Chapitre[1] Concavité et convexité des fonctions de plusieurs variables

Samir Kenouche - Département des Sciences de la Matière - UMKB

Module : Méthodes Mathématiques et Algorithmes pour la Physique
Version corrigée, améliorée et actualisée le 10/10/2020

Résumé

Ce chapitre débute par un bref rappel sur des notions élémentaires portant notamment sur la caractérisation de la convexité, de la concavité ainsi que les points stationnaires. Ces notions fondamentales sont indispensables et nécessaires afin d'appréhender le fonctionnement et le fondement des algorithmes d'optimisation traités dans le deuxième chapitre. Les problèmes rencontrés en physique, sont très souvent complexes nécessitant des modèles mathématiques multidimensionnels, il est donc fréquent de recourir à des fonctions de plusieurs variables. Ainsi, l'étude de la concavité et de convexité de ces fonctions repose pleinement sur la détermination de la matrice *Hessienne*. Par ailleurs, ce chapitre se veut avant tout introductif afin de tester et de consolider les connaissances mathématiques des étudiants (es).

Table des matières

I. Introduction

L'optimisation numérique de problèmes complexes, issus de la physique, a connu un essor considérable ces dernières années grâce notamment au développement de l'informatique et des moyens de calcul très puissants. Une opération d'optimisation se présente systématiquement comme l'ultime étape après avoir mis en équation ou modéliser le problème physique en question. Nous comprenons ainsi aisément que la compréhension du formalisme inhérent à la détermination ainsi qu'à l'analyse des points critiques (encore appelés points stationnaires) est une étape cruciale et indispensable afin de mener à bien cette opération d'optimisation. Il convient de préciser que ce premier chapitre se veut introductif pour le deuxième, dans lequel une étude avancée et détaillée des algorithmes d'optimisation sera conduite.

II. Notations différentielles

Nous entamons ces notes de cours en rappelant succinctement la notion de la dérivée. Cette quantité mathématique joue un rôle central dans la définition et la détermination des points stationnaires. Soit la fonction continue d'une seule variable $f : [a, b] \subset \mathbb{R} \longrightarrow \mathbb{R}$. On définit sa fonction dérivée par l'expression :

$$\frac{df(x)}{dx} = \left[\frac{\Delta f(x)}{\Delta x}\right]_{\Delta x \approx 0} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = f^{'}(x) \tag{1}$$

S. Kenouche est docteur en Physique de l'Université des Sciences et Techniques de Montpellier et docteur en Chimie de l'Université A. Mira de Béjaia.
page web personnelle : http://www.sites.univ-biskra.dz/kenouche

La variation totale absolue de f engendrée par un accroissement de $x + \Delta x$ s'écrit $\Delta f = f(x + \Delta x) - f(x)$. Par conséquent, la variation totale relative est la quantité :

$$\frac{\Delta f(x)}{\Delta x} \tag{2}$$

Rappelons que pour une variation infinitésimale ($\Delta x \longrightarrow 0$), nous pouvons écrire :

$$\Delta x \longrightarrow dx$$

$$\Delta f \longrightarrow df$$

Ayant un point x_0 appartenant à l'intervalle de définition de f, le nombre $f^{'}(x_0)$ exprime la pente, au point $(x_0, f(x_0))$, de la droite tangente à la courbe $y = f(x)$. En effet, le rapport :

$$f^{'}(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \tag{3}$$

vaut la pente de la droite passant par les points $(x_0, f(x_0))$ et $(x_0 + \Delta x, f(x_0 + \Delta x))$. Cette pente exprime le taux de changement de $f(x)$ causé par la variation Δx. Il en ressort que l'équation de la droite tangente à la courbe $y = f_T(x)$ au point (x_0, y_0) s'écrit simplement :

$$y - y_0 = f^{'}(x_0)(x - x_0) \Longrightarrow f_T(x) = y_0 + f^{'}(x_0)(x - x_0) \tag{4}$$

Cette approximation est d'autant plus vraie que $\Delta x \longrightarrow 0$. Lorsque la limite de l'Eq. (3) existe, on dira que f est dérivable en x_0. Dans le cas où la fonction f est dérivable pour tout $x \in [a, b] \subset \mathbb{R}$, on dira qu'elle est dérivable sur $[a, b] \subset \mathbb{R}$ et on aura par conséquent la fonction $f^{'}(x)$. Cette dernière peut également être dérivable et on aura $f^{''}(x)$ ... etc. Notons que la notion de la *dérivée directionnelle* sera définie dans le prochain chapitre.

