Optimisation

AMAL Youssef

ENSAT
Google Classroom: kzgrskw
yamal@uae.ac.ma

1445/2023

- Introduction
- Existence et unicité de minimum
- Algorithmes de minimisation sans contraintes
- Algorithmes de minimisation avec contraintes
- Calculs par le logiciel R

Référence Recommandée :

- *Numerical Optimization* by Jorge Nocedal et Stephen Wright.

- L'optimisation consiste en la recherche du minimum (ou du maximum) d'une certaine quantité sans ou avec contraintes :

$$(P) \quad \inf_{x \in C} f(x)$$

- On dit que problème (P) admet une solution s'il existe $x_0 \in C$ tel que

$$\forall x \in C,\ f(x_0) \leq f(x)$$

- Dans ce cas, $f(x_0) = \inf_{x \in C} f(x)$ est un minimum de f sur C.
- Les valeurs maximales de fonctions f sont obtenues en remplacant f par $-$f :

$$\sup_{x \in C} f(x) = \inf_{x \in C} f(x)$$

- **Exemple :** Une entreprise de production d'ordinateurs de types : laptops et ordinateurs de bureau. Les laptops se vendent à 10M Dhs chacun, tandis que les ordinateurs de bureau se vendent à 16M Dhs chacun. Cependant, la production de ces ordinateurs nécessite des ressources limitées en composants techniques.
- **Modèle Mathématique :** $$\begin{cases} \max(10Mx_1 + 16Mx_2) \\ \text{contraintes :} \\ x_1 + x_2 \leq 400 \\ x_1 + 2x_2 \leq 600 \\ x_1, x_2 > 0 \end{cases}$$
  où x_1 et x_2 sont les quantités respectives de laptops et d'ordinateurs de bureau à produire.
- **Objectif :** Maximiser les revenus de la production d'ordinateurs portables et de bureau, en respectant les limitations des ressources disponibles.

- Existence et Unicité du minimum,
  - liée à la continuité,
  - liée à la convexité (stricte).
- **Résolution du problème :**
  - Étude analytique,
  - Étude approchée par méthodes numériques.

- On se place dans $\mathbb{R}^N$ muni de la norme euclidienne $\|.\|$ et du produit scalaire $\langle ., . \rangle$ avec $\langle x, x \rangle = \|x\|^2$.
- Soit $U \subset \mathbb{R}^N$ un ouvert et $f : U \to \mathbb{R}$ une application scalaire. Soit $a \in U$.
- On dit que f est différentiable au point a s'il existe une application linéaire $df_a \in \mathcal{L}(\mathbb{R}^N, \mathbb{R})$ tels que $\exists \alpha > 0, \forall \|h\| \leq \alpha$,

$$f(a+h) = f(a) + df_a(h) + \|h\|\varepsilon(h)$$

avec $\lim\limits_{h \to 0} \varepsilon(h) = 0$.

- **Formule de Taylor - Young à l'ordre 2 :**

$$f(a+h) = f(a) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(a) + \sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a) + o(\|h\|^2)$$

où $\langle \nabla f(a), h \rangle = \sum_{i=1}^{n} h_i \dfrac{\partial f}{\partial x_i}(a)$

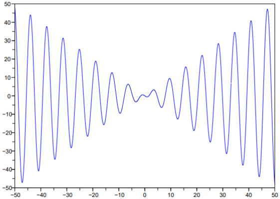
et $\sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j \dfrac{\partial^2 f}{\partial x_i \partial x_j}(a) =^t h.H_f(a).h$

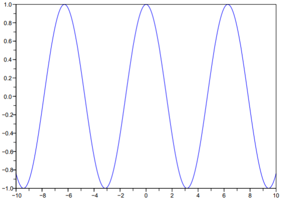
- Soit $x_0 \in K$. On dit que la fonction f admet
  - un minimum global sur K au point x_0, si

$$\forall x \in K,\ f(x_0) \leq f(x)$$

  - un minimum local sur K au point x_0, si

$$\exists r > 0,\ x \in B(x_0, r) \cap K,\ f(x_0) \leq f(x).$$



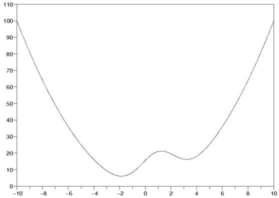


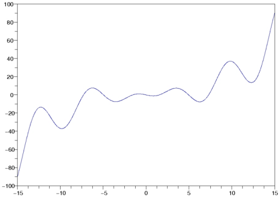
- Soit K un compact de $\mathbb{R}^N$ et $f : K \to \mathbb{R}$ une application continue sur K. Alors f est bornée et atteint ses bornes :

$$\exists x_0 \in K \text{ tel que } \inf_{x \in K} f(x) = f(x_0)$$

- Contre-exemples:
  - Soient $X = \mathbb{R}$ et $f(x) = x^2 + 1$ sur $\mathbb{R}^*$ et $f(0) = 3$, on a $\inf_{x \in \mathbb{R}} f(x) = ?$
  - Soient $X =]0; 1]$ et $f(x) = x^2 + 1$, on a $\inf_{x \in \mathbb{R}} f(x) = ?$

- **Fonctions coercives** : Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dite coercive si $\lim\limits_{\|x\| \to +\infty} f(x) = +\infty$.





