Optimisation Non-Linéaire
IENAC 1ère année
Notes de cours

Emmanuel Rachelson

2007

Avertissement : Ces notes de cours ne constituent pas un polycopié de cours à part entière. Elles sont distribuées afin de faciliter les révisions pour le test et de fournir un support en complément du cours. Par ailleurs il se peut qu'il subsiste des fautes de frappe ou de petites erreurs : utilisez-donc ces notes de cours avec un regard critique compte-tenu de ce qui a été vu en cours.
Bonnes révisions !

Chapitre 1

Différentiabilité et Convexité

1.1 Différentiabilité

1.1.1 Fonctions multivariables

Définition (fonction multivariable)**.** *Soient X et Y, 2 espaces vectoriels de dimensions n et m. On prendra $Y = \mathbb{R}^m$*

$$f : \left\{ \begin{array}{ccc} X & \to & Y \\ x & \mapsto & \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix} \end{array} \right.$$
$$\text{avec } \forall i \in [\![1;m]\!] \ : \ f_i : X \to \mathbb{R}$$

1.1.2 Plusieurs notions de différentiabilité

Définition (Dérivée partielle)**.** *Soient $\hat{x} \in X$, une base $(e_1, \dots, e_n)$ de X, $i \in [\![1;n]\!]$ et $j \in [\![1;m]\!]$. f admet une dérivée partielle $\frac{\partial f_j}{\partial x_i}(\hat{x}) \in \mathbb{R}$ si :*

$$\exists o_{ij} : \mathbb{R} \to \mathbb{R} \ / \left\{ \begin{array}{l} f_j(\hat{x} + te_i) = f_j(\hat{x}) + \frac{\partial f_j}{\partial x_i}(\hat{x}) \cdot t + o_{ij}(t) \\ \lim\limits_{t \to 0^+} \frac{o_{ij}(t)}{t} = 0 \end{array} \right.$$

Définition (Dérivée directionnelle)**.** *Soit $x \in X$ et $v \in X$. f est différentiable en $\hat{x}$ dans la direction v si la limite suivante existe :*

$$\lim_{t \to 0^+} \frac{f(\hat{x} + tv) - f(\hat{x})}{t} = D_v f(\hat{x})$$

Ce qu'on peut réécrire :

$$\exists \left\{ \begin{array}{l} a_v \in Y \\ o_v : \mathbb{R} \to Y \end{array} \right. \ / \left\{ \begin{array}{l} f(\hat{x} + tv) = f(\hat{x}) + a_v \cdot t + o_v(t) \\ \lim\limits_{t \to 0^+} \frac{\|o_v(t)\|}{t} = 0 \end{array} \right.$$

avec $a_v = D_v f(\hat{x})$.

Définition (Dérivée de Gâteaux)**.** *Soit $\hat{x} \in X$. f différentiable au sens de Gâteaux en $\hat{x}$ si f est différentiable dans toutes les directions $v \in X$. Donc si la limite suivante existe pour tout $v \in X$:*

$$\lim_{t \to 0^+} \frac{f(\hat{x} + tv) - f(\hat{x})}{t} = D_G f(\hat{x}, v)$$

Que l'on peut réécrire en :

$\exists f'_G(\hat{x}) : X \to Y / \ \forall v \in X, \forall t \in \mathbb{R}^+, \begin{cases} f(\hat{x}+tv) = f(\hat{x}) + f'_G(\hat{x})(v) \cdot t + o_v(t) \\ \lim\limits_{t \to 0^+} \frac{\|o_v(t)\|}{t} = 0 \end{cases}$ *On assimile l'opérateur de différentiation selon Gâteaux appliqué à f en $(\hat{x}, v)$ et l'image de v par l'application $f'_G(\hat{x})$:* $D_G f(\hat{x}, v) = f'_G(\hat{x})(v)$.

Propriété (Homogénéité de $D_G f(x, \cdot)$)**.** *L'opérateur $D_G f(x, \cdot)$ est homogène,* ie. $D_G f(x, \alpha v) = \alpha D_G f(x, v)$.

Preuve. Soit f différentiable au sens de Gâteaux en $\hat{x}$ et soit $D_G f(\hat{x}, v)$ sa différentielle. Soit $\alpha \in \mathbb{R}$. On a :

$$\begin{aligned} D_G f(\hat{x}, \alpha v) &= \lim_{t \to 0} \frac{f(\hat{x} + t\alpha v) - f(\hat{x})}{t} \\ &= \lim_{t' \to 0} \frac{f(\hat{x} + t'v) - f(\hat{x})}{\frac{t'}{\alpha}} \text{ avec } t' = \alpha t \\ &= \alpha \lim_{t' \to 0} \frac{f(\hat{x} + t'v) - f(\hat{x})}{t'} \\ &= \alpha D_G f(\hat{x}, v) \end{aligned}$$

□

Définition (continûment différentiable)**.** *On dit que f est continûment différentiable en $\hat{x}$ si f différentiable en $\hat{x}$ et $f'_G(\hat{x}) : X \to Y$ continue en $\hat{x}$.*

Définition (Dérivée de Fréchet)**.** *Soit $\hat{x} \in X$. On dit que f est différentiable au sens de Fréchet en $\hat{x}$ s'il existe un opérateur linéaire borné $A(\cdot) = f'_F(\hat{x})(\cdot)$ tel que :*

$$\lim_{v \to 0} \frac{\|f(\hat{x}+v) - f(\hat{x}) - A(v)\|}{\|v\|} = 0$$

Ce que l'on peut réécrire du point de vue fonctionnel comme :

$$\exists f'_F(\hat{x}) \in \mathcal{L}(X, Y)\ ^1 / \ \forall v \in X, \begin{cases} f(\hat{x}+v) = f(\hat{x}) + f'_F(\hat{x})(v) + o(v) \\ \lim\limits_{v \to 0} \frac{\|o(v)\|}{\|v\|} = 0 \end{cases}$$

Propriétés.

1. *Si f différentiable au sens de Fréchet en $\hat{x}$, alors f continue en $\hat{x}$.*
2. *Si f différentiable au sens de Fréchet en $\hat{x}$, alors $f'_F(\hat{x}) : \begin{cases} X & \to & Y \\ v & \mapsto & f'_F(\hat{x})(v) = \frac{\partial f}{\partial x}(\hat{x}) \cdot v \end{cases}$ avec $\frac{\partial f}{\partial x}(\hat{x}) = \left[\frac{\partial f}{\partial x_1}(\hat{x}), \ldots, \frac{\partial f}{\partial x_n}(\hat{x})\right]$, qu'on appelle la matrice Jacobienne de f en x.*
3. *Si f différentiable au sens de Fréchet, alors f différentiable au sens de Gâteaux et $f'_G(\hat{x}) = f'_F(\hat{x})$.*

Preuve.

1. Soient u et v deux éléments de X,

$$\begin{aligned} \|f(u) - f(v)\| &= \|f(u) - f(u + v - u)\| \text{ or } f \text{ est différentiable au sens de Fréchet donc :} \\ &= \|f'_F(u)(v-u) + o(v-u)\| \text{ et donc :} \\ &\leq \|f'_F(u)(v-u)\| + \|o(v-u)\| \text{ or } f'_F(\hat{u})(\cdot) \text{ est un opérateur linéaire borné donc :} \\ &\leq \|f'_F(u)\|\|u-v\| + \|o(v-u)\| \end{aligned}$$

On a $\lim\limits_{\|v-u\| \to 0} \frac{\|o(v-u)\|}{\|v-u\|} = 0$ donc :
$\forall m \in \mathbb{R}, \forall (v,u) \in X^2 / \|v-u\| \leq m, \exists M \in \mathbb{R}^{+*} / \frac{\|o(v-u)\|}{\|v-u\|} \leq M$
donc $\forall (v,u) \in X^2 / \|v-u\| \leq m, \|f(u) - f(v)\| \leq (\|f'_F(u)\| + M)\, \|u-v\|$ Ce qui assure la continuité de f.

2. Rappel : les colonnes de la matrice d'une application linéaire sont les coordonnées dans la base d'arrivée des images des vecteurs de la base de départ. Soient $i \in [\![1; n]\!]$ et $j \in [\![1; m]\!]$, soit $\hat{x} \in X$. f est différentiable au sens de Fréchet donc, en prenant $v = t \cdot e_i$:

1. $\mathcal{L}(X, Y)$ désigne l'ensemble des applications linéaires de X dans Y.

$f(\hat{x}+te_i) = f(\hat{x}) + f'_F(\hat{x})(te_i) + o(te_i)$

donc : $< f(\hat{x}+te_i)|e_j > = < f(\hat{x})|e_j > + t < f'_F(\hat{x})(e_i)|e_j > + < o(te_i)|e_j >$

et donc : $f_j(\hat{x}+te_i) = f_j(\hat{x}) + t\left(f'_F(\hat{x})(e_i)\right)^T \cdot e_j + \left(o(te_i)\right)^T \cdot e_j$

On a $\frac{o(te_i)^T \cdot e_j}{\|te_i\|} \leq \frac{\|o(te_i)\|}{t}$, donc en posant :

$o' : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ t & \mapsto & \|o(te_i)\| \end{array} \right.$, on a :

$f_j(\hat{x}+te_i) = f_j(\hat{x}) + t\left(f'_F(\hat{x})(e_i)\right)^T \cdot e_j + o'(t)$ avec $\lim\limits_{t\to 0^+} \frac{o'(t)}{t} = 0$

Ce qu'on vient d'écrire est la définition de la dérivée partielle donc :

$$\left(f'_F(\hat{x})(e_i)\right)^T \cdot e_j = \frac{\partial f_j}{\partial x_i}(\hat{x})$$

En d'autres termes, la coordonnée sur e_j de l'image par $f'_F(\hat{x})(\cdot)$ de e_i est $\frac{\partial f_j}{\partial x_i}(\hat{x})$, donc la matrice de l'application $f'_F(\hat{x})(\cdot)$ de X dans Y est la matrice des $\frac{\partial f_j}{\partial x_i}(\hat{x})$. Donc la Jacobienne est bien la matrice de l'application $f'_F(\hat{x})(\cdot)$.

3. Soit f différentiable au sens de Fréchet, $(x,v) \in X^2$ et $t \in \mathbb{R}$.

$$\begin{aligned} f(\hat{x}+tv) &= f(\hat{x}) + f'_F(\hat{x})(tv) + o(\|tv\|) \\ &= f(\hat{x}) + tf'_F(\hat{x})(v) + o_v(t) \end{aligned}$$

On a $\lim\limits_{t\to 0^+} \frac{\|o_v(t)\|}{t} = \lim\limits_{tv\to 0} \frac{\|o(tv)\|}{\|tv\|} = 0$.
Donc f est différentiable au sens de Gâteaux et :

$$\forall(\hat{x},v) \in X^2, f'_F(\hat{x})(v) = f'_G(\hat{x})(v)$$

□

1.1.3 Dérivée seconde

Définition (Dérivée seconde)**.** *Si f différentiable au sens de Fréchet et f'_F différentiable au sens de Fréchet en tout $x \in X$ alors on dit que f est deux fois différentiable et on note $f''(x)$ sa dérivée seconde en x.*

Propriétés. *1. $f''(\hat{x}) \in \mathcal{L}(X, \mathcal{L}(X,Y))$*

2. Soit $f : X \to \mathbb{R}$. La matrice $\frac{\partial^2 f}{\partial x^2}(\hat{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\hat{x}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\hat{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\hat{x}) & \dots & \frac{\partial^2 f}{\partial x_n^2}(\hat{x}) \end{bmatrix}$ s'appelle le Hessien de f en $\hat{x}$ et on a : $\forall(w,v) \in X^2, f''(x)(w)(v) = < \frac{\partial^2 f}{\partial x^2}(\hat{x})w, v >$.

3. On peut identifier les applications $f'_F : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathcal{L}(\mathbb{R}^n, \mathbb{R}) \\ \hat{x} & \mapsto & f'_F(\hat{x}) : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ v & \mapsto & \frac{\partial f}{\partial x}(\hat{x})v \end{array} \right. \end{array} \right.$ et le gradient $\frac{\partial f}{\partial x} : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \hat{x} & \mapsto & \left(\frac{\partial f}{\partial x}(\hat{x})\right)^T \end{array} \right.$. De même, on peut identifier $f''(\cdot)(\cdot)(\cdot)$ et $\frac{\partial^2 f}{\partial x^2} : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^{n\times n} \\ \hat{x} & \mapsto & \frac{\partial^2 f}{\partial x^2}(\hat{x}) \end{array} \right.$.

Preuve. 1. Il suffit d'écrire la définition de la différentiabilité : $f'_F(\hat{x}) : X \to \mathcal{L}(X,Y)$ est différentiable au sens de Fréchet donc il existe une application de X dans $\mathcal{L}(X,\mathcal{L}(X,Y))$ qui est sa différentielle de Fréchet et qu'on a nommé f''. Donc $f''(\hat{x}) \in \mathcal{L}(X,\mathcal{L}(X,Y))$.