Pour une fonction de deux variables $f(x, y) \in \mathbb{R}$, lorsque les deux variables sont modifiées de $x + \Delta x$ et $y + \Delta y$, la variation totale de f s'obtient selon :

$$\Delta f = f(x + dx, y + dy) - f(x, y) \tag{5}$$

En développant le premier terme du second membre en *séries de Taylor*, il vient :

$$f(x + dx, y + dy) = f(x, y) + \frac{\partial f(x, y)}{\partial x}\, dx + \frac{\partial f(x, y)}{\partial y}\, dy + O(dx^2, dy^2) \tag{6}$$

En substituant (6) dans (5), nous obtenons :

$$\Delta f = \frac{\partial f(x, y)}{\partial x}\, dx + \frac{\partial f(x, y)}{\partial y}\, dy \tag{7}$$

La variation totale de la fonction $f(x, y)$ s'écrit ainsi en fonction des variations partielles de f par rapport aux deux variables.

III. Convexité et concavité

Nous rappelons que la dérivée traduit les variations, croissance ($f^{'}(x) > 0$) et décroissance ($f^{'}(x) < 0$), d'une fonction. Une dérivée nulle, en un point x_0, est caractérisée par une tangente horizontale et donc la fonction admet un point stationnaire : un minimum, un maximum ou un point selle. Cette condition est nécessaire pour l'existence d'un extremum, mais elle n'est pas suffisante. A titre d'exemple, la fonction $f(x) = x^3$ n'admet pas d'extremum local en $x_0 = 0$ bien que sa dérivée $f^{'}(x) = 3\, x^2$ s'y annule, ce point est *singulier*.

Définition 1 *: La courbe de $f(x)$ est dite* **convexe** *si tous les points de la courbe $y = f(x)$ se trouvant au dessous de la tangente*[1] *en un point quelconque de l'intervalle de définition de la fonction, autrement dit : $\forall\, x\, \in [a, b]\; f(x) < f_T(x)$. Cette condition est satisfaite si $f^{''}(x) < 0$.*

Définition 1' *: La courbe de $f(x)$ est dite* **concave** *si tous les points de la courbe $y = f(x)$ se trouvant au dessus de la tangente en un point quelconque de l'intervalle de définition de la fonction, autrement dit : $\forall\, x\, \in [a, b]\; f(x) > f_T(x)$. Cette condition est satisfaite si $f^{''}(x) > 0$.*

Ces résultats se démontrent aisément en développant la fonction f en série de *Taylor* au voisinage du point critique x_0, soit :

$$f(x_0 + \delta x) = f(x_0) + f^{'}(x_0)\,\delta x + f^{''}(x_0)\,\delta x^2 + \epsilon(\delta x^3) \tag{8}$$

Avec x_0 est un point critique par conséquent $f^{'}(x_0) = 0$, il vient :

$$\Rightarrow \Delta f = f^{''}(x_0)\,\delta x^2 + \epsilon(\delta x^3) \tag{9}$$

Ainsi la nature du point critique est déterminée par le signe de $f^{''}(x_0)$. Si $f^{''}(x_0) > 0 \Rightarrow \Delta f > 0$ alors $f(x_0 + \delta x) > f(x_0)$. Nous en déduisons que la fonction f prend des valeurs supérieures au voisinage $(x_0 + \delta x)$ alors il s'agit nécessairement d'un minimum. Nous concluons pour les fonctions d'une seule variable, les conditions $f^{'}(x_0) = 0$ et $f^{''}(x_0) > 0$ sont suffisantes pour que la fonction $f(x)$ admette un minimum local (concavité). Les conditions $f^{'}(x_0) = 0$ et $f^{''}(x_0) < 0$ prouvent l'existence d'un maximum local (convexité).

