

Chapitre 1 : Optimisation sans contraintes

H. Ben Majdoub

Janvier 2023

L'optimisation est une branche des mathématiques, dont le but est de trouver analytiquement ou numériquement, la meilleure solution (l'optimale) à un problème donné.

De nos jours, l'optimisation joue un rôle très important dans différents domaines de la vie.

- Transport et livraisons.
- Fabrication et production.
- Agriculture et génie civil.
- Finance, vente et marketing
- Gestion de stock
- Recherche et gestion des bases de données.

En optimisation, on parle de la fonction coût (objectif). C'est la fonction à minimiser/maximiser, après formulation mathématique du problème. On distingue deux grandes familles de techniques d'optimisation, et cela suivant le problème posé :

- Techniques d'optimisation sans contraintes.
- Techniques d'optimisation sous contraintes.

Définition :

Un problème d'optimisation s'écrit généralement sous la forme suivante :

$$\begin{cases} \text{trouver } x^* \in \Omega \text{ tel que} \\ f(x^*) \leq f(x) , \quad \forall x \in \Omega \end{cases}$$

Autrement dit,
$$f(x^*) = \min_{x \in \Omega} f(x)$$

Avec $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue

f est appelée fonction objectif

Théorème :

Soit la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable sur \mathbb{R}^n si f a un minimum local (ou maximum local) au point x^* alors :

$$\nabla f (x^*) = 0$$

Rappel : le gradient d'une fonction $f(x_1, x_2, \dots, x_n)$ à plusieurs

Variables est le vecteurs de composantes les dérivées partielles $\frac{\partial f}{\partial x_i} (i = 1, 2, \dots, n)$

C'est-à-dire ,

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Définition :

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ différentiable. Tout point $x \in \mathbb{R}^n$ vérifiant :

$$\nabla f(x) = 0$$

est appelé point critique de f .

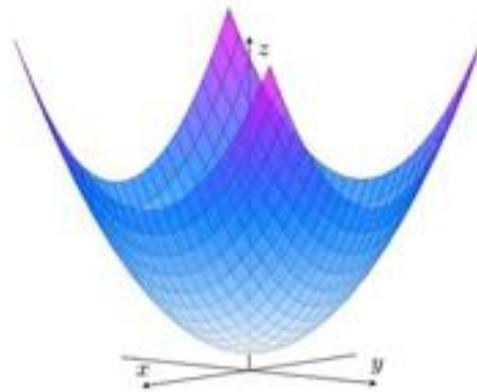
Exemples :

- La fonction $x \mapsto x^2$ admet un point critique en $x = 0$ qui est aussi un minimum local.
- La fonction $x \mapsto -(x-1)^2$ admet un point critique en $x = 1$ qui est aussi un maximum local.
- La fonction $x \mapsto x^3$ admet un point critique en $x = 0$ qui n'est ni un minimum local, ni un maximum local.

Exemple 1

On considère la fonction $f(x, y) = 1 + x^2 + y^2$ définie sur \mathbb{R}^2

Il est facile à deviner que f admet un minimum global en $(x_0, y_0) = (0, 0)$



Vérifions le théorème en montrant que ce point est un point critique de f .

On calcule donc les dérivées partielles :

$$\frac{\partial f}{\partial x}(x, y) = 2x \quad , \quad \frac{\partial f}{\partial y}(x, y) = 2y$$

Donc : $\frac{\partial f}{\partial x}(0, 0) = 0$ et $\frac{\partial f}{\partial y}(0, 0) = 0$ ce qui confirme le théorème !

Exemple 2

Cet exemple démontre qu'un point critique n'est pas toujours un extremum.

