最优化理论作业参考答案 [1]

Question 1. (Problem 2.16, *Convex Optimization*)

Show that if S_1 and S_2 are convex sets in $R^{m\times n}$, then so is their partial sum

$$S = \{(x, y_1 + y_2) | x \in R^m, y_1, y_2 \in R^n, (x, y_1) \in S_1, (x, y_2) \in S_2\}. \tag{1}$$

解：

考虑两个点 $(\hat{x}, \hat{y_1} + \hat{y_2}), (\widetilde{x}, \widetilde{y_1} + \widetilde{y_2}) \in S$，也即

$$(\hat{x}, \hat{y_1}) \in S_1, \quad (\hat{x}, \hat{y_2}) \in S_2, \quad (\widetilde{x}, \widetilde{y_1}) \in S_1, \quad (\widetilde{x}, \widetilde{y_2}) \in S_2. \tag{2}$$

对任意 $0 \leq \theta \leq 1$,

$$\theta(\hat{x}, \hat{y_1} + \hat{y_2}) + (1-\theta)(\widetilde{x}, \widetilde{y_1} + \widetilde{y_2}) = (\theta\hat{x} + (1-\theta)\widetilde{x}, \theta\hat{y_1} + (1-\theta)\widetilde{y_1} + \theta\hat{y_2} + (1-\theta)\widetilde{y_2}). \tag{3}$$

由 S_1 和 S_2 的凸性

$$(\theta\hat{x} + (1-\theta)\widetilde{x}, \theta\hat{y_1} + (1-\theta)\widetilde{y_1}) \in S_1, \tag{4}$$
$$(\theta\hat{x} + (1-\theta)\widetilde{x}, \theta\hat{y_2} + (1-\theta)\widetilde{y_2}) \in S_2. \tag{5}$$

因此

$$\theta(\hat{x}, \hat{y_1} + \hat{y_2}) + (1-\theta)(\widetilde{x}, \widetilde{y_1} + \widetilde{y_2}) \in S, \tag{6}$$

从而 S 是一个凸集。

[1] 感谢王筝同学的 Tex 模板和作业参考。

Question 2. (Problem 3.3, *Convex Optimization*)

Inverse of an increasing convex function. Suppose $f : R \to R$ is increasing and convex on its domain (a,b). Let g denote its inverse, *i.e.*, the function with domain $(f(a), f(b))$ and $g(f(x)) = x$ for $a < x < b$. What can you say about convexity of concavity of g?

g 是凹函数

解：

g 的定义域 $(f(a), f(b))$ 显然是凸集。

考虑 $y_1, y_2 \in (f(a), f(b))$，定义 x_1, x_2 为

$$x_1 = g(y_1), \qquad x_2 = g(y_2), \tag{7}$$

根据逆函数的定义，我们有

$$y_1 = f(x_1), \qquad y_2 = f(x_2). \tag{8}$$

从 f 的凸性得

$$\theta f(x_1) + (1-\theta) f(x_2) \geq f(\theta x_1 + (1-\theta) x_2). \tag{9}$$

从 f 的单调性得，$x_1 \leq x_2$ 当且仅当 $f(x_1) \leq f(x_2)$，也即 $g(y_1) \leq g(y_2)$ 当且仅当 $y_1 \leq y_2$，这意味着 g 是单调增函数。

考虑任意 $0 \leq \theta \leq 1$

$$\begin{aligned} g(\theta y_1 + (1-\theta) y_2) &= g(\theta f(x_1) + (1-\theta) f(x_2)) && (10) \\ &\geq g(f(\theta x_1 + (1-\theta) x_2)) && (11) \\ &= \theta x_1 + (1-\theta) x_2 && (12) \\ &= \theta g(y_1) + (1-\theta) g(y_2), && (13) \end{aligned}$$

因此 g 是凹函数。

注意：本题题干没有说明 f 的可微性，因此不能使用凸函数的一阶或二阶条件来证明 g 的凹凸性。但在 f 二阶可导时使用二阶条件证明比较简洁，下面给出在 f 二阶可导时的一种推导过程，供参考。

解：（二阶条件）

因为 $g(f(x)) = x \Rightarrow f(g(x)) = x, g(x) \in [a, b]$ ，对其求一阶导数得到 $f'(g(x))g'(x) = 1$ ，变形成

$$f'(g(x)) = \frac{1}{g'(x)} \tag{14}$$

因为 $f(x)$ 在定义域内是单调递增的，所以 $f'(x) > 0, x \in [a, b]$ 恒成立。推断 $g'(x) > 0$

对等式 14 再次求导

$$f''(g(x))g'(x) = -\frac{g''(x)}{g'^2(x)} \Rightarrow f''(g(x)) = -\frac{g''(x)}{g'^3(x)} \tag{15}$$