Propriétés *: Si x_0 est un optimum local alors c'est un optimum global. Ainsi, le plus petit des minima locaux est un minimum global. De la même façon, le plus grand des maxima locaux est un maximum global.*

A. ***Généralisation aux fonctions de plusieurs variables***

Réécrivons les résultats obtenus pour le cas monodimensionnel au cas multidimensionnel. Soit f une fonction de classe $\mathcal{C}^2\,([\mathcal{D} \subset \mathbb{R}^n])$. La détermination de la nature des extremums sera conduite en analysant la série de *Taylor* généralisée au cas des fonctions de plusieurs variables. Afin de simplifier le formalisme, prenons $n = 2$ et travaillons avec $X \in \mathbb{R}^2$ et $f(X) \in \mathbb{R}$.

a) ***Théorème****: Soit $X^* = (x_0, y_0)^T \in \mathbb{R}^2$ un point critique de $f(X) \Rightarrow \nabla f(X^*) = 0$. Nous posons $D = (\delta a, \delta b)^T$, c'est l'accroissement infinitésimal de f alors $\exists\, \eta > 0$ tel que pour tous $(\delta a, \delta b)^T \in \mathcal{D} \subset \mathbb{R}^2$ vérifiant $\parallel (\delta a, \delta b)^T \parallel_2 < \eta$, nous avons :*

$$f(X + D) = \sum_{n=0}^{\infty} \frac{(D \cdot \nabla)^n\, f(X)}{n!} \tag{10}$$

Autrement dit, trouver la nature d'un point critique (maximum, minimum, selle, singulier) revient à étudier *localement le comportement* de la fonction f dans son voisinage $\parallel (\delta a, \delta b)^T \parallel_2 < \eta$, c'est-à-dire : $x_0 + \delta a \in [x_0 - \eta\,,\, x_0 + \eta]$ et $y_0 + \delta b \in [y_0 - \eta\,,\, y_0 + \eta]$. Tenant compte de (10), la série de *Taylor* au voisinage du point critique $(x_0 + \delta a, y_0 + \delta b)^T$ se développe comme suit :

$$\begin{aligned} f(x_0 + \delta a, y_0 + \delta b) = f(x_0, y_0) \,+\, \frac{(D \cdot \nabla)\, f(X)}{1!}\, f(x_0, y_0) \,+\, \frac{(D \cdot \nabla)^2\, f(X)}{2!}\, f(x_0, y_0) \,+\, \cdots\cdots \\ +\, \frac{(D \cdot \nabla)^p\, f(X)}{p!}\, f(x_0, y_0) \,+\, \cdots\cdots +\, \frac{(D \cdot \nabla)^{(n+1)}\, f(X)}{(n+1)!}\, f(x_0, y_0) \end{aligned}$$

1. Nous rappelons que l'équation de la tangente au point x_0 vaut : $f_T(x) = f(x_0) + f^{'}(x_0)\,\underbrace{(x - x_0)}_{\delta x}$.

Tronquons la série de *Taylor* à l'ordre deux, il vient :

$$f(x_0 + \delta a, y_0 + \delta b) = f(x_0, y_0) \, + \, \frac{(D \cdot \nabla)\, f(X)}{1!}\, f(x_0, y_0) \, + \, \frac{(D \cdot \nabla)^2\, f(X)}{2!}\, f(x_0, y_0) \, + \epsilon(\delta a^3, \delta b^3) \tag{11}$$

Avec $\nabla = \left(\dfrac{\partial}{\partial x}, \dfrac{\partial}{\partial y}\right)$ est l'opérateur différentiel. Calculons les termes linéaire et quadratique de la série (11) :

$$(D \cdot \nabla)\, f = \left((\delta a, \delta b) \cdot \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)\right)\, f \tag{12}$$

$$= \left(\delta a \, \frac{\partial f}{\partial x} \, + \, \delta b \, \frac{\partial f}{\partial y}\right) \tag{13}$$

$$\tag{14}$$

De façon analogue pour l'ordre quadratique,

$$(D \cdot \nabla)^2\, f = \left((\delta a, \delta b) \cdot \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)\right)^2\, f \tag{15}$$

$$= \left(\delta a \, \frac{\partial f}{\partial x} \, + \, \delta b \, \frac{\partial f}{\partial y}\right)^2 \tag{16}$$

$$= \delta a^2 \, \frac{\partial^2 f}{\partial x^2} \, + \, + 2\, \delta a \, \delta b \, \frac{\partial^2 f}{\partial x \, \partial y} \, + \, \delta b^2 \, \frac{\partial^2 f}{\partial y^2} \tag{17}$$

$$\tag{18}$$

Substituons (14) et (18) dans la série (11), nous obtenons :

