Optimisation non linéaire sans contraintes

Un problème d'optimisation non linéaire sans contraintes s'écrit :

$$\min f(x)$$

$$x \in R^n$$

$$f: R^n \to R$$

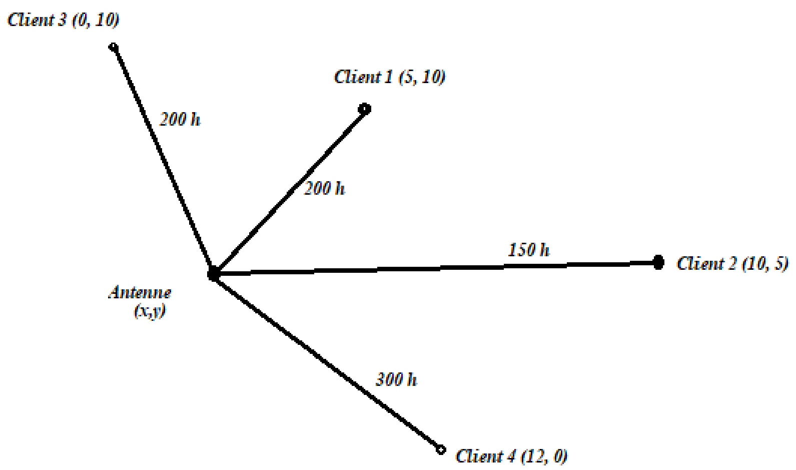
f : fonction non linéaire

<u>Exemple :</u>

On voudrait installer une antenne pour connecter quatre clients importants. Cette antenne doit être se trouver au plus proche de chaque client, en donnant propriété aux meilleurs clients.

Chaque client, on connait :

- Sa localisation (ses coordonnées *x,y*).
- Le nombre d'heure de communication par mois.



$$Min\, f(x,y) = 200\sqrt{(x-5)^2+(y-10)^2} + 150\sqrt{(x-10)^2+(y-5)^2} + 200\sqrt{x^2+(y-10)^2} + 300\sqrt{(x-12)^2+y^2}$$

<u>Convexité</u>

Soit *x1* et *x2* deux vecteurs de R^n . Le segment de droite joignant les extrémités de ces vecteurs est l'ensemble des points :

$$D = \{\boldsymbol{x} \in IR^n \mid \boldsymbol{x} = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2,\ \ 0 \leq \lambda \leq 1\}$$

Exemple : considérons les deux vecteurs $x1 = (4\ \ 1)^T$ et $x2 = (2\ \ 5)^T$
Notons $\overline{P_1P_2}$ Le segment de droite joignant les extrémités de $x1$ et de $x2$

$$\overline{\boldsymbol{P}_1\boldsymbol{P}_2} = \left\{\boldsymbol{x} \in IR^2 \mid \boldsymbol{x} = \lambda\begin{pmatrix}4\\1\end{pmatrix} + (1-\lambda)\begin{pmatrix}2\\5\end{pmatrix},\ \ 0 \leq \lambda \leq 1\right\}$$

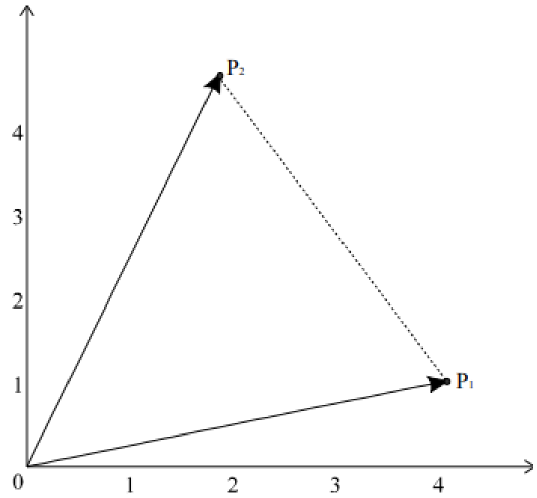
En faisant varier λ de 0 à 1, on obtient :

$\lambda = 0 \qquad x = \binom{2}{5}$, $\qquad\qquad \lambda = 1/4 \qquad x = \binom{5/2}{4}$,