- Soient U une partie non vide fermée non bornée de $\mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}$ une fonction coercive et continue. Alors il existe au moins un élément $x_0 \in U$ tel que $\inf\limits_{x \in U} f(x) = f(x_0)$.
- C.exp : Soit $X = \mathbb{R}$ et $f(x) = -x^2$, alors $\inf\limits_{x \in \mathbb{R}} f(x) = ?$

- Soit $K \subset \mathbb{R}^N$. L'ensemble X est dite convexe si :

$$\forall (x,y) \in K^2,\ \forall t \in [0,1] \text{ tels que } tx + (1-t)y \in K$$

- Exemple : $\mathbb{R}^N$ est un convexe.
- Soit $K \subset \mathbb{R}^N$ convexe et $f : K \to \mathbb{R}$.
  - f est dite convexe si :

$$\forall (x,y) \in K^2,\ \forall t \in]0,1[,\ f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

  - f est dite strictement convexe si :

$$\forall (x,y) \in K^2,\ x \neq y,\ \forall t \in]0,1[,\ f(tx+(1-t)y) < tf(x)+(1-t)f(y)$$

- Exemple : toute fonction affine, $f(x) = ax + b$, est convexe mais non strictement convexe.

- **Critéres de convexité :** On suppose que f est deux fois différentiable en tout point de K. On a équivalence entre :
    1. f convexe sur K.
    2. $\forall (u,v) \in K^2$, ${}^t(v-u).H_f(u).(v-u) \geq 0$ ($H_f(u)$ est semi-définie positive).
- **Critéres de convexité stricte :** :
    - Si $\forall (u,v) \in K^2, u \neq v$, ${}^t(v-u).H_f(u).(v-u) > 0$ ($H_f(u)$ est définie positive), alors f est strictement convexe sur K.
- **Théorème (critère de Sylvester) :** Pour qu'une matrice $H_f = (a_{ij})_{1\leq i,j\leq n}$ réelle symétrique soit définie positive, il faut et suffit que les n sous matrices mineurs principaux $H_f^p = (a_{ij})_{1\leq i,j\leq p}$ de H_f aient leur déterminant strictement positif pour tout $p = 1, \ldots, n$.

- Soient K un ouvert de $\mathbb{R}^N$ et $f : K \to \mathbb{R}$ différentiable en a,
- On dit qu'un point a est un point critique de f si $\nabla f(a) = 0$.
- Soit f une fonction de classe C^2 sur un **voisinage** de a. $H_f(a)$ est alors une matrice symétrique réelle dont les valeurs propres, nécessairement réelles, sont ordonnées comme suit: $\lambda_{min} = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$. On a alors :
  - $H_f(a)$ est semi-définie positive si et seulement si $\lambda_{min} \geq 0$.
  - $H_f(a)$ est définie positive si et seulement si $\lambda_{min} > 0$.
- Si $\lambda_{min} > 0$ alors f admet un minimum local en a.

- Si f est convexe sur K et si elle admet un point critique en $a \in K$ vérifiant $\nabla f(a) = 0$, alors f admet un minimum local et global en a sur K.
- **C. Exemple :** fonctions affines.
- Soit $K \subset \mathbb{R}^N$ un ensemble non vide et convexe. Si $f : K \to \mathbb{R}$ admet un minimum local en u sur K. Alors,
  - Si f est convexe alors f admet un minimum global en u sur K.
  - Si f est strictement convexe alors alors u est l'unique point de minimum global de f sur K.
- **C. Exemple :** $f(x) = e^x$.

