1 Notion de Convexité

1.1 Ensembles Convexes

Definition 1 (Ensemble Convexe) *Soit $C \subset \mathbb{R}^n$, C est dit **convexe** si*

$$\forall \lambda \in [0,1], \forall x, y \in C : (1-\lambda) x + \lambda y \in C. \tag{1}$$

ou d'une manière équivalente

$$\forall \alpha, \beta \in \mathbb{R}_+, \text{ avec } \alpha + \beta = 1, \forall x, y \in C : \ \alpha x + \beta y \in C.$$

Pour $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, on appele le segment relient x à y l'ensemble

$$[x, y] = \{(1-\lambda) x + \lambda y : \lambda \in [0,1]\}.$$

Donc $C \subset \mathbb{R}^n$ est convexe SSI $\forall x, y \in C$, $[x, y] \subset C$.

Example 2 *Un sous-espace vectoriel est évidemment convexe, de même pour un sous espace affine qui n'est rien d'autre que le translaté d'un sous-espace vectoriel.*

Example 3 *Soit l'ensemble $\Pi = \left\{x \in \mathbb{R}^3 : x_1 + 2x_2 - x_3 = 4\right\}$. L'ensemble Π peut aussi s'écrir comme suit: $\Pi = \left\{x \in \mathbb{R}^3 : x^\top p = 4\right\}$, avec $p = (1, 2, 3)^\top$. L'ensemble Π est appelé hyperplan[1], le vecteur p est dit vecteur normal. L'ensemble Π est-il convexe? En effet: Soient $\lambda \in [0,1]$ et $x, y \in \Pi$. Montrons que $(1-\lambda) x + \lambda y \in \Pi$ i.e. $p^T ((1-\lambda) x + \lambda y) \stackrel{?}{=} 4$. On a, $\langle p, (1-\lambda) x + \lambda y\rangle = (1-\lambda) \langle p, x\rangle + \lambda \langle p, y\rangle = (1-\lambda) 4 + \lambda 4 = 4$.Donc, $(1-\lambda) x + \lambda y \in \Pi$. La convexité de Π s'ensuit. D'une manière générale un hyperplan H de $\mathbb{R}^n$ est défini à partir d'un vecteur normal p et une constante α, c-à-d $H = \left\{x \in \mathbb{R}^n : x^\top p = \alpha\right\}$. L'ensemble $H_\leq = \left\{x \in \mathbb{R}^n : x^\top p \leq \alpha\right\}$ est appelé demi espace, d'une manière équivalente on définit l'autre demi espace $H_\geq = \left\{x \in \mathbb{R}^n : x^\top p \geq \alpha\right\}$ associer à l'hyperplan H. Le lecteur peut vérifier que $H_\leq$ et $H_\geq$ sont aussi des ensembles convexes.*

Exercise 4 *Montrer que H est un sous-espace affine si et seulement si $\forall (x, y) \in H^2, \forall \alpha, \beta \in \mathbb{R}_+ : \alpha x + \beta y \in H$.*

Exercise 5 *Montrer que C est convexe si et seulement si*

$$C = \left\{ \sum_{i=1}^{k} \alpha_i c_i : k \in \mathbb{N}, c_i \in C \text{ et } \alpha_i \geq 0 \text{ pour } i = 1, ..., k, \text{ avec } \sum_{i=1}^{k} \alpha_i = 1 \right\}.$$

Proposition 6 *Soit $\{C_i\}_{i \in I}$ une famille d'ensemble convexe (I est un ensemble d'indice quelconque dénombrable ou non). Alors, $\cap_{i \in I} C_i$ est convexe.*

Proof. Soient $x, y \in \cap_{i \in I} C_i$ et $\lambda \in [0,1] : (1-\lambda) x + \lambda y \stackrel{??}{\in} \cap_{i \in I} C_i$. On a $x, y \in \cap_{i \in I} C_i \iff x \in C_i, \, y \in C_i, \, \forall i \in I$.Les C_i sont tous convexes pour tout $i \in I$, donc $(1-\lambda) x + \lambda y \in C_i, \forall i \in I \iff (1-\lambda) x + \lambda y \in \cap_{i \in I} C_i$. ■

[1] C'est aussi un sous espace affine c-à-d $\Pi = F + a$, où F est un sous-espace vectoriel de $\mathbb{R}^n$ et a un vecteur fixé dans $\mathbb{R}^n$.