$\lambda = 1/2 \qquad x = \binom{3}{3}$, $\qquad\qquad \lambda = 3/4 \qquad x = \binom{7/2}{2}$

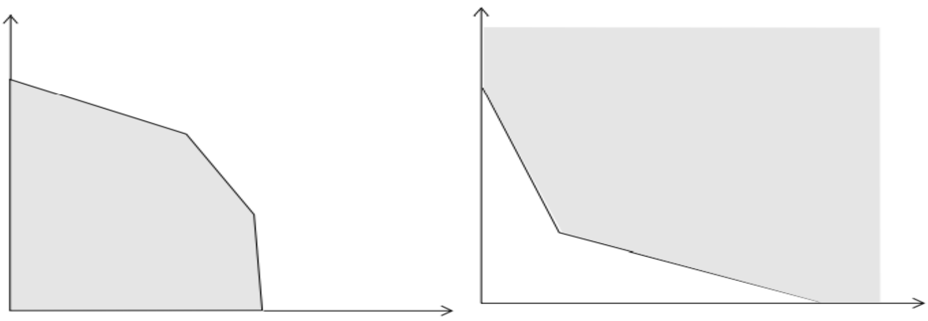
$\lambda = 1 \qquad x = \binom{4}{1}$

L'ensemble des points du segment sont représentées sur la figure ci-dessous

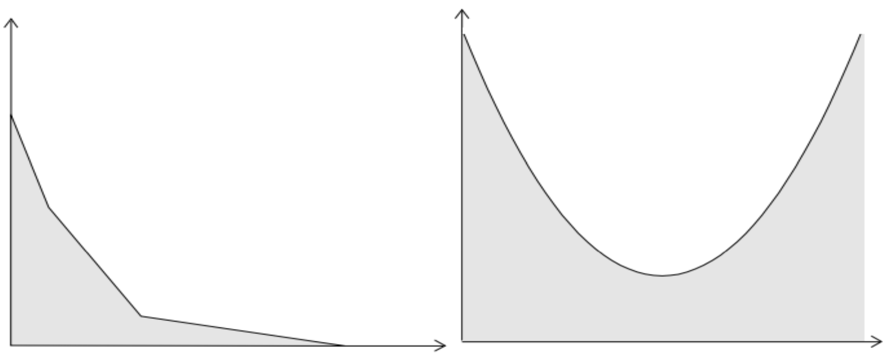


Définition : un ensemble S non vide de R^n est dit convexe si le segment de droite joignant deux point quelconque de S se trouve dans S.

$$\text{si } \boldsymbol{x}_1, \boldsymbol{x}_2 \in S \text{ et } 0 \leq \lambda \leq 1, \text{ alors } \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in S$$



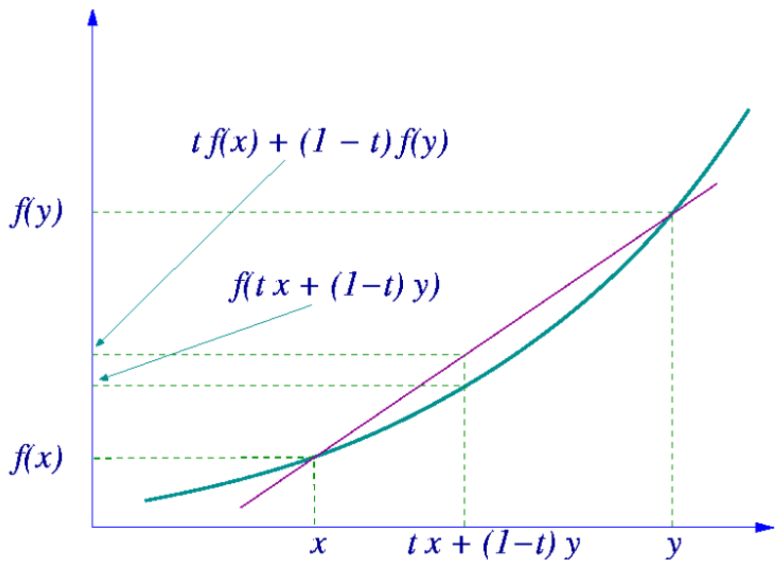
Ensembles convexes



Ensembles non convexes

Une fonction définie sur un intervalle réel D est dite convexe lorsque, pour tous x et y D et tout $t \in [0,1]$ on a :

$$f(tx + (1-t)y) \leq tf(y) + (1-t)f(y)$$



Le gradient

Soit $f: R^n \rightarrow R$ admet en un point x des dérivées partielles du premier ordre. Le gradient de f au point x

$$\nabla f(x) = (\frac{\partial f}{\partial x_1}(x),\ \frac{\partial f}{\partial x_2}(x),\ \frac{\partial f}{\partial x_3}(x), \dots \dots, \frac{\partial f}{\partial x_n}(x))$$

