作业 6

李邹

人工智能一班 (2020 级)

最优化方法 课程作业

LANZHOU UNIVERSITY

2022 年 5 月 9 日

作业题 1

假设 $f : R \longrightarrow R$ 是凸函数，$a, b \in \textbf{dom} f, a < b$,

(a) 证明对任意 $x \in [a, b]$, 下式成立

$$f(x) \le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$

(b) 证明对任意 $x \in (a, b)$, 下式成立

$$\frac{f(x)-f(a)}{x-a} \le \frac{f(b)-f(a)}{b-a} \le \frac{f(b)-f(x)}{b-x}$$

画一个草图来描述此不等式。

(c) 假设函数 f 可微。利用 (b) 中的结果来证明

$$f'(a) \le \frac{f(b)-f(a)}{b-a} \le f'(b)$$

注意到这些不等式也可以通过 (3.2) 得到：

$$f(b) \ge f(a) + f'(a)(b-a)$$

$$f(a) \ge f(b) + f'(b)(a-b)$$

(d) 假设函数 f 二次可微。利用 (c) 中的结论证明 $f''(a) \ge 0$ 以及 $f''(b) \ge 0$。

证明

(a) 由题，f 是凸函数，$a, b \in \textbf{dom} f$，则有，

$$f(\theta a + (1-\theta)b) \le \theta f(a) + (1-\theta) f(b) \tag{1}$$

令 $\theta = \frac{b-x}{b-a}$，由于 $a < b$，则 $\theta \in [0, 1]$ 一定成立。

代入 (1) 式，则有，

$$f\left(\frac{b-x}{b-a}a + \frac{x-a}{b-a}b\right) = f(x) \le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$

□

(b) 将 (1) 式去分母化，得

$$\begin{aligned} &(b-a)f(x) \le (b-x)f(a) + (x-a)f(b) \\ \Longleftrightarrow &(b-a)f(x) - (b-a)f(a) \le (a-x)f(a) + (x-a)f(b) \\ \Longleftrightarrow &\frac{f(x)-f(a)}{x-a} \le \frac{f(b)-f(a)}{b-a} \end{aligned} \tag{2}$$

同理可得：

$$\begin{aligned}&(b-a)f(x)-(b-a)f(b)\le(b-x)f(a)+(x-b)f(b)\\ \Longleftrightarrow&\frac{f(b)-f(x)}{b-x}\le\frac{f(b)-f(a)}{b-a}\end{aligned}\tag{3}$$