$$f(x_0 + \delta a, y_0 + \delta b) = f(x_0, y_0) \, + \, \left(\delta a \, \frac{\partial f}{\partial x} \, + \, \delta b \, \frac{\partial f}{\partial y}\right) \, + \, \frac{1}{2}\left[\delta a^2 \, \frac{\partial^2 f}{\partial x^2} \, + \, 2\, \delta a \, \delta b \, \frac{\partial^2 f}{\partial x \, \partial y} \, + \, \delta b^2 \, \frac{\partial^2 f}{\partial y^2}\right] + \epsilon(\delta a^3, \delta b^3) \tag{19}$$

$$\Rightarrow \; \Delta f = \left(\delta a \, \frac{\partial f}{\partial x} \, + \, \delta b \, \frac{\partial f}{\partial y}\right) \, + \, \frac{1}{2}\left[\delta a^2 \, \frac{\partial^2 f}{\partial x^2} \, + \, 2\, \delta a \, \delta b \, \frac{\partial^2 f}{\partial x \, \partial y} \, + \, \delta b^2 \, \frac{\partial^2 f}{\partial y^2}\right] + \epsilon(\delta a^3, \delta b^3) \tag{20}$$

Le vecteur $X^* = (x_0, y_0)^T$ étant un point critique alors :

$$\frac{\partial f(x_0, y_0)}{\partial x} = \frac{\partial f(x_0, y_0)}{\partial y} = 0 \tag{21}$$

Substituons (21) dans (20), il en découle :

$$\Delta f = \frac{1}{2}\left[\delta a^2 \, \frac{\partial^2 f}{\partial x^2} \, + \, 2\, \delta a \, \delta b \, \frac{\partial^2 f}{\partial x \, \partial y} \, + \, \delta b^2 \, \frac{\partial^2 f}{\partial y^2}\right] + \epsilon(\delta a^3, \delta b^3) \tag{22}$$

Sous forme matricielle cette dernière relation s'écrit :

$$\Delta f = \frac{1}{2}\, (\delta a, \delta b)^T \times \begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \, \partial y} \\ & \\ \dfrac{\partial^2 f}{\partial y \, \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{pmatrix} \times \begin{pmatrix} \delta a \\ \delta b \end{pmatrix} + \epsilon(\delta a^3, \delta b^3) \tag{23}$$

Cette formulation met en exergue la matrice carrée des dérivées partielles d'ordre 2, dite matrice *Hessienne*. Le *déterminant* et la *trace* de cette matrice jouent un rôle fondamental dans la détermination de la nature des points critiques. La généralisation de cette matrice à $\mathbb{R}^{n\times n}$ est immédiate :

$$H_{ij}(f) = \frac{\partial^2 f}{\partial x_i\, x_j} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\, x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \tag{24}$$

Bien entendu cela suppose l'existence des dérivées partielles secondes de la fonction. Nous remarquons que la matrice *Hessienne* de f est symétrique du fait des dérivées secondes mixtes $\partial_{xy} f$ et $\partial_{yx} f$. On rappellera que pour une fonction de deux variables $f(x_1, x_2)$, nous avons la possibilité de constituer quatre dérivées partielles.

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right) \quad ; \quad \frac{\partial^2 f}{\partial x_1\,\partial x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right) \quad ; \quad \frac{\partial^2 f}{\partial x_2\,\partial x_1} = \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)$$

Cependant, seulement trois de ces dérivées sont distinctes. Ainsi, on démontre que

$$\frac{\partial^2 f}{\partial x_1\,\partial x_2} = \frac{\partial^2 f}{\partial x_2\,\partial x_1}$$

On saisit à ce propos que pour une dérivée partielle du second ordre mixte, l'ordre dans lequel on calcule les dérivées n'a pas d'impact sur le résultat. Notons aussi que le graphe d'une fonction de plusieurs variables, par exemple en deux dimensions, est défini formellement :

$$\mathcal{G}_f \equiv \{(x, y, z) = f(x, y)) \mid (x, y) \in \mathcal{D}\}$$

Ainsi, à tout point $(x, y) \in \mathcal{D}$ ayant l'image $f(x, y) \in \mathbb{R}$ correspond un point du graphe $(x, y, f(x, y)) \in \mathbb{R}^3$. Ce champ de points formera le relief de la fonction en question.

