Chapitre 1

Introduction à l'optimisation

1.1 Problématique

1.1.1 Cadre

Un problème d'optimisation consiste, étant donnée une fonction $f : S \to \mathbb{R}$, à trouver :

1) son minimum v (resp. son maximum) dans S

2) un point $x_0 \in S$ qui réalise ce minimum (resp. maximum) i.e. $f(x_0) = v$.

Vocabulaire

- f est la *fonction objectif*
- v est la *valeur optimale*
- x_0 est la *solution optimale*
- $S = \{$*solutions réalisables du problème*$\}$
- *écriture du problème* : $\min_{x \in S} f(x)$ resp. $\max_{x \in S} f(x)$

Remarque 1) L'optimisation est une branche des mathématiques. Dans la pratique, on part d'un problème concret, on le modélise et on le résoud mathématiquement (analytiquement : problème d'optimisation, numériquement : programme mathématique).

2) Lien minimum/maximum : soit f une fonction dont on veut trouver le maximum. Le problème $\max_{x \in S} f(x)$ renvoit (x_0, v) alors que le problème $\min_{x \in S} f(x)$ renvoit $(x_0, -v)$. D'où ce lien. Ainsi la recherche d'un maximum peut toujours se ramener à la recherche d'un minimum.

1.1.2 Applications

L'optimisation intervient dans de nombreux domaines :

• en recherche opérationnelle (problème de transport, économie, gestion de stocks...)

• en analyse numérique (approximation/résolution de systèmes linéaires, non linéaires...)

• en automatique (modélisation de systèmes, filtrage...)

• en ingénierie (dimensionnement de structures, conception optimale de systèmes (réseaux, ordinateurs...))

1.2 Différents types d'optimisation

1.2.1 Classification des problèmes d'optimisation

- optimisation linéaire
  f est une fonction linéaire : $f(x) = < c,x >$
  S est défini par des fonctions affines : $ax + b \geq 0$

- optimisation linéaire quadratique
  f est une fonction convexe quadratique : $f(x) = \frac{1}{2} < Ax,x > + < b,x >$
  A est une matrice symétrique semi-définie positive
  S est défini par des fonctions affines : $ax + b \geq 0$

- optimisation convexe
  f est une fonction convexe et S un domaine convexe

- optimisation différentiable
  f est une fonction différentiable
  S est défini par des fonction (=contraintes) différentiables

- optimisation non différentiable
  ex : $f(x) = \max\{f_1(x),..,f_m(x)\}$

- optimisation en dimension infinie
  ex : problèmes variationnels $J(x) = \int_0^T L(t,x(t),\dot{x}(t))dt$ avec x dans un ensemble de fonctions X, $x(0) = x_0$ et $x(T) = x_T$

Nous étudierons, dans ce cours, uniquement des problèmes d'optimisation non linéaire.

1.2.2 Optimisation non linéaire

On distingue trois types de problèmes :

- problème sans contraintes : $\min_{x \in \mathbb{R}^n} f(x)$

- problème avec contraintes de type égalité : $\min_{x \in S} f(x)$ avec S de la forme $S = \{x \in \mathbb{R}^n \ \text{ t.q. } \ g_i(x) = 0 \ \text{ pour } \ i = 1..l\}$ avec $g_i : \mathbb{R}^n \to \mathbb{R}$

- problème avec contraintes de type inégalité : $\min_{x \in S} f(x)$ avec S de la forme $S = \{x \in \mathbb{R}^n \ \text{ t.q. } \ h_i(x) \geq 0 \ \text{ pour } \ i = 1..l\}$ avec $h_i : \mathbb{R}^n \to \mathbb{R}$

1.3 Exemples de problèmes d'optimisation non linéaire

1. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ qui au couple (a,b) associe le réel $a^2 + b^2 + ab - 5a + 2b$. C'est un problème sans contrainte, chaque variable a, b varie dans $\mathbb{R}$ tout entier.

2. Un problème de production

   Une usine a besoin de n produits $a_1,..,a_n$ en quantité $x_1,..,x_n$ pour fabriquer un produit fini b en quantité $h(x_1,..,x_n)$.
   Soit p_i le prix unitaire du produit i et p le prix de vente du produit fini.
   On désire calculer la production optimale. Il s'agit donc de maximiser la fonction $ph(x_1,..,x_n) - \sum_{i=1}^{n} x_i p_i$

3. Un problème avec contrainte de type égalité

   On veut trouver le parallélépipède de volume maximal et de surface donnée. Soint a, b, c les différentes longueurs des côtés.
   On cherche donc à maximiser la fonction abc sous la contrainte $2(ab + ac + bc) = S$.

4. Un problème avec contrainte de type inégalité

   Une firme fabrique des objets A et des objets B en utilisant des

matières premières m_1 et m_2.
Pour fabriquer une unité de A, il faut 2 unités de m_1 et 1 de m_2. Et pour fabriquer une unité de B, il faut 1 unité de m_1 et 2 de m_2.
On dispose de 8 unités de m_1 et 7 de m_2.
Enfin, le bénéfice est de 4 par unité de A et de 5 par unité de B.
La patron de la firme voudrait optimiser son bénéfice.

