Rappels sur l'optimisation libre

(A lire attentivement)

On parle d'optimisation "libre" (ou sans contraintes) lorsqu'on cherche à minimiser ou maximiser une fonction sur tout son domaine de définition. Tout ce qui suit a été abordé en 1ère année éventuellement avec d'autres notations. Le but de ce texte est de :

1. résumer des résultats qui doivent être connus (avant d'aborder la question plus difficile de l'optimisation sous contraintes)
2. fixer les notations que nous emploierons.

1 Quelques définitions et notations

On considère une fonction de n variables $f : \vec{x} = (x_1, \ldots, x_n) \in \Omega \subset \mathbb{R}^n \mapsto f(x_1, \ldots, x_n) \in \mathbb{R}$. Dans tout ce qui suit, nous supposerons satisfaites les deux conditions suivantes :

1. L'ensemble Ω ("omega") sur lequel est définie f est "ouvert". Cela veut dire que *tous les points suffisamment "proches" d'un point donné de Ω sont aussi dans Ω.*
2. La fonction f est de classe $\mathcal{C}^2$ sur Ω i.e. elle admet des dérivées partielles d'ordre 1 et 2 en tout point de Ω, et ces dérivées partielles sont continues.

1.1 Ouverts

Précisons comment la condition 1. s'exprime mathématiquement. On appelle *boule ouverte* de centre $\vec{x}$ et de rayon $r > 0$ l'ensemble $B(\vec{x}, r) = \{\vec{y} \in \mathbb{R}^n / \|\vec{x} - \vec{y}\| < r\}$. C'est l'ensemble des points (ou vecteurs) *à distance inférieure à r de $\vec{x}$.* La définition d'un ensemble ouvert est alors

Définition 1 *Un sous-ensemble Ω de $\mathbb{R}^n$ est dit* ouvert *si, pour tout $\vec{x} \in \Omega$, on peut trouver $r > 0$ tel que $B(\vec{x}, r) \subset \Omega$.*

Exemple 2 *Si $n = 1$, pour tout $x \in \mathbb{R}$ et $r > 0$, $B(x, r) =]x - r, x + r[$. L'intervalle $]a, b[, a < b$, est ouvert mais $]a, b]$ ne l'est pas. Pourquoi ?*

Exemple 3 *Dans $\mathbb{R}^2$, $B(\vec{x}, r)$ est le disque de centre $\vec{x}$ et de rayon r (bord exclu). On peut montrer que c'est un ensemble ouvert.*

1.2 Extrema

Lorsque Ω est ouvert, les definitions suivantes ont un sens:

1. **Maximum local** : $\vec{x} \in \Omega$ est un maximum local de f si, "localement" autour de $\vec{x}$, $f(\vec{x}) \geq f(\vec{y})$. Plus précisément, s'il existe $r > 0$ tel que, pour tout $\vec{y} \in B(\vec{x}, r) \subset \Omega$, $f(\vec{x}) \geq f(\vec{y})$.
2. **Maximum global** : $\vec{x} \in \Omega$ est un maximum global de f si, pour tout $\vec{y} \in \Omega$, $f(\vec{x}) \geq f(\vec{y})$.

3. **Minimum local, global**: on change les inégalités $\geq$ par des $\leq$ dans les définitions précédentes. Les valeurs de f sont donc supérieures à $f(\vec{x})$ autour de $\vec{x}$ (minimum local) ou partout dans Ω (minimum global).

4. **Extremum local, global**: un extremum est un minimum ou un maximum.

1.3 Formule de Taylor

L'étude des extrema locaux repose sur la formule de Taylor.

Théorème 4 (Formule de Taylor à l'ordre 2) *Soit $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$ sur un ouvert Ω de $\mathbb{R}^n$ et soit $\vec{x} \in \Omega$. Alors, il existe $r > 0$ (tel que $B(\vec{x}, r) \subset \Omega$) et une fonction $\epsilon : B(\vec{0}, r) \subset \mathbb{R}^n \to \mathbb{R}$ (i.e. une fonction définie autour de $\vec{0}$) qui tend vers 0 en $\vec{0}$, pour lesquels on a, pour tout $\vec{u} = (u_1, \ldots, u_n) \in B(\vec{0}, r)$,*

$$f(\vec{x} + \vec{u}) = f(\vec{x}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{x})\, u_i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x})\, u_i\, u_j + \epsilon(\vec{u})\, \|\vec{u}\|^2 . \tag{1}$$