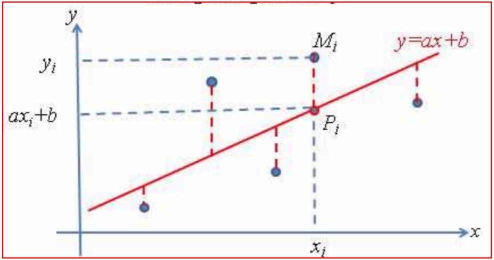
- **Exemple :** Étudier le problème de minimisation des fonctions suivantes :
  1. $f(x, y) = x^2 + y^2 + 3xy - y.$
  2. $g(x, y) = x^2 + y^3 - 2xy - y.$

- Considérons un ensemble de données expérimentales représentant le temps nécessaire à l'exécution d'un algorithme en fonction de la taille de l'entrée.
- Modéliser cette relation à l'aide de la régression linéaire afin de prédire le temps d'exécution pour de nouvelles tailles d'entrée :

Tailles d'entrée	Temps d'exécution
10	5
20	12
30	21
40	35
50	48

Objectif : Trouver une droite de régression $y = ax + b$ pour un ensemble de données $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$.

- **Fonction Objectif :** $f(a,b) = \sum_{i=1}^{n}(y_i - (ax_i + b))^2$
- Le problème est la minimisation de la fonction f : $\inf\limits_{(a,b)\in\mathbb{R}^2} f(a,b)$,

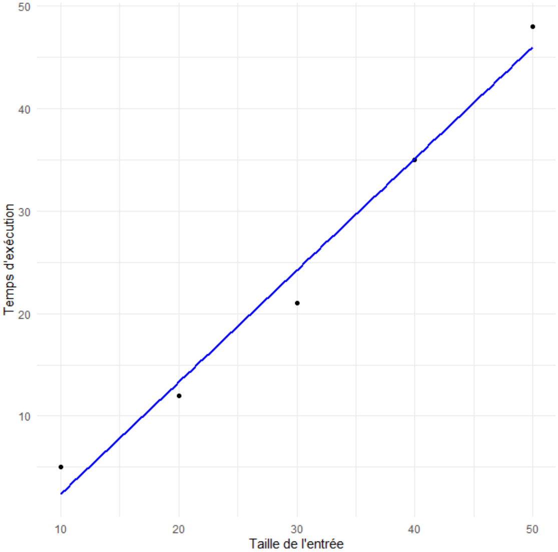


- Point critique solution de $\nabla f(a,b) = (0,0)$:

$$a_0 = \frac{\overline{xy} - \overline{x}.\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{Cov(x,y)}{V(x)},\ b_0 = \overline{y} - a\overline{x}$$

- On a $H_f(a,b) = \begin{pmatrix} 2\overline{x^2} & 2\overline{x} \\ 2\overline{x} & 2 \end{pmatrix}$ qui est définie positive pour tout $(a,b) \in \mathbb{R}^2$ avec $a \neq b$.
- (a_0, b_0) est l'unique point de minimum global de f sur $\mathbb{R}^2$.

- **Exemple :** Le calcul donne $a = 1.09$ et $b = -8.5$.



- **Hypothèses** : $f : \mathbb{R}^n \to \mathbb{R}^n$ strictement convexe, de classe C^2.
- **Objectif** : Trouver numériquement $x^* \in \mathbb{R}^n$ tel que $f(x^*) = \min_{x \in \mathbb{R}^n} f(x)$.
- Le problème se ramène à résoudre le système $\nabla f(x) = 0_{\mathbb{R}^n}$.

- **Rappel (Dérivée directionnelle) :**
  - $\lim\limits_{t \to 0} \dfrac{f(x+t.d) - f(x)}{t} = D_d f(x)$,
  - $D_d f(x) = \langle \nabla f(x), d \rangle$.
- **Direction de descente :** Une direction de descente de f en x est un vecteur $d \in \mathbb{R}^n$ tel que

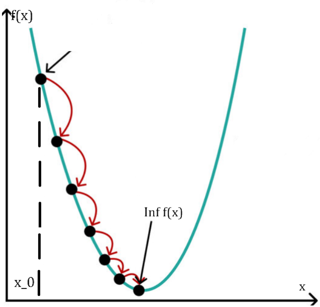
$$D_d f(x) < 0 \text{ ou encore } \langle \nabla f(x), d \rangle < 0$$

- c.à.d : $\exists \alpha > 0, \forall 0 < t < \alpha : f(x+td) < f(x)$.

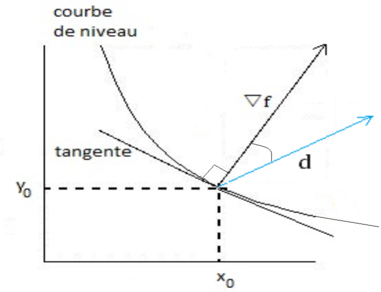
- Construction d'une suite $(x_k)_{k\in\mathbb{N}}$ vérifiant

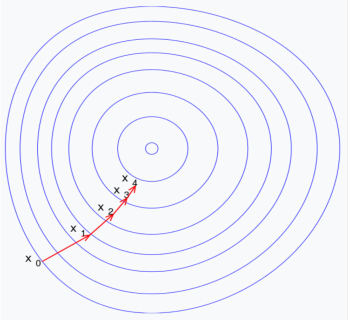
$$f(x_{k+1}) \leq f(x_k)$$

- Pour x_0 choisi arbitrairement, on a : $x_{k+1} = x_k + \rho_k d_k$, avec
  - d_k : la direction descente.
  - ρ_k : le pas de la k-ième itération.



