Optimisation Convexe 1 suite

Trouver l'extremum d'une parabole

$f(x) = ax^2 + bx + c \qquad a > 0$
$f'(x) = 2ax + b$
$x^* \qquad tqf'(x^*) = 0$
$2ax^* + b = 0$
$x^* = -\frac{-b}{2a}$

$$\begin{aligned} f^* &= f(x^*) \\ &= a\left(-\frac{b}{2a}\right)^2 + b\left(-\frac{b}{2a}\right) + c \\ &= \frac{b^2}{4a} - \frac{b^2}{2a} + c \\ &= -\frac{b^2}{4a} + c \end{aligned}$$

$f : \mathbb{R} \longrightarrow \mathbb{R}$
f est dérivable en x_0 : $\lim\limits_{h\to 0}\frac{f(x+h)-f(x)}{h}$ est finie.

Et $\lim\limits_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h} = f'(x_0)$
$\lim\limits_{h\to x_0}\frac{f(x)-f(x_0)}{x-x_0}$
$f = o_{x_0}(g)$ f est négligeable par rapport à g en x_0.
$\Leftrightarrow$ Il existe une fonction $\varepsilon : \mathbb{R} \to R$ avec $\varepsilon(x) \underset{x\to x_0}{\longrightarrow} 0$
Et $f(x) = \varepsilon(x)g(x)$ au voisinage de x_0.

Si g ne s'annule pas au voisinage de x_0
$f = o_{x_0}(g) \Leftrightarrow \lim\limits_{x\to x_0}\frac{f(x)}{g(x)} = 0$

$\lim\limits_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h} = f'(x_0)$
$\lim\limits_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h} - \frac{hf'(x_0)}{h} = 0$
soit $\varepsilon : \mathbb{R} \longrightarrow \mathbb{R}$
tq $\varepsilon(h) \underset{h\to 0}{\to} 0$
$\lim\limits_{h\to 0}\frac{f(x_0+h)-f(x_0)-hf'(x_0)}{h} = \lim\limits_{h\to 0}\varepsilon(h) = 0$

$\frac{f(x_0+h)-f(x_0)-hf'(x_0)}{h} = \varepsilon(h)$
$f(x_0+h) - f(x_0) - hf'(x_0) = h\varepsilon(h)$

$f(x_0+h) = f(x_0) + hf'(x_0) + h\varepsilon(h) = f(x_0) + hf'(x_0) + o_0(h)$
$f(x) = f(x) + (x - x_0)f'(x_0) + (x - x_0)\underbrace{\varepsilon(x - x_0)}_{o_0(x-x_0)}$

$f : \mathbb{R}^n \longrightarrow \mathbb{R}$
$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longmapsto f(x_1, \ldots, x_n)$$

$f(x_1, \ldots, x_n) = x_1 + x_2 + \ldots + x_n$
La $k^{ième}$ dérivée partielle de f existe en $x_0 \in \mathbb{R}^n$
$\Leftrightarrow$ la fonction $\underset{t\to f(x_0,\ldots,x_n)}{\varphi : \ \mathbb{R} \to \mathbb{R}}$ est dérivable en 0

et $\varphi'(0) = \frac{\partial f}{\partial x_k}(x_0) \Leftrightarrow \partial k f(x_0)$

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0) \end{cases}$$

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \frac{\partial}{\partial x}\left(\frac{xy}{x^2+y^2}\right) \\ &= y\frac{\partial}{\partial x}\left(\frac{x}{x^2+y^2}\right) \\ &= y\frac{x^2+y^2-x(2x)}{(x^2+y^2)^2} \\ &= y\frac{y^2-x^2}{(x^2+y^2)^2} \end{aligned}$$

$\frac{\partial f}{\partial x}(t,0) = 0 \qquad \frac{\partial f}{\partial y}(0,t) = 0$

$\frac{\partial f}{\partial x}(x,y) \Leftrightarrow$ on dérive selon l'axe (o_x)

$\Leftrightarrow$ on dérive selon le vecteur $e_x = (1, 0)$

$\frac{\partial f}{\partial y}(x, y) \Leftrightarrow$ on dérive selon l'axe (o_y)

La derivee directionnelle

Dans le cas de n variables :
$f : \mathbb{R}^n \longrightarrow \mathbb{R}$
$\frac{\partial f}{\partial x_k}(x)$ = on dérive par rapport à la $k^{\text{ième}}$ variable
$\Leftrightarrow$ on dérive selon la $k^{\text{ième}}$ variable
$\Leftrightarrow$ on dérive selon le vecteur $ek = (0, \ldots, o, \underbrace{1}_{k^{\text{ième}}}, 0, \ldots, 0)$

Définition : On appelle dérivée directionnelle de f en x_0 suivant le vecteur $h \in \mathbb{R}^2$ et on note $D_h f(x_0)$ la dérivée en 0 de la fonction

$$\varphi : \begin{array}{l} \mathbb{R} \longrightarrow \mathbb{R} \\ t \longmapsto f(x_0 + th) \end{array}$$

$\frac{\partial f}{\partial x}(x_0) \equiv$ derivee de φ : $\begin{array}{l} \mathbb{R} \longrightarrow \mathbb{R} \\ t \longmapsto \underbrace{f(x_{01}+, \ldots, x_{0k+t}, \ldots, x_{0n})}_{\begin{pmatrix} x_{01} \\ \vdots \\ x_{0n} \end{pmatrix} + t \begin{pmatrix} 0 \\ \vdots \\ 1 \to k^e \\ \vdots \\ 0 \end{pmatrix}} \end{array}$

