Chapitre 1

Rappels mathématiques

1.1 Introduction

Lorsqu'on parle de l'optimisation nous devons impérativement passer par les mathématiques qui présentent l'outil indispensable afin de Les techniques d'optimisation se basent sur

1.1.1 Notations

Dans ce qui suit, on considère la notation des variables suivantes :

— $n, m, p, ...$, des scalaires qui indiquent souvant les dimensions,
— $x, y, t, \lambda, \alpha ...$, des variables scalaires,
— $\mathbf{x}, \mathbf{y}, \mathbf{X}, \mathbf{Y}, ...$, des vecteurs colonne,
— $\mathbf{A}, \mathbf{M}, \mathbf{P}, ...$, des matrices.
— $\mathbf{E}, \mathbf{I}, ...$, ensembles ou intervalles

1.1.2 Vecteur

Le vecteur réel $\mathbf{X}$ de dimension n, se met sous la forme :

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \tag{1.1}$$

ou encore de la manière suivante :

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}' \tag{1.2}$$

— deux vecteurs sont égaux ssi tous leurs éléments sont égaux.

— Le produit scalaire de deux vecteurs $\mathbf{X}$ et $\mathbf{Y}$ est défini par :

$$\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{X} = \sum_{i=1}^{n} x_i y_i \tag{1.3}$$

— Les deux vecteurs $\mathbf{X}$ et $\mathbf{Y}$ sont orthogonaux si :

$$\mathbf{X}'\mathbf{Y} = 0 \tag{1.4}$$

— La norme d'un vecteur $\mathbf{X}$ est :

$$|\mathbf{X}| = \sqrt{\mathbf{X}'\mathbf{X}} \tag{1.5}$$

— L'inégalité de *Cauchy-Schwarz* est vérifiée pour les deux vecteurs $\mathbf{X}$ et $\mathbf{Y}$:

$$|\mathbf{X}'\mathbf{Y}| \leq |\mathbf{X}|.|\mathbf{Y} \tag{1.6}$$

1.1.3 Matrice

Dans ce cours on ne considère que les matrice réelles. Une matrice $\mathbf{A}$ de dimension $m \times n$ est donnée par :

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \tag{1.7}$$

La matrice transposée de la matrice $\mathbf{A}$ qu'on note $\mathbf{A}'$ ou $\mathbf{A}^T$, est de dimension $n \times m$ est la suivante :

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} \tag{1.8}$$

La somme (différence) de deux matrice de dimensions identiques est une matrice de même dimension, dont les élements sont égaux à la somme (différence) des éléments de même position.

Le produit de deux matrice $\mathbf{A}$ $(m \times p)$ et $\mathbf{B}$ $(p \times n)$ est une matrice $\mathbf{C}$ $(m \times n)$ dont les élements sont :

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj} \qquad i = 1, ..., m \quad j = 1, ..., n \tag{1.9}$$

La multiplication des matrices requière une attention partculière, le nombre de colonne de la première doit être égale au nombre de ligne de la deuxième. Prenant l'exemple illsutratif suivant : soit un vecteur $\mathbf{X}$ de dimension n alors,

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \begin{bmatrix} x_1 & x_2 & ... & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} \\ &= x_1^2 + x_2^2 + ... + x_n^2 \end{aligned}$$

Le résultat du produit $\mathbf{X}'\mathbf{X}$ est un scalaire.

$$\begin{aligned} \mathbf{X}\mathbf{X}' &= \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} \begin{bmatrix} x_1 & x_2 & ... & x_n \end{bmatrix} \\ &= \begin{bmatrix} x_1^2 & x_1 x_2 & ... & x_1 x_n \\ x_2 x_1 & x_2^2 & ... & x_2 x_n \\ ... & ... & ... & ... \\ x_n x_1 & x_n x_2 & ... & x_n^2 \end{bmatrix} \end{aligned}$$

Le résultat du produit $\mathbf{X}\mathbf{X}'$ est une matrice.

Une matrice est dite carrée dans le cas particulier où $n = m$. On défini la matrice identité $\mathbf{I}$ de la manière suivante :

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{bmatrix} \tag{1.10}$$

Souvent on indique la dimension n de la matrice identité $\mathbf{I}_n$.

Exercice 1. *Opérations sur les matrices et les vecteurs*
. Soit les deux vecteurs $\mathbf{A} = [1, 2, 3]'$ et $\mathbf{B} = [-1, 0, 2]'$. Effectuer les opérations suivantes :

a. $\mathbf{AB}'$.

b. $\mathbf{A}'\mathbf{B}$.

c. $\mathbf{AB}'\mathbf{B}$

1.1.4 Valeurs propres

On appel le scalaire λ valeur propre de la matrice $\mathbf{A}$ s'il existe un vecteur non nul $\mathbf{V}$ tel que :

$$\mathbf{AV} = \lambda\mathbf{V} \tag{1.11}$$

Le vecteur $\mathbf{V}$ est appelé vecteur propre de la matrice $\mathbf{A}$ associé à la valeur propre λ (**Exp.** $\mathbf{A} = [3\ 2\ ;\ 3\ \ -2], \mathbf{V} = [2\ 1]', \lambda = 4$).

Les valeurs propres présentent la solution de l'équation caractéristique suivante :

$$|\mathbf{A} - \lambda\mathbf{I}_n| = 0 \tag{1.12}$$

Le déterminant d'une matrice $\mathbf{A}$ peut être donné par le produit de ses valeurs propres qu'on note $\{\lambda_i\}_{i=1}^n$.

$$|\mathbf{A}| = \prod_{i=1}^{n} \lambda_i \tag{1.13}$$

Si la matrice carrée $\mathbf{A}$ est symétrique alors :

1. Les valeurs propres sont positifs.
2. Les vecteurs propres associées aux différentes valeurs propres sont orthogonaux.
3. Il existe une base orthogonale où chaque élément présente un vecteur propre de $\mathbf{A}$.

La matrice carrée $\mathbf{A}$ est inversible si son déterminant est différent de zéro. La matrice inverse est notée $\mathbf{A}^{-1}$, on a alors :

$$\mathbf{AA}^{-1} = \mathbf{I}_n \tag{1.14}$$

La trace de $\mathbf{A}$ est la somme de ses éléments diagonaux :

$$Tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} \tag{1.15}$$

La trace d'une matrice carrée est égale à la somme de ces valeurs propres $Tr(\mathbf{A}) = \sum \lambda_i$.

