(1)

Technique d'optimisation

I- Formulation des problèmes d'optimisation en Machine learning :

Def :

L'optimisation est une branche des mathématique cherchant à modéliser, à analyser et résoudre numériquement ou analytiquement

▷ Def :

une solution x^* est minimum globale de f sur $D_f = S$ si :

$$f(x^*) \leq f(x) \quad \forall x \in S$$

▷ Def

une Solution x^* est minimum locale de f sur S si :

$$f(x^*) \leq f(x) \quad \forall x \in S \cap B_\varepsilon(x^*)$$

tq

$$B_\varepsilon(x^*) = \{x \in \mathbb{R}^n : \|x - x^*\| \leq \varepsilon\}$$

- Si $\nabla f(x^*) = 0$, alors x^* est un point critique

→ exemple de détermination des points critique :

Soit $f(x,y) = xy(x+y-1)$

(1) trouver ∇f

$$\nabla f(x,y) = \begin{pmatrix} y(x+y-1) + xy \\ x(x+y-1) + xy \end{pmatrix}$$

⚠ essaye toujours de factoriser

$$\Rightarrow \nabla f(x,y) = \begin{pmatrix} y(2x+y-1) \\ x(2y+x-1) \end{pmatrix}$$

(2) $\nabla f = 0$

$$\Leftrightarrow \begin{cases} y = 0 \text{ ou } (2x+y-1) = 0 \\ x = 0 \text{ ou } 2y+x-1 = 0 \end{cases}$$

1er cas :

$y = 0$ et $x = 0 \Rightarrow (0,0)$

ou

$y = 0$ et $0 + x - 1 = 0 \Rightarrow (1,0)$

2ème cas :

$x = 0$ et $2x+y-1 = 0 \Rightarrow (0,1)$

$\Leftrightarrow y = 1$

ou

$y = 1 - 2x$ donc $2(1-2x) + x - 1 = 0$

$\Rightarrow 2 - 4x + x - 1 = 0$

$\Rightarrow -3x + 1 = 0$

$\Rightarrow x = \frac{1}{3}$

$\Rightarrow y = 1 - 2\left(\frac{1}{3}\right) = \frac{1}{3}$

$\Rightarrow \left(\frac{1}{3}, \frac{1}{3}\right)$

Donc les p.c sont : $(0,0)$; $(0,1)$; $(1,0)$; $\left(\frac{1}{3}, \frac{1}{3}\right)$

Théorème (conditions nécessaire) :

→ Si x^* est un minimum locale de fonction f sur $\mathbb{R}^n$ alors $\nabla f(x^*) = 0$

→ Rappel Def : (Voir Analyse 2)

Soit A de $n \times n$:

Semi défini ≥ 0 $y^T A y \geq 0 \ \forall y \in \mathbb{R}^n$

Défini > 0 $y^T A y > 0 \ \forall y \in \mathbb{R}^n \setminus \{0\}$

Semi-définie ≤ 0 $y^T A y \leq 0 \ \forall y \in \mathbb{R}^n$

Défini < 0 $y^T A y < 0 \ \forall y \in \mathbb{R}^n$

→ théorème : (condition nécessaire de 2^{nd} ordre)

Si x^* est un min locale de f, alors $\nabla f(x^*) = 0$ et $y^T \nabla^2 f(x^*) y \geq 0 \ \forall y \in \mathbb{R}^2$ ($\Leftrightarrow \nabla^2 f(x^*)$ est semi défini ≥ 0)

→ théorème : (condition suffisante du 2^{nd} ordre)

Soit $x^* \in \mathbb{R}^n$. Si $\nabla f(x^*) = 0$ et si $y^T \nabla^2 f(x^*) y > 0$ $\forall y \in \mathbb{R}^n - \{0\}$, alors x^* est un minimum locale de la fonction f sur $\mathbb{R}^n$

→ Définition d'un point Selle :