合并 (2)(3) 式，即得，

$$\frac{f(x)-f(a)}{x-a}\le\frac{f(b)-f(a)}{b-a}\le\frac{f(b)-f(x)}{b-x}$$

□

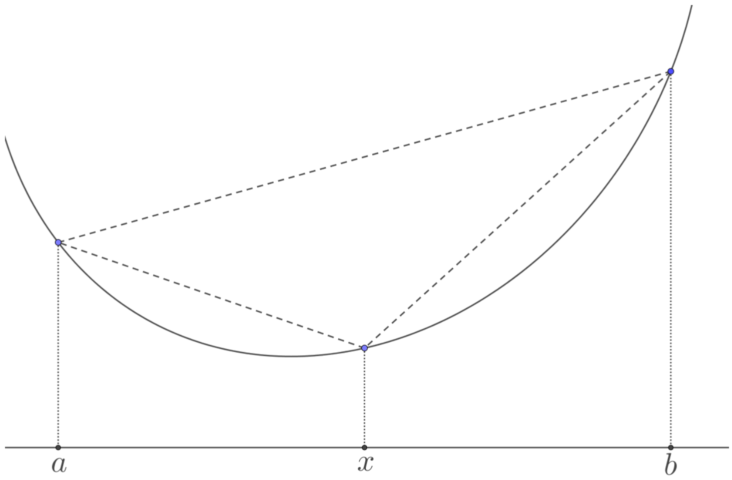

图 1: 描述 (b) 中不等式的草图

(c) 当 $x\to a$ 时，由微分的定义，有

$$\frac{f(x)-f(a)}{x-a}=f'(a)$$

代入 (2) 式，有

$$f'(a)\le\frac{f(b)-f(a)}{b-a}\tag{4}$$

同理可得：

$$\frac{f(b)-f(a)}{b-a}\le f'(b)\tag{5}$$

综合 (4)(5) 式，可得

$$f'(a)\le\frac{f(b)-f(a)}{b-a}\le f'(b)\tag{6}$$

□

(d) 由 (6) 有 $f'(a) \leq f'(b)$，由于 $a < b$，则有：

$$\frac{f'(b) - f'(a)}{b - a} \geq 0 \tag{7}$$

当 $a \to b$ 时，(7) 式可化为：

$$f''(b) \geq 0$$

同理可得：

$$f''(a) \geq 0$$

$\square$

作业题 2

推导下列函数的共轭函数

(a) **最大值函数**。函数 $\max_{i=1,\ldots,n} x_i$，定义在 $\mathbf{R}^n$ 上。

(b) **最大若干分量之和**。函数 $f(x) = \sum_{i=1}^{r} x_{[i]}$，定义域为 $\mathbf{R}^n$。

(c) **定义在 R 上的线性分片函数**。定义在 $\mathbf{R}$ 上的线性分片函数 $f(x) = \max_{i=1,\ldots,m}(a_i x + b_i)$。在求解过程中，可以假设 a_i 按升序排列，即 $a_1 \leq \cdots \leq a_m$，且每一个函数 $a_i x + b_i$ 都不是冗余的，即任选 k，至少存在一点 x 使得 $f(x) = a_k x + b_k$。

(d) **幂函数**。定义在 $\mathbf{R}_{++}$ 上的函数 $f(x) = x^p$，其中 $p > 1$。如果 $p < 0$ 呢？

(e) **几何平均**。定义在 $\mathbf{R}^n_{++}$ 上的几何平均函数 $f(x) = -\left(\prod x_i\right)^{1/n}$。

(f) **二阶锥上的负广义对数**。函数 $f(x,t) = -\log\left(t^2 - x^T x\right)$，定义域为 $\{(x,t) \in \mathbf{R}^n \times \mathbf{R} \mid \|x\|_2 < t\}$。

解答

(a) 由定义有，

$$f^*(y) = \sup_x \left\{ y^T x - \max x \right\}$$

首先考虑 y 分量中是否存在负值，若存在负值，则对应的 $x^T \to \infty$ 时，函数无上界，因此 $y \geq 0$ 必然成立。其次考虑 y 是否存在大于 1 的分量，若存在，则对应的 $x_i \to \infty$ 时，函数无上界，因此 $y \leq 1$ 必然成立。

综上，我们将定义域限制在 $0 \leq y \leq 1$ 的范围中。
设 $x \in \mathbf{R}^n$ 的各分量按从大到小排序为 $x_1 \geq x_2 \geq \cdots \geq x_n$

于是有：

$$f^*(y) = \sup_x \left\{ \sum_{i=1}^n y_i x_i - x_1 \right\} = \sup_x \left\{ (y_1 - 1) x_1 + \sum_{i=2}^n y_i x_i \right\}, \quad \forall i \geq 2 \quad x_i \leq x_1$$

考虑使函数逼近上界的 x 取值，不妨取各个分量相等的向量，即 $x = t\mathbf{1}$，则有：

$$f^*(y) = \sup_x \left\{ (y_1 - 1) x_1 + \sum_{i=2}^n y_i x_i \right\} = \sup_t \left\{ t \left(\sum_{i=1}^m y_i - 1 \right) \right\}$$

容易发现，当且仅当 $\sum_i y_i = 1$ 时，函数值才不受变量 t 的控制，上界一定存在。此时，

$$f^*(y) = 0$$

综上所述：

$$f^*(y) = \begin{cases} 0 & \text{if } y \geq 0, \quad \mathbf{1}^T y = 1 \\ \infty & \text{otherwise} \end{cases}$$

(b) 与 (a) 类似，不难确定本小问的定义域为 $0 \leq y \leq 1$
设 $x \in \mathbf{R}^n$ 的各分量按从大到小排序为 $x_1 \geq x_2 \geq \cdots \geq x_n$
于是有：

$$f^*(y) = \sup \left\{ \sum_{i=1}^r (y_i - 1) x_i + \sum_{i=r+1}^n y_i x_i \right\}$$

考虑最大化函数值的 x 取值，不妨取各个分量相等的向量，即 $x = t\mathbf{1}$，则有：

$$f^*(y) = \sup \left\{ \sum_{i=1}^r (y_i - 1) x_i + \sum_{i=2}^n y_i x_i \right\} = \sup \left\{ t \left(\sum_{i=1}^m y_i - r \right) \right\}$$

容易发现，当且仅当 $\sum_i y_i = r$ 时，函数值才不受变量 t 的控制，上界一定存在。此时，

$$f^*(y) = 0$$

综上所述：

$$f^*(y) = \begin{cases} 0 & \text{if} \quad 0 \leq y \leq 1, \quad \mathbf{1^T} y = r \\ \infty & \text{otherwise} \end{cases}$$