Une matrice est dite symétrique si elle vérifie la condition :

$$\mathbf{A} = \mathbf{A}' \tag{1.16}$$

Dans ce qui suit on ne considère que le cas des matrices carrées et symétriques, sauf si on l'indique clairement.

Exercice 2.
Soit la matrice carrée A dont les valeurs propres sont : $\lambda_1 = 1, \lambda_2 = 2$ et $\lambda_3 = 5$.

1. Quelles sont les dimensions de la matrice A ?
2. Que peut on dire sur la positivité de la matrice A ?
3. Comment déterminer l'équation caractéristique de A ?
4. Calculer le déterminant de A.
5. La matrice A est elle inversible ? Pourquoi ?
6. Calculer la trace de la matrice A.

1.2 Positivité

1.2.1 Définition

— Une fonction $f(x)$ est dite positive ssi $f(x) > 0, \quad \forall x \in \Re$.
— Une matrice symétrique $\mathbf{A}$ est définie positive ssi $\mathbf{X}'\mathbf{A}\mathbf{X} > 0, \quad \forall \mathbf{X} \neq \mathbf{0}$.
— Une matrice symétrique $\mathbf{A}$ de dimension n est définie positive si toutes les valeurs propres sont positives $\lambda_i > 0, \quad \forall i = 1, ..., n$.
— Une matrice $\mathbf{A}$ de dimension n est définie positive si tous les mineurs principaux sont positifs.

1.2.2 Exemples

1. La fonction $f(x) = x^2 - 4x + 5$ est définie positive sur l'ensemble $\Re$.
2. La matrice : $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ est définie positive, car ses valeurs propres sont positives. Les valeurs propres sont les solutions de l'équation caractéristique $|\mathbf{A} - \lambda \mathbf{I}_2| = 0$.

Exercice 3. *Matrice positive*
. Déterminer parmi les matrices suivantes celles qui sont définies positives :

a.

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

b.

$$\mathbf{B} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Utiliser la définition ensuite les valeurs propres.

1.3 Convexité

1.3.1 Fonction convexe

Soit la fonction f, définie sur un intervalle $\mathbf{I}$ de $\Re^n$, telle que $\forall(x_1, x_2) \in \mathbf{I}^2$ et $\forall \alpha \in [0, 1]$, on dit que la fonction f est convexe (ou plus exactement strictement convexe) si elle vérifie l'inégalité suivante :

$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2) \tag{1.17}$$

Remarque Si la fonction $f(x) \in \Re$, elle est convexe si on a $f''(x) \geq 0$ sur le domaine de définition.

1.3.2 Ensemble convexe

Un ensemble $\Omega \subset \Re^n$ est dit convexe ssi :

$$\forall u, v \in \Omega \quad et \;\; \forall \alpha \in [0, 1], \quad on\ a \quad x = \alpha u + (1 - \alpha)v \in \Omega \tag{1.18}$$

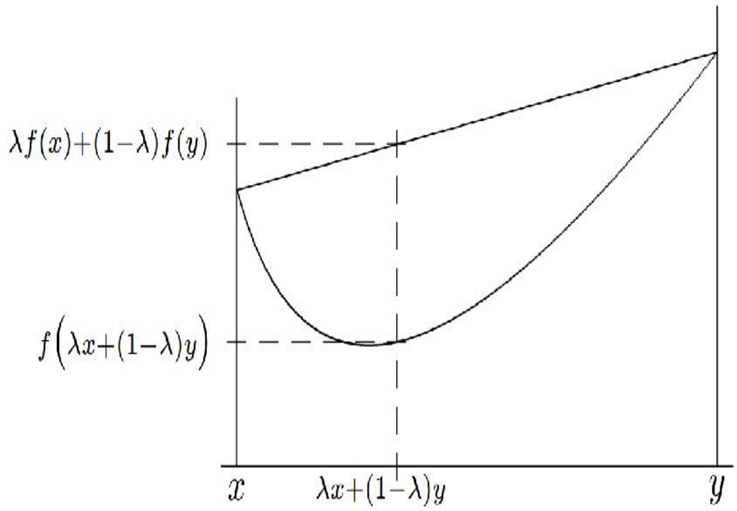

FIGURE 1.1 – Exemple d'une fonction convexe

Exercice 4.
Démontrer que l'ensemble définie par $\mathbf{\Omega} = \{\ \mathbf{X}\ \in\ \Re^n\ : \mathbf{a}'\mathbf{X}\ \geq\ \beta\}$, où $\mathbf{a} \in \Re^n$ et $\beta \in \Re$, est convexe, $\mathbf{\Omega}$ est appelé demi-espace *halfspace*.

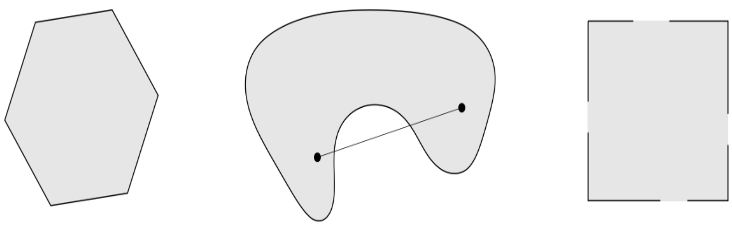

FIGURE 1.2 – Illustration d'ensembles convexes et non vonvexes

1.4 Le gradient

Soit $f(\mathbf{X})$ une fonction scalaire de n variables. Le gradient de la fonction f qu'on note ∇f est un vecteur ligne dont les élements sont les dérivées partielles de la fonction f suivant chaque variable.

$$\nabla f = \frac{\partial f}{\partial \mathbf{X}} = \left[\ \frac{\partial f}{\partial x_1}\ \ \frac{\partial f}{\partial x_2}\ \ \dots\ \ \frac{\partial f}{\partial x_n}\ \right] \tag{1.19}$$

Exemple Trouver le gradient de la fonction :

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

Solution

Les dérivées partielles suivant les variables x et y sont :

$$\frac{\partial f}{\partial x} = 2x + y,$$
$$\frac{\partial f}{\partial y} = x + 2y,$$

Le vecteur du gradient est donnée par :

$$\frac{\partial f}{\partial \mathbf{X}} = [2x + y \ , \ x + 2y]$$

1.5 Le Hessien

Soit $f(\mathbf{X})$ une fonction scalaire de n variables. Le Hessien de la fonction f, qu'on note $\mathbf{H}_f$ est une matrice de dimension $n \times n$, dont les élements sont les dérivées partielles d'ordre deux de la fonction f.