Le point x est un point Selle si $\nabla f(x) = 0$ et la matrice hessienne $\nabla^2 f(x)$ est indéfini

exemple : $f(x,y) = x^2 - xy + y^2$

$$\nabla f = \begin{pmatrix} 2x - y \\ -x + 2y \end{pmatrix} \Rightarrow \cancel{\nabla f(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}}$$

$$\cancel{y^T \nabla f(0,0) = (0 \ 0)\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0}$$

$$\Rightarrow y^T \nabla f(0,0) \, y = 0 \cdot \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

$$\nabla^2 f = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

$$(x^* \ y^*) \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} = (2x^* - y^* \quad -x^* + 2y^*)$$

$$\begin{pmatrix} x^* \\ y^* \end{pmatrix} = (2x - y)x + y(2y - x)$$

$$= 2x^2 - xy - xy + 2y^2$$

$$= 2x^2 + 2y^2 - 2xy$$

$$= 2(x^2 + y^2 - xy)$$

$$= 2\left(x^2 + \cancel{\ldots} + y^2 + \frac{y^2}{4} - \frac{y^2}{4} - 2x\frac{y}{2}\right)$$

$$= 2\left(x^2 - 2x\frac{y}{2} + \frac{y^2}{4} - \frac{y^2}{4} + y^2\right)$$

$$= 2\left(\left(x + \frac{y}{2}\right)^2 + \frac{3}{4}y^2\right) > 0$$

Donc $(0,0)$ est un minimum local

$\circ\, f(x,y) = x^3 + 2xy^2 - y^4 + x^2 + 3xy + y^2 + 10$ ②

$$\nabla f = \begin{pmatrix} 3x^2 + 2y^2 + 2x + 3y \\ 4xy - 4y^3 + 3x + 2y \end{pmatrix}$$

$$\nabla^2 f = \begin{pmatrix} 6x+2 & 4y+3 \\ 4y+3 & 4x - 12y^2 + 2 \end{pmatrix}$$

$$\Rightarrow \nabla^2 f(0,0) = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} u \\ v \end{pmatrix}^T \nabla^2 f(0,0) = (u\ v) \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$$

$$= (2u+3v \quad 3u+2v)$$

$$\Rightarrow \begin{pmatrix} u \\ v \end{pmatrix} = (2u+3v)u + (3u+2v)v$$

$$= 2u^2 + 3uv + 3uv + 2v^2$$

$$= 2u^2 + 6uv + 2v^2$$

$$= 2(u^2 + 3uv + v^2)$$

$$= \cancel{2((u+v)^2 + uv)}$$

$$= 2\left(\left(u + \frac{3}{2}v\right)^2 - \frac{9}{4}v^2 + v^2\right)$$

$$= 2\left(\left(u + \frac{3}{2}v\right)^2 - \frac{5}{4}v^2\right)$$

Donc $\nabla^2 f$ est indéfini (suffit de prendre des points qui donne des signes différents comme $(0,1)$ et $(-1,1)$ dans notre cas)

Donc $(0,0)$ est un point selle

Pour le maximum, même théorème que ceux de minimum sauf que

$$y^T \nabla^2 f(x^*) y \le 0 \quad \forall y \in \mathbb{R}^n$$

Cas Particulier ($n = 2$)

Soit $x_0 = (a, b)$ un point critique d'une fonction :

on pose $\Delta = rt - s^2$

avec $r = \dfrac{\partial f(x_0)}{\partial x^2}$, $t = \dfrac{\partial f(x_0)}{\partial y^2}$

$$s = \frac{\partial f(x_0)}{\partial x \partial y} = \frac{\partial f(x_0)}{\partial y \partial x}$$

Si $\Delta > 0$: ~~rest~~

f admet un min locale si $r > 0$

f " " max " si $r < 0$

Si $\Delta < 0$:

f admet un point selle

Si $\Delta = 0$; on ne peut rien dire

Exemple : cherchez les valeurs minimales/maximales et les points selles de la fonction

$$f(x,y) = 4x^2 - xy + y^2 - x^3$$

Réponse :