$f : \mathbb{R}^2 \longrightarrow R \qquad x_0 = (1, 2)$
$(x, y) \longmapsto x^2 - y^2 \qquad h = (3, 5)$

$\varphi : \mathbb{R} \longrightarrow \mathbb{R}$
$t \longmapsto f(x_0 + th)$

$\varphi(t) = f\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} + t\begin{pmatrix} 3 \\ 5 \end{pmatrix}\right)$
$= f(1 + 3t, 2 + 5t)$
$= (1 + 3t)^2 - (2 + 5t)^2$
$= 1 + 6t + 9t^2 - (4 - 20t + 25t^2)$
$= -3 - 14t - 16t^2$

$\varphi'(t) = -14 - 32t$
$\varphi'(0) = -14 = D_h(x)$

$h \leftrightarrow \alpha h$
$D_{\alpha h} f(x_0) = \alpha D_h f(x_0)$
On parle de dérivée directionnelle selon la direction de $h \in \mathbb{R}^n \backslash \{0\}$ uniquement quand h est unitaire (par opposition à la dérivée directionnelle selon le vecteur h).

Malheureusement, l'existence de derivees directionnelles en Vn point selon tout vecteur n'implique pas la continuité en ce point.

$f(x, y) = \begin{cases} \frac{y^2}{x} & x \neq 0 \\ y & x = 0 \end{cases}$

En $(0, 0)$ soit $h = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \neq (0, 0)$

$$\begin{aligned} \varphi(t) = f(th) = f(th_1, th_2) &= \begin{cases} \frac{(th_2)^2}{th_1} & h \neq 0 \\ th_2 & h = 0 \end{cases} \\ &= \begin{cases} t\frac{h_2^2}{h_1} & h \neq 0 \\ th_2 & h = 0 \end{cases} \end{aligned}$$

$\varphi'(t) = \begin{cases} \frac{h_2^2}{h_1} & h_1 \neq 0 \\ h_2 & h_1 = 0 \end{cases}$ = $\varphi'(0)$

si g est continue en 0 et f continue en $g(0)$ alors $f \circ g$ est continue en 0
$g : \mathbb{R} \longrightarrow \mathbb{R}^2$
$t \longmapsto (t^2, t)$

$f \circ g : \mathbb{R} \longrightarrow \mathbb{R}^2$
$t \longmapsto f \circ g(t) = f(g(t))$

$f(g(t)) = f(t^2, t) = \begin{cases} \frac{t^2}{t^2} & t \neq 0 \\ 0 & t = 0 \end{cases}$

$f \circ g(t) = \begin{cases} 1 & t \neq 0 \\ 0 & t = 0 \end{cases}$

Donc $f \circ g$ pas continue en 0
$\Rightarrow f$ pas continue en $g(0) = (0,0)$

$$f(x_0+h) = f(x_0) + hf'(x_0) + \begin{cases} o_0(h) \\ h\varepsilon(h) \text{ avec } \varepsilon(h) \end{cases} \qquad \varepsilon(h) \underset{h\to 0}{\longrightarrow} 0$$

> **Définition** : On dit que $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est differentiable de x_0 ssi il existe une application linéaire $d_{x_0}f$ (aussi noté df_{x_0}) tq $f(x_0+h) = f(x_0) + d_{x_0}f(h) + \underset{||H||\varepsilon(h)}{o_0(h)}$

$h \mapsto hf'(x_0)$ est linéaire $\varepsilon : \mathbb{R}^n \longrightarrow \mathbb{R}$
$d_{x_0}f : h \mapsto d_{x_0}f(h) = h \times f'(x_0)$

> **Propriété** : Si f est différentiable en x_0 alors f est continue en x_0
> **Propriété** : Si f est différentiable en x_0 alors f admet des dérivées directionnelles selon tout vecteur $h \in \mathbb{R}^n \backslash \{0\}$, et la dérivée directionnelle vaut $D_h f(x_0) = d_{x_0}f(h)$

Soit f différentiable en x_0.
Donc les dérivées partielles $\frac{\partial f}{\partial x_k}$ existent en x_0

Soit $h \in \mathbb{R}^n \backslash \{0\}$ et $(e_1, \dots, e_n)$ la base

$$h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \sum_{i=1}^n h_i e_i$$

$$D_h f(x_0) = d_{x_0}f(h) = d_{x_0}f(\sum_{i=1}^n h_i e_i) = \sum_{i=1}^n h_i d_{x_0}f(e_i) = \sum_{i=1}^n h_i \frac{\partial f}{\partial x_i} x_0$$

$d_{x_0}f(h) = \langle \nabla f(x_0), h \rangle$

Soit $f : \mathbb{R}^n \to \mathbb{R}$
on définit le vecteur gradient de f en x_0 par

$$\nabla f(x_0) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_0) \end{pmatrix}$$

Si f différentiable en x_0, alors $d_{x_0}f : h \longmapsto \langle \nabla f(x), h \rangle$
$d_{x_0} : h \longmapsto hf(x_0)$

