Chapitre 2

Optimisation différentiable sans contraintes

Dans ce chapitre, nous allons traiter de problèmes de minimisation (ou de maximisation) d'une fonction définie sur tout l'espace (donc pas de contrainte) ou encore que ce minimum est atteint à l'intérieur de l'ensemble des contraintes et par conséquent joue un rôle secondaire.

2.1 Rappels sur le calcul différentiel dans $\mathbb{R}^n$

On notera par $\{e_i\}_{i=1}^{n}$ la base canonique de $\mathbb{R}^n$

$$e_1 = (1, 0, 0, \ldots, 0)$$
$$e_2 = (0, 1, 0, \ldots, 0)$$
$$\vdots$$
$$e_n = (0, 0, 0, \ldots, 0, 1)$$

Soit f une fonction numérique définie sur un ouvert U de $\mathbb{R}^n$

$$f : U \longrightarrow \mathbb{R}$$

Définition : dérivées partielles

On dit que f admet des dérivées partielles au point $x_0 \in U$ si les limites suivantes existent

$$\frac{\partial f}{\partial x_i}(x_0) = \lim_{t \to 0} \frac{f(x_0 + t e_i) - f(x_0)}{t} \qquad \forall i = 1, 2, \ldots n$$

L'existence des dérivées partielles ne signifie nécessairement que la fonction possèdera de bonnes propriétés telles que

i) la continuité de f

ii) l'existence d'une approximation linéaire f au voisinage du pt x_0

Pour plus de détails, on consultera le chapitre 2 du manuel de M. Delfour. Dans ce qui suit, nous allons faire l'hypothèse que toutes les dérivées existent et sont continues.

Définition : fonction de classe C^1

Soit $f: U \to \mathbb{R}$ définie sur un ouvert de $\mathbb{R}^n$.
On dit que f est de classe C^1 sur U si

1) $\frac{\partial f}{\partial x_i}(x)$ existent $\forall x \in U$ et $\forall i = 1, \ldots, n$

2) les $\frac{\partial f}{\partial x_i} : U \to \mathbb{R}$ sont des fonctions continues

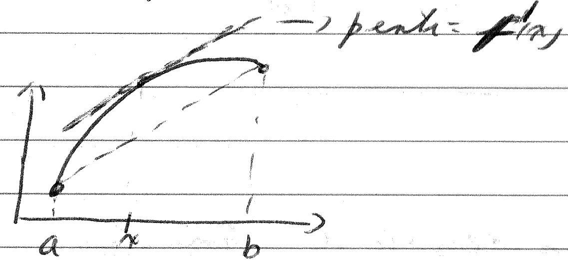
Remarque : pour cette classe de fonction, il n'est pas nécessaire d'introduire différents concepts de dérivabilité comme les dérivées de Fréchet, de Gâteaux ou d'Hadamard (voir M. Delfour). En fait, tous ces concepts sont équivalents si la fonction f est de classe C^1

Rappellons deux théorèmes qui seront importants pour la suite.

Théorème de la moyenne pour une fonction d'une variable

Si $F : [a,b] \to \mathbb{R}$ est continue et dérivable partout dans $]a,b[$, il existe un point $x \in]a,b[$ tel que

$$F(b) - F(a) = (b-a)\, F'(x)$$



Théorème de Taylor

Soit $F : [a,b] \to \mathbb{R}$ admettant des dérivées $F', F^{(II)}, F^{(III)}, \ldots, F^{(N)}$ continues dans $]a,b[$ et que la dérivée $F^{(N+1)}_{(t)}$ existe $\forall t \in]a,b[$.

Si x_0 et x sont deux points de $]a,b[$, alors il existe un point α entre x_0 et x tel que

$$F(x) = F(x_0) + F'(x_0)(x-x_0) + F''(x_0)\frac{(x-x_0)^2}{2} + \cdots$$

$$\cdots F^{(N)}_{(x_0)} \frac{(x-x_0)^N}{N!} + F^{(N+1)}_{(\alpha)} \frac{(x-x_0)^{N+1}}{(N+1)!}$$

Soit $f : U \to \mathbb{R}$ une fonction numérique définie sur un ouvert U de $\mathbb{R}^n$, de classe $\mathcal{C}^1$.