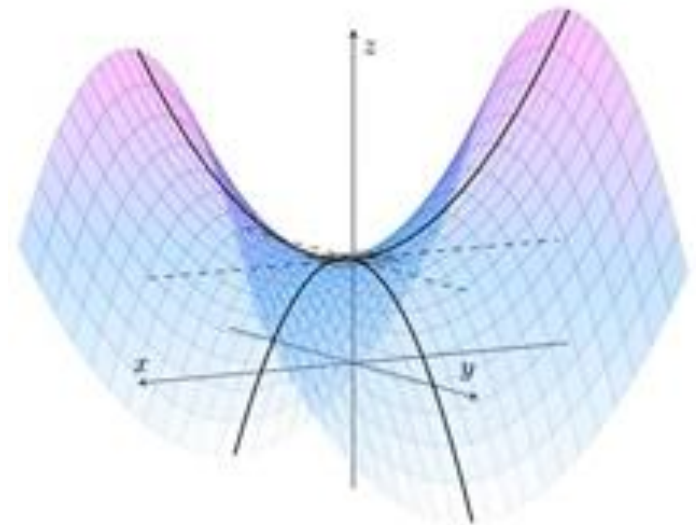
Soit $f(x, y) = 1 + x^2 - y^2$ définie sur \mathbb{R}^2

On a : $\frac{\partial f}{\partial x}(x, y) = 2x$ et $\frac{\partial f}{\partial y}(x, y) = -2y$

Donc : $\frac{\partial f}{\partial x}(0, 0) = 0$ et $\frac{\partial f}{\partial y}(0, 0) = 0$

D'où $(x_0, y_0) = (0, 0)$ est un point critique de f .
Cependant, on voit clairement que ce point ne correspond pas à un extremum.

On dit que $(0, 0)$ est un point **selle** de f



Théorème : (Existence)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application et soit le problème $(P) : \min_{x \in \mathbb{R}^n} f(x)$

Si f est **continue** et **coercive** (*c'est-à-dire* $f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$)

Alors (P) admet au moins une solution .

Théorème : (unicité)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application et soit le problème $(P) : \min_{x \in \mathbb{R}^n} f(x)$

Si f est **strictement convexe** alors (P) admet au plus une solution.

Théorème (existence et unicité)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et le problème $(P) : \min_{x \in \mathbb{R}^n} f(x)$

Si f est **continue, coercive et strictement convexe**

Alors le problème (P) admet une **unique** solution.

Rappel (convexité) soit $f : E \subset \mathbb{R}^n \rightarrow \mathbb{R}$

- f est convexe ssi :

$$\forall (x, y) \in E^2, \quad \forall \lambda \in]0, 1[\quad , \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

- f est strictement convexe ssi :

$$\forall (x, y) \in E^2, x \neq y \quad , \quad \forall \lambda \in]0, 1[\quad , \quad f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$$

Théorème: (caractérisation différentielle de la convexité)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^2 . On note H_f sa matrice hessienne.

- Si H_f est semi-définie positive pour tout $x \in \mathbb{R}^n$, alors f est convexe.
- Si H_f est définie positive pour tout $x \in \mathbb{R}^n$, alors f est strictement convexe.

La matrice hessienne de la fonction f est une matrice de dimension $n \times n$ dont les éléments sont les dérivées partielles d'ordre 2 de f .

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Exemple

- Soit $f(x, y) = x^2 + xy + y^2$ et le problème $(P): \min_{(x,y) \in \mathbb{R}^2} f(x, y)$

Montrer que le problème (P) admet une unique solution et déterminer cette solution.

- f est continue et coercive car :

$$2xy \geq -x^2 - y^2 \Rightarrow xy \geq \frac{1}{2}(-x^2 - y^2) \quad \text{donc} \quad f(x, y) \geq \frac{1}{2}\|(x, y)\|^2$$

d'où l'existence de la solution de (P)

- La matrice hessienne de f est $H_f = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ qui est définie positive car ses valeurs propres sont 1 et 3 d'où l'unicité de la solution de (P)
- La solution de (P) vérifie l'équation :

$$\nabla f(x, y) = 0 \Rightarrow \begin{cases} 2x + y = 0 \\ 2y + x = 0 \end{cases} \Rightarrow (x, y) = (0, 0)$$

Définition :

On appelle fonction quadratique une fonction de \mathbb{R}^n dans \mathbb{R} définie par :

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

Où $A \in M_n(\mathbb{R})$, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$