1.2 Fonctions Convexes

Definition 7 (Fonction Convexe) *Soit C un ensemble convexe de $\mathbb{R}^n$ et $f : C \to \mathbb{R}$. f est dite convexe si*

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

ou aussi

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

ou aussi

$$\forall p, q \geq 0,\ p + q = 1 : f(px + qy) \leq pf(x) + qf(y),$$

Example 8 *Soit $L : \mathbb{R}^n \to \mathbb{R}$ une application linéaire, et $\alpha \in \mathbb{R}$, on définit une application affine h par $h(x) = L(x) + \alpha$.Soient $\lambda \in [0,1]$, $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$, on a $h((1-\lambda)x + \lambda y) = L((1-\lambda)x + \lambda y) + \alpha$. La linéarité de L donne $h((1-\lambda)x + \lambda y) = (1-\lambda)L(x) + \lambda L(y) + \alpha = (1-\lambda)L(x) + \lambda L(y) + (1-\lambda)\alpha + \lambda\alpha = (1-\lambda)(L(x) + \alpha) + \lambda(L(y) + \alpha) = (1-\lambda) f(x) + \lambda f(y)$. D'où f est convexe.*

Definition 9 (Fonction Concave) *Soit C un ensemble convexe de $\mathbb{R}^n$ et $f : C \to \mathbb{R}$. f est dite concave si $(-f)$ est convexe i.e.*

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

Remark 10 *Toute application affine est convexe et concave à la fois.*

Definition 11 (Fonction Strictement Convexe) *Soit C un ensemble convexe de $\mathbb{R}^n$ et $f : C \to \mathbb{R}$. La fonction f est dite strictement convexe si*

$$\forall\lambda \in \left]0,1\right[, \forall x, y \in C \text{ et } x \neq y : f((1-\lambda)x + \lambda y) < (1-\lambda) f(x) + \lambda f(y).$$

Remark 12 *Une fonction strictement convexe est évidemment convexe. La réciproque n'est pas vraie, il suffit de prendre l'application affine comme contre exemple*

La notion d'ensemble suivante met bien le lien entre fonction et ensemble convexe.

Definition 13 (Epigraphe) *L'ensemble*

$$\{(x,\alpha) : f(x) \leq \alpha\} \subset \mathbb{R}^n \times \mathbb{R} \quad (3)$$

est appelé épigraphe de f. On le note $epif$. L'ensemble

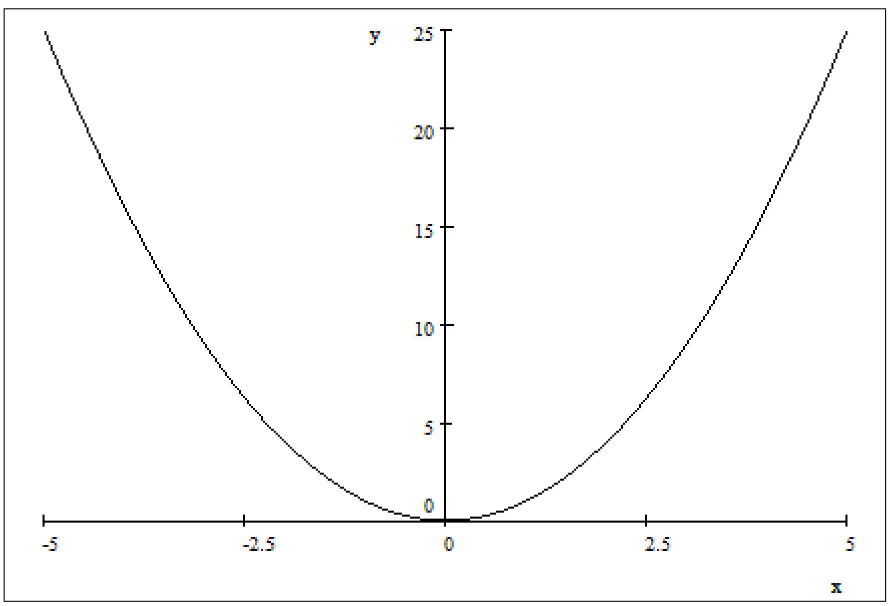
$$\{(x,\alpha) : f(x) \geq \alpha\} \subset \mathbb{R}^n \times \mathbb{R}$$

est appelé hypographe.