Il s'agit donc de maximiser sur S la fonction $4a + 5b$ où a (resp. b) représente la quantité de produit A (resp. B) fabriquée avec S défini par

$$S = \{(a,b) \in \mathbb{N}^2 \ \text{ t.q. } \ 2a + b \leq 8, \ \ a + 2b \leq 7\}$$

<u>Résolution graphique</u>

$S = \{(a,b) \in \mathbb{N}^2 \ \text{ t.q. } \ b \leq -2a + 8, \ \ b \leq -\frac{a}{2} + \frac{7}{2}\}$

On trace les lignes de niveau

$$L_c = \{(a,b) \ \text{ t.q. } \ 4a + 5b = c\}$$

et on cherche les coordonnées (a_0,b_0) du point où c est maximal et $(a_0,b_0) \in L_c \cap S \neq \emptyset$.

On trouve $x_0 = (3,2)$ et $v = c = 4 * 3 + 5 * 2 = 23$.

Remarque La résolution graphique devient impossible pour des problèmes à plus de 3 variables. D'où la nécessité d'obtenir des outils pour résoudre les problèmes dans le cas général.

1.4 Existence d'une solution

Théorème 1.4.1 *Théorème de Weierstrass*
Soit f une fonction continue sur $S \subset \mathbb{R}^n$ avec S fermé borné.
Alors

$\exists \ x_0 \in S$ *tel que* $f(x_0) = \min_{x \in S} f(x)$
$\exists \ x_1 \in S$ *tel que* $f(x_1) = \max_{x \in S} f(x)$

Cela signifie que f est bornée sur S ($f(x_0) \leq f(x) \leq f(x_1)$) et atteint ses bornes. Le maximum et le minimum sont atteints dans ce cadre-là.

Exemple 1) Soit f de $\mathbb{R}$ dans $\mathbb{R}$ définie par $f(x) = x$. Soit $S = [0,1]$. Il est clair que le maximum est atteint en 1 et que $\max_{x \in S} f(x) = 1 = f(1)$. S est bien un fermé borné. Les conditions du théorème sont bien vérifiées.

2) Par contre, si on étudie la même fonction mais cette fois-ci sur $S =]0,1[$. On a aussi $\max_{x \in S} f(x) = 1$. Mais le maximum n'est pas atteint : $\nexists \; x_0 \in S$ tel que $f(x_0) = \max_{x \in S} f(x)$.
Ici les conditions du théorème ne sont pas vérifiées : S n'est pas un fermé borné.

Remarque Pour l'unicité, il faut utiliser des propriétés propres à la fonction objectif, comme la stricte convexité par exemple.

Chapitre 2

Méthodes de résolution des problèmes d'optimisation non linéaire sans contrainte

2.1 Quelques définitions

2.1.1 Définitions

On veut résoudre $\min_{x\in S\subset\mathbb{R}^n} f(x)$ ou $\max_{x\in S\subset\mathbb{R}^n} f(x)$ i.e. on cherche v valeur optimale et x_0 tel que $f(x_0) = v$.

Définition 2.1.1 *Soit $x_0 \in S$.*

1. *On dit que x_0 est un minimum local si*

$$\exists V \in \mathcal{V}(x_0) \quad \textit{tel que} \quad \forall x \in V, \quad f(x) \geq f(x_0)$$

2. *On dit que x_0 est un minimum local strict si*

$$\exists V \in \mathcal{V}(x_0) \quad \textit{tel que} \quad \forall x \in V, \quad f(x) > f(x_0)$$

3. *On dit que x_0 est un minimum global si*

$$\forall x \in S, \quad f(x) \geq f(x_0)$$

4. *On dit que x_0 est un minimum global strict si*

$$\forall x \in S, \quad f(x) > f(x_0)$$

On définit de la même façon un *maximum local*, *maximum local strict*,*maximum global*,*maximum global strict*, en renversant les inégalités.

Remarque Bien sûr, la qualité du minimum local dépend de la "taille" de V. Plus V est "grand", mieux c'est. Et si $S = V$, alors le minimum local est un minimum global.

Propriétés 1) Si x_0 est un minimum global, alors c'est un minimum local.
2) Si on connait tous les minima locaux alors le plus petit est le minimum global.

2.1.2 Dans la pratique

Nous noterons (P_S) le problème $\min_{x\in S} f(x)$ et (P) le problème $\min_{x\in\mathbb{R}^n} f(x)$.

Lemme 2.1.1 *Soit $S \subset \mathbb{R}^n$ un <u>ouvert</u>. Alors*

$$x_0 \textit{ est solution de } (P_S) \Leftrightarrow \begin{cases} x_0 \quad \textit{solution de} \quad (P) \\ x_0 \in S \end{cases}$$

En pratique, pour résoudre (P_S), on résoud d'abord (P), c'est plus simple puis

1. si (P) a une solution $x_0 \in S$ alors x_0 est solution de (P_S). Sinon, si $x_0 \notin S$ alors (P_S) n'a pas de solution.

2. si (P) n'a pas de solution, on ne sait pas pour (P_S).