b) ***Théorème****: Soit f une fonction de classe $\mathcal{C}^2\left([\mathcal{D} \subset \mathbb{R}^n]\right)$. Soit $X^* = (x_0, y_0, z_0, \cdots) \in \mathbb{R}^n$ un point critique de $f(X) \Rightarrow \nabla f(X^*) = 0$. La matrice Hessienne $H_{ij}(f)$ au point $X^* = (x_0, y_0, z_0, \cdots)$ est symétrique dont les valeurs propres $\{\lambda_i\}_{i=1,n} \in \mathbb{R}$ et $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p \leq \cdots \leq \lambda_n$.*

- *si $\{\lambda_i\}_{i=1,n} > 0 \Rightarrow f$ admet un minimum local au point critique $X^* = (x_0, y_0, z_0, \cdots)$. Dans ce cas de figure, la matrice $H_{ij}(f)$ est dite définie positive.*
- *si $\{\lambda_i\}_{i=1,n} < 0 \Rightarrow f$ admet un maximum local au point critique $X^* = (x_0, y_0, z_0, \cdots)$. Dans ce cas de figure, la matrice $H_{ij}(f)$ est dite définie négative.*
- *si $\{\lambda_i\}_{i=1,n} = 0 \Rightarrow$ il faudra analyser la série de Taylor a des ordres ≥ 3 pour pouvoir conclure sur le comportement de f au voisinage du point critique.*
- *si $\forall j \neq i$ tel que $\lambda_j = 0$, pour conclure sur la nature du point critique, il faudra déterminer le signe de la trace soit : $T_{H(f)}$.*

Comme il a été mentionné précédemment, la matrice *Hessienne* $H_{ij}(f)$, est symétrique donc elle est diagonalisable. Le signe est conservé par changement de base. Cela signifie concrètement que les signes de $\{\lambda_i\}_{i=1,n}$ de la matrice diagonale sont identiques à ceux des éléments diagonaux de la *Hessienne*. Par conséquent, il n'est pas nécessaire de calculer les valeurs propres pour conclure sur la nature d'un point critique. Du fait de la conservation du signe, après changement de base, nous nous baserons uniquement sur la matrice $H_{ij}(f)$. Nous en déduisons les cas de figure ci-dessous :

A- si $\mid H_{ij}(f) \mid > 0 \Rightarrow$ les $\{\lambda_i\}_{i=1,n}$ ont le même signe.

A-1 si la trace $T_{H(f)} > 0$ alors les $\{\lambda_i\}_{i=1,n} > 0$. De la formule de *Taylor* (23), il en découle $\Delta f > 0 \Rightarrow f(x_0 + \delta a, y_0 + \delta b) > f(x_0, y_0)$ alors la fonction f prend des valeurs supérieures au voisinage $(x_0 + \delta a, y_0 + \delta b)$. La fonction croît, nous concluons que le point critique (x_0, y_0) est nécessairement un minimum local.

A-2 si la trace $T_{H(f)} < 0$ alors les $\{\lambda_i\}_{i=1,n} < 0$. De la formule de *Taylor* (23), il en découle $\Delta f < 0 \Rightarrow f(x_0 + \delta a, y_0 + \delta b) < f(x_0, y_0)$ alors la fonction f prend des valeurs inférieures au voisinage $(x_0 + \delta a, y_0 + \delta b)$. La fonction diminue, nous concluons que le point critique (x_0, y_0) est nécessairement un maximum local.

B- si $\mid H_{ij}(f) \mid < 0 \Rightarrow$ les $\{\lambda_i\}_{i=1,n}$ sont de signes contraires alors la fonction f admet un point *selle*.

C- si $\mid H_{ij}(f) \mid = 0 \Rightarrow \forall j \neq i,\ \lambda_j = 0$. Pour conclure sur la nature du point critique, il faut déterminer le signe de la trace de la *Hessienne*.

C-1 si la trace $T_{H(f)} > 0$ alors f admet un minimum local.

C-2 si la trace $T_{H(f)} < 0$ alors f admet un maximum local.