Exemple :

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

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} e^{x_1} + 2x_1x_3 - x_2x_3 \\ -x_1x_3 \\ x_3 - x_1x_2 \end{pmatrix}$$

La matrice hessienne

Si les dérivées secondes de f existent et sont continues, alors la matrice hessienne en x s'écrit :

$$\mathrm{H} = \nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{ij}$$

Ou bien

$$H = \nabla^2 f(x) = \begin{vmatrix} \frac{\partial^2 f}{\partial^2 x_1}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) \dots\dots\dots & \frac{\partial^2 f}{\partial x_1 \partial x_m}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial^2 x_2}(x) \dots\dots\dots & \frac{\partial^2 f}{\partial x_2 \partial x_m}(x) \\ & \vdots & \\ \frac{\partial^2 f}{\partial x_m \partial x_1}(x) & \frac{\partial^2 f}{\partial x_m \partial x_2}(x) \dots\dots\dots & \frac{\partial^2 f}{\partial^2 x_m}(x) \end{vmatrix}$$

H est une matrice carrée et symétrique

Condition nécessaire d'optimalité du premier ordre

Si f possède un extremum local $x = x^*$ et si le gradient de $f(x)$ existe en $x = x^*$ alors $\nabla f(x^*) = 0$ (x^* est un point critique).

Note : ce n'est pas une condition suffisante : un point critique peut être un minimum local, un maximum local ou bien un point de selle.

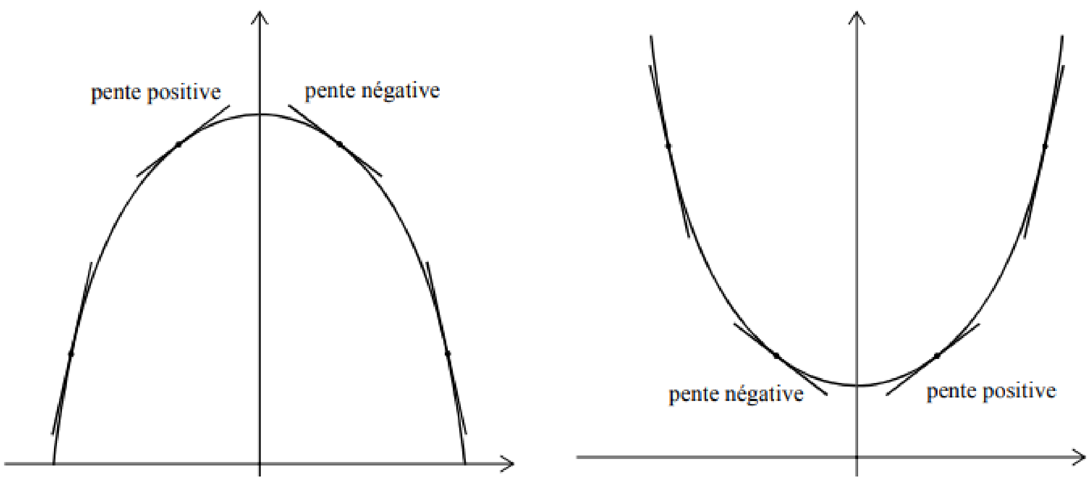
Condition nécessaire d'optimalité du second ordre ordre

Si x^* est minimum local de f dans R^n alors $\nabla f(x^*) = 0$ et la matrice Hessienne $\nabla^2 f(x^*)$ est semi définie positive