- $\nabla^2 f = \begin{pmatrix} 8-6x & -1 \\ -1 & 2 \end{pmatrix}$

$$\Delta = (8-6x)(2) - 1$$

$$= 15 - 12x$$

① cherchons les p.c :

$$\nabla f(x,y) = 0 \Leftrightarrow \begin{cases} 8x - y - 3x^2 = 0 \\ 2y = x \end{cases}$$

$$\Rightarrow y = 0 \text{ ou } y = \frac{15}{12} = \frac{5}{4}$$

Si $y = 0 \Rightarrow x = 0 \Rightarrow (0,0)$

Si $y = \frac{5}{4} \Rightarrow x = \frac{5}{2} \Rightarrow \left(\frac{5}{2}, \frac{5}{4}\right)$

② la nature des points trouvées :

- Pour $(0,0)$

$\Delta = 15 > 0$: $r = 8 > 0$

Donc $(0,0)$ est un mini locale

- Pour $\left(\frac{5}{2}, \frac{5}{4}\right)$:

$\Delta = -15 < 0$: donc $\left(\frac{5}{2}, \frac{5}{4}\right)$ est un point Selle

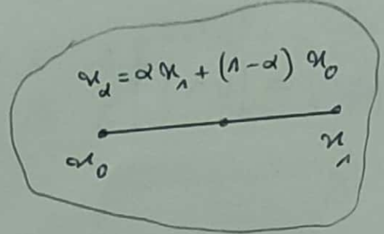
Ensemble convexe.

- Un ensemble $C \subset \mathbb{R}^d$ est dit convexe si :

$$\forall x_0, x_1 \in C, \forall \alpha \in [0,1] :$$

$$x_\alpha = \alpha x_1 + (1-\alpha) x_0 \in C$$

- Interpretation : toute droite construit par $x_0, x_1 \in C$ reste dans C



- ~~Ensemble~~ fonction convexe

- $f : \mathbb{R}^d \to \mathbb{R}$ est convexe si elle vérifie $\forall x_0, x_1 \in \mathbb{R}^d$, $\forall \alpha \in [0,1]$, $f(\alpha x_1 + (1-\alpha)(x_0)) \leq \alpha f(x_1) + (1-\alpha) f(x_0)$

- Si $(-f)$ est convexe

$\Rightarrow f$ concave

- fct strictement convexe

- $f : \mathbb{R}^d \to \mathbb{R}$ est strictement convexe si elle vérifie

$\forall x_0 \neq x_1 \in \mathbb{R}^d$, $\forall \alpha \in]0,1[$,

$$f(\alpha x_1 + (1-\alpha)(x_0)) < \alpha f(x_1) + (1-\alpha) f(x_0)$$

interpretation :
sur l'axe "x" dans l'intervalle $[x_0, x_1]$, si on varie x_α entre x_0 et x_1, le Cf reste toujours $f(x_0)$ $f(x_1)$

- exemple de fct convexe :

③

- fct cst : $x \to C$
- fct affine : $x \to \langle a, x \rangle + C$

- combinaison convexe / moyenne pondéré

soit $x_1, \dots, x_m \in \mathbb{R}^d$

tout point qui s'écrit $\sum_{i=1}^{n} \alpha_i x_i$ pour $\alpha_i \geq 0$ et $\sum_{i=1}^{n} \alpha_i = 1$ est appelé moyenne pondéré

- $f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$

théorème :

Si $f : \mathbb{R}^n \to \mathbb{R}$ est deux fois différentiable, on a :

f convexe $\Leftrightarrow$ $\nabla^2 f$ est semi-défini positive

. exemple :