$$\mathbf{H}_f = \frac{\partial^2 f(\mathbf{X})}{(\partial \mathbf{X})^2} = \begin{bmatrix} \frac{\partial^2 f(\mathbf{X})}{(\partial x_1)^2} & \frac{\partial^2 f(\mathbf{X})}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{X})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{X})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{X})}{(\partial x_2)^2} & \dots & \frac{\partial^2 f(\mathbf{X})}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f(\mathbf{X})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{X})}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(\mathbf{X})}{(\partial x_n)^2} \end{bmatrix} \tag{1.20}$$

Exemple

1. Trouver le Hessien de la fonction :

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

2. Démontrer que cette matrice est définie positive.

Solution

1. Calcule du Hessien :

$$\frac{\partial f}{\partial x} = 2x + y, \qquad \longrightarrow \frac{\partial^2 f}{(\partial x)^2} = 2$$
$$\longrightarrow \frac{\partial^2 f}{\partial x \partial y} = 1$$
$$\frac{\partial f}{\partial y} = x + 2y, \qquad \longrightarrow \frac{\partial^2 f}{(\partial y)^2} = 2$$
$$\longrightarrow \frac{\partial^2 f}{\partial y \partial x} = 1$$

$$\mathbf{H}f = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

2. Pour démontrer que cette matrice est définie positive, on doit calculer les valeurs propres correspondantes.
   L'équation caractéristique dans ce cas est donnée par :
$$(2-\lambda)^2 - 1 = 0$$
$$(2-\lambda-1)(2-\lambda+1) = 0$$
$$(1-\lambda)(3-\lambda) = 0$$
   Les deux valeurs propres correspondantes sont $\lambda_1 = 1,\ \lambda_2 = 3$ et sont positives, ainsi la matrice est définie positive.

Remarque

Une fonction $f(\mathbf{X})$ est convexe ssi $\mathbf{H}_f(\mathbf{X})$ est semi-definie positif pour tout $\mathbf{X} \in \Re^n$.

Exercice 5. *Gradient et Hessien*

. Considérons la fonction scalaire suivante :

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

a. Calculer les dérivées partielles $f_x = \frac{\partial f}{\partial x}$ et $f_y = \frac{\partial f}{\partial y}$.

b. Trouver les extrémums en résolvant les équations $f_x = 0$ et $f_y = 0$.

c. Calculer le Hessein de la fonction f.

d. Déduire la nature des extrémums.

1.6 Exercices de synthèse

Exercice 6. *Régression linéaire*
. Soit un ensemble de n points dans le plan $\{\mathbf{X}_i = [x_i, y_i]'\}_{i=1}^{n}$, ces points sont souvent les résultats d'une experience. Les points ne sont pas sur la même ligne, Cependant on veut trouver la ligne approchant au mieux ce nuage de points. Pour cela on utlise la méthode des moindres carrées, et on choisis la relation suivante $y = ax + b$ avec $(a, b) \in \Re^2$. La fonction objective est la somme des carrées des erreurs (cette fonction est convexe) :

$$f(a,b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

a. Claculer le gradient de la fonction $f(a, b)$.

b. Ecrire le système d'équations qui donne les extremums.

c. Donner l'écriture matricielle correspondante à ce système.

d. Trouver les coefficients (a, b) qui minimisent cette fonction.

Exercice 7.
Soit la fonction réelle à deux variables :

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

1. Calculer le Jacobien,
2. Calculer le Hessien,
3. Déterminer l'équation caractéristique du Hessien,
4. Donner les valeurs propres de cette matrice,
5. Cette matrice est elle définie positive, pourquoi ?
6. Calculer le déterminant en utilisant les valeurs propres.
7. Trouver analytiquement les optimums de cette fonction. Quelle sont leurs natures ?

1.7 Solutions des exercices

Solution Ex. 1.

—

$$
\begin{aligned}
\mathbf{AB}' &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} [-1\ 0\ 2] \\
&= \begin{bmatrix} -1 & 0 & 2 \\ -2 & 0 & 4 \\ -3 & 0 & 6 \end{bmatrix}
\end{aligned}
$$

—

$$
\mathbf{A}'\mathbf{B} = [1\ 2\ 3] \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = 5
$$

— Le calcul de l'expression $\mathbf{AB}'\mathbf{B}$ peut se faire de deux façon différentes, suivant l'ordre dans lequel nous calculons le produit :

- $$
\begin{aligned}
(\mathbf{AB}')\mathbf{B} &= \begin{bmatrix} -1 & 0 & 2 \\ -2 & 0 & 4 \\ -3 & 0 & 6 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \\
&= \begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix}
\end{aligned}
$$

- $$
\mathbf{A}(\mathbf{B}'\mathbf{B}) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times 5 = \begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix}
$$

Remarque
Lors de calcul du produit de plusieurs matrices il est souvent préférable de commencer par le calcule des produits qui minimisent le nombre d'opérations de multiplications.

Solution Ex. 2.
La matrice carrée A dispose des 3 valeurs propres $\lambda_1 = 1, \lambda_2 = 2$ et $\lambda_3 = 5$. alors :

1. La matrice A est de dimension 3×3 (car elle est carrée et dispose de 3 valeurs propres).
2. La matrice A et positive car toutes les valeurs propres sont positives.
3. L'équation caractéristique de A est donnée par la résolution de l'équation :
$$det(A - \lambda I) = 0 \tag{1.21}$$
Dans notre cas, on a en plus $(\lambda - \lambda_1)(\lambda - \lambda_3)(\lambda - \lambda_3) = 0$
4. Le déterminant de A est égale à :
$$\det(A) = \prod_{i=1}^{3} \lambda_i = \lambda_1 \times \lambda_2 \times \lambda_3 = 10 \tag{1.22}$$
5. La matrice A est inversible, car le déterminant est non nul.
6. La trace de la matrice A est égale à la somme des éléments de sa diagonale. elle est égale aussi à la somme des valeurs propres.
$$trace(A) = \sum_{i=1}^{3} \lambda_i = \lambda_1 + \lambda_2 + \lambda_3 = 8 \tag{1.23}$$

Solution Ex. 3.
Deux méthodes permettent de déterminer si une matrice réelle $\mathbf{C}$ de dimension $(n \times n)$ est définie positive ou non :

1. **Utilisation de la définition**
La matrice $\mathbf{C}$ est définie positive si on a $\mathbf{X'CX} > 0, \forall \mathbf{X} \in \Re^{*n}$.
   (a) Pour la matrice $\mathbf{A}$
$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$
$$\begin{aligned} \mathbf{X'AX} &= [x\ y] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= 2x^2 + 2xy + 2y^2 \\ &= x^2 + y^2 + (x+y)^2 > 0, \quad \forall (x,y) \in \Re^{*2} \end{aligned}$$

Ainsi, la matrice **A** est définie positive.

