Optimization Methods **Fall 2020**

Homework 2

Instructor: Lijun Zhang *Name:* 张逸凯, *StudentId:* 171840708

Notice

- The submission email is: **opt4grad@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.

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.

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

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

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

- **(a) 解:**

  由严格凸的二阶条件, 往证:

$$\nabla^2 f(x) \succ 0$$

  $\because\ \nabla f(x) = \left(-\frac{1}{x_1}, -\frac{1}{x_2}, \cdots, -\frac{1}{x_n}\right)$, 求Hessian矩阵时$i,j$位置的元素即为对上述 $\nabla f(x)$ 第 i 个元素关于 x_j 求导, 我们有:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{1}{x_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_n^2} \end{pmatrix}$$

$$f : \mathbb{R}^n_{++} \to \mathbb{R},\ \therefore\ |\nabla^2 f(x)| = \frac{1}{x_1^2 x_2^2 \cdots x_n^2} > 0$$

  即Hessian矩阵是正定阵, 此时严格凸.

- **(b) 解:**
  - “ $\Rightarrow$ ”

    $\because$ f 是凸函数且二阶可微, 由一阶条件我们有:

$$\begin{cases} f(y) \geqslant f(x) + \nabla f(x)^\top (y-x) \\ f(x) \geqslant f(y) + \nabla f(y)^\top (x-y) \end{cases}$$

    以上两式取反并相加得:

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

  - “ $\Leftarrow$ ”

    $\because (\nabla f(x) - \nabla f(y))^\top (x-y) > 0, \forall x, y \quad (*)$

    不妨令 $h(t) = f(x + t(y-x)),\ t \in [0,1]$, 我们有:

$$\begin{aligned} & h'(t) = \nabla_t f^\top (y-x) \\ & h'(1) = \nabla_t f(y)^\top (y-x),\ h'(0) = \nabla_t f(x)^\top (y-x) \\ & \therefore\ h'(1) - h'(0) > 0 \quad (**) \end{aligned}$$

    $(*)\ \Rightarrow\ x \nearrow \nabla f(x) \nearrow$, 即此时 $h'(t)$ 单调.

    由 $(**)$ 式得 $h'(t) - h'(0) > 0$, 又 h 连续:

$$\begin{aligned} & \therefore f(y) = h(1) = h(0) + \int_0^1 h'(t) dt \geqslant h(0) + h'(0) \\ & \therefore f(y) \geqslant f(x) + \nabla f(x)^\top (y-x) \text{ 得证.} \end{aligned}$$

  - **(c) 解:**

    $\because\ \mathrm{dom}(g) = \left\{(x,t) \in \mathbb{R}^{n+1} : \frac{x}{t} \in \mathrm{dom}(f), t > 0\right\}$

    往证: 对于$\forall \left(x_1, t_1\right), \left(x_2, t_2\right) \in \mathrm{dom}\, g,\ \forall \theta \in [0,1]$

    有:

$$g\left(\theta x_1 + (1-\theta) x_2,\ \theta_1 t_1 + (1-\theta) t_2\right) \leqslant \theta g\left(x_1,\ t_1\right) + (1-\theta) g\left(x_2,\ t_2\right)$$

    以下为证明, 由Jensen不等式:

$$\begin{aligned} & g\left(\theta x_1 + (1-\theta) x_2,\ \theta t_1 + (1-\theta) t_2\right) \\ = & \left(\theta t_1 + (1-\theta) t_2\right) f\left(\frac{\theta x_1 + (1-\theta) x_2}{\theta t_1 + (1-\theta) t_2}\right) \\ = & \ \left(\theta t_1 + (1-\theta) t_2\right) \cdot f\left(\frac{\theta_1 t_1}{\theta t_1 + (1-\theta) t_2} \cdot f\left(\frac{x_1}{t_1}\right) + \frac{(1-\theta) t_2}{\theta t_1 + (1-\theta) t_2} \cdot f\left(\frac{x_2}{t_2}\right)\right) \\ \leqslant & \ \left(\theta t_1 + (1-\theta) t_2\right) \cdot \left\{\underbrace{\frac{\theta t_1}{\theta t_1 + (1-\theta) t_2}}_{A} f\left(\frac{x_1}{t_1}\right) + \underbrace{\frac{(1-\theta) t_2}{\theta t_1 + (1-\theta) t_2}}_{B} f\left(\frac{x_2}{t_2}\right)\right\} \\ = & \theta t_1 f\left(\frac{x_1}{t_1}\right) + (1-\theta) t_2 f\left(\frac{x_2}{t_2}\right) \\ = & \theta g\left(x_1,\ t_1\right) + (1-\theta) g\left(x_2,\ t_2\right) \end{aligned}$$