2. L'idée est la même que pour la Jacobienne.

   Soit $f : X \to \mathbb{R}$ deux fois différentiable. On a : $f''(\hat{x}+w) = f'(\hat{x}) + f''(\hat{x})(w) + \omega(w)$ avec $\lim\limits_{\|w\|\to 0} \frac{\|\omega(w)\|}{\|w\|} = 0$.

   Or $f'(\hat{x})(v) = \frac{\partial f}{\partial x}(\hat{x})v$, donc, en utilisant une notation vectorielle :

   $\forall v \in X,\ \dfrac{\partial f}{\partial x}(\hat{x}+w)\cdot v = \dfrac{\partial f}{\partial x}(\hat{x})\cdot v + f''(\hat{x})(w)\cdot v + \omega(w)\cdot v$ ou encore :

   $\dfrac{\partial f}{\partial x}(\hat{x}+w) = \dfrac{\partial f}{\partial x}(\hat{x}) + f''(\hat{x})(w) + \omega(w)$

   $\forall i \in [\![1;n]\!],\ \dfrac{\partial f}{\partial x}(\hat{x}+w)\cdot e_i = \dfrac{\partial f}{\partial x}(\hat{x})\cdot e_i + f''(\hat{x})(w)\cdot e_i + \omega(w)\cdot e_i$

   Or $\dfrac{\partial f}{\partial x}(\hat{x}+w)\cdot e_i$ est la $i^{\text{ème}}$ composante du vecteur $\dfrac{\partial f}{\partial x}(\hat{x}+w)$ donc :

   $\dfrac{\partial f}{\partial x_i}(\hat{x}+w) = \dfrac{\partial f}{\partial x_i}(\hat{x}) + f''(\hat{x})(w)\cdot e_i + \omega(w)\cdot e_i$

   On prend $w = te_j$ avec $j \in [\![1;n]\!]$,

   $\dfrac{\partial f}{\partial x_i}(\hat{x}+te_j) = \dfrac{\partial f}{\partial x_i}(\hat{x}) + f''(\hat{x})(te_j)\cdot e_i + \omega(te_j)\cdot e_i$

   On reconnait l'écriture de la dérivée partielle de $\dfrac{\partial f}{\partial x_i}$ par rapport à la $j^{\text{ème}}$ variable :

   $f''(\hat{x})(e_j)\cdot e_i = \dfrac{\partial \frac{\partial f}{\partial x_i}}{\partial x_j}(\hat{x}) = \dfrac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x})$

   Donc $f''(\hat{x})(e_j)$ correspond au vecteur des $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x})$ pour i allant de 1 à n. Et donc la matrice de l'application bilinéaire $f''(\hat{x})$ est l'application dont les colonnes sont les vecteurs $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\hat{x})$ pour i allant de 1 à n. La matrice de l'application $f''(\hat{x})$ est donc bien le Hessien.

3. C'est l'identification classique d'une application linéaire à son coefficient directeur ou d'une application bilinéaire à sa matrice. Ainsi, quand on cherche la dérivée seconde de $f : X \to \mathbb{R}$, on peut chercher directement le Hessien et non l'application linéaire de X dans $\mathcal{L}(X,\mathcal{L}(X,Y))$. Pour s'en convaincre, il suffit d'écrire, par exemple, la dérivée de Fréchet de f'_F :
   On pose :

$$\frac{\partial f}{\partial x} : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^n \\ \hat{x} & \mapsto & \left(\dfrac{\partial f}{\partial x}(\hat{x})\right)^T \end{array} \right.$$

   On a :

$$\frac{\partial f}{\partial x}(x+u) = \frac{\partial f}{\partial x}(x) + \left(\frac{\partial f}{\partial x}\right)'(x)(u) + o_\alpha(u) \quad \text{avec : } \lim_{u\to 0} \frac{\|o_\alpha(u)\|}{\|u\|} = 0$$

Donc :

$$\begin{aligned}
f'_F(x+u) &= \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ v & \mapsto & \frac{\partial f}{\partial x}(x+u)v \end{array} \right. \\
&= \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ v & \mapsto & \frac{\partial f}{\partial x}(x)v + \left(\frac{\partial f}{\partial x}\right)'(x)(u)v + o_\alpha(u)v \end{array} \right. \\
&= \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ v & \mapsto & \frac{\partial f}{\partial x}(x)v \end{array} \right. + \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ v & \mapsto & \left(\frac{\partial f}{\partial x}\right)'(x)(u)v \end{array} \right. + \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ v & \mapsto & o_\alpha(u)v \end{array} \right. \\
&= f'_F(x) + f''(x)(u) + o_\beta(u)
\end{aligned}$$

et on a $\frac{\|o_\beta(u)\|}{\|u\|} = \frac{\sup_{\|v\|\leq 1} \|o_\alpha(u)v\|}{\|u\|} \leq \frac{\|o_\alpha(u)\|}{\|u\|}$. Le dernier terme tend vers 0 quand u tend vers 0 donc le premier aussi.

Donc dériver l'opérateur $f'_F(x)$ revient à dériver la fonction $\frac{\partial f}{\partial x}(x)^T$ (attention à la transposition pour le sens d'écriture des vecteurs : le gradient est un vecteur colonne, c'est le vecteur transposé de la Jacobienne).

□

Théorème (Théorème de Schwartz)**.** *(admis) Le Hessien est une matrice symétrique.*

Propriété (Formule de Taylor)**.** *Soit $f : X \to \mathbb{R}$ deux fois différentiable :*

$$\begin{aligned}
f(\hat{x}+v) &= f(\hat{x}) + f'(\hat{x})(v) + \frac{1}{2}f''(\hat{x})(v)(v) + o_2(v) \\
&= f(\hat{x}) + \frac{\partial f}{\partial x}(\hat{x})(v) + \frac{1}{2} < \frac{\partial^2 f}{\partial x^2}(\hat{x})v, v > + o_2(v)
\end{aligned} \qquad \textit{avec } \lim_{v\to 0} \frac{\|o_2(v)\|}{\|v\|^2} = 0$$

Preuve. On utilise la formule de Taylor avec reste intégral pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$.

Soit $F : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ t & \mapsto & f(\hat{x}+tv) \end{array} \right.$.

On a

$$\begin{aligned}
F(1) &= F(0) + F'(0) + \int_0^1 (1-t)F''(t)dt \\
f(\hat{x}+v) &= f(\hat{x}) + \frac{\partial f}{\partial x}(\hat{x}) \cdot v + \int_0^1 (1-t)v^T \frac{\partial^2 f}{\partial x^2}(\hat{x}+tv)vdt \\
&= f(\hat{x}) + f'(\hat{x}) \cdot v + \int_0^1 (1-t)v^T f''(\hat{x})vdt + \int_0^1 (1-t)v^T \left[f''(\hat{x}+tv) - f''(\hat{x})\right] vdt
\end{aligned}$$

On pose : $o_2(v) = \int_0^1 (1-t)v^T \left[f''(\hat{x}+tv) - f''(\hat{x})\right] vdt$

On a donc : $f(\hat{x}+v) = f(\hat{x}) + f'(\hat{x})(v) + \frac{1}{2}f''(\hat{x})(v)(v) + o_2(v)$

On a : $\left|\int_0^1 (1-t)v^T \left[f''(\hat{x}+tv) - f''(\hat{x})\right] vdt\right| \leq \int_0^1 (1-t) \|f''(\hat{x}+tv) - f''(\hat{x})\| \|v\|^2 dt.$

Donc : $\dfrac{\left|\int_0^1 (1-t)v^T \left[f''(\hat{x}+tv) - f''(\hat{x})\right] vdt\right|}{\|v\|^2} \leq \int_0^1 (1-t) \|f''(\hat{x}+tv) - f''(\hat{x})\| dt.$

Le terme de droite tend vers zéro quand v tend vers 0. Donc $\lim_{\|v\|\to 0} \frac{\|o_2(v)\|}{\|v\|^2} = 0$. Ce qui achève la démonstration.

□

1.2 Convexité

1.2.1 Définitions

Définition (Ensemble convexe)**.** *$S \subset X$ convexe* $\Leftrightarrow\ \forall (x,y) \in S^2,\ \forall \lambda \in [0,1],\ \lambda x + (1-\lambda)y \in S$

Propriété. *$S \subset X$ convexe* $\Leftrightarrow\ \forall k \in \mathbb{N}^*,\ \forall (x^1,\dots,x^k) \in S^k,\ \forall (\lambda^1,\dots,\lambda^k) \in {\mathbb{R}^+}^k$ *tels que* $\sum\limits_{i=1}^{k} \lambda^i = 1,\ \sum\limits_{i=1}^{k} \lambda^i x^i \in S$

Preuve. CN Supposons S convexe. On montre l'implication par récurrence sur k.

On constate immédiatement que la propriété est vraie pour $k=1$.

Soit $k \in \mathbb{N}$. Supposons la propriété vraie pour $k-1$.

Soient $(x^1,\dots,x^k) \in S^k$ et $(\lambda^1,\dots,\lambda^k) \in {\mathbb{R}^+}^k$ tels que $\sum\limits_{i=1}^{k} \lambda^i = 1$.

Comme $\sum\limits_{i=1}^{k} \lambda^i = 1$ il existe au moins un λ^i différent de 1. On suppose que $\lambda^1 \neq 1$.

$\sum\limits_{i=2}^{k} \frac{\lambda^i}{1-\lambda^1} x^i \in S$ grace à la propriété de récurrence et $x^1 \in S$.

On a $\lambda^i \leq 1$ car tous les λ^i sont positifs et leur somme vaut 1. Donc $0 \leq \lambda^1 \leq 1$ et donc, parce que S est un ensemble convexe : $\lambda^1 x^1 + (1-\lambda^1)\left(\sum\limits_{i=2}^{k} \frac{\lambda^i}{1-\lambda^1} x^i\right) \in S$.

Donc finalement $\sum\limits_{i=1}^{k} \lambda^i x^i \in S$.

CS Il suffit de prendre $k=2$ pour avoir l'implication inverse.

□

Définition (Fonction coercive)**.** *Une fonction f est dite* coercive *si elle vérifie* $\lim\limits_{|x| \to \infty} f(x) = +\infty$

Définition (Fonction convexe)**.** *$f : X \to \mathbb{R}$ convexe sur un sous ensemble convexe S de X si :*

$$\forall (x,y) \in S^2, \forall \lambda \in [0;1],\ f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

f est dite strictement convexe si l'inégalité est stricte.

Définition (Enveloppe convexe)**.** *Soit S un sous-ensemble de X, l'enveloppe convexe de S, $Co(S)$, est le plus petit ensemble convexe contenant S (plus petit au sens de l'inclusion).*

Propriétés. *1. A convexe contenant $S \Rightarrow Co(S) \subseteq A$*

2. $Co(S) = \bigcup\limits_{k=1}^{\infty} \left\{ \sum\limits_{i=1}^{k} \lambda^i x^i \ / (x^1,\dots,x^k) \in S^k, (\lambda^1,\dots,\lambda^k) \in {\mathbb{R}^+}^k, \sum\limits_{i=1}^{k} \lambda^i = 1 \right\}$

Preuve. 1. C'est la traduction directe de la définition de $Co(S)$.

2. On montre d'abord l'inclusion de $Co(S)$ dans $\bigcup\limits_{k=1}^{\infty} \{\}$. Pour cela on prouve que $\bigcup\limits_{k=1}^{\infty} \{\}$ est convexe et contient S.

$S = \left\{ \sum\limits_{i=1}^{k} \lambda^i x^i \ / (x^1,\dots,x^k) \in S^k, (\lambda^1,\dots,\lambda^k) \in {\mathbb{R}^+}^k, \sum\limits_{i=1}^{k} \lambda^i = 1 \right\}$ pour $k=1$.

Donc $S \subset \bigcup\limits_{k=1}^{\infty} \{\}$.

Soient a et b appartenant à $\bigcup\limits_{k=1}^{\infty} \{\}$.

$\exists k_a \in \mathbb{N}^*,\ \exists(\lambda_a^1, \ldots, \lambda_a^{k_a}) \in \mathbb{R}^{+^{k_a}}$ tels que $\sum\limits_{i=1}^{k_a} \lambda_a^i = 1, \exists(x_a^1, \ldots, x_a^{k_a}) \in S^{k_a}$ tels que :

$$a = \sum_{i=1}^{k_a} \lambda_a^i x_a^i$$

De même, $b = \sum\limits_{j=1}^{k_b} \lambda_b^j x_b^j$.

