Université de Med Boudiaf M'sila
Faculté des Technologies
Département de Génie Mécanique

Année Universitaire 2016/2017
Option : Fabrication et Construction
Master 1ère Année

Chapitre 2 : Optimisation non linéaire sans contraintes

2.1 Introduction :

Soit un domaine $D \subseteq R^n$ et une fonction $f : D \rightarrow R$ de n variables. Dans un problème de minimisation, on cherche un point $x \in D$ qui minimise la valeur de f (x) sur D. On écrit :

$\min f(x) \quad (x \in D)$

La fonction **f** est appelée **la fonction objectif**.

En **optimisation sans contraintes**, on considère $D = R^n$. On cherche donc à résoudre

$\min f(x). \; x \in R^n$

L'étude des problèmes d'optimisation sans contraintes trouve des applications dans la résolution des systèmes non linéaires. Une grande classe d'algorithmes que nous allons considérer dans la résolution du problème d'optimisation sans contraintes, ils ont la forme générale suivante :

x_0 étant donnée, calculer $x_{k+1} = x_k + \alpha_k d_k$;

Ou, le vecteur d_k s'appelle la direction de descente,
α_k le pas de la méthode à la k-iéme itération.

En pratique, on s'arrange presque toujours pour avoir l'inégalité suivante :

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

qui assure la décroissance suffisante de la fonction objectif f. De tels algorithmes sont souvent appelés méthodes de descente. La différence essentielle entre ces algorithmes réside dans le choix de la direction de descente dk. Cette direction étant choisie et nous sommes plus où moins ramenés à un problème unidimensionnel pour la détermination de α_k: Pour s'approcher de la solution optimale du problème (P) (dans le cas général, c'est un point en lequel ont lieu peut être avec une certaine précision les conditions nécessaires d'optimalité de f), on se déplace naturellement à partir du point x_k dans la direction de la décroissance de la fonction f.
L'optimisation sans contraintes a les propriétés suivantes :

- toutes les méthodes nécessitent un point de départ x_0 ;
- les méthodes déterministes converges vers le minimum local le plus proche ;
- plus vous saurez sur la fonction (gradient, hessien) plus la minimisation sera efficace.

2.2 Quelques rappelles :

Nous allons ici rappeler brièvement les définitions de quelques notions importantes auxquelles nous ferons appel par la suite ainsi que quelques propriétés.

2.2.1 Dérivée partielle

Soit $f : \mathbb{R}^n \to \mathbb{R}$ *une fonction continue. La fonction notée* $\nabla_i f(x)$: $\mathbb{R}^n \to \mathbb{R}$, *également notée* $\partial f/\partial x_i$ *est appelée* $i^{ième}$ *dérivée partielle de* f *et est définie par*

$$\lim_{\alpha \to 0} \frac{f(x_1, ..., x_i + \alpha, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{\alpha}.$$

Cette limite peut ne pas exister.

2.2.2 Gradient

Si les dérivées partielles $\partial f/\partial x_i$ exixtent pour tout i, le gradient de f est défini de la façon suivante.

$$(\nabla f(x))^T = \left(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}\right)_{(x)},$$

le gradient de f *au point* $x = (x_1, .., x_n)$.

Le gradient jouera un role essentiel dans le développement et l'analyse des algorithmes d'optimisation.

Exemple 1 :

Calculer le gradient de la fonction f :

$f(x_1, x_2, x_3) = e^{x_1} + x_1^2 x_3 - x_1 x_2 x_3$.

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} e^{x_1} + 2x_1 x_3 - x_2 x_3 \\ -x_1 x_3 \\ x_1^2 - x_1 x_2 \end{pmatrix}.$$

2.2.3 Matrice Hessien

On appelle Hessien de f *la matrice symétrique de* $M_n(\mathbb{R})$

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

Alors

$$H(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

Exemple 2 :

Calculer la matrice Hessien de la fonction f :

$f(x_1, x_2, x_3) = e^{x_1} + x_1^2 x_3 - x_1 x_2 x_3$.

$$H(x) = \begin{pmatrix} e^{x_1} + 2x_3 & -x_3 & 2x_1 - x_2 \\ -x_3 & 0 & -x_1 \\ 2x_1 - x_2 & -x_1 & 0 \end{pmatrix}.$$

2.2.4 Ensembles et fonctions convexes

(ensemble convexe). Soit l'ensemble $C \subseteq \mathbb{R}^n$. C est *convexe* si

$$\lambda x + (1-\lambda)y \in C, \qquad \forall x, y \in C,\ \forall \lambda \in [0,1].$$

(fonction convexe). Soit l'ensemble convexe $C \subseteq \mathbb{R}^n$. Une fonction $f : C \mapsto \mathbb{R}$ est *convexe* si

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y), \qquad \forall x, y \in C,\ \forall \lambda \in [0,1].$$

Une fonction f est *concave* si $-f$ est convexe. Une fonction f est *strictement convexe* si l'inégalité ci-dessus est stricte pour tout $x, y \in C$ tels que $x \neq y$ et tout $\lambda \in (0,1)$.

Propriétée :

Si $C_1, \ldots, C_r$ sont des ensembles convexes, alors l'intersection $\bigcap_{i=1}^{r} C_i$ est convexe.

