Optim : Homework 1

Name : Zinelo Semane

Problem 1 :

1/ $A = \{x \in \mathbb{R}^n \mid \alpha \leq a^T x \leq \beta\}$ is convex.

For $x, y \in A$ and $t \in [0,1]$, we have :

$$a^T(t x + (1-t)y) = t.(a^T x) + (1-t).a^T y$$

As $x, y \in A$, we have : $\begin{cases} \alpha \leq a^T x \leq \beta & (1) \\ \alpha \leq a^T y \leq \beta & (2) \end{cases}$

$$t(1) + (1-t)(2) \underset{\substack{\text{(because } t\geq 0 \\ \text{and } 1-t \geq 0)}}{\Longrightarrow} t\alpha + (1-t)\alpha \leq t.a^T x + (1-t)a^T y \leq t\beta + (1-t)\beta$$

$$\Rightarrow \alpha \leq a^T(tx + (1-t)y) \leq \beta$$

$$\Rightarrow tx + (1-t)y \in A.$$

2/ $A = \{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i\}$ is convex

For $x, y \in A$ and $t \in [0,1]$ we have :

$$tx + (1-t)y = \begin{bmatrix} tx_1 + (1-t)y_1 \\ tx_2 + (1-t)y_2 \\ \vdots \\ tx_n + (1-t)y_n \end{bmatrix}$$

We know that for i in $[\![1, n]\!]$: $\begin{cases} \alpha_i \leq x_i \leq \beta_i \\ \alpha_i \leq y_i \leq \beta_i \end{cases}$

Then $t\alpha_i + (1-t)\alpha_i \leq t x_i + (1-t)y_i \leq t\beta_i + (1-t)\beta_i$ (because $t \geq 0$ and $1-t \geq 0$)

Hence $\alpha_i \leq tx_i + (1-t)y_i \leq \beta_i$, $\forall i \in [\![1, n]\!]$

Conclusion : $tx + (1-t)y \in A$.

3/ $A = \{x \mid \|x - x_0\|_2 \leq \|x-y\|_2, \forall y \in S\}$, $S \in \mathbb{R}^n$

$$= \bigcap_{y \in S} \{x \mid \|x - x_0\|_2 \leq \|x-y\|_2\}$$

$$= \bigcap_{y \in S} \{x \mid (x-x_0)^T(x-x_0) \leq (x-y)^T(x-y)\}$$ (because $\|x-x_0\|_2 \leq \|x-y\|_2 \iff \|x-x_0\|_2^2 \leq \|x-y\|_2^2$)

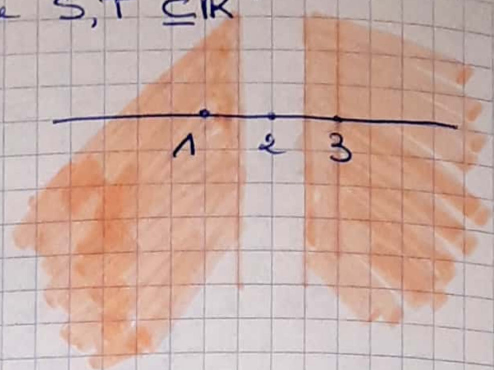
$$= \bigcap_{y \in S} \{x \mid \cancel{x^T x} - 2x_0^T x + x_0^T x_0 \leq \cancel{x^T x} - 2y^T x + y^T y\}$$

$$= \bigcap_{y \in S} \{x \mid 2(y - x_0)^T x \leq y^T y - x_0^T x_0\}$$

For a fixed $y \in S$, $y - x_0$ is a fixed vecter and $y^T y - x_0^T x_0 = \|y\|_2^2 - \|x_0\|_2^2$ is a fixed scalar, then $\{x \mid 2(y-x_0)^T x \leq \|y\|_2^2 - \|x_0\|_2^2\}$ is a half space wich is convex.

As A is the intersection of convex sets, A is convex.