$\to f(x,y) = x^4 + x^2y^2 + y^4 - 3x - 8y$

$$\nabla f(x,y) = \begin{pmatrix} 4x^3 + 2xy^2 - 3 \\ 2yx^2 + 4y^3 - 8 \end{pmatrix}$$

$$\nabla^2 f(x,y) = \begin{pmatrix} 12x^2 + 2y^2 & 4xy \\ 4xy & 2x^2 + 12y^2 \end{pmatrix}$$

on pose $u = \begin{pmatrix} x \\ y \end{pmatrix}$

$(x \ y) \nabla^2 f(x,y)$

$$= \left((12x^2 + 2y^2)x + 4xy^2 \quad 4x^2y + y(2x^2 + 12y^2) \right.$$

$$= \left((12x^2 + 2y^2)x + 4xy^2 \quad 4x^2y + y(2x^2 + 12y^2) \right)$$

$$(x\,y)\nabla^2 f\begin{pmatrix} x \\ y \end{pmatrix} = x^2(12x^2 + 2y^2) + 4x^2y^2 + 4x^2y^2 + y^2(2x^2 + 12y^2)$$

$$= 12x^4 + 2x^2y^2 + 4x^2y^2 + 4x^2y^2 + 2y^2x^2 + 12y^4$$

$$= 12x^4 + 12y^2x^2 + 12y^4 \geq 0$$

$\boxed{f \text{ est convexe}}$

- Unicité et optimisation

Soit le problème $\inf_{x \in K} (f(x))$ avec f et K convexes (éventuellement de dimension fini), alors :

① tout minimum locale est un minimum globale

② si f est stric convexe, alors il y a au plus un minimum

→ démonstration d'unicité :

I) tout minimum locale est un minimum globale.

① Supposons par absurde que :

Supposons que x^* est un minimum locale

et $\exists y$ tq $f(y) < f(x^*)$

c.à.d on a supposé que y est un minimum globale

② alors on peut écrire :

$$y_\alpha = \alpha y + (1-\alpha) x^* \text{ avec } \alpha \in]0,1[$$

remarque que si α est proche de 0, on s'approche de x^*,

(Si $\alpha = 0 \Rightarrow y_0 = x^*$)

~~Donc cela donne que~~ :

Donc on a : $f(y_\alpha) \geq f(x^*)$

③ d'après la convexité, on a :

$$f(x^*) \leq f(y_\alpha) \leq \alpha f(y) + (1-\alpha) f(y) \leq f(y)$$

au voisinage de $\alpha = 1$, on peut majorer par $f(y)$

on obtient $f(x^*) \leq f(y)$

or d'après la proposition : $f(y) < f(x^*)$

$$\left.\begin{array}{l} f(x^*) \leq f(y) \\ f(y) < f(x^*) \end{array}\right\} \quad f(y) < f(x^*) \leq f(y)$$

Donc absurde

Def :

① Domaine d'une fct : $dom(f) = \{ x \in \mathbb{R}^n , f(x) < +\infty \}$

on suppose dans le reste du cours que $dom(f) \neq \emptyset$

② def : Soit $f : E \to \overline{\mathbb{R}}$. L'épigraphe de f est le sous-ensemble de $E \times \overline{\mathbb{R}}$ définie par :

$$epi(f) = \{ (x, \alpha) \in E \times \overline{\mathbb{R}} \ / \ f(x) \leq \alpha \}$$

l'épi stricte :

$$epi_s(f) = \{ (x, \alpha) \in E \times \overline{\mathbb{R}} \ / \ f(x) \leq \alpha \}$$

③ def : Soit $f : E \to \overline{\mathbb{R}}$.

- Si $\liminf_{y \to x} (f(y)) \geq f(x)$ en $x \in E$ alors f est semi-continue inférieurement au point x { une fonction est (SCI) si elle est (SCI) en tout point de E

Remarque : f est S.C.I en $x_0 \in E$ si $\liminf_{x \to x_0} (f(x)) = f(x_0)$

convexité et différentiabilité

① def : Sous gradient et sous-différentiel
. Soit f convexe. Un vecteur $\eta \in \mathbb{R}^n$ est appelé sous-gradient de f au point $x_0 \in dom(f)$ si :

$$\forall x \in dom(f) , f(x) \geq f(x_0) + \langle \eta , x - x_0 \rangle$$

l'ensemble de tous sous-gradients en x_0 est appelé sous-différentielle de f. il est notée $\partial f(x_0)$

