
   (a) Se Q é múltiplo da identidade, então as direções Q-conjugadas são ortogonais
   (b) Se Q é diagonal, então as direções Q-conjugadas são ortogonais
   (c) Autovetores de Q associados a autovalores distintos são Q-conjugados.

5. Seja Q uma matriz $n \times n$ simétrica definida positiva. Mostre que o método de gradiente conjugados linear resolve o sistema $Ax = b$, usando como máximo n iterações.

6. Seja Q uma matriz simétrica definida positiva e considere $\{v^1, \ldots, v^n\}$ uma família de vetores linearmente independentes. Defina
$$d^1 := v^1 \quad \text{e} \quad d^{k+1} := v^{k+1} - \sum_{i=1}^{k} \theta_i^{k+1} d^i, \quad \text{para} \quad k = 1, 2, \ldots, n-1,$$
onde $\theta_i^{k+1} = (v^{k+1})^T Q d^i / (d^i)^T Q d^i$. Mostre que $\{d^1, \ldots, d^n\}$ são Q-conjugados.

7. 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 $g^k := \nabla q(x^k)$.

8. Dado $n \in \mathbb{N}$. Considere a matriz A $n \times n$ e o vetor $b \in \mathbb{R}^n$ definidos como $b_i = 1$, $\forall i$ e $A_{ij} = \frac{1}{i+j-1}, \forall i, j$ (a matriz A é chamada de *matriz de Hilbert*). Use o método de Gradiente conjugado, para resolver $Ax = b$, com ponto inicial $x^0 = 0$ para diferente valores de $n \in \mathbb{N}$, em especial, $n = 5, 10, 20$.

   Observação: A matriz de Hilbert é o exemplo clássico de matriz mal condicionada.

9. Mostre que d^* é a solução do problema
$$\min\ m(d) := f + g^T d + \frac{1}{2} d^T B d \quad \text{sujeito a} \quad \|d\|_2 \leq \Delta,$$
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 otima d^* está no interior da bola $\{d : \|d\| \leq \Delta\}$ temos que $\nabla m(d^*) = 0$ e $B \succeq 0$ [1].

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

[1] Essa observação foi discutida em aula

10. Considere o problema quadrático $\min\ m(d) := f + g^T d + \frac{1}{2} d^T B d$ sujeito a $\|d\|_2 \leq \Delta$.

(a) Considere a decomposição espectral de B, $B = U^T \Lambda U$, onde U é uma matriz ortogonal, Λ é uma matriz diagonal e $\lambda_i := \Lambda_{ii}$. Prove que se $\lambda > -\lambda_{min}(B)$, o sistema linear $(B + \lambda I)d = -g$ admite uma única solução denotado por $d(\lambda)$ e

$$\|d(\lambda)\|^2 = \sum_{i=1}^{n} \frac{|(Ug)_i|^2}{(\lambda_i + \lambda)^2}.$$

(b) *Nas mesmas hipóteses do item anterior.* Certamente, $\phi(\lambda) := \|d(\lambda)\|$ tem finitos polos mas não possui zeros. Assim, $\psi(\lambda) := 1/\phi(\lambda)$ tem zeros mas não tem polos. Portanto em lugar de resolver $\|d(\lambda)\| - \Lambda = 0$ é conveniente resolver $\frac{1}{\|d(\lambda)\|} - \frac{1}{\Lambda} = 0$ (dita equação é chamada de *equação secular*). Usualmente para resolver a equação secular é usado o método de Newton. Assim,

i. Verifique que $\nabla_\lambda d(\lambda) = -(B + \lambda)^{-1} d(\lambda)$

ii. Mostre que a derivada de $\psi'(\lambda)$ é

$$\psi'(\lambda) = -\frac{\langle d(\lambda), \nabla_\lambda d(\lambda)\rangle}{\|d(\lambda)\|^3}$$

iii. Calcule a segunda derivada de $\psi''(\lambda)$

$$\psi''(\lambda) = -\frac{3(\langle d(\lambda), \nabla_\lambda d(\lambda)\rangle^2 - \|d(\lambda)\|^2 \|\nabla_\lambda d(\lambda)\|^2)}{\|d(\lambda)\|^5}$$

iv. Descreva o método de Newton desse caso.

11. Resolva o sistema $Ax = b$ usando o método de gradientes conjugados com $x^0 = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$, onde $A = \begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix}$ e $b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

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

(a) Seja d_G^k solução otima de

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

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

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

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

13. 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)

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

14. 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

15. Considere a função $f : \mathbb{R} \to \mathbb{R}$ $f(x) := x^3 - x$.

    (a) Encontre todos os pontos máximos e mínimos (locais e globais) de f

    (b) Faça duas iterações do método de região de confiança para o problema $\min f(x)$. Use $\Delta_0 = 1/4$ e $x^0 = 0$.

16. Seja $f(x, y) = y^2 + \frac{1}{2}x^2$ e $x^0 = (1, 1)^T$. Para $\Delta_0 = 1$ e $\Delta_0 = 5/4$. Use o método dogleg para encontrar x^1.

17. 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 $i \neq j$.

    (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.

18. (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^k_C 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 $\liminf \lambda_{min}\nabla^2 f(x^k) \geq 0$.