最优化第九次作业

张晋 15091060

2017 年 11 月 9 日

1.4 (a) $\nabla(\boldsymbol{a}^T\boldsymbol{x}) = (\nabla\boldsymbol{a}^T)\boldsymbol{x} + (\nabla\boldsymbol{x}^T)\boldsymbol{a} = \boldsymbol{a}$

$\nabla^2(\boldsymbol{a}^T\boldsymbol{x}) = \nabla\boldsymbol{a} = 0$

(b) $\nabla(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) = (\nabla\boldsymbol{x}^T\boldsymbol{A})\boldsymbol{x} + (\nabla\boldsymbol{x}^T)(\boldsymbol{x}^T\boldsymbol{A})^T = (\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{x}$

$\nabla^2(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}+\boldsymbol{A}^T$

(c) $\nabla(\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$

$\nabla^2(\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}) = \nabla(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) = \boldsymbol{A}$

(d) $\nabla(\boldsymbol{r}^T\boldsymbol{r}) = (\nabla\boldsymbol{r}^T)\boldsymbol{r} + (\nabla\boldsymbol{r}^T)\boldsymbol{r} = 2\boldsymbol{A}^T\boldsymbol{r}$

$\nabla^2(\boldsymbol{r}^T\boldsymbol{r}) = \nabla(2\boldsymbol{A}^T\boldsymbol{r}) = 2(\nabla\boldsymbol{A}^T)\boldsymbol{r} + 2\boldsymbol{A}^T\boldsymbol{A}$

1.6

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x'}) + (\boldsymbol{x}-\boldsymbol{x}')^T\nabla\boldsymbol{f}(\boldsymbol{x}') + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}')^T\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}') + o(\|\boldsymbol{x}-\boldsymbol{x}'\|^2) \tag{1}$$

1.7 函数 $\boldsymbol{g}$ 沿 $\boldsymbol{p}$ 方向的斜率为 $\boldsymbol{p}^T\nabla\boldsymbol{g}(x') = \boldsymbol{p}^T\boldsymbol{g}'$

$$\boldsymbol{p}^T\boldsymbol{g}' = \|\boldsymbol{p}^T\boldsymbol{g}'\|_2 \leq \|\boldsymbol{p}^T\|_2 \cdot \|\boldsymbol{g}'\|_2 = \|\boldsymbol{g}'\|_2$$

故当 $\boldsymbol{p}(\boldsymbol{g}')^T = \boldsymbol{p}^T\boldsymbol{g}' = \|\boldsymbol{g}'\|_2$ 时,斜率取到最大值 $\|\boldsymbol{g}'\|_2$,此时 $\boldsymbol{p}(\boldsymbol{g}')^T\boldsymbol{g}' = \boldsymbol{g}'\|\boldsymbol{g}'\|_2$，得 $\boldsymbol{p} = \boldsymbol{g}'/\|\boldsymbol{g}'\|_2$

4.2 (a) $\mathbb{R}^n$ 空间经过 $\boldsymbol{A}$ 变换后形成的新的 m 维空间到点 $\boldsymbol{B}$ 的最短欧式距离。

(b) 设 $g(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$

$$\begin{aligned}
\nabla g &= \nabla \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 && (2)\\
&= \nabla (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) && (3)\\
&= 2[\nabla (\boldsymbol{x}^T \boldsymbol{A}^T - \boldsymbol{b}^T)](\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) && (4)\\
&= 2\boldsymbol{A}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) && (5)
\end{aligned}$$

$\nabla^2 g = 2\boldsymbol{A}^T\boldsymbol{A}$ 为半正定矩阵，故其必要条件是 $\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x} - b) = \boldsymbol{0}$. 显然，它也是充分条件。

(c) 不唯一,满足 $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$ 的解 $\boldsymbol{x}$ 都是最优解,当 $rank(\boldsymbol{A}^T\boldsymbol{A}) \leq n$ 时可能会出现多解.

