Convex Optimization. Exercise 2. Chao Ni

1. Convex functions

(a) $f_1(x) = \ln(e^{x_1} + \dots + e^{x_n})$. where $x = (x_1, \dots x_n)$

Let $z_k = e^{x_k}$.

$e^{x_1} + \dots + e^{x_n} = 1^T z$

$$\nabla f_1(x) = \left(\frac{z_1}{1^T z}, \frac{z_2}{1^T z}, \dots \frac{z_n}{1^T z} \right)^T.$$

$$\nabla^2 f_1(x) = \frac{1}{1^T z} \cdot \begin{pmatrix} z_1 & & \\ & z_2 & \\ & & \ddots \\ & & & z_n \end{pmatrix} - \frac{1}{(1^T z)^2} z z^T.$$

$\forall x \in \mathbb{R}^n$. $x^T \nabla^2 f_1(x) x = \frac{1}{1^T z} \cdot \sum_{i=1}^n z_i x_i^2 - \frac{1}{(1^T z)^2} \left(\sum_{i=1}^n z_i x_i \right)^2 = \frac{\sum_{i=1}^n z_i x_i^2 \cdot \sum_{i=1}^n z_i - \left(\sum_{i=1}^n z_i x_i \right)^2}{(1^T z)^2} \geq 0.$

Since $\left(\sum_{i=1}^n z_i x_i \right)^2 \leq \sum_{i=1}^n z_i x_i^2 \cdot \sum_{i=1}^n z_i$.

Hence $\nabla^2 f_1(x) \succeq 0$. $f_1(x)$ is convex

(b). $f_2(x) = \|x\|^p$ with $p \geq 1$

Consider the function $h(t) = t^p$.

Hence $f_2(x) = \|x\|^p = h(\|x\|)$.

Because $p > 1$. $h'(t) = pt^{p-1}$. $h''(t) = p(p-1)t^{p-2} > 0$.

$h(t)$ is convex, and also $h(t)$ is non-decreasing.

$\|x\|$ is convex.

Hence By composition rule. $f_2(x)$ is convex

(c). $f_3(x) = \frac{1}{f(x)}$. where $f: \mathbb{R}^n \to \mathbb{R}$ is concave and $f(x) > 0$. $\forall x$

$$f_3(\theta x + (1-\theta) y) = \frac{1}{f(\theta x + (1-\theta) y)} \leq \frac{1}{\theta f(x) + (1-\theta) f(y)} \qquad \theta \in [0,1]$$

To show $\frac{1}{\theta f(x) + (1-\theta) f(y)} \leq \theta f_3(x) + (1-\theta) f_3(y) = \frac{\theta}{f(x)} + \frac{1-\theta}{f(y)} = \frac{\theta f(y) + (1-\theta) f(x)}{f(x) f(y)}$

$\Leftarrow$ $[\theta f(x) + (1-\theta) f(y)][\theta f(y) + (1-\theta) f(x)] \geq f(x) f(y)$

$\Leftrightarrow$ $\theta(1-\theta) \cdot [f(x) - f(y)]^2 \geq 0$

which is obvious. Hence. $f_3(x)$ is convex.

(d) $f_4(x) = \|Ax-b\|_2^2 + \|x\|_1$

Let $h(x) = \|x\|_2^2 = x^T x$.

$\nabla h(x) = x$, $\nabla^2 h(x) = I \succ 0$.

Hence $h(x)$ is convex.

As $Ax-b$ is affine wrt x., $\Rightarrow$ $\|Ax-b\|_2^2$ is convex

and $\|x\|_1$ is a norm. it's also convex.

Hence $f_4(x) = \|Ax-b\|_2^2 + \|x\|_1$ is convex.

(e) $f_5(x) = \sum_{i=1}^{n} (\max(0, 1+x_i)) + \|x\|_2^2$, where $x = (x_1, \dots, x_n)$

Consider function $h_i : x \longmapsto \max(0, 1+x_i)$

$$
\begin{aligned}
h_i(\theta x + (1-\theta) y) &= \max(0, 1 + (\theta x_i + (1-\theta) y_i)) \\
&= \max(0, \theta(1+x_i) + (1-\theta)(1+y_i)) \\
&\le \max(0, \theta(1+x_i)) + \max(0, (1-\theta)(1+y_i)) \\
&= \theta \max(0, 1+x_i) + (1-\theta) \cdot \max(0, 1+y_i) \\
&= \theta \cdot h_i(x) + (1-\theta) h_i(y)
\end{aligned}
$$

Hence h_i is convex.

By (d) we know $\|x\|_2^2$ is convex.

Hence $f_5(x) = \sum_{i=1}^{n} h_i(x) + \|x\|_2^2$ is also convex

(f) $f_6(x) = \sup_y (x^T y - g(y))$ with $g: \mathbb{R}^n \to \mathbb{R}$ not necessarily convex.