Soit

$$f : \begin{matrix} \mathbb{R}^n \longmapsto \mathbb{R}^p \\ x = (x_1, \dots, x_n) \longmapsto f(x) = (f_1(x), \dots, f_p(x)) \end{matrix}$$

Soit $x_0 \in \mathbb{R}^n$ et f différentiable en x_0
Les $f_1, \dots, f_p$ sont différentiables en x_0

Soit $h \in \mathbb{R}^n$ $\quad \overbrace{f(x+h)}^{\in\mathbb{R}^p} = \overbrace{f(x_0)}^{\in\mathbb{R}^p} + \overbrace{d_{x_0}f(h)}^{\in\mathbb{R}^P} + o_0(h)$

$$\begin{pmatrix} f_1(x_0+h) \\ \vdots \\ f_p(x_0+h) \end{pmatrix} = \begin{pmatrix} f_1(x_0) \\ \vdots \\ f_p(x_0) \end{pmatrix} + \begin{pmatrix} dx_0 f_1(h) \\ \vdots \\ dx_0 f_p(h) \end{pmatrix} + o_0(h)$$

$$f(x_0+h) = f(x_0) + \begin{pmatrix} \langle \nabla f_1(x_0), h \rangle \\ \vdots \\ \langle \nabla f_p(x_0), h \rangle \end{pmatrix} + o_0(h)$$

$$\langle \nabla f_i(x_0), h \rangle = \nabla f_i(x_0)^T h = \left(\frac{\partial f_i}{\partial x_1}(x_0), \dots, \frac{\partial f_i}{\partial x_n}(x_0) \right) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \qquad \frac{\partial f_i}{\partial x_j}(x_0) = \partial_j f_i(x_0)$$

$$f(x_0+h) = f(x_0) + \begin{pmatrix} (\partial_1 f_1(x_0) \dots\dots \partial_n f_1(x_0))h \\ \vdots \\ (\partial_1 f_p(x_0) \dots\dots \partial_n f_p(x_0))h \end{pmatrix} + o_0(h)$$

$$\text{les p composantes de f} : \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \dots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(x_0) & \dots & \frac{\partial f_p}{\partial x_n}(x_0) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

On appelle jacobienne de f en $x_0 = (u_1, \dots, u_n) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ la matrice :

$$\mathcal{J}_{x_0} f = \left[\frac{\partial f_i}{\partial x_j}(x_0) \right]_{\substack{i=1,\dots,p \\ j=1,\dots,n}}$$

Telle que $\underbrace{f(x_0+h)}_{\in \mathbb{R}^p} = \underbrace{f(x_0)}_{\in \mathbb{R}^p} + \underbrace{\underbrace{\mathcal{J}_{x_0} f}_{\in \mathbb{M}_{p,n}(\mathbb{R})} \times \underbrace{h}_{\in \mathbb{R}^p}}_{\in \mathbb{R}^p} + o_0(h)$

$d_{x_0} f : h \longmapsto \mathcal{J}_{x_0} f \times h$ est bien linéaire

Soit $f : \mathbb{R}^n \to \mathbb{R}^p$ differentiable en $x_0 \in \mathbb{R}^n$
Soit $g : \mathbb{R}^p \to \mathbb{R}^n$ differentiable en $f(x_0) \in \mathbb{R}^p$

Alors la composee $g \circ f = d_{f(x_0)} g \circ d_{x_0} f$
Avec les jacobiennes $\mathcal{J}_{x_0} g \circ f = \mathcal{J}_{f(x_0)} g \times^{\text{produit matriciel}} \mathcal{J}_{x_0} f$

$(g \circ f)' = f' \times (g' \circ f)$
$(g \circ f)(x) = g(f(x))$
$(g \circ f')(x) = f'(x) \times g'(f(x))$

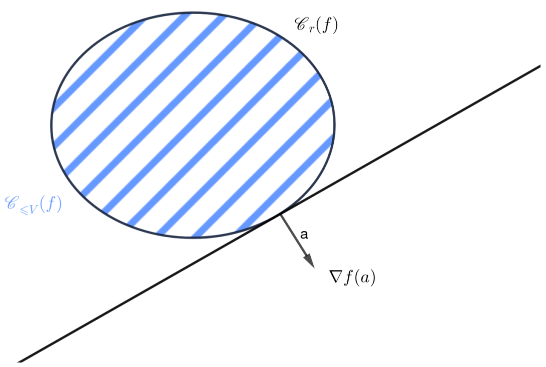
Interprétation géométrique du gradient

On se limite désormais au cas des fonctions convexes.
Quand les resultats énoncés s'appliquent à un cadre plus général que l'on spécifiera. Les questions auxquelles on n'a pas encore de réponses générales :