Pour le cas d'une fonction de trois variables $X \in \mathbb{R}^3$, d'après la théorie développée précédemment, un déterminant $\mid H_{ij}(f) \mid > 0$ signifie un produit positif des valeurs propres $\Pi_{i=1}^{3} \lambda_i > 0$. Par conséquent nous pouvons dégager deux séquences de signes (+, +, +) ou (-, -, +) sachant que d'après le théorème ci-dessus nous obtenons toujours $\lambda_1 \leq \lambda_2 \leq \lambda_3$. Afin de conclure sur la nature des points critiques nous devons calculer la trace de la *Hessienne*.

D-1 si la trace $T_{H(f)} < 0$ alors le point critique est un *point-selle*.

D-2 si la trace $T_{H(f)} > 0$ c'est un cas ambigu. Il faudra calculer les valeurs propres.

D-1-1 si $\Pi_{i=1}^{3} \lambda_i > 0$ alors f admet un minimum local.

D-1-2 si $\Pi_{i=1}^{3} \lambda_i > 0$ alors si toutes les valeurs propres sont positives, f admet un minimum local. Si au moins une des valeurs propres est négative alors il s'agit d'un *point-selle*.

Une autre manière de vérifier si $H_{ij}(f)$ est définie positive (existence d'un minimum local), ou définie négative (existence d'un maximum local) est de considérer :

$$H_{ij}(f) \text{ est définie positive} \Longrightarrow \forall X \neq 0 \in \mathbb{R}^n : X^T\, H_{ij}(f)\, X > 0 \tag{25}$$

$$H_{ij}(f) \text{ est définie négative} \Longrightarrow \forall X \neq 0 \in \mathbb{R}^n : X^T\, H_{ij}(f)\, X < 0 \tag{26}$$

Une matrice $H_{ij}(f) \in \mathbb{R}^{n \times n}$ ne vérifiant aucunes de ces propriétés est dite *indéfinie*[2]. Un moyen pratique de vérifier si une matrice est définie positive est de calculer les n-déterminants[3] des sous-matrices de $H_{ij}(f)$,

2. Faudra-il encore vérifier que $H_{ij}(f)$ est semi-définie positive $\Longrightarrow \forall X \in \mathbb{R}^n : X^T\, H_{ij}(f)\, X \geq 0$ et $H_{ij}(f)$ est semi-définie négative $\Longrightarrow \forall X \in \mathbb{R}^n : X^T\, H_{ij}(f)\, X \leq 0$. Par ailleurs, une matrice symétrique vérifie $A = A^T$, ses valeurs propres sont réelles. De plus si A est définie positive alors elle est inversible et ses valeurs propres $\{\lambda_i\}_{i=1,n}$ sont positives

3. Ces n-déterminants sont appelés les *mineurs principaux dominants*.

soit :

$$\triangle_1 = | \, a_{11} \, | \qquad \triangle_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \qquad \triangle_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \quad \cdots\cdots \quad \triangle_n = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}$$

1) Si $\{\triangle_i\}_{i=1,n} > 0 \Rightarrow H_{ij}(f)$ est définie positive.
2) Si $\{\triangle_i\}_{i=1,n} < 0 \Rightarrow H_{ij}(f)$ est définie négative.

Concluons cette analyse sur la nature d'un point critique par le théorème ci-dessous.

Théorème *(conditions suffisantes d'optimalité) : Soit f une application $\mathbb{R}^n \longrightarrow \mathbb{R}$ de classe $\mathcal{C}^2$ sur $\mathbb{R}^n$. Si $X^* \in \mathbb{R}^n$ est un minimum local (respectivement un maximum local) de f alors :*

$$\nabla f(X^*) = 0 \qquad\qquad \textit{Condition d'optimalité du premier ordre}$$

et de plus,

$$H_{ij}(f(X^*)) = \nabla^2 f(X^*) \qquad\qquad \textit{Condition d'optimalité du second ordre}$$

est définie positive (respectivement définie négative)

La condition du premier ordre sur le gradient de la fonction, $\nabla f(X^*) = 0$, permet d'identifier un point stationnaire sans que l'on sache s'il s'agit d'un minimum, d'un maximum ou d'un point de selle. La condition du second ordre portant sur le *Hessien* de f, $\nabla^2 f(X^*)$, détermine la nature de ce point critique.