(b) Pour la matrice **B**

$$\mathbf{B} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\begin{aligned} \mathbf{X}'\mathbf{B}\mathbf{X} &= [x\ y] \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= 2x^2 - 2xy + 2y^2 \\ &= x^2 + y^2 + (x-y)^2 > 0 \quad \forall (x,y) \in \Re^{*2} \end{aligned}$$

Ainsi, la matrice **B** est définie positive.

2. **Utilisation des valeurs propres**

La matrice est positive si toutes ces valeurs propres sont positives.

(a) Pour la matrice **A**

$$\begin{aligned} \det(\mathbf{A} - \lambda \mathbf{I}_2) &= \det\left(\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \\ &= \det\left(\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \right) \\ &= (2-\lambda)^2 - 1 \end{aligned}$$

Le déterminant s'annule pour $\lambda_1 = 1$ et $\lambda_2 = 3$

(b) Pour la matrice **B**

$$\begin{aligned} \det(\mathbf{B} - \lambda \mathbf{I}_2) &= \det\left(\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \\ &= \det\left(\begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix} \right) \\ &= (2-\lambda)^2 - 1 \end{aligned}$$

Le déterminant s'annule pour $\lambda_1 = 1$ et $\lambda_2 = 3$

Les valeurs propres sont positives alors les deux matrices **A** et **B**, sont définies positives.

Solution Ex. 4.
Soit $\mathbf{Z} = \lambda\mathbf{X} + (1-\lambda)\mathbf{Y}$ pour $\lambda \in [0,1]$. Alors :

$$\begin{aligned}
\mathbf{a}'\mathbf{Z} &= \mathbf{a}'\left(\lambda\mathbf{X} + (1-\lambda)\mathbf{Y}\right) \\
&= \lambda\mathbf{a}'\mathbf{X} + (1-\lambda)\mathbf{a}'\mathbf{Y} \\
&\geq \lambda\beta + (1-\lambda)\beta \ = \ \beta \ \ (\ on\ a\ \lambda \geq 0\ et\ (1-\lambda) \geq 0\ ) \\
\mathbf{a}'\mathbf{Z} &\geq \beta \Rightarrow \mathbf{Z} \ \in \ \mathbf{\Omega}
\end{aligned}$$

par conséquent, $\mathbf{\Omega}$ est un ensemble convexe.

Solution Ex. 5. Soit la fonction

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

1. Calcul des gradients

$$f_x = \frac{\partial f}{\partial x} = 1 - 2xy + y^2$$
$$f_y = \frac{\partial f}{\partial y} = -1 - x^2 + 2xy$$

2. En sommant les deux équations $f_x = 0$ et $f_y = 0$, on obtient :

$$y^2 - x^2 = 0 \Rightarrow \quad y = \pm x$$

— Soit $y = x$ alors on a :

$$1 - 2x^2 + x^2 = 0 \rightarrow 1 - x^2 = 0 \rightarrow x = \pm 1$$

ainsi on a deux points : $(1,1)$ et $(-1,-1)$

— Soit $y = -x$ alors on a :

$$1 + 2x^2 + x^2 = 1 + 3x^2 = 0$$

cette équation n'a pas de solution dans $\Re$.

3. Les dérivées partielles d'ordre deux et le Hessien sont :

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

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

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

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

$$\mathbf{H}_f = \begin{bmatrix} -2y & -2x + 2y \\ \\ -2x + 2y & 2x \end{bmatrix}$$

4. La nature des extrémums, pour les deux points sont :
   — pour le point $(x, y) = (1, 1)$

$$\mathbf{H}_f(1,1) = \begin{bmatrix} -2 & 0 \\ \\ 0 & 2 \end{bmatrix} \longrightarrow |\mathbf{H}f(1,1)| = -4 < 0$$

   ainsi ce point est un point de selle.
   — pour le point $(x, y) = (-1, -1)$

$$\mathbf{H}_f(-1,-1) = \begin{bmatrix} 2 & 0 \\ \\ 0 & -2 \end{bmatrix} \longrightarrow |\mathbf{H}f(-1,-1)| = -4 < 0$$

   ainsi ce point est aussi un point de selle.

Solution Ex. 6.
Nous avons la fonction fà deux variables a et b :

$$f(a,b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

1. Les gradients suivant les variables a et b sont comme suit :

$$\begin{aligned}
\frac{\partial f}{\partial a} &= \sum_{i=1}^{n} 2(-y_i + ax_i + b)x_i \\
&= 2a\sum_{i=1}^{n} x_i^2 + 2b\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} y_i x_i \\
\frac{\partial f}{\partial b} &= \sum_{i=1}^{n} 2(-y_i + ax_i + b) \\
&= 2a\sum_{i=1}^{n} x_i + 2nb - 2\sum_{i=1}^{n} y_i
\end{aligned}$$

2. Sachant que cette fonction est convexe, l'extremum et donné par le système d'équation suivant :

$$\begin{aligned}
\sum_{i=1}^{n} x_i^2\ a + \sum_{i=1}^{n} x_i\ b &= \sum_{i=1}^{n} y_i x_i \\
\sum_{i=1}^{n} x_i\ a + n\ b \qquad &= \sum_{i=1}^{n} y_i
\end{aligned}$$

3. Soit :

$$\begin{cases}
S_{11} = \sum_{i=1}^{n} x_i^2 \\
S_{12} = S_{21} = \sum_{i=1}^{n} x_i \\
S_{22} = n
\end{cases}$$

$$\begin{cases} Z_1 = \sum_{i=1}^{n} y_i x_i \\ Z_2 = \sum_{i=1}^{n} y_i \end{cases}$$

L'écriture matricielle correspondante est :