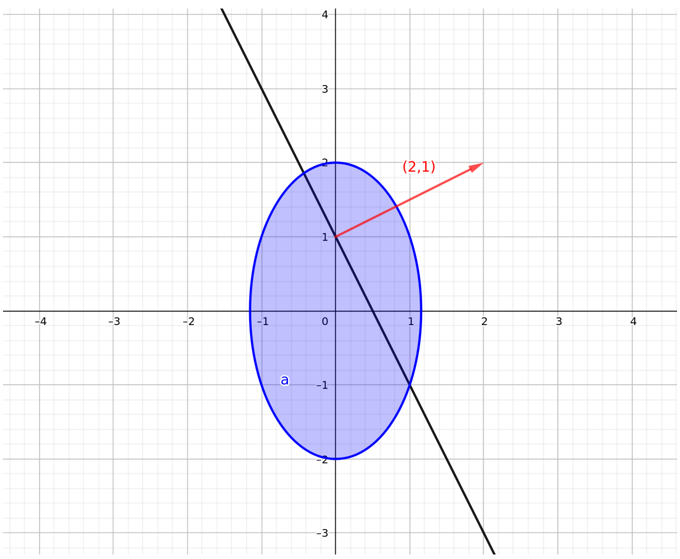
- "Direction" de minimisation d'une fonction objectif
- Trouver des hyperplans d'appui au lieu admissible d'un problème d'optimisation

C'est la proposition suivante qui permet d'apport une réponse à ces 2 questions :

Proposition :
Soit $f : U \subset \mathbb{R}^N \longrightarrow \mathbb{R}$ fonction convexe et différentiable en $a \in U$. $\nabla f(a)$ définit un hyperplan d'appui à $\mathcal{C}_{\leq r}(f)(r = f(a))$



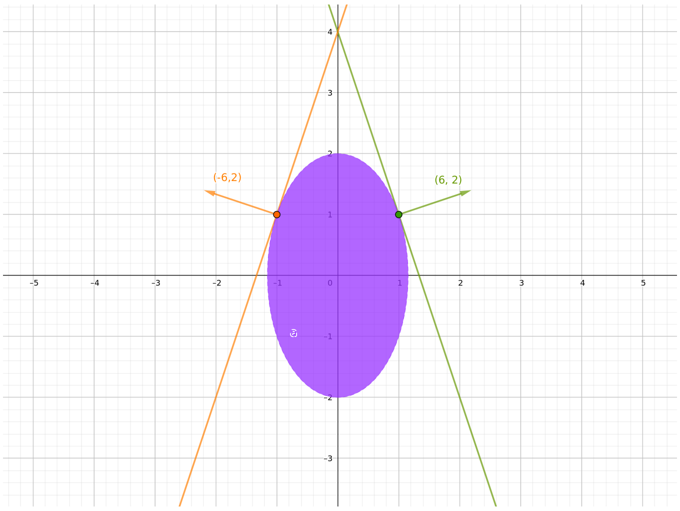
[Q 4-33]
On cherche à résoudre le problème de minimisation : $\min f_0(x,y) = 2x + y$
sujet à : $3x^2 + y^2 \leqslant 4$



Pour minimiser f_0 on part vers la “gauche” du dessin ; vers la direction opposée au gradient de f_0.
La position “limite” de ces courbes de niveaux (la courbe de niveau qui réalise la valeur optimale) correspond à un hyperplan d’appui.

$\mathbb{A}$ est le sous-niveau de niveau 4 de $f_1(x, y) = 3x^2 + y^2$

$$\nabla f_1(x, y) = \begin{pmatrix} 6x \\ 2y \end{pmatrix}$$



On cherche donc un point (x, y) tel que :

$$\nabla f_1(x, y) + \lambda \nabla f_0(x, y) = 0 \qquad \text{avec } \lambda \geqslant 0$$

Pour trouver (x,y) on cherche a resoudre:
$$\begin{cases} \begin{pmatrix} 6x \\ 2y \end{pmatrix} = -\lambda \begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ 3x^2 + y^2 = 4 \end{cases}$$

Des deux premières équations on obtient:

$$x = -\frac{\lambda}{3}, y = -\frac{\lambda}{2}$$

En réinjectant dans la deuxième équation :

$$x = -\frac{1}{\sqrt{21}}, y = -2\sqrt{\frac{3}{7}}$$

Définition :
Avec les notations de la proposition on appelle espace tangeant à $\mathcal{C}_r(f)$ en a l’espace affine.

$$\begin{aligned} T_a(f) &= a + \nabla f(a)^{\perp} \\ &= a + \{x | \nabla f(a)^{\top} x = 0\} \end{aligned}$$

La proposition donne, telle quelle, la réponse à la question 2. posée ci-dessus. Elle suggère également une direction vers laquelle minimiser la valeur objectif de f. On suppose que $\nabla f(a) \neq 0$, on regarde $-t\nabla f(a)$ avec $t > 0$.

Pour t proche de 0,

$$f(u|\nabla f(a)) - f(a) = \nabla f(a)^{\top}(-t\nabla f(a)) + ||t\nabla f(a)||\epsilon(\nabla f(a))$$

le $O_o(t)$ est négligable devant $-t\nabla f(a)^{\top}\nabla f(a) = -t||\nabla f(a)||_2^2$
L’expression $f(a - t\nabla f(a)) = f(a)$ est du signe de $-t||\nabla f(a)||_2^2$. Pour t assez pertit

$$f(a - t\nabla f(a)) \leqslant f(a)$$

Remarque :
1. L’étude précédente est contrainte par le fait "t assez petit". Ca donne une idée de direction du min, pas une garantie.
2. L’étude ci-dessus ne nécessite pas de convexité.