- **Remarque :** $d = -\nabla f(x)$ est la direction de plus forte descente.





- **Algorithme de gradient :**
  1. Choisir x_0 , $tol > 0$, $k \leq k_{max}$
  2. Calculer $x_{k+1} = x_k - \rho_k \nabla f(x_k)$, et $k = k+1$,
     tant que $\|x_{k+1} - x_k\| > tol$ (ou $\|\nabla f(x_k)\| > tol$) et $k \leq k_{max}$.
- Reste à préciser ρ_k :
  - Algorithme de gradient à pas fixe : $\rho_k = \rho$
  - Algorithme de gradient à pas optimal :

$$f(x_k - \rho_k \nabla f(x_k)) = \min_{t>0} f(x_k - t\nabla f(x_k)).$$

- Théoriquement, pour un pas $\rho_k > 0$ assez petit, la suite $(x_k)_k$ converge et la convergence est au moins géométrique, c.à.d :

$$\exists \beta \in]0, 1[\text{ tel que } \|x_k - x^*\| \leq \beta^k \|x_0 - x^*\|, \ \forall k \in \mathbb{N}$$

- Si le pas est trop grand, la descente vers le minimum est rapide mais le risque d'osciller autour du minimum sans converger est élevé.
- Le pas devrait être assez petit pour converger vers le minimum mais pas trop petit sinon le coût numérique sera très élevé.
- Le choix du pas fixe convenable est obtenu par le test de plusieurs valeurs.

- **Recherche exacte de pas optimal ρ_k solution de :**
  $\min_{t>0} f(x_k - t\nabla f(x_k))$
- **Exemple :** Soit $f(x,y) = \frac{1}{2}x^2 + \frac{7}{2}y^2$.
  - $\nabla f(x_k, y_k) = (x_k, 7y_k)$
  - $\rho_k = \dfrac{x_k^2 + 49y_k^2}{x_k^2 + 343y_k^2}$
  - $x_{k+1} = x_k - \rho_k x_k$
  - $y_{k+1} = y_k - \rho_k y_k$

- Résoudre le problème d'optimisation suivant :

$$\inf_{x \in \mathbb{R}^2} f(x)$$

où

$$f(x,y) = x^4 - 2x^3 + \frac{1}{6}y^2 - 3x - xy$$

- Considérons la Fonction objectif de forme quadratique :

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$$

  oú A est une matrice symétrique définie positive.
- La fonction f est strictement convexe et de classe $C^{+\infty}$,
- Le minimum global de f est atteint en x^* tel que $Ax^* = b$, en effet :
  - $\nabla f(x) = Ax - b$ et $H_f(x) = A$.
- La méthode du gradient conjugué est une méthode itérative directe pour résoudre l'équation $Ax = b$.

- **Algorithme du gradient conjugué :** Pour tout $k \geq 0$, $r_k = -\nabla f(x_k) = b - Ax_k$,
  1. Choisir un vecteur initial x_0, poser $d_{-1} = 0 \in \mathbb{R}^n$ et calculer $r_0 = b - Ax_0$,
  2. Pour $k = 0, 1, \ldots, n$ ou jusqu'à la convergence :
  3. Si $r_k > tol$ alors $x_k = x^*$ arrêt. Sinon,
  4. $d_k = r_k + \alpha_{k-1} d_{k-1}$ avec α_{k-1} tel que $\langle Ad_k, d_{k-1} \rangle = 0$,
  5. $x_{k+1} = x_k + t_k d_k$ où t_k est tel que $f(x_k + t_k d_k) = \min\limits_{t \geq 0} f(x_k + td_k)$.

  Il en résulte que :
- $\alpha_k = -\dfrac{\langle Ar_k, d_k \rangle}{\langle Ad_k, d_k \rangle}$ et $t_k = \dfrac{\langle r_k, d_k \rangle}{\langle Ad_k, d_k \rangle}$.
- La méthode consiste à construire une suite $(d_k)_k$ A-orthogonale formant une base de $\mathbb{R}^n$ dans le cas où la convergence est atteint en n itérations.

- Minimiser la fonction suivante, en utilisant la méthode des gradients conjugués :

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