    如上 A, B 为Jensen不等式($f\left(A x_1 + B x_2\right) \leq A f\left(x_1\right) + B f\left(x_2\right)$) 中的参数, 且 $A + B = 1$.

    综上所述, 透视函数g的凸性得证.

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.

- **解:** 由二阶条件证明 f 是凹的, 一阶梯度为:

$$\nabla f(x) = \left(x_1^{p-1}\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-1}, x_2^{p-1}\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-1}, \cdots, x_n^{p-1}\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-1}\right)^\top$$

  二阶梯度为:

$$\begin{aligned}\nabla^2 f(x) = (p-1)\cdot\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}-2} \cdot \Bigg\{ &\operatorname{diag}\left(\left(\sum_{i=1}^{n} x_i^p\right)\cdot x_1^{p-2}, \cdots, \left(\sum_{i=1}^{n} x_i^p\right)\cdot x_n^{p-2}\right) \\ &- \begin{pmatrix} x_1^{p-1} \\ x_2^{p-1} \\ \vdots \\ x_n^{p-1} \end{pmatrix} \left(x_1^{p-1}, x_2^{p-1}, \cdots, x_n^{p-1}\right) \Bigg\}\end{aligned}$$

  由柯西不等式, 由矩阵负定的定义, $t^\top \nabla^2 t < 0,\ \forall\, t \neq 0, p < 1, p \neq 0$.

  所以原函数的凹性得证.

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 $w_1, \ldots, w_n \in W$,

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

- **解:**

  数学归纳法:

  - $i = 1, 2$ 时, 由凸函数定义知原不等式成立.
  - 假设 $i = k,\ k = 1, 2, \cdot$ 时原不等式成立.
  - 当 $i = k+1$ 时, 由 Jensen 不等式知:

$$\begin{aligned} f\left(\sum_{i=1}^{k+1} \lambda_i x_i\right) &= f\left(\lambda_{k+1} x_{k+1} + \sum_{i=1}^{k} \lambda_i x_i\right) \\ &= f\left(\lambda_{k+1} x_{k+1} + (1-\lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i\right) \end{aligned}$$

    注意到:

$$\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} = 1$$

我们有:

$$f\left(\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i\right) \leqslant \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} f(x_i)$$

结合第 k 步的假设:

$$\begin{aligned}
\therefore\ f\left(\sum_{i=1}^{k+1} \lambda_i x_i\right) &= f\left(\lambda_{k+1} x_{k+1} + (1-\lambda_{k+1}) \sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i\right) \\
&\leqslant \lambda_{k+1} f\left(x_{k+1}\right) + (1-\lambda_{k+1}) f\left(\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x_i\right) \\
&\leqslant \lambda_{k+1} f\left(x_{k+1}\right) + (1-\lambda_{k+1}) \frac{\sum_i^k \lambda_i}{1-\lambda_{k+1}} f\left(x_i\right)
\end{aligned}$$

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

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

- **(a) 解:**

  由题意, 做如下化简:

  $$\Pi_X(x) = \arg\min_{t \in X} \tfrac{1}{2} \|t-x\|_2^2 \Rightarrow \Pi_X(x) = x$$

  不妨令 $u = \Pi_X(y)$.