Caracteristique du premier ordre de la convexite :

$f : T \subset \mathbb{R}^b \longmapsto \mathbb{R}$ est convexe si:

- U est convexe
- $\forall x, y \in U, f(y) - f(x) \geqslant \nabla f(x)^T(y-x)$

En supposant cette caracterisation VRAIE :

Soit $y \in \mathcal{C}_{\leq r}(f)$; on veut montrer $\nabla f(x)^T(y-x) \leq 0$

Or comme f est convexe on a :
$\nabla f(x)^T(y-x) \leqslant \underbrace{f(y)}_{\leq r}\underbrace{(fx)}_{=r}$
d' ou $\nabla f(x)^T(y-x) \leqslant 0$

Preuve de la caractérisation de convexité

Convexe $\Leftrightarrow$ (∇ convexe)

Soient $x, y \in U, t \in [0,1]$
On regarde la fonction

$$g(t) = f((1-t)x + ty)$$

La definition de convexite de f :

$$\begin{array}{ccl} & f((1-t)x+t(y)) & \leqslant (1-t)f(x) + tf(y) \\ \Leftrightarrow & g(t) & \leqslant (1-t)g(0) + g(1) \\ \Leftrightarrow & g(t) - g(0) & \leqslant t(g(1) - g(0)) \\ \Leftrightarrow & \dfrac{g(t)-g(0)}{t} & \leqslant g(1) - g(0) \\ \Rightarrow & g(0) & \leqslant f(y) - f(x) \end{array}$$

Or $g(0)\nabla f(x)^T(y-x)$
D'ou $\boxed{\nabla f(x)'(y-x) \leqslant f(y) - f(x)}$

(∇ convexe) et U convexe $\Rightarrow$ convexe

Soient $x, y \in U \qquad z_t = (1-t)x + ty$

$$\begin{array}{c} t \times [f(y) - f(z_t)] \geqslant \nabla f(z_t)^\top (y - z_t) \\ (1-t) \times [f(x) - f(z_t) \geqslant \nabla f(z_t)(x - z_t)] \\ tf(y) + (1-t)f(x) + tf(z_t) - (1-t)f(z_t) \geqslant \nabla f(z_t)(t(y-z_t) + (1-t)(x-z_t))(D) \end{array}$$

(D) :
$\nabla f(z_t)(ty + (Lt)x - z_t) = 0$

$(D) = 0$

(G) :

$$\begin{array}{l} tf(y) + (1-t)f(x) + f(z_t) \\ = tf(y) + (1-t)f(x) + f(ty + (1-t)x) \end{array}$$

> **Exercice :**
> Trouver les points sur le paraboloïde $z = 4x^2 + y^2$ où le plan tangent est parallèle au plan $x + 2y + z = 6$.
> De même pour le plan $3x + 5y - 2z = 5$

Problèmes d'optimisation

Catégorie des problèmes convexes

> **Convention :**
> pour simplifier la notation on note $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction qui n'est pas nécessairement, définie sur $\mathbb{R}^n$

> **Définition :**
> Un problème d'optimisation convexe est un problème qui s'exprime sous la forme $\min f_0(x)$, sujet à:
>
> $$\begin{array}{l} f_i(x) \leqslant 0 \, \forall i \in \{1, \ldots, m\} \\ f_j(x) = 0 \, \forall j \in \{1, \ldots, p\} \end{array}$$
>
> Où $f_0, f_i, h_j : \mathbb{R}^n \longrightarrow \mathbb{R}$ sont convexes et de plus les h_j sont affines.

On peut en particulier réécrire (P) sous la forme: $\min f_0(x)$ sujet à :

$$\begin{array}{c} f_i(x) \leqslant 0 \, \forall i \in \{1, \ldots, m\} \\ Ax = b \end{array}$$

où $\mathcal{A} \in M_{p,n}(\mathbb{R}); b \in M_{p,1}(\mathbb{R})$

On dit qu'un point x est admissible s'il satisfait les contraintes définies par (P). Le lieu admissible $\mathcal{A}$ de (P) correspond aux points admissibles de (P). On fait remarquer que sous nos hypotheses, $\mathcal{A}$ est convexe. On note p^* la valeur optimale de (P):

$$p^* = \underset{x \in \mathcal{A}}{inf}\{f_0(x)\}$$

Par convention si $\mathcal{A} = \emptyset; p^* = +\infty$. Dans le cas sur $p^* = -\infty$ on dit que (P) est non borné.
On appelle enfin point optimal x^* de (P) tout point tq $f_0(x^*) = p^*$. Un tel point n'existe pas toujours; par exemple c'est le cas $\underset{x \in \mathbb{R}_+^*}{min} \frac{1}{x}$. De plus, il n'existe pas en général qu'un seul point optimal (quand il y en a); prendre par exemple le problème:

$$\underset{x \in \mathbb{R}}{min} 10$$

L'écriture de (P) dans la définition est appelée standard d'un problème d'optimisation. Il existe une notion théorique d'équivalence de problème d'optimisation, On ne rentrera pas dans le détail, sachez qu'elle consiste à réexprimer un problème d'optimisation de façon à le résoudre plus facilement.

Exemple :
$\min |x|$ sujet à:

$$\begin{aligned} x - 2 \leqslant 0 \qquad (P_1) \\ -x - 2 \leqslant 0 \qquad\quad \end{aligned}$$

P_1 est équivalent à:

$\min -x^2$ sujet à:

$$\begin{aligned} x - 2 \leqslant 0 \\ -x - 2 \leqslant 0 \end{aligned}$$

Pourquoi la convexité ?

unicité du minimum

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction convexe, alors f :

- n'admet pas de maximum locaux strictes.
- Admet au plus un minimum local stricte.

