Mémento sur l'optimisation

1 Définitions générales

Résoudre un problème d'optimisation consiste à trouver $X \in \mathbb{R}$ qui minimise $f(x)$, où $f : \mathbb{R}^n \to \mathbb{R}$. L'ensemble d'arrivée doit être ordonné. De tels problèmes n'ont pas toujours de solution.
On dit *optimiser* ou *minimiser* alors qu'optimiser peut signifier minimiser ou maximiser. On privilégie la minimisation, par convention, sachant qu'un problème de maximisation se ramène à un problème de minimisation.

$$max_{X \in \mathbb{R}^n} f(X) = min_{X \in \mathbb{R}^n} - f(X)$$

f est appelée la **fonction objectif** (recherche opérationnelle) ou **fonctionnelle** (mathématiques) ou **énergie** (mécanique) ou **fonction de coût** (finance).

— Le problème s'écrit :

$$min_{X\mathbb{R}^n} f(X)$$

— une des solutions du problème :

$$X^* = argmin_{X \in \mathbb{R}^n} f(X)$$

— la valeur minimale de $f(X)$: $f(X^*) = min f(X)$ minimum.
On appelle souvent **minimum** le minimiseur.

2 Fonctions de n variables réelles

Soit f une fonction de n variables réelles à valeurs réelles.

$$\begin{array}{llll} f: & E \subset \mathbb{R}^n & \to & F \subset \mathbb{R} \\ & x = (x_1, ..., x_n) & \mapsto & f(x) = f(x_1, ..., x_n) \end{array}$$

2.1 Dérivées partielles d'ordre 1

2.1.1 Notation

Soit $\bar{x} = (\bar{x}_1, ..., \bar{x}_n) \in \mathbb{R}^n$ fixé. On note alors $\forall i \in \{1, ..., n\}$:

$$\bar{x}_{\setminus i} = (\bar{x}_1, ..., \bar{x}_{i-1}, x_i, \bar{x}_{i+1} ..., \bar{x}_n)$$

$\varphi_{i,\bar{x}_{\setminus i}}$ la fonction d'une variable réelle à valeurs réelles

$$\begin{array}{llll} \varphi_{i,\bar{x}_{\setminus i}}: & D_{\varphi_i,\bar{x}} \subset \mathbb{R} & \to & \mathbb{R} \\ & x & \mapsto & \phi(\bar{x}_1, ..., x_i, ..., \bar{x}_n) \end{array}$$

où $D_{\varphi_i,\bar{x}} = \{x_i \in \mathbb{R} / (\bar{x}_1, ..., x_i, ..., \bar{x}_n) \in \mathcal{D}_f$.

$\varphi_{i,\bar{x}_{\setminus i}}$ est la fonction qui, à la ième variable x_i, associe la valeur de la fonction f en considérant chacune des variables $x_k, k \neq i$, égale à $\bar{x}_k$ fixé.

Pour des raisons de légèreté de notation, nous ne préciserons pas les domaines de définition des fonctions par la suite.

2.1.2 Définition : Dérivées partielles, vecteur gradient, différentielle

Pour $i \in \{1, ..., n\}$, la **dérivée partielle** de f par rapport à x_i est la fonction de n variables

$$\frac{\partial f}{\partial x_i} \qquad \bar{x} = (\bar{x}_1, ..., \bar{x}_n) \mapsto \varphi'_{i,\bar{x}_{\setminus i}}(\bar{x}_i)$$

Autre notation : $\frac{\partial f}{\partial x_i} = \partial_i f$.

Si les dérivées partielles de f sont définies en $\bar{x}$, on définit le **vecteur gradient** de f en $\bar{x}$:

$$\nabla f(\bar{x}) = \left(\frac{\partial f}{\partial x_1}(\bar{x}), ..., \frac{\partial f}{\partial x_n}(\bar{x}) \right)^T \in \mathbb{R}^n$$

parfois noté $\vec{grad} f(\bar{x})$.

La **différentielle** de la fonction f au point $\bar{x}$ est la fonction de n variables définie sur $\mathbb{R}^n$

$$\begin{array}{lrcl} Df : & \mathbb{R}^n & \to & \mathbb{R} \\ & h = (h_1, ..., h_n) & \mapsto & \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(\bar{x}) h_i \right) = Df_{\bar{x}}(h) \end{array}$$

Remarque Pour le calcul de $\frac{\partial f}{\partial x_i}$, il suffit de dériver l'expression définissant f en la considérant comme une fonction de x_i, tous les autres x_j étant considérés comme des constantes.

2.2 Dérivée partielle d'ordre 2

Définition : Sous réserve d'existence des dérivées partielles de f, on définit les n^2 dérivées partielles d'ordre 2 par les fonctions de n variables

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i} \right) \qquad (i, j) \in \{1, ..., n\} \times \{1, ..., n\}$$

parfois noté $\partial_j \partial_i f$.

On note

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_i} \right) \quad \text{noté aussi } \partial_i^2 f$$

On définit la **matrice Hessienne** de f en $\bar{x} \in \mathbb{R}^n$ où toutes ses dérivées partielles d'ordre 2 sont bien définies, comme la matrice $n \times n$ réelle

$$H^f(\bar{x}) = \nabla^2 f(\bar{x}) = \left(\frac{\partial^2 f}{\partial x_j \partial x_i}(\bar{x}) \right)_{1 \leq i,j \leq n}$$

3 Optimisation sans contrainte

Les résultats ci-dessous sont présentés dans le cas de fonction à 2 variables mais ils peuvent être généralisés à n variables de la même façon.

