

Homework 2

Instructor: Lijun Zhang*Name:* 方盛俊, *StudentId:* 201300035

Notice

- The submission email is: zhangzhenyao@lamda.nju.edu.cn.
- Please use the provided Latex file as a template.
- If you are not familiar with LaTeX, you can also use Word to generate a **PDF** file.

Problem 1: Convex functions

(a)

令 $g(x) = -\log x$, 求导可得 $g'(x) = -\frac{1}{x}$, $g''(x) = \frac{1}{x^2} > 0$

所以 $g(x) = -\log x$ 是严格凸的.

而 $f_i(x) = -\log x_i = -\log A_i^T x = g(A_i^T x)$, 其中 A_i 第 i 分量为 1, 其他分量为 0, 可以看出 $f_i(x)$ 是凸函数的仿射映射函数, 也是严格凸函数.

因此 $f(x) = \sum_{i=1}^n f_i(x)$ 为严格凸函数的和, 结果也是严格凸的.

(b)

\Rightarrow :

因为 f 是一个二阶可微的凸函数, 因此 $\forall x, y$ 有

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

$$f(x) \geq f(y) + \nabla f(y)^T (x - y)$$

两式相加可得

$$\nabla f(x)^T (x - y) \geq \nabla f(y)^T (x - y)$$

最后有 $(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$ 成立

⇐:

因为我们有 $(\nabla f(x) - \nabla f(y))^T(x - y) \geq 0$

令 $g(t) = f(tx + (1 - t)y)$, 则 $g'(t) = \nabla f(tx + (1 - t)y)^T(x - y)$

即证 $g'(t) \geq g'(0)$, 即 $[\nabla f(tx + (1 - t)y)^T - \nabla f(y)^T](x - y) \geq 0$

经过观察, 计算 $tx + (1 - t)y - y = tx - ty = t(x - y)$, 那么我们只需将 $tx + (1 - t)y$ 带入 x 的位置, 根据 $(\nabla f(x) - \nabla f(y))^T(x - y) \geq 0$ 有

$(\nabla f(x) - \nabla f(y))^T t(x - y) \geq 0$ 成立, 因为 $t \geq 0$, 可知 $g'(t) \geq g'(0)$

最后有 $f(x) = g(1) = g(0) + \int_0^1 g'(t)dt \geq g(0) + g'(0) = g(y) + \nabla f(y)(x - y)$

(c)

因为 f 是凸函数, 因此我们有 $f(\frac{\theta x_1 + (1 - \theta)x_2}{\theta t_1 + (1 - \theta)t_2}) \leq \theta f(\frac{x_1}{t_1}) + (1 - \theta)f(\frac{x_2}{t_2})$

$$\begin{aligned} & g(\theta x_1 + (1 - \theta)x_2, \theta t_1 + (1 - \theta)t_2) \\ &= (\theta t_1 + (1 - \theta)t_2) f\left(\frac{\theta x_1 + (1 - \theta)x_2}{\theta t_1 + (1 - \theta)t_2}\right) \\ &= \theta t_1 f\left(\frac{\theta x_1 + (1 - \theta)x_2}{\theta t_1 + (1 - \theta)t_2}\right) + (1 - \theta)t_2 f\left(\frac{\theta x_1 + (1 - \theta)x_2}{\theta t_1 + (1 - \theta)t_2}\right) \\ &\leq \theta t_1 [\theta f(\frac{x_1}{t_1}) + (1 - \theta)f(\frac{x_2}{t_2})] + (1 - \theta)t_2 [\theta f(\frac{x_1}{t_1}) + (1 - \theta)f(\frac{x_2}{t_2})] \\ &= \theta t_1 f(\frac{x_1}{t_1}) + (1 - \theta)t_2 f(\frac{x_2}{t_2}) \\ &= \theta g(x_1, t_1) + (1 - \theta)g(x_2, t_2) \end{aligned}$$

因此 g 也是凸函数.

