Optimization Methods **Fall 2020**

Homework 2

Instructor: Lijun Zhang *Name:* Fuyun Wong, *StudentId:* 191300051

Notice

- The submission email is: **njuoptfall2019@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: Convex functions

a) Prove that the function $f : \mathbb{R}^n_{++} \to \mathbb{R}$, defined as

$$f(x) = -\sum_{i=1}^{n} \log(x_i),$$

is strictly convex.

Solution. Let $f_i(x) = -\log x_i$. Then we have $f_i(x)$ is a convex function.

Thus $f(x) = f_1(x) + f_2(x) + \cdots + f_n(x)$. Since nonnegative weighted sums is an operation that preserves the convexity, $f(x)$ is a convex function.

In the other way, we could caculate the Hessian Matrix of the function. it's easy to find that the Hessian Matrix of the function is positive semidefinite,because it is a diagonal matrix and all the elements are nonnegative.

$\square$

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

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0,$$

for all x, y.

Solution. If f is differentialble and convex, then

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad (1), f(x) \geq f(y) + \nabla f(y)^T (x - y) \quad (2)$$

.

Then (1) + (2) we have $(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$.

If ∇f is monotone,then let $g(t) = f(x + t(y - x))$ and $g(t)' = \nabla f(x + t(y - x))^T (y - x)$.

we can find that $g'(t) \geq g'(0)$ for $t \geq 0$ and $t \in \mathrm{dom}$,because $t(g'(t) - g'(0))^T (y - x) \geq 0$.

hence

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

According to the first-order condition for convexity, f is convex.

$\square$

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

$$g(x, t) = tf(x/t),$$

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

Solution. let $0 \leq \theta \leq 1$, then we have

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

□

Problem 2: Concave function
Suppose $p < 1, p \neq 0$. Show that the function

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

with dom $f = \mathbb{R}_{++}$ is concave.

Solution. first derivatives:

$$\frac{\partial f}{\partial x_i} = (\frac{f}{x_i})^{1-p}$$

second derivatives:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{1-p}{f}(\frac{f^2}{x_i x_j})^{1-p}$$

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{1-p}{f}(\frac{f^2}{x_i^2})^{1-p} - \frac{1-p}{x_i}(\frac{f(x)}{x_i})^{1-p}$$

.

then we can denote the matrix as

$$(\frac{(1-p)f}{f^{2p}}(\frac{1}{x_i x_j})^{1-p})_{n\times n} - \frac{(1-p)f}{f^{2p}} \frac{1}{x_i^2} f^p diag(x_i^p)$$

let $A = diag(1/x_1, 1/x_2, \ldots, 1/x_n)$ and $y = (x_1^p, x_2^p, \ldots, x_n^p)$.
then we can denote the matrix as

$$\frac{f(1-p)}{f^{2p}} A^T (yy^T - f^p diag(y))A$$

so $\forall z \in dom$, we have

$$
\begin{aligned}
&z^T \frac{f(1-p)}{f^{2p}} A^T (yy^T - f^p diag(y))Az \\
&= \frac{f(1-p)}{f^{2p}} (g^T yy^T g - g^T f^p diag(y) g) \\
&= \frac{f(1-p)}{f^{2p}} (\sum_{i=1}^{n} (g_i x_i^p)^2 - \sum_{i=1}^{n} x_i^p \sum_{i=1}^{n} x_i^p g_i^2)
\end{aligned}
$$

with Cauchy–Schwarz inequality, we have

$$\sum_{i=1}^{n} (g_i x_i^p)^2 \leq \sum_{i=1}^{n} x_i^p \sum_{i=1}^{n} x_i^p g_i^2$$

since $\frac{f(1-p)}{f^{2p}} \geq 0$,denote the matrix M and then we have

$$\forall z \in \text{dom} \quad \text{and} \quad z \neq 0, z^T M z \leq 0$$

.

then the matrix M which is the hessian matrix of function f, is a positive semidefinite matrix.
therefore, f is concave.

□

Problem 3: Convexity
Let $f : W \mapsto \mathbb{R}$ be a convex function and $\lambda_1, \ldots, \lambda_n$ be n positive numbers with $\sum_{i=1}^n \lambda_i = 1$. Prove that for any $w_1, \ldots, w_n \in W$,

$$f\left(\sum_{i=1}^n \lambda_i w_i\right) \leq \sum_{i=1}^n \lambda_i f(w_i). \tag{1}$$

Solution. According to Jensen Inequality, if the function f is convex, and $\lambda_1, \ldots, \lambda_n \geq 0$ and $\sum_{i=1}^n \lambda_i = 1$, we have

$$f(\sum_{i=1}^n \lambda_i w_i) \leq \sum_{i=1}^n \lambda_i f(w_i)$$

.

□

Problem 4: Projection
For any point y, the projection onto a nonempty and closed convex set X is defined as

$$\Pi_X(y) = \operatorname*{argmin}_{x \in X} \frac{1}{2}\|x - y\|_2^2. \tag{2}$$

a) Prove that $\|\Pi_X(x) - \Pi_X(y)\|_2^2 \leq \langle \Pi_X(x) - \Pi_X(y), x - y\rangle$.