Soit $\lambda \in \mathbb{R}$. On a :

$$\begin{aligned}\lambda a + (1-\lambda)b &= \sum_{i=1}^{k_a} \lambda\lambda_a^i x_a^i + \sum_{j=1}^{k_b}(1-\lambda)\lambda_b^j x_b^j \\ &= \sum_{l=1}^{k_a+k_b} \tilde{\lambda}^l \tilde{x}^l\end{aligned}$$

Avec $\tilde{\lambda}^l = \begin{cases} \lambda\lambda_a^l & \text{si } l \in [\![1; k_a]\!] \\ (1-\lambda)\lambda_b^{l-k_a} & \text{si } l \in [\![k_a+1; k_a+k_b]\!] \end{cases}$,
et $\tilde{x}^l = \begin{cases} x_a^l & \text{si } l \in [\![1; k_a]\!] \\ x_b^{l-k_a} & \text{si } l \in [\![k_a+1; k_a+k_b]\!] \end{cases}$.

Les λ^l sont dans $\mathbb{R}^+$ et les x^l sont dans S donc :
$\lambda a + (1-\lambda)b \in \left\{ \sum\limits_{i=1}^{k} \lambda^i x^i \ / (x^1, \ldots, x^k) \in S^k, (\lambda^1, \ldots, \lambda^k) \in \mathbb{R}^{+^k}, \sum\limits_{i=1}^{k} \lambda^i = 1 \right\}$ pour $k = k_a + k_b$.
Donc $\lambda a + (1-\lambda)b \in \bigcup\limits_{k=1}^{\infty} \{\}$. Et donc $\bigcup\limits_{k=1}^{\infty} \{\}$ est convexe. Au final : $Co(S) \subset \bigcup\limits_{k=1}^{\infty} \{\}$.

Inversement, on montre que $\bigcup\limits_{k=1}^{\infty} \{\} \subset Co(S)$.

Soit $a \in \bigcup\limits_{k=1}^{\infty} \{\}$. On a $a = \sum\limits_{i=1}^{k} \lambda^i x^i$.

Les x^i sont dans S donc dans $Co(S)$ or $Co(S)$ est convexe donc $a \in Co(S)$. Donc $\bigcup\limits_{k=1}^{\infty} \{\}$. $a = \sum\limits_{i=1}^{k} \lambda^i x^i \subset Co(S)$. Ce qui achève la démonstration.

□

1.2.2 Convexité et Différentiabilité

Théorème (Convexité et dérivée première). *Une application $f : \mathbb{R}^n \to \mathbb{R}$ continûment différentiable est convexe sur un sous ensemble convexe Ω de $\mathbb{R}^n$ ssi :* $\forall(\hat{x}, \hat{y}) \in \Omega^2,\ f(\hat{y}) \geq f(\hat{x}) + \frac{\partial f}{\partial x}(\hat{x})(\hat{y} - \hat{x})$.
"strictement convexe", si l'inégalité est stricte.

Preuve. CN Soit f convexe. On a :

$$\begin{aligned}f\left(x + \lambda\left(y - x\right)\right) &= f\left(\left(1-\lambda\right)x + \lambda y\right) \\ &\leq \left(1-\lambda\right)f(x) + \lambda f(y) \\ &\leq f(x) + \lambda\left(f(x) - f(y)\right)\end{aligned}$$

Donc : $f(y) - f(x) \geq \dfrac{f(x + \lambda(y-x)) - f(x)}{\lambda}$.
Or $f(x + \lambda(y-x)) = f(x) + \dfrac{\partial f}{\partial x}(x)(\lambda(y-x)) + o(\lambda(y-x))$ avec $\lim\limits_{\|v\|\to 0} \frac{\|o(v)\|}{\|v\|} = 0$.

Donc $\dfrac{f(x+\lambda(y-x))-f(x)}{\lambda} = \dfrac{\partial f}{\partial x}(x)(y-x) + \dfrac{o(\lambda(y-x))}{\lambda}$.

Donc $f(y)-f(x) \geq \dfrac{\partial f}{\partial x}(x)(y-x) + \dfrac{o(\lambda(y-x))}{\lambda}$.

Le terme de gauche ne dépend pas de λ donc l'inégalité reste vraie quel que soit λ, en particulier à la limite, quand on fait tendre λ vers zéro. Or $\lim\limits_{\lambda\to 0} \dfrac{o(\lambda(y-x))}{\lambda} = 0$. On conclut donc que :

$$\text{f convexe} \;\Rightarrow\; f(y) \geq \frac{\partial f}{\partial x}(x)(y-x) + f(x)$$

Le raisonnement est le même dans le cas d'une fonction strictement convexe, l'inégalité est alors stricte.

CS Supposons que $\forall (x,y) \in {\mathbb{R}^n}^2$. Soient $(x,y) \in {\mathbb{R}^n}^2$ et soit $\lambda \in [0,1]$. On pose $z = \lambda x + (1-\lambda)x$ et on a :

$$\begin{aligned} f(y) &\geq f(z) + \frac{\partial f}{\partial x}(z)(y-z) \qquad && \times (1-\lambda) \\ f(x) &\geq f(z) + \frac{\partial f}{\partial x}(z)(x-z) \qquad && \times \lambda \end{aligned}$$

Donc : $(1-\lambda)f(y) + \lambda f(x) \geq f(z) + \frac{\partial f}{\partial x}(z)\left((1-\lambda)\,y + \lambda x - z\right)$.

Or $(1-\lambda)\,y + \lambda x - z = 0$.

Donc $(1-\lambda)f(y) + \lambda f(x) \geq f(\lambda x + (1-\lambda)x)$. Donc f convexe. Ce qui prouve l'équivalence (le raisonnement est le même dans le cas strict).

□

Théorème (Convexité et dérivée seconde)**.** *Soit $f : X \to \mathbb{R}$ deux fois continûment différentiable.*

1. *f convexe sur X $\Leftrightarrow$ $\forall \hat{x} \in X,\ \frac{\partial^2 f}{\partial x^2}(\hat{x}) \geq 0$*
2. *f strictement convexe sur X $\Leftrightarrow$ $\forall \hat{x} \in X,\ \frac{\partial^2 f}{\partial x^2}(\hat{x}) > 0$*
3. *f est strictement convexe dans un voisinage de $\hat{x} \in X$ $\Leftrightarrow$ $\frac{\partial^2 f}{\partial x^2}(\hat{x}) > 0$*

Preuve. Le troisième cas est la version locale du second cas. Comme précédemment, les raisonnements pour le premier et le second cas sont similaires, les inégalités strictes induisant la convexité stricte.

Prouvons l'implication pour le premier cas.

Soit f convexe sur X. D'après la formule de Taylor, on a :

$$\forall (x,y) \in X^2,\ f(y) = f(x) + \frac{\partial f}{\partial x}(x)(y-x) + \frac{\partial^2 f}{\partial x^2}(x)(y-x)(y-x) + o_2(y-x)$$
$$\text{avec : } \lim_{\|v\|\to 0} \frac{\|o_2(v)\|}{\|v\|^2} = 0$$

Or d'après le théorème précédent $f(y) \geq f(x) + \dfrac{\partial f}{\partial x}(x)(y-x)$, donc :

$$\forall (x,y) \in X^2,\ \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x)(y-x)(y-x) + o_2(y-x) \geq 0$$

Soient x et v deux éléments de X et soit $\lambda \in \mathbb{R}^{+*}$. L'inégalité précédente étant vraie pour tous (x,y), on l'applique pour $(x, x+\lambda v)$. On a donc :

$$\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x)(\lambda v)(\lambda v) + o_2(\lambda v) \geq 0$$
$$\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x)(v)(v) + \frac{o_2(\lambda v)}{\lambda^2} \geq 0$$

On a $\lim\limits_{\lambda \to 0^+} \frac{|o_2(\lambda v)|}{\lambda^2} = 0$, donc :

$$\forall (x,v) \in X^2 \frac{\partial^2 f}{\partial x^2}(x)(v)(v) \geq 0$$
$$\text{Donc } \forall x \in X, \frac{\partial^2 f}{\partial x^2} \geq 0.$$

Si f strictement convexe alors l'inégalité est stricte, ce qui signifie que le Hessien est défini positif.

Prouvons à présent l'implication inverse. Supposons que $\forall x \in X,\ \frac{\partial^2 f}{\partial x^2}(\hat{x}) \geq 0$.
Soient x et y deux éléments de X. On pose $g : \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ t & \mapsto & f(x + t(y-x)) \end{array} \right.$.

$$\forall t \in \mathbb{R},\ g'(t) = \frac{\partial f}{\partial x}(x + t(y-x))(y-x)$$
$$\forall t \in \mathbb{R},\ g''(t) = \frac{\partial^2 f}{\partial x^2}(x + t(y-x))(y-x)(y-x)$$

$\frac{\partial^2 f}{\partial x^2}$ positive donc $\forall t \in \mathbb{R},\ g''(t) \geq 0$ et donc g convexe. Et donc :

$\forall \lambda \in [0,1],\ g(\lambda \times 0 + (1-\lambda) \times 1) \leq \lambda g(0) + (1-\lambda) g(1)$

Or $g(1-\lambda) = f(x + (1-\lambda)(y-x)) = f(\lambda x + (1-\lambda)y)$, $g(0) = f(x)$ et $g(1) = f(y)$. Donc :
$$\forall \lambda \in [0;1], f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

Donc f est convexe. On a donc bien l'équivalence (et le raisonnement est le même pour le cas strict). $\square$

Chapitre 2

Optimisation sans contraintes

2.1 Définitions

Définition (Problème d'optimisation). *Optimiser, c'est maximiser une fonction de gain ou minimiser une fonction de coût f. On définit une forme standard de problème à optimiser :*

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x)$$

De façon plus générale, pour un domaine $\Omega \subset \mathbb{R}^n$:

$$(P_\Omega) \quad \min_{x \in \Omega} f(x)$$

Définition (Solution locale, globale).
1. *$\hat{x}$ est solution locale de (P_Ω) s'il existe $\epsilon > 0$ tel que $\forall y \in \mathcal{B}_\epsilon(\hat{x}) \cap \Omega,\ f(y) \geq f(\hat{x})$.*
2. *$\hat{x}$ est solution globale de (P_Ω) si $\forall y \in \Omega,\ f(y) \geq f(\hat{x})$.*
3. *$\hat{x}$ est solution locale stricte de (P_Ω) (respectivement globale stricte) s'il existe $\epsilon > 0$ tel que $\forall y \in \mathcal{B}_\epsilon(\hat{x}) \cap \Omega$(respectivement Ω), $f(y) \geq f(\hat{x})$.*

Théorème (Weierstrass). *Soient K un compact de $\mathbb{R}^n$ et f une fonction continue de K dans $\mathbb{R}$. Il existe $\hat{x} \in K$ tel que $\forall x \in K,\ f(\hat{x}) \leq f(x)$*

2.2 Conditions du premier ordre

Théorème (Condition d'optimalité du premier ordre). *Soit f différentiable et $\hat{x} \in \mathbb{R}^n$.*
1. *$\hat{x}$ solution de $(P) \Rightarrow \frac{\partial f}{\partial x}(\hat{x}) = 0$*
2. *Si f convexe sur $\mathbb{R}^n$, alors : $\hat{x}$ solution de $(P) \Leftrightarrow \frac{\partial f}{\partial x} = 0$*
3. *Si f strictement convexe sur $\mathbb{R}^n$, et $\hat{x}$ solution de (P), alors $\hat{x}$ est la solution globale stricte de (P) sur $\mathbb{R}^n$.*

Preuve. 1. Soit $\hat{x}$ solution locale de (P) :
$\exists \epsilon > 0 \ / \ \hat{x} = \arg \min_{y \in \mathcal{B}_\epsilon(\hat{x})} f(y)$.

Soit $v \in \mathcal{B}_\epsilon(0)$ et $\lambda \in [-1; 1]$, on a : $f(\hat{x} + \lambda v) \geq f(\hat{x})$.

Donc $\lambda \frac{\partial f}{\partial x}(\hat{x})(v) + o(\lambda v) \geq 0$.

Donc, $\forall v \in \mathcal{B}_\epsilon(0) \begin{cases} \forall \lambda \in]0,1],\ \frac{\partial f}{\partial x}(\hat{x})(v) + \frac{o(\lambda v)}{\lambda} \geq 0 & \text{donc } \frac{\partial f}{\partial x}(\hat{x})(v) \geq 0 \\ \forall \lambda \in [-1,0[,\ \frac{\partial f}{\partial x}(\hat{x})(v) + \frac{o(\lambda v)}{\lambda} \leq 0 & \text{donc } \frac{\partial f}{\partial x}(\hat{x})(v) \leq 0 \end{cases}$

Donc $\forall v \in \mathcal{B}_\epsilon(0) \frac{\partial f}{\partial x}(\hat{x})(v) = 0$.

On a donc le résultat sur $\mathcal{B}_\epsilon(0)$, on va l'étendre simplement à tous les vecteurs de $\mathbb{R}^n$ en les ramenant à des vecteurs de $\mathcal{B}_\epsilon(0)$:
Soit $w \in \mathbb{R}^n$:

$$\begin{aligned}
\frac{\partial f}{\partial x}(\hat{x})(w) &= \frac{\partial f}{\partial x}(\hat{x})\left(\frac{\epsilon}{2\|w\|}\frac{2\|w\|}{\epsilon}w\right) \\
&= \frac{2\|w\|}{\epsilon}\frac{\partial f}{\partial x}(\hat{x})\left(\frac{\epsilon}{2\|w\|}w\right) \\
&= \frac{2\|w\|}{\epsilon}\cdot 0 \qquad \text{car } \left\|\frac{\epsilon w}{2\|w\|}\right\| = \frac{\epsilon}{2} < \epsilon \\
&= 0
\end{aligned}$$

Donc $\frac{\partial f}{\partial x}(\hat{x}) = 0$.

