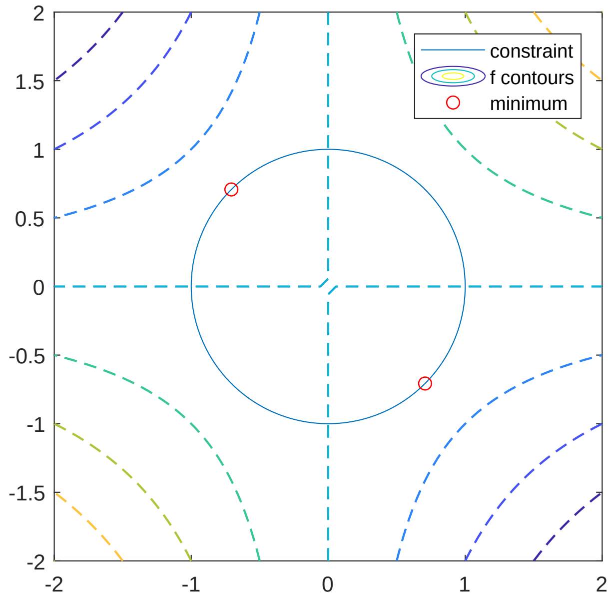
最优化第十七次作业

张晋 15091060

2017 年 12 月 21 日

7.9 函数图像如下:



$$\mathcal{L}(\boldsymbol{x}, \lambda) = x_1 x_2 + \lambda(x_1^2 + x_2^2 - 1)$$

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda) = \begin{bmatrix} 2\lambda x_1 + x_2 \\ x_1 + 2\lambda x_2 \end{bmatrix}$$

$$\nabla^2_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda) = \begin{bmatrix} 2\lambda & 1 \\ 1 & 2\lambda \end{bmatrix}$$

KKT 条件:

$$2\lambda x_1 + x_2 = 0 \tag{1}$$
$$x_1 + 2\lambda x_2 = 0 \tag{2}$$
$$x_1^2 + x_2^2 - 1 = 0 \tag{3}$$

解以上方程得:

$$\begin{aligned}
\boldsymbol{x}^{(1)} &= (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T, \qquad \lambda^{(1)} = -\frac{1}{2} \\
\boldsymbol{x}^{(2)} &= (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})^T, \qquad \lambda^{(2)} = \frac{1}{2} \\
\boldsymbol{x}^{(3)} &= (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T, \qquad \lambda^{(3)} = \frac{1}{2} \\
\boldsymbol{x}^{(4)} &= (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})^T, \qquad \lambda^{(4)} = -\frac{1}{2}
\end{aligned}$$

对 $\boldsymbol{x}^{(1)}$:

$$\nabla_{\boldsymbol{x}}^2 \mathcal{L}(\boldsymbol{x}^{(1)}, \lambda) = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \boldsymbol{W}_1$$
$$G(\boldsymbol{x}^{(1)}, \lambda^{(1)}) = k(1, -1)^T = \boldsymbol{p}_1$$
$${\boldsymbol{p}_1}^T W \boldsymbol{p}_1 = -4k^2 < 0$$

故 $\boldsymbol{x}^{(1)}$ 非局部最优解。

对 $\boldsymbol{x}^{(2)}$:

$$\nabla_{\boldsymbol{x}}^2 \mathcal{L}(\boldsymbol{x}^{(2)}, \lambda) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \boldsymbol{W}_2$$
$$G(\boldsymbol{x}^{(2)}, \lambda^{(2)}) = k(1, 1)^T = \boldsymbol{p}_2$$
$${\boldsymbol{p}_2}^T W \boldsymbol{p}_2 = 4k^2 > 0$$

故 $\boldsymbol{x}^{(2)}$ 是局部最优解。

对 $\boldsymbol{x}^{(3)}$：

$$\nabla_{\boldsymbol{x}}^2 \mathcal{L}(\boldsymbol{x}^{(3)}, \lambda) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \boldsymbol{W}_3$$

$$G(\boldsymbol{x}^{(3)}, \lambda^{(3)}) = k(1,1)^T = \boldsymbol{p}_3$$

$$\boldsymbol{p}_3{}^T W \boldsymbol{p}_3 = 4k^2 > 0$$

故 $\boldsymbol{x}^{(3)}$ 是局部最优解。

对 $\boldsymbol{x}^{(4)}$：

$$\nabla_{\boldsymbol{x}}^2 \mathcal{L}(\boldsymbol{x}^{(4)}, \lambda) = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \boldsymbol{W}_4$$

$$G(\boldsymbol{x}^{(4)}, \lambda^{(4)}) = k(1,-1)^T = \boldsymbol{p}_4$$

$$\boldsymbol{p}_4{}^T W \boldsymbol{p}_4 = -4k^2 < 0$$

故 $\boldsymbol{x}^{(4)}$ 非局部最优解。

综上: 该问题极小点为 $\boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}$，极小值为 -0.5