Problem 2: Concave function

先证明不等式 $\left(\sum_{i=1}^n (x_i + y_i)^p\right)^{\frac{1}{p}} \geq \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n y_i^p\right)^{\frac{1}{p}}$

令 $a_i = x_i^p, b_i = y_i^p$, 则 $x_i = a_i^{\frac{1}{p}}, y_i = b_i^{\frac{1}{p}}$, 且 $x_i, y_i \geq 0$

原式两边乘 p 次方可转化为

$$\sum_{i=1}^n (a_i^{\frac{1}{p}} + b_i^{\frac{1}{p}})^p \geq \left(\left(\sum_{i=1}^n a_i \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n b_i \right)^{\frac{1}{p}} \right)^p$$

用 $\frac{1}{p}$ - Norm 范数表示该不等式即为

$$\sum_{i=1}^n \left\| \begin{pmatrix} a_i \\ b_i \end{pmatrix} \right\|_{\frac{1}{p}} \geq \left\| \sum_{i=1}^n \begin{pmatrix} a_i \\ b_i \end{pmatrix} \right\|_{\frac{1}{p}}$$

由范数的三角不等式 $\|x + y\|_{\frac{1}{p}} \leq \|x\|_{\frac{1}{p}} + \|y\|_{\frac{1}{p}}$ 即可知该式成立.

因此我们带入 $\theta x + (1 - \theta)y$ 即可知

$$\begin{aligned} \left(\sum_{i=1}^n (\theta x_i + (1 - \theta)y_i)^p \right)^{\frac{1}{p}} &\geq \left(\sum_{i=1}^n (\theta x_i)^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n ((1 - \theta)y_i)^p \right)^{\frac{1}{p}} = \\ &\theta \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}} + (1 - \theta) \left(\sum_{i=1}^n y_i^p \right)^{\frac{1}{p}} \end{aligned}$$

即 $f(\theta x + (1 - \theta)y) \geq \theta f(x) + (1 - \theta)f(y)$ 成立.

因此 $f(x) = \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}}$ 在 $\text{dom}(f) = \mathbb{R}_{++}$ 时是一个凹函数.

Problem 3: Convexity

(a)

首先对 $x \neq y$ 的情况进行分析.

因为 ψ 是一个严格凸函数, 根据定义有 $\psi(\theta x + (1 - \theta)y) < \theta\psi(x) + (1 - \theta)\psi(y)$

我们考虑过 x, y 两点的函数 $g(t) = \psi(ty + (1 - t)x), t \in [0, 1]$

我们求导可得 $g'(t) = \nabla\psi(ty + (1 - t)x)^T(y - x)$

因为 ψ 是严格凸函数, 因此 g 也是严格凸函数, 我们有 $g(0) > g(1) + g'(1) \cdot (0 - 1)$

即 $\psi(x) > \psi(y) - \nabla\psi(y)^T(y - x) = \psi(y) + \nabla\psi(y)^T(x - y)$

即有 $\Delta_\psi(x, y) = \psi(x) - \psi(y) - \langle \nabla\psi(y), x - y \rangle > 0$

对于 $x = y$ 的情况, 带入即可知

$$\Delta_\psi(x, y) = \psi(x) - \psi(x) - \langle \nabla \psi(y), x - x \rangle = 0$$

综上所述有 $\psi(x, y) \geq 0, \forall x, y \in \Omega$ 且当且仅当 $x = y$ 时取到等号.

(b)

$$\text{要证 } L(y) + \Delta_\psi(y, x_0) \geq L(x^*) + \Delta_\psi(x^*, x_0) + \Delta_\psi(y, x^*)$$

$$\text{即证 } L(y) + \psi(y) - \psi(x_0) - \nabla \psi(x_0)^T(y - x_0) \geq L(x^*) + \psi(x^*) - \psi(x_0) - \nabla \psi(x_0)^T(x^* - x_0) + \psi(y) - \psi(x^*) - \nabla \psi(x^*)^T(y - x^*)$$

$$\text{即证 } L(y) \geq L(x^*) + [\nabla \psi(x_0) - \nabla \psi(x^*)]^T(y - x^*)$$

$$\text{由 } L(y) \text{ 是凸函数可知 } L(y) \geq L(x^*) + \nabla L(x^*)^T(y - x^*)$$

令 $f(x) = L(x) + \Delta_\psi(x, x_0) = L(x) + \psi(x) - \psi(x_0) - \nabla \psi(x_0)^T(x - x_0)$, 因为其是数个凸函数相加, 结果仍然是凸函数

求梯度得 $\nabla f(x) = \nabla L(x) + \nabla \psi(x) - \nabla \psi(x_0)$, 因为在 $x = x^*$ 处取得最小值, 因此有 $\nabla f(x^*) = \nabla L(x^*) + \nabla \psi(x^*) - \nabla \psi(x_0) = 0$

$$\text{因此 } \nabla L(x^*) = \nabla \psi(x_0) - \nabla \psi(x^*), \text{ 带入 } L(y) \geq L(x^*) + \nabla L(x^*)^T(y - x^*)$$

$$\text{可知 } L(y) \geq L(x^*) + [\nabla \psi(x_0) - \nabla \psi(x^*)]^T(y - x^*) \text{ 成立}$$

因此原式成立.