Exemple : $f(x, y, z) = \frac{1}{2}x^2 + \frac{1}{2}y^2 + \frac{3}{2}z^2 + xz + yz - x - y - z$

$$f(x, y, z) = \frac{1}{2} \left\langle \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle - \left\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle$$

Donc f est quadratique avec $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ et $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Proposition :

Soit f une fonction quadratique tel que $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c$

Si A est symétrique définie et positive alors le problème quadratique

$(P) : \min_{x \in \mathbb{R}^n} f(x)$ admet une unique solution x^* vérifiant : $Ax^* = b$

Exemple : Soit $f(x, y, z) = x^2 + y^2 + z^2 - xy - yz + 3x - y - 2z$

Montre que f admet un unique minimum sur \mathbb{R}^3 . Calculer le minimum de f

$$f(x, y, z) = \frac{1}{2} \left\langle \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle - \left\langle \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right\rangle$$

Donc f est quadratique avec :

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \quad \text{et} \quad b = \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix}$$

Problème de minimisation quadratique

$$\begin{aligned}\det(A - \lambda I_3) &= \begin{vmatrix} 2-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 2-\lambda \end{vmatrix} = (2-\lambda) \begin{vmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{vmatrix} + \begin{vmatrix} -1 & -1 \\ 0 & 2-\lambda \end{vmatrix} \\ &= (2-\lambda)((2-\lambda)^2 - 1) - (2-\lambda) = (2-\lambda)(2-\sqrt{2}-\lambda)(2+\sqrt{2}-\lambda)\end{aligned}$$

Donc les valeurs propres de A sont : $\lambda_1 = 2$, $\lambda_2 = 2 - \sqrt{2}$ et $\lambda_3 = 2 + \sqrt{2}$

Qui sont toutes strictement positives donc A est symétrique définie positive

D'où f admet un unique minimum solution de l'équation :

$$A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = b \Leftrightarrow \begin{pmatrix} 2x - y \\ -x + 2y - z \\ -y + 2z \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix} \Leftrightarrow \begin{cases} 2x - y = -3 \\ -x + 2y - z = 1 \\ -y + 2z = 2 \end{cases}$$

Donc le minimum de f est atteint au point : $\begin{pmatrix} -5/4 \\ 1/2 \\ 5/4 \end{pmatrix}$

Définition : soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction, et soit $x \in \mathbb{R}^n$.

On dit que $d \in \mathbb{R}^n \setminus \{0\}$ est une direction de descente de f en x s'il existe $\alpha_0 > 0$

tel que :

$$f(x + \alpha d) \leq f(x) \quad , \quad \forall \alpha \in [0, \alpha_0]$$

Ainsi une méthode de descente pour la recherche de x^* solution du problème

$(P) : \min_{x \in \mathbb{R}^n} f(x)$ consiste à construire $(x^{(k)})_{k \in \mathbb{N}}$ de la manière suivante :

(1) Initialisation de $x^{(0)} \in \mathbb{R}^n$

(2) Itération k ($k \geq 0$)