B. ***Définition du point-selle***

Soient $\mathcal{D}_1$ et $\mathcal{D}_2$ deux sous-ensembles et $f : \mathcal{D}_1 \times \mathcal{D}_2 \longmapsto \mathbb{R}$. Nous disons que le point critique $(x_s, y_s) \in \mathcal{D}_1 \times \mathcal{D}_2$ est un point-selle de f sur $\mathcal{D}_1 \times \mathcal{D}_2$ si :

$$\forall (x, y) \in \mathcal{D}_1 \times \mathcal{D}_2 : f(x_s, y) \leq f(x_s, y_s) \leq f(x, y_s) \tag{27}$$

Dans ce cas $f(x_s, y_s)$ est appelée la valeur-selle de $f(x, y)$. Autrement dit, $y \longmapsto f(x_s, y)$ atteint un maximum en y_s sur $\mathcal{D}_2$ et $x \longmapsto f(x, y_s)$ atteint un minimum en x_s sur $\mathcal{D}_1$.

Comme exemple d'application, considérons :

$$f : \underbrace{[-2 \; ; \; 2]}_{\mathcal{D}_1} \times \underbrace{[-2 \; ; \; 2]}_{\mathcal{D}_2} \longmapsto \mathbb{R}$$
$$(x, y) \longmapsto f(x, y) = x^2 - y^2$$

Ci-dessous le script Matlab® correspondant.

```
clear all ; clc ; close all ;

% Samir KENOUCHE Le 08/08/2019
% VISUALISATION DU POINT-SELLE DE F(X,Y) = X^2 - Y^2
lB = -2 ; uB = 2 ; n = 50 ; h = (uB - lB)/n ;

% POINT CRITIQUE => GRADIENT NUL AU POINT X* = (xs,ys) = (0,0)
xs = 0 ; ys = 0 ;  % POINT CRITIQUE OU STATIONNAIRE

[x,y] = meshgrid(lB :h: uB) ; fun = x.^2 - y.^2 ;
figure('color',[1 1 1]) ; surf(x,y,fun) ; hold on ;
plot(xs,ys,'ro','MarkerSize',22,'LineWidth',2) ;
plot(xs,ys,'rx','MarkerSize',22,'LineWidth',2) ; xlabel('x') ; ylabel('y') ;
title('POINT-SELLE DE F(X,Y) = X^2 - Y^2 AVEC (Xs,Ys) = (0,0)') ;
view(-32,14) ;
```

Exercice ❶ ☞ ®

Soit la fonction définie sur un ouvert $\mathcal{D} \subset \mathbb{R}^2$ et deux fois continument dérivable.

$$f(X) = x^4 + y^4 - 4\,x\,y + 1 \tag{28}$$

1) Déterminer les points critiques de f.
2) Identifier la nature de ces points critiques.

Solution: *(a) les points critiques sont déterminés en annulant le gradient* $\nabla f(X) = 0$ *:*

$$\begin{cases} \dfrac{\partial f}{\partial x} = 4\,x^3 - 4\,y = 4\,(x^3 - y) = 0 \\[2ex] \dfrac{\partial f}{\partial y} = 4\,y^3 - 4\,x = 4\,(y^3 - x) = 0 \end{cases} \tag{29}$$

Ce qui donne,

$$\frac{\partial f}{\partial x} = (x^3 - y) = 0 \tag{30}$$

$$\frac{\partial f}{\partial y} = (y^3 - x) = 0 \tag{31}$$

De l'équation (30), *nous déduisons*

$$y = x^3 \tag{32}$$

Par substitution dans l'équation (31) *nous obtenons*[4] *:*

$$x^9 - x = 0 \Leftrightarrow x\,(x^8 - 1) = x\,(x^4 - 1)\,(x^4 + 1) = 0 \tag{33}$$

$$\Rightarrow x\,(x^2 - 1)\,(x^2 + 1)\,(x^4 + 1) = 0 \tag{34}$$

4. Nous utilisons la propriété suivante : $(a^2 - b^2) = (a - b) \times (a + b)$

$$\Rightarrow x = 0 \quad ; \quad x = 1 \quad ; \quad x = -1 \tag{35}$$

La solution pour laquelle $x^2 + 1 = 0 \Rightarrow x = \sqrt{j^2\,1} = \pm j \in \mathbb{C}$ *est refusée car* $X = (x, y) \in \mathbb{R}^2$*. A partir de* (32) *nous en déduisons les points critiques suivants :*

$$x = 0 \quad \Rightarrow \quad y = 0 \quad \Rightarrow \quad X_c^{(1)} = (0\ ;\ 0)^T$$

$$x = 1 \quad \Rightarrow \quad y = 1 \quad \Rightarrow \quad X_c^{(2)} = (1\ ;\ 1)^T$$

$$x = -1 \quad \Rightarrow \quad y = -1 \quad \Rightarrow \quad X_c^{(3)} = (-1\ ;\ -1)^T$$