4/ $A = \{x \mid dist(x, S) \leq dist(x, T)\}$, where $S, T \subseteq \mathbb{R}^n$



For example, we take:

$S = \{1, 3\}$, $T = \{2\}$

$dist(1, S) = 0$ (because $1 \in S$)

$dist(1, T) = 1$

then $1 \in A$

$dist(3, S) = 0$ (because $3 \in S$)

$dist(3, T) = 1$

then $3 \in A$

if we take $t = 1/2$, we have $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 3 = 2$

But $dist(2, S) = 1$, $dist(2, T) = 0$ and $1 > 0$

Then $2 \notin A$

Conclusion: A is not convex

5/ $B = \{x \in \mathbb{R}^n, Ax \leq b\}$ for some $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ (Polyhedron) Convex

We can verify that B is convex by using the definition, or by writing B as intersection of half spaces.

let a_i be the i^{th} row for A, and b_i the i^{th} coefficient for the vector b, then:

$$B = \bigcap_{i \in [\![1, m]\!]} \{x \in \mathbb{R}^n, \ a_i x \leq b_i\}$$

For a fixed $i \in [\![1, m]\!]$, the set $\{x \in \mathbb{R}^n, \ a_i x \leq b_i\}$ define a half space, then it's convex. By using the property of intersection of convex sets, we conclude that B is convex.

6/ $A = \bigcap_i S_i$, where $S_i \in \mathbb{R}^n$ collection of convex sets.

For $x, y \in A$, $t \in [0, 1]$, we have:

$\forall i$, $x, y \in S_i$, then $\forall i$, $tx + (1-t)y \in S_i$ (because S_i convex)

Hence $tx + (1-t)y \in \bigcap_i S_i \Rightarrow tx + (1-t)y \in A$

Conclusion A is convex.

Problem 2 :

1/. $f(x) = x^4$

- We have $\nabla f(x) = 4x^3$, and $\nabla^2 f(x) = 12x^2 \geq 0$, and $\mathrm{dom} f = \mathbb{R}$ is convex then f is convex

- For $g(x) = x^2$, $\nabla^2 g(x) = 2 > 0$ then g is strictly convex

For $x, y \in \mathbb{R}$, $t \in]0,1[$, $x \neq y$;

$g(tx + (1-t)y) < t x^2 + (1-t)y^2$

$\Rightarrow g \circ g(tx + (1-t)y) < g(tx^2 + (1-t)y^2)$ (because g is strictly increasing on $\mathbb{R}^*_+$)

$\Rightarrow f(tx + (1-t)y) < t g(x^2) + (1-t) g(y^2)$

$\Rightarrow f(tx + (1-t)y) < tx^4 + (1-t)y^4$

$\Rightarrow$ f is strictly convex

- For $m > 0$, we consider $g(x) = f(x) - \frac{m}{2}\|x\|^2 = f(x) - \frac{m}{2}x^2$

$\nabla^2 g(x) = \nabla^2 f(x) - m = 12x^2 - m$

For $x = 0$, $\nabla^2 g(x) = -m < 0$

then g it's not convex

Conclusion : f isn't strongly convex.

2/. $f(x) = \sum_i x_i \log(x_i)$, $x > 0$

* $\nabla_i f(x) = x_i \cdot \frac{1}{x_i} + 1 \cdot \log x_i = 1 + \log x_i$

$\nabla^2_{i,j} f(x) = \frac{1}{x_i} 1\{i = j\}$

Then $\nabla^2 f(x) = \mathrm{diag}\left(\frac{1}{x_i}\right)_{i=1,\dots,m} > 0$

Conclusion : f is strictly convex and then it's also convex.

* For $m > 0$, $g(x) = \frac{m}{2}\|x\|_2^2$, $\nabla_i g(x) = m x_i$, $\nabla^2_{i,j} g(x) = m 1\{i=j\}$

then $\nabla^2 g(x) = \mathrm{diag}(m, \dots, m)$

$\nabla^2(f(x) - g(x)) = \mathrm{diag}\left(\frac{1}{x_1} - m, \dots, \frac{1}{x_n} - m\right)$

For $i \in [\![1, n]\!]$, and $x'_i < \frac{1}{m}$, $\frac{1}{x'_i} - m < 0$

$\Rightarrow \nabla^2(f - g)(x')$ not semi definite positive.

Conclusion : f not strongly convex.

3/ $f(x) = \log(1+e^x)$

* $\nabla f(x) = \frac{e^x}{1+e^x}$, $\nabla^2 f(x) = \frac{e^x(1+e^x) - e^{2x}}{(1+e^x)^2} = \frac{e^x}{(1+e^x)^2} > 0$

f is strictly convex and then also convex

* For $m > 0$, $\nabla^2 f(x) - m = \frac{e^x}{(1+e^x)^2} - m$ not positive for a $m>0$

Then f isn't strongly convex.

