Optimization Methods **Fall 2022**

Homework 2

Instructor: Lijun Zhang *Name:* 张运吉, *StudentId:* 211300063

Notice

- The submission email is: optfall2022@163.com.
- 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: Convexity

a) Suppose f and g are both convex, nondecreasing (or nonincreasing), and positive real-valued functions defined on $\mathbb{R}$, prove that fg is convex on $\mathbf{dom}(f) \cap \mathbf{dom}(g)$.

b) Let f be twice differentiable, with $\mathbf{dom}(f)$ convex. Prove that f is convex if and only if

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geqslant 0,$$

for all x, y.

c) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Its perspective transform $g : \mathbb{R}^{n+1} \to \mathbb{R}$ is defined by

$$g(x, t) = tf\left(\frac{x}{t}\right),$$

with domain $\mathbf{dom}(g) = \{(x, t) \in \mathbb{R}^{n+1} : x \in \mathbf{dom}(f), t > 0\}$. Use the definition of convexity to prove that if f is convex, then so is its perspective transform g.

Solution.

(a)

$\because f, g$ 是凸函数，$\therefore dom(f), dom(g)$ 是凸集合, $\therefore dom(f) \cap dom(g)$ 是凸集合。
$\forall x, y \in dom(f \cap g), \forall \theta \in [0, 1]$, $g(\theta x + (1-\theta)y) \leq \theta g(x) + (1-\theta)g(y)$
再用 f 作用于不等式两边，并且由于 f 是非递减的，
得到 $f[g(\theta x + (1-\theta)y)] \leq f[\theta g(x) + (1-\theta)g(y)] \leq \theta fg(x) + (1-\theta)fg(x)$,
最后一个不等式是由于 f 是凸函数。
$\therefore fg$ 是凸函数。

(b)

凸函数的一阶条件：$f(x) \geq f(y) + \nabla f(y)^\top (x-y) ... (1)$
$f(y) \geq f(x) + \nabla f(x)^\top (y-x) ... (2), \forall x, y \in dom(f)$
$(\nabla f(x) - \nabla f(y))^\top (x-y) \geq 0 \iff \nabla f(x)^\top (x-y) - \nabla f(y)^\top (x-y) \geq 0$
$\iff \nabla f(x)^\top (x-y) - (f(x) - f(y)) \geq 0$ 由 (1),
$\iff f(y) - [f(x) + \nabla f(x)^\top (y-x)] \geq 0$
由 (2), 上式成立。

(c)

因为 $dom(f)$ 是凸集合，易知 $\mathbf{dom}(g) = \{(x,t) \in \mathbb{R}^{n+1} : x \in \mathbf{dom}(f), t > 0\}$ 也是凸集合。
$\forall (x,t), (y,t) \in dom(g)$,

$$
\begin{aligned}
&g(\theta(x,t) + (1-\theta)(y,t)) \\
&= g((\theta x, \theta t) + ((1-\theta)y, t(1-\theta))) \\
&= g(\theta x + (1-\theta)y, t) \\
&= tf(\frac{\theta x + (1-\theta y)}{t}) \\
&= tf(\theta\frac{x}{t} + (1-\theta)\frac{y}{t}) \\
&\leq t\theta f(\frac{x}{t}) + t(1-\theta)f(\frac{y}{t}) \\
&= \theta g(x,t) + (1-\theta)g(y,t)
\end{aligned}
$$

$\therefore g$ 是凸函数。

□

Problem 2: Convex Functions

a) Suppose $f_i : \mathbb{R}^n \mapsto \mathbb{R}$ are convex for all $i \in \{1, 2, \cdots, n\}$. Prove that the function $f : \mathbb{R}^n \mapsto \mathbb{R}$, defined as

$$f(x) = \sum_{i=1}^{n} e^{-1/f_i(x)},$$

is convex on $\mathbf{dom}(f) = \{x \in \mathbb{R}^n | f_i(x) < 0, i = 1, 2, \cdots, n\}$.

b) Show that the logarithmic barrier function for the second-order cone, defined as

$$f(x,t) = -\log(t^2 - x^\top x)$$

is convex on $\mathbf{dom}(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} | t > \|x\|_2\}$. (*Hint: consider the function* $-\log(t - (1/t)u^\top u)$)

c) Suppose $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n$. Show that the function

$$f(x) = \frac{\|Ax - b\|_2^2}{1 - x^\top x}$$

is convex on $\mathbf{dom}(f) = \{x \in \mathbb{R}^n | \|x\|_2 \leq 1\}$.