3.1 Condition nécessaire d'optimalité

Théorème : CN 1er ordre

Soit D un domaine ouvert de $\mathbb{R}^2$, f une fonction une fois dérivable sur D. Si (a, b) est un extremum, alors le gradient de f en (a, b) est nul :

$$\nabla f(a, b) = 0 \Leftrightarrow \begin{cases} \frac{\partial f}{\partial x}(a, b) = 0 \\ \frac{\partial f}{\partial y}(a, b) = 0 \end{cases}$$

La nullité du gradient en (a, b) est une condition nécessaire, mais insuffisante, pour que f admette un extremum en (a, b). Les points où le gradient s'annule sont appelés **points critiques**.

3.2 Condition suffisante d'optimalité

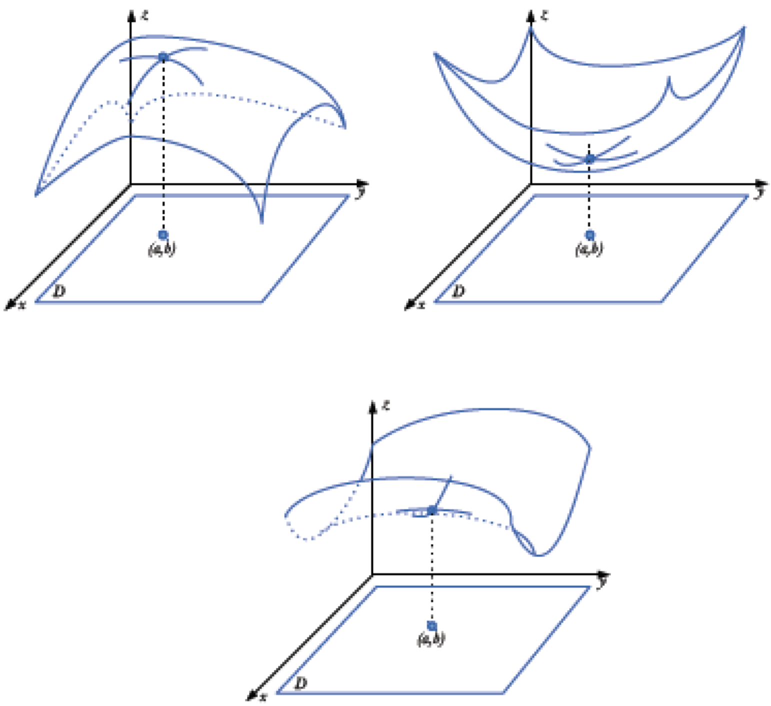
Théorème : CS 2ème ordre
Soit D un domaine ouvert de $\mathbb{R}^2$, f une fonction deux fois continûment différentiable sur D et (a, b) un point de D.

1. Si $\nabla f(a, b) = 0$ et si $H_f(a, b)$ a toutes ses valeurs propres strictement négatives, alors (a, b) est un **maximum local** pour f.

2. Si $\nabla f(a, b) = 0$ et si $H_f(a, b)$ a toutes ses valeurs propres strictement positives, alors (a, b) est un **minimum local** pour f.

3. Si $\nabla f(a, b) = 0$ et si $H_f(a, b)$ a certaines de ses valeurs propres strictement positives, toutes les autres strictement négatives alors (a, b) est un **point selle** ou un **point col** pour f.

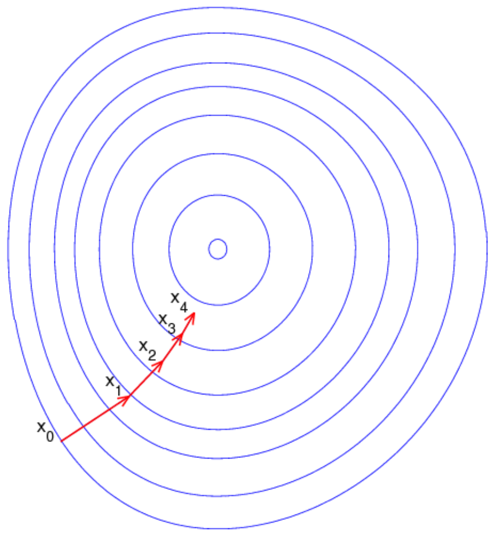
4. Si $\nabla f(a, b) = 0$ et si certaines valeurs propres de H_f sont nulles, alors on ne peut pas conclure.



Exemples de points critiques : maximum, minimum et point col

3.3 Algorithme du gradient

L'idée pour minimiser une fonction est de partir d'un point donné, $x^{(0)}$ et de suivre la plus grande pente (en descendant) pour aboutir à un point $x^{(1)}$ puis de recommencer jusqu'à aboutir à un point critique (où le gradient est nul). On rappelle que le vecteur gradient est perpendiculaire aux courbes de niveau de la fonction, dirigé dans le sens des niveaux croissants.



Exemple d'algorithme de descente de gradient : courbes de niveau en bleu, vecteur gradient en rouge.

3.3.1 Algorithme à pas fixe

Algorithme : $x^{(0)}, \epsilon, f, \nabla f, t$ donnés, $k = 0$.
Répéter :

1. $x^{(k+1)} = x^k - t\nabla f(x^{(k)})$
2. $k = k + 1$

jusqu'à $|f(x^{(k)}) - f(x^{k-1})| < \epsilon$.

3.3.2 Algorithme à pas optimal

Algorithme : $x^{(0)}, \epsilon, f, \nabla f$, et $k = 0$.
Répéter :

1. Calculer de $\nabla f(x^{(k)})$.
2. Trouver t qui minimise la coupe :

$$t \in \mathbb{R} \mapsto \gamma_k(t) = f(x^{(k)} + t\nabla f(x^{(k)})).$$

3. $x^{(k+1)} = x^k - t\nabla f(x^{(k)})$
4. $k = k + 1$

jusqu'à $|f(x^{(k)}) - f(x^{k-1})| < \epsilon$.