Cette formule, que vous avez vue l'année dernière, peut sembler compliquée au premier abord. Mais ce qu'elle signifie est simple. C'est dire qu'**une fonction régulière se comporte à l'ordre 2** (i.e. à un reste près qui tend plus vite vers 0 que le carré de la norme) et **autour d'un point** $\vec{x}$ fixé, **comme la somme d'une constante, d'une partie linéaire et d'une partie quadratique**. Bien comprendre ceci est fondamental. Pour mieux le voir, nous allons réécrire la formule (1) avec de nouvelles notations.

Définition 5 (Gradient) *On appelle **gradient** de f en $\vec{x}$, noté $Df(\vec{x}) \in \mathbb{R}^n$, le vecteur des dérivées partielles premières de f en $\vec{x}$:*

$$Df(\vec{x}) = \left(\frac{\partial f}{\partial x_1}(\vec{x}), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}) \right).$$

Définition 6 (Hessienne) *On appelle **matrice hessienne** de f en $\vec{x}$ la matrice $D^2 f(\vec{x}) \in M_n(\mathbb{R})$ des dérivées partielles secondes de f en $\vec{x}$:*

$$D^2 f(\vec{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}) \right)_{i,j=1,\ldots,n} \in M_n(\mathbb{R}).$$

Nous noterons parfois $H(\vec{x})$ ou même H pour $D^2 f(\vec{x})$. Si $\vec{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ est considéré comme un vecteur ligne (i.e. $\vec{u} \in M_{1,n}(\mathbb{R})$) alors ${}^t\vec{u}$ est un vecteur colonne et en explicitant le produit de matrice, on trouve

$$\vec{u}\, D^2 f(\vec{x})\, {}^t\vec{u} = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x})\, u_i\, u_j \in \mathbb{R},$$

qui n'est autre que

- la forme quadratique en $\vec{u}$ définie par la matrice $H = D^2 f(\vec{x})$.
- la partie du second ordre dans la formule de Taylor (au $\frac{1}{2}$ près).

On peut donc réécrire la formule de Taylor (1) à l'ordre 2 sous la forme plus "simple" (ou plutôt plus synthétique)

$$\boxed{f(\vec{x} + \vec{u}) = \underbrace{f(\vec{x})}_{\text{constante}} + \underbrace{\langle Df(\vec{x}), \vec{u} \rangle}_{\text{linéaire}} + \frac{1}{2} \underbrace{\vec{u}\, \left(D^2 f(\vec{x}) \right)\, {}^t\vec{u}}_{\text{quadratique}} + \underbrace{\epsilon(\vec{u})\, \|\vec{u}\|^2}_{\text{reste}} .} \tag{2}$$

Remarquez que cette formule est très proche de celle obtenue pour les fonctions d'une variable et est donc assez facile à retenir.

2 Détermination des extrema locaux

2.1 Condition nécessaire du 1er ordre

Définition 7 (Point stationnaire) *Un point $\vec{x}$ de Ω est dit point* ***stationnaire*** *(ou* ***critique****) de f si $Df(\vec{x}) = \vec{0}$. C'est donc un point où le gradient s'annule.*

Proposition 8 *Si f admet un extremum local en $\vec{x}$, $\vec{x}$ est un point stationnaire.*

Pour déterminer les extrema locaux, on cherche donc tout d'abord à déterminer les points stationnaires (nous noterons leur ensemble $\mathcal{S}(f)$). Une étude de la hessienne permet alors de conclure, dans les cas favorables, sur leur *nature* i.e. de savoir s'il s'agit bien (ou non) d'un maximum ou d'un minimum local. C'est ce qu'il nous reste à faire.