Solution.

(a)

记 $h(z) = -\frac{1}{z}, dom(h) = \{z < 0\}$,$h(z)$ 是凸函数且非递减的，$f_i(x) < 0$ 且是凸函数，由 p1(1) 的结论知 $h(f_i(x)) = -\frac{1}{f_i(x)}$ 是凸函数, 又因为 e^x 是凸函数且非递减的，同理可知 $e^{-\frac{1}{f_i(x)}}$ 是凸函数, 由于非负加权求和保持凸性，所以 $f(x) = \sum_{i=1}^n e^{-1/f_i(x)}$ 是凸函数。

(b)

$f(x,t) = -\log(t^2 - x^\top x) = -\log(t - \frac{1}{t}x^\top x) - \log t$
设 $g(x,t) = t - \frac{1}{t}x^\top x$,$h(x) = 1 - x^\top x$,
则 $g(x,t) = th(\frac{x}{t})$
易知 h 是凹函数，所以 g 是凹函数。(透视运算保持凸性)
$\log g(x,t)$ 是凹函数。(复合保持凸性)
$\therefore f = -\log g(x,t) - \log t$ 是凸函数。

(c)

先证明 $h(x) = \frac{\|Ax-b\|_2^2}{c^\top x+d}$ 是凸函数，
$h(x) = g(y,t) = \frac{y^\top y}{t}$, 令 $(y,t) = (Ax-b, c^\top x + d)$
可以看出 h 是 g 经过仿射变换得到
而 g 是函数 $x^\top x$ 的透视函数，由于 $x^\top x$ 是图的，所以 g 是凸函数
$\therefore h$ 是凸的
最后在 h 中令 $c = -x, d = 1$ 可得 $f(x) = \frac{\|Ax-b\|_2^2}{1-x^\top x}$ 是凸的

□

Problem 3: Concave Function

Suppose $0 < p \leq 1$. Show that the function

$$f(x) = \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}$$

with $\mathbf{dom}(f) = \mathbb{R}_+^n$ is concave.

Solution. Write your answer here.

$\frac{\partial f(X)}{\partial x_i} = (\sum_{i=1}^n x_i^p)^{\frac{1}{p}-1} x_i^{p-1} = (\frac{f(x)}{x_i})^{1-p}$
$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = (1-p)(\frac{f(x)}{x_i})^{-p}(\frac{f(x)}{x_i})' = (1-p)(\frac{f(x)}{x_i})^{-p}\frac{1}{x_i}\frac{\partial f(x)}{\partial x_j} = \frac{1-p}{f(x)}(\frac{f(x)^2}{x_i x_j})^{1-p}$
$\forall y \in \mathbb{R}_+^n, y^\top(\nabla^2 f(x))y = \frac{1-p}{f(x)}((\sum_{i=1}^n \frac{y_i f(x)^{1-p}}{x_i^{1-p}})^2 - \sum_{i=1}^n \frac{y_i^2 f(x)^{2-p}}{x_i^{2-p}})$
由柯西不等式 $|a| \cdot |b| \geq |a|$,
令 $a_i = (\frac{f(x)}{x_i})^{-\frac{p}{2}}, b_i = y_i(\frac{f(x)}{x_i})^{1-\frac{p}{2}}$
得 $(\sum_{i=1}^n \frac{y_i f(x)^{1-p}}{x_i^{1-p}})^2 \leq \sum_{i=1}^n \frac{y_i^2 f(x)^{2-p}}{x_i^{2-p}}$
$\therefore \nabla^2 f(x) \preceq 0$, 即 f 是凹函数。

□

Problem 4: Convexity and Conjugate Function

Let $R : \Omega \mapsto \mathbb{R}$ be a strictly convex and continuously differentiable function defined on a closed convex set Ω. Denote by $\Delta_R(x, y)$ the *Bregman divergence* with respect to the function R, defined as

$$\Delta_R(x, y) = R(x) - R(y) - \langle \nabla R(y), x - y \rangle, \quad \forall x, y \in \Omega. \tag{1}$$