Définition : gradient de f

On définit le vecteur gradient au pt $x_0 \in U$ par le vecteur

$$\nabla f(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \frac{\partial f}{\partial x_2}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0) \right) \in \mathbb{R}^n$$

Définition : dérivée directionnelle

Soit $f : \underset{\text{ouvert}}{U} \xrightarrow{\mathcal{C}^1} \mathbb{R}$, $x_0 \in U$ et $v \in \mathbb{R}^n$ (une direction)

La dérivée directionnelle de f dans la direction v est définie par

$$D_v f(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

On a la formule suivante pour le calcul de $D_v f(x_0)$

$$\underline{D_v f(x_0) = \left(\nabla f(x_0), v \right)}$$

Théorème de la moyenne pour une fonction de plusieurs variables.

Soit $f : \underset{\text{ouvert}}{U} \to \mathbb{R}$ de classe $\mathcal{C}^1$ et $x, y \in U$.

Si le segment de droite $[x, y] \subset U$, alors $\exists \bar{t} \in]0, 1[$ tel que

$$f(y) = f(x) + \left(\nabla f(x + \bar{t}(y - x)), y - x \right)$$

Preuve: on considère la fonction d'une variable

$$F(t) = f(x + t(y-x))$$

On a que: $F(0) = f(x)$

$F(1) = f(y)$

et

$F'(t) = \left(\nabla f(x + t(y-x)), y-x\right)$

On applique le théorème de la moyenne pour F sur l'intervalle $[0,1]$

$\Rightarrow \exists \bar{t} \in]0,1[$ t.q.

$$F(1) = F(0) + F'(\bar{t})$$

$$\Leftrightarrow f(y) = f(x) + \left(\nabla f(x + \bar{t}(y-x)), y-x\right)$$

On peut aussi étendre le théorème de Taylor dans le cas de fonctions de plusieurs variables. Nous allons nous limiter au développement d'ordre 2 ce qui devrait suffire dans les applications traitées par la suite. Mais on peut aisément l'étendre jusqu'à un ordre quelconque.

Définition: f de classe C^k

$f : U \to \mathbb{R}$ où U est un ouvert de $\mathbb{R}^n$.

On dit que f est de classe C^k sur U si

$\frac{\partial f}{\partial x_i}$ soit de classe C^{k-1} $\quad \forall i = 1, 2, \dots, n$

sur U

Par exemple, pour le cas de la classe C^2, ceci signifie que

i) f continue

ii) $\frac{\partial f}{\partial x_i}$ continue $\forall i$

iii) $\frac{\partial^2 f}{\partial x_j \partial x_i}$ continue $\forall i, j$

Définition: Hessien

Soit $f: U \to \mathbb{R}$ de classe C^2 et $x_0 \in U$.

La matrice Hessienne au pt. x_0 est définie par

$$(H_f)_{ij} = H_{ij}(x_0) = \frac{\partial^2 f}{\partial x_i \partial x_j} \left(= \frac{\partial^2 f}{\partial x_j \partial x_i} \right)$$

Ex:

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

$$\frac{\partial f}{\partial x} = 12x^3 - 8xy$$

$$\frac{\partial f}{\partial y} = -4x^2 + 2y$$

$$H_f = \begin{pmatrix} 36x^2 - 8y & -8x \\ -8x & 2 \end{pmatrix}$$

au pt $(0,0)$

$$H_f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

Remarque: la matrice H_f est toujours symétrique.

Théorème de Taylor pour une fonction de plusieurs variables.

Soit $f: U \to \mathbb{R}$ de classe C^2 définie sur un ouvert $U \subset \mathbb{R}^n$. On se donne 2 pts $x, y \in U$.