7.10 **必要性: 若 $[f, C]$ 是凸集**

设

$$\big(\boldsymbol{x}_0, f(\boldsymbol{x}_0)\big), \big(\boldsymbol{x}_1, f(\boldsymbol{x}_1)\big) \in [f, C]$$

因为 $[f, C]$ 是凸集，那么

$$\big(\boldsymbol{x}_\theta = (1-\theta)\boldsymbol{x}_0 + \theta \boldsymbol{x}_1, (1-\theta) f(\boldsymbol{x}_0) + \theta f(\boldsymbol{x}_1)\big) \in [f, C]$$

根据 $[f, C]$ 的定义，有:

$$f(\boldsymbol{x}_\theta) \leq (1-\theta) f(\boldsymbol{x}_0) + \theta f(\boldsymbol{x}_1)$$

故 $f(\boldsymbol{x})$ 是凸函数.

充分性: 若 $f(\boldsymbol{x})$ 是凸函数

那么对 $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_\theta$ 满足

$$\big(\boldsymbol{x}_0, f(\boldsymbol{x}_0)\big), \big(\boldsymbol{x}_1, f(\boldsymbol{x}_1)\big), \big(\boldsymbol{x}_\theta, f(\boldsymbol{x}_\theta)\big) \in [f, C]$$

对 $\boldsymbol{x}_\theta = (1-\theta)\boldsymbol{x}_0 + \theta\boldsymbol{x}_1$，有:

$$f(\boldsymbol{x}_\theta) \le (1-\theta)f(\boldsymbol{x}_0) + \theta f(\boldsymbol{x}_1)$$

即:

$$\big(\boldsymbol{x}_\theta, (1-\theta)f(\boldsymbol{x}_0) + \theta f(\boldsymbol{x}_1)\big) \in [f, C]$$

故有:

$$(1-\theta)\big(\boldsymbol{x}_0, f(\boldsymbol{x}_0)\big) + \theta\big(\boldsymbol{x}_1, f(\boldsymbol{x}_1)\big) \in [f, C]$$

即 $[f, C]$ 是凸集.

7.11 (a) 由次梯度的定义有 $f(\boldsymbol{x}+\alpha\boldsymbol{h}) - f(\boldsymbol{x}) \ge \alpha\boldsymbol{p}^T\boldsymbol{h}$ 故有:

$$\lim_{a\to 0^+} \frac{f(\boldsymbol{x}+\alpha\boldsymbol{h}) - f(\boldsymbol{x})}{\alpha} \ge \boldsymbol{p}^T\boldsymbol{h}$$

$$\lim_{a\to 0^-} \frac{f(\boldsymbol{x}+\alpha\boldsymbol{h}) - f(\boldsymbol{x})}{\alpha} \le \boldsymbol{p}^T\boldsymbol{h}$$

又 f 在 $\boldsymbol{x}$ 可微，则有

$$\lim_{a\to 0} \frac{f(\boldsymbol{x}+\alpha\boldsymbol{h}) - f(\boldsymbol{x})}{\alpha} = \boldsymbol{h}^T\nabla f(\boldsymbol{x})$$

且由可微性的定义:

$$\nabla f(\boldsymbol{x}) = \boldsymbol{p}^T\boldsymbol{h}$$

即 $\boldsymbol{p} = \nabla f(\boldsymbol{x}), \quad \partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$

(b) 对于 $x \in (-\infty, 0)$ 时，$f(x) = -x$ 可微，$\partial f(x) = -1$

对于 $x \in (0, \infty)$ 时，$f(x) = x$ 可微，$\partial f(x) = 1$

对于 $x = 0$ 时，由次梯度的定义有:$|\Delta x| \ge p\Delta x$ 成立，

易得 $-1 \le p \le 1$, 即

$$\partial f(x) = \begin{cases} \{-1\}, & x \in (-\infty, 0) \\ \{1\}, & x \in (0, \infty) \\ [-1, 1], & x = 0 \end{cases}$$