  往证:

  $$\begin{aligned}
  &\|x-u\|_2^2 \leqslant \langle x-u,\ x-y \rangle \\
  \Rightarrow &(x-u)^\top (x-u-x+y) \leqslant 0 \\
  \Rightarrow &(x-u)^\top (y-u) \leqslant 0
  \end{aligned}$$

  以下为证明, 由 Π_X 定义知 $y-u$ 与在u处的支撑超平面正交, 反之不然, 垂直支撑超平面的范数更小, 这与u的定义相违.

  又 $x,\ u \in X$, X 为闭凸集. 存在支持超平面:

  $$\forall\, x, (x-u)^\top (y-u) = 0$$

  且对于 $\forall\, x'$, 满足 $(x'-u)^\top (y-u) > 0$, 则 $x' \notin X$.

  综上所述, $(x-u)^\top (y-u) \leqslant 0$ 得证.

  即:

  $$\|x-u\|_2^2 \leqslant \langle x-u,\ x-y \rangle$$

- **(b) 解:**

  由题意, 两边平方并化简, 往证:

$$
\begin{aligned}
&\|\pi_X(x)-\pi_X(y)\|_2^2-\|x-y\|_2^2 \\
&=(2x-u-y)^\top(y-u) \\
&=(x-u)^\top(y-u)+(x-y)^\top(y-u)\leqslant 0 \\
&\because\ (x-u)^\top(y-u)\leqslant 0 \text{ 在 (a) 中证明.} \\
&\Rightarrow (x-y)^\top(y-u)\leqslant 0
\end{aligned}
$$

  反证法, 假设 $(x-y)^\top(y-u)>0$, 则有:

$$
(y-x)^\top(y-u)+(x-u)^\top(y-u)=(y-u)^\top(y-u)<0, \text{矛盾}.
$$

  综上所述, 原不等式得证.

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

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

- **(a) 解:**

  因为 ψ 严格凸, 由一阶条件:

$$
\forall\ x,\ y\in\operatorname{dom}\psi, x\neq y,\ \psi(x)>\psi(y)+\nabla\psi(y)^\top(x-y)\Rightarrow\Delta_\psi(x,y)>0
$$

  当且仅当 $x=y$ 时:

$$
\psi(x)=\psi(y)+\nabla\psi(y)^\top(x-y)\Rightarrow\Delta_\psi(x,y)=0
$$

- **(b) 解:**

  展开 $\Delta_\psi\left(y,x_0\right)$, $\Delta_\psi\left(x^*,x_0\right)$, $\Delta_\psi\left(y,x_*\right)$ 三项, 我们有如下简化, 往证:

$$
L(y)-\nabla\psi\left(x_0\right)^\top\left(y-x_0\right)-\Big(L\left(x^*\right)-\nabla\psi\left(x_0\right)^\top\left(x^*-x_0\right)\Big)\geqslant-\nabla\psi\left(x^*\right)^\top\left(y-x^*\right)
$$

  以下为证明:

  不妨令 $T_1(y)=L(y)-\nabla\psi\left(x_0\right)^\top\left(y-x_0\right)$, 我们有:

  由凸函数定义知: $T_1(y)$ 是凸函数.

由一阶条件:

$$L(y)-\nabla\psi\left(x_0\right)^{\top}\left(y-x_0\right)-\Big(L\left(x^*\right)-\nabla\psi\left(x_0\right)^{\top}\left(x^*-x_0\right)\Big)\geqslant\Big(\nabla L\left(x^*\right)^{\top}-\nabla\psi\left(x_0\right)^{\top}\Big)\left(y-x^*\right)\quad(*)$$

不妨令 $T_2(x)=L(x)+\Delta_\psi\left(x,x_0\right)$, 这是关于 x 的凸函数, 且 x^* 是其最小值, 对于任意 $y\in C$, 凸性保证了保号性, 我们有:

$$\Big(\nabla L\left(x^*\right)^{\top}+\nabla\psi\left(x^*\right)^{\top}-\nabla\psi\left(x_0\right)^{\top}\Big)\left(y-x^*\right)\geqslant 0$$

将上式代入 $(*)$ 式, 即可证得原不等式.