2.2.5 Points critiques : maximums, minimums

Dans un ensemble ordonné, le plus grand élément (ou le plus petit) d'une partie de cet ensemble est un extremum maximum (ou minimum) s'il est supérieur (ou inferieur) à tous les autres éléments de la partie. Ce groupe d'éléments sont connus sous le nom de points critiques ou points extremum définis sur un domaine d'étude D (espace topologique).

f(a) est un **maximum global** si $\forall x \in D\ f(x) \leq f(a)$

f(a) est un **minimum global** si $\forall x \in D\ f(x) \geq f(a)$

- f(a) est un **maximum local (ou relatif)** s'il existe un voisinage V de a tel que $\forall x \in V$, $f(x) \leq f(a)$
- f(a) est un **minimum local (ou relatif)** s'il existe un voisinage V de a tel que $\forall x \in V$, $f(x) \geq f(a)$

Pour trouver les maximums et les minimums d'une fonction, on utilise le calcul différentiel, et là ou la dérivée de la fonction s'annule, on trouve soit un maximum ou un minimum. Un minimum local est facile à trouver, mais il est difficile de trouver un minimum absolu.
Pour trouver le type de point critique, on utilise les deuxièmes dérivées évaluées dans le point d'étude. Pour le cas d'une fonction à une variable f(x),

- si f''(a)>0, on trouve un minimum local en ce point,
- si f''(a)<0 on trouve un maximum local

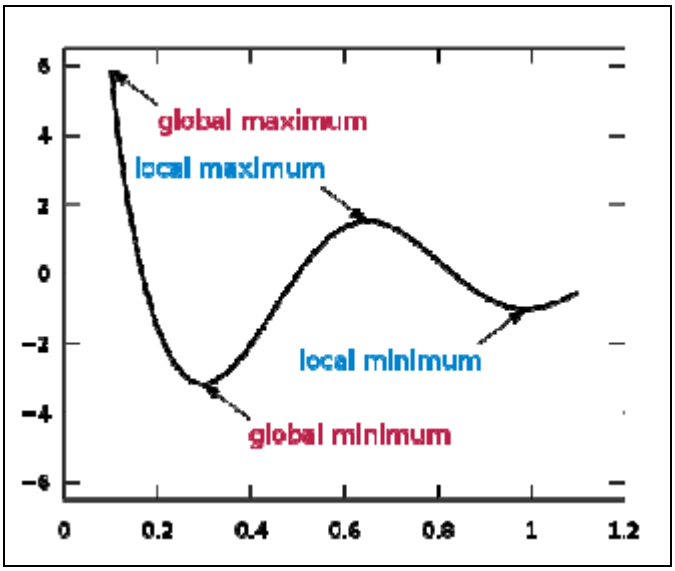


Figure1 : Points extrêmes (maximums et minimums locaux et globaux) sur une fonction

2.3 Conditions nécessaires

Etant donné un vecteur x*, nous souhaiterions être capables de déterminer si ce vecteur est un minimum local ou global de la fonction f. La propriété de différentiabilité continue de f fournit une première manière de caractériser une solution optimale. Enonçons tout d'abord un théorème, sur lequel s'appuiera le corollaire suivant pour établir une première condition nécessaire d'optimalité :

Théorème :

Soient $f : \mathbb{R}^n \mapsto \mathbb{R}$ continuellement différentiable et $x^* \in \mathbb{R}^n$. S'il existe un vecteur d tel que la dérivée directionnelle de f dans la direction d au point x^* (notée $f'(x^*; d)$) est strictement inférieure à zéro, alors d est une direction de descente de f en x^*

En outre, selon, une propriété connue de l'analye,

$$f'(x, d) = \nabla f(x) \cdot d.$$

Le théorème peut donc être également énoncé sous la forme équivalente suivante : s'il existe un vecteur d tel que $\nabla f(x^*) \cdot d < 0$, alors d est une direction de descente de f en x^*.

Corollaire (condition nécessaire du premier ordre). Soient $f : \mathbb{R}^n \mapsto \mathbb{R}$ continuellement différentiable et $x^* \in \mathbb{R}^n$. Si x^* est un minimum local de f, alors $\nabla f(x^*) = 0$.

Ce corollaire établit une première condition nécessaire pour que x^* soit un minimum (local ou global) de la fonction f, qui est $\nabla f(x^*) = 0$. Elle fait intervenir le vecteur gradient de f, dont les composantes sont ses premières dérivées partielles ; c'est la raison pour laquelle cette condition est appelée *condition nécessaire du premier ordre.*

Remarque : si f est convexe, la condition nécessaire du premier ordre est également suffisante pour que x^* soit un minimum global.

2.4 Conditions suffisantes

Les conditions données précédemment sont nécessaire (si f n'est pas convexe), c'est-à-dire qu'elles doivent être satisfaites pour tout minimum local ; cependant, tout vecteur vérifiant ces conditions

n'est pas nécessairement un minimum local. Le théorème suivant établit une condition suffisante pour qu'un vecteur soit un minimum local, si f est deux continuellement différentiable.

Théorème (condition suffisante du second ordre) Soient $f : \mathbb{R}^n \mapsto \mathbb{R}$ deux fois continuellement différentiable et $x^* \in \mathbb{R}^n$. Si $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*)$ est définie positive, alors x^* est un minimum local de f.

2.5 Méthodes itératives d'optimisation sans contraintes

Nous considérons ici les méthodes permettant de résoudre un problème d'optimisation sans Contraintes (appelées aussi parfois méthodes d'optimisation directe), soit le problème

$$\text{minimiser } f(x)$$
$$\text{avec } x \in \mathbb{R}^n$$

Il existe plusieurs méthodes :

1. Méthodes locales
2. Méthodes de recherche unidimensionnelle
3. Méthodes du gradient
4. Méthodes des directions conjuguées
5. Méthode de Newton
6. Méthodes quasi-Newton

Dans ce qui suit, nous allons présenter quelques méthodes par algorithmes permettant de calculer (de manière approchée) la ou les solutions du problème (P) de départ. Bien entendu, nous ne pouvons pas être exhaustifs ; nous présentons les méthodes "de base" les plus classiques. Toutefois, la plupart de ces algorithmes exploitent les conditions d'optimalité dont on a vu qu'elles permettaient (au mieux) de déterminer des minima locaux.

2.5.1 Définition d'un algorithme :

Un algorithme est défini par une application A de R^n dans R^n permettant la génération d'une suite d'éléments de R^n par la formule :

$$\begin{cases} x_o \in \mathbb{R}^n \text{ donné}, \ k = 0 & \textit{Etape d'initialisation} \\ x_{k+1} = \mathcal{A}(x_k), \ k = k + 1 & \textit{Itération } k \ . \end{cases}$$

2.5.2 Définition convergence d'un algorithme :

On dit que l'algorithme A converge si la suite $(x_k)_{k \in N}$ engendrée par l'algorithme converge vers une limite x*.