Lemme :
Le sous-diff $\partial f(x_0) = \{ \eta \in \mathbb{R}^n / \forall x \in \mathbb{R}^n ; f(x) \geq f(x_0) + \langle \eta , x - x_0 \rangle \}$ est un ensemble convexe fermée

∘ démonstration de lemme :

Soit $g_1, g_2 \in \partial f(x_0)$, Soit $\alpha \in [0,1]$

on a $\forall x_0 \in \mathbb{R}^n$, $f(x) \geq f(x_0) + \langle g_1, x - x_0 \rangle$ ①

$\forall x_0 \in \mathbb{R}^n$, $f(x) \geq f(x_0) + \langle g_2, x - x_0 \rangle$ ②

on multiplie : ① par α

② par $(1-\alpha)$

① $\alpha f(x) \geq \alpha f(x_0) + \alpha \langle g_1, x - x_0 \rangle$

② $(1-\alpha) f(x) \geq (1-\alpha) f(x_0) + (1-\alpha) \langle g_2, x - x_0 \rangle$

$\overset{①+②}{\Rightarrow} f(x) \geq f(x_0) + \alpha \langle g_1, x - x_0 \rangle + (1-\alpha) \langle g_2, x - x_0 \rangle$

$\Rightarrow f(x) \geq f(x_0) + \langle \alpha g_1 + (1-\alpha) g_2, x - x_0 \rangle$

donc $\alpha g_1 + (1-\alpha) g_2 \in \partial f(x_0)$

Def : derivée directionnel

Soit f définie au $V(x)$ et Soit $v \in \mathbb{R}^n$

> Si $f'(x; v) = \lim_{t \to 0^+} \frac{f(x+tv) - f(x)}{t}$ existe dans $\mathbb{R}$

alors f admet une derivée directionnel vers v

Remarque : Si f est différentiable en x alors $f'(x; v)$ existe pour Toute direction v et on a : $f'(x; v) = \langle \nabla f(x), v \rangle$

Def : Soit f définie au $V(x)$. On dit que f est Gâteau-diffb en x si elle admet des dérivée dans toute les directions et s'il existe $y \in \mathbb{R}^n$ tq que $\forall v$: $f'(x;v) = \langle y, v \rangle$

on pose alors $y = f'(x)$

◦ remarque : Si f est diff en x alors f est Gâteau-diff en x et $f'(x) = \nabla f(x)$

Lemme : Soit f convexe et $x \in dom(f)$

$\forall$ direction v, $f'(x,v)$ existe et $\exists y \in \partial f(x)$ tq

$$f'(x;v) = \langle y, v \rangle$$

▷ Proposition :

Soit U un ouvert convexe de $\mathbb{R}^n$ et $f : U \longrightarrow \mathbb{R}$ est Gâteaux-diff sur U.

$$\underbrace{\begin{aligned} f \text{ convexe sur } U &\Leftrightarrow f(y) \geq f(x) + \langle f'(x), y - x \rangle \; \forall x, y \in U \\ &\Leftrightarrow \langle f'(y) - f'(x), y - x \rangle \geq 0 \quad \forall x, y \in U \end{aligned}}_{\text{Faire démonstration}}$$

▷ Proposition :

Soit f convexe et $x \in dom(f)$. Alors f est diff en x si $\partial f(x)$ est un singleton. On a alors :

$$\partial f(x) = \{ \nabla f(x) \}$$

transformation de legendre-Fenchel :

utiliser pour : ▷ convexifier une fonction

▷ calculer le sous-diff d'une fct convexe

▷ calculer des problèmes dits « duaux ». ces problèmes apportent bcq d'info sur les problèmes « primaux » ceux que l'on souhaite résoudre

▷ passer de la mécanique lagrangienne à la méca hamiltonienne.