2. Il faut montrer l'implication inverse dans le cas où f est convexe.
Si f est convexe et $\hat{x}$ est tel que $\frac{\partial f}{\partial x}(\hat{x}) = 0$, alors :
$\forall y \in \mathbb{R}^n,\ f(y) \geq f(x) + \frac{\partial f}{\partial x}(\hat{x})(y - x)$.
Donc $f(y) \geq f(x)$, donc $\hat{x}$ est solution de (P).
3. La preuve est la même que pour le point précédent.

□

2.3 Condition du second ordre

Théorème (Condition du second ordre)**.** *Soit f deux fois différentiable et $\hat{x} \in \mathbb{R}^n$.*

1. *$\hat{x}$ solution de (P)* $\Rightarrow\ \forall v \in \mathbb{R}^n,\ \frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) \geq 0$
2. $\left.\begin{array}{l} \frac{\partial f}{\partial x}(\hat{x}) = 0 \\ \forall v \in \mathbb{R}^n \setminus \{0\}, \frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) > 0 \end{array}\right\} \Rightarrow \hat{x}$ *solution locale stricte de (P)*

Preuve. 1. Soit $\hat{x}$ une solution de (P). On a :
$\exists \epsilon > 0\ /\ \forall v \in \mathcal{B}_\epsilon(0), \forall \lambda \in [-1; 1],\ f(\hat{x} + \lambda v) \geq f(\hat{x})$

Or la formule de Taylor nous indique que :
$f(\hat{x} + \lambda v) = f(\hat{x}) + \lambda \frac{\partial f}{\partial x}(\hat{x})(v) + \frac{1}{2}\lambda^2 \frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) + o_2(\lambda v)$

$\hat{x}$ est solution de (P) donc $\frac{\partial f}{\partial x}(\hat{x})(v) = 0$, donc :
$\forall \lambda \in [-1; 1], \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) + \frac{o_2(\lambda v)}{\lambda^2} \geq 0$

Donc $\frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) \geq 0$.
Comme à la preuve précédente, on passe de $\mathcal{B}_\epsilon(0)$ à $\mathbb{R}^n$:
Soit $w \in \mathbb{R}^n$.

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2}(\hat{x})(w)(w) &= \frac{4\|w\|^2}{\epsilon^2}\frac{\partial^2 f}{\partial x^2}(\hat{x})\left(\frac{\epsilon}{2\|w\|}w\right)\left(\frac{\epsilon}{2\|w\|}w\right) \\
&\geq 0 \qquad \text{car } \left\|\frac{\epsilon w}{2\|w\|}\right\| = \frac{\epsilon}{2} < \epsilon
\end{aligned}$$

Donc $\frac{\partial^2 f}{\partial x^2}(\hat{x}) \geq 0$.

2. Soit $v \in \mathbb{R}^n$. La formule de Taylor s'écrit :
$f(\hat{x} + v) = f(\hat{x}) + \frac{\partial f}{\partial x}(\hat{x})(v) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) + o_2(v)$

Et comme $\frac{\partial f}{\partial x}(\hat{x}) = 0$:
$f(\hat{x} + v) = f(\hat{x}) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) + o_2(v)$

Si on note m la plus petite valeur propre de $\frac{\partial^2 f}{\partial x^2}(\hat{x})$, on a $m > 0$ car $\frac{\partial^2 f}{\partial x^2}(\hat{x})$ est définie positive par hypothèse.

Donc, en notant λ_i les valeurs propres de $\frac{\partial^2 f}{\partial x^2}(\hat{x})$, on a, dans la base diagonale :
$\frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) = \sum\limits_{i=1}^{n} \lambda_i v_i^2$.

Donc $\frac{\partial^2 f}{\partial x^2}(\hat{x})(v)(v) \geq m\|v\|^2$.

Donc $\forall v \in \mathbb{R}^n, f(\hat{x}+v) - f(\hat{x}) \geq \frac{1}{2}m\|v\|^2 + o_2(v)$.

Donc $\frac{f(\hat{x}+v)-f(\hat{x})}{\|v\|^2} \geq \frac{1}{2}m + \frac{o_2(v)}{\|v\|^2}$.

Or, par définition de $o_2(v)$:
$\exists \epsilon > 0 \ / \ \forall v \in \mathcal{B}_\epsilon(0),\ \frac{-1}{4}m < \frac{o_2(v)}{\|v\|^2} < \frac{1}{4}m$

Donc $\forall v \in \mathcal{B}_\epsilon(0), 0 < \frac{1}{4}m < \frac{1}{2}m + \frac{o_2(v)}{\|v\|^2}$.

Et donc : $\forall v \in \mathcal{B}_\epsilon(0), \frac{f(\hat{x}+v)-f(\hat{x})}{\|v\|^2} > 0$.

Ou encore : $\forall v \in \mathcal{B}_\epsilon(0), f(\hat{x}+v) > f(\hat{x})$. Donc $\hat{x}$ est un minimum strict.

$\square$

2.4 Cas des ensembles ouverts

Théorème. *Soit Ω un ouvert de $\mathbb{R}^n$: $\hat{x}$ solution de (P_Ω)* $\Leftrightarrow \left\{ \begin{array}{c} \hat{x} \in \Omega \\ \hat{x} \textit{ solution de } (P) \end{array} \right.$

Preuve.

CN Soit $\hat{x}$ une solution de (P_Ω).
$\exists \epsilon > 0 / \forall y \in \mathcal{B}_\epsilon(\hat{x}) \cap \Omega, f(y) \geq f(\hat{x})$
Or $\mathcal{B}_\epsilon(\hat{x}) \cap \Omega$ est un ouvert donc :
$\exists \epsilon' > 0 / \mathcal{B}_{\epsilon'}(\hat{x}) \subset \mathcal{B}_\epsilon(\hat{x}) \cap \Omega$.
Donc $\forall y \in \mathcal{B}_{\epsilon'}(\hat{x}), f(y) \geq f(\hat{x})$.
Donc $\hat{x}$ est une solution locale de (P).

CS Soit $\hat{x} \in \Omega$ et $\hat{x}$ solution de (P).
$\exists \epsilon > 0 / \forall y \in \mathcal{B}_\epsilon(\hat{x}), f(y) \geq f(\hat{x})$
Or $\mathcal{B}_\epsilon(\hat{x}) \cap \Omega \neq \emptyset$ puisqu'il contient $\hat{x}$.
Donc : $\forall y \in \mathcal{B}_\epsilon(\hat{x}) \cap \Omega, f(y) \geq f(\hat{x})$.
Donc $\hat{x}$ est solution de (P_Ω).

$\square$

Chapitre 3

Méthodes numériques pour l'optimisation sans contraintes de fonctions différentiables

3.1 Pourquoi des méthodes numériques ?

— Certains problèmes ne sont pas quadratiques, la Jacobienne peut être complexe et difficile à inverser pour trouver des points candidats à l'optimum.

— Certains problèmes sont simplement trop gros en nombre de variables pour envisager un traitement efficace de l'inversion de la Jacobienne.

Méthode alternative : On cherche à construire une suite de points $(x_i)_{i\in\mathbb{N}}$ de X qui converge vers un optimum local de f.

Définition (convergence globale)**.** *On dit qu'un algorithme converge globalement si* $\forall x_0 \in X$, $\lim\limits_{i\to\infty} x_i$ *existe.*

Définition (convergence locale)**.** *On dit qu'un algorithme converge localement si pour tout optimum local* $\hat{x}$ *il existe* $\epsilon > 0$ *tel que* $\forall x_0 \in \mathcal{B}_\epsilon(\hat{x})$, $\lim\limits_{i\to\infty} x_i = \hat{x}$.

Idée générale des méthodes vues dans les sections suivantes :

$$x_{k+1} = x_k + \alpha_k d_k \text{ avec } \left\{ \begin{array}{c} \alpha_k > 0 \\ \|d_k\| = 1 \end{array} \right.$$

On se déplace dans X selon les directions successives d_k, avec un pas α_k.

On va voir deux types de méthodes :

1. "Direction de descente" : On prend d'abord une direction de descente d_k telle que $\nabla f(x_k) \cdot d_k < 0$ puis on minimise la fonction d'une variable $f(x_k + td_k)$ pour trouver le pas t_k et le point x_{k+1}.
2. "Points stationnaires", "méthode de Newton" : On cherche directement les points x_k qui annulent le gradient : $\nabla f(x)$.

3.2 Méthode de plus forte pente

Définition (Direction de plus forte pente)**.** *Soit* $a \in X/\nabla f(a) \neq 0$. *Une direction de plus forte pente est un vecteur* $d \in X$ *de norme fixée qui minimise* $\nabla f(a).d$.

Propriété. *Le vecteur* $d = -\nabla f(a)$ *est une direction de plus forte pente de norme* $\|\nabla f(a)\|$.

Algorithme (Méthode de plus forte pente)**.**

$$\left| \begin{array}{l} x_0 \textit{ donné} \\ \texttt{Tant que } \nabla f(x_k) \neq 0 \texttt{ faire :} \\ \quad d_k = -\nabla f(x_k) \\ \quad t_k = \underset{t \in \mathbb{R}}{argmin}\, f(x_k + t d_k) \\ \quad x_{k+1} = x_k + t_k d_k \end{array} \right.$$

Théorème (Convergence de la méthode de plus forte pente)**.**

$$\left.\begin{array}{l} f \in C^1(\mathbb{R}^n, \mathbb{R}) \\ \lim\limits_{\|x\| \to \infty} f(x) = \infty \textit{ (coercivité)} \\ t_k \;/\; \forall t \in \mathbb{R} f(x_k + t_k d_k) \leq f(x_k + t d_k) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} x_k \textit{ bornée} \\ \quad \textit{(donc admet des points d'accumulation)} \\ \textit{les points d'accumulation sont des points} \\ \quad \textit{stationnaires de } f \end{array}\right.$$

Si en plus, f est strictement convexe, alors x_k converge vers l'unique minimum global sur $\mathbb{R}^n$.

Propriété (Orthogonalité des d_k)**.** $d_{k+1} \perp d_k$

Preuve. En posant $g(t) = f\left(x_k - t\frac{\partial f}{\partial x}(x_k)\right)$.
$f(x_{k+1}) = g(t_k)$
$\frac{\partial g}{\partial t}(t) = -\frac{\partial f}{\partial x}\left(x_k - t\frac{\partial f}{\partial x}(x_k)\right) \cdot \left[\frac{\partial f}{\partial x}(x_k)\right]^T$
Or t_k correspond au minimum de g donc $\frac{\partial g}{\partial t}(t_k) = 0$. donc : $\frac{\partial f}{\partial x}(x_{k+1}) \cdot \left[\frac{\partial f}{\partial x}(x_k)\right]^T = 0$
Ou encore $d_{k+1} \perp d_k$. □

Propriété (vitesse de convergence)**.** *Si on prend une fonction quadratique $f(x) = \frac{1}{2}x^TAx - b^Tx$ et on suppose $A \geq 0$. On applique la méthode de plus forte pente et on pose : $E_k = (x_k - \hat{x})^T A(x_k - \hat{x})$. On a alors, en notant α et β la plus petite et la plus grande valeur propre de A :*

$$E_k \leq E_0 \left(\frac{\beta - \alpha}{\beta + \alpha}\right)^{2k}$$

Remarque. *Ainsi, plus la matrice A est mal conditionnée (plus on a une fonction en forme de vallée encaissée)(plus β est grand devant α), plus la convergence est mal bornée en théorie (on verra en BE que l'on peut trouver une borne plus intéressante).*
La propriété précédente met en évidence le fait que la direction $-\nabla f$ n'est pas nécessairement la plus intéressante, Cf. le cas des longues vallées encaissées (exemple du Puy de Dôme vs. Grand Canyon).

3.3 Méthode des gradients conjugués

3.3.1 Cas des fonctions quadratiques (Hestenes-Stiefel, 1952)

On pose $f(x) = \frac{1}{2}x^TAx - b^Tx$.

L'idée est de reprendre la méthode de plus forte pente mais d'éviter les "mauvaises" directions de descente trouvées précédemment qui faisient "rebondir" le chemin de descente sur les parois de f. Pour les éviter, on cherche des directions de descente d_i mutuellement conjuguées par rapport à A pour prendre en compte la géométrie de la fonction f. A chaque itération, on trouve la nouvelle direction de descente par combinaison de $\frac{\partial f}{\partial x}(x_k)$ et des d_i précédents.