2.2 Conditions suffisantes du second ordre

La formule de Taylor à l'ordre 2 permet d'obtenir le résultat suivant

Théorème 9 *Soit $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$ sur un ouvert Ω de $\mathbb{R}^n$ et soit $\vec{x} \in \Omega$ un point stationnaire de f. Notons $H = H(\vec{x}) = D^2 f(\vec{x})$ la matrice hessienne de f en $\vec{x}$ et q_H (ou $q_{H(\vec{x})}$ lorsque nécessaire) la forme quadratique associée i.e. $q_H(\vec{u}) = \vec{u}\, H\, {}^t\vec{u}$.*

i) Si q_H est définie positive, *$\vec{x}$ est un minimum local.*

ii) Si q_H est définie négative, *$\vec{x}$ est un maximum local.*

iii) Si $\vec{0}$ est un point col de $q_{H(\vec{x})}$, $\vec{x}$ est un point col de f et n'est donc pas un extremum local.

iv) Dans tous les autres cas, on ne peut pas conclure (à ce stade).

Ce théorème s'énonce brièvement. Encore faut-il savoir ce que cela veut dire. C'est ce que nous allons préciser maintenant.

2.3 Précisions sur les formes quadratiques

Définition 10 (Définie positive/négative) *La forme quadratique q_H est dite définie positive (resp. définie négative) si $q_H(\vec{u}) > 0$ (resp. < 0) pour tout vecteur $\vec{u} \neq \vec{0}$.*

Définition 11 (Point col d'une forme quadratique) *$\vec{0}$ est point col de q_H si on peut trouver 2 vecteurs $\vec{u}, \vec{v} \in \mathbb{R}^n$ tels que $q_H(\vec{u}) > 0$ et $q_H(\vec{v}) < 0$.*

Pour déterminer si la forme quadratique q_H (non nulle) est définie positive, définie négative ou bien admet $\vec{0}$ comme point col, on utilise la **réduction de Gauss**. Vous avez vu sur des exemples l'année dernière comment cette réduction s'opère en pratique. Et c'est ce qui importe. Nous allons néanmoins rappeler comment définir cette réduction en général.

1. On peut toujours écrire q_H sous la forme "réduite" suivante

$$q_H(\vec{u}) = \sum_{i=1}^{p} \alpha_i \langle \vec{b}_i, \vec{u} \rangle^2 \tag{3}$$

   où $1 \leq p \leq n$, $(\vec{b}_1, \dots \vec{b}_p)$ est une famille libre de vecteurs de $\mathbb{R}^n$ et $\alpha_1, \dots, \alpha_p \in \mathbb{R}^*$. Dans cette formule, le crochet désigne le produit scalaire.

2. La *signature* de q_H est le couple (n_1, n_2) où n_1 est le nombre de coefficients $\alpha_i > 0$ et n_2 est le nombre de coefficients $\alpha_i < 0$. On a $0 \leq n_1 \leq p$, $0 \leq n_2 \leq p$ et $n_1 + n_2 = p$.

Il y a bien des façons d'écrire q_H sous la forme (3) (en fait, il y en a une infinité...) mais le couple (n_1, n_2) sera le même pour chacune d'elles et ne dépend donc que de q_H. C'est ce couple (n_1, n_2) qui permet de dire si q_H est définie positive, définie négative ou admet $\vec{0}$ comme point col.

Proposition 12
1) q_H est définie positive ssi sa signature est $(n, 0)$.
2) q_H est définie négative ssi sa signature est $(0, n)$.
3) $\vec{0}$ est un point col de q_H ssi ($n_1 \geq 1$ et $n_2 \geq 1$).

2.4 Conclusion

Ce qu'il faut retenir de tout cela est que pour déterminer les extrema locaux d'une fonction régulière, on cherche d'abord ses points stationnaires. Puis, que chercher la signature de la forme quadratique associée à la matrice hessienne en un point stationnaire permet de conclure sur sa nature (dans les bons cas au moins).

La raison en est que, grâce à la formule de Taylor à l'ordre 2, autour d'un point stationnaire, la fonction f se comporte comme q_H à l'ordre 2:

$$f(\vec{x} + \vec{u}) = f(\vec{x}) + \frac{1}{2} q_{H(\vec{x})}(\vec{u}) + \text{reste},$$

(puisque le gradient est nul en $\vec{x}$). Dans les trois cas "favorables" du théorème 9, q_H l'*emporte* sur le reste et f se comporte effectivement comme q_H. Dans tous les autres cas, l'étude du reste (et donc un développement de Taylor à des ordres supérieurs à 2) est nécessaire.
Les trois cas "favorables" sont fondamentaux. On représentera les graphes de ces 3 modèles en dimension 2 au prochain cours.
Finissons en annonçant que les résultats pour l'optimisation sous contraintes seront similaires avec des subtilités supplémentaires dues à l'introduction des contraintes.