① def : Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexe sur $\mathbb{R}^n$.

Sa transformé de legendre est défini par : $f^*(s) = \sup_{x \in \mathbb{R}^n} (\langle s, x \rangle - f(x))$

- l'app $f \to f^*$ est appelé conjugaison ou transformation de Legendre-Fenchel
- f^* est appelé conjuguée convexe, transformé de Fenchel ou transformé de legendre-Fenchel de f.

proposition : f^* est convexe sur $\mathbb{R}^n$

② def : inégalité de Fenchel (i) $f(x) + f^*(s) \geq \langle s, x \rangle$

~~cette inégalité~~ $s \in \partial f(x) \Leftrightarrow f(x) + f^*(s) = \langle s, x \rangle$

(ii) $f^*(0) = \sup(-f) = -\inf(f)$

③ def : biconjugué :

soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction. Sa biconjugée est défini par : $f^{**}(x) = \sup_{s \in \mathbb{R}^n} (\langle x, s \rangle - f^*(s))$

▷ th1 :

f^{**} est la plus grande fonction convexe fermé inférieure à f

▷ th2 :

f^{**} satisfait $f^{**} = f \Leftrightarrow f$ est convexe fermée

cherche sa signification et graphe

Démonstration de : f^* est une fonction convexe sur $\mathbb{R}^n$

Soit $s, t \in \mathbb{R}^n$, $\alpha \in [0,1]$, Mq $f^*(\alpha t + (1-\alpha) s) \leq \alpha f^*(t) + (1-\alpha) f^*(s)$

$\rightarrow$ On a $f^*(\alpha t + (1-\alpha) s) = \sup\limits_{x \in \mathbb{R}^n} \left(\langle \alpha t + (1-\alpha) s, x \rangle - f(x) \right)$

$$= \sup_{x \in \mathbb{R}^n} \left(\langle \alpha t, x \rangle + (1-\alpha) \langle s, x \rangle - f(x) \right)$$

$$= \sup_{x \in \mathbb{R}^n} \left(\alpha \langle t, x \rangle - \alpha f(x) + (1-\alpha) \langle s, x \rangle - (1-\alpha) f(x) \right)$$

Donc

$$f^*(\alpha t + (1-\alpha) S) \leq \sup_x \left(\alpha \langle t, x \rangle - \alpha f(x) \rangle \right) \quad x \in \mathbb{R}^n$$
$$+ \sup_x \left((1-\alpha) \langle s, x \rangle - (1-\alpha) f(x) \right) \quad x \in \mathbb{R}^n$$

$$f^*(\alpha t + (1-\alpha) s) \leq \alpha f^*(t) + (1-\alpha) f^*(s)$$

Donc f^* est convexe

Proposition : Soit f convexe fermée. $\forall (s, x) \in \mathbb{R}^n \times \mathbb{R}^n$:

$$s \in \partial f(x) \Leftrightarrow x \in \partial f^*(s)$$

Exercice à faire :

① Dém que $\partial f(x)$ convexe fermée

② Dém équivalence de page 3

③ dém : $\partial f(x) = \{\nabla f(x)\}$

④ : $f(x) + f^*(s) = \langle s, x \rangle \Leftrightarrow s \in \partial f(x)$

⑤ : Dém th2 page 4

⑥ Soient I un ensemble d'indices et $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$; $i \in I$

Mq $\left(\inf_{i \in I} f_i\right)^* = \sup_{i \in I} f_i^*$

→ Demonstration du Proposition de Legend-Fenchel (Ex 4)

on a $s \in \partial f(x)$

$\Leftrightarrow f^*(s) = \langle s, x \rangle - f(x)$

$\Leftrightarrow f^*(s) + f(x) = \langle s, x \rangle$

$\Leftrightarrow f^*(s) + f^{**}(x) = \langle s, x \rangle$

$\Leftrightarrow x \in \partial f^*(s)$

Remarque :