2.2 Condition du premier ordre

2.2.1 Dérivabilité-notion de gradient

Définition 2.2.1 *Si $f : \mathbb{R}^n \to \mathbb{R}$. Si f est différentiable, alors on définit le gradient de f en x par $[\frac{\partial f}{\partial x_1}(x),..,\frac{\partial f}{\partial x_n}(x)]$. On le note $\nabla f(x)$.*

Proposition 2.2.1 *On a le développement de Taylor suivant,*

$$f(x + v) = f(x) + \nabla f(x).v + \epsilon(v)$$

avec $\frac{||\epsilon(v)||}{||v||} \to 0$ quand $||v|| \to 0$.

Exemple Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x_1,x_2) = x_1 + x_2^2$.
Alors

$$\nabla f(\bar{x}_1,\bar{x}_2) = [\frac{\partial f}{\partial x_1}(\bar{x}), \frac{\partial f}{\partial x_2}(\bar{x})] = [1 \quad 2\bar{x}_2]$$

où $\bar{x} = (\bar{x}_1,\bar{x}_2)$.

2.2.2 Recherche des points stationnaires

On recherche d'abord les points stationnaires.

Définition 2.2.2 *x_0 est un point stationnaire ssi $\nabla f(x_0) = 0$.*

Proposition 2.2.2 *Condition nécessaire d'optimalité.*
Supposons que S est ouvert et que f est différentiable sur S.
Si x_0 est un minimum ou maximum local alors $\nabla f(x_0) = 0$.

Corollaire 2.2.1 *Les minima et les maxima locaux sont des points stationnaires et par conséquent, le minimum global et le maximum global aussi.*

Remarque

1. S doit être ouvert pour pouvoir appliquer la proposition.
   Cex : Soit $S = [0,1]$ et $f(x) = x$. Alors $f'(x) = 1 \neq 0 \ \forall x$ et pourtant 1 est un maximum.
2. La condition nécessaire exige seulement que f soit différentiable en $\bar{x}$ mais comme on ne sait pas a priori si les points candidats $\bar{x}$ se trouvent dans la zone de différentiabilité, on exige une condition plus forte : f différentiable sur tout S. Toutefois, il n'en est pas toujours ainsi :

   Propriétés
   • Soit $f : S \to \mathbb{R}$ telle que $f(x) = \min[f_1(x),..,f_m(x)]$ avec les f_i différentiables sur S. Alors f est différentiable en tout point $\bar{x}$ minimum local de f et donc $\nabla f(\bar{x}) = 0$.

   • Dans certains cas, on peut conclure même si f n'est pas différentiable en $\bar{x}$.
   Soit $\bar{x} \in S$ point en lequel f est continue.
   Si $\exists V \in \mathcal{V}(\bar{x})$ ouvert tel que
   1) f est différentiable sur $V \ \{\bar{x}\}$
   2) $< \nabla f(x), x - \bar{x} > \geq 0$ (resp > 0) $\forall x \in V \ \{\bar{x}\}$
   Alors $\bar{x}$ est un minimum local (resp strict) de f.

Exemple : Soit f définie par $f(x,y) = x^{\frac{2}{3}} + y^{\frac{4}{5}} + 1$. Alors f n'est pas différentiable en (0,0) mais c'est un minimum strict de f sur $\mathbb{R}^2$.

3. Attention, la réciproque est fausse : on peut avoir $\nabla f(\bar{x}) = 0$ sans que $\bar{x}$ ne soit ni un minimum ni un maximum.

Conclusion : Dans les points stationnaires, il y a les minima et les maxima (locaux et globaux) et il y a aussi d'autres points.

Définition 2.2.3 *Un point stationnaire qui n'est ni un minimum ni un maximum est un point singulier.*

Remarque Si le nombre de variables est :
- 1 alors c'est un *point d'inflexion.* Ex : $f(x) = x^3$.

- 2 ça se complique. Ex : $f(x_1,x_2) = x_1^2 - x_2^2$. Alors

$\rightarrow x_0 = (0,0)$ est un maximum pour $\mathcal{C}_1 = \{(x_1,x_2)/x_1 = 0, y = x_1^2 - x_2^2\}$

$\rightarrow x_0 = (0,0)$ est un minimum pour $\mathcal{C}_2 = \{(x_1,x_2)/x_2 = 0, y = x_1^2 - x_2^2\}$

Définition 2.2.4 *Un point selle ou col pour $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est un point $x_0 = (x_0^1, x_0^2)$ tel que $\exists \Delta_1, \Delta_2$ droites passant par x_0 telles que $f(\Delta_1)$ présente un maximum en x_0 et $f(\Delta_2)$ présente un minimum en x_0.*

Remarque Tous les points singuliers ne sont pas des points selle.

2.3 Conditions du second ordre

2.3.1 La matrice Hessienne

Définition 2.3.1 *Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est de classe $\mathcal{C}^2$ alors on définit la matrice Hessienne de f en x par*

$$H_f(x) = \frac{\partial^2 f}{\partial x^2}(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

Proposition 2.3.1 *1. $H_f(x)$ est une matrice symétrique car $\frac{\partial^2 f}{\partial x_i \partial xj} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ dès que f est deux fois dérivable (Schwarz).*

2. *On a le développement de Taylor suivant :*

$$f(x+v) = f(x) + \nabla f(x).v + \frac{1}{2} v^t H_f(x) v + \epsilon(v)$$

avec $\frac{||\epsilon(v)||}{||v||} \to 0$ *quand* $||v||^2 \to 0$.

Exemple : Soit $f : \mathbb{R}^n \to \mathbb{R}$ telle que $f(x_1,x_2) = x_1 + x_2^2$.