Essayons de justifier le premier point. Supposons qu'il existe un voisinage $\mathcal{B}(x, \varepsilon)$ pour $\varepsilon > 0$ tq :

$$\forall y \in \mathcal{B}(x, \varepsilon), y \neq x f(y) < f(x)$$

Soient $y_1, y_2 \in \mathcal{B}(x, \varepsilon) \backslash \{x\}$
$\forall t \in [0, 1]$
(conv):$f(ty_1 + (1-t)y_2) \leqslant tf(y_1) + (1-t)f(y_2)$
Donc $tf(y_1) + (1-t)f(y_2) \leqslant f(x)$
car $f(y_1) \leqslant f(x)$ et $f(y_2) \leqslant f(x)$

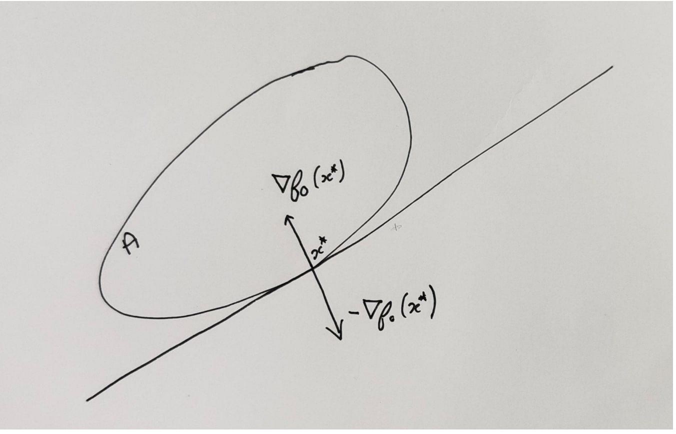
La condition précédente exprime le fait que la sécante au graphe de f sur $\mathcal{B}(x, \epsilon)$ est en dessous de celui ci, donc pas dans l'épigraphe de f. Dans ce cas f n'est pas convexe.

Pour le second point:

si y_1, y_2 sont 2 minimaux locaux et différents, on retrouve la situation qui contredit la convexité.

Condition d'existance d'un minimum sous contraintes

Si on est dans la situation suivante



Propriété :
Un point $x \in \mathcal{A}$ est optimal si:

$$\nabla f(x^*)^\top (y.x) \geqslant 0$$

$-\nabla f_0(x^*)^\top (y-x) \leqslant 0$
$-\nabla f_0(x^*)$ définit un hyperplan d'appui en x^* à $\mathcal{A}$.

Preuve :
Supposons x^* satisfait (op).
D'après les inégalités de convexité sur f_0 on a:

$$\forall y \in \mathcal{A}; \nabla f_0(x^*)^\top (y-x^*) \leqslant f(y) - f(x)$$

D'après (op):

$$f(y) - f(x^*) \geqslant 0 \Leftrightarrow f(y) \geqslant f(x^*)$$

La réciproque se fait par contraposition. On la laisse de côté pour cette fois.

Est-ce que l'hypothèse de convexité de (P) sur tout son domaine de définition est important ?

$\rightarrow$ Matheux dans sa tête : OUI
$\rightarrow$ Informaticien (matheux qui fait calculer) : BAH … CON vexe ?

Cas sans contrainte

On s'intéresse en un premier temps au problème d'optimisation de la forme:

$$\min_{x \in \mathbb{R}^n} f_0(x)$$

avec f_0 différentiable

Propriété :
Si x^* est un point optimal de f_0 alors:

$$\nabla f_0(x^*) = \underline{0}$$

Preuve :
On se place sur un voisinage $\mathcal{B}(x^*, \varepsilon)$ pour $\varepsilon > 0$ où $\forall y \in \mathcal{B}(x^*, \varepsilon); f_0(y) \geqslant f_0(x^*)$

En particulier pour le h assez proche de 0:

$$\begin{aligned} f_0(x^*+h) - f_0(x^*) &\geqslant 0 \\ \Rightarrow \nabla f_0(x^*)^T h + \theta_0 &\geqslant 0 \end{aligned}$$

Ainsi $\forall h \in \mathcal{B}(\underline{0}, \eta)$ pour $\eta > 0$

$$\nabla f_0(x^*)^T \geqslant 0$$

La seule application linéaire qui est possible sur un voisinage $\mathcal{B}(\underline{0}, \eta)$ est l'application nulle.

Dans le cas f_0 convexe, l'annulation du gradient en un point va nous limiter à un sous-lieu de points optimaux à étudier. En réalité, on a en général la situation suivante:

Les points critiques d'une fonction f_0 quelconque sont de l'une des trois formes suivantes:

1. Minimum locaux.
2. Maximums locaux.
3. Points selles.

Dans le cas convexe on a que des points du premier type. Dans ce cas l'étude des points critiques se confond avec celle des points minimaux.