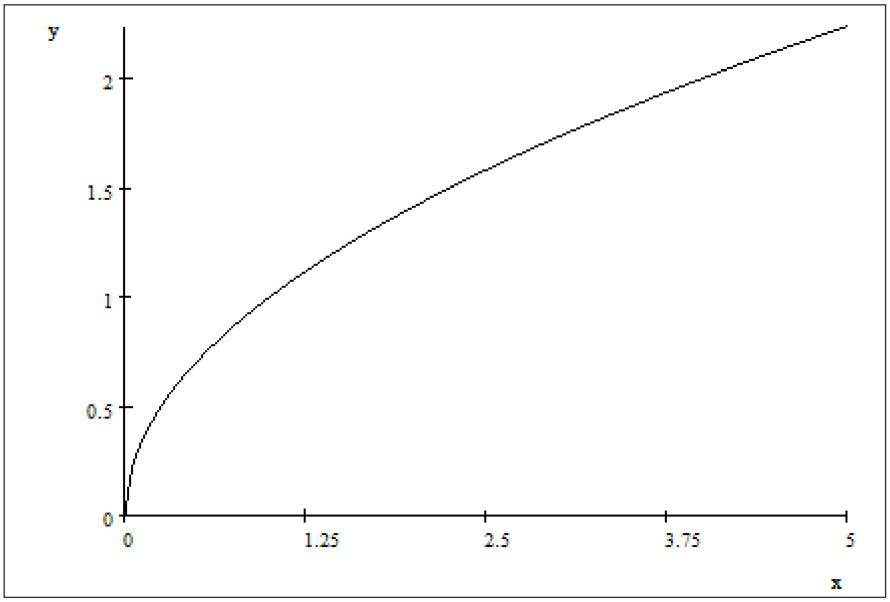
Proposition 14 *Soit $f : C \subset \mathbb{R}^n \to \mathbb{R}$, avec C convexe. Alors, f est convexe (concave) si et seulement si son épigraphe (hypographe) est convexe.*
Proof. *Montrons la nécessité et supposons que f est convexe. Soient $(x_1, \alpha_1), (x_2, \alpha_2) \in epif$: $(1-\lambda)(x_1, \alpha_1) + \lambda(x_2, \alpha_2) \overset{??}{\in} epif \Leftrightarrow ((1-\lambda)x_1 + \lambda x_2, (1-\lambda)\alpha_1 + \lambda\alpha_2) \overset{??}{\in} epif \iff f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)\alpha_1 + \lambda\alpha_2$. En effet: f est convexe, donc $f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2) \leq (1-\lambda)\alpha_1 + \lambda\alpha_2$, la dérnière inégalité est du au fait que $(x_1, \alpha_1) \in epif$ et $(x_2, \alpha_2) \in epif$, ainsi $epif$ est convexe. Pour montrer la suffisance, on suppose que $epif$ est convexe et montrons la convexité de f. Soient $x_1, x_2 \in C$ et $\lambda \in [0,1]$, On a $(x_1, f(x_1)) \in epif$ et $(x_2, f(x_2)) \in epif$. Par convexité de l'épigraphe on déduit que, $(1-\lambda)(x_1, f(x_1)) + \lambda(x_2, f(x_2)) \in epif$. Ceci se traduit par $(1-\lambda)(x_1, f(x_1)) + \lambda(x_2, f(x_2)) = ((1-\lambda)x_1 + \lambda x_2, (1-\lambda)f(x_1) + \lambda f(x_2)) \in epif \implies f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2)$. D'où le résultat.* ■

Example 15 *Les fonctions suivantes sont convexes ou concaves selon la convexité de leurs épigraphe ou hypographe.*