4/ $f(x,y) = |x| + |y| + 2x - 2$

We consider $g(x,y) = |x| + |y|$, $h(x,y) = A\begin{pmatrix} x \\ y \end{pmatrix} + b$ where $A = [2, 0]$, $b = -2$

g is norm 1 then it's convex

h is affine function then it's convex.

* $f(x,y) = g(x,y) + h(x,y) = |x| + |y| + 2x - 2$ is convex as nonnegative linear combination of g and h.

* For $(x_1, y_1), (x_2, y_2) > 0$ and different, we have:

for $t \in]0,1[$: $g(t(x_1,y_1) + (1-t)(x_2,y_2)) = g(tx_1 + (1-t)x_2, ty_1 + (1-t)y_2)$

$= |tx_1 + (1-t)x_2| + |ty_1 + (1-t)y_2|$

$= tx_1 + (1-t)x_2 + ty_1 + (1-t)y_2$

$= t\,g(x_1, y_1) + (1-t)(x_2, y_2)$

Then g is not strictly convex, and also h isn't strictly convex

Conclusion: f isn't strictly convex, and then not strongly convex.

5/ $f(x,y) = x^2 - xy + 2y^2 + 3$

* $\nabla f = \begin{pmatrix} 2x - y \\ 4y - x \end{pmatrix}$, $\nabla^2 f = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$

$|\nabla^2 f| = 7$, $tr(\nabla^2 f) = 6 \Rightarrow \lambda_1, \lambda_2 > 0$

eigen values of $\nabla^2 f$

Conclusion: f is strictly convex and also convex.

* For $m>0$, $g(x,y) = \frac{m}{2} \|(x,y)\|^2 = \frac{m}{2}(x^2 + y^2)$

$\nabla^2 g = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}$

Then $\nabla^2(f-g) = \begin{bmatrix} 2-m & -1 \\ -1 & 4-m \end{bmatrix}$

$|\nabla^2 (f-g)| = (2-m)(4-m) - 1 = 7 - 6m + m^2$, $tr(\nabla^2(f-g)) = 6 - 2m$

For $m = 1$, $|\nabla^2(f-g)| = 2$, $tr(\nabla^2(f-g)) = 4 \Rightarrow \lambda_1', \lambda_2' > 0$

Conclusion f is strongly convex

eigen values of $\nabla^2(f-g)$

6/ $f(x,y) = 2xy + 3x^2 + y^2$

$\nabla f = \begin{pmatrix} 2y + 6x \\ 2x + 2y \end{pmatrix}$, $\nabla^2 f = \begin{bmatrix} 6 & 2 \\ 2 & 2 \end{bmatrix}$

For $m > 0$, $g(x,y) = \frac{m}{2}\|(x,y)\|^2 = \frac{m}{2}(x^2+y^2)$

$\nabla^2(f-g) = \begin{bmatrix} 6-m & 2 \\ 2 & 2-m \end{bmatrix}$

$|\nabla^2(f-g)| = (6-m)(2-m) - 4 = 8 - 8m + m^2$, $tr(\nabla^2(f-g)) = 8 - 2m$

For $m = 1$, $|\nabla^2(f-g)| = 1$, $tr(\nabla^2(f-g)) = 6 \Rightarrow \underset{\text{eigen values of } \nabla^2(f-g)}{\lambda_1, \lambda_2} > 0$

Conclusion: f is strongly convex, then it's also convex and strictly convex.

7/ $f(x,y) = e^{x^2 - xy + y^2 + 2} + \frac{1}{\sqrt{x} + 2y}$, $x, y > 0$

* $x^2 - xy + y^2 + 2 = (x,y)\,Q\begin{pmatrix} x \\ y \end{pmatrix} + 2$ where $Q = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \succcurlyeq 0$

then it's convex (quadratic function)

As $x \mapsto e^x$ is convex and non decreasing, $x \mapsto e^{x^2 - xy + y^2 + 2}$ is convex

* for $(x,y) \xrightarrow{h} \frac{1}{\sqrt{x} + 2y}$

$\nabla h(x,y) = \left[\frac{-1}{2\sqrt{x}(\sqrt{x}+2y)^2} \quad \frac{-2}{(\sqrt{x}+2y)^2} \right]^T$