Si le segment de droite $[x, y] \subset U$, alors $\exists\, \xi \in]0,1[$ tel que

$$f(y) = f(x) + (\nabla f(x), y-x) + \frac{1}{2}\left(H_f(x+\xi(y-x))(y-x), y-x\right)$$

Preuve: on introduit la fonction d'une variable

$$F(t) = f(x + t(y-x))$$

On sait que:

$$F(0) = f(x)$$
$$F(1) = f(y)$$
$$F'(t) = (\nabla f(x+t(y-x)), y-x)$$

Calculons $F''(t)$

$$F''(t) = \frac{d}{dt} \sum_i \frac{\partial f}{\partial x_i}(x+t(y-x))(y_i - x_i)$$

$$= \sum_i \sum_j \frac{\partial^2 f}{\partial x_j \partial x_i}(x+t(y-x))(y_j - x_j)(y_i - x_i)$$

$$= \sum_{i,j} H_{ij}(y_j - x_j)(y_i - x_i)$$

$$= (H_f(y-x), y-x)$$

où H est évalué au pt $x + t(y-x)$

On applique le théorème de Taylor à $F \Rightarrow \exists \xi \in]0,1[$

$$F(1) = F(0) + F'(0) + \frac{F''(\xi)}{2}$$

$\Leftrightarrow$

$$f(y) = f(x) + (\nabla f(x), y-x) + \frac{(H_f(y-x), y-x)}{2}$$

où H est évalué au pt $x + \xi(y-x)$.

2.2 Dérivabilité et convexité

Dans cette section, nous allons caractériser la convexité d'une fonction en termes de notions de dérivabilité. Nous obtiendrons des critères plus simples pour vérifier la convexité et la stricte convexité d'une fonction donnée.

Théorème: critère de convexité du 1er ordre

Soit $f: U \to \mathbb{R}$ une fonction numérique de classe C^1 définie sur ouvert convexe $U \subset \mathbb{R}^n$.

a) f est convexe dans U $\iff$ $f(y) \geq f(x) + (\nabla f(x), y-x)$ $\forall x \forall y \in U$

b) f est strictement convexe dans U $\iff$ $f(y) > f(x) + (\nabla f(x), y-x)$ $\forall x \neq y \in U$

Preuve

a) $\Rightarrow$

f convexe, $0 \leq t \leq 1$ $\Rightarrow$ $f(x+t(y-x)) \leq (1-t)f(x) + t f(y)$
$\leq f(x) + t(f(y) - f(x))$

$$\Rightarrow \frac{f(x+t(y-x)) - f(x)}{t} \leq f(y) - f(x)$$

$$\lim_{t \to 0} \frac{f(x+t(y-x)) - f(x)}{t} \leq f(y) - f(x)$$

$$(\nabla f(x), y-x) \leq f(y) - f(x)$$

$\Leftarrow$ Par hypothèse, on a que

$$f(z) \geq f(u) + \langle \nabla f(u), z-u \rangle \qquad \forall u, z \in U$$

On pose $z = x$ et $u = x + t(y-x)$

(1) $\Rightarrow$ $f(x) \geq f(x+t(y-x)) - t\langle \nabla f(x+t(y-x)), y-x \rangle$

et avec $z = y$ et $u = x+t(y-x)$

(2) $f(y) \geq f(x+t(y-x)) + (1-t)\langle \nabla f(x+t(y-x)), y-x \rangle$

La combinaison convexe de (1) et (2) donne

$$(1-t) f(x) + t f(y) \geq \underbrace{(1-t+t)}_{} f(x+t(y-x))$$

$$\geq f((1-t)x + ty) \qquad \forall x, y \in U$$

Preuve de b)

($\Rightarrow$) Soit $x \neq y$ et $t \in]0,1[$
Si f est strict. convexe $\Rightarrow$ convexe
selon a), on a que

$$f(x+t(y-x)) \geq f(x) + t\langle \nabla f(x), y-x \rangle$$

$\Leftrightarrow$ $t\langle \nabla f(x), y-x \rangle + f(x) < (1-t) f(x) + t f(y)$
car f est strict. convexe

$\Leftrightarrow$ $t\langle \nabla f(x), y-x \rangle < t(f(y) - f(x))$

$\therefore \langle \nabla f(x), y-x \rangle < f(y) - f(x)$

$\Leftarrow$ On refait la même preuve que a) $\Leftarrow$ avec inégalité stricte.