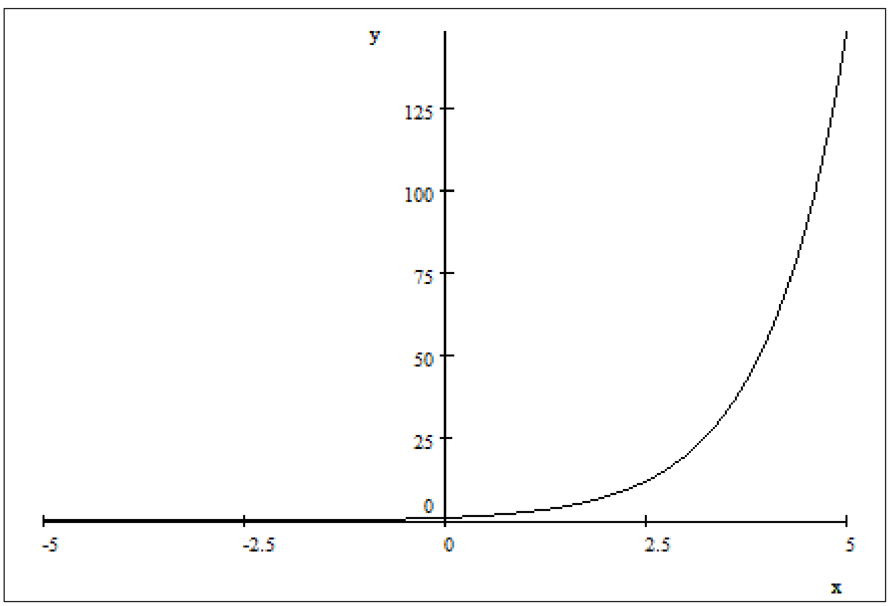
- $x \mapsto x^2$ *est convexe*



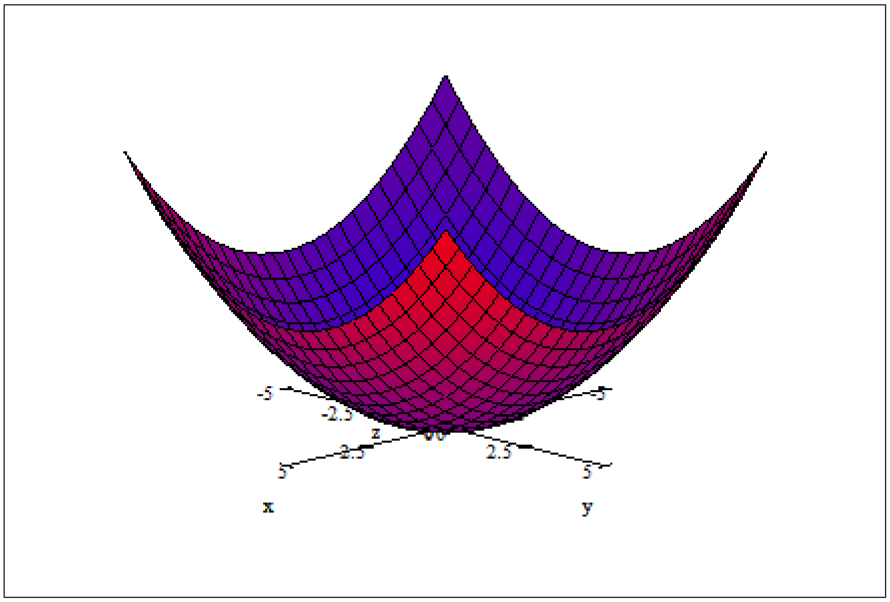
- $x \mapsto \sqrt{x}$ *est concave.*



- $x \mapsto \exp(x)$ *est convexe.*



- $(x_1, x_2) \mapsto \frac{1}{2}\left(x_1^2 + x_2^2\right)$ *est convexe.*



1.3 Propriétés des fonctions convexes

Proposition 16 *Soient $f, g : C \subset \mathbb{R}^n \to \mathbb{R}$, avec C convexe, f et g deux fonctions convexes.*

1. *$f + g$ est convexe.*
2. *Si $\alpha > 0$, αf est convexe.*
3. *La fonction $h(x) = \max(f(x), g(x))$ est convexe.*

Les fonction convexes ne peuvent avoir un point de discontinuité qu'aux points frontières de leurs domaine, comme le montre le théorème suivant.

Theorem 17 *Soient C un convexe de $\mathbb{R}^n$ d'intérieur non vide et $f : C \to \mathbb{R}$ une fonction convexe. Alors, f est continue sur $\mathrm{int}C$ (intérieur de C).*

Theorem 18 (Caractérisation de la convexité par le gradient) *Soient $C \subset \mathbb{R}^n$ un ouvert convexe et $f : C \to \mathbb{R}$, différentiable. Alors, f est convexe sur C si et seulement si*

$$\forall x, y \in C : f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle . \tag{4}$$