Problem 4: Projection

(a)

因为 $\Pi_X(x)$ 是在凸集 X 上离 x 最近的点, 因此与 $x - \Pi_X(x)$ 垂直的, 过点 $\Pi_X(x)$ 的超平面 S_x 是 X 的一个支撑超平面, 同理 S_y 也是 X 的一个支撑超平面.

过 $\Pi_X(x), \Pi_X(y), x$ 三点作一个二维平面 P , 将 $y - \Pi_X(y)$ 直线投影至 P 上得 $y' - \Pi_X(y)$, 其中的 $x - y'$ 与 $\Pi_X(x) - \Pi_X(y)$ 平行. 并且 P 分别与 S_x, S_y 形成了两条切线 l_x, l_y , P 与 X 的交集形成了一个新的二维凸集 X' .

通过这种方式, 根据点乘的几何意义即可将问题转化为 $\|\Pi_X(x) - \Pi_X(y)\|_2^2 \leq [\Pi_X(x) - \Pi_X(y)]^T(x - y')$

对于 $x, y', \Pi_X(x), \Pi_X(y)$ 在同一条直线上时易知成立.

对于不在同一条直线上的情况, 设 $\Pi_X(x)$ 和 $\Pi_X(y)$ 的中点为 O , 且以 $O - \Pi_X(y)$ 为横坐标轴正方向建立坐标系, 且凸集 X' 位于横坐标轴下方, x, y' 位于横坐标轴上方. 接下来证明位于 O 左侧的过 $\Pi_X(x)$ 的切线 l_x 斜率大于等于零, l_y 斜率小于等于零.

使用反证法, 假设 l_x 斜率小于零, 即 l_x 向右下方倾斜, 那么 O 就会处于 l_x 的下方, 再根据支撑超平面的性质可知, O 点不在凸集 X 上, 这与 $O = \frac{\Pi_X(x) + \Pi_X(y)}{2}$ 位于凸集 X 上矛盾. 因此假设不成立, l_x 斜率大于等于零, 同理 l_y 斜率小于等于零.

然后观察图像, 我们可知 $\Pi_X(x) - \Pi_X(y), x - y'$ 同向, 并且 $\Pi_X(x) - \Pi_X(y), x - y', l_x, l_y$ 形成了一个梯形, 因此我们有 $\|\Pi_X(x) - \Pi_X(y)\|_2 \leq \|x - y\|_2$

可知 $\|\Pi_X(x) - \Pi_X(y)\|_2^2 \leq [\Pi_X(x) - \Pi_X(y)]^T (x - y') = \|\Pi_X(x) - \Pi_X(y)\|_2 \|x - y\|_2$ 成立.

综上所述, $\|\Pi_X(x) - \Pi_X(y)\|_2^2 \leq \langle \Pi_X(x) - \Pi_X(y), x - y \rangle$ 成立.

(b)

由 (a) 有 $\|\Pi_X(x) - \Pi_X(y)\|_2^2 \leq \langle \Pi_X(x) - \Pi_X(y), x - y \rangle$

而由点乘的几何意义我们可知 $\langle \Pi_X(x) - \Pi_X(y), x - y \rangle = \|\Pi_X(x) - \Pi_X(y)\|_2 \cdot \|x - y\|_2 \cdot \cos \theta$

因此我们有 $\|\Pi_X(x) - \Pi_X(y)\|_2 \leq \|x - y\|_2 \cdot \cos \theta \leq \|x - y\|_2$

Problem 5:

(a)