Il est bien entendu très important d'assurer la convergence d'un algorithme via les hypothèses ad-hoc, mais la vitesse de convergence et la complexité sont aussi des facteurs à prendre en compte lors de l'utilisation (ou de la génération) d'un algorithme ; on a en effet "intérêt » à ce que la méthode soit la plus rapide possible tout en restant précise et stable. Un critère de mesure de la vitesse (ou taux) de convergence est l'évolution de l'erreur commise
$(e_k = \|x_k - x^*\|)$.

Soit $(x_k)_{k\in\mathbb{N}}$ une suite de limite x^* définie par la donnée d'un algorithme convergent $\mathcal{A}$. On dit que la convergence de $\mathcal{A}$ est

- ***linéaire*** *si l'erreur $e_k = \|x_k - x^*\|$ décroit linéairement :*

$$\exists C \in [0,1[\ , \exists k_0, \quad \forall k \geq k_0 \qquad e_{k+1} \leq C e_k\ .$$

- ***super-linéaire*** *si l'erreur e_k décroit de la manière suivante:*

$$e_{k+1} \leq \alpha_k\, e_k\ ,$$

où α_k est une suite positive convergente vers 0. Si α_k est une suite géométrique, la convergence de l'algorithme est dite ***géométrique****.*

- ***d'ordre p*** *si l'erreur e_k décroit de la manière suivante:*

$$\exists C \geq 0\ , \exists k_0, \quad \forall k \geq k_0 \qquad e_{k+1} \leq C\,[e_k]^p\ .$$

Si $p = 2$, la convergence de l'algorithme est dite ***quadratique****.*

Enfin, la convergence est dite ***locale*** *si elle n'a lieu que pour des points de départ x_0 dans un voisinage de x^*. Dans le cas contraire la convergence est globale.*

1.Méthodes locales

1.1 Définition

Une **méthode locale** est une méthode qui :

- part d'un point initial x_0,
- utilise uniquement des informations **locales** (gradient, Hessienne, valeurs proches)
- converge vers un **minimum local** (pas forcément global)

Classification des méthodes locales

On distingue principalement :

1. Méthodes sans dérivées (ordre 0)

- Recherche directe (sans dérivés)

On explore le voisinage du point sans utiliser le gradient.

- Section dorée (en 1D)
- Méthode de Hooke-Jeeves

Exemple détaillé (Hooke-Jeeves simplifié)

Soit la fonction suivante à minimiser :

$f(x,y)=(x-2)^2+(y-1)^2$

✔ **Étape 1 : point initial**

$(x_0,y_0)=(0,0)$

✔ **Étape 2 : exploration**

On teste :

- (1,0) → f=2
- (0,1) → f=4

→ meilleur point : (1,0)

✔ **Étape 3 : nouvelle exploration**

- (2,0) → f=1
- (2,1) → f=0

✔ Solution trouvée :

(x∗,y∗)=(2,1)(x^*,y^*) = (2,1)(x∗,y∗)=(2,1)

2. Méthodes du premier ordre

Minimiser :

$f(x,y)=x^2+y^2$

Itération :

$x_{k+1}=x_k-\alpha\nabla f(x_k)$

✔ Convergence vers (0,0)

Descente de gradient

- Gradient à pas optimal

3. Méthodes du second ordre

- Newton
- Quasi-Newton (BFGS, DFP)

Ces familles sont séparées, mais **elles font toutes partie des méthodes locales**.

2.Méthodes de recherche unidirectionnelle

1. **La méthode analytique :**

Théorème 1: (Condition suffisante de minimalité locale)

Soit f : R → R une fonction C^2. Si x* ϵR vérifie les conditions :

f'(x*) = 0

f''(x*) >0

Alors, f(x*) est un minimum local strict de f.

Théorème 2: (Condition suffisante de maximalité locale)

Soit f : R → R une fonction C^2. Si x* ϵR vérifie les conditions :

f'(x*) = 0

f''(x*) <0

Alors, f(x*) est un maximum local strict de f.

Théorème 3: (Condition suffisante de point critique locale)

Soit f :R → R une fonction C^2. Si x* ϵR vérifie les conditions :

f'(x*) = 0

f''(x*) =0

Alors, x*est un point critique de f.

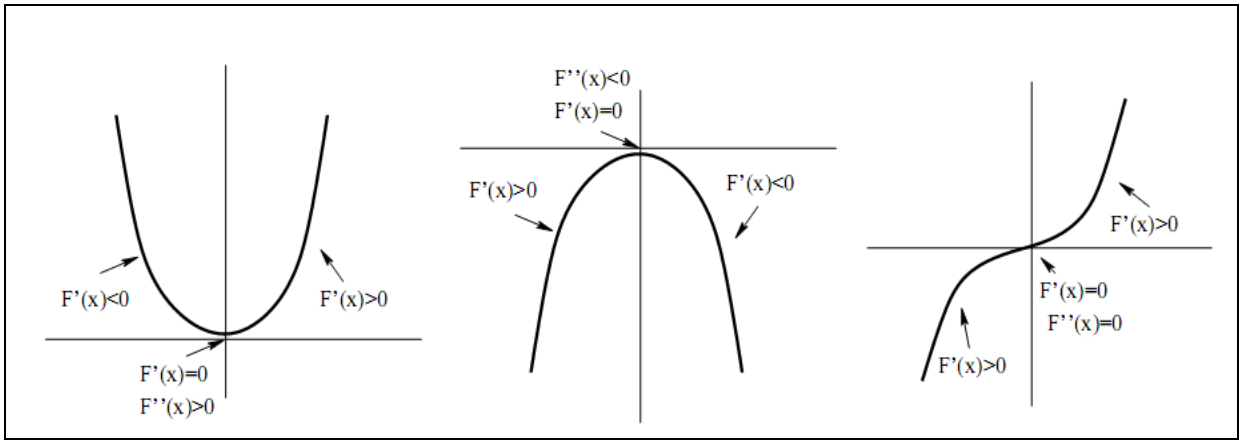


Figure 1: Exemples de problème d'optimisation.