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

4. Les coefficients a et b sont données en résolvant le système d'équation précédent, on obtient alors :

$$a = \frac{S_{12}Z_2 - S_{22}Z_1}{S_{12}^2 - S_{11}S_{22}}$$

$$b = \frac{S_{12}Z_1 - S_{11}Z_2}{S_{12}^2 - S_{11}S_{22}}$$

Programme Matlab

La figure 1.3 montre le résultat après exécution du programme suivant :

```
%% Construction d'un nuage de points
a=2;
b=3;
N=100;
x=3+10*rand(1,N);
y=a*x+b+3*randn(1,N);
plot(x,y,'.'), hold on
%% Calcul des parametres
S11=sum(x.^2);
S12=sum(x);
S21=S12;
S22=N;
Z1=sum(y.*x);
Z2=sum(y);
```

```

a=(S12*Z2-S22*Z1)/(S12^2-S11*S22);
b=(S21*Z1-S11*Z2)/(S12^2-S11*S22);
%% Dessin des resultats
x_min=min(x);
x_max=max(x);
pas=(x_max-x_min)/20;
x=(x_min:pas:x_max);
y_est=a*(x_min:pas:x_max)+b;

plot(x,y_est,'r','LineWidth',3)
xlabel('x')
ylabel('y')
legend('Nuage de points','Droite estimee')
```

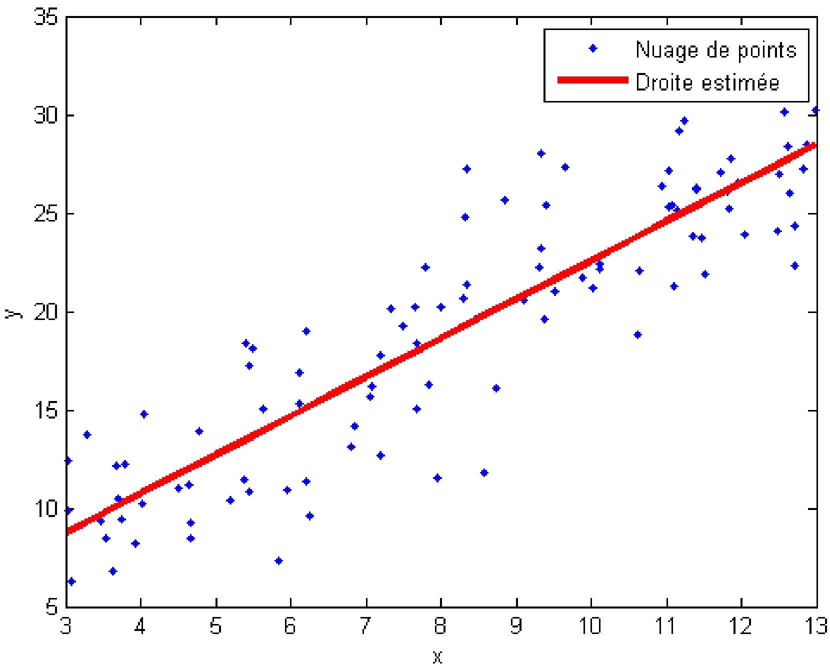

FIGURE 1.3 – Résultat de la régression linéaire

Solution Ex. 7.

Soit la fonction réelle à deux variables :

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

1. Calcule du Jacobien,

$$\begin{aligned}\mathbf{J}_f = \nabla f' &= \begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{bmatrix} \\ &= \begin{bmatrix} 1 - 2xy + y^2 \\ \\ -x^2 + 2xy \end{bmatrix}\end{aligned}$$

2. Le Hessien

$$\mathbf{H}_f = \begin{bmatrix} -2y & -2x + 2y \\ -2x + 2y & 2x \end{bmatrix}$$

3. L'équation caractéristique du Hessien est donnée par :

$$\begin{aligned}\det(H - \lambda I) = 0 \\ \rightarrow\ & (-2y - \lambda)(2x - \lambda) - (-2x + 2y)^2 = 0 \\ & \lambda^2 + 2(y - x)\lambda - (-2x + 2y)^2 - 4xy = 0\end{aligned}$$

4. Les valeurs propres sont les solutions de l'équation caractéristique de cette matrice, c'est une équation du deuxième ordre alors :

$$\begin{aligned}\Delta &= 4(y - x)^2 + 16(y - x)^2 + 16xy \\ &= 20(y - x)^2 + 16xy\end{aligned}$$

   Les solution sont alors :

$$\begin{aligned}\lambda_1 &= \frac{-2(y - x) + \sqrt{\Delta}}{2} \\ &= x - y + \sqrt{5(y - x)^2 + 4xy} \\ \lambda_2 &= \frac{-2(y - x) - \sqrt{\Delta}}{2} \\ &= x - y - \sqrt{5(y - x)^2 + 4xy}\end{aligned}$$

5. Pour que cette matrice soit définie positive, il faut que $\lambda_1 > 0$ et $\lambda_2 > 0$. Ainsi la positivité de dépend des valeurs attribuées aux variables x et y.
6. Le déterminant est donné par le produit des valeurs propres :

$$\begin{aligned} \det(\mathbf{H}_f) &= \lambda_1 \times \lambda_2 \\ &= \left(x - y + \sqrt{5(y-x)^2 + 4xy}\right)\left(x - y - \sqrt{5(y-x)^2 + 4xy}\right) \\ &= -4(x^2 + y^2 - xy) \end{aligned}$$

7. L'optimum de cette fonction est obtenu analytiquement comme suit :

$$\mathbf{J}_f = \nabla f' = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 - 2xy + y^2 \\ \\ -x^2 + 2xy_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \quad \begin{matrix} (1) \\ \\ (2) \end{matrix}$$

De la deuxième équation on tire :

$$x^2 - 2xy = 0 \quad \to \quad x(x - 2y) = 0$$

on élimine la solution $x = 0$ car elle ne vérifie pas la première équation. Alors on prend $x = 2y$ et on le remplace dans la première équation on trouve :

$$1 - 4y^2 + y^2 = 0 \quad \to \quad y = \pm\sqrt{\frac{1}{3}}$$

$$\begin{aligned} y = \sqrt{\tfrac{1}{3}} \quad &\to \quad x = \tfrac{2}{\sqrt{3}} \quad \to X_1 = \left[\tfrac{2}{\sqrt{3}}\ , \sqrt{\tfrac{1}{3}}\ \right]' \\ y = -\sqrt{\tfrac{1}{3}} \quad &\to \quad x = -\tfrac{2}{\sqrt{3}} \quad \to X_2 = \left[\tfrac{-2}{\sqrt{3}}\ , -\sqrt{\tfrac{1}{3}}\ \right]' \end{aligned}$$

La nature des points dépend de la valeur de la Hessienne à ces points :

$$\det(H) = -4(x^2 + y^2 - xy)$$

- pour X_1 on a $\det(H(X = X_1)) = -\frac{4}{3}$
- pour X_2 on a $\det(H(X = X_2)) = -\frac{4}{3}$

pour les deux points on a $\det(H) < 0$ alors on a deux points de celle.

