CM106 - Otimização I

12 de Abril de 2018 - Prova 1

Gabarito

1. $\boxed{15}$ Considere a função $f(x,y,z) = x^3 - 3x + y^3 - 12y + z^3 - 27z$. Calcule seus pontos críticos e classifique-os.

> **Solution:** O gradiente de f é $\nabla f(x,y,z) = \left\langle 3x^2 - 3, 3y^2 - 12, 3z^2 - 27 \right\rangle$. Os 8 pontos críticos são $(\pm 1, \pm 2, \pm 3)$.
>
> A Hessiana é $\nabla^2 f(x,y,z) = \text{diag}(6x, 6y, 6z)$, então os autovalores são $6x$, $6y$ e $6z$. Sendo assim, a Hessiana só é definida positiva se x, y, z são positivos, e análogo para negativa. Logo $(1,2,3)$ é minimizador local, $(-1,-2,-3)$ é maximizador local e os outros 6 pontos são pontos de sela.

2. Faça o que se pede:

   (a) $\boxed{20}$ A partir de $x^{(0)} = (2,1)$ para $f(x) = x_1^2 + 2x_2^2$, calcule $x^{(1)}$ usando o Método do Gradiente com busca por *backtracking* com parâmetro 0.5, e satisfazendo a condição de Armijo com parâmetro 0.5.

   > **Solution:** $f(x^0) = 6$. $\nabla f(x) = (2x_1, 4x_2)^T$. $\nabla f(x^0) = (4,4)^T$. $d_0 = (-4,-4)^T$. $\nabla f(x^0)^T d_0 = -32$.
   >
   > $t = 1$
   >
   > $f(2-4, 1-4) = f(-2,-3) = 22 > 6$
   >
   > $t = 1/2$
   >
   > $f(2-2, 1-2) = f(0,-1) = 2$
   >
   > $f(2,1) - 1/2 \times 1/2 \times 32 = 6 - 8 = -2 < 2$
   >
   > $t = 1/4$
   >
   > $f(2-1, 1-1) = f(1,0) = 1$
   >
   > $f(2,1) - 1/4 \times 1/2 \times 32 = 6 - 4 = 2 > 1$
   >
   > $x^{(1)} = x^{(0)} + \frac{1}{4} d_0 = (1,0)^T$.

   (b) $\boxed{10}$ O Método SR1 para aproximar a Hessiana da função f é dado pela atualização

   $$B_{k+1} = B_k + \frac{v_k v_k^T}{v_k^T s_k},$$

   onde $v_k = y_k - B_k s_k$, $s_k = x^{(k+1)} - x^{(k)}$, e $y = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$. Considerando $B_0 = I$, calcule B_1, usando $x^{(0)}$ e $x^{(1)}$ do passo anterior.

Solution: $s_0 = (1,0)^T - (2,1)^T = (-1,-1)^T$. $y_0 = \nabla f(1,0) - \nabla f(2,1) = (2,0)^T - (4,4)^T = (-2,-4)^T$.
$v_0 = \begin{bmatrix} -2 \\ -4 \end{bmatrix} - I \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -3 \end{bmatrix}$. $v_0^T s_0 = (-1,-3)(-1,-1)^T = 4$.

$$B_1 = I + \frac{1}{4}\begin{bmatrix} -1 \\ -3 \end{bmatrix}\begin{bmatrix} -1 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{4}\begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 5 & 3 \\ 3 & 13 \end{bmatrix}.$$

3. $\boxed{15}$ O problema de Quadrados Mínimos Regularizado consiste de minimizar a função

$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \frac{\lambda}{2}\|x\|^2,$$

onde $A \in \mathbb{R}^{m\times n}$, com $m > n$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$ e $\lambda \geqslant 0$ é um parâmetro de regularização. Mostre que esse problema sempre tem minimizador global único para $\lambda > 0$, e descreva esse minimizador.

Solution:

$$\nabla f(x) = A^T(Ax - b) + \lambda x = (A^TA + \lambda I)x - A^Tb.$$
$$\nabla^2 f(x) = A^TA + \lambda I.$$

A matriz $A^TA + \lambda I$ é definida positiva se $\lambda > 0$:

$$y^T(A^TA + \lambda I)y = \|Ay\|^2 + \lambda\|y\|^2 > 0, \forall y \neq 0.$$

Logo, ela é inversível, e o sistema $\nabla f(x) = 0$ terá solução. Como é definida positiva, $\nabla^2 f(x)$ é definida positiva, e essa solução é minimizador local. Como $\nabla^2 f(x)$ é definida positiva para todo x, então f é convexa, e seu minimizador local é global.

4. Faça o que se pede:

(a) $\boxed{10}$ Mostre que a função $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(x) = \sum_{i=1}^n |x_i|$ é convexa, mas que a função $g : \mathbb{R}^2 \to \mathbb{R}$ dada por $g(x,y) = |xy|$ não é convexa.

Solution: Sejam $x, y \in \mathbb{R}^n$. Daí,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \sum_{i=1}^n |\lambda x_i + (1-\lambda)y_i| \leqslant \sum_{i=1}^n \lambda|x_i| + (1-\lambda)|y_i| \\ &= \lambda\sum_{i=1}^n |x_i| + (1-\lambda)\sum_{i=1}^n |y_i| = \lambda f(x) + (1-\lambda)f(y). \end{aligned}$$

Por outro lado, $g(3,1) = g(1,3) = |3| = 3$, mas $g(2,2) = |2\times 2| = 4$. $(2,2) = \frac{1}{2}\Big((3,1) + (1,3)\Big)$, mas $g(2,2) > \frac{1}{2}\Big(f(3,1) + f(1,3)\Big)$.