Proof. *Soit f est convexe i.e.*

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

En retranchant $f(x)$ des deux termes de l'inigalité et en dévisant par λ, on obtient

$$\frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \leq f(y) - f(x).$$

En faisant tendre λ vers 0, on a la nécessité

$$\lim_{\lambda \to 0} \frac{f(x + \lambda(y - x)) - f(x)}{\lambda} = \langle \nabla f(x), y - x \rangle \leq f(y) - f(x).$$

Soient à présent $x, y \in C$ et $\lambda \in [0,1]$. Posons $x_\lambda = x + \lambda(y - x)$. En retranchant respectivement x et y de x_λ on obtient

$$\begin{cases} x_\lambda - x = \lambda(y - x) \\ x_\lambda - y = (1 - \lambda)(x - y). \end{cases}$$

En appliquant la relation (4) à la fonction f respectivement aux point (x, x_λ) et (y, x_λ), on obtient

$$\begin{cases} f(x) \geq f(x_\lambda) - \lambda \langle \nabla f(x_\lambda), y - x \rangle \\ f(y) \geq f(x_\lambda) + (1 - \lambda) \langle \nabla f(x_\lambda), y - x \rangle \end{cases}.$$

En multipliant respectivement ces deux inégalités par $(1 - \lambda)$ et λ

$$\begin{cases} (1 - \lambda) f(x) \geq (1 - \lambda) f(x_\lambda) - \lambda(1 - \lambda) \langle \nabla f(x_\lambda), y - x \rangle \\ \lambda f(y) \geq \lambda f(x_\lambda) + \lambda(1 - \lambda) \langle \nabla f(x_\lambda), y - x \rangle \end{cases}$$

En somment les deux inégalités on obtient la relation (2) assurant la convexité de f. ■

Theorem 19 (Caractérisation de la convexité par le Hessien) *Soit $C \subset \mathbb{R}^n$ un ouvert convexe; $f : C \to \mathbb{R}$, de classe C^2. Alors,*

f est convexe sur C si et seulement si pour tout $x \in C, \nabla^2 f(x)$ est semi-définie positive sur C i.e.

$$\forall y \in \mathbb{R}^n : \left\langle y, \nabla^2 f(x) y \right\rangle \geq 0. \tag{5}$$

f est strictement covexe sur C si $\nabla^2 f(x)$ est définie positive.

Proof. Supposons que f est convexe, et soit $x \in C$. On veut montrer la relation (5). Vue que C est ouvert, alors pour tout $y \in \mathbb{R}^n$, il existe λ suffisamment petite avec $|\lambda| \neq 0$ et $x + \lambda y \in C$. Du théorème précédent et la différentiabilté d'ordre 2 de f on a

$$f(x + \lambda y) \geq f(x) + \lambda \langle \nabla f(x), y \rangle, \tag{6}$$

$$f(x + \lambda y) = f(x) + \lambda \langle \nabla f(x), y \rangle + \frac{1}{2} \lambda^2 \left\langle \nabla^2 f(x) y, y \right\rangle + \lambda^2 \|x\|^2 \varepsilon(\lambda x). \tag{7}$$

En substituant, (7) dans (6) on obtien

$$\frac{1}{2} \lambda^2 \left\langle \nabla^2 f(x) y, y \right\rangle + \lambda^2 \|x\|^2 \varepsilon(\lambda x) \geq 0. \tag{8}$$

En divisons la relation (8) par λ^2 et en faisant tendre $\lambda \to 0$, d'où la relation (5). Inversement, supposons que la matrice Hessienne et semidéfinie positif en tout point dans C. Considérons x et y dans C, du développement de Taylor avec reste de la Moyenne, on a

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \left\langle \nabla^2 f(\widehat{y}) y - x, y - x \right\rangle \tag{9}$$

où $\widehat{y} = \lambda x + (1 - \lambda) y$ pour un certain $\lambda \in]0, 1[$. Notons que $\widehat{y} \in C$ par convexité, de la semidéfinie positivité de la matrice Hessienne $\nabla^2 f(\widehat{y})$ et de (9), on déduit que

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle .$$

Puisque x et y sont quelconque dans C, la convexité de f se déduit du théorème précédent. ■

Considérons la forme quadratique q définie par la relation (**??**), on sait déjà que sa matrice Hessienne est H. Ainsi, du précédent théorème, en déduit que q est convexe (strictement convexe) si et seulement si la matrice Hessienne H est semidéfinie positive (définie positive). Le théorème suivant, un classique de l'agèbre linéaire nous donne une caractérisation pratique.