(c) 对于 $f(x) = \max_{i=1,\ldots,m}(a_i x + b_i)$，存在 $m-1$ 个拐点，其表达式为:

$$x_i = \frac{b_i - b_{i+1}}{a_{i+1} - a_i}, \quad i = 1, \ldots, m-1$$

因此，不难求得：

$$f^*(y) = -b_i - \frac{b_{i+1} - b_i}{a_{i+1} - a_i}(y - a_i)$$

其中，$y \in [a_i, a_{i+1}]$

(d) **先考虑 $p > 1$ 的情况**

由 $y = \nabla f(x) = px^{p-1}$ 可得：

$$x = \left(\frac{y}{p}\right)^{\frac{1}{p-1}} \tag{8}$$

当 $y > 0$ 时，将 (8) 式代入共轭函数定义式中，则有：

$$\begin{aligned} f^*(y) &= y\left(\frac{y}{p}\right)^{\frac{1}{p-1}} - \left(\frac{y}{p}\right)^{\frac{p}{p-1}} \\ &= (p-1)\left(\frac{y}{p}\right)^{\frac{p}{p-1}} \end{aligned}$$

当 $y \leq 0$ 时，显然有 $f^*(y) = 0$. 综上所述：

$$f^*(y) = \begin{cases} 0 & y \leq 0 \\ (p-1)\left(\frac{y}{p}\right)^{\frac{p}{p-1}} & y > 0 \end{cases}$$

再考虑 $p < 0$ 的情况

此时定义域变换为 $-\mathrm{R}_{++}$，同理可得：

$$\begin{aligned} f^*(y) &= -y\left(\frac{-y}{p}\right)^{\frac{1}{p-1}} + \left(\frac{-y}{p}\right)^{\frac{p}{p-1}} \\ &= (1-p)\left(\frac{-y}{p}\right)^{\frac{p}{p-1}} \end{aligned}$$

(e) 假设 y 中存在一个大于 0 的分量，设其为 $y_k > 0$

令 $x_k = t, x_i = 1\,(i \neq k)$，当 $t \to \infty$ 时有：

$$f^*(y) = ty_k + t^{\frac{1}{n}} \to \infty$$

因此，$y \notin \mathbf{dom}\, f$

当 y 中不存在一个大于 0 的分量，且 $\left(\prod_i (-y_i)\right)^{\frac{1}{n}} < \frac{1}{n}$ 时。

不妨取 $x = -\frac{t}{y}$，且 $t \to \infty$，则有：

$$f^*(y) = -\mathrm{tn} - t\left(\prod_i \left(-\frac{1}{y_i}\right)\right)^{\frac{1}{n}} > t \to \infty$$

此时，同样有 $y \notin \mathbf{dom}\, f$

当 y 中不存在一个大于 0 的分量，且 $\left(\prod_i (-y_i)\right)^{\frac{1}{n}} \geq \frac{1}{n}$ 时，则有：

$$x^T y = \sum_{i=1}^{n} x_i y_i \leq -n\left(\prod_{i=1}^{n} -x_i y_i\right)^{\frac{1}{n}} \leq -n \cdot \frac{1}{n}\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = f(x)$$

从而有

$$f^*(y) = x^T y - f(x) \leq 0$$

综上所述，$f^*(y) = 0$

(f) 分两种情况讨论。

当 $\|y\|_2 \geq u$ 时

不妨取 $x = sy$, $t = s\left(\|y\|_2 + 1\right) > su$, $s \to \infty$，则有：

$$x^T y + tu + \log\left(t^2 - x^T x\right) > su^2 + su^2 + \log\left(s^2\right) \to \infty$$

上界不存在，因此 $y \notin \mathbf{dom}\, f$

当 $\|y\|_2 < u$ 时

此时 $(y, u) = \nabla f(x, t), f(x, t) = -\log\left(t^2 - x^T x\right)$，计算可得：

$$u = -\frac{2t}{t^2 - x^T x}, y = \frac{2x}{t^2 - x^T x}$$

下面尝试反解出 x, t，不难得到：

$$x = \frac{2y}{u^2 - y^T y}, t = -\frac{2u}{u^2 - y^T y} \tag{9}$$

将 (9) 式代入 $g(y, u)$ 中，即得：

$$\begin{aligned} g(y, u) &= x^T y + tu + \log\left(t^2 - x^T x\right) \\ &= -2 + \log\left(\frac{4}{u^2 - y^T y}\right) \end{aligned}$$

注 在《Convex Optimization Solutions Manual》中，编者给出本小问的参考答案为 $2 + \log 4 - \log\left(y^2 - u^t u\right)$，显然 y 不应为平方的形式，而应当以转置相乘的形式参与运算。值得注意的是，参考答案中 $y^2 - u^t u$ 似乎不在 log 函数所要求的定义域中。