(d) 若 $\boldsymbol{A}^T\boldsymbol{A}$ 是严格正定的，则最优解 $\boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$

(e)

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 6 & 2 \\ 1 & 2 & 2 \end{bmatrix}, \qquad \boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} = \begin{bmatrix} 3 \\ 3/2 \\ -5/2 \end{bmatrix}$$

4.3

$$\nabla q(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^T)\boldsymbol{x}-\boldsymbol{b}$$

$$\nabla^2 q(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^T)$$

此处题目中没写 $\boldsymbol{G}$ 是否为对称阵，如果 $\boldsymbol{G}$ 不是对称阵，那么第一问的结论将不成立 (下面给出证明)，感觉题目的意思应该是约定俗成了 $\boldsymbol{G}$ 是对称阵，那么此题我继续按照已知 $\boldsymbol{G}$ 是对称阵的条件去做。

(a)（若 $\boldsymbol{G}$ 为对称阵）$\nabla q(\boldsymbol{x}) = \boldsymbol{G}\boldsymbol{x}-\boldsymbol{b}, \nabla^2 q(\boldsymbol{x}) = \boldsymbol{G}$，若 $\boldsymbol{G}$ 半正定，$\boldsymbol{G}\boldsymbol{x}=\boldsymbol{b}$ 有解，那么该解 $\boldsymbol{x}^*$ 即为极小点。同时若 $\boldsymbol{x}^*$ 为极小点，那么也必同时包含以上条件。

（在没有 $\boldsymbol{G}$ 为对称阵的条件下）：若 $\boldsymbol{G}$ 半正定，且 $\boldsymbol{G}\boldsymbol{x}=\boldsymbol{b}$ 有解，则容易得到 $\boldsymbol{G}+\boldsymbol{G}^T$ 半正定的结论，但无法推出 $\frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^T)\boldsymbol{x}=\boldsymbol{b}$ 有解的结论。可以举出例子如下:

$$\boldsymbol{G} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \text{满足半正定条件}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \text{解得}: \boldsymbol{x} = \boldsymbol{G}^{-1}\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{但对于方程}\frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^T)\boldsymbol{x}=\boldsymbol{b}: \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}\text{是无解的}.$$

同时，$\frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^T)\boldsymbol{x}=\boldsymbol{b}$ 有解、$\frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^T)$ 半正定的条件也无法推出 $\boldsymbol{G}\boldsymbol{x}=b$. 举例如下:

$$\boldsymbol{G} = \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}, \quad \frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^T) = \begin{bmatrix} 2 & 1/2 \\ 1/2 & 0 \end{bmatrix}\text{满足半正定条件},$$

$$\text{若}b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \text{则}\frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^T)\boldsymbol{x}=b\text{有解，但}\boldsymbol{G}\boldsymbol{x}=b: \quad \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}\text{无解}.$$

(b) $\boldsymbol{G}$ 正定时，$q(\boldsymbol{x})$ 严格为凸函数，方程 $\boldsymbol{G}\boldsymbol{x}=\boldsymbol{b}$ 有唯一解 $\boldsymbol{x}=\boldsymbol{G}^{-1}\boldsymbol{b}$，该点即为唯一的极小点.

(c) 若 $\boldsymbol{G}$ 半正定，$q(\boldsymbol{x})$ 为凸函数，满足方程 $\boldsymbol{G}\boldsymbol{x}=\boldsymbol{b}$ 的点都为局部极小点，由凸性可知每个局部极小点都为全局极小点.

4.6 先证必要性: 若 f 为凸函数，则对 $\forall x_1, x_n \in \mathbb{R}^n$, 有

$$f(\theta_1 x_1 + (1-\theta_1)x_n) \leq \theta_1 f(x_1) + (1-\theta_1)f(x_n) \tag{6}$$

那么对于

$$f(\theta_1 x_1 + \theta_2 x_2 + (1-\theta_1-\theta_2)x_n) \tag{7}$$

令

$$x' = \frac{\theta_1 x_1 + \theta_2 x_2}{\theta_1 + \theta_2} \tag{8}$$

代入式 (7) 得:

$$f((\theta_1+\theta_2)x' + (1-\theta_1-\theta_2)x_n) \leq (\theta_1+\theta_2)f(x') + (1-\theta_1-\theta_2)f(x_n) \tag{9}$$

根据凸函数性质 (6)，有:

$$f(x') = (\frac{\theta_1}{\theta_1+\theta_2}x_1 + \frac{\theta_2}{\theta_1+\theta_2}x_2) \leq \frac{\theta_1}{\theta_1+\theta_2}f(x_1) + \frac{\theta_2}{\theta_1+\theta_2}f(x_2) \tag{10}$$

将式 (10) 代入式 (9) 得:

$$f(\theta_1 x_1 + \theta_2 x_2 + (1-\theta_1-\theta_2)x_n) \leq \theta_1 f(x_1) + \theta_2 f(x_2) + (1-\theta_1-\theta_2)f(x_n) \tag{11}$$

同理，可将该不等式从 2 扩充到任意 k, 即为所证。

取 $k=2$，充分性即可得证.