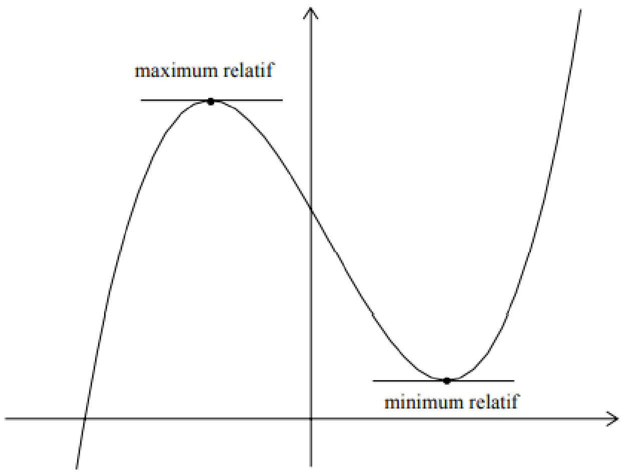
Condition suffisante d'optimalité du second ordre ordre

Si $x^* \in R^n$ est un point critique et si $\nabla^2 f(x^*)$ est :

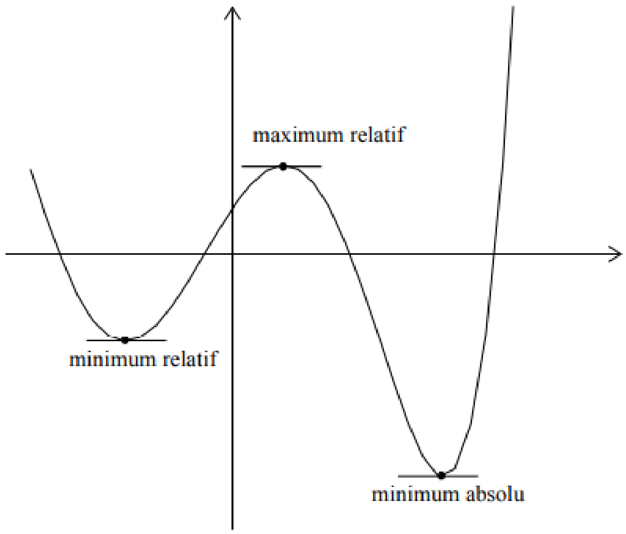
- Définie positive : x^* est un minimum local
- Définie négative : x^* est un maximum local
- Indéfinie : x^* est un point de selle
- Semi définie positive ou semi définie négative : on ne peut rien dire.



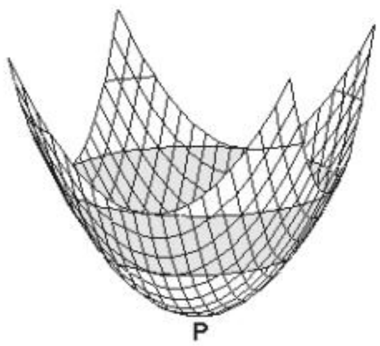
Maximum et minimum d'une fonction à une variable



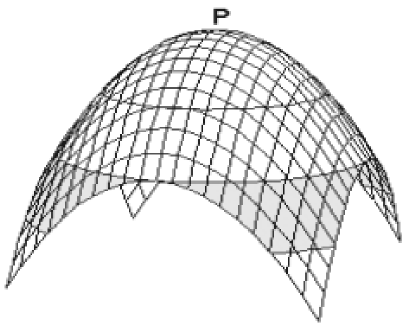
Extrema relatifs (locaux)



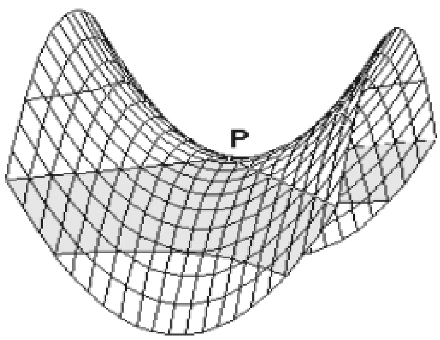
Extrema relatifs et absolus (locaux et globaux)



Minimum au point P



Maximum au point P



point-selle en P

Mineurs principaux d'une matrice

On appelle mineurs principaux de la matrice *H* (matrice carrée *nxn*), notés Δ_i, les déterminants des sous-matrices de *H* obtenues en lui retirant ses $n - i$ dernières lignes et colonnes (*i = 1,...,n*).
Soit P le point en lequel :

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \cdots \dots \dots = \frac{\partial f}{\partial x_n} = 0$$