Exemple1 :

Soit la fonction

$f(x) = x^2$

Trouver l'optimum de la fonction f.

f'(x)= 2x =0, x*=0

f''(x)= 2 > 0

Donc f(0) est un minimum de la fonction f.

Exemple2 :

Soit la fonction

$f(x) = x^3$

Trouver l'optimum de la fonction f.

f'(x)= $3x^2$ =0, x*=0

f''(x)= 6x et f''(x*) = 0 . Le point x* n'est pas minimum local ni maximum local.

Dans ce cas f''(x) <0 pour x<x* et f''(x)>0 pour x>x*.

Exemple3 :

Soit la fonction

f(x) =x^4

Trouver l'optimum de la fonction f.

f'(x)= $4x^3$ =0, x*=0

f''(x)= $12x^2$, f''(x*) = 0

Dans ce cas, x* est un minimum local de la fonction f(x), f'(x) < 0 pour x < x* et f'(x) > 0 pour x > x*

2. La méthode du second ordre (méthode de Newton) :

Il est possible d'utiliser la méthode de Newton pour rechercher un minimiseur. Le principe est de calculer une approximation p de f autour d'un point x_k par son développement de Taylor du second ordre :

$$p(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}f''(x_k)(x - x_k)^2$$

On calcule alors un nouveau point, x_{k+1}, qui minimise p soit :

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

D'un point de vue pratique, cette méthode souffre de nombreux inconvénients :

- La méthode peut diverger si le point de départ est trop éloigné de la solution,
- la méthode n'est pas définie si f''(x_k)=0,
- la méthode peut converger indifféremment vers un minimum, un maximum ou un poit selle.

Exemple :

Soit la fonction

f(x) = 2sinx – x^2/10 avec x_0 =2.5

La première dérivée de la fonction f est :

$$f'(x) = 2\cos x - \frac{2x}{10} = 2\cos x - \frac{x}{5}$$

La deuxième dérivée de la fonction f est :

$$f''(x) = -2\sin x - \frac{1}{5}$$

Suivant la méthode de Newton le point xi+1 s'écrit comme suit :

$$x_{i+1} = x_i - \frac{2\cos x_i - \frac{x_i}{5}}{-2\sin x_i - \frac{1}{5}}$$

La solution de la fonction f est x = 1.469.

$x_0 = 2.5,\ x_1 = 0.995,\ x_2 = 1.469.$

3. La méthode du premier ordre (Dichotomie) :

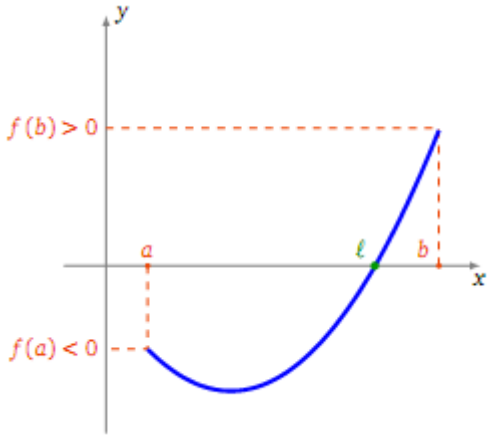
On cherche à trouver un minimum local x* avec une précision donnée par la réduction itérative de l'intervalle de recherche : dichotomie (c'est une optimisation exacte). Le principe de dichotomie repose sur la version suivante du théorème des valeurs intermédiaires :

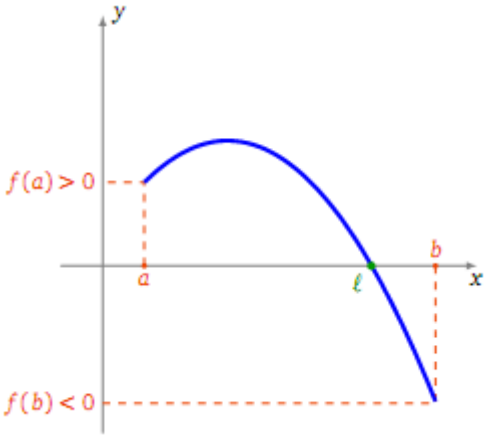
Théorème 1.

Soit f : [a, b] → R une fonction continue sur un segment.

Si f (a)· f (b) ≤ 0, alors il existe $l \in [a, b]$ tel que f (l) = 0

La condition f(a)·f(b) ≤ 0 signifie que f(a)et f(b) sont de signes opposés (ou que l'un des deux est nul). L'hypothèse de continuité est essentielle.



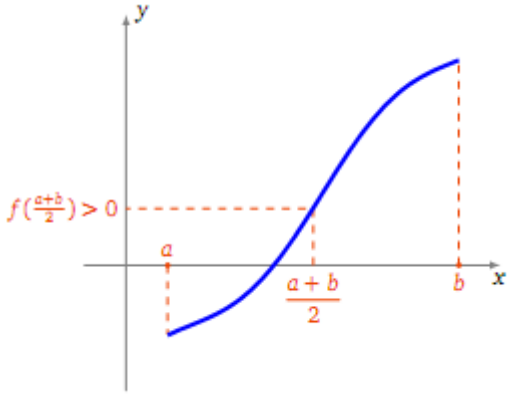


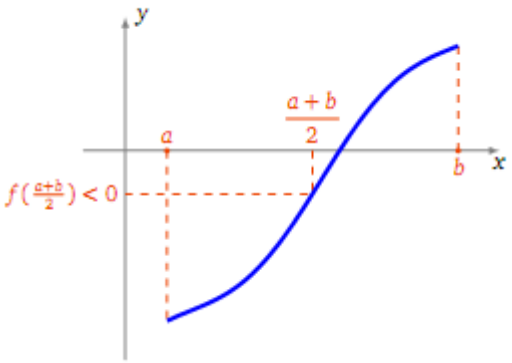
On part d'une fonction f : [a, b] → R continue, avec a < b, et f (a)· f (b) ≤0.

Voici la première étape de la construction : on regarde le signe de la valeur de la fonction f appliquée au point milieu a+b/ 2 .

