PSI 2533
Lista 1 — Módulo de Filtragem Adaptativa

Prof. Vítor H. Nascimento

15 de outubro de 2014

Nesta lista vamos estudar um pouco o problema de achar o mínimo de uma função diferenciável, o algoritmo do gradiente, e recordar algumas propriedades de processos estocásticos.

Antes, vamos precisar de duas definições e de um teorema:

Definição: A matriz $\boldsymbol{A} \in \mathbb{R}^{M\times M}$ é dita *positiva-definida* se, para qualquer vetor $\boldsymbol{x} \in \mathbb{R}^M$ *diferente do vetor nulo* valer

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0.$$

A matriz é *positiva semi-definida* se valer, para qualquer vetor $\boldsymbol{x} \in \mathbb{R}^M$,

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0.$$

Definição: A *hessiana* de uma função de várias variáveis $f(x_1, \dots, x_M)$ é a matriz $\boldsymbol{H}$

$$\boldsymbol{H}(x_1, \dots, x_M) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

Teorema: Toda matriz $n \times n$ simétrica (isto é, $\boldsymbol{A} = \boldsymbol{A}^T$) pode ser diagonalizada por uma matriz $\boldsymbol{Q}$ tal que $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$ (a matriz identidade — diz-se que $\boldsymbol{Q}$ é *ortogonal* ou *unitária*). Além disso, todos os seus autovalores são números reais. Assim, temos:

$$\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_i) = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_M \end{bmatrix},$$

com $\lambda_i \in \mathbb{R}$, $1 \leq i \leq M$.

Exercícios:

1. Mostre que as colunas de $\boldsymbol{Q}$ são autovetores de $\boldsymbol{A}$. Quais são os autovalores correspondentes?

2. Considere a função de várias variáveis

$$f(x_1, x_2, \dots, x_M) = \begin{bmatrix} x_1 & x_2 & \dots & x_M \end{bmatrix} \boldsymbol{A} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix},$$

em que $\boldsymbol{A} \in \mathbb{R}^{M\times M}$ é uma matriz qualquer. Mostre que sempre é possível reescrever f em termos de uma matriz $\boldsymbol{R}$ simétrica. Mostre como obter $\boldsymbol{R}$ a partir de $\boldsymbol{A}$. Dica: lembre que $y^T = y$ se y for um escalar. Qual é expressão da transposta de f?

3. Mostre que a matriz simétrica $\boldsymbol{R} \in \mathbb{R}^{M\times M}$ é positiva-definida se e somente se todos os seus autovalores forem positivos. A matriz será positiva semi-definida se os autovalores forem maiores ou iguais a zero. Dica: use a forma diagonalizada da matriz.

4. Considere a função de várias variáveis

$$f(\boldsymbol{x}) = c + 2\boldsymbol{x}^T\boldsymbol{b} + \boldsymbol{x}^T\boldsymbol{R}\boldsymbol{x},$$

em que c é uma constante (real), $\boldsymbol{b} \in \mathbb{R}^M$ é um vetor constante e $\boldsymbol{R} \in \mathbb{R}^{M\times M}$ é uma matriz simétrica.

   (a) Mostre que o ponto em que o gradiente de f se anula é $\boldsymbol{x}_* = -\boldsymbol{R}^{-1}\boldsymbol{b}$.

   (b) Calcule a hessiana de $f(\boldsymbol{x})$.

   (c) Mostre que $\boldsymbol{x}_*$ é um ponto de mínimo se e somente $\boldsymbol{R}$ for positiva semi-definida, e um ponto de máximo se $-\boldsymbol{R}$ for positiva semi-definida.

   (d) Considere $c = 1$, $\boldsymbol{b} = \begin{bmatrix}1 & -1\end{bmatrix}^T$ e três casos para $\boldsymbol{R}$:

      i. $\boldsymbol{R} = \begin{bmatrix}1 & 0{,}5\\ 0{,}5 & 1\end{bmatrix}$,

      ii. $\boldsymbol{R} = \begin{bmatrix}1 & 0{,}5\\ 0{,}5 & 0{,}2\end{bmatrix}$,

      iii. $\boldsymbol{R} = \begin{bmatrix}-1 & 0{,}5\\ 0{,}5 & -0{,}5\end{bmatrix}$.

   Para cada caso, determine se o $\boldsymbol{x}_*$ é um ponto de mínimo ou de máximo. Use a função `contour` do Matlab para desenhar as curvas de nível de f em cada caso, e ver o que acontece.

   (e) Para o(s) caso(s) em que $\boldsymbol{x}_*$ é um ponto de mínimo, determine a faixa de passo μ para a qual o algoritmo do gradiente converge para $\boldsymbol{x}_*$, começando de uma condição inicial qualquer. Teste o resultado no programa que você fez em aula.

5. Suponha agora que $f(\boldsymbol{x})$ seja uma função de várias variáveis, com gradiente $\nabla f(\boldsymbol{x})$ e hessiana $\boldsymbol{H}(\boldsymbol{x})$. A função pode ser expandida em série de Taylor em torno de um ponto $\boldsymbol{x}$ qualquer, assim:

$$f(\boldsymbol{y}) \approx f(\boldsymbol{x}) + (\boldsymbol{y}-\boldsymbol{x})^T\nabla f(\boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^T\boldsymbol{H}(\boldsymbol{x})(\boldsymbol{y}-\boldsymbol{x}).$$

Mostre que os mínimos de f ocorrem nos pontos $\boldsymbol{x}_*$ tais que $\nabla(\boldsymbol{x}_*) = \boldsymbol{0}$, e $\boldsymbol{H}(\boldsymbol{x}_*)$ é positiva definida.

Recordação de processos estocásticos:

1. Considere o vetor de variáveis aleatórias $\boldsymbol{x} = \begin{bmatrix}x_1 & \dots & x_M\end{bmatrix}^T$. Mostre que a matriz de autocorrelação $\boldsymbol{R} = \mathrm{E}\{\boldsymbol{x}\boldsymbol{x}^T\}$ é sempre positiva semi-definida. Dica: calcule $\mathrm{E}\{y^2\}$, em que $y = \boldsymbol{b}^T\boldsymbol{x}$, e $\boldsymbol{b}$ é um vetor constante.

2. Considere um ruído branco Gaussiano $x[n]$ com média nula e potência média 2. Imagine que $x[n]$ passa por um filtro com resposta ao impulso $h[n]$, em que $h[0] = 1$, $h[1] = 0.5$ e $h[n] = 0$ para todos os outros valores de n. Calcule a média e a função de autocorrelação de $y[n]$.