$\nabla^2 h(x,y) = \begin{bmatrix} \frac{1}{4x(\sqrt{x}+2y)^2}\left[\frac{1}{\sqrt{x}} + \frac{2}{\sqrt{x}+2y}\right] & \frac{2}{\sqrt{x}(\sqrt{x}+2y)^3} \\ \frac{2}{\sqrt{x}(\sqrt{x}+2y)^3} & \frac{8}{(\sqrt{x}+2y)^3} \end{bmatrix}$

$|\nabla^2 h(x,y)| = \frac{2}{x(\sqrt{x}+2y)^5}\left(\frac{1}{\sqrt{x}} + \frac{2}{\sqrt{x}+2y}\right) - \frac{4}{x(\sqrt{x}+2y)^6} = \frac{2}{x(\sqrt{x}+2y)^5}\left(\frac{1}{\sqrt{x}} + \cancel{\frac{2}{\sqrt{x}+2y}} - \cancel{\frac{2}{\sqrt{x}+2y}}\right)$

$= \frac{2}{x\sqrt{x}(\sqrt{x}+2y)^5} > 0$ where $x, y > 0$

$tr(\nabla^2 h(x,y)) = \frac{1}{(\sqrt{x}+2y)^2}\left[\frac{1}{4x\sqrt{x}} + \frac{1}{2x(\sqrt{x}+2y)} + \frac{8}{\sqrt{x}+2y}\right] > 0$

conclusion: eigen values of $\nabla^2 h$ are strictly positive, then h is strictly convex

Hence f as nonnegative linear combination of convex functions, it's strictly convex

8/ $f(x) = \dfrac{\|Ax-b\|_2^2}{1-x^Tx}$ Where $\{x \mid \|x\|_2 < 1\}$ (least squares loss)

We consider $\cdot\ g_1(x) = \|Ax-b\|_2^2$: convex as quadratic function

$\cdot\ g_2(x) = 1 - x^Tx$: Concave (because norm is convex)

The function $h : (x,y) \longmapsto \dfrac{g_1(x)}{y}$ is convex (as quadratic over the linear function) in (x,y).

$\nabla h^T = \left[g_1'(x) \quad -\dfrac{g_1(x)}{y^2}\right]$, Then h is nonincreasing in y

Using composition rule we conclude that $f(x) = h(x, g_2(x))$ is convex.

Problem 3:

(a) $1 \Rightarrow 2]$

∇f is L lipschitz $\Leftrightarrow \forall x,y \in \mathbb{R}^n,\ \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$

$\Rightarrow \forall x,y \in \mathbb{R}^n, x \neq y,\ \dfrac{\|\nabla f(x) - \nabla f(y)\|}{\|x-y\|} \leq L I_n$

$\Rightarrow \lim\limits_{x \to y} \dfrac{\|\nabla f(x) - \nabla f(y)\|}{\|x-y\|} \leq L I_n$

$\Rightarrow \nabla^2 f(y) \leq L I_n$

$2 \Leftrightarrow 3]$

We consider $g(x) = \dfrac{L}{2} x^Tx - f(x)$

$$A = g(y) - g(x) - \nabla g(x)^T(y-x) = \frac{L}{2}\left[y^Ty - x^Tx\right] + f(x) - f(y) - Lx^T(y-x) + \nabla f^T(x)(y-x)$$

$$= \frac{L}{2}\left[y^Ty + x^Tx - 2x^Ty\right] - \left(f(y) - f(x) - \nabla f^T(x)(y-x)\right)$$

$$= \frac{L}{2}(x-y)^T(x-y) - \left(f(y) - f(x) - \nabla f^T(x)(y-x)\right)$$

$2 \Leftrightarrow f(y) \leq f(x) + \nabla f(x)^T(y-x) + \dfrac{L}{2}\|x-y\|_2^2$