Theorem 20 *Une matrice $H \in S_n(\mathbb{R})$, est définie positive (semidéfinie positive) si et seulement si ses valeurs propres sont strictement positives (positives).*

Corollary 21 *Une forme quadratique q définie par la relation (**??**), est convexe (strictement convexe) si et seulement si les valeurs propres de la matrice Hessienne H sont positives (strictement positives).*

La convexité d'une forme quadratique nécessite le calcul de tous ces valeurs propres, ceci peut s'avérer couteux point de vu numérique. Le dernier résultat de ce chapitre connu sous le nom du Critère de Sylvester, donne une caractérisation moins couteuse numériquement. Donnons d'abord quelques définitions.

Definition 22 *Soit $H = (h_{ij})_{i,j} \in S_n(\mathbb{R})$, Une sous-matrice $k \times k$ formée, à partir de A, en éliminant $n-k$ colonnes, disons les colonnes $i_1, i_2, ..., i_{n-k}$ et les mêmes $n - k$ lignes $i_1, i_2, ..., i_{n-k}$, est appellée une sous-matrice de A, d'ordre principal k. Le déterminant d'une sous-matrice principale $k \times k$ est appellé le mineur principal d'ordre k de la matrice H.*

Example 23 *Soit H une matrice* 3×3

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}.$$

Les mineurs principaux de premier ordre sont tous les termes qui sont sur la diagonale. Pour une matrice 3×3, on a donc 3 mineurs principaux de premier ordre: h_{11}, h_{22} et h_{33}. Les mineurs principaux de second ordre sont les déterminants de toutes les sous-matrices 2×2 obtenues en éliminant la même ligne et colonne. On a trois mineurs principaux de second ordre

$$\det \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}, \det \begin{bmatrix} h_{11} & h_{13} \\ h_{31} & h_{33} \end{bmatrix} \text{ et } \det \begin{bmatrix} h_{22} & h_{23} \\ h_{32} & h_{33} \end{bmatrix}.$$

Un seul Mineur principal de troisième ordre; qui est le déterminant de la matrice 3×3.

Definition 24 *Soit* $H = (h_{ij})_{i,j} \in S_n(\mathbb{R})$*, Une sous-matrice* $k \times k$ *formée, à partir de* A*, en éliminant les* $n-k$ *dernières colonnes et les memes* $n-k$ *dernières lignes, est appellée une sous-matrice de* A*, d'ordre principal dominant* k*. Le déterminant d'une sous-matrice principale dominante* $k \times k$ *est appellé le mineur principal dominant d'odre* k *de la matrice* H*.*

Example 25 *Soit* H *une matrice* 3×3

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}.$$

Les mineurs principaux dominants de premier ordre est Seulement h_{11}*, obtenu en éliminant les 2 dernières lignes et colonnes. les mineurs principaux dominants de second ordre On a seulement*

$$\det \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

où la sous-matrice est obtenue en éliminant la troisième ligne et colonne. Le mineurs principal dominant de troisième ordre est un seul et c'est le determinant de la matrice H*.*

Theorem 26

1. *Une matrice* $H \in S_n(\mathbb{R})$*, est semi-définie positive si et seulement si chacun des mineurs principaux de* H *est* ≥ 0*.*

2. *Une matrice* $H \in S_n(\mathbb{R})$*, est semi-définie négative si et seulement si chacun des mineurs principaux d'ordre impaire de* H *est* ≤ 0*, et chacun des mineurs principaux d'ordre paire de* H *est* ≥ 0*.*

3. *Une matrice* $H \in S_n(\mathbb{R})$*, est définie positive si et seulement si chacun des mineurs principaux dominants de* H *est* > 0*.*

4. *Une matrice* $H \in S_n(\mathbb{R})$*, est définie négative si et seulement si ses* n *mineurs principaux dominant alternent en signe de la manière suivante:*

$$h_{11} < 0,\ \det \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} > 0,\ \det \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} < 0 \quad ect....$$

Example 27 *la matrice*

$$H = \begin{pmatrix} -4 & 4 \\ 4 & -6 \end{pmatrix}$$

n'est pas semidéfinie positive, est-elle définie négative? Il se trouve que $h_{11} < 0$ *et* $\det \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} > 0$*, d'où* H *est définie négative.*