3 Un exemple

Voici maintenant un exemple de ce qu'il faut savoir en pratique pour déterminer les extrema d'une fonction.

Soit la fonction $f(x, y) = x^3 + y^3 - 3x\,y$ définie sur $\Omega = \mathbb{R}^2$.

3.1 Détermination des points stationnaires

On calcule le gradient de f en $\vec{x} = (x, y)$: $Df(\vec{x}) = \left(\frac{\partial f}{\partial x}(\vec{x}), \frac{\partial f}{\partial y}(\vec{x})\right) = (3x^2 - 3y, 3y^2 - 3x)$. Ainsi,

$$(x, y) \in \mathcal{S}(f) \Leftrightarrow \left| \begin{array}{l} 3x^2 - 3y = 0 \\ 3y^2 - 3x = 0 \end{array} \right. \Leftrightarrow \left| \begin{array}{l} x^2 = y \\ y^2 = x \end{array} \right. \Leftrightarrow \left| \begin{array}{l} x^4 - x = 0 \\ y = x^2 \end{array} \right.$$

$$\Leftrightarrow \left| \begin{array}{l} x(x^3 - 1) = 0 \\ y = x^2 \end{array} \right. \Leftrightarrow \left| \begin{array}{l} x = 0 \text{ ou } x = -1 \text{ ou } x = 1 \\ y = x^2 \end{array} \right. \Leftrightarrow \left| \begin{array}{l} (x, y) = (0, 0) \\ \text{ ou} \\ (x, y) = (1, 1) \end{array} \right.$$

Ainsi, $\mathcal{S}(f) = \{(0, 0); (1, 1)\}$. Ces deux points stationnaires sont les seuls extrema éventuels.

3.2 Nature du point stationnaire $(0,0)$

On calcule les dérivées partielles secondes pour déterminer la matrice hessienne :

$$D^2 f(\vec{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(\vec{x}) & \frac{\partial^2 f}{\partial x \partial y}(\vec{x}) \\ \frac{\partial^2 f}{\partial y \partial x}(\vec{x}) & \frac{\partial^2 f}{\partial y^2}(\vec{x}) \end{pmatrix} = \begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}. \tag{4}$$

Ainsi, $H = D^2 f(0,0) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}$ et donc, pour $\vec{u} = (u,v)$, $q_H(\vec{u}) = -6uv$. Or, en posant

$$\left| \begin{array}{l} a+b=u \\ a-b=v \end{array} \right. \quad \text{i.e.} \quad \left| \begin{array}{l} a = \frac{u+v}{2} \\ b = \frac{u-v}{2} \end{array} \right.$$

on obtient

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

puisque $uv = (a+b)(a-b) = a^2 - b^2$. Finalement, $q_H(u,v) = 6\left(\frac{u-v}{2}\right)^2 - 6\left(\frac{u+v}{2}\right)^2$ est de signature $(1,1)$. Ainsi, $(0,0)$ est un point col de f : ça n'est pas un extremum local de f.

3.3 Nature du point stationnaire $(1,1)$

On utilise le calcul (4) pour $(x,y) = (1,1)$. On a $H = H(1,1) = D^2 f(1,1) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}$. Ainsi, $q_H(u,v) = 6u^2 - 6uv + 6v^2 = 6(u^2 - uv + v^2)$. Or,

$$u^2 - uv + v^2 = \underbrace{\left[u^2 - 2u\left(\frac{v}{2}\right)\right]}_{\text{début d'un carré}} + v^2 = \left[\left(u - \frac{v}{2}\right)^2 - \left(\frac{v}{2}\right)^2\right] + v^2 = \left(u - \frac{v}{2}\right)^2 + \frac{3}{4}v^2.$$

D'où l'on tire que $q_H(u,v) = 6\left(u - \frac{v}{2}\right)^2 + \frac{9}{2}v^2$ est de signature $(2,0)$ et est donc définie positive : le point $(1,1)$ est un minimum local de f.