$\Leftrightarrow A \geq 0$

$\Leftrightarrow g$ convex

$\Leftrightarrow \nabla^2 g(x) \geq 0$

$\Leftrightarrow LI - \nabla^2 f(x) \geq 0$

$\Leftrightarrow \nabla^2 f(x) \leq LI$

$3 \Rightarrow 1]$ From 3, we have: $\begin{cases} f(y) \leq f(x) + \nabla f(x)^T(y-x) + \dfrac{L}{2}\|x-y\|_2^2 \\ f(x) \leq f(y) + \nabla f(y)^T(x-y) + \dfrac{L}{2}\|x-y\|_2^2 \end{cases}$

by summing two inequalities, we get:

$$\left(\nabla f(x)^T - \nabla f(y)^T\right)(y-x) + L\|x-y\|_2^2 \geq 0$$

$$\Rightarrow (\nabla f(x) - \nabla f(y))^T(x-y) \leq L\|x-y\|_2^2$$

$$\Rightarrow \|\nabla f(x) - \nabla f(y)\| \cdot \|x-y\| \leq L\|x-y\|_2^2 \quad \text{(Cauchy Schwarz)}$$

$$\Rightarrow \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$$

We have $1 \Rightarrow 2 \Leftrightarrow 3 \Rightarrow 1$, Then the 3 propositions are equivalents.

b/ For $k=0$, $\|x(0) - x_*\| \leq \beta^0 . \|x(0) - x_*\|$ True

For $k \in \mathbb{N}$, we suppose that this is true.

$$\|x(k+1) - x_*\| = \|x(k) - t\nabla f(x(k)) - x_*\|$$
$$= \|g(x(k)) - (x_* - t\nabla f(x_*))\| \quad \left(\nabla f(x_*) = 0 \text{ because } x^* \text{ denotes its global minimum}\right)$$
$$= \|g(x(k)) - g(x_*)\|$$
$$\leq \beta \|x(k) - x_*\|$$
$$\leq \beta . \beta^k \|x(0) - x_*\|$$
$$\leq \beta^{k+1} \|x(0) - x_*\|$$

By induction, we include that:

$$\|x(k) - x_*\|_2 \leq \beta^k \|x(0) - x_*\|_2$$

Problem 4:

1/ $f(x,y) = x \log x + y \log y$, $x, y > 1$

$$\nabla^2 f(x,y) = \begin{bmatrix} \frac{1}{x} & 0 \\ 0 & \frac{1}{y} \end{bmatrix} \geq 0 \quad \text{Not strongly convex}$$

f convex and differentiable. Then:

a. We can use: gradient, subgradient and proximal gradiant

b. Convergence rates are respectively: $O(\frac{1}{\varepsilon})$, $O(\frac{1}{\varepsilon^2})$ and $O(\frac{1}{\varepsilon})$

c. Best algorithm: gradient descent:

$(x,y)^+ = (x,y) - t\nabla f(x,y)$, t should be less Than $\frac{1}{L}$ where $L = 1$ because f it's 1-Smooth (using 2nd property from Problem 3.a)

2/ $f(x,y) = |x| + |y| + 2x - 2y + 3$

f is convex and non differentiable (L_1-regularizer not differentiable)

a. We can use: subgradient and proximal gradiant

b. Convergence rates are respectively: $O(\frac{1}{\varepsilon^2})$, $O(\frac{1}{\varepsilon})$

c. As f is composed of smooth and non-smooth parts, proximal gradient is the best for it.

$$f(x,y) = \underbrace{2x - 2y + 3}_{g(x,y)} + \underbrace{|x| + |y|}_{h(x,y)}$$

- We find the step by minimizing smooth part ($\nabla^2 g(x,y) = 0$ for $L = 1$, g is L-smooth)
- We update: $(x,y)^* = prox_h((x,y) - t\nabla g(x,y))$

(t should be smaller than 1, for example)

3/ $f(x,y) = x^2 - xy + 2y^2$

$\nabla^2 f(x,y) = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix} \succeq 0$ strongly convex for $m = 1$.

a. We can use: gradient descent, subgradient and proximal gradient

b. Convergence rates are respectively: $O(\log(\frac{1}{\varepsilon}))$, $O(\frac{1}{\varepsilon^2})$, $O(\frac{1}{\varepsilon})$

c. Best choice is gradient descent.

$(x,y)^+ = (x,y) - t\nabla f(x,y)$

$\nabla^2 f(x,y) - \lambda I = \begin{bmatrix} 2-\lambda & -1 \\ -1 & 4-\lambda \end{bmatrix} = 0 = \lambda^2 - 6\lambda + 7$

$\Delta = 36 - 4.7 = 8$

$\lambda_1 = \frac{6 - 2\sqrt{2}}{2} = 3 - \sqrt{2}$, $\lambda_2 = 3 + \sqrt{2}$

$\max(\lambda_1, \lambda_2) < 5$

Then we can take $t \leq \frac{1}{5}$