$x^T y - g(y)$ is convex wrt x.

Hence $f_6(x)$ is pointwise supremum of convex functions from family $\{x^T y - g(y)\}_y$.

Hence $f_6(x)$ is convex.

(g). $\phi_C(x) = \frac{1}{2}(\|x\|_2^2 - d_C^2(x))$, where $C \subset \mathbb{R}^n$ is non-empty compact set.

$d_C(x) = \min_{y \in C} \|x-y\|_2$ is the distance function to C

$$d_C^2(x) := \min_{y \in C} \|x-y\|_2^2 = \min_{y \in C} \left[\|x\|_2^2 + \|y\|_2^2 - 2x^Ty\right]$$
$$= \|x\|_2^2 + \min_{y \in C}\left[\|y\|_2^2 - 2x^Ty\right]$$

Hence, $\phi_C(x) = \frac{1}{2}\left(-\min_{y \in C} \|y\|_2^2 - 2x^Ty\right)$
$$= \frac{1}{2}\max_{y \in C}\left(2x^Ty - \|y\|_2^2\right)$$

By (f) we know, $\phi_C(x)$ is convex.

(h) $g(x) = \inf_{u \in \mathbb{R}^n} \{h_1(u) + h_2(x-u)\}$, where $h_1, h_2: \mathbb{R}^n \to (-\infty, \infty]$ are proper convex functions that have a common affine minorant. (s, r).

$h_i(x) \geq s^Tx - r$ for $i=1,2$. $\forall x$.

$\forall x, \forall u \in \mathbb{R}^n$.

$$h_1(u) + h_2(x-u) \geq s^Tu - r + s^T(x-u) - r = s^Tx - 2r > -\infty.$$

Hence. $g(x) > -\infty$, $\forall x \in \mathbb{R}^n$

Let $f(x,u) = h_1(u) + h_2(x-u)$.

As h_1, h_2 are convex functions, $f(x,u)$ is the convex wrt x, u.

To show this

$$f(\theta x + (1-\theta)y, u) = h_1(u) + h_2(\theta x + (1-\theta)y - u)$$
$$= \theta h_1(u) + (1-\theta)h_1(u) + h_2\left[\theta(x-u) + (1-\theta)(y-u)\right]$$
$$\leq \theta h_1(u) + (1-\theta)h_1(u) + \theta h_2(x-u) + (1-\theta)h_2(y-u)$$
$$= \theta\left[h_1(u) + h_2(x-u)\right] + (1-\theta)\left[h_1(u) + h_2(y-u)\right]$$
$$= \theta f(x,u) + (1-\theta) f(y,u).$$

$$f(x, \theta u + (1-\theta)v) = h_1(\theta u + (1-\theta)v) + h_2(x - (\theta u + (1-\theta)v))$$
$$\leq \theta h_1(u) + (1-\theta)h_1(v) + h_2(\theta(x-u) + (1-\theta)(x-v))$$
$$\leq \theta h_1(u) + (1-\theta)h_1(v) + \theta h_2(x-u) + (1-\theta)h_2(x-v)$$
$$= \theta f(x,u) + (1-\theta) f(x,v)$$

Here $f(x,u)$ is convex wrt u, v. $\mathbb{R}^n$ is nonempty convex set

By partial minimization rule, $g(x)$ is convex.

2. Strictly Convex Functions and Unique Minimizers,

(a). For the sake of Contradiction, $\exists\, x \neq x^*$ st $f(x) = f(x^*)$

$$\Rightarrow \quad f(\tfrac{1}{2}x + \tfrac{1}{2}x^*) < \tfrac{1}{2}f(x) + \tfrac{1}{2}f(x^*) = f(x).$$

Hence $y = \frac{1}{2}x + \frac{1}{2}x^* \in C$ is the ~~...~~ better point with smaller value.

contradicts to the fact that x^* is global minimizer.

Hence. $\nexists\, x \neq x^*$ st $f(x) = f(x^*)$. x^* is unique if it's global minimizer.

(b). $f(x) = \frac{1}{x}$, with $\mathrm{dom}(f) = \mathbb{R}_{++}$

3.

(a). $g(x) = \sum_{i=1}^{n} g_i(x_i)$, $x = (x_1, \dots x_n)$.

Show that $s \in \partial g(x) \iff s_i \in \partial g_i(x_i)$, where $s = (s_1, \dots, s_n)$

1° $\Rightarrow$.

Assume. $s \in \partial g(x)$, then $g(y) \geq g(x) + s^T(y - x)$ $\forall y \in \mathbb{R}^n$.

$\forall j \in [n]$. Let $y_i = x_i$, $i \in [n] \setminus \{j\}$.

Then we have $g_j(y_j) \geq g_j(x_j) + s_j^T(y_j - x_j)$

$$\Rightarrow \quad s_j \in \partial g_j(x_j). \quad \forall j \in [n].$$

2° $\Leftarrow$ Assume. $s_i \in \partial g_i(x_i)$. $\forall i \in [n]$

then. $g_i(y_i) \geq g_i(x_i) + s_i^T(y_i - x_i)$ $\forall i \in [n]$.