(i) On cherche $d^{(k)}$ direction de descente en $x^{(k)}$

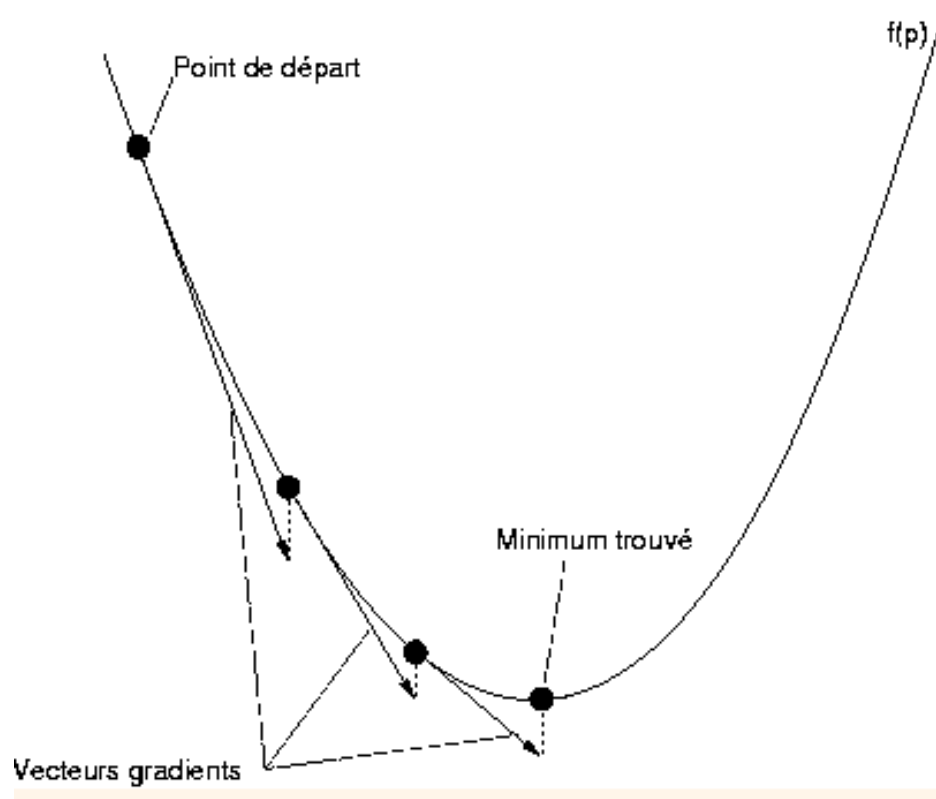
(ii) On calcule $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$ avec $\alpha_k > 0$ (le pas de descente)

Caractérisation des directions de descente

Proposition : Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction différentiable ,

$$x \in \mathbb{R}^n \text{ et } d \in \mathbb{R}^n \setminus \{0\} .$$

- (1) Si d est une direction de descente en x alors : $\langle \nabla f(x), d \rangle \leq 0$
- (2) Si $\nabla f(x) \neq 0$ alors $d = -\nabla f(x)$ est une direction de descente en x .



Algorithme du gradient à pas fixe

La méthode du gradient à pas fixe $\alpha > 0$ consiste à choisir comme direction à l'étape $k+1$, $d^{(k+1)} = -\nabla f(x^{(k)})$ ainsi l'algorithme s'écrit comme suit :

$$\begin{cases} \text{Initialisation : } x^{(0)} \in \mathbb{R}^n \\ \text{Pour } k \geq 0 : x^{(k+1)} = x^{(k)} - \alpha \nabla f(x^{(k)}) \end{cases}$$

Exemple : $f(x, y) = \ln(x^2 + y^2 + 1)$

En prenant $(x_0, y_0) = (1, 1)$ donner trois itérations de l'algorithme du gradient à pas fixe $\alpha = 0,5$

$$\nabla f(x, y) = \begin{pmatrix} \frac{2x}{x^2 + y^2 + 1} \\ \frac{2y}{x^2 + y^2 + 1} \end{pmatrix} \text{ donc } X^{(k+1)} = X^{(k)} - \alpha \nabla f(X^{(k)}) = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \begin{pmatrix} \frac{x_k}{x_k^2 + y_k^2 + 1} \\ \frac{y_k}{x_k^2 + y_k^2 + 1} \end{pmatrix}$$

Algorithme du gradient à pas fixe

$$\text{D'où} \quad X^{(k+1)} = \begin{pmatrix} x_k \left(1 - \frac{1}{x_k^2 + y_k^2 + 1}\right) \\ y_k \left(1 - \frac{1}{x_k^2 + y_k^2 + 1}\right) \end{pmatrix}$$

$$\text{Et par suite :} \quad X^{(1)} = \begin{pmatrix} x_0 \left(1 - \frac{1}{x_0^2 + y_0^2 + 1}\right) \\ y_0 \left(1 - \frac{1}{x_0^2 + y_0^2 + 1}\right) \end{pmatrix} = \begin{pmatrix} 2/3 \\ 2/3 \end{pmatrix}$$

$$X^{(2)} = \begin{pmatrix} x_1 \left(1 - \frac{1}{x_1^2 + y_1^2 + 1}\right) \\ y_1 \left(1 - \frac{1}{x_1^2 + y_1^2 + 1}\right) \end{pmatrix} = \begin{pmatrix} 16/51 \\ 16/51 \end{pmatrix}$$

$$X^{(3)} = \begin{pmatrix} x_2 \left(1 - \frac{1}{x_2^2 + y_2^2 + 1}\right) \\ y_2 \left(1 - \frac{1}{x_2^2 + y_2^2 + 1}\right) \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix}$$