Définition (directions conjuguées)**.** *On dit que deux directions de $\mathbb{R}^n$ x et y sont conjuguées par rapport à $A \in \mathbb{R}^{n \times n}$ si $x^TAy = 0$.*

Algorithme (Méthode des gradients conjugués)**.**

$$\left| \begin{array}{l} x_0 \textit{ donné} \\ d_0 = -\nabla f(x_0) \\ \texttt{Itération } k \\ \quad t_k = \frac{-(\nabla f(x_k))^T \nabla f(x_k)}{d_k^T A d_k} \\ \quad x_{k+1} = x_k + t_k d_k \\ \quad d_{k+1} = -\nabla f(x_{k+1}) + \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2} d_k \end{array} \right.$$

Propriétés. *1.* $\forall i \neq j, d_i^T A d_j = 0$ *: les directions de descentes sont toutes mutuellement conjuguées.*

2. L'algorithme des gradients conjugués converge en au plus n itérations.

3.3.2 Cas des fonctions quelconques

Algorithme (Fletcher, Reeves (1964))**.** *L'algorithme est le même, t_k est obtenu en minimisant $f(x_k + td_k)$.*

Algorithme (Polak, Ribière (1969))**.** *On remplace la famille des d_k par celle des :*

$$d_{k+1} = -\nabla f(x_{k+1}) + \frac{(\nabla f(x_{k+1}))^T(\nabla f(x_{k+1}) - \nabla f(x_k))}{\|\nabla f(x_k)\|^2}$$

Remarque. *1. Convergence en un temps constant pour les fonctions quadratiques*

2. Si f est quadratique les deux algorithmes précédents se ramènent au cas général quadratique.

3. Peu d'information à stocker entre deux itérations (x_k, $\nabla f(x_k)$, d_k)

4. Peu de calculs à effectuer (il faut pouvoir évaluer le gradient en x_{k+1}).

5. $\rightarrow$ méthode fréquemment utilisée.

3.4 Méthode de Newton

Plutôt que de chercher des points qui font décroître la valeur de f on cherche des points qui annulent ∇f. On pose $F(x) = \nabla f(x)$. L'idée est d'avoir $F(x) = 0$. On a :
$F(x_k + d) = F(x_k) + DF(x_k) \cdot d + o(d)$
Pour faire converger F vers 0 on prend d tel que $F(x_k) + DF(x_k) \cdot d = 0$.

Algorithme (Méthode de Newton).

> *x_0 donné*
> `Itération` k
> *Si $DF(x_k)$ inversible, $d_k = -(DF(x_k))^{-1} F(x_k)$*
> $x_{k+1} = x_k + d_k$

Théorème (Convergence locale de la méthode de Newton)**.**

Si $\begin{cases} F \in C^1(\mathbb{R}^n, \mathbb{R}) \text{ (donc } f \in C^2(\mathbb{R}^n, \mathbb{R})) \\ \exists \hat{x} \in \mathbb{R}^n \ / \ F(\hat{x}) = 0 \\ DF(\hat{x}) \text{ inversible} \\ DF \text{ L-Lipschitzienne } (\|DF(x) - DF(x')\| \leq \|x - x'\|) \end{cases}$ *, alors :*

$$\exists r > 0 \ / \ \left. \begin{array}{l} \forall x_0 \in \mathcal{B}_r(\hat{x}) \\ x_{k+1} = x_k - (DF(x_k))^{-1} F(x_k) \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} x_k \in \mathcal{B}_r(\hat{x}) \\ \lim\limits_{i \to \infty} x_i = \hat{x} \end{array} \right.$$

Propriété. *La convergence de la méthode de Newton est quadratique :*
$\forall k \in \mathbb{N}^* \|x_{k+1} - \hat{x}\| \leq c\|x_k - \hat{x}\|^2$

Remarque. *1. Pour une fonction f quadratique, la méthode de Newton converge en une unique itération.*

2. La convergence est uniquement locale.

3. Contrairement aux méthodes de descente précédentes on ne sépare pas la recherche de la direction de descente du pas de descente : d est trouvé en un unique calcul. On peut réécrire :

$$x_{k+1} = x_k - \left(\frac{\partial^2 f}{\partial x^2}(x_k)\right)^{-1} \nabla f(x_k)$$

4. Inconvénient de la méthode : il faut pouvoir calculer le Hessien $\frac{\partial^2 f}{\partial x^2}$ en chaque x_k puis inverser le sytème : cela peut être coûteux en ressources de calcul (temps et/ou mémoire).

Méthodes approchées pour les résolutions type Newton : l'idée est d'approcher $\left[\frac{\partial^2 f}{\partial x^2}\right]^{-1}$ par une suite de matrices symétriques définies positives H_k.

Algorithme (Davidson, Fletcher, Power (1959-1963)).

$$\left|\begin{array}{l} x_0 \text{ donné} \\ H_0 \text{ donné} \\ \texttt{Itération } k \\ \quad d_k = -H_k F(x_k) \\ \quad \lambda_k \ \texttt{minimise} \ f(x_k + t d_k) \\ \quad \Delta_k = \lambda_k d_k \\ \qquad x_{k+1} = x_k + \Delta_k \\ \quad \sigma_k = F(x_{k+1}) - F(x_k) \\ \quad H_{k+1} = H_k + \frac{\Delta_k \Delta_k^T}{\Delta_k^T \sigma_k} - \frac{H_k \sigma_k \sigma_k^T H_k}{\sigma_k^T H_k \sigma_k} \end{array}\right.$$

Algorithme (Broyden, Fletcher, Goldfarb, Shannon (1969-1970)). *Comme DFP, mais on prend :*

$$H_{k+1} = H_k + \left(1 + \frac{\sigma_k^T H_k \sigma_k}{\sigma_k^T \Delta_k}\right) \frac{\Delta_k \Delta_k^T}{\Delta_k^T \sigma_k} - \frac{1}{\sigma_k^T \Delta_k} \left(\Delta_k \sigma_k^T H_k + H_k \sigma_k \Delta_k^T\right)$$

Remarque. *1. Il n'y a plus d'inversion lourde à effectuer*

2. Il y a beaucoup d'information à stocker entre deux itérations (il faut stocker tout H_k)

3. Pour être sûrs de converger, il faut réinitialiser régulièrement H_k

Chapitre 4

Optimisation sous contraintes égalité

4.1 Domaine admissible

Définition (Problème d'optimisation contraint). *On écrit les problèmes d'optimisation non-linéaire sous contrainte sous une forme standard :*

$$(P_\Omega) : \left\{ \begin{array}{c} \textit{minimiser } f(x), \\ x \textit{ appartenant au domaine } \Omega \\ \textit{défini par les contraintes :} \\ \forall i \in [\![1, q]\!],\ g_i(x) = 0 \\ \forall j \in [\![1, p]\!],\ h_j(x) \leq 0 \end{array} \right.$$

Définition (Domaine admissible). $\Omega = \{x \in \mathbb{R}^n / \forall i \in [\![1; q]\!], g_i(x) = 0 \textit{ et } \forall j \in [\![1; l]\!], h_j(x) \leq 0\}$

Définition (Direction admissible). $v \in \mathbb{R}^n$ *est une direction admissible en* $\hat{x}$ *si :*

$$\exists \phi \in C^1(\mathbb{R}^+, \mathbb{R}^n) / \left\{ \begin{array}{c} \phi(0) = \hat{x} \\ \phi'(0) = v \\ \phi(t) \in \Omega \textit{ pour } t \textit{ petit} \end{array} \right.$$

Définition (Ensemble des directions admissibles). $\mathcal{D}_{\hat{x}}(\Omega)$.

4.2 Régularité des contraintes égalité

Définition (Régularité des contraintes). *Soit* $\hat{x} \in \Omega$. *Les contraintes* $\{g_i, i \in [\![1; q]\!]\}$ *sont régulières en* $\hat{x}$ *si la famille des* $\left\{\frac{\partial g_i}{\partial x}(\hat{x}), i \in [\![1; q]\!]\right\}$ *est libre, ie. si la Jacobienne de* g *est de rang* q.

Propriété. $\mathcal{D}_{\hat{x}}(\Omega) \subset \ker\left[\frac{\partial g}{\partial x}(\hat{x})\right] = \left\{v \in \mathbb{R}^n / \frac{\partial g}{\partial x}(\hat{x}) \cdot v = 0\right\}$

Preuve. Soit $v \in \mathcal{D}_{\hat{x}}(\Omega)$.

v est une direction admissible donc : $\exists \phi \in C^1(\mathbb{R}^+, \mathbb{R}^n)\ /\ \left\{ \begin{array}{c} \phi(0) = \hat{x} \\ \phi'(0) = v \\ \exists \eta > 0\ /\ \forall t \in [0; \eta[,\ \phi(t) \in \Omega \end{array} \right.$

Donc $\forall t \in [0; \eta[, \forall i \in [\![1; q]\!],\ g_i(\phi(t)) = 0$

En dérivant l'expression précédente pour t dans $[0; \eta[$: $\frac{\partial g_i}{\partial x}(\phi(t)) \cdot \phi'(t) = 0$

Et donc : $\frac{\partial g_i}{\partial x}(\hat{x}) \cdot v = 0$ en prenant $t = 0$.

La ligne précédente est vraie pour tout i dans $[\![1; q]\!]$ donc : $\frac{\partial g}{\partial x}(\hat{x}) \cdot v = 0$.

Et donc $v \in \ker \left[\frac{\partial g}{\partial x}(\hat{x})\right]$.

Donc $\mathcal{D}_{\hat{x}}(\Omega) \subset \ker \left[\frac{\partial g}{\partial x}(\hat{x})\right]$ □

Théorème (Caractérisation des contraintes régulières)**.** *Si les* $\{g_i = 0, i \in [\![1;q]\!]\}$ *sont régulières en* $\hat{x}$*, alors* $\mathcal{D}_{\hat{x}}(\Omega) = \ker \left[\frac{\partial g}{\partial x}(\hat{x})\right]$.

Preuve. Il faut montrer que si les contraintes sont régulières, alors $\ker \left[\frac{\partial g}{\partial x}(\hat{x})\right] \subset \mathcal{D}_{\hat{x}}(\Omega)$ (l'inclusion inverse est toujours vraie comme on l'a vu plus haut). En d'autres termes, pour tout vecteur v de $\ker \left[\frac{\partial g}{\partial x}(\hat{x})\right]$ il faut pouvoir trouver une application s qui permette d'affirmer que v est une direction admissible. Cette application s illustre le fait que l'on peut se "déplacer" dans le domaine dans la direction v sans sortir du domaine. L'idée de cette preuve est d'utiliser le fait que, comme la Jacobienne est de rang q, au voisinage du point considéré, on peut exprimer certaines coordonnées des points du domaine en fonction des autres (on peut en exprimer q en fonction des $n-q$ autres comme on le voit dans le théorème des fonctions implicites). On a donc localement un domaine décrit par un "arc paramétré" en dimension $n-q$: $\psi(u) = (u, \phi(u))$, avec $\phi : \mathbb{R}^{n-q} \to \mathbb{R}^q$ l'équation de l'arc. Si on montre alors que pour tout élément v de $\ker \left[\frac{\partial g}{\partial x}(\hat{x})\right]$ on a un antécédant par $\frac{\partial \psi}{\partial u}$ parmi les u, alors en évoluant sur u le long de cet antécédant, l'arc paramétré restera dans le domaine, on évoluera selon v dans $\mathbb{R}^n$ et on passera par le point $\hat{x}$. Il y a donc deux points clés à cette démonstration : il faut dans un premier temps exprimer le domaine admissible comme un "arc paramétré" de dimension $n-q$ puis montrer que cet arc coïncide bien avec le noyau de la Jacobienne de g.

On peut écrire les contraintes de la façon suivante :

$$g : \left\{ \begin{array}{ccc} \mathbb{R}^{n-q} \times \mathbb{R}^q & \to & \mathbb{R}^q \\ (u,v) & \mapsto & g(u,v) \end{array} \right.$$

On pose $\hat{x} = (u_0, v_0)$.

$g(\hat{x}) = g(u_0, v_0)$ est nulle dans un voisinage de (u_0, v_0) dans Ω puisque $\hat{x} \in \Omega$.
Par ailleurs, la sous-matrice $\frac{\partial g}{\partial v}(u_0, v_0)$ de la Jacobienne de g en (u_0, v_0) est inversible car cette dernière est carrée et de rang q (par définition de la régularité des contraintes).