1- Si les mineurs principaux de la matrice hessienne au point P sont tous strictement positifs, il s'agit d'un minimum.
2- Si les mineurs principaux de la matrice hessienne au point P sont de signes alternés, le premier étant strictement négatif, il s'agit d'un maxmaximum.
3- Si les mineurs principaux ne vérifient pas l'une des conditions ci-dessus prises au sens large (c'est-à-dire respectivement "positif ou nul" et "négatif ou nul"), il ne s'agit ni d'un minimum ni d'un maximum, mais d'un point-selle.
4- si les conditions (1) et (2) se vérifient au sens large, alors on ne peut pas conclure.

Dans le cas d'une fonction à deux variables *(x,y)* on :

$$\Delta_1 = \frac{\partial^2 f}{\partial x^2}$$

$$\Delta_2 = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial x}\right)^2$$

Ainsi :

$$\begin{cases} \Delta_1 > 0 \text{ et } \Delta_2 > 0 \Longrightarrow \text{minimum.} \\ \Delta_1 < 0 \text{ et } \Delta_2 > 0 \Longrightarrow \text{maximum.} \\ \Delta_1 \text{ quelconque et } \Delta_2 < 0 \Longrightarrow \text{point-selle.} \\ \Delta_1 \text{ quelconque et } \Delta_2 = 0 \Longrightarrow \text{on ne peut pas conclure.} \end{cases}$$

<u>Exemple 1 :</u>

Soit la fonction $f(x,y) = x^2 + y^2$

$$\frac{\partial f}{\partial x} = 2x = 0 \Longrightarrow x = 0$$

$$\frac{\partial f}{\partial y} = 2y = 0 \Longrightarrow y = 0$$

Il existe un point critique $x_0 = y_0 = 0$

La matrice hessienne est donc définie par :

$$\boldsymbol{H} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Ici, $\Delta_1 = 2$ et $\Delta_2 = 4$. Par conséquent, *f* possède un minimum en (0; 0; 0)

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

<u>Exemple 2:</u>

Soit la fonction $z = f(x,y) = 4x^2 - xy + y^2 - x^3$

On a :

$$\frac{\partial f}{\partial x} = 8x - y - 3x^2 \quad \frac{\partial f}{\partial y} = -x + 2y$$

$$\frac{\partial^2 f}{\partial x^2} = 8 - 6x \qquad \frac{\partial^2 f}{\partial y^2} = 2 \qquad \frac{\partial^2 f}{\partial x \partial y} = -1$$

Les points candidats s'obtiennent en résolvant les deux équations :

$$\begin{aligned} 8x - y - 3x^2 &= 0 \\ -x + 2y &= 0 \end{aligned}$$

Il existe donc deux points critiques *P(x ;y ; z)* : *P1(0; 0 ;0)* et *P2(5/ 2 ; 5 /4 ; 125 /16).*

La matrice hessienne est définie par :

$$\boldsymbol{H} = \begin{pmatrix} 8 - 6x & -1 \\ -1 & 2 \end{pmatrix}$$

Evaluons la matrice hessienne pour le premier point $x_1 = 0$ et $y_1 = 0$

$$\boldsymbol{H} = \begin{pmatrix} 8 & -1 \\ -1 & 2 \end{pmatrix}$$

Comme $\Delta_1 = 8 > 0$ et $\Delta_2 = 15 > 0$, il s'agit d'un minimum.
Pour le second point critique $x_2 = 5/2$ et $y_1 = 5/4$, on obtient :

$$\boldsymbol{H} = \begin{pmatrix} -7 & -1 \\ -1 & 2 \end{pmatrix}$$

Comme $\Delta_1 = -7 < 0$ et $\Delta_2 = -15 < 0$, il ne s'agit ni d'un minimum ni d'un maximum, mais d'un point de selle.