Alors $\begin{cases} \nabla f(x) = [1 \;\; 2x_2] \\ H_f(x) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \end{cases}$

2.3.2 Rappels sur les matrices

Définition 2.3.2 *Soit* $M \in \mathbb{R}^{n \times n}$ *symétrique.*

1. *On dit que* M *est semi-définie positive ssi* $x^t M x \geq 0 \;\; \forall x \in \mathbb{R}^n$
2. *On dit que* M *est définie positive ssi* $x^t M x \geq 0 \;\; \forall x \in \mathbb{R}^n \{0\}$

Remarque On définit de la même façon les matrices semi-définie négative et définie négative en renversant les inégalités.

Rappel Les mineurs principaux d'une matrice M sont les déterminants des sous-matrices $k \times k$ de M obtenues en barrant un sous-ensemble de $(n-k)$ colonnes et le sous-ensemble correspondant des $(n-k)$ lignes.

2.3.3 Conditions du second ordre

Rappel Les points stationnaires sont de deux types : optimaux locaux/globaux et points singuliers (points selle et autres). Or ce que l'on recherche c'est un optimum global et on sait que s'il existe il fait partie des extrema locaux. Donc une fois que l'on a déterminé les points stationnaires ($\nabla f(x) = 0$) on doit faire le tri. D'où la nécessité de faire une étude du second ordre.

En effet si x_0 est tel que $\nabla f(x_0) = 0$ alors

$$f(x_0 + v) = f(x_0) + \frac{1}{2} v^t H f(x) v + \epsilon(||v||^2)$$

il faut donc déterminer le signe de $v^t H f(x) v$.

Si Hf est définie positive, $f(x_0 + v) > f(x_0)$ et x_0 est un minimum local strict.

Si Hf est semie définie positive, $f(x_0 + v) \geq f(x_0)$ et x_0 est un minimum local.

Si le signe de $v^t Hf(x)v$ varie, x_0 est un point singulier.

Plus précisément,

Théorème 2.3.1 *Condition nécessaire*
Si $x_0 \in S$ ouvert est un minimum local de f et si f est $\mathcal{C}^2$, alors

$$\begin{cases} \nabla f(x_0) = 0 \\ y^t Hf(x)y \geq 0 \forall y \in \mathbb{R}^n \end{cases}$$

La deuxième condition signifie que Hf est semi-définie positive.

On a aussi

Théorème 2.3.2 *Condition suffisante*
Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ est $\mathcal{C}^2$ en $x_0 \in S$.
Si $\nabla f(x_0) = 0$ et $y^t Hf(x)y \geq 0 \forall y \in \mathbb{R}^n$, alors x_0 est un minimum local strict de f.

Remarque 2.3.1 *Si Hf est seulement positive et non définie positive, on n'est pas assuré d'avoir un minimum local (ex: $f : \mathbb{R} \to \mathbb{R} \quad x \to x^3$, alors $\bar{x} = 0$, $Hf(\bar{x}) = 0 \geq 0$ et pourtant on n'a pas d'extremum en 0).*

Remarque 2.3.2 *En dimension 2, soit $A = \begin{pmatrix} r & s \\ s & t \end{pmatrix} \in \mathcal{M}_2(\mathbb{R})$*

1. *Si $\det(A) = rt - s^2 > 0$ alors les vap λ_1 et λ_2 de A sont de même signe car $\det(A) = \lambda_1\lambda_2$. De plus, $tr(A) = \lambda_1 + \lambda_2 = r + t$ et comme r et t sont de même signe $\det(A) = rt - s^2 > 0$, on en déduit que $\lambda_1 + \lambda_2$ a le signe de r. Les vap ont donc le signe de r.*
   Donc
   - *$\det(A) > 0$, A est définie positive $\Leftrightarrow r > 0$, A est définie négative $\Leftrightarrow r < 0$*
   - *$\det(A) < 0$, λ_1 et λ_2 sont de signes opposés donc A n'est ni définie positive ni définie négative*

2. *Considérons* $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ $\mathcal{C}^2$ *telle que* $\nabla f(\bar{x},\bar{y}) = 0$ *avec* $(\bar{x},\bar{y}) \in U$.

   Posons $r = \frac{\partial^2 f}{\partial x^2}(\bar{x},\bar{y})$, $r = \frac{\partial^2 f}{\partial y^2}(\bar{x},\bar{y})$ *et* $s = \frac{\partial^2 f}{\partial x \partial y}(\bar{x},\bar{y})$.

