

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU
FACULTE DES SCIENCES



MÉMOIRE DE FIN D'ÉTUDES

En vue de l'obtention du Diplôme de Master II en
Spécialité : Recherche opérationnelle

Thème **Optimisation non linéaire et applications**

Soutenu le : 03/11/2021

Présenté par :

DJERBA Rabah

TERMACHE Rahmoune

Devant le jury :

Président : CHEBBAH Mohamed
Promoteur : OUANES Mohamed
Examineur : ZIDELMAL Nacera

Année universitaire : 2020 /2021

Dédicaces

C'est avec un respect que je dédie ce travail :

A mon défunt père et ma chère maman, qui m'ont toujours poussé
et motivé durant mon parcours d'étude.

Et sans oublier mes chers amis sans exception, et les étudiants RO
de la promotion 2020/2021.

Rabah

Dédicaces

**Je dédie ce modeste travail à tous ceux
qui me sont chers.**

Rahmoune

Remerciements

En premier lieu, nous remercions Dieu de nous avoir donné du courage, de la volonté et d'avoir guidé nos pas vers les portes du savoir.

En deuxième lieu, nous tenons à exprimer notre profonde gratitude pour notre promoteur Monsieur OUANES Mohand, de nous avoir encadré, pour la confiance qu'il nous a toujours témoignée. Son expérience dans la recherche qu'il a partagé avec nous, ses inestimables conseils et encouragements, sa disponibilité et sa ponctualité nous ont été très précieux et nous ont permis de mener à bien notre travail.

Ensuite, nous exprimons nos sincères remerciements à Monsieur CHEBBAH Mohamed et Madame ZIDELMAL Nacera, de nous avoir fait honneur d'être les membres de notre jury, afin d'évaluer notre travail.

Nous ne saurons oublier le grand mérite des enseignants qui ont contribué à notre cursus particulièrement ceux du département « Mathématiques » et qu'ils trouvent ici le témoignage de notre profonde reconnaissance.

Finalement, nous remercions nos amis : yahia, amar, nadir et toute personne ayant contribué de près ou de loin à l'élaboration de ce mémoire.

Table des matières

Optimisation non linéaire et applications	1
Introduction générale	4
1 Optimisation non linéaire	6
1.1 Introduction à \mathbb{R}^n	6
1.1.1 L'espace euclidien de \mathbb{R}^n :	6
1.1.2 Structure d'espace vectoriel :	6
1.1.3 Inégalité de Cauchy Schwarz :	8
1.2 Notions topologiques :	8
1.3 Matrices	9
1.3.1 Signe d'une matrice	10
1.4 Fonctions de plusieurs variables	12
1.4.1 Dérivées partielles	12
1.4.2 Dérivées d'ordre supérieurs à 1	13
1.4.3 Formule de Taylor au point X^* à l'ordre 2	14
1.5 Convexité	15
1.5.1 Ensembles convexes	15
1.5.2 Fonctions convexes	16
1.5.3 Fonctions coercives	17
1.6 Optimisation sans contraintes [2]	18
1.6.1 Résultat d'existence et d'unicité	18
1.6.2 Conditions nécessaires et suffisantes d'optimalité de 1 ^{er} ordre	19
1.6.3 Conditions nécessaires et suffisantes d'optimalité de 2 ^{ème} ordre	20
1.6.4 Algorithme	23
1.7 Optimisation avec contraintes [7]	26
1.7.1 Optimisation avec contraintes d'égalités	26
1.7.2 Optimisation avec contraintes d'inégalités [1]	31
1.7.3 Optimisation avec contraintes mixtes	34

2	Application des moindres carrés	37
2.1	Régression linéaire	37
2.1.1	Critère des moindres carrés :	38
2.1.2	Forme standard :	39
2.1.3	Recherche d'une solution :	39
2.2	Ajustement [9]	40
2.2.1	Position de problème	41
2.2.2	Ajustement par la méthode des moindres carrés	42
2.3	Corrélation[9]	47
2.3.1	position de problème	47
2.3.2	Droites de régression	48
2.3.3	Coefficient de corrélation	49
3	Simulation numérique avec MATLAB	52
	Conclusion générale	62
	Bibliographie	63

Table des figures

1.1	Ensemble convexe et non convexe	15
1.2	Méthode de Gradient	24
2.1	Exemple de nuage de points	40
2.2	Ajustement linéaire	43
2.3	projection de b sur $Im(a)$	45
2.4	Le nuage de points	45
2.5	Le poids d'individus en fonction de taille	47
2.6	S_1, S_2 : Nuage de points	47
2.7	Les droites de régressions (D) et (D')	48
2.8	Ajustement affine	50
2.9	L'ajustement affine et l'ajustement exponentielle	51
3.1	Programme matlab 01	53
3.2	Résultat 01	54
3.3	L'ajustement de programme 01	55
3.4	Programme matlab 02	56
3.5	Résultat 02	57
3.6	L'ajustement linéaire	58
3.7	Programme matlab 03	59
3.8	Résultat 03	60
3.9	L'ajustement linéaire et l'ajustement exponentielle	61

Introduction générale

L'optimisation, vue comme branche des mathématiques est essentiellement un outil d'aide à la décision au sein de l'entreprise, mais aussi pour des individus. Le terme optimal est souvent trompeur. Ce n'est pas un jugement de valeur absolu, c'est plutôt une information sur l'approche méthodologique utilisée.