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fct convexe de classe C^1. Alors la fonction $x \to \langle p, x \rangle - f(x)$ atteint son supremum sur $\mathbb{R}^n$ en un point unique $x(p) \in \mathbb{R}^n$. Ainsi :

$$\left[\nabla(\langle p, x \rangle - f(x))\right]_{x = x(p)} = 0$$

implique

$$\left[\nabla(\langle p, x\rangle)\right]_{x=x(p)} = \left[\nabla(f(x))\right]_{x=x(p)}$$

implique

$$\left[\nabla(\langle p, x(p)\rangle)\right] = \nabla(f(x(p)))$$

$$\Rightarrow \nabla\left(\sum p_i x_i\right)_{x=x(p)} = \left[\nabla(f(x(p)))\right]$$

$$\Rightarrow \boxed{p = \left[\nabla f(x(p))\right]}$$

en effet Soit $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ $p = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_L \end{pmatrix}$

$$\nabla\left(\sum p_i x_i\right) = \begin{pmatrix} \frac{\partial p_1 x_1}{\partial x_1} \\ \frac{\partial p_2 x_2}{\partial x_2} \\ \vdots \\ \frac{\partial p_n x_n}{\partial x_n} \end{pmatrix} = p \begin{pmatrix} \frac{\partial x_1}{\partial x_1} \\ \vdots \\ \frac{\partial x_n}{\partial x_n} \end{pmatrix} = p$$

Donc $\left[\nabla\left(\sum p_i x_i\right)\right]_{x=x(p)} = p$

Remarque :

Pour $n=1$, Si $x \in \mathbb{R}$ et $f : \mathbb{R} \to \mathbb{R}$ convexe de C^1 Sur $\mathbb{R}$, Sa transformé de Legendre est la fonction g de variable p défini comme suit :

$$g(p) = \sup_{x \in \mathbb{R}} \left(px - f(x) \right)$$

$g(p)$ represente la distance maximal entre f et la droite $y = px$.

g est la transformé de Legendre de la fonction f par rapport à x donc $g(p) = p \cdot x(p) - f(x(p))$ où $x(p)$ est défini par la condition extrémale $\frac{\partial}{\partial x}\left(px - f(x)\right) = 0$, càd $f'(x(p)) = p$. Et puisque f est convexe, $\underline{x(p) \text{ est unique}}$

Exemple : ! tu dois savoir d'abord si la fct est convexe ou non !

- $f(x) = x^2$, cherchons $g(p)$

on doit chercher $x(p)$, $x(p)$ vérifie $f'(x(p)) = p$

$$\Rightarrow 2x(p) = p \Rightarrow \boxed{x(p) = \frac{p}{2}}$$

Donc

$$g(p) = p \cdot x(p) - f(x(p))$$
$$= \frac{p^2}{2} - \frac{p^2}{4} = \frac{p^2}{4}$$

- $f(x) = \frac{m x^2}{2}$; $m \in]0, +\infty[$

$\to f'(x) = 2xm$

$\to f'(x(p)) = p \Leftrightarrow x(p)m = p$

$$\Leftrightarrow \boxed{x(p) = \frac{p}{m}}$$

$$g(p) = p \cdot \frac{p}{m} - \frac{m}{2} x^2(p)$$
$$= \frac{p^2}{m} - \frac{m}{2}\frac{p^2}{m^2}$$
$$= \frac{p^2}{m} - \frac{p^2}{2m} = \frac{p^2}{2m}$$

- $f(x) = \frac{a}{2}x^2 + bx + C$

$$g(p) = \frac{(p-b)^2}{2a} - C$$

$\to$ transformation de Legendre des formes quadratiques :

Soit $f(X) = \frac{1}{2} X^T A X$ une forme quadratique définie positive d'ordre n, i.e $f(X) > 0$ $\forall X \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$, où A est <u>une matrice Symetrique definie positif</u> d'ordre n.

on a

$$\nabla f(X) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(X) \\ \vdots \\ \frac{\partial f}{\partial x_n}(X) \end{pmatrix} = AX$$