(b) $\boxed{10}$ Mostre que se $\{x^{(k)}\}$ é uma sequência gerada pelo Método do Gradiente com Busca Exata, então $\nabla f(x^{(k)})$ é ortogonal à $\nabla f(x^{(k+1)})$.

> **Solution:** $x^{(k+1)} = x^{(k)} + t_k d_k$, onde $d_k = -\nabla f(x^{(k)})$ e $t_k = \arg\min_t f(x^{(k)} + t d_k)$. Seja $\varphi(t) = f(x^{(k)} + t d_k)$. Logo, $\varphi'(t_k) = 0$. Mas,
>
> $$\varphi'(t) = d_k^T \nabla f(x^{(k)} + t d_k),$$
>
> de modo que
>
> $$0 = \varphi'(t_k) = d_k^T \nabla f(x^{(k+1)}) = -\nabla f(x^{(k)})^T \nabla f(x^{(k+1)}).$$

(c) $\boxed{10}$ Considere $f(x) = (x_1 - 1)^2 x_2$, e $\bar{x} = (1, 0)$. Mostre que $\bar{x}$ é um ponto crítico e descubra se é minimizador, maximizador ou ponto de sela, utilizando características da função.

> **Solution:** Temos $\nabla f(x) = \begin{bmatrix} 2(x_1 - 1)x_2 \\ (x_1 - 1)^2 \end{bmatrix}$, logo $\nabla f(1, 0) = 0$, ou seja $\bar{x}$ é ponto crítico.
> Temos $\nabla^2 f(x) = \begin{bmatrix} 2x_2 & 2(x_1 - 1) \\ 2(x_1 - 1) & 0 \end{bmatrix}$, logo $\nabla^2 f(0) = 0$, que não ajuda nada.
> Por outro lado, $f(1 + t, t) = t^3$, que para $t > 0$ é positivo, e para $t < 0$ é negativo. Como $f(1, 0) = 0$, então $(1, 0)$ é ponto de sela.

(d) $\boxed{10}$ Prove ou dê um contra-exemplo: "Se $\bar{x}$ é minimizador local, mas $\nabla^2 f(\bar{x})$ não é definida positiva, então o Método de Newton Puro nunca converge para $\bar{x}$, a partir de qualquer $x^{(0)} \neq \bar{x}$.

> **Solution:** A afirmação é falsa. Considere $f(x) = \sum_{i=1}^n x_i^4$. Temos $\nabla f(x) = 4(x_1^3, \dots, x_n^3)^T$ e $\nabla^2 f(x) = \text{diag}(12x_1^2, \dots, 12x_n^2)$. A iteração de Newton vira
>
> $$x_i^+ = x_i - \frac{4x_i^3}{12x_i^2} = \frac{2}{3} x_i,$$
>
> ou seja $x^{(k+1)} = \dfrac{2}{3} x^{(k)} \to 0$, que é o minimizador local de f. Note que a velocidade de convergência é linear, que é o problema gerado por não ser definida positiva.

5. **(Bônus)** Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável até segunda ordem. Um método de otimização para encontrar um minimizador de f é o seguinte:

   1. Dado $x^{(0)} \in \mathbb{R}^n$, $\sigma_0 > 0$, faça $k = 0$;
   2. Defina
   $$m_k(d) = \frac{1}{2} d^T B_k d + d^T g_k + \frac{\sigma_k}{3} \|d\|^3,$$

onde $B_k = \nabla^2 f(x^{(k)})$ e $g_k = \nabla f(x^{(k)})$;

3. Encontre d_k minimizador de m_k;
4. Se houve decréscimo suficiente, faça $x^{(k+1)} = x^{(k)} + d^{(k)}$ e escolha $\sigma_{k+1} \leqslant \sigma_k$;
5. Senão, faça $x^{(k+1)} = x^{(k)}$ e escolha $\sigma_{k+1} > \sigma_k$.
6. Incremente k e volte ao passo 2.

Faça o que se pede:

(a) $\boxed{10}$ A parte difícil do método é encontrar o minimizador de m_k. Ao invés, podemos buscar um minimizador para $\varphi(t) = m_k(-tg_k)$, com $t > 0$. Mostre que $-g_k$ é uma direção de descida para m_k no ponto $d = 0$.

Solution: Podemos calcular ∇m_k. Veja que $\|d\|^3$ não é simplesmente cair a potência, pois $\|d\|^3 = (d^T d)^{3/2}$.

$$\nabla m_k(d) = B_k d + g_k + \frac{\sigma_k}{3}\frac{3}{2}(d^T d)^{1/2} 2d = B_k d + g_k + \sigma_k \|d\| \, d.$$

Daí,

$$\nabla m_k(0) = g_k,$$

e $-g_k$ é obviamente direção de descida.

A outra maneira de fazer isso é olhar para $m_k(-tg_k)$:

$$m_k(-tg_k) = \frac{t^2}{2} g_k^T B_k g_k - t g_k^T g_k + \frac{\sigma_k t^3}{3} \|g_k\|^3 .$$

Daí, $\dfrac{m_k(-tg_k) - m_k(0)}{t} \to -\|g_k\|^2 < 0$.

(b) $\boxed{10}$ Prove que m_k sempre tem um minimizador para $\sigma_k > 0$.

Solution: Uma maneira de mostrar isso é ver que $m_k(d) \to \infty$ quando $\|d\| \to \infty$. Logo, dado $M > 0$ suficientemente grande, o conjunto $L = \{d \in \mathbb{R}^n \mid m_k(d) \leqslant M\}$ é não vazio e limitado. Note que L é fechado, logo é compacto, e portanto vale o Teorema de Weierstrass do valor extremo: existe um minimizador de m_k em L.

O aluno pode usar coercividade para provar diretamente.