• Si f (a)· f (a+b 2) ≤0, alors il existe c ∈ [a, a+b 2] tel que f (c) = 0.

• Si f (a)· f (a+b 2) > 0, cela implique que f (a+b 2)· f (b) ≤0, et alors il existe c ∈ [a+b 2 , b] tel que f (c) = 0





Nous avons obtenu un intervalle de longueur moitié dans lequel l'équation (f (x) = 0) admet une solution. On itère alors le procédé pour diviser de nouveau l'intervalle en deux. Voici le processus complet :

• **Au rang 0 :**

On pose $a_0 = a$, $b_0 = b$. Il existe une solution x0 de l'équation (f (x) = 0) dans l'intervalle [a_0 , b_0].

• **Au rang 1 :**

- Si f (a_0)x f ((a_0+b_0)/2) ≤ 0, alors on pose $a_1 = a_0$ et b_1 = (a_0+b_0)/2 ,
- sinon on pose a_1 = (a_0+b_0)/2 et $b_1 = b$.

- Dans les deux cas, il existe une solution x_1 de l'équation (f (x) = 0) dans l'intervalle [a_1 , b_1].

• ..

• Au rang n :

supposons construit un intervalle [a_n , b_n], de longueur $(b-a)/2^n$, et contenant une solution x_n de l'équation (f (x) = 0). Alors :

- Si f (a_n)· f ((a_n+b_n)/ 2) ≤0, alors on pose $a_{n+1} = a_n$ et $b_{n+1} = (a_n+b_n)/2$,
- sinon on pose $a_{n+1} = (a_n+b_n)/2$ et $b_{n+1} = b_n$.
- Dans les deux cas, il existe une solution x_{n+1} de l'équation (f (x) = 0) dans l'intervalle [a_{n+1} , b_{n+1}].

À chaque étape on a

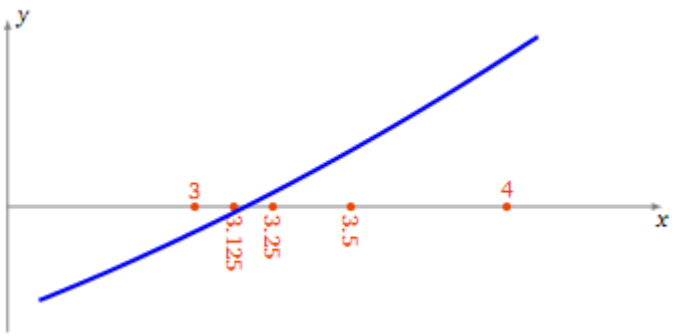
$$a_n \leq x_n \leq b_n$$

On arrête le processus dès que $b_n - a_n = (b-a)/2^n$ est inférieur à la précision souhaitée.

Exemple :

Soit la fonction f définie par f (x) = $x^2 - 10$, c'est une fonction continue sur R qui s'annule en ± √10. De plus √10 est l'unique solution positive de l'équation (f (x) = 0). Nous pouvons restreindre la fonction f à l'intervalle [3,4].

En d'autre termes f (3) 60 et f (4) >0, donc l'équation (f (x) = 0) admet une solution dans l'intervalle [3, 4] d'après le théorème des valeurs intermédiaires, et par unicité c'est √ 10, donc √10 ∈ [3, 4].



Voici les toutes premières étapes :

1. On pose a0 = 3 et b0 = 4, on a bien f (a0) 60 et f (b0) >0. On calcule a0+b0 2 = 3,5 puis f (a0+b0 2) : f (3,5) = 3, 52 − 10 = 2, 25 >0. Donc p 10 est dans l'intervalle [3; 3, 5] et on pose a1 = a0 = 3 et b1 = a0+b0 2 = 3, 5.
2. On sait donc que f (a1) 60 et f (b1) >0.

On calcule

f (a1+b1 2) = f (3,25) = 0,5625 >0,

on pose a2 = 3 et b2 = 3, 25. 3.

On calcule

f (a2+b2 2) = f (3,125) = −0,23. . . 60.

Comme f (b2) >0 alors cette fois f s'annule sur le second intervalle [a2+b2 2 , b2] et on pose a3 = a2+b2 2 = 3, 125 et b3 = b2 = 3, 25.

À ce stade, on a prouvé : 3, 125 6 ≤ 10 63, 25.

Voici la suite des étapes : a0 = 3 b0 = 4

a1 = 3 b1 = 3, 5

a2 = 3 b2 = 3, 25

a3 = 3, 125 b3 = 3, 25

a4 = 3, 125 b4 = 3, 1875

a5 = 3, 15625 b5 = 3, 1875

a6 = 3, 15625 b6 = 3, 171875

a7 = 3, 15625 b7 = 3, 164062 .
a8 = 3, 16015 . . . b8 = 3, 164062 . . .

Donc en 8 étapes on obtient l'encadrement : 3, 160 6 p 10 63, 165 En particulier, on vient d'obtenir les deux premières décimales : √ 10 = 3, 16 .

4. **La méthode de la section d'or :**

La méthode du nombre d'or est plus générale et plus simple à mettre en œuvre.

On impose un **rapport de réduction fixe** de l'intervalle à chaque itération.

$$\frac{\Delta_1}{\Delta_2}=\frac{\Delta_2}{\Delta_3}=\ldots\frac{\Delta_k}{\Delta_{k+1}}=\frac{\Delta_{k+1}}{\Delta_{k+2}}=\ldots=\gamma \quad \text{avec} \quad \Delta_k=\Delta_{k+1}+\Delta_{k+2}$$

$$\Rightarrow \frac{\Delta_k}{\Delta_{k+1}}=1+\frac{\Delta_{k+2}}{\Delta_{k+1}} \Rightarrow \gamma=1+\frac{1}{\gamma} \Rightarrow \gamma^2-\gamma-1=0$$

On obtient pour le rapport γ le nombre d'or : $\boxed{\gamma=\frac{1+\sqrt{5}}{2}\approx 1.618034}$