   Alors $Hf(\bar{x},\bar{y}) = A$ *et donc d'après le théorème et le point précédent, on a la proposition suivante :*

   Proposition 2.3.2 – *Si* $rt - s^2 > 0$ *et* $r > 0$, f *a un minimum local en* $(\bar{x},\bar{y})$.
   - *Si* $rt - s^2 > 0$ *et* $r < 0$, f *a un maximum local en* $(\bar{x},\bar{y})$.
   - *Si* $rt - s^2 = 0$, *on ne peut pas conclure.*

Exemple 2.3.1 *Soit* $f(x_1,x_2) = mx_1^2 + x_2^2$ *définie de* $\mathbb{R}^2$ *dans* $\mathbb{R}$, *avec* $m \neq 0$. *Alors* $\nabla f(x_1,x_2) = (2mx_1,2x_2)$.
On cherche les $x = (x_1,x_2)$ *tels que* $\nabla f(x) = 0$ *qui équivaut à* $\left\{ \begin{array}{l} 2mx_1 = 0 \\ 2x_2 = 0 \end{array} \right.$
La seule possibilité est $x_0 = (0,0)$. *Dans ce cas,*

$$Hf(x_0) = \begin{pmatrix} 2m & 0 \\ 0 & 2 \end{pmatrix} \in \mathcal{M}_2(\mathbb{R})$$

- *si* $m > 0$, *alors* $Hf(x_0)$ *est définie positive (vap* > 0*) donc par la condition suffisante* $x_0 = (0,0)$ *est un minimum local strict.*
- *si* $m < 0$, *alors les vap n'ont pas le même signe donc par la condition nécessaire,* $x_0 = (0,0)$ *ne peut pas être un minimum local.*

2.4 Algorithmes

Rappel : Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ avec S un ouvert. Alors si f est différentiable sur S et $\bar{x} \in S$ un extremum local alors $\nabla f(\bar{x}) = 0$.
Parfois, on ne peut pas résoudre cette équation explicitement donc on fait appel à une méthode nimérique. C'est la *méthode de Newton.*

Soit $g : S \subset \mathbb{R}^n \to \mathbb{R}$. On veut résoudre $g(x) = 0$. Notons a une racine.

L'idée est de remplacer la courbe représentative de g par sa tangente au point x_0 où x_0 est une approximation grossière de a :

$$y = g'(x_0)(x - x_0) + g(x_0) \sim g(x)$$

Posons $x_1 = x_0 - \frac{g(x_0)}{g'(x_0)}$. x_1 est une meilleure approximation de a que x_0. On itère alors la fonction $\phi(x) = x - \frac{g(x)}{g'(x)}$.

Théorème 2.4.1 *Soit g une fonction $\mathcal{C}^2$ définie de $I = [a-r,a+r]$ dans $\mathbb{R}$ telle que $g'(x) \neq 0 \quad \forall x \in I$.*

Posons $M = \max_I |\frac{g''(x)}{g'(x)}|$ et $h = \min(r, \frac{1}{M})$.

Alors
- $\forall x \in [a-h,a+h]$, $|\phi(x) - a| \leq M|x-a|^2$
- $\forall x_0 \in [a-h,a+h]$, $|x_p - a| \leq \frac{1}{M}|x_0 - a|^{2^p}$

Chapitre 3

Méthodes de résolution des problèmes d'optimisation avec contraintes

3.1 Contraintes de type égalités

Cadre : on veut minimiser la fonction différentiable $f : \mathbb{R}^n \to \mathbb{R}$ sur $C = \{x \in \mathbb{R}^n \text{ t.q. } h_i(x) = 0 \forall i = 1..m\}$ avec h_i de classe C^1.

On a donc ici un ensemble de contraintes "pas épais" (pelure de $\mathbb{R}^n$).

Théorème 3.1.1 *Soient f et $(h_i)_{i=1..m}$ C^1 au voisinage de $\bar{x}$. Si $\bar{x} \in \mathbb{R}^n \cap C$ est un minimum local de f sur C et si les vecteurs $\nabla h_1(\bar{x})$, .., $\nabla h_m(\bar{x})$ sont linéairement indépendants, alors $\exists(\lambda_1,..,\lambda_m) \in \mathbb{R}^m$ tel que*

$$(*) \quad \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) = 0$$

Les λ_i sont uniques.

Définition 3.1.1 *Les λ_i sont appelés les multiplicateurs de Lagrange.*

Remarque 1) Le théorème ne fournit pas d'information sur le signe des λ_i. (On aurait très bien pu faire $h_i \to -h_i$.)

2) On a le même type de conditions d'optimsation pour un maximum local.

Cex : Si les ∇h_i ne sont pas linéairement indépendants, le théorème ne s'applique pas.
Soit $f(x,y) = x + y^2$ à minimiser sur $x^3 - y^2 = 0$. Il est facile de voir que $\bar{x} = (0,0)$ est le minimum global de f sous $h(x) = x^3 - y^2 = 0$. Mais on ne peut pas avoir (*).

Interprétation analytique

Soit $\mathcal{L}(x,\lambda) = f(x)+\sum_{i=1}^m \lambda_i h_i(x)$ fonction de Lagrange associée au problème de minimisation de f sur C. Alors

$$(*) \Leftrightarrow \nabla_x \mathcal{L}(\bar{x},\bar{\lambda}) = 0$$

Or $\nabla_\lambda \mathcal{L}(\bar{x},\bar{\lambda}) = (h_1(\bar{x}),..,h_m(\bar{x})) = 0$, condition de type égalité.
Le théorème dit que en un minimum local (maximum local) $\bar{x}$ de f sur C, $\exists \bar{\lambda}$ tel que

$$\nabla \mathcal{L}(\bar{x},\bar{\lambda}) = (\nabla_x \mathcal{L}(\bar{x},\bar{\lambda}),\nabla_\lambda \mathcal{L}(\bar{x},\bar{\lambda})) = 0$$

L'idée de Lagrange a donc consisté à étendre la condition d'optimalité "$\nabla f(\bar{x}) = 0$" (problème sans contrainte) à "$\nabla \mathcal{L}(\bar{x},\bar{\lambda}) = 0$" et ce moyennant l'intervention des multiplicateurs λ.