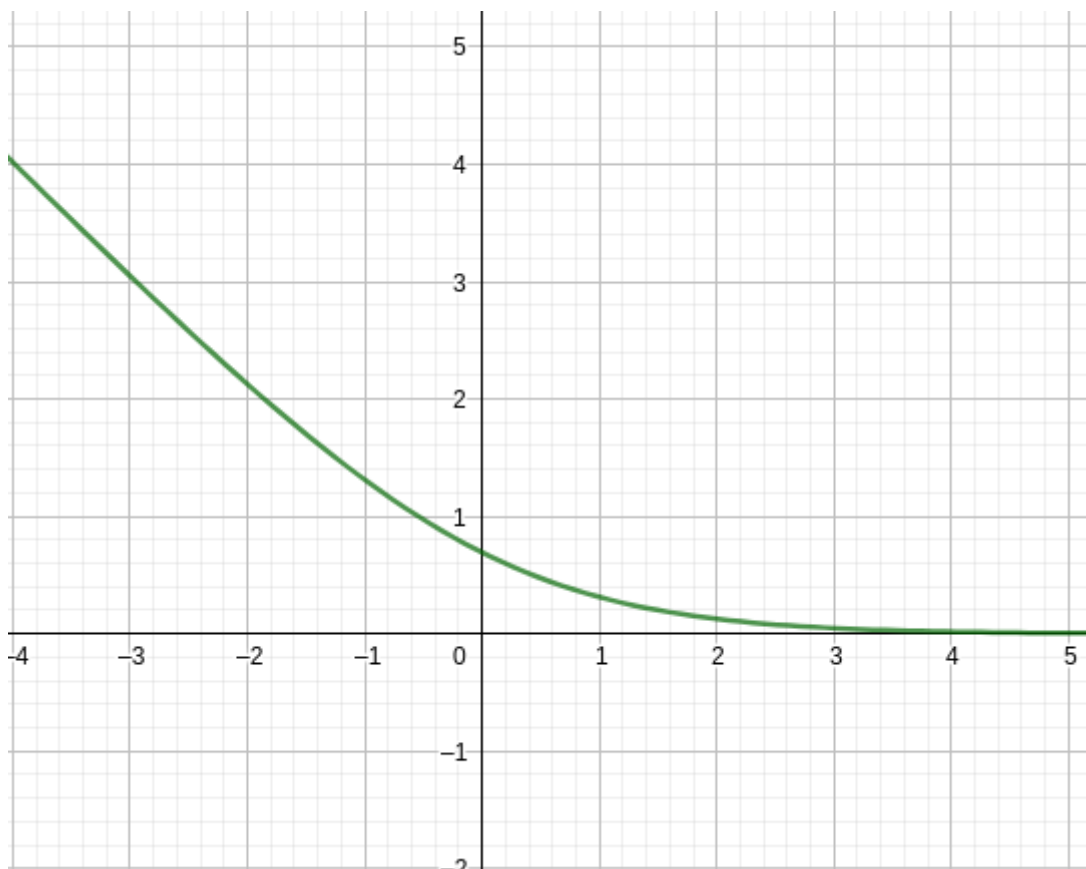
$$f^*(y) = \sup_{x \in \text{dom} f} (yx - \max\{0, 1 - x\})$$

显然, $f^*(y)$ 的定义域为 $[-1, 0]$, 均为在 $x = 1$ 处取得最大值, 即

$$f^*(y) = y - \max\{0, 1 - 1\} = y$$

(b)

$f(x) = \ln(1 + e^{-x})$ 的图像如图所示



因为 $\lim_{x \rightarrow -\infty} \frac{\ln(1 + e^{-x})}{-x} = 1, \lim_{x \rightarrow +\infty} \ln(1 + e^{-x}) = 0$

所以 $y = -x$ 和 $y = 0$ 是 $f(x) = \ln(1 + e^{-x})$ 的两条渐近线.

因此 $f^*(y)$ 的定义域为 $(-1, 0)$, $(yx - \ln(1 + e^{-x}))' = y + \frac{e^{-x}}{1 + e^{-x}}$

即 $y + (y + 1)e^{-x} = 0 \Rightarrow x = -\ln \frac{-y}{y + 1}$ 时有最大值

$$f^*(y) = y \cdot \left(-\ln \frac{-y}{y + 1}\right) - \ln\left(1 + \frac{-y}{y + 1}\right) = (y + 1) \ln(y + 1) - y \ln(-y)$$