Chapitre 2

Optimisation sans contraintes : méthodes locales

2.1 Définitions

2.1.1 Définition 1

Soit $f(\mathbf{X})$ une fonction de $\Re^n \to \Re^n$; Le problème d'optimisation sans contraintes s'écrit de la façon suivante :

$$\text{Minimiser la fonction} f(\mathbf{X}) \text{ avec } \mathbf{X} \in \Re^n$$

La fonction $f(\mathbf{X})$ est souvent appelée :
* fonction coût,
* fonction objectif,
* critère d'optimisation.

2.1.2 Définition 2

On dit que $\mathbf{X}^*$ est un point qui minimise la fonction $f(\mathbf{X})$ sur l'ensemble $\Re^n$ ssi :

$$\forall \mathbf{X} \in \Re^n, \quad f(\mathbf{X}^*) \leq f(\mathbf{X})$$

2.2 Théorèmes d'existence de l'optimum

Avant d'entamer la recherche de optimum d'une fonction, nous devons s'assurer qu'il existe. A cet effet, on utilise une technique similaire à celle de la recherche du maximum ou du minimum d'une fonction scalaire. Ce qui nous amène à utiliser les dérivées partielles d'une fonction à plusieurs variables.

Théorème 1.
Soit la fonction $f(\mathbf{X}) : \Re^n \to \Re$ *différentiable sur* $\Re^n$*, on dit que* $f(\mathbf{X})$ *à un optimum au point* $\mathbf{X}^*$ *si :*

$$\nabla f(\mathbf{X}^*) = 0$$

Remarques

- $\mathbf{X}^*$ est appelé point stationnaire de la fonction $f(\mathbf{X})$.
- La relation $\nabla f(\mathbf{X}^*) = 0$ est aussi appelée équation d'*Euler*.
- Ce théorème n'a aucun sens si la fonction n'est pas différentiable (Exemple fonction en V, comme $f(x) = |x|$).

Théorème 2.
Soit la fonction $f(\mathbf{X}) : \Re^n \to \Re$ *convexe et continument différentiable à l'ordre deux sur* $\Re^n$*, on dit que* $f(\mathbf{X})$ *à un minimum global au point* $\mathbf{X}^*$ *ssi :*

$$\nabla f(\mathbf{X}^*) = 0$$

L'égalité précédente engendre un ensemble de n équations qui peuvent être, dans certains cas, résolues analytiquement.

Cependant, on a souvent recours à résoudre cet ensemble d'équation numériquement et par des algorithmes itératifs. On construit ainsi, une suite de solutions qui converge vers la solution optimale

$$\mathbf{X}_0 \to \mathbf{X}_1 \to ... \to \mathbf{X}_\infty = \mathbf{X}^*$$

Le sens physique de l'infini signifie qu'à un instant donné ou après plusieurs itérations, nous devons accepter une certaine tolérance. Cela va engendrer un critère d'arrêt lorsqu'on atteint la tolérance acceptée ou bien arrêter les itérations dans le cas des boucles infinies.

2.3 Le développement en série de Taylor

Dans le cas unidimensionnel, la dérivée d'une fonction en un point, est comme suit :

$$\begin{aligned} f'(a) &= \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \\ &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \end{aligned}$$

on peut faire l'approximation suivante :

$$f(x) - f(a) \approx f'(a)(x - a)$$

soit :

$$f(x) \approx f(a) + f'(a)(x - a)$$

ou encore

$$f(a + \Delta x) \approx f(a) + f'(a)\Delta x$$

pour plus de précision, on ajoute les termes :

$$\begin{aligned} f(x) &\approx f(a) + f'(a)(x - a) + \frac{1}{2}f''(a)(x - a)^2 + ... + \frac{1}{m!}f^{(m)}(a)(x - a)^m \\ &= \sum_{i=0}^{m} \frac{1}{i!} f^{(i)}(a)(x - a)^i \qquad (2.1) \end{aligned}$$

Cette dernière expression est appelée développement en série de *Taylor* d'ordre m au voisinage du point a.

Dans le cas multidimensionnel, le développement en série de *Taylor* du premier ordre de la fonction $f(\mathbf{X})$ au voisinage du vecteur $\mathbf{A}$, est donné par :

$$\begin{aligned} f(\mathbf{A} + \Delta\mathbf{X}) &\approx f(\mathbf{A}) + \nabla f(\mathbf{A})\Delta\mathbf{X} \qquad (2.2) \\ &= f(\mathbf{A}) + \sum_{i=1}^{n} \left.\frac{\partial f(X)}{\partial x_i}\right|_{x_i = a_i} (x_i - a_i) \qquad (2.3) \end{aligned}$$

Le développement en série de *Taylor* du second ordre est :

$$\begin{aligned} f(\mathbf{A}+\Delta\mathbf{X}) &\approx f(\mathbf{A}) + \nabla f(\mathbf{A})\Delta\mathbf{X} + \frac{1}{2!}\Delta\mathbf{X}'\mathbf{H}f(\mathbf{A})\Delta\mathbf{X} \\ &= f(\mathbf{A}) + \sum_{i=1}^{n} \left.\frac{\partial f(X)}{\partial x_i}\right|_{x_i=a_i} (x_i - a_i) \\ &\quad + \frac{1}{2!}\sum_{i,j=1}^{n} \left.\frac{\partial^2 f(X)}{\partial x_i \partial x_j}\right|_{\substack{x_i=a_i\\ x_j=a_j}} (x_i - a_i)(x_j - a_j) \end{aligned} \tag{2.4}$$

2.4 Résolution analytique (cas unidimensionnel)

Certains problèmes d'optimisation à plusieurs variables, peuvent être amenés à un problème équivalent à une seul variables uniquement. Dans cette partie nous allons examiner certains problèmes de ce type.

Ci-dessous les on présente la démarche pour résoudre ce type d'exercice.

1. Lire attentivement le problème
2. Faire des figures d'illustration
3. Identifier les variables
4. Définir la fonction à optimiser $f(x, y, z, ...)$
5. Reformuler la fonction suivant une seule variable $f(x)$
6. Résoudre l'équation $\frac{df(x)}{dx} = 0$
7. Réécrire la solution suivant les variables initiales.