<u>Exemple 3:</u>

Soit la fonction $z = f(x, y, z) = x^4 - 17x^2 - 2xy + 2y^2 + z^2 - 2zy + 81$
Annulons les premières dérivées partielles :

$$\begin{aligned} \frac{\partial f}{\partial x} &= 4x^3 - 34x - 2y = 0 \\ \frac{\partial f}{\partial y} &= 4y - 2x - 2z = 0 \\ \frac{\partial f}{\partial z} &= 2z - 2y = 0 \end{aligned}$$

De la troisième équation on tire :

$$z = y$$

De la deuxième équation et avec la substitution de $x = y$:

$$\begin{aligned} 4y - 2x - 2y = 0 \\ 2y - 2x = 0 \\ x = y \end{aligned}$$

En remplaçant dans la première équation on trouve :

$$\begin{aligned} 4x^3 - 34x - 2x &= 0 \\ 4x^3 - 36x &= 0 \\ 4x(x^2 - 9) &= 0 \end{aligned}$$

Cette dernière équation à trois solutions : $x_1 = 0$, $x_2 = -3$ et $x_3 = 3$
Les valeurs correspondantes de y et z sont :

$$\begin{aligned} y_1 &= 0 \quad y_2 = -3 \quad y_3 = 3 \\ z_1 &= 0 \quad z_2 = -3 \quad z_3 = 3 \end{aligned}$$

Pour le point $(x_1, y_1, z_1) = (0, 0, 0)$

$$\boldsymbol{H} = \begin{pmatrix} -34 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 2 \end{pmatrix}$$

$$\begin{aligned} \triangle_1 &= -34 \\ \triangle_2 &= -140 \\ \triangle_3 &= -144 \end{aligned}$$

Le point $(x_1, y_1, z_1) = (0, 0, 0)$ est un point de selle

Pour les points $(x_2, y_2, z_2) = (-3, -3, -3)$ et $(x_3, y_3, z_3) = (3, 3, 3)$ la matrice hessienne est la même

$$\boldsymbol{H} = \begin{pmatrix} 74 & -2 & 0 \\ -2 & 4 & -2 \\ 0 & -2 & 2 \end{pmatrix}$$

$$\begin{aligned} \triangle_1 &= 74 \\ \triangle_2 &= 292 \\ \triangle_3 &= 288 \end{aligned}$$

Ainsi, pour chacun de ces points, la fonction présente un minimum.

Minimum d'une fonction convexe

Soit $f: R^n \to R$ une fonction convexe possédant des dérivées, f admet un minimum global en un point x^* si et seulement si $\nabla f(x^*) = 0$

Algorithmes de minimisation usuels

1- Algorithme de recherche dichotomique

La méthode est décrite comme suit :

Soit : $f: [a, b] \to R$, une fonction unimodale dérivable sur $[a, b]$. Si $f'(a) \times f'(b) < 0$, alors il existe au moins un $x^* \in [a, b]$ pour lequel $f'(x^*) = 0$. On prend $x_m = \frac{a+b}{2}$ telle que :

- Si $f'(x_m) = 0$ c'est le minimum de la fonction f
- Si $f'(x_m) \times f'(a) < 0$ le minimum se trouve dans l'intervalle $[a, x_m]$
- Si $f'(x_m) \times f'(a) > 0$ le minimum se trouve dans l'intervalle $[x_m, b]$

Le processus de division par deux de l'intervalle de la fonction dérivée est réitéré jusqu'à la convergence par la tolérance considérée.

2- Méthode de Newton

Soit $f: U \to R$ une fonction de classe C^2 définie sur un ensemble ouvert $U \subset R^n$

Une condition nécessaire (et suffisante si *f* est convexe) pour qu'un point $x^* \in U$ soit un minimum est :

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

Par conséquent, le point x^* doit vérifier le système de n équations à n variables.

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

Une approche très populaire pour résoudre $F(x) = 0$ est la méthode de Newton.

$$F(x) = 0 \Longleftrightarrow \left\{ \begin{array}{lcl} F_1(x_1, x_2, x_3, \ldots, x_n) & = & 0 \\ F_2(x_1, x_2, x_3, \ldots, x_n) & = & 0 \\ F_3(x_1, x_2, x_3, \ldots, x_n) & = & 0 \\ \vdots & = & \vdots \\ F_n(x_1, x_2, x_3, \ldots, x_n) & = & 0 \end{array} \right.$$

La matrice Jacobienne est définie par :

$$dF(x) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \frac{\partial F_n}{\partial x_2} & \cdots & \frac{\partial F_n}{\partial x_n} \end{pmatrix}$$