(b) Identifions désormais la nature des ces points critiques. Nous devons donc déterminer la matrice Hessienne $\in \mathbb{R}^{2\times 2}$*. Calculons les éléments de cette matrice :*

$$\begin{cases} \dfrac{\partial f}{\partial x} = 4\,x^3 - 4\,y \quad \Rightarrow \quad \dfrac{\partial^2 f}{\partial x^2} = 12\,x^2 \\[2ex] \dfrac{\partial f}{\partial y} = 4\,y^3 - 4\,x \quad \Rightarrow \quad \dfrac{\partial^2 f}{\partial y^2} = 12\,y^2 \end{cases} \tag{36}$$

Pour les éléments hors de la diagonale principale :

$$\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial^2 f}{\partial y\,\partial x} = -4 \tag{37}$$

D'où la matrice Hessienne,

$$\begin{pmatrix} 12\,x^2 & -4 \\ -4 & 12\,y^2 \end{pmatrix} \tag{38}$$

Calculons son déterminant,

$$|H_{2\times 2}(f)| = \begin{vmatrix} 12\,x^2 & -4 \\ -4 & 12\,y^2 \end{vmatrix} = 12\,x^2 \times 12\,y^2 - 16 \tag{39}$$

$$- X_c^{(1)} = (0\ ;\ 0)^T \quad \Rightarrow \quad |H_{2\times 2}(f)| = -16 < 0 \quad \Rightarrow \quad X_c^{(1)} \quad \textit{est un point selle}$$
$$- X_c^{(2)} = (1\ ;\ 1)^T \quad \Rightarrow \quad |H_{2\times 2}(f)| = 128 > 0 \quad \textit{et} \quad T_{H(f)} = 12\,x^2 + 12\,y^2 = 24 > 0$$
$$\Rightarrow \quad X_c^{(2)} \quad \textit{est un minimum local}$$
$$- X_c^{(3)} = (-1\ ;\ -1)^T \quad \Rightarrow \quad |H_{2\times 2}(f)| = 128 > 0 \quad \textit{et} \quad T_{H(f)} = 12\,x^2 + 12\,y^2 = 24 > 0$$
$$\Rightarrow \quad X_c^{(3)} \quad \textit{est un minimum local}$$

Exercice ❷ ☞ Ⓢ

Considérons les fonctions de trois variables, $f(X)$ tel que $X \in \mathbb{R}^3$, suivantes :

$$f_1(X) = x_1^2 + 2\,x_2^2 + 2\,x_3^2 + x_1\,x_2 + 2\,x_2\,x_3 + 4\,x_1 + 6\,x_2 + 12\,x_3$$
$$f_2(X) = 6\,x_1^2 + x_2^3 + 3\,x_2^2 + 6\,x_1\,x_2 + \frac{1}{4}\,x_3^4 - \frac{1}{3}\,x_3^3$$

$$f_3(X) = x_1^2 - 2\,x_2^2 - 3\,x_3^2 + 4\,x_1\,x_2 + 6\,x_1\,x_3 - 8\,x_2\,x_3$$

1) Identifier analytiquement les points stationnaires.
2) Déterminer la nature de ces points stationnaires.
3) Écrire chaque fonction sous la forme quadratique :

$$f(X) = \frac{1}{2}\,X^T\,A\,X - B^T\,X \quad \text{tel que} \quad \mathrm{A} \in \mathbb{R}^{n\times n} \quad \text{et} \quad B^T \in \mathbb{R}^n \tag{40}$$

4) Trouver la forme quadratique associée à la matrice :

$$\begin{pmatrix} 1 & 2 & 4 \\ 2 & 2 & 6 \\ 4 & 6 & 4 \end{pmatrix}$$

5) Tenant compte de (40), montrer que la forme quadratique algébrique s'écrit sous la forme :

$$f(X) = \sum_i a_i\,x_i^2 + 2\,\sum_i\,\sum_{j\neq i} b_{ij}\,x_i\,x_j \tag{41}$$