Problème du dual

Soit P le problème d'optimisation:

$$\min f_0(x)$$

sujet à

$$\begin{aligned} f_i(x) \leqslant 0 \\ h_j(x) = 0 \end{aligned}$$

Si on voulait ramener l'étude de (P) à la minimisaion d'une seule fonction on pourrait étudier:

$$\Phi(x) = f_0(x) + \sum_{i=1}^{n} I_+(f_i(x)) + \sum_{j=0}^{p} I_0(f_j(x))$$

où

$$I_+(x) = \begin{cases} 0 \text{ si } x \leqslant 0 \\ +\infty \text{ sinon} \end{cases}$$
$$I_0(x) = \begin{cases} 0 \text{ si } x = 0 \\ +\infty \text{ sinon} \end{cases}$$

Problème d'optimisation équivalent à (P) mais inutilisable

Définition :
On appelle Lagrangien du problème (P) la fonction:

$$\mathcal{L}_P(\underset{\in\mathbb{R}^n}{x}, \underset{\in\mathbb{R}^m}{\lambda}, \underset{\in\mathbb{R}^p}{\nu}) = f_0(x) + \sum_{i=1}^{n} \lambda_i f_i(x) + \sum_{j=0}^{p} \nu_j f_j(x)$$

On définit le problème dual $(\check{P})$ de (P) comme suit: on note:

$$g(\lambda, \nu) = \underset{x\in\mathbb{R}^n}{inf} \mathcal{L}(x, \lambda, \nu)$$

avec cette notation:

$$\underset{\lambda,\nu}{max} g(\lambda, \nu) \qquad (\check{P})$$
$$\text{sujet à} \quad \lambda \geqslant 0$$

Remarque: $g(\lambda, \nu)$ pour $\lambda \geqslant 0$ est l'inf d'une fonction concave. (affines en les λ et ν) c'est donc concave (exo bribes de géometries)
Donc $(\check{P})$ est toujours un problème convexe.

Propriété :

$$\forall \lambda \geqslant 0; \text{on a} : g(\lambda, \nu) \leqslant p^*$$

Preuve :
Soit x un point admissible de (P). on a donc $f_i \leqslant 0$ et $h_j(x) = 0$.
Donc

$$\sum_{i=1}^{m} d_i f_i(x) + \sum_{j=1}^{P} \nu_j h_j(x) \leqslant 0 \,\forall \qquad \lambda \geqslant 0$$

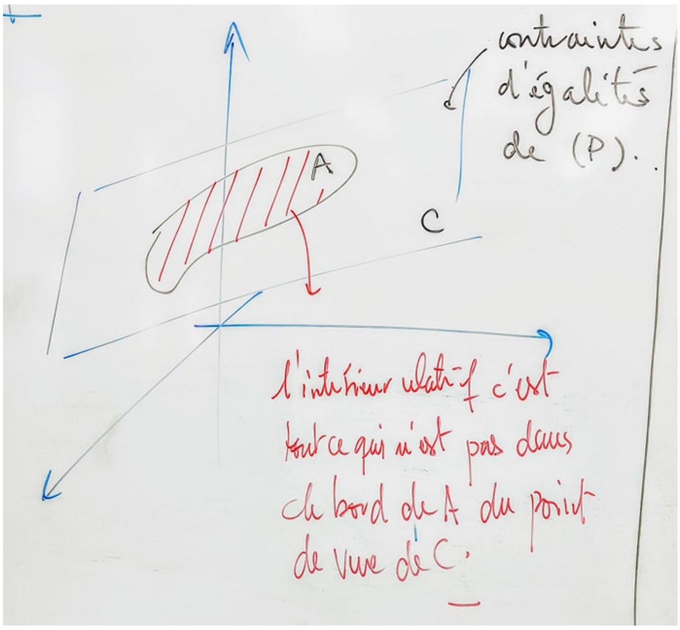
D'où:

$$\begin{aligned} \mathcal{L}_p(x, \lambda_1 \nu) \quad &= f_0(x) + \sum_{i=1}^{n} \lambda_i f_i(x) + \sum_{j=0}^{p} \nu_j f_j(x) \quad &\leqslant f_0(x) \\ \Rightarrow \underset{x}{inf} \mathcal{L}(x, \lambda, \nu) \quad &= g(\lambda, \nu) &\leqslant f_0(x) \\ \Rightarrow g(\lambda, \nu) \quad &\leqslant p^* \end{aligned}$$

Corollaire :
Si on note d^* la valeur optimale du dual on a : $d^* \leqslant p^*$

Question : Est ce qu'on a l'égalité ?
Dans la situation d'égalité on dit qu'on a une dualité forte entre § et $(\check{P})$

Condition de Slater : Si (P) est convexe et il existe un pt dans l'intèrieur relatif du domaine de définition de (P) tq :
$f_i(x) < 0$
$Ax = b$
alors (P) et $(\check{P})$ sont en dualité forte:



> **Définition :**
> On dit qu'un couple (λ, ν) est de t dual admissible si $\lambda \geqslant 0$ et $g(\lambda, \nu) > -\infty$. Les points (λ^*, ν^*) optimaux pour $\check{\mathcal{P}}$ sont parfois appelés multiplicateurs de Lagrange.