D'après le théorème des fonctions implicites, il existe un voisinage U de u_0, un voisinage V de v_0 et une application $\phi \in C^1(U, V)$ tels que :

$$\forall (u,v) \in U \times V, \ g(u,v) = 0 \Leftrightarrow v = \phi(u)$$

Si on pose $\psi(u) = (u, \phi(u))$, on a :

$$\left\{ \begin{array}{l} \psi(u_0) = (u_0, \phi(u_0)) = (u_0, v_0) = \hat{x} \\ \forall u \in U, \ \psi(u) \in \Omega \end{array} \right.$$

Si on peut montrer que pour tout $v \in \ker \left[\frac{\partial g}{\partial x}(\hat{x})\right]$, il existe un antécédant z à v par $\frac{\partial \psi}{\partial u}(u_0)$, alors on pourra construire une application $s(t) = \psi(u_0 + tz)$ qui vaut $\hat{x}$ en $t = 0$, donc la dérivée vaut v en $t = 0$ et pour laquelle $s(t)$ est dans Ω pour t petit (tant que $u_0 + tz$ ne sort par de U). Pour montrer qu'un tel antécédant existe toujours, il faut montrer que $\ker \left[\frac{\partial g}{\partial x}(\hat{x})\right] \subset \Im \left[\frac{\partial \psi}{\partial u}(u_0)\right]$.

On a, $\forall u \in U, \ g(\psi(u)) = 0$.

Donc (en dérivant par rapport à u) : $\frac{\partial g}{\partial x}(\psi(u)) \cdot \left(\frac{\partial \psi}{\partial u}(u)\right) = 0$.

Et en appliquant l'équation précédente en u_0 : $\frac{\partial g}{\partial x}(\hat{x}) \cdot \left(\frac{\partial \psi}{\partial u}(u_0)\right) = 0$.

Donc $\Im \left[\frac{\partial \psi}{\partial u}(u_0)\right] \subset \ker \left[\frac{\partial g}{\partial x}(\hat{x})\right]$.

Or $\frac{\partial\psi}{\partial u}(u_0) = \begin{bmatrix} I_{n-q} \\ \frac{\partial\phi}{\partial u}(u_0) \end{bmatrix}$. Cette matrice a n lignes et $n-q$ colonnes et, grace à la sous-matrice I_{n-q}, on a la garantie qu'elle est de rang $n-q$. Donc : $\dim\left(\Im\left[\frac{\partial\psi}{\partial u}(u_0)\right]\right) = n-q$.

Or, d'après le théorème du rang, on a :

$$\dim\left(\ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]\right) + \dim\left(\Im\left[\frac{\partial g}{\partial x}(\hat{x})\right]\right) = n$$

Donc : $\dim\left(\ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]\right) = n-q$.

On a : $\begin{cases} \dim\left(\ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]\right) = \dim\left(\Im\left[\frac{\partial\psi}{\partial u}(u_0)\right]\right) \\ \Im\left[\frac{\partial\psi}{\partial u}(u_0)\right] \subset \ker\left[\frac{\partial g}{\partial x}(\hat{x})\right] \end{cases}$

Donc $\Im\left[\frac{\partial\psi}{\partial u}(u_0)\right] = \ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]$.

On a donc montré qu'à tout élément v de $\ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]$, on pouvait faire correspondre un antécédant par $\frac{\partial\psi}{\partial u}(u_0)$. On déroule alors le raisonnement initial :

Soit $v \in \ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]$.

On a $v \in \Im\left[\frac{\partial\psi}{\partial u}(u_0)\right]$, donc il existe $z \in \mathbb{R}^{n-q}$ tel que $v = \frac{\partial\psi}{\partial u}(u_0)(z)$.

Posons $s(t) = \psi(u_0 + tz)$. On a alors :

$$\begin{cases} s(0) = \psi(u_0) = \hat{x} \\ s'(0) = \frac{\partial\psi}{\partial u}(u_0)(z) = v \\ s(t) \in \Omega \text{ pour } t \text{ petit et tant que } u_0 + tz \text{ reste dans } U. \end{cases}$$

Donc $v \in \mathcal{D}_{\hat{x}}(\Omega)$. □

4.3 Optimisation sous contraintes égalité

4.3.1 Lagrangien

Définition (Lagrangien)**.** *Le Lagrangien est un opérateur :*

$$L : \begin{cases} \mathbb{R}^n \times \mathbb{R}^q & \rightarrow & \mathbb{R} \\ (x, \lambda) & \mapsto & f(x) + \sum_{i=1}^{q} \lambda_i g_i(x) \end{cases}$$

Les λ_i sont appellés paramètres de Lagrange.

4.3.2 Condition du premier ordre

Théorème (CN et CS d'optimalité)**.**

CN Soit $\hat{x} \in \Omega$. Soient f et g_i différentiables en $\hat{x}$ et les g_i régulières en $\hat{x}$:

$$\hat{x} \text{ solution de } (P_\Omega) \Rightarrow \exists \hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_q) \in \mathbb{R}^q / \frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda}) = \frac{\partial f}{\partial x}(\hat{x}) + \sum_{i=1}^{q} \hat{\lambda}_i \frac{\partial g_i}{\partial x}(\hat{x}) = 0$$

Les contraintes étant régulières, les $\hat{\lambda}_i$ sont uniques.

CS Si $\begin{cases} g_i \text{ affines} \\ f \text{ convexe} \end{cases}$, alors la relation précédente est une équivalence.

Preuve.

CN Les contraintes sont qualifiées en $\hat{x}$, donc $\mathcal{D}_{\hat{x}}(\Omega) = \ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]$.

Soit $v \in \ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]$: $\exists \phi \in C^1(\mathbb{R}^+, \mathbb{R}^n)$ tel que $v = \phi'(0)$ et $\hat{x} = \phi(0)$.

Soit $\psi(t) = f(\phi(t))$.

On a : $\psi(0) = f(\hat{x})$ donc 0 est un minimum de ψ. Donc $\psi'(0) = 0$.

Or $\psi'(t) = \frac{\partial f}{\partial x}(\phi(t))(\phi'(t))$.

Donc $\psi'(0) = \frac{\partial f}{\partial x}(\hat{x})v$.

Donc $\left[\frac{\partial f}{\partial x}(\hat{x})\right]^T$ appartient à l'orthogonal à $\ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]$ qu'on note : $\left[\ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]\right]^\perp$.

On a $\left[\ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]\right]^\perp = \left[\Im\left[\frac{\partial g}{\partial x}(\hat{x})\right]\right]^T$.

Donc $\exists(\hat{\lambda}_1, \ldots, \hat{\lambda}_q) \ / \ \frac{\partial f}{\partial x}(\hat{x}) + \sum_{i=1}^{q} \hat{\lambda}_i \frac{\partial g_i}{\partial x}(\hat{x}) = 0$.

Les $\hat{\lambda}_i$ sont uniques car $\frac{\partial g}{\partial x}(\hat{x})$ est de rang q (les contraintes sont régulières).

CS Supposons les g_i affines, f convexe, et prenons $\hat{x}$ et les $\hat{\lambda}_i$ tels que :

$$\frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda}) = 0$$

f est convexe donc on a : $\forall y \in \mathbb{R}^n,\ f(y) \geq f(x) + \frac{\partial f}{\partial x}(\hat{x})(y - \hat{x})$.

Donc $f(y) \geq f(\hat{x}) - \sum_{i=1}^{q} \hat{\lambda}_i \frac{\partial g_i}{\partial x}(\hat{x})(y - x)$.

Or les g_i sont affines : $g_i(x) = a_i^T x + b$ et $g_i(x) = 0$ dans le domaine.

Donc $\frac{\partial g_i}{\partial x}(\hat{x})(y - x) = a_i^T(y - x) = -b + b = 0$.

Donc $\forall y \in \mathbb{R}^n, f(y) \geq f(\hat{x})$.

□

La condition suffisante d'optimalité est très restrictive, pour caractériser l'optimalité dans le cas général il nous faut la condition du second ordre.

4.3.3 Condition du second ordre

Théorème (CN et CS d'optimalité)**.** *Soit $\hat{x} \in \Omega$. Soient f et g_i différentiables en $\hat{x}$ et les g_i régulières en $\hat{x}$:*

CN $\hat{x}$ est solution de $(P_\Omega) \Rightarrow \frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda}) \geq 0$ dans $\mathcal{D}_{\hat{x}}(\Omega)$.

CS Si $\exists \hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_q) \in \mathbb{R}^q / \frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda}) = 0$ et si $\frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda}) \geq 0$ sur $\mathcal{D}_{\hat{x}}(\Omega)$ alors $\hat{x}$ est une solution locale stricte de (P_Ω).

Preuve.

CN On prend $\hat{x}$ solution de (P_Ω) avec la famille des $\left(\hat{\lambda}_i\right)_{i=1\ldots q}$ et $v \in \mathcal{D}_{\hat{x}}(\Omega)$ et on veut montrer que $\frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda}) \geq 0$.

v est une direction admissible donc $\exists \phi \in C^1(\mathbb{R}^+, \mathbb{R}^n) \ / \ v = \phi'(0)$.

On écrit la formule de Taylor pour $(L \circ \phi)$ en 0 :

$L\left(\phi(t), lam\hat{b}da\right) - L\left(\phi(0), lam\hat{b}da\right) = \frac{\partial L}{\partial x}(\hat{x}, lam\hat{b}da)t + \frac{1}{2}\frac{\partial^2 L}{\partial x^2}(\hat{x}, lam\hat{b}da)t^2 + o_2(t)$

Or : $\begin{cases} L\left(\phi(0), lam\hat{b}da\right) = L(\hat{x}, lam\hat{b}da) \\ L(\phi(t), lam\hat{b}da) - L(\hat{x}, lam\hat{b}da) \geq 0 \text{ car } \hat{x} \text{ est un minimum de } L(\cdot, lam\hat{b}da) \\ \frac{\partial L}{\partial x}(\hat{x}, lam\hat{b}da) = 0 \text{ d'après la condition du premier ordre} \end{cases}$

Donc $\forall t \geq 0,\ \frac{1}{2}\frac{\partial^2 L}{\partial x^2}(\hat{x}, lam\hat{b}da)t^2 + o_2(t) \geq 0$

En divisant par t^2 et en faisant tendre t vers 0, on obtient :

$$\frac{\partial^2 L}{\partial x^2}(\hat{x}) \geq 0$$

CS Raisonnons par l'absurde et supposons que $\hat{x}$ n'est pas une solution locale stricte de (P_Ω). Il existe donc une suite de points x_k de $\mathbb{R}^n \cap \Omega$ qui converge vers $\hat{x}$ et dont les images sont plus petites que $f(\hat{x})$. L'idée est de montrer que dans la direction formée par $\lim\limits_{k\to\infty} \frac{x_k - \hat{x}}{\|x_k - \hat{x}\|}$ la fonction n'est pas convexe et donc le Hessien n'est pas positif, ce qui est en contradiction avec les hypothèses.

On a donc une suite x_k avec :
$\forall k \in \mathbb{N}, \exists x_k \in \mathbb{R}^n \cap \Omega \ / \ \|x_k - \hat{x}\| < \frac{1}{k}$ et $f(x_k) \leq f(\hat{x})$

On pose : $v_k = \frac{x_k - \hat{x}}{\|x_k - \hat{x}\|}$.

$(v_k)_{k\in\mathbb{N}}$ est une suite d'éléments du compact $\{x \in \mathbb{R}^n \ / \ \|x\| = 1\}$ donc, d'après le théorème de Bolzano-Weierstrass il existe une sous-suite convergente $v_{\phi(k)}$ de v_k. On note v_∞ la limite de cette suite.

On a donc : $\begin{cases} \lim\limits_{k\to\infty} x_{\phi(k)} = \hat{x} \\ \lim\limits_{k\to\infty} v_{\phi(k)} = v_\infty \end{cases}$

Montrons dans un premier temps que v_∞ est une direction admissible.

$$\begin{aligned} g\left(x_{\phi(k)}\right) &= g\left(\hat{x}\right) + \frac{\partial g}{\partial x}(\hat{x})\left(x_{\phi(k)} - \hat{x}\right) + o\left(x_{\phi(k)} - \hat{x}\right) \\ 0 &= 0 + \frac{\partial g}{\partial x}(\hat{x})\left(\frac{x_{\phi(k)} - \hat{x}}{\|x_{\phi(k)} - \hat{x}\|}\right) + \frac{o\left(x_{\phi(k)} - \hat{x}\right)}{\|x_{\phi(k)} - \hat{x}\|} \text{ car } x_{\phi(k)} \in \Omega \end{aligned}$$

Donc $\frac{\partial g}{\partial x}(\hat{x}) \cdot v_{\phi(k)} + \frac{o\left(x_{\phi(k)} - \hat{x}\right)}{\|x_{\phi(k)} - \hat{x}\|} = 0$.

Or $\lim\limits_{\|u\|\to 0} \frac{o(u)}{u} = 0$.

Donc $\frac{\partial g}{\partial x}(\hat{x}) \cdot v_\infty = 0$. En d'autres termes $v_\infty \in \ker\left[\frac{\partial g}{\partial x}(\hat{x})\right]$. Or les contraintes sont qualifiées en $\hat{x}$ donc $v_\infty \in \mathcal{D}_{\hat{x}}(\Omega)$.

On va maintenant montrer que $\frac{\partial^2 L}{\partial x^2}(\hat{x})(v_\infty)(v_\infty) \leq 0$ pour en tirer une contradiction avec les hypothèses.