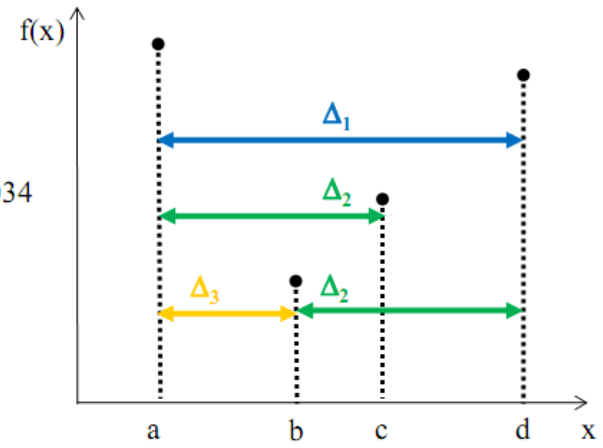
→ **Méthode du nombre d'or** (ou de la **section dorée**)

La Méthode du nombre d'or commence par le positionnement des points comme illustré sur cette figure .

$$\frac{\Delta_1}{\Delta_2}=\frac{\Delta_2}{\Delta_3}=\gamma=\frac{1+\sqrt{5}}{2}\approx 1.618034$$

$$\Rightarrow \begin{cases}\Delta_2=\frac{1}{\gamma}\Delta_1 &= r\Delta_1\\ \Delta_3=\frac{1}{\gamma^2}\Delta_1=r^2\Delta_1\end{cases} \quad \text{avec } r=\frac{1}{\gamma}=\frac{\sqrt{5}-1}{2}\approx 0.618034$$

$$\Rightarrow \begin{cases} b=a+r^2\Delta_1\\ c=a+r\Delta_1\\ d=a+\Delta_1\end{cases}$$

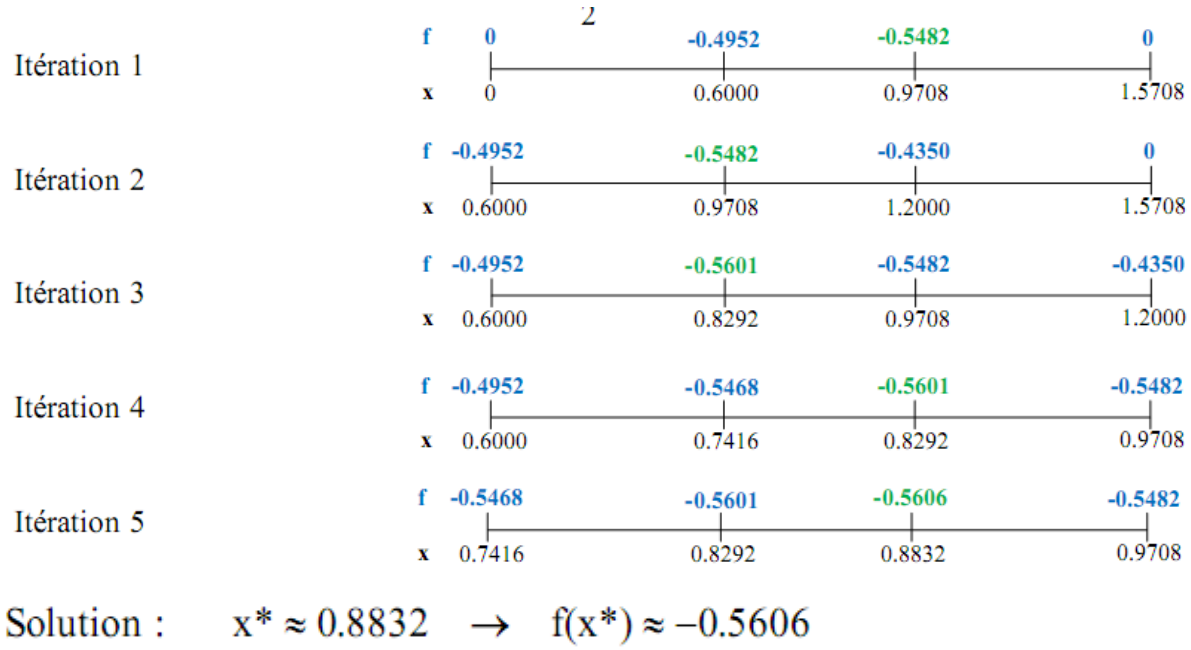


Itération

$$\text{Si } f(b)<f(c) \rightarrow \begin{cases} d \leftarrow c\\ c \leftarrow b\\ \Delta_1 \leftarrow \Delta_2\\ b=a+r^2\Delta_1\end{cases} \qquad \text{Si } f(b)>f(c) \rightarrow \begin{cases} a \leftarrow b\\ b \leftarrow c\\ \Delta_1 \leftarrow \Delta_2\\ c=a+r\Delta_1\end{cases}$$

Exemple :

Minimisation de f(x) = - x cos(x) avec $0 \le x \le 2/\pi$



Solution : $x^* \approx 0.8832 \rightarrow f(x^*) \approx -0.5606$

Méthodes du gradient

La méthode (ou algorithme) du Gradient fait partie d'une classe plus grande de méthodes numériques appelées méthodes de descente. Cette méthode est la plus simple à mettre en œuvre pour trouver un minimum local sur une fonction. Comme son nom l'indique, on utilise le gradient en un point donné de courbe pour donner la direction de la descente. La distance entre le point x_k et le point x_{k+1} est calculé en fonction de la valeur du gradient d_k et d'un pas déterminé à l'avance ρ.

$x_{k+1} = x_k + \rho\, g_k$

Elle consiste à minimiser une fonction J. Pour cela on se donne un point de départ arbitraire xo. Pour construire l'itéré suivant x_1 il faut penser qu'on veut se rapprocher du minimum de J; on veut donc que $J(x_1) < J(xo)$.
On cherche alors x1 sous la forme :

$x_1 = xo + \rho_1 d_1$

Avec :
d_1 est un vecteur non nul de R^n
ρ_1 un réel strictement positif.