→ Voir démonstration (Rappel)

Exemple :

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 3 & -4 \\ 2 & -4 & 11 \end{pmatrix}, \quad f(X) = \frac{1}{2} X^T A X$$

Soit $g(Y)$ la transformée de Legendre de $f(X)$ par rapport à X, d'après ce qui précède on en déduit ;

$$\boxed{g(Y) = \frac{1}{2} Y^T . A^{-1} . Y} \longleftarrow \text{Voir démonstration}$$

$$= \frac{1}{2} \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix} . \frac{1}{10} \begin{pmatrix} 17 & 3 & -2 \\ 3 & 7 & 2 \\ -2 & 2 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$= \frac{1}{20} \left(17 y_1^2 + 6 y_1 y_2 - 4 y_1 y_3 + 7 y_2^2 + 4 y_2 y_3 + 2 y_3^2 \right)$$

Démonstration de $\nabla f(X) = AX$

$\rightarrow f(X) = \frac{1}{2} X^T A X = \frac{1}{2} \sum_{i=1}^{n} a_{ij} x_i x_j$ avec $a_{ij} = a_{ji}$

$$= \frac{1}{2}\left(\sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{i<j} a_{ij} x_i x_j \right)$$

$$= \frac{1}{2}\Big(a_{11} x_1^2 + a_{22} x_2^2 + \dots + a_{nn} x_n^2 + 2a_{12} x_1 x_2 + 2a_{13} x_1 x_3 + 2a_{1n} x_1 x_n + 2a_{23} x_2 x_3 + 2a_{24} x_2 x_4 + \dots + 2a_{2n} x_2 x_n + \dots + 2a_{(n-1)n} x_{n-1} x_n \Big)$$

$$\rightarrow \nabla f(X) = \begin{pmatrix} \frac{\nabla f(X)}{\partial x_1} \\ \vdots \\ \frac{\nabla f(X)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left(2a_{11} x_1 + 2a_{12} x_2 + \dots + 2a_{1n} x_n\right) \\ \frac{1}{2}\left(2a_{22} x_2 + 2a_{12} x_1 + \dots + 2a_{2n} x_n\right) \\ \vdots \\ \frac{1}{2}\left(2a_{nn} x_n + 2a_{1n} x_1 + 2a_{2n} x_2 + \dots + 2a_{(n-1)n} x_{n-1}\right) \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n \\ a_{12} x_2 + a_{22} x_2 + \dots + a_{2n} x_n \\ \vdots \\ a_{1n} x_1 + a_{2n} x_2 + \dots + a_{nn} x_n \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = AX$$

Donc $\nabla f(X) = AX$

et $\nabla^2 f(X) = A$

soit $g(Y)$ la transformé de Legendre de la fonction f pour $X \in \mathbb{R}^n$

$$g(Y) = \sup\left(\langle Y, X \rangle - f(X)\right)$$

$$= \langle Y, X(Y) \rangle - f(X(Y))$$

on a $\nabla f(X(Y)) = A.X(Y) \Rightarrow X(Y) = A^{-1}.Y$

Donc

$$g(Y) = Y^T . X(Y) - f\left(X(Y)\right)$$

$$= Y^T . X(Y) - \frac{1}{2} X^T(Y) . A . X(Y) \quad \text{avec } f(X) = \frac{1}{2} X^T . A . X$$

$$= Y^T A^{-1} Y - \frac{1}{2} (A^{-1} Y)^T . A . A^{-1} . Y$$

$$= Y^T A^{-1} Y - \frac{1}{2} Y^T . (A^{-1})^T . Y$$

$$= Y^T A^{-1} Y - \frac{1}{2} Y^T . (A^T)^{-1} . Y$$

$$= Y^T A^{-1} Y - \frac{1}{2} Y^T A^{-1} . Y \quad \text{Puisque } A \text{ est Symétrique } A^T = A$$

$$\boxed{g(Y) = \frac{1}{2} Y^T . A^{-1} . Y}$$