Algorithme du gradient à pas fixe

Données : Un point initial $x^{(0)}$, un seuil de tolérance $\varepsilon > 0$ et un pas fixe $\alpha > 0$

Résultat : Un point $x \in \mathbb{R}^n$ proche de x^*

Initialiser x :

$$x \leftarrow x^{(0)}$$

$$k \leftarrow 0$$

tant que $\|\nabla f(x)\| > \varepsilon$ **faire**

Mettre à jour x avec le pas fixe α dans la direction de descente $-\nabla f(x^{(k)})$

$$x \leftarrow x - \alpha \nabla f(x)$$

$$k \leftarrow k + 1$$

Fin

Remarques : les conditions d'arrêt qu'on peut définir sont :

(1) $\|\nabla f(x^{(k)})\| < \varepsilon$ où ε est le seuil de tolérance

(2) $\|x^{(k+1)} - x^{(k)}\| < \text{précision tolérée}$

Algorithme du gradient à pas optimal

L'idée de l'algorithme du gradient à pas optimal est d'essayer de calculer à chaque itération le pas α_k qui minimise la fonction $\alpha \mapsto f\left(x^{(k)} + \alpha d^{(k)}\right)$

Donc le calcul du pas α_k revient à chercher α tel que :

$$f(x + \alpha d) \leq f(x + r d) \quad , \quad \forall r \in \mathbb{R}$$

$$\left\{ \begin{array}{l} \textit{Initialisation} : x^{(0)} \in \mathbb{R}^n \\ \textit{Pour } k \geq 0 : \textit{On cherche } \alpha_k \textit{ qui minimise la fonction} \\ \qquad \qquad \qquad \varphi(\alpha) = f\left(x^{(k)} - \alpha \nabla f\left(x^{(k)}\right)\right) \\ \qquad \qquad \qquad x^{(k+1)} = x^{(k)} - \alpha_k \nabla f\left(x^{(k)}\right) \end{array} \right.$$

Algorithme du gradient à pas optimal

Exemple : $f(x, y) = 4x^2 + 6y^2 + 6xy + 3x + 4y + 6$

En prenant $X^{(0)} = (0, 0)$ donner la première itération de l'algorithme du gradient à pas optimal,

$$\nabla f(x, y) = \begin{pmatrix} 8x + 6y + 3 \\ 12y + 6x + 4 \end{pmatrix}$$

$$X^{(1)} = X^{(0)} - \alpha_0 \nabla f(X^{(0)}) = \begin{pmatrix} -3\alpha_0 \\ -4\alpha_0 \end{pmatrix} \quad \text{où } \alpha_0 \text{ minimise la fonction}$$

$$\varphi(\alpha) = f(-3\alpha, -4\alpha) = 204\alpha^2 - 25\alpha + 6$$

$$\varphi'(\alpha) = 408\alpha - 25 \Rightarrow \alpha_0 = \frac{25}{408} = 0,0613$$

$$\text{D'où } X^{(1)} = \begin{pmatrix} -0,1838 \\ -0,245 \end{pmatrix}$$

Algorithmes du gradient pour une fonction quadratique

Proposition : Si $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c$ où A est une matrice symétrique définie positive, alors la méthode du gradient à pas fixe converge ssi le pas $\alpha \in \left] 0, \frac{2}{\rho(A)} \right[$ où $\rho(A) = \sup_{\lambda_i \in Sp(A)} \lambda_i$