En pratique donc, on cherche d_1 et ρ_1 pour que $J(xo+\rho_1 d_1) < J(xo)$. On ne peut pas toujours trouver d_1. Quand d_1 existe on dit que c'est une direction de descente et ρ_1 est le pas de descente. La direction et le pas de descente peuvent être fixé ou changer à chaque itération. Le schéma général d'une méthode de descente est le suivant :

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ donné} \\ x_{k+1} = x_k + \rho_k\, d_k,\ d_k \in \mathbb{R}^n - \{0\},\ \rho_k \in \mathbb{R}^{+*}\ , \end{cases}$$

où ρ_k et d_k sont choisis de telle sorte que $J(x_k + \rho_k\, d_k) \leq J(x_k)$.

Une idée naturelle pour trouver une direction de descente est de faire un développement de Taylor (formel) à l'ordre 2 de la fonction J entre deux itérés x_k et $x_{k+1} = x_k + \rho_k\, d_k$:

$$J(x_k + \rho_k\ d_k) = J(x_k) + \rho_k(\nabla J(x_k), d_k) + o(\rho_k d_k).$$

Comme on veut $J(x_k + \rho_k\ d_k) < J(x_k)$, on peut choisir en première approximation $d_k = -\nabla J(x_k)$. La méthode ainsi obtenue s'appelle l'algorithme du **Gradient**. Le pas ρ_k est choisi constant ou variable.

Algorithme du Gradient

1. **Initialisation**

   $k = 0$: choix de x_0 et de $\rho_0 > 0$

2. **Itération** k

   $x_{k+1} = x_k - \rho_k\ \nabla J(x_k)$;

3. **Critère d'arrêt**

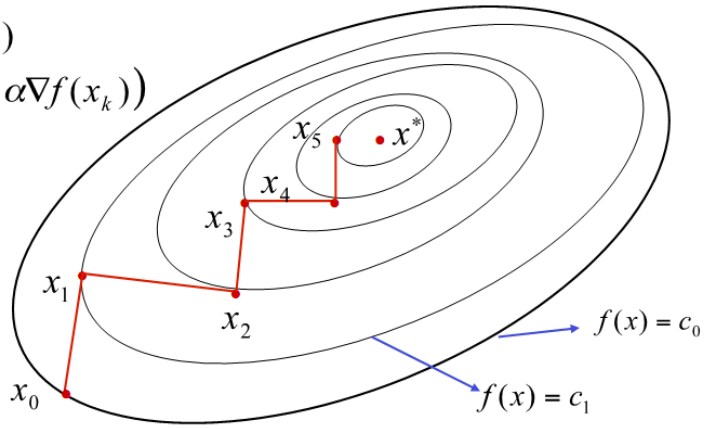
   Si $\|x_{k+1} - x_k\| < \varepsilon$, STOP

   Sinon, on pose $k = k + 1$ et on retourne à 2.

- Le gradient de la fonction J Indique la direction avec le plus grand taux de décroissance de f au point

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$\alpha_k = \arg\min_{\alpha \geq 0} f\left(x_k - \alpha \nabla f(x_k)\right)$$



Propriétés :

1° $(x_{k+1} - x_k) \perp (x_k - x_{k-1})$ $\qquad c_1 < c_0$

2° si $\nabla f(x_k) \neq 0$ alors $f(x_{k+1}) < f(x_k)$

Cette méthode a pour avantage d'être très facile à mettre en œuvre. Malheureusement, les conditions de convergence sont assez lourdes (c'est essentiellement de la stricte convexité) et la méthode est en général assez lente. Nous donnons ci-dessous un critère de convergence.

On utilise le plus souvent la méthode du gradient à pas constant ($\rho_k = \rho$= constant). Toutefois, on peut faire varier le pas à chaque itération : on obtient alors la méthode du gradient à pas variable.
La méthode du gradient `a pas optimal propose un choix du pas qui rend la fonction cout minimale le long de la direction de descente choisie. Plus précisément, l'´étape 2 devient :

2'.- **Itération** k

$$x_{k+1} = x_k - \rho_k \nabla J(x_k)$$

où ρ_k réalise le minimum sur $\mathbb{R}^+$ de la fonction Φ_k définie par

$$\Phi_k(\rho) = J(x_k - \rho \nabla J(x_k)) \ .$$

En pratique, on ne cherche pas le minimum de Φ_k et on détermine ρ_k en effectuant une **recherche linéaire** de pas optimal suivant une règle de la forme suivante par exemple :

Exemple 1(à pas constant)
Le problème est de trouver la valeur de x qui minimise E(x).
Dans cette exemple, on connaît l'expression analytique de la fonction E : E(x) = (x-1)(x-2)(x-3)(x-5) = (x2 -3x +2)(x2 -8x +15)
= x4 -11x^3 +41x^2 -61x +30

On connaît aussi sa dérivée :
E'(x) = 4x3 - 33x2 + 82x – 61

Pour trouver analytiquement le minimum de la fonction E, il faut trouver les racines de l'équation E'(x) = 0, donc trouver les racines d'un polynôme de degré 3, ce qui est « difficile ». Donc on va utiliser la DG.
La DG consiste à construire une suite de valeurs xi (avec x0 fixé au hasard) de manière itérative:

xi+1 = xi – η E'(xi)

si x_0 = 4,5 et η = 0.01,
x_1= x_0- η E'(x_0)

Exemple 2(à pas variable)
Soit la fonction à deux variables suivante :
f(x,y) = 4x^2 -4xy +2y^2
Déterminer la solution optimale en utilisant la méthode du gradient à pas variable si x_0(2,3)

- La première itération :

On a :
∇f(x_0,y_0) = t (8x_0 −4y_0,−4x_0 + 4y_0)
∇f(x(0)) = t (4,4)
x(0) −ρ∇f(x(0))= t(2,3)−ρ t(4,4) = t(2−4ρ,3−4ρ)
Ainsi
f(x(0) −ρ∇f(x(0)))=4(2−4ρ)2 −4(2−4ρ)(3−4ρ) + 2(3−4ρ)2 =10−32ρ + 32ρ^2
Cette dernière fonction est un trinôme du second degré, qui est minimale en ρ = 1/2.
On trouve ainsi
x(1) = t(0,1)