Pour illustrer cela, nous présentons deux exercices, le premier concerne la détermination des dimension d'une boite et l'objectif étant la maximisation du volume. Le second concerne un problème typique d'une aire maximale.

Exercice 8. *Boite à volume maximal*

.

En utilisant une feuil de dimensions $L \times H$, on veut déterminer les dimensions correspondantes d'une boite (sans couvercle) dont le volume est maximal.

a. Dessiner sur une feuil le croquis de cette boite.

b. Déterminer la fonction à minimiser correspondante à ce problème suivant la hauteur h.

c. Déterminer les dimension de la boite à volume maximal.

d. Donner le résultat pour une page de dimension $A4$ (21×29.7).

e. Trouver le même résultat en utilisant Matlab (dessiner la fonction de volume ainsi que sa dérivée en fonction de la variable h.

Solution Voir le lien suivant 8.

Exercice 9. *Aire maximale*

.
La parabole d'équation :

$$y = -\frac{2}{9}x^2 + 8,$$

coupe l'axe des abscisses en A et B.

Le point $P(x, y)$ se déplace sur la parabole entre A et B. Soit A' la projection du point P sur la droite (AB).

a. Représenter graphiquement cet énoncé.

b. Déterminer les coordonnées du point P pour que l'aire du triangle rectangle $A'PB$ soit maximale.

Solution Voir le lien suivant 9.

2.5 Les méthodes de descente

Les méthodes de descente sont des techniques itératives permettant de trouver le minimum d'une fonction donnée. Le principe est de proposer un point initiale $\mathbf{X}_0$, ensuite on cherche à trouver un autre point $\mathbf{X}_1$ qui vérifie la condition $f(\mathbf{X}_1) < f(\mathbf{X}_0)$. Par la suite, $\mathbf{X}_1$ devient le point initiale pour la nouvelle itération et ainsi de suite jusqu'à l'obtention du point minimale. Le point $\mathbf{X}_1$ peut être mis sous la forme suivante :

$$\mathbf{X}_1 = \mathbf{X}_0 + \rho_0 \Delta \mathbf{X}_0$$

où $\rho_0 \in \Re^{*+}$ est appelé le *pas de descente* et $\Delta \mathbf{X}_0$ un vecteur de $\Re^n$ appelé *direction de descente*. Ainsi, à la première itération on doit trouver ρ_0 et $\Delta \mathbf{X}_0$ qui satisfont la condition $f(\mathbf{X}_1) < f(\mathbf{X}_0)$. Et en générale, à la $k^{ième}$ itération, on a :

$$\mathbf{X}_{k+1} = \mathbf{X}_k + \rho_k \Delta \mathbf{X}_k \tag{2.5}$$

et on doit trouver $\rho_k \in \Re^{*+}$ et $\Delta \mathbf{X}_k$ qui satisfont la condition $f(\mathbf{X}_{k+1}) < f(\mathbf{X}_k)$.

Remarque
Le pas de descente ρ est déterminé soit analytiquement en minimisant la fonction :

$$\underset{s \geq 0}{argmin} f(X + s\Delta X).$$

Soit itérativement en minimisant la fonction $f(\mathbf{X} + \rho \Delta \mathbf{X})$, suivant la variable ρ. L'algorithme correspondant dans ce cas est comme suit :

* Connaissant la direction $\Delta \mathbf{X}$ et soit deux réelles $0 < \alpha < 0.5$ et $0 < \beta < 1$.
* $\rho \leftarrow 1$.
* **Tant que** $f(\mathbf{X} + \rho \Delta \mathbf{X}) > f(\mathbf{X}) + \alpha \rho \nabla f(\mathbf{X}) \Delta \mathbf{X}$,
  faire $\rho :\, = \beta \rho$.

Remarque
En pratique, souvent, on prend :

$$\alpha \in [0.01 \ , \ 0.3]$$
$$\beta \in [0.1 \ , \ 0.8]$$

.

2.6 Méthode du gradient

2.6.1 Principe

La méthode du gradient est une technique de descente (itérative) permettant de trouver le minimum d'une fonction donnée. La façon la plus évidente de trouver une direction est l'utilisation de la dérivée. Ainsi, pour trouver la direction de descente, on utilise le sens inverse de la dérivée. A cet effet, on prend $\Delta\mathbf{X} = -\nabla f(\mathbf{X})'$. La relation itérative correspondante à la méthode du gradient est la suivante :

$$f(\mathbf{X}_{k+1}) \approx f(\mathbf{X}_k) + \rho_k \nabla f(\mathbf{X}_k)\Delta\mathbf{X}_k \tag{2.6}$$

Si ρ_k est constant, cette méthode est qualifiée de méthode du gradient à pas fixe. Lorsque ρ_k varie, on parle de la méthode du gradient à pas variable.

2.6.2 Algorithme

- connaissant la valeur initiale $\mathbf{X}_0$,
- Répéter
  1. $\Delta\mathbf{X} \leftarrow -\nabla f(\mathbf{X})'$.
  2. Choisir le pas ρ (s'il est fixe on ignore cette étape).
  3. Mettre à jours $\mathbf{X} \leftarrow \mathbf{X} + \rho\Delta\mathbf{X}$.
- Arrêter lorsque la condition de convergence est satisfaite.

La condition de convergence est souvent pris comme : $\|\nabla f(\mathbf{X})\|_2 \leq \eta$ avec η un nombre petit et positif.

2.6.3 Exemple

Soit une fonction cout $f(\mathbf{X})$ quadratique en $\Re^2$ sous la forme :

$$f(\mathbf{X}) = x^2 + 3y^2,$$

avec $\mathbf{X} = \begin{bmatrix} x \\ y \end{bmatrix}$. Il est évident que l'origine est le point minimisant cette fonction. Cette fonction est illustrée sur les figures 2.3 et 2.2.