L'optimisation non linéaire s'occupe principalement des problèmes d'optimisation dont la fonction objectif et/ou au moins l'une des contraintes est non linéaire, mais sont aussi différentiables autant de fois que nécessaire pour l'établissement des outils théoriques, comme les conditions d'optimalité qui sont un ensemble d'équations, d'inéquations et d'expressions diverses vérifiées par une solution d'un problème d'optimisation (conditions nécessaires d'optimalité) ou qui permettent d'affirmer qu'un point qui les vérifie est solution du problème d'optimisation considéré (conditions suffisantes d'optimalité). Ces expressions analytiques de l'optimalité sont utiles entre autres pour : calculer les solutions d'un problème d'optimisation, vérifier l'optimalité d'un point donné, ou concevoir des algorithmes de résolution.

Parmi les méthodes d'optimisations, on trouve la méthode des moindres carrés qui permet de fournir un outil d'interprétation des données. Plus précisément, lorsqu'on dispose de données dépendantes de deux paramètres X et Y , on peut les représenter dans le plan muni d'un repère, en marquant X en abscisse et Y en ordonnée ; si le "nuage de points" que l'on obtient a l'allure d'une droite, on veut savoir quelle est l'équation de cette droite, c'est-à-dire quelle loi relie les deux paramètres de la mesure donc on cherche la droite (représentée par son équation $y = aX + b$) qui minimise la somme

des carrés des distances verticales des points à la droite.

L'application de la méthode des moindres carrés est une méthodologie connue pour effectuer des tests en théories quantitatives et elle semble plus appropriée pour établir un lien entre les variables que nous aurons choisies notamment la variable expliquée à caractère aléatoire et les variables explicatives mesurées sans erreurs.

Ce mémoire comporte trois chapitres.

-Dans le premier, on présente quelques résultats fondamentaux qu'on utilise pour étudié les problèmes d'optimisations non linéaires sans contraintes puis avec contraintes d'égalité et d'inégalité.

-Le deuxième chapitre est consacré à l'application de la méthode des moindres carrés.

-En dernier, dans la partie pratique on a fait une simulation numérique avec le logiciel MATLAB de la méthode des moindres carrés.

Optimisation non linéaire

1.1 Introduction à \mathbb{R}^n

1.1.1 L'espace euclidien de \mathbb{R}^n :

Définition 1.1

\mathbb{R}^n est l'ensemble des points (*n*-uplets), $P = (x_1, x_2, \dots, x_n)$ où $x_i \in \mathbb{R}$, $i = 1, \dots, n$.

x_i est appelé *i^{ème}* coordonnée ou composante de P .

P est appelé vecteur.

P est appelé point (la valeur de la fonction en un point).

1.1.2 Structure d'espace vectoriel :

Soient :

$$P = (x_1, x_2, \dots, x_i, \dots, x_n)^T \quad Q = (y_1, y_2, \dots, y_i, \dots, y_n)^T$$

.

- On définit l'**addition** dans \mathbb{R}^n par :

$$P + Q = (x_1 + y_1, x_2 + y_2, \dots, x_i + y_i, \dots, x_n + y_n)^T \in \mathbb{R}^n$$

.

- On définit la **multiplication** par un scalaire λ :

$$\lambda P = (\lambda x_1, \lambda x_2, \dots, \lambda x_i, \dots, \lambda x_n)^T$$

on montre facilement que $(\mathbb{R}^n, +, \cdot)$ est un **espace vectoriel**.

- On définit le **produit scalaire** dans \mathbb{R}^n par :

$$\langle P, Q \rangle = P \cdot Q = \sum_{i=1}^n x_i y_i \in \mathbb{R}$$

Remarque : Le produit scalaire satisfait les propriétés suivantes :

- * $P \cdot Q = Q \cdot P.$
- * $P(Q_1 + Q_2) = PQ_1 + PQ_2 .$
- * $(\lambda P) \cdot Q = \lambda(P \cdot Q).$
- * $P \cdot P = \sum_{i=1}^n x_i^2 = \|P\|^2.$

• Si on a deux vecteurs $x, y \in \mathbb{R}^n$ et $\langle x, y \rangle = 0$, alors les deux vecteurs sont **orthogonaux** (on notera $x \perp y$).

Définition 1.2

On appelle espace **euclidien**, un espace vectorielle muni d'un produit scalaire $\langle \cdot, \cdot \rangle$.