That is, the difference between the value of R at x and the first order Taylor expansion of R around y evaluated at point x.

a) Derive the *Bregman divergence* of $f(x) = \sum_{i=1}^n x_i \log(x_i)$ with $\Omega = \mathbb{R}^n_+$.

b) Let L be a convex and differentiable function defined on Ω and $C \subset \Omega$ be a convex set. Let $x_0 \in \Omega - C$ and define

$$x^* = \arg\min_{x \in C} L(x) + \Delta_R(x, x_0).$$

Prove that for any $y \in C$,

$$L(y) + \Delta_R(y, x_0) \geqslant L(x^*) + \Delta_R(x^*, x_0) + \Delta_R(y, x^*).$$

c) Recall that the definition of conjugate function is

$$f^\star(y) = \sup_{x \in \mathbf{dom}(f)} \left(y^\top x - f(x) \right).$$

Let $R(x) = \frac{1}{2} x^\top Q x$ defined on $\mathbb{R}^n$, where $Q \in \mathbb{S}^n_{++}$. Derive the conjugates $R^\star$ and $R^{\star\star}$. Verify that $(\nabla R^\star)(\nabla R(x)) = x$ and $\Delta_R(x, y) = \Delta_{R^\star}(\nabla R(y), \nabla R(x))$.

Solution.

(a)

$\Delta_f(x, y) = f(x) - f(y) - \langle \nabla f(x), x - y \rangle$
$= \sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n y_i \log y_i - \sum_{i=1}^n (x_i - y_i)(\log y_i + 1)$
$= \sum_{i=1}^n (x_i \log \frac{x_i}{y_i} - (x_i - y_i))$

(b)

要证 $L(y) + \Delta_f(y, x_0) \geqslant L(x^*) + \Delta_f(x^*, x_0) + \Delta_f(y, x^*)$

即证 $L(y) + f(y) - f(x_0) - \nabla f(x_0)^T (y - x_0) \geqslant L(x^*) + f(x^*) - f(x_0) - \nabla f(x_0)^T (x^* - x_0) + f(y) - f(x^*) - \nabla f(x^*)^T (y - x^*)$

即证 $L(y) \geqslant L(x^*) + [\nabla f(x_0) - \nabla f(x^*)]^T (y - x^*)$

由 $L(y)$ 是凸函数可知 $L(y) \geqslant L(x^*) + \nabla L(x^*)^T (y - x^*)$

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

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

因此 $\nabla L(x^*) = \nabla f(x_0) - \nabla f(x^*)$, 带入 $L(y) \geqslant L(x^*) + \nabla L(x^*)^T (y - x^*)$

可知 $L(y) \geqslant L(x^*) + [\nabla f(x_0) - \nabla f(x^*)]^T (y - x^*)$ 成立

因此原式成立.

(c)

$R^{\star}(y) = \sup_{x \in \mathbf{dom}(R)} \left(y^{\top} x - R(x) \right)$
令 $h(x) = y^{\top} x - R(x)$, $\nabla h(x) = y - Qx$
令 $\nabla h(x) = y - Qx = 0$, 得 $x = Q^{-1}y$(也即 $h(x)$ 的最大值点)
$\therefore R^{\star}(y) = y^{\top} Q^{-1} y - \frac{1}{2}(Q^{-1}y)^{\top} Q Q^{-1} y = \frac{1}{2} y^{\top} Q^{-1} y$
$\because \nabla^2 R(x) = Q$ 是正定矩阵
$\therefore R$ 是凸函数
$\because$ 凸函数的共轭函数的共轭函数是其本身
$\therefore R^{\star\star}(x) = R(x) = \frac{1}{2} x^{\top} Q x$
$(\nabla R^{\star})(\nabla R(x)) = Q^{-1}(Qx) = x$
$\Delta_{R^{\star}}(\nabla R(y), \nabla R(x)) = R^{\star}(\nabla R(y)) - R^{\star}(R(x)) - \langle \nabla R^{\star}(\nabla R(x)), \nabla R(y) - \nabla R(x) \rangle$
$= \frac{1}{2} y^{\top} Q y - \frac{1}{2} x^{\top} Q x - x^{\top} Q(y - x) = \frac{1}{2} x^{\top} Q x - \frac{1}{2} y^{\top} Q y + (y^{\top} - x^{\top}) Q y$
$\Delta_R(x, y) = \frac{1}{2} x^{\top} Q x - \frac{1}{2} y^{\top} Q y - Qy(x - y)$
$= \frac{1}{2} x^{\top} Q x - \frac{1}{2} y^{\top} Q y + (y^{\top} - x^{\top}) Q y$
$\therefore \Delta_R(x, y) = \Delta_{R^{\star}}(\nabla R(y), \nabla R(x))$