— **Résolution avec un pas fixe**
Dans ce cas on recherche un pas ρ optimale c'est celui qui minimise une certaine fonction, on a :

$$\begin{aligned} \nabla f(\mathbf{X}) &= [\ 2x\ ,\ 6y] \\ \xi(\rho) &= \mathbf{X} - \rho \nabla f(\mathbf{X})' \\ &= \begin{bmatrix} (1-2\rho)x \\ (1-6\rho)y \end{bmatrix} \end{aligned} \tag{2.7}$$

Pour l'obtention du paramètre ρ, considérons la fonction $g(\rho)$ qu'on doit minimiser suivant ρ :

$$\begin{aligned} g(\rho) &= f(\xi(\rho)) \\ &= (1-2\rho)^2 x^2 + 3\,(1-6\rho)^2 y^2, \end{aligned} \tag{2.8}$$

La dérivé de cette fonction suivant la variable ρ donne :

$$g'(\rho) = -4(1-2\rho)x^2 - 36(1-6\rho)y^2$$

la paramètre ρ qui minimise la fonction $g(\rho)$ est celui qui satisfait $g'(\rho) = 0$, soit :

$$\rho = \frac{x^2 + 9y^2}{2x^2 + 54y^2} \tag{2.9}$$

L'équation itérative de résolution de ce problème est donc :

— Initialisation $\mathbf{X} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$
— Répéter
  1. $\Delta \mathbf{X} \leftarrow - \begin{bmatrix} 2x \\ 6y \end{bmatrix}$.
  2. $\rho = \frac{x^2+9y^2}{2x^2+54y^2}$
  3. $\mathbf{X} \leftarrow \mathbf{X} + \rho \Delta \mathbf{X}$.
— Arrêter lorsque $\|\Delta \mathbf{X}\|_2^2 \leq 10^{-3}$.

— **Résolution avec un pas itératif**
Lorsqu'on choisis un pas itératif alors, la deuxième étape dans l'algorithme est remplacée par la boucle suivante :
soit $\alpha = 0.25$, $\beta = 0.5$ $\rho = 1$,
alors :
pour une direction donnée $\Delta\mathbf{X} = -\begin{bmatrix} 2\mathbf{X}(1) \\ 6\mathbf{X}(2) \end{bmatrix}$,

— faire :

$$\mathbf{Y} = \mathbf{X} + \rho\Delta\mathbf{X},$$
$$f(\mathbf{Y}) = \mathbf{Y}(1)^2 + 3\mathbf{Y}(2)^2,$$
$$\nabla f(\mathbf{X}) = [2\mathbf{X}(1)\ ,\ 6\mathbf{X}(2)],$$

— si $f(\mathbf{Y}) < f(\mathbf{X}) + \alpha\rho\nabla f(\mathbf{X})\Delta\mathbf{X}$
fin
— sinon
$\rho = \beta\rho$
— fin

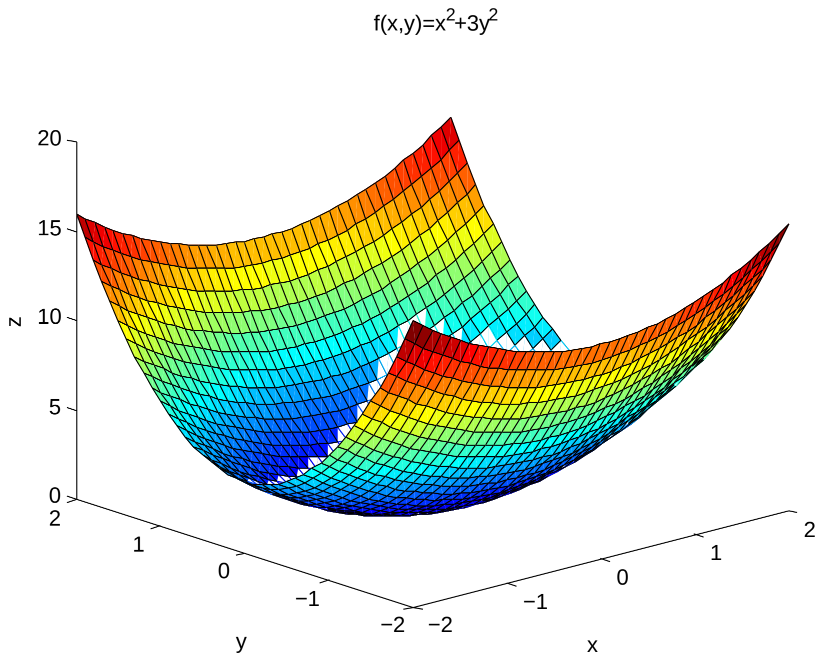

FIGURE 2.1 – Allure de la fonction $f(x, y) = x^2 + 3y^2$

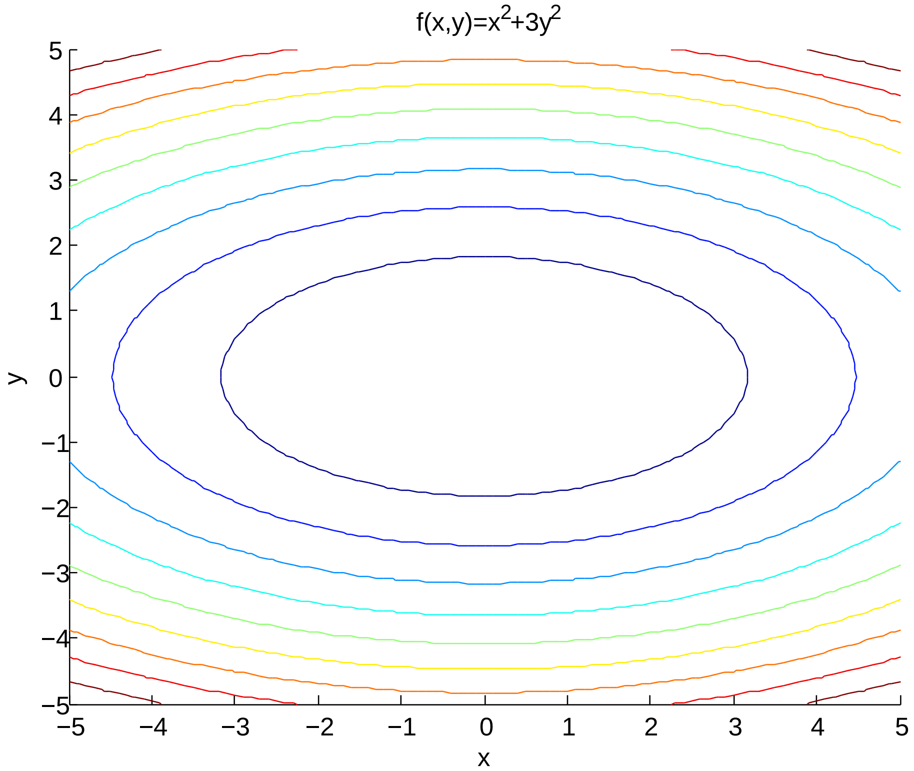

FIGURE 2.2 – Vue de haut de la fonction $f(x, y) = x^2 + 3y^2$