Toutefois, il faut bien se garder de croire que la nature des points est conservée : si $\bar{x}$ est un minimum local de f sur C, il est certes stationnaire pour $\mathcal{L}(.,\bar{\lambda})$ mais pas nécessairement un minimum local.

Ex : $\bar{x} = -1$ est un minimum de $f(x) = x^3$ sous la contrainte $h(x) = x + 1 = 0$. Le multiplicateur associé est $\lambda = -3$ mais $\bar{x}$ est un maximum local de $\mathcal{L}(.,\bar{\lambda})$ sur $\mathbb{R}$.

En effet, $\nabla f(x) = 3x^2$ et $\nabla h(x) = 1 \neq 0 \forall x$.
On doit résoudre $\nabla f(x) + \lambda \nabla h(x) = 0$ sous la contrainte $x + 1 = 0$. On obtient $\bar{x} = -1$ et $\bar{lambda} = -3$.

On a aussi $\mathcal{L}(x,\bar{\lambda}) = x^3 + \bar{\lambda}(x+1) = x^3 - 3(x+1)$ et $\nabla_x \mathcal{L}(x,\bar{\lambda}) = 3x^2 + \bar{\lambda} = 3(x^2 - 1)$. Ainsi $\nabla_x \mathcal{L}(\bar{x},\bar{\lambda}) = 0$ et c'est donc bien un point stationnaire.
Pourtant, $\nabla_x^2 \mathcal{L}(x,\bar{\lambda}) = 6x$ et $\nabla_x^2 \mathcal{L}(\bar{x},\bar{\lambda}) = -6 \leq 0$. C'est donc un maximum local!!!

Interprétation géométrique

L'idée de différenciation du premier ordre est d'approcher $f(x)$ au voisinage de $\bar{x}$ par la fonction affine $x \to f(\bar{x}) + < \nabla f(\bar{x}), x - \bar{x} >$ et d'approcher $h - i(x)$ au voisinage de $\bar{x}$ par $< \nabla h_i(\bar{x}), x - \bar{x} >$ (car $h_i(\bar{x}) = 0$). D'où l'idée d'approcher l'ensemble contrainte C au voisinage de $\bar{x}$ par le sous-espace affine $\{x \in \mathbb{R}^n / < \nabla h_i(\bar{x}), x - \bar{x} > \quad i = 1..m\}$.
Ceci traduit la notion de tangence à C en $\bar{x}$. Un élément x de $\{x \in \mathbb{R}^n / < \nabla h_i(\bar{x}), x - \bar{x} > \quad i = 1..m\}$ est de la forme $x = \bar{x} + d$ où $d \in \{d \in \mathbb{R}^n / < \nabla h_i(\bar{x}), d > \quad i = 1..m\}$.

3.2 Contraintes de type inégalités

Cadre : on veut minimiser la fonction différentiable $f : \mathbb{R}^n \to \mathbb{R}$ sur $C = \{x \in \mathbb{R}^n \text{ t.q. } g_j(x) = 0 \forall j = 1..p\}$ avec g_j de classe C^1.

On a donc ici un ensemble contrainte "épais" (c'est à dire d'intérieur non vide).

Remarque 1) Si $\bar{x} \in C$ est candidat à être un minimum de f et est à l'intérieur de C alors $\nabla f(\bar{x}) = 0$.

2) Si $\bar{x}$ se trouve sur la frontière de C alors interviendront nécessairement les fonctions g_j.
Pour $\bar{x} \in C$, posons $J(\bar{x}) = \{j / \quad g_j(\bar{x}) = 0\}$.

Définition 3.2.1 *On appelle les $j \in J(\bar{x})$ les indices des contraintes* actives *ou* saturées *en $\bar{x}$.*

Si $j \notin J(\bar{x})$ i.e. $g_j(\bar{x}) < 0$ alors au voisinage de $\bar{x}$ par continuité, on a $g_j(x) < 0$. Donc la contrainte $g_j(x) \leq 0$ ne joue localement aucun rôle.
Si $j \in J(\bar{x})$ i.e. $g_j(\bar{x}) = 0$ alors il est probable que $\bar{x}$ est à la frontière de C et que l'on sort nécessairement de C en bougeant un petit peu autour de $\bar{x}$.

Donc seules les fonctions g_j correspondant aux indices $j \in J(\bar{x})$ interviendront dans les CN vérifiées par un minimum local $\bar{x}$ de f.

Définition 3.2.2 *Un point $x \in C$ est* régulier *quand les gradients des contraintes saturées en x forment un système libre.*

Exemple On cherche à calculer le max de $(-(x_1-2)^2-(x_2-1)^2))$ sous $C=\{-1\leq x_1\leq 1 \ \text{ et } \ 0\leq x_2\leq \frac{1}{2}\}$ ce qui revient à minimiser

$$(x_1-2)^2+(x_2-1)^2)$$

sur

$$C=\{g_1(x)\leq 0,g_2(x)\leq 0,g_3(x)\leq 0,g_4(x)\leq 0\}$$

avec $g_1(x)=x_1-1$, $g_2(x)=-x_1-1$, $g_3(x)=x_2-\frac{1}{2}$, $g_4(x)=-x_2$.
En $A=(1,0)$, 1 et 4 sont saturés mais $\nabla g_1(A)=(1,0)$ et $\nabla g_4(A)=(0,-1)$ forment un système libre donc A est régulier.