□

Problem 5: Projection

For any point y, the projection onto a nonempty and closed convex set $\mathcal{X}$ is defined as

$$\Pi_{\mathcal{X}}(y) = \arg\min_{x \in \mathcal{X}} \frac{1}{2} \|x - y\|_2^2.$$

a) Prove that $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2^2 \leqslant \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle$.

b) Prove that $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2 \leqslant \|x - y\|_2$.

c) If we choose $\Pi_{\mathcal{X}}(y) = \arg\min_{x \in \mathcal{X}} \Delta_R(x, y)$, where $\mathcal{X}$ is the n-dimensional simplex $\{x \in \mathbb{R}_+^n \,|\, \sum_{i=1}^n x_i = 1\}$, $\Delta_R(x, y)$ is defined in (**??**) and $R(x) = \sum_{i=1}^n x_i \log(x_i)$. Prove that $\Pi_{\mathcal{X}}(y) = \frac{y}{||y||_1}$ when $y \in \mathbb{R}_{++}^n$. (*Hint: you may use the Jensen' s inequality*)

Solution.

(a)

注意到:
$\langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - \Pi_{\mathcal{X}}(x) \rangle \geq 0$
$\langle \Pi_{\mathcal{X}}(y) - \Pi_{\mathcal{X}}(x), y - \Pi_{\mathcal{X}}(y) \rangle \geq 0$
两式相加可得:
$\langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y - \Pi_{\mathcal{X}}(x) + \Pi_{\mathcal{X}}(y) \rangle \geq 0$
上式可变成:
$\langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle - \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y) \rangle \geq 0$
$\therefore \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2^2 \leqslant \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle$

(b)

$\langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle = \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}\|_2 \|x - y\|_2 \cos\theta$
$\geq \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2^2$
$\therefore \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2 \leqslant \|x - y\|_2 \cos\theta \leq \|x - y\|_2$

(c)

$\nabla R(y) = (\log(y_1) + 1, ..., log(y_n) + 1)^\top$
$\Delta_R(x, y) = \sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n y_i \log y_i - \sum_{i=1}^n (x_i \log x_i - y_i \log y_i + x_i - y_i)$
$= \sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n x_i \log y_i - \sum_{i=1}^n (x_i - y_i)$
$= \sum_{i=1}^n x_i \log \frac{x_i}{y_i} - (1 - \|y\|_1)$
$\therefore \Pi_{\mathcal{X}}(y) = \arg\min\limits_{x\in\mathcal{X}} \Delta_R(x, y) = \arg\min\limits_{x\in\mathcal{X}} \sum_{i=1}^n x_i \log \frac{x_i}{y_i}$
令 $f(k) = \log \frac{1}{k}$, 则 $f''(k) = \frac{1}{k^2} > 0$,$\therefore f(k)$ 是凸函数且非递减的
令 $k_i = \frac{y_i}{x_i}$, 则 $\sum_{i=1}^n x_i \log \frac{x_i}{y_i} = \sum_{i=1}^n x_i f(\frac{1}{k_i})$
有 Jensen's inequality 得:
$\sum_{i=1}^n x_i f(\frac{1}{k_i}) \geq f(\sum_{i=1}^n \frac{x_i}{k_i})$
当且仅当 $\frac{x_i}{y_i} = \frac{x_j}{y_j}(i \neq j)$ 时取等号
我们取 $x_i = \frac{y_i}{\|y\|_1}$, 发现正好满足 $\frac{x_i}{y_i} = \frac{x_j}{y_j}(i \neq j)$ 并且 $\sum_{i=1}^n x_i = 1$
$\therefore x = (\frac{y_1}{\|y\|_1}, ..., \frac{y_n}{\|y\|_1}) = \frac{y}{\|y\|_1}$
即证明了 $\Pi_{\mathcal{X}}(y) = \frac{y}{\|y\|_1}$

$\square$