Sum these inequalities up.

$$g(y) \geq g(x) + s^T(y - x).$$

$$\Rightarrow \quad s \in \partial g(x) \quad \square$$

(b). The normal cone to C: $N_C(x) = \{s \in \mathbb{R}^n \mid s^T(y-x) \le 0,\ \forall y \in C\}$.

indicator function ~~set~~ of a nonempty set C. $I_C(x)$

$$I_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}.$$

The subdifferental of the $I_C(x)$ satisfies that.

$$I_C(y) \ge I_C(x) + s^T(y-x). \quad \forall y \in \mathbb{R}^n \qquad (*)$$

Case 1: $x \in C$. $I_C(x) = 0$.

If $y \in C$. $0 \ge 0 + s^T(y-x)$.

$\partial I_C(x) = \{s : s^T(y-x) \le 0\} = N_C(x)$.

If $y \notin C$. $s \in \mathbb{R}^n$.

Hence. for all y, $(*)$ must hold $\Rightarrow \partial I_C(x) = \{s : s^T(y-x) \le 0, \forall y \in C\} = N_C(x)$

Case 2: $x \notin C$. $I_C(x) = \infty$.

If ~~let~~ $y \in C$. then $(*)$ can't hold.

$\Rightarrow$ ~~$s = \phi$~~, $\nexists s$ such that $(*)$ holds.

Hence $\partial I_C(x) = \phi$.

~~Also, because $x \notin C$. Hence~~

Overall, $\partial I_C(x) = N_C(x)$. $\square$.

4. Conjugate functions.

(a). $f(x) = \|x\|_2$.

(a) $f^*(x) = \sup_y \left(y^T x - \|y\|_2 \right)$

Let $y = x$., $y^T x - \|y\|_2 = 0$.

Hence $f^*(x) = \sup_y (y^T x - \|y\|_2) \ge y^T x - \|y\|_2 \Big|_{y=x} = 0$.

(b) For all y. $y^T x - \|y\|_2 \le |y^T x| - \|y\|_2 \le \|y\|_2 \|x\|_2 - \|y\|_2 \le \|y\|_2 - \|y\|_2 = 0$

Hence $f^*(x) \le 0$ with $\|x\|_2 \le 1$

(c) For all x with $\|x\|_2 > 1$.

Let $y = tx$.

$$f^*(x) \ge t\|x\|_2^2 - t\|x\|_2 = t\|x\|_2(\|x\|_2 - 1)$$

Hence. $f^*(x) \to +\infty$ when $t \to \infty$. $f^*(x) = \infty$.

(d). $f^*(x) = \begin{cases} 0 & \|x\|_2 \le 1 \\ +\infty & \|x\|_2 > 1 \end{cases}$

(b). Use the conjugate to compute the subdifferential of $f(x)=\|x\|_2$

$$f(y) \geq f(x) + s^T(y-x) \quad . \forall y.$$

$$\Rightarrow \quad s^T y - f(y) \leq s^T x - f(x) \quad \forall y$$

$$\Rightarrow \quad \sup_y \left(s^T y - f(y)\right) \leq s^T x - f(x)$$

$$\Rightarrow \quad f^*(s) \leq s^T x - f(x).$$

Because $f(x)$ is differentiable at $x \neq 0$, s is the unique at $x \neq 0$.

Hence. take $\partial f(x) = \arg\max_s \; s^T x - f^*(s).$

$$= \arg\max_{\|s\|_2 \leq 1} (s^T x). \qquad x \neq 0.$$

For $x=0$. $f^*(s) \leq s^T x - f(x) \Big|_{x=0} = 0.$

Based on (a). $\|s\|_2 \leq 1. \Rightarrow \partial f(0) = \{s : \|s\|_2 \leq 1\}$

For $x \neq 0$.

$$\partial f(x) = \arg\max_{\|s\|_2 \leq 1} s^T x$$

$$\leq \arg\max_{\|s\|_2 \leq 1} \|s\|_2 \cdot \|x\|_2 \qquad \text{--- Cauchy Schwarz inequality.}$$

the equality holds when $\langle s, x \rangle = \|s\|_2 \cdot \|x\|_2 \cdot \cos 0°.$ ($s \parallel x$.).

Because $\|s\|_2 \leq 1$, Hence ~~$\partial f(x)$ takes~~ $s^T x$ reaches its maximum when $\|s\|_2 = 1$ and $s \parallel x$.

$$\Rightarrow \quad s = \frac{x}{\|x\|_2}.$$

Overall

$$\partial f(0) = \{s : \|s\|_2 \leq 1\}$$

$x \neq 0$, $\partial f(x) = \{ \frac{x}{\|x\|_2}$.

$\square$