Les conditions KKT *(Karush-Kuhn-Tucker)*

Supposons que les valeurs optimales, primale et duale, soient atteintes et egales, en particulier on a une dualite forte. On designe par x^* (respectivement (λ^*, ν^*)) un point optimal de $\mathcal{P}$ (respectivement $\check{\mathcal{P}}$)

On a :

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf(\mathcal{L}_p(x, \lambda^*, \nu^*)) \\ &\leq \mathcal{L}_p(x^*, \lambda^*, \nu^*) \\ &= f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{j=1}^{p} \nu_j^* h_j(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

Toutes les inégalités qui apparaissent précédemment sont donc des égalités. On en déduit :

1. x^* minimise $\mathcal{L}_p(x, \lambda^*, \nu^*)$
2. $\sum_{i=1}^{m} \underbrace{\lambda_i^* f_i(x^*)}_{\leq 0} = 0$
   $\Rightarrow \forall i \in \{1, \ldots, m\}; \lambda_i^* f_i(x^*) = 0$

La fonction $x \longmapsto \mathcal{L}_p(x, \lambda^*, \nu^*)$ est convexe des que (P) l'est. Dire que x^* minimise $x \longmapsto \mathcal{L}_p(x, \lambda^*, \nu^*)$ est equivalent a dire que

$$\nabla_x \mathcal{L}_p(x^*, \lambda^*, \nu^*) = 0$$
$$\Leftrightarrow \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{p} \nabla j^* \nabla hj(x^*) = 0$$

Pour resumer (x^*, λ^*, ν^*) verifient les contraintes :
$f_i(x^*) \leqslant 0 \qquad \forall i \in \{1, \ldots, m\}$
$h_j(x^*) = 0 \qquad \forall j \in \{1, \ldots, p\}$
$\lambda_i^* \geqslant 0 \qquad \forall i \in \{1, \ldots, m\}$ (KKT)
$\lambda_i^* f_i(x^*) = 0 \qquad \forall i \in \{1, \ldots, m\}$

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(x^*) = 0$$

> Propriété : Quand (P) est un problème convexe et dans le cas de forte dualité, (condition de Slater satisfaite, par exemple) les conditions KKT sont nécessaires et suffisantes pour avoir une pair primal-dual optimale.

Exercices :
Résoudre en utilisant les conditions KKT

1

$$min_{x\in\mathbb{R}^2} \quad \frac{1}{2}({x_1}^2+{x_2}^2) \qquad tq \quad x_1-2x_2 \leqslant -2$$

Correction :

$$f_0(x_1,x_2)=\frac{1}{2}(x_1^2+x_2^2)$$

$$\begin{aligned}\mathcal{L}(x_1,x_2,\lambda)&=f_0(x_1,x_2)+\lambda f_1(x_1,x_2)\\&=\frac{1}{2}\left(x_1^2+x_2^2\right)+\lambda(x_1-2x_2+2)\end{aligned}$$

Pour que:
$\left(x_1^*,x_2^*\right)$ soit optimal, il faut que:

$$\nabla_x\mathcal{L}(x^*,\lambda)=0$$
$$\nabla_x\mathcal{L}(x,\lambda)=0 \Leftrightarrow \begin{cases}\frac{\partial\mathcal{L}}{\partial x_1}=0\\ \frac{\partial\mathcal{L}}{\partial x_2}=0\end{cases}$$
$$\begin{cases}\frac{\partial\mathcal{L}}{\partial x_1}=x_1+\lambda=0\\ \frac{\partial\mathcal{L}}{\partial x_1}=x_2-2\lambda=0\end{cases} \Leftrightarrow \begin{cases}x_1=-\lambda\\ x_2=2\lambda\end{cases}$$

La fonction objective duale est:

$$\begin{aligned}g(\lambda,\nu)&=\underset{x\in\mathbb{R}^n}{inf}\,\mathcal{L}(x,\lambda,\nu)\\ g(\lambda)&=\frac{1}{2}((-\lambda)^2+(2\lambda)^2)+\lambda(-\lambda-4\lambda+2)\\&=\frac{1}{2}(\lambda^2+4\lambda^2)-\lambda^2-4\lambda^2+2\lambda\\&=\frac{5}{2}\lambda^2-5\lambda^2+2\lambda\\&=-\frac{5}{2}\lambda^2+2\lambda\end{aligned}$$

Problème dual p^2

$$\underset{\lambda\geqslant 0}{max} \qquad g(\lambda,\nu)$$

On cherche $\underset{\lambda\geqslant 0}{max}\underbrace{(-\frac{5}{2}\lambda^2+2\lambda)}_{g(\lambda)}$
On cherche

$$\lambda^* \quad tq \quad \begin{cases}\nabla g(\lambda^*) & =0\\ \lambda^* & \geqslant 0\end{cases}$$

2)

$$\begin{aligned}\nabla g(\lambda) \quad &=-5\lambda+2\\ \nabla g(\lambda^*) \quad &=0\\ \Leftrightarrow -5\lambda^*+2 \quad &=0\\ \Leftrightarrow\lambda^* \quad &=\frac{2}{5}\geqslant 0\end{aligned}$$

et $x^*=(x_1^*,x_2^*)=\left(-\frac{2}{5},\frac{4}{5}\right)$

2 … Apres avoir mis sous forme matricielle

$$min_{x\in\mathbb{R}^3} \quad \frac{1}{2}(x_1^2+x_2^2+x_3^2) \qquad tq \quad \begin{matrix}x_1+x_2+2x_3=1\\ x_1+4x_2+2x_3=3\end{matrix}$$

Correction :
//FIXME

Suite Ici (https://hackmd.io/GiWgSaP7RrOeMOnqCgVrag)