La méthode de Newton est basée sur l'approximation linéaire de *F* autour d'un point x_0

$$F(x_0 + \Delta x) \approx F(x_0) + dF(x_0)\Delta x$$

On désire calculer la correction Δx de sorte que :

$$0 = F(x_0 + \Delta x) \approx F(x_0) + dF(x_0)\Delta x \Longrightarrow \Delta x = -[dF(x_0)]^{-1}F(x_0)$$

Etant donné l'approximation, il est nécessaire d'itérer ce qui conduit à l'algorithme suivant :

Méthode de Newton

1. Etant donné une approximation initiale x_0,
2. Résoudre le système linéaire: $dF(x_k)\ \Delta x = -F(x_k) = R(x_k)$,
3. Mettre à jour la solution: $x_{k+1} = x_k + \Delta x$,
4. Si $\frac{\|\Delta x\|}{\|x_{k+1}\|} < \epsilon_1$ et/ou $\|F(x_{k+1})\| < \epsilon_2$, la convergence est atteinte.

Soit $f: U \rightarrow R$ une fonction de classe C^2 définie sur un ensemble ouvert $U \subset R^n$, pour calculer le minimum de f l'algorithme de Newton s'écrit :

$$\begin{cases} x_0 \in \mathbb{R}^n, \\ x_{k+1} = x_k - [H_f(x_k)]^{-1}\ \nabla f(x_k). \end{cases}$$

Avec :

$$dF(x) = \frac{\partial F_i}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j} = H_f(x)$$

Qui est toujours symétrique.

3- ***Méthodes du gradient :***

Il s'agit d'une famille de méthodes itératives qui s'appliquent à des fonctions dérivables. Elles sont des méthodes de descente.

On appelle méthode de descente toute méthode où, à chaque étape, on pose :

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

x_{k+1} : Nouvelle solution

x_k : Solution antérieure

ρ_k : Le pas tel que $\rho_k > 0$

d_k : Direction de descente $d_k = -\nabla f(x_k)$.

Partant de x_k dans la direction d_k, on descend alors $f(x_{k+1}) < f(x_k)$

Algorithme du gradient:

- $x_0 \in \mathbb{R}^n$ donné
- $x_{k+1} = x_k - \rho_k \nabla f(x_k)$ où $\rho_k > 0$.

Si $\rho_k = \rho$, l'algorithme est dit à pas constant.

Dans la pratique, la méthode du gradient à pas constant converge très lentement. Pour améliorer la convergence il est préférable de choisir les ρ_k de manière optimale d'où l'algorithme :

- $x_0 \in \mathbb{R}^n$ donné
- $x_{k+1} = x_k - \rho_k \nabla f(x_k)$
  où $\rho_k > 0$ est choisi de sorte que

$$\min_{\rho} f(x_k - \rho \nabla f(x_k))$$

Critère de convergence :

On peut utiliser un (ou une combinaison) des critères suivants pour arrêter les itérations d'un algorithme de descente.

- Etant donné que $\nabla f(x_k) \to \nabla f(\bar{x}) = 0$, on pose

$$\|\nabla f(x_k)\| < \epsilon_1.$$

- On a $x_k \to \bar{x}$, on peut aussi prendre

$$\|x_{k+1} - x_k\| < \epsilon_2.$$

- Finalement, on a $f(x_k) \to f(\bar{x})$, on peut prendre

$$\|f(x_{k+1}) - f(x_k)\| < \epsilon_3.$$

Exemple (Méthode de gradient à pas fixe)

Cherchons la minimisation de la fonction objectif :

$$f(x) = 4x^2 + e^x$$

Cette fonction est dérivable sur *R*. Exécutant le schéma numérique de l'algorithme de descente de gradient à pas fixe pour $x_0 = 1.5,\quad \rho = 0.02\quad$ et $\varepsilon = 10^{-3}$

$$d = -\nabla f(x) = 8x + e^x$$

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

$$\begin{aligned}
x^{(1)} &= 1.500000 - 0.02\,\nabla f(1.500000) = 1.170366\\
x^{(2)} &= 1.170366 - 0.02\,\nabla f(1.170366) = 0.918644\\
x^{(3)} &= 0.918644 - 0.02\,\nabla f(0.918644) = 0.721543\\
x^{(4)} &= 0.721543 - 0.02\,\nabla f(0.721543) = 0.564944\\
x^{(5)} &= 0.564944 - 0.02\,\nabla f(0.564944) = 0.439366\\
\vdots\ & \quad\ \vdots \qquad\qquad\quad \vdots \qquad\qquad\quad \vdots\\
\vdots\ & \quad\ \vdots \qquad\qquad\quad \vdots \qquad\qquad\quad \vdots\\
x^{(30)} &= -0.10695 - 0.02\,\nabla f(-0.10695) = -0.10781
\end{aligned}$$