En $B=(1,\frac{1}{2})$, 1 et 3 sont saturés mais $\nabla g_1(A)=(1,0)$ et $\nabla g_3(A)=(0,1)$ forment un système libre donc B est régulier.

Théorème 3.2.1 *Soient f et $(g_j)_{j=1..p}$ C^1 au voisinage de $\bar{x}$. Si $\bar{x}$ est régulier et si $\bar{x}\in\mathbb{R}^n\cap C$ est un minimum local de f sur C alors $\exists(\mu_1,..,mu_p)\in\mathbb{R}^p$ tel que*

$$(*)\quad \nabla f(\bar{x})+\sum_{j=1}^{p}\mu_j\nabla g_j(\bar{x})=0$$

De plus, $\mu_j\geq 0forallj=1..p$ et $\mu_j g_j(\bar{x})=0forallj=1..p$

3.3 Contraintes de type égalités et inégalités

Soit $C=\{x\in mathbbR^n/h_i(x)=0 \ \ i=1..m,g_j(x)\leq 0 \ \ j=1..p\}$ où les fonctions sont C^1.

Théorème 3.3.1 *Soient f, $(g_j)_{j=1..p}$ et $(h_i)_{i=1..m}$ C^1 au voisinage de $\bar{x}$. Si $\bar{x}\in\mathbb{R}^n\cap C$ est un minimum local de f sur C et si $\{\nabla g_1(\bar{x}),..,\nabla g_p(\bar{x})\}\cup\{\nabla h_1(\bar{x}),..,\nabla h_m(\bar{x})\}$ sont linéairement indépendants, alors $\exists(\mu_1,..,\mu_p)\in\mathbb{R}^p_+$ et $(\lambda_1,..,\lambda_m)\in\mathbb{R}^m$ tels que*

$$(*)\quad \nabla f(\bar{x})+\sum_{j=1}^{p}\mu_j\nabla g_j(\bar{x})+\sum_{i=1}^{m}\lambda_i\nabla h_i(\bar{x})=0$$

et

$$\mu_j\nabla g_j(\bar{x})=0 \quad \textit{pour} \quad j=1..p$$

C'est un mélange des deux théorèmes précédents.

Chapitre 4

Optimisation convexe

4.1 Rappels

Définition 4.1.1 *Soit $S \subset \mathbb{R}^n$. On dit que*

$$\begin{aligned} S \quad \textit{est} \text{ convexe} \quad &\Leftrightarrow \quad \lambda x + (1-\lambda)y \in S \forall x,y \in S \quad and \quad \forall \lambda \in [0,1] \\ &\Leftrightarrow \quad \sum_{i=1}^{k} \lambda_i x_i \in S \forall (x_i)_{i=1..k} \in S \quad and \quad \forall \lambda_i \geq 0 \quad t.q. \quad \sum_{i=1}^{k} \lambda_i = 1 \end{aligned}$$

Exemple

Définition 4.1.2 *Soit $S \subset \mathbb{R}^n$ convexe. Soit $f : S \to \mathbb{R}$. On dit que*

1) f *est* convexe *sur* S

$$\Leftrightarrow \quad \forall x,y \in S \quad and \quad \forall \lambda \in [0,1] f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

2) f *est* strictement convexe *sur* S

$$\Leftrightarrow \quad \forall x,y \in S \quad and \quad \forall \lambda \in [0,1] f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

Exemple : $f(x) = x^4$ et $f(x) = x_1^2 + x_2^2$ sont strictement convexes sur $\mathbb{R}$ et $\mathbb{R}^2$.

Définition 4.1.3 *Soit $S \subset \mathbb{R}^n$ convexe. Soit $f : S \to \mathbb{R}$. On dit que*

$$\begin{array}{ll} 1) & f \quad \textit{est } \text{concave } \textit{sur} \quad S \\ & \Leftrightarrow \quad \forall x,y \in S \quad \textit{and} \quad \forall \lambda \in [0,1] f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y) \end{array}$$

$$\begin{array}{ll} 2) & f \quad \textit{est } \text{strictement concave } \textit{sur} \quad S \\ & \Leftrightarrow \quad \forall x,y \in S \quad \textit{and} \quad \forall \lambda \in [0,1] f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda) f(y) \end{array}$$

Remarque 4.1.1 *Si f est concave alors $-f$ est convexe.*

<u>Caractérisation du premier ordre</u>

Théorème 4.1.1 *Soit $f : \mathbb{R}^n \to \mathbb{R}$ C^1. Soit $\Omega \subset \mathbb{R}^n$ convexe.*

1) f est convexe *sur* $\Omega \Leftrightarrow \forall x,y \in \Omega f(y) \geq f(x) + (y-x)\frac{\partial f}{\partial x}(x)$

2) Si

$$\forall x \neq y \in \Omega f(y) > f(x) + (y-x)\frac{\partial f}{\partial x}(x)$$

alors f est strictement convexe sur Ω.