这里因为 f 是一个凸函数，因此 $f''(x) \geq 0, x \in [a, b]$ 恒成立，由此可以推断出 $g''(x) < 0$ ，因此 g 是一个凹函数。

Question 3. 一范数规范化最小二乘问题

考虑一范数规范化最小二乘问题

$$\min_{x} \quad \frac{1}{2}||Ax-y||^2+\lambda||x||_1 \tag{16}$$

其中 $x\in R^n, A\in R^{m\times n}, y\in R^m, \lambda\in R_+$ ，存在 $\lambda_{\max}$ 使得，当 $\lambda>\lambda_{\max}$ 时，上述优化问题的最优解一定是 $x=0$

解：

定义 $f(x)=\frac{1}{2}||Ax-y||^2+\lambda||x||_1$，则 $f(x)$ 的次梯度为 $A^T(Ax-y)+\lambda\partial||x||_1$。由一阶最优性条件，某个点 x 为最优点的条件为：

$$0\in A^T(Ax-y)+\lambda\partial||x||_1. \tag{17}$$

由次梯度的定义，一范数的次梯度的第 i 维 $[\partial||x||_1]_i$ 为

$$[\partial||x||_1]_i=\begin{cases}1, & x_i>0,\\ [-1,1], & x_i=0, \qquad i\in[1,n].\\ -1, & x_i<0,\end{cases} \tag{18}$$

由式子(17)和(18)得，只需要令 $A^Ty\in\lambda\partial||0||_1$ 就可以使 $x=0$ 称为优化问题的最优解。此时若

$$-\lambda\le[A^Ty]_i\le\lambda, \tag{19}$$

则条件 $A^Ty\in\lambda\partial||x||_1$ 成立。从而得到最优解是 $x=0$。此时只需满足 $\lambda_{\max}=||A^Ty||_\infty$ 即可，即 $\lambda_{\max}$ 为 A^Ty 绝对值最大的元素的绝对值。

注 1：有同学使用证明 0 是临近点梯度法一个稳定点的方法得出了相同的结论，但这种方法是不对的：优化问题最优解的性质不应该与具体的优化方法相关。

注 2：本题要求只需给出 $\lambda_{\max}=||A^Ty||_\infty$。事实上，当 $\lambda\ge\lambda_{\max}$ 时，0 是一个（可能不唯一的）最优解，当 $\lambda>\lambda_{\max}$ 时，0 是唯一解。部分同学同时了证明解的唯一性，因此此处也给出一个证明供参考。

证：定义 $g(z)=\frac{1}{2}||z-y||^2$，以及 $f(x)=\frac{1}{2}||Ax-y||^2+\lambda||x||_1=g(Ax)+\lambda||x||_1$。

记 x^* 为 $f(x)$ 某一个最优解，同时定义 f^* 为最优值，即 $f(x^*)=f(0)=f^*$。

我们先证 $Ax^*=0$。用反证法，若 $Ax^*\neq0$，注意到 $g(z)$ 是严格凸的函数，$||x||_1$ 为凸函数，这意味着对任意 $0<\theta<1$，

$$\begin{aligned} f(\theta x^*+(1-\theta)0) &= g(A(\theta x^*+(1-\theta)0))+\lambda||\theta x^*+(1-\theta)0||_1 & (20)\\ &< \theta g(Ax^*)+(1-\theta)g(A0)+\theta\lambda||x^*||_1+(1-\theta)\lambda||0||_1 & (21)\\ &= \theta f(x^*)+(1-\theta)f(0) & (22)\\ &= f^*, & (23)\end{aligned}$$

与 f^* 为函数的最小值矛盾，因此 $Ax^*=0$。

再证 $x^*=0$。注意到

$$f(x^*)-f(0)=g(Ax^*)-g(A0)+\lambda\|x^*\|_1-\lambda\|0\|_1=\lambda(\|x^*\|_1-\|0\|_1), \tag{24}$$

但 $f(x^*) - f(0) = 0$，因此当 $\lambda > \lambda_{max} \geq 0$ 时，$\|x^*\|_1 = \|0\|_1 = 0$，即 $x^* = 0$。
当 $\lambda = \lambda_{max}$ 时，解有可能不唯一，例如当 $\lambda = \lambda_{max} = 0$ 时，所有令 $Ax = 0$ 的非零 x 都是原问题的最优解。