On a :

$$f\left(x_{\phi(k)}\right) = f(\hat{x}) + \frac{\partial f}{\partial x}(\hat{x})\left(x_{\phi(k)} - \hat{x}\right) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(\hat{x})\left(x_{\phi(k)} - \hat{x}\right)\left(x_{\phi(k)} - \hat{x}\right) + o_2\left(x_{\phi(k)} - \hat{x}\right) \quad (4.1)$$

Et par ailleurs :

$$g\left(x_{\phi(k)}\right) = g(\hat{x}) + \frac{\partial g}{\partial x}(\hat{x})\left(x_{\phi(k)} - \hat{x}\right) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(\hat{x})\left(x_{\phi(k)} - \hat{x}\right)\left(x_{\phi(k)} - \hat{x}\right) + o_2\left(x_{\phi(k)} - \hat{x}\right)$$

Avec $g\left(x_{\phi(k)}\right) = g(\hat{x}) = 0$, on obtient :

$$0 = \frac{\partial g}{\partial x}(\hat{x})\left(x_{\phi(k)} - \hat{x}\right) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(\hat{x})\left(x_{\phi(k)} - \hat{x}\right)\left(x_{\phi(k)} - \hat{x}\right) + \tilde{o}_2\left(x_{\phi(k)} - \hat{x}\right) \tag{4.2}$$

En additionnant (4.1) et $\hat{\lambda}^T(4.2)$, on obtient :

$$\begin{aligned} f\left(x_{\phi(k)}\right) = f(\hat{x}) + \frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda})\left(x_{\phi(k)} - \hat{x}\right) + \frac{1}{2}\frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda})\left(x_{\phi(k)} - \hat{x}\right)\left(x_{\phi(k)} - \hat{x}\right) \\ + o_2\left(x_{\phi(k)} - \hat{x}\right) + \hat{\lambda}^T\tilde{o}_2\left(x_{\phi(k)} - \hat{x}\right) \end{aligned}$$

Or $\frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda}) = 0$ par hypothèse et $f(x_{\phi(k)}) \leq f(\hat{x})$, donc on a :
$\frac{1}{2}\frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda})\left(x_{\phi(k)} - \hat{x}\right)\left(x_{\phi(k)} - \hat{x}\right) + o_2\left(x_{\phi(k)} - \hat{x}\right) + \hat{\lambda}^T\tilde{o}_2\left(x_{\phi(k)} - \hat{x}\right) \leq 0$

En divisant par $\|x_{\phi(k)} - \hat{x}\|^2$:
$\frac{1}{2}\frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda})(v_k)(v_k) + \frac{o_2\left(x_{\phi(k)} - \hat{x}\right)}{\|x_{\phi(k)} - \hat{x}\|^2} + \hat{\lambda}^T\frac{\tilde{o}_2\left(x_{\phi(k)} - \hat{x}\right)}{\|x_{\phi(k)} - \hat{x}\|^2} \leq 0$

Donc, en faisant tendre k vers l'infini, on a :
$\frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda})(v_\infty)(v_\infty) \leq 0$.

Donc $\frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda})$ n'est pas définie positive sur $\mathcal{D}_{\hat{x}}(\Omega)$. donc l'hypothèse est absurde, $\hat{x}$ est bien un minimum local strict de (P_Ω).

□

Chapitre 5

Optimisation sous contraintes inégalité et égalité - Cas général

5.1 Cône d'admissibilité et qualification des contraintes

Définition (Cône d'admissibilité)**.** *Le cône d'admissibilité en $\hat{x}$ ou cône tangent en $\hat{x}$, est l'enveloppe convexe de l'ensemble des directions admissibles.* $\mathcal{C}_{\hat{x}}(\Omega) = Co(\mathcal{D}_{\hat{x}}(\Omega))$.

Ou encore : $\mathcal{C}_{\hat{x}}(\Omega) = \bigcup\limits_{k=1}^{\infty} \left\{ \sum\limits_{i=1}^{k} \lambda_i v_i, \lambda_i \geq 0, v_i \in \mathcal{D}_{\hat{x}}(\Omega) \right\}$

Définition (Contrainte active)**.** *Une contrainte inégalité h_i est dite active ou saturée en $\hat{x}$ si $h_i(\hat{x}) = 0$. On note $\mathcal{A}(\hat{x})$ l'ensemble des indices des contraintes saturées en $\hat{x}$.*

Définition (Ensemble $G_{\hat{x}}(\Omega)$)**.** $G_{\hat{x}}(\Omega) = \left\{ v \in \mathbb{R}^n / \ \forall i \in \mathcal{A}(\hat{x}), \frac{\partial h_i}{\partial x}(\hat{x}) \cdot v \leq 0 \right\}$

Propriété. $\mathcal{C}_{\hat{x}}(\Omega) \subset G_{\hat{x}}(\Omega)$

Preuve. Soit $v \in \mathcal{D}_{\hat{x}}(\Omega)$.

$$\exists \phi \in C^1(\mathbb{R}, \mathbb{R}^n) \ / \ \begin{cases} \phi(0) = \hat{x} \\ \phi'(0) = v \\ \exists \eta > 0 \ / \ \forall t \in [0; \eta[, \ \phi(t) \in \Omega \end{cases}$$

Donc : $\exists \eta > 0 \ / \ \forall t \in [0; \eta[, \forall i \in [\![1, p]\!], \ h_i(\phi(t)) \leq 0$.

Donc en particulier : $\forall i \in A(\hat{x}), h_i(\phi(t)) \leq 0$.

Donc $\forall i \in A(\hat{x}), h_i(\hat{x} + vt + o(t)) \leq 0$.

Ou encore $h_i(\hat{x}) + \frac{\partial h_i}{\partial x}(\hat{x})(vt + o(t)) + \tilde{o}(vt + o(t))$.

Or $h_i(\hat{x}) = 0$ car $i \in A(\hat{x})$, donc : $\frac{\partial h_i}{\partial x}(\hat{x})\, v + \frac{\partial h_i}{\partial x}\left(\frac{o(t)}{t}\right) + \frac{\tilde{o}(vt+o(t))}{t} \leq 0$

Donc en faisant tendre t vers 0, on obtient : $\forall i \in A(\hat{x}), \ \frac{\partial h_i}{\partial x}(\hat{x})\, v \leq 0$, donc $v \in G_{\hat{x}}$. $\square$

Définition (Qualification des contraintes)**.** *Les contraintes $\{h_i(x) \leq 0, i \in [\![1, p]\!]\}$ sont qualifiées en $\hat{x}$ ssi $\mathcal{C}_{\hat{x}}(\Omega) = G_{\hat{x}}(\Omega)$.*

Théorème (de qualification)**.** *Les h_i sont qualifiées en $\hat{x}$ si l'une des conditions suivantes est vérifiée :*

1. *Pour des contraintes égalité seules, les contraintes sont qualifiées en $\hat{x}$ si elles sont régulières en $\hat{x}$.*
2. *Les contraintes sont qualifiées en $\hat{x}$ si $G_{\hat{x}}^{-} = \left\{ v \in \mathbb{R}^n / \ \forall i \in \mathcal{A}(\hat{x}), \frac{\partial h_i}{\partial x}(\hat{x}) \cdot v < 0 \right\} \neq \emptyset$.*
3. *Si les vecteurs $\left\{ \frac{\partial g_i}{\partial x}(\hat{x}), i \in [\![1; q]\!] \right\} \cup \left\{ \frac{\partial h_i}{\partial x}(\hat{x}), i \in \mathcal{A}(\hat{x}) \right\}$ forment une famille libre alors les contraintes sont qualifiées en $\hat{x}$.*

4. *Si les $\{h_i, i \in \mathcal{A}(\hat{x})\}$ et les $\{g_i\}$ sont affines et distinctes alors les contraintes sont qualifiées en $\hat{x}$.*
5. *Si les h_i sont convexes et $\exists x \in \Omega / \forall i \in [\![1;q]\!],\ h_i(x) < 0$ (intérieur non vide), alors les contraintes sont qualifiées en $\hat{x}$.*

Preuve.

1. Le (a) est une définition qu'on peut voir comme une conséquence du (b), en effet, si on a un problème (P) avec uniquement q contraintes égalité $g_i(x) = 0$ et si ces dernières sont régulières, alors, d'après le (b), pour le problème d'optimisation sous contraintes inégalités (P') ayant $2q$ contraintes inégalité : $g_i(x) \leq 0$ et $-g_i(x) \leq 0$, les contraintes sont qualifiées.
2. On montre tout d'abord que les directions de $G_{\hat{x}}^-$ sont des directions admissibles, puis on utilise ce résultat pour montrer que, quel que soit u dans $G_{\hat{x}}$, u appartient à $\mathcal{C}_{\hat{x}}(\Omega)$.

   Supposons que $G_{\hat{x}}^-$ soit non vide. On sait alors que :

   $\exists v \in \mathbb{R}^n\ /\ \forall i \in A(\hat{x}),\ \frac{\partial h_i}{\partial x}(\hat{x})v < 0$
   Or $\forall i \in A(\hat{x}),\ h_i(\hat{x} + tv) = h_i(x) + t\frac{\partial h_i}{\partial x}(\hat{x})v + o(t)$.

   Donc, pour t petit, on a : $\frac{\partial h_i}{\partial x}(\hat{x})v + \frac{o(t)}{t}$.

   Donc, pour t petit, $h_i(\hat{x} + tv) < 0$. Donc, pour t petit, $\hat{x} + tv \in \Omega$. Donc en prenant une fonction $\phi(t) = \hat{x} + tv$, on montre que v est une direction admissible en $\hat{x}$ de Ω.
   Donc $G_{\hat{x}}^-(\Omega) \subset \mathcal{C}_{\hat{x}}(\Omega)$. On va montrer que cela suffit pour que $G_{\hat{x}} \subset \mathcal{C}_{\hat{x}}(\Omega)$.

   Prenons un élément u de $G_{\hat{x}}$. On va contruire une suite d'éléments de $\mathcal{C}_{\hat{x}}(\Omega)$ (qui est un fermé) qui converge dans $\mathcal{C}_{\hat{x}}(\Omega)$ et qui converge vers u.

   On a : $\forall i \in A(\hat{x}) \begin{cases} \frac{\partial h_i}{\partial x}(\hat{x})v < 0 \\ \frac{\partial h_i}{\partial x}(\hat{x})u \leq 0 \end{cases}$.

   Soit $\lambda \in]0;1]$: $\frac{\partial h_i}{\partial x}(\hat{x})[\lambda v + (1-\lambda)u] < 0$ donc $\lambda v + (1-\lambda)u \in G_{\hat{x}}^-$. Donc, d'après le résultat précédent, $\lambda v + (1-\lambda)u \in \mathcal{C}_{\hat{x}}(\Omega)$.

   Donc $\exists w_\lambda \in \mathcal{C}_{\hat{x}}(\Omega)\ /\ w = \lambda v + (1-\lambda)u$.

   Ce w_λ existe pour tout $\lambda \in]0;1[$ et on a $u = \frac{1}{1-\lambda}w_\lambda - \frac{\lambda}{1-\lambda}v$.

   Prenons une suite λ_n convergeant vers 0. De la suite des w_λ correspondants on peut extraire une suite convergente dans $\mathcal{C}_{\hat{x}}(\Omega)$ car $\mathcal{C}_{\hat{x}}(\Omega)$ est un fermé. Notons (par abus de notation) w_{λ_n} cette suite, on a : $\lim\limits_{n\to\infty} w_{\lambda_n} = w_\infty$ avec $w_\infty \in \mathcal{C}_{\hat{x}}(\Omega)$.

   Or, par contruction, $u = \frac{1}{1-\lambda_n}w_{\lambda_n} - \frac{\lambda_n}{1-\lambda_n}v$. Donc en faisant tendre n vers l'infini, on obtient : $u = w_\infty$. Donc $u \in \mathcal{C}_{\hat{x}}(\Omega)$.
   Donc $G_{\hat{x}} \subset \mathcal{C}_{\hat{x}}(\Omega)$ (les contraintes sont qualifiées).

□

5.2 Optimisation non linéaire : cas général (optimisation sous contraintes égalité et inégalité)

5.2.1 Lagrangien

Définition (Lagrangien)**.** *Le Lagrangien est un opérateur :*

$$L : \left\{ \begin{array}{ccc} \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^{+p} & \rightarrow & \mathbb{R} \\ (x, \lambda, \mu) & \mapsto & f(x) + \lambda^T g(x) + \mu^T h(x) \end{array} \right.$$

Les λ et μ sont appellés paramètres de Karush, Kuhn et Tucker (KKT). En développant l'écriture ci-dessus, on a :

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{q} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

5.2.2 Conditions du premier ordre

Théorème (CN et CS d'optimalité)**.**

CN Soit $\hat{x} \in \Omega$. Soient f, g_i et h_j différentiables en $\hat{x}$. Si les contraintes sont qualifiées en $\hat{x}$ on peut écrire :

$$\hat{x} \text{ solution de } (P_\Omega) \;\Rightarrow \exists \left(\hat{\lambda}, \hat{\mu}\right) \in \mathbb{R}^q \times \mathbb{R}^{+p} \;/\; \left\{ \begin{array}{c} \frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda}, \hat{\mu}) = 0 \\ \forall j \in [\![1; p]\!],\; \hat{\mu}_j h_j(\hat{x}) = 0 \end{array} \right.$$

Note : $\frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda}, \hat{\mu}) = \frac{\partial f}{\partial x}(\hat{x}) + \hat{\lambda}^T \frac{\partial g}{\partial x}(\hat{x}) + \hat{\mu}^T \frac{\partial h}{\partial x}$.