Remarque : dans ce théorème, nous avons utilisé seulement la dérivée première pour caractériser la convexité ! Dans ce qui va suivre, nous allons donner un critère basé sur la dérivée seconde (bien connu en une variable réelle).

Théorème : critère de convexité du 2^e ordre

Soit $f : U \to \mathbb{R}$ une fonction de classe C^2 définie sur un ouvert convexe $U \subset \mathbb{R}^n$.

a) f est convexe dans U $\Leftarrow$ $H_f(x) \geq 0 \quad \forall x \in U$

La notation $A \geq 0$ signifie que $(Au, u) \geq 0 \quad \forall u \in \mathbb{R}^n$

b) s'il existe un pt $x \in U$ t.q. $H_f(x) > 0$ alors on peut trouver un voisinage de x (une boule centrée en x) tel que f soit strictement convexe dans ce voisinage.

Preuve de a)

$\Rightarrow$ On utilise le théorème de Taylor

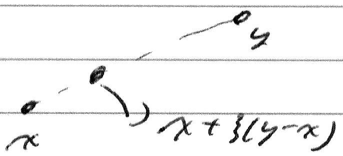
$$f(y) = f(x) + (\nabla f(x), y-x) + \frac{1}{2}(H_f(x+\xi(y-x))\, y-x, y-x)$$

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

grâce au critère de convexité du 1^e ordre

$$\Rightarrow (*) \quad \big(Hf(x+\xi(y-x))\, y-x,\, y-x\big) \geq 0$$

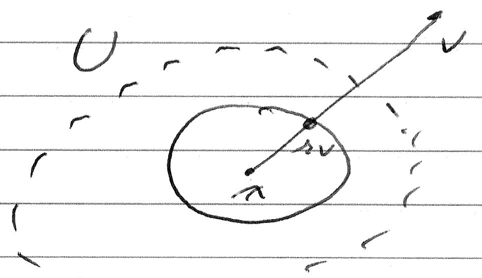
avec



Mais l'inégalité (*) est encore vraie si y s'approche de x tout en restant sur le segment de droite.

$$\Rightarrow \big(Hf(x)\, y-x,\, y-x\big) \geq 0$$

Cela est vrai $\forall y \in U \Rightarrow y = x + rv$ où $r>0$ est petit et $\forall v \in \mathbb{R}^n$ car U est ouvert et convexe.



Donc $\quad \big(Hf(x)\, rv,\, rv\big) \geq 0$

$$\Leftrightarrow \big(Hf(x)\, v,\, v\big) \geq 0$$

$$\Leftrightarrow Hf(x) \geq 0$$

$\Leftarrow$

De nouveau, par Taylor, on a que

$$f(y) = f(x) + \big(\nabla f(x), y-x\big) + \frac{1}{2}\big(Hf(x+\xi(y-x))\, y-x,\, y-x\big)$$

$$\geq f(x) + \big(\nabla f(x), y-x\big) \quad \text{car } Hf(z) \geq 0 \ \forall z \in U$$

donc $z = x + \xi(y-x)$

d'où le résultat par le critère de convexité du 1er ordre.

Preuve de b)

$\exists V$ = voisinage convexe de x t.q.

$(\Rightarrow)$ Si $H_f(x) > 0 \Rightarrow H_f(y) > 0 \ \forall y \in V$

Par Taylor

$y, x \in V$ $\Rightarrow f(y) = f(x) + (\nabla f(x), y-x) + \frac{1}{2}\left(H_f(x+\xi(y-x))(y-x), y-x\right)$

$> f(x) + (\nabla f(x), y-x)$

d'où f est strictement convexe