Exemple : $f(x, y) = 4x^2 + 6y^2 + 6xy + 3x + 4y + 6$

Pour quelles valeurs du pas α on a convergence de l'algorithme du gradient à pas fixe ?

$$f(x, y) = \frac{1}{2} \langle AX, X \rangle - \langle b, X \rangle + c \quad \text{où} \quad X = \begin{pmatrix} x \\ y \end{pmatrix}, \quad A = \begin{pmatrix} 8 & 6 \\ 6 & 12 \end{pmatrix}, \quad b = \begin{pmatrix} -3 \\ -4 \end{pmatrix} \quad \text{et} \quad c = 6$$

$$\det(A - \lambda I_2) = \begin{vmatrix} 8 - \lambda & 6 \\ 6 & 12 - \lambda \end{vmatrix} = (8 - \lambda)(12 - \lambda) - 36 = \lambda^2 - 20\lambda + 60$$

$$\text{Donc} \quad Sp(A) = \{10 - 2\sqrt{10}, 10 + 2\sqrt{10}\}$$

$$\text{D'où l'algorithme du gradient à pas fixe converge ssi} \quad \alpha \in \left] 0, \frac{2}{10 + 2\sqrt{10}} \right[$$

Algorithmes du gradient pour une fonction quadratique

Proposition : Si $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c$ où A est une matrice symétrique définie positive, alors le pas optimal α_k à l'itération k pour l'algorithme du gradient à pas optimal est donné par :

$$\alpha_k = \frac{-\langle Ax^{(k)} - b, d^{(k)} \rangle}{\langle Ad^{(k)}, d^{(k)} \rangle} \quad \text{où} \quad d^{(k)} = -\nabla f(x^{(k)})$$

Exemple : $f(x, y) = \frac{3}{2}x^2 + \frac{3}{2}y^2 + 2xy - x - y$

En prenant $X^{(0)} = (0, 0)$ donner la première itération de l'algorithme du gradient à pas optimal,

f est une fonction quadratique avec $A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$ et $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$\nabla f(x, y) = \begin{pmatrix} 3x + 2y - 1 \\ 3y + 2x - 1 \end{pmatrix} \quad \text{donc} \quad d^{(0)} = -\nabla f(0, 0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Algorithmes du gradient pour une fonction quadratique

$$\text{Donc } \alpha_0 = \frac{-\langle Ax^{(0)} - b, d^{(0)} \rangle}{\langle Ad^{(0)}, d^{(0)} \rangle}$$

$$Ax^{(0)} - b = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \quad \text{donc } -\langle Ax^{(0)} - b, d^{(0)} \rangle = 2$$

$$\text{et } Ad^{(0)} = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \end{pmatrix} \quad \text{donc } \langle Ad^{(0)}, d^{(0)} \rangle = 10$$

$$d'où \quad \alpha_0 = \frac{2}{10} = 0,2$$

$$x^{(1)} = x^{(0)} - \alpha_0 \nabla f(x^{(0)}) = \begin{pmatrix} 0,2 \\ 0,2 \end{pmatrix}$$

Généralement, les méthodes de gradient ne sont pas très performantes parce qu'elles ne tiennent pas compte de la courbure (ou de la Hessienne) qui est une information de second ordre.

Méthode de Newton :

$$\left\{ \begin{array}{l} \textit{Initialisation} : x^{(0)} \in \mathbb{R}^n \\ \textit{Pour } k \geq 0 : x^{(k+1)} = x^{(k)} - \left(H_f(x^{(k)}) \right)^{-1} \nabla f(x^{(k)}) \end{array} \right.$$

En pratique, lorsque la matrice Hessienne est de très grande taille et mal conditionnée, le calcul de son inverse est un peu difficile, donc on peut utiliser la méthode suivante :

$$x^{(k+1)} = x^{(k)} + d^{(k)}$$

où $d^{(k)}$ est l'unique solution du système linéaire : $H_f(x^{(k)}) d^{(k)} = -\nabla f(x^{(k)})$

$d^{(k)}$ est appelée direction de Newton

Algorithme de descente de gradient à pas fixe