Solution. First of all , I prove a lemma.

lemma 1. *$\forall y \in X$, we have $\langle x - \Pi_X(x), y - \Pi_X(x)\rangle \leq 0$.*

proof: let $0 < \lambda < 1$, then we have

$$\begin{aligned}
&\|x - \Pi_X(x)\|^2 \\
\leq &\|x - \Pi_X(x) - \lambda(y - \Pi_X(x))\|^2 \\
= &\|x - \Pi_X(x)\|^2 - 2\lambda(x - \Pi_X(x))^T(y - \Pi_X(x)) + \lambda\|y - \Pi_X(x)\|^2.
\end{aligned}$$

this implies that $2\langle x - \Pi_X(x), y - \Pi_X(x)\rangle \leq \lambda\|y - \Pi_X(x)\forall\lambda, 0 < \lambda < 1\|$.
when $\lambda \to 0$, we have $\langle x - \Pi_X(x), y - \Pi_X(x)\rangle \leq 0$.
With lemma1,we can get that

$$\langle \Pi_X(y) - \Pi_X(x), x - \Pi_X(x)\rangle \leq 0, \forall x, y \tag{1}$$

and

$$\langle \Pi_X(x) - \Pi_X(y), y - \Pi_X(y)\rangle \leq 0, \forall x, y \tag{2}$$

sum the equation(1) and equation(2), we have

$$\langle \Pi_X(x) - \Pi_X(y), (\Pi_X(x) - \Pi_X(y) - (x - y))\rangle \leq 0$$

then

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

□

b) Prove that $\|\Pi_X(x) - \Pi_X(y)\|_2 \leq \|x - y\|_2$.

Solution. since we have proved that $\|\Pi_X(x)-\Pi_X(y)\|_2^2 \le \langle \Pi_X(x)-\Pi_X(y), x-y\rangle = \|\Pi_X(x)-\Pi_X(y)\|_2\|x-y\|_2\cos\theta \le \|\Pi_X(x)-\Pi_X(y)\|_2\|x-y\|_2$.

if $\|\Pi_X(x)-\Pi_X(y)\|_2 = 0$, it's trivial that $\|x-y\|_2 \ge \|\Pi_X(x)-\Pi_X(y)\|_2$.

else $\|\Pi_X(x)-\Pi_X(y)\|_2 > 0$, we still have $\|x-y\|_2 \ge \|\Pi_X(x)-\Pi_X(y)\|_2$.

□

Problem 5: Convexity

Let $\psi : \Omega \mapsto \mathbb{R}$ be a strictly convex and continuously differentiable function. We define

$$\Delta_\psi(x,y) = \psi(x) - \psi(y) - \langle \nabla\psi(y), x-y\rangle, \quad \forall x,y \in \Omega.$$

a) Prove that $\Delta_\psi(x,y) \ge 0, \forall x,y \in \Omega$ and the equality holds only when $x = y$.

Solution. According to the First-order conditions of strictly and continuously differentiable function, we have

$$\psi(x) > \psi(y) + \nabla f(y)^T(x-y)\forall x,y(x\ne y) \in \Omega$$

. thus it's trivial that

$$\Delta_\psi(x,y) \ge 0, \forall x,y \in \Omega$$

and the equality holds only when $x = y$.

□

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

$$x^* = \underset{x\in C}{\arg\min}\ L(x) + \Delta_\psi(x,x_0).$$

Prove that for any $y \in C$,

$$L(y) + \Delta_\psi(y,x_0) \ge L(x^*) + \Delta_\psi(x^*,x_0) + \Delta_\psi(y,x^*). \tag{3}$$

Solution. let $f(x) = L(x) + \nabla_\psi(x,x_0) = L(x) + \psi(x) - \psi(x_0) - \nabla\psi(x_0)(x-x_0)$.

then $f'(x) = \nabla L(x) + \nabla\psi(x) - \nabla\psi(x_0)$

since x^* minimizes f over C, we have

$$\langle \nabla L(x^*) + \nabla\psi(x^*) - \nabla\psi(x_0), x-x^*\rangle \ge 0\forall x \in C$$

with first order condition, we have

$$L(y) \ge L(x^*) + \langle \nabla L(x^*), y-x^*\rangle$$

then we have

$$\begin{aligned} L(y) &\ge L(x*) + \langle \nabla L(x^*), y-x^*\rangle \\ &\ge \langle \nabla\psi(x_0) - \nabla\psi(x^*), y-x\rangle \end{aligned} \tag{*}$$

simplify the $L(x^*) + \Delta_\psi(x^*,x_0) + \Delta_\psi(y,x^*) - \Delta_\psi(y,x_0)$

$$\begin{aligned} &L(x^*) + \Delta_\psi(x^*,x_0) + \Delta_\psi(y,x^*) - \Delta_\psi(y,x_0) \\ =&L(x^*) + \psi(x^*) - \psi(x_0) + \psi(y) - \psi(x^*) - \psi(y) + \psi(x_0) + \nabla\psi(x_0)^T(x^*-x_0) + \nabla\psi(x^*)^T(y-x^*) - \nabla\psi(x_0)^T(y-x_0) \\ =&(\nabla\psi(x_0) - \nabla\psi(x^*))^T(y-x) = \langle \nabla\psi(x_0) - \nabla\psi(x^*), y-x\rangle \end{aligned} \tag{**}$$

compare (*) and (**), it's obvious that $L(y) + \Delta_\psi(y,x_0) \ge L(x^*) + \Delta_\psi(x^*,x_0) + \Delta_\psi(y,x^*)$

□