CS Si $\left\{ \begin{array}{c} g_i \text{ affines} \\ f, h_j \text{ convexes} \end{array} \right.$*, alors la relation précédente est une équivalence.*

5.2.3 Conditions du second ordre

Théorème (CN et CS d'optimalité)**.**
Soit $\hat{x} \in \Omega$. Soient f, g_i et h_j différentiables deux fois en $\hat{x}$. On suppose les contraintes qualifiées en $\hat{x}$. On pose $V(\hat{x}) = \left\{ z \in \Omega / \frac{\partial g_i}{\partial x}(\hat{x}) \cdot z = 0, i \in [\![1; q]\!] \text{ et } \frac{\partial h_j}{\partial x}(\hat{x}) \cdot z = 0 \text{ pour } j \in \mathcal{A}(\hat{x}) \right\}$

CN $\hat{x}$ *est solution de* $(P_\Omega) \;\Rightarrow\; \exists(\hat{\lambda}, \hat{\mu}) \in \mathbb{R}^q \times \mathbb{R}^{+p} / \; v^T \frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda}, \hat{\mu}) \cdot v \geq 0$ *pour tout* $v \in V(\hat{x})$.

CS Si $\exists(\hat{\lambda}, \hat{\mu}) \in \mathbb{R}^q \times \mathbb{R}^{+p}$ *tels que :*

$$\left. \begin{array}{c} \frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda}, \hat{\mu}) = 0 \\ \forall i \in [\![1; p]\!],\; \hat{\mu}_i h_i(\hat{x}) = 0 \\ \forall v \in V(\hat{x}),\; v^T \frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda}, \hat{\mu}) v > 0 \\ \forall i \in \mathcal{A}(\hat{x}),\; \mu_i > 0 \end{array} \right\} \text{ alors } \hat{x} \text{ est une solution stricte de } (P_\Omega).$$

Il existe une version "faible" de ce théorème (*ie.* qui ne permet pas de traiter tous les cas de figure : dans certains cas particuliers on n'arrive pas à prouver la CS) qu'on utilise en TD :

Théorème (CN et CS d'optimalité)**.** *Soit $\hat{x} \in \Omega$. Soient f, g_i et h_j différentiables deux fois en $\hat{x}$. On suppose les contraintes qualifiées en $\hat{x}$.*

CN $\hat{x}$ *est solution de* $(P_\Omega) \;\Rightarrow\; \exists(\hat{\lambda}, \hat{\mu}) \in \mathbb{R}^q \times \mathbb{R}^{+p} / \; v^T \frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda}, \hat{\mu}) \cdot v \geq 0$ *pour tout* $v \in \mathcal{D}_{\hat{x}}(\Omega)$.

CS Si $\exists(\hat{\lambda}, \hat{\mu}) \in \mathbb{R}^q \times \mathbb{R}^{+p}$ *tels que :*

$$\left. \begin{array}{c} \frac{\partial L}{\partial x}(\hat{x}, \hat{\lambda}, \hat{\mu}) = 0 \\ \forall i \in [\![1; p]\!],\; \hat{\mu}_i h_i(\hat{x}) = 0 \\ \forall v \in \mathcal{D}\hat{x}(\Omega),\; v^T \frac{\partial^2 L}{\partial x^2}(\hat{x}, \hat{\lambda}, \hat{\mu}) v > 0 \end{array} \right\} \text{ alors } \hat{x} \text{ est une solution stricte de } (P_\Omega).$$

Chapitre 6

Dualité

6.1 Problème primal

On cherche à résoudre :

$$(P_\Omega) : \begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x) \\ h_j(x) \leq 0, \; i \in [\![1;p]\!] \end{cases}$$

Proposition (Problème primal)**.** *Soit $\hat{x} \in \Omega$. $\hat{x} \in \Omega$ est solution de (P_Ω) si et seulement si $\hat{x}$ est solution du problème primal :*

$$\inf_{x \in \mathbb{R}^n} \sup_{\mu \geq 0} L(x, \mu)$$

Preuve. Posons $p : \begin{cases} \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \sup\limits_{\mu \geq 0} L(x, \mu) \end{cases}$ On veut montrer que la borne inférieure de la fonction $p(x)$ est atteint au même point que la solution du problème (P_Ω). On peut étudier pour cela la fonction p dans le domaine Ω et en dehors.

Pour $x \in \Omega$, on a : $\forall j \in [\![1;p]\!], \; h_j(x) \leq 0$, donc $\forall \mu_j \in \mathbb{R}^+, \; \mu_j h_j(x) \leq 0$.

Par conséquent, $\sum\limits_{j=1}^{p} \mu_j h_j(x) \leq 0$.

Or pour $\mu = (0, \dots, 0)$, on a $\sum\limits_{j=1}^{p} \mu_j h_j(x) = 0$.

Donc $\sup\limits_{\mu \geq 0} \sum\limits_{j=1}^{p} \mu_j h_j(x) = 0$,

et enfin : $p(x) = f(x)$

Par ailleurs, pour $x \notin \Omega$, il existe $i_0 \in [\![1;p]\!]$ tel que $h_{i_0}(x) > 0$.

En écrivant que $p(x) = \sup\limits_{\mu \geq 0} f(x) + \sum\limits_{\substack{j=1 \\ j \neq i_0}}^{p} \mu_j h_j(x) + \mu_{i_0} h_{i_0}(x)$, il apparait que le dernier terme n'est pas borné. On a donc, pour $x \notin \Omega$, $p(x) = +\infty$.

On peut donc écrire que la borne inférieure de $p(x)$ (si elle existe) est trouvée dans Ω. En effet : $\inf\limits_{x \in \mathbb{R}^n} p(x) = \min \left\{ \inf\limits_{x \in \Omega} p(x), \inf\limits_{x \notin \Omega} p(x) \right\} = \min \left\{ \inf\limits_{x \in \Omega} p(x), \infty \right\} = \inf\limits_{x \in \Omega} p(x)$.

Or $\inf\limits_{x \in \Omega} p(x) = \min\limits_{x \in \Omega} f(x) = f(\hat{x})$. Donc il est bien équivalent de dire que $\hat{x}$ est solution de (P_Ω) ou du problème primal. $\square$

6.2 Point selle d'une fonction multivariables

Définition (Point selle)**.** *Soit $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. Le point $(\hat{x}, \hat{y})$ est un point selle de F si et seulement si*

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^m, \; F(\hat{x}, y) \leq F(\hat{x}, \hat{y}) \leq F(x, \hat{y})$$

Propriété (Dualité faible)**.** *Soit* $F : X \times Y \to \mathbb{R}$ *où* $X \subset \mathbb{R}^n$ *et* $y \subset \mathbb{R}^m$*. On a :*

$$\sup_{y \in Y} \inf_{x \in X} F(x, y) \leq \inf_{x \in X} \sup_{y \in Y} F(x, y)$$

Preuve. C'est quasiment immédiat :

$\forall (x, y) \in (X \times Y),\ \inf_{x \in X} F(x, y) \leq F(x, y) \leq \sup_{y \in Y} F(x, y)$, donc en particulier :

$\forall (x, y) \in (X \times Y),\ \inf_{x \in X} F(x, y) \leq \sup_{y \in Y} F(x, y)$,

et donc $\forall y \in Y\ \inf_{x \in X} F(x, y) \leq \inf_{x \in X} \sup_{y \in Y} F(x, y)$,

par conséquent $\sup_{y \in Y} \inf_{x \in X} F(x, y) \leq \inf_{x \in X} \sup_{y \in Y} F(x, y)$. □

Définition (Saut de dualité)**.** *On appelle* saut de dualité *la quantité* $\inf_{x \in X} \sup_{y \in Y} F(x, y) - \sup_{y \in Y} \inf_{x \in X} F(x, y)$.

Théorème (Caractérisation des points selles)**.** *Soit* $F : X \times Y \to \mathbb{R}$ *où* $X \subset \mathbb{R}^n$ *et* $y \subset \mathbb{R}^m$*.*
$(\hat{x}, \hat{y})$ *est un point selle de* F *si et seulement si le saut de dualité est nul. En d'autres termes, si et seulement si :*

$$\inf_{x \in X} \sup_{y \in Y} F(x, y) = \sup_{y \in Y} \inf_{x \in X} F(x, y)$$

On peut en tirer une formulation séparée :

CN $(\hat{x}, \hat{y})$ *est un point selle de* $F \Rightarrow \inf_{x \in X} \sup_{y \in Y} F(x, y) = \sup_{y \in Y} \inf_{x \in X} F(x, y) = F(\hat{x}, \hat{y})$

CS $\inf_{x \in X} \sup_{y \in Y} F(x, y) = \sup_{y \in Y} \inf_{x \in X} F(x, y) \Rightarrow (\hat{x}, \hat{y})$ *est un point selle de* F*, avec* $\begin{cases} \hat{x} = \arg \inf_{x \in X} \sup_{y \in Y} F(x, y) \\ \hat{x} = \arg \sup_{y \in Y} \inf_{x \in X} F(x, y) \end{cases}$

Preuve. CN Supposons que $(\hat{x}, \hat{y}) \in X \times Y$ soit un point selle de F.

On a : $F(\hat{x}, y) \leq F(\hat{x}, \hat{y}) \leq F(x, \hat{y})$,

Donc $\sup_{y \in Y} F(\hat{x}, y) \leq F(\hat{x}, \hat{y}) \leq \inf_{x \in X} F(x, \hat{y})$.

Or $\sup_{y \in Y} F(\hat{x}, y) \geq \inf_{x \in X} \sup_{y \in Y} F(x, y)$,

et $\inf_{x \in X} F(x, \hat{y}) \leq \sup_{y \in Y} \inf_{x \in X} F(x, y)$.

Donc au final $\inf_{x \in X} \sup_{y \in Y} F(x, y) \leq \sup_{y \in Y} \inf_{x \in X} F(x, y)$.

Or d'après la propriété de dualité faible, on a $\inf_{x \in X} \sup_{y \in Y} F(x, y) \geq \sup_{y \in Y} \inf_{x \in X} F(x, y)$.

L'égalité est donc vérifiée.

CS On suppose que $\inf_{x \in X} \sup_{y \in Y} F(x, y) = \sup_{y \in Y} \inf_{x \in X} F(x, y)$.

Posons $\begin{cases} \hat{x} = \arg \inf_{x \in X} \sup_{y \in Y} F(x, y) \\ \hat{y} = \arg \sup_{y \in Y} \inf_{x \in X} F(x, y) \end{cases}$

On a $\begin{cases} F(\hat{x}, y) \leq \sup_{y \in Y} F)(\hat{x}, y) \\ F(x, \hat{y}) \geq \inf_{x \in X} F(x, \hat{y}) \end{cases}$.

Or, par définition de $(\hat{x}, \hat{y})$, on a $\begin{cases} \inf_{x \in X} \sup_{y \in Y} F(x, y) = \sup_{y \in Y} F(\hat{x}, y) \\ \sup_{y \in Y} \inf_{x \in X} F(x, y) = \inf_{x \in X} F(x, \hat{y}) \end{cases}$.

Donc en particulier $\begin{cases} F(\hat{x}, y) \leq \inf_{x \in X} \sup_{y \in Y} F(x, y) \\ F(x, \hat{y}) \geq \sup_{y \in Y} \inf_{x \in X} F(x, y) \end{cases}$.

Or les deux quantités de droite sont égales par hypothèse, donc on a : $F(\hat{x}, y) \leq F(\hat{x}, \hat{y}) \leq F(x, \hat{y})$.
$(\hat{x}, \hat{y})$ est bien un point selle de F.

□

6.3 Dualité

Définition (Problème dual)**.** *On pose le problème dual :*

$$\sup_{\mu \geq 0} \inf_{x \in \mathbb{R}^n} L(x, \mu)$$

Théorème (de la dualité)**.** *Soient le problème (P_Ω), $\hat{x} \in \Omega$. On suppose les contraintes qualifiées en $\hat{x}$. On suppose que f et $(h_j)_{1 \leq j \leq p}$ sont convexes. On a alors :*
$\hat{x}$ solution de (P_Ω) si et seulement si $\exists \hat{\mu} \in \mathbb{R}^{+p}$ tel que $(\hat{x}, \hat{\mu})$ soit un point selle de L. En d'autres termes, si et seulement si $(\hat{x}, \hat{\mu})$ est solution du problème dual.