- La deuxième itération :

∇f(x(0)) = t (−4,4)
d'où x(1) −ρ∇f(x(1))= t(0,1)−ρ t(−4,4) = t(4ρ,1−4ρ)
Ainsi f(x(1) −ρ∇f(x(1)))=4(4ρ)2 −4(4ρ)(1−4ρ) + 2(1−4ρ)2 =2−32ρ + 160ρ2

Cette dernière fonction est un trinôme du second degré, qui est minimale en $\rho = 1/10$. On trouve ainsi
$x(2) = t(2/5,3/5)$

- La troisième itération :

on a $\nabla f(x(2)) =$
$t(4/5,4/5)$
d'où $x(2) - \rho \nabla f(x(2)) =$
$t(2/5,3/5) - \rho\, t(4/5,4/5) = t(2/5 - 4\rho/5,\ 3\ 5 - 4\rho/5)$
Ainsi $f(x(2) - \rho \nabla f(x(2))) = 4(2/5 -$
$4\rho/5)2 - 4(2/5 - 4\rho/5)(3/5 - 4\rho/5) + 2(3/5 - 4\rho/5)2 = 2/5 - 32/25\rho + 32/25\rho^2$
Cette dernière fonction est un trinôme du second degré, qui est minimale en $\rho = 1/2$. On trouve ainsi
$x(3) = t(0, 1/5)$

Valeurs de f(x(k)) L'algorithme minimise bien la fonctionnelle f ; on a en effet :
$f(x(0)) = 10$
$f(x(1)) = 2$
$f(x(2)) = 2/5$
$f(x(3)) = 2/25$

Méthodes des directions conjuguées

Les **méthodes des directions conjuguées** sont des méthodes d'optimisation locales utilisées pour minimiser une fonction, particulièrement efficaces pour les **fonctions quadratiques**.
Elles constituent une amélioration de la **méthode du gradient**, en évitant les oscillations et en accélérant la convergence.

On considère :

$\text{Min } f(x) \quad x \in R^n$

Dans le cas quadratique :

$f(x) = (1/2)\, x^T A\, x - b^T x$

où :

- A est une matrice symétrique définie positive
- $B \in R^n$

Contrairement à la descente de gradient :

- on ne descend pas simplement selon $-\nabla f$
- on construit des directions conjuguées d_k

Définition

Deux directions d_i et d_j sont **conjuguées** si :

$d_i^T A d_j = 0$ pour $i \neq j$

Interprétation

Les directions sont indépendantes vis-à-vis de la géométrie de la fonction

✔ Cela évite de "revenir en arrière"
✔ Permet une convergence très rapide

Algorithme (Gradient conjugué)

Initialisation

x_0 donne
$g_0 = Ax_0 - b$
$d_0 = -g0$

Itérations

Pour k=0,1,2,...

- **Pas optimal**

$$\alpha k = (g_k T g_k)/(d_k T A d_k)$$

- **Mise à jour**

$$x_{k+1} = x_k + \alpha_k d_k$$

- **Nouveau gradient**

$$g_{k+1} = Ax_{k+1} - b$$

- **Coefficient de conjugaison**

$$\beta_k = (g_{k+1} T g_{k+1})/(g_k T g_k)$$

- **Nouvelle direction**

$$d_{k+}1 = -g_{k+1} + \beta_k d_k$$

Méthodes de Newton

Son principe repose sur une approximation quadratique de f, donc la direction est donnée par :

$$dk = -[\nabla 2f(x_k)]^{-1} \nabla f(x_k)$$

Dans ce cas les ittirations sont effectuées par :

$$x_{k+1} = x_k + dk$$

Propriétés

- Convergence quadratique locale
- Très rapide près de la solution
- Nécessite :
  - calcul de la Hessienne
  - inversion matricielle

Limites

- Coût élevé pour grandes dimensions
- Instabilité si Hessienne non définie positive

Méthodes quasi Newton

Les méthodes quasi-Newton sont des méthodes d'optimisation locales utilisées pour minimiser une fonction non linéaire. Elles constituent une amélioration de la méthode du gradient et une alternative à la méthode de Newton.

- Newton utilise la Hessienne exacte : $\nabla^2 f(x)$

- Quasi-Newton : approxime la Hessienne dans le but de :

- garder la rapidité de Newton
- éviter le coût de calcul élevé

On considère :

Min $f(x)$ $x \in R^n$

Itération générale

$$x_{k+1} = x_k - H_k \nabla f(x_k)$$

où :

- $H_k \approx [\nabla^2 f(x_k)]^{-1}$
- H_k est une approximation de l'inverse de la Hessienne

Idée clé : condition de sécante

On impose :

$$H_{k+1} y_k = s_k$$

avec :

$s_k = x_{k+1} - x_k$ et $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ Cela garantit une bonne approximation de la Hessienne

Algorithme général

Initialisation

x_0 donne´, H0=I

À chaque itération

- **Direction**

$d_k = -H_k \nabla f(x_k)$

- **Pas optimal**

$\alpha_k = \arg\min f(x_k + \alpha d_k)$

- **Mise à jour**

$x_{k+1} = x_k + \alpha_k d_k$

- **Calcul**

$s_k = x_{k+1} - x_k$

$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$

- **Mise à jour de Hk**

Dépend de la méthode choisie, la plus utilisée est la Méthode BFGS ;

$H_{k+1} = (H_k + y_k y_k^T)/(y_k^T s_k) - (H_k s_k s_k^T H_k)/(s_k^T H_k s_k)$

✔ Stable
✔ Très performante
✔ Standard industriel

Interprétation géométrique

- Gradient → descend selon la pente
- Newton → utilise la courbure exacte
- Quasi-Newton → **apprend progressivement la courbure**, comme si l'algorithme "comprend" la forme de la fonction

Exemple d'application :

On considère la fonction quadratique :

$f(x,y) = x^2 + 3y^2$ avec point initial : $x_0 = (1,1)$

On applique les différentes méthodes d'optimisation précédentes.