Le test d'arrêt $|x_{k+1} - x_k| < \varepsilon$

$$|x_{30} - x_{29}| = 0.00085 < \varepsilon = 10^{-3}$$

L'algorithme converge vers la solution $x^* = -0.10781$ correspondant au minimum de la fonction objectif

Le tableau suivant donne le nombre d'itérations nécessaires pour approcher le minimum $(0,0)$ de la forme quadratique elliptique : $f = x^2 + 2\,y^2$ sur $\mathbb{R}^2$ à 10^{-6} près, à partir de l'initialisation : $(x_0, y_0) = (1,1)$, en fonction du pas choisi :

pas	0.5	0.45	0.4	0.33	0.1	0.01
	divergence	60	30	13	60	685

Exemple 2 :

Utiliser la méthode de la plus forte pente pour trouver le minimum de la fonction f définie de R^2 dans R comme suit :

$$f(x_1, x_2) = x_1^4 - 2x_1^3 + \frac{1}{6}x_2^2 - 3x_1 - x_1x_2$$

$$\nabla f(x_1, x_2) = \begin{pmatrix} 4x_1^3 - 6x_1^2 - x_2 - 3 \\ \frac{1}{3}x_2 - x_1 \end{pmatrix}$$

Soit : $x_0 = \binom{0}{0}$, on trouve la direction de la pente la plus forte

$$d_0 = -\nabla f(x_0) = \binom{3}{0}$$

On minimise donc :

$$g_0(\rho) = f(x_0 + \rho d_0) = 81\rho^4 - 54\rho^3 - 9\rho$$

On trouve

$$\rho_0 \approx 0.582$$

Notre premier pas nous conduit à :

$$x_1 = x_0 + \rho_0 d_0 = \binom{1.746}{0}$$

On réitère pour obtenir les valeurs suivantes :

i	x^i	$-\nabla f(x^i)$	$g_i(\lambda)$	λ_i	x^i
0	$\binom{0}{0}$	$\binom{3}{0}$	$81\lambda^4 - 54\lambda^3 - 6\lambda$	0.582	$\binom{1.746}{0}$
1	$\binom{1.746}{0}$	$\binom{0}{1.746}$	$0.508\lambda^2 - 3.049\lambda - 6.590$	3	$\binom{1.746}{5.238}$
2	$\binom{1.746}{5.238}$	$\binom{5.238}{0}$	$752\lambda^4 + 716\lambda^3 + 214\lambda^2 - 27\lambda - 11$	0.05	$\binom{2.01}{5.238}$
3	$\binom{2.01}{5.238}$	$\binom{0}{0.264}$	$0.012\lambda^2 - 0.07\lambda - 11.904$	3	$\binom{2.01}{6.03}$

<u>Exemple 2 :</u>

Utiliser la méthode de la plus forte pente pour trouver le minimum de la fonction f définie de R^2 dans R comme suit :

$$f(x_1, x_2) = 4x^2 - 4xy + 2y^2$$

Avec $x_0 = \binom{2}{3}$

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

- Première itération

$$\rho_0 = 0.5$$

$$x_1 = \binom{0}{1}$$

- Deuxième itération

$$\rho_0 = 0.1$$

$$x_2 = \begin{pmatrix} \frac{2}{5} \\ \frac{3}{5} \end{pmatrix}$$

- Troisième itération

$$\rho_0 = 1/2$$

$$x_3 = \begin{pmatrix} 0 \\ \\ \frac{1}{5} \end{pmatrix}$$

L'algorithme minimise bien la fonction f ; on a en effet :

$f(x_0) = 10$, $f(x_1) = 2$, $f(x_2) = 2/5$, $f(x_3) = 2/25$