<u>Caractérisation du second ordre</u>

Théorème 4.1.2 *Soit $f : \mathbb{R}^n \to \mathbb{R}$, C^2.*

1) f est convexe sur $\mathbb{R}^n \Leftrightarrow \forall x \in \mathbb{R}^n \frac{\partial^2 f}{\partial^2 x}(x) \geq 0$
2) f est strictement convexe sur $\mathbb{R}^n \Rightarrow \forall x \in \mathbb{R}^n \frac{\partial^2 f}{\partial^2 x}(x) > 0$
3) f est strictement convexe sur un voisinage de $\bar{x} \Rightarrow \frac{\partial^2 f}{\partial^2 x}(\bar{x}) > 0$

Définition 4.1.4 *Un problème d'optimisation est dit* convexe *si on veut* $\min_{x \in S} f(x)$ *ou* $\max_{x \in S} f(x)$ *avec f convexe et S convexe.*

4.2 Optimisaion sans contrainte

1) Dans les chapitres précédents, on a vu qu'on n'a pas toujours des CNS mais seulement des CN, CS pour les minima. On va voir que dans le cadre des fonctions convexes, les choses se passent beaucoup mieux.

2) Point fort des fonctions convexes : min local = min global.

4.2.1 Conditions du premier ordre

La minimisation On a vu en introduction une CS d'existence d'un minimum et d'un maximum : si f est une fonction continue sur S fermé borné alors $\exists x_1 \in S$ et $x_2 \in S$ t.q. $\min_{x\in S} f(x) = f(x_1)$ et $\max_{x\in S} f(x) = f(x_2)$.

En fait, dans le cadre des fonctions convexes, on a même un CS d'unicité de la solution :
Théorème 4.2.1 *Soit S un convexe ouvert et f strictement convexe sur S. Alors il existe au plus un $\bar{x} \in S$ minimisant f sur S.*

Exemple :

Corollaire 4.2.1 *Si S est un convexe fermé borné et si f est une fonction strictement convexe sur un ouvert convexe contenant S
alors il existe un unique $\bar{x} \in S$ minimisant f sur S.*

Théorème 4.2.2 *CNS du premier ordre.*
Soit $f : S \subset \mathbb{R}^n \to \mathbb{R}$ avec S convexe ouvert. Supposons que f est différentiable sur S et convexe.
Alors les conditions suivantes sont équivalentes :

1. *$\bar{x}$ est un minimum de f sur S*
2. *$\bar{x}$ est un minimum local de f sur S*
3. *$\bar{x}$ est un point stationnaire de f i.e. $\nabla f(\bar{x}) = 0$.*

Exemple : Soient A une matrice carrée symétrique, définie positive et b un vecteur de $\mathbb{R}^n$. Soit f de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par

$$f(x) = \frac{1}{2} < Ax,x > + < b,x > +c$$

Déterminer le minimum de f s'il existe.

4.3 Optimisation avec contrainte

Il suffit de reprendre les théorèmes et conditions du chapitre concerné dans le cas des fonctions convexes.

1. Contraintes de type inégalités

Théorème 4.3.1 *Soient* f *et* $(g_j)_{j=1..p}$ *convexes. Si* $\bar{x}$ *est régulier et si* $\bar{x} \in \mathbb{R}^n \cap C$ *est un minimum local de* f *sur* C *alors* $\exists(\mu_1,..,mu_p) \in \mathbb{R}^p$ *tel que*

$$(*) \quad \nabla f(\bar{x}) + \sum_{j=1}^{p} \mu_j \nabla g_j(\bar{x}) = 0$$

De plus, $\mu_j \geq 0 forall j = 1..p$ *et* $\mu_j g_j(\bar{x}) = 0 forall j = 1..p.$

2. Contraintes de type 'egalités

Théorème 4.3.2 *Soient* f *convexe et* $(h_i)_{i=1..m}$ *affines. Si* $\bar{x} \in \mathbb{R}^n \cap C$ *est un minimum local de* f *sur* C *et si les vecteurs* $\nabla h_1(\bar{x})$*, ..,* $\nabla h_m(\bar{x})$ *sont linéairement indépendants, alors* $\exists(\lambda_1,..,\lambda_m) \in \mathbb{R}^m$ *tel que*

$$(*) \quad \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) = 0$$

Les λ_i *sont uniques.*

3. Contraintes de type 'egalités et in'egalités

Théorème 4.3.3 *Soient* f *convexe,* $(g_j)_{j=1..p}$ *convexes et* $(h_i)_{i=1..m}$ *affines. Si* $\bar{x} \in \mathbb{R}^n \cap C$ *est un minimum local de* f *sur* C *et si* $\{\nabla g_1(\bar{x}), .., \nabla g_p(\bar{x})\} \cup \{\nabla h_1(\bar{x}), .., \nabla h_m(\bar{x})\}$ *sont linéairement indépendants, alors* $\exists(\mu_1,..,\mu_p) \in \mathbb{R}^p_+$ *et* $(\lambda_1,..,\lambda_m) \in \mathbb{R}^m$ *tels que*

$$(*) \quad \nabla f(\bar{x}) + \sum_{j=1}^{p} \mu_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\bar{x}) = 0$$

et

$$\mu_j \nabla g_j(\bar{x}) = 0 \quad pour \quad j = 1..p$$