Définition 1.3

pour tout $x \in \mathbb{R}^n$ on note $\|x\|$ **la norme euclidienne** de x donnée par :

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^n x_i^2}$$

qui vérifie les inégalités suivantes :

1. $\|x\| \geq 0$, où $\|x\| = 0$ si $x = 0$.
2. $\|\alpha x\| = |\alpha| \cdot \|x\|.$
3. $\|x + y\| \leq \|x\| + \|y\|.$

1.1.3 Inégalité de Cauchy Schwarz :

Soient P et Q deux vecteur dans \mathbb{R}^n alors :

$$|P \cdot Q| \leq \|P\| \cdot \|Q\|.$$

Preuve :

$$P \cdot Q = \|P\| \cdot \|Q\| \cdot \cos(P \cdot Q)$$

$$|P \cdot Q| = \|\|P\| \cdot \|Q\| \cdot \cos(P \cdot Q)\|$$

$$|P \cdot Q| = \|P\| \cdot \|Q\| \cdot |\cos(P \cdot Q)| \quad \text{avec : } \cos(P \cdot Q) \leq 1$$

$$\text{donc : } |P \cdot Q| \leq \|P\| \cdot \|Q\|$$

1.2 Notions topologiques :

Définition 1.4

Soit $r > 0$, l'ensemble des points :

$$\mathcal{B}_r(A) = \{P \in \mathbb{R}^n / \|P - A\|_2 < r\} \Rightarrow \text{Boule de centre } A \text{ et de rayon } r.$$

$$\mathcal{S}_r(A) = \{P \in \mathbb{R}^n / \|P - A\|_2 = r\} \Rightarrow \text{Sphère de centre } A \text{ et de rayon } r.$$

Remarque :

$\|P - A\|_2$ définit la distance de P à A notée $d(P, A)$.

Exemple 1.1

Dans \mathbb{R}^2 :

$$1. \quad \mathcal{B}_r(A) = \{P \in \mathbb{R}^2 / \|P - A\|_2 < r\}$$

est le disque sans la frontière.

$$2. \quad \mathcal{S}_r(A) = \{P \in \mathbb{R}^2 / \|P - A\|_2 = r\}$$

est le cercle de centre A et de rayon r .

Définition 1.5

Soit $\mathcal{S} \subset \mathbb{R}^n$

- Un point P de \mathcal{S} est appelé point intérieur à \mathcal{S} s'il existe $r > 0$ telle que : $\mathcal{B}_r(P) \subset \mathcal{S}$.

- L'ensemble des points intérieurs à \mathcal{S} est appelé intérieur de \mathcal{S} noté $\overset{o}{\mathcal{S}}$.

Définition 1.6

1. On dit que S est un ouvert si $S = \overset{o}{S}$.
2. Une partie F dite fermée si son complémentaire est ouvert.

Remarques :

* \mathbb{R}^n et \emptyset sont ouverts et fermés en même temps.

On a des ensembles ni ouvert ni fermé comme $[a, b[$.

* $\mathcal{B}_r(A)$ est appelé voisinage de A pour un r petit.

Définition 1.7

Un ensemble $S \subset \mathbb{R}^n$ est dit compact si il est **fermé** et **borné** en même temps.

Théorème 1.2.1 Théorème de Weierstrass[1]

Si f une fonction réelle continue sur $\mathbb{K} \subset \mathbb{R}^n$, compact (\mathbb{K} fermé borné) alors le problème d'optimisation :

$$\begin{cases} \min f(X) \\ X \in \mathbb{K} \end{cases}$$

a une solution optimale $X^* \in \mathbb{K}$. Ce résultat reste valide si on suppose seulement f semi-continue inférieurement (S.C.I).

1.3 Matrices

Définition 1.8

On définit la matrice $A_{m \times n}$ par un tableau avec : $a_{i,j} \in \mathbb{R}$; $\begin{cases} i = 1, \dots, m \\ j = 1, \dots, n \end{cases}$

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}$$

$a_{i,j}$ se trouve à la $i^{\text{ème}}$ ligne et la $j^{\text{ème}}$ colonne.

Le vecteur $(a_{i,1}, a_{i,2}, \dots, a_{i,n})$ est appelé $i^{\text{ème}}$ ligne.

Le vecteur $\begin{pmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{pmatrix}$ est appelé la $j^{\text{ème}}$ colonne.

-On définit la transposée de A par : A^T .

Remarque :

Si $m = n$ A est appelé **matrice carrée**.

Définition 1.9

Soit A une matrice $(m \times n)$

On définit le rang de A par le membre maximum de lignes ou colonnes linéairement indépendantes, on le note $rg(A)$.

C'est le plus grand ordre de déterminants non nulles qu'on peut extraire de la matrice A .

1.3.1 Signe d'une matrice

Définition 1.10

1. Soit A une matrice carrée $(m \times m)$ symétrique.

- A est dite définie positive si $X^T A X > 0$, $\forall X \in \mathbb{R}^n$, $X \neq 0$.
- A est dite semi-définie positive si $X^T A X \geq 0$, $\forall X \in \mathbb{R}^n$.
- A est dite définie négative si $X^T A X < 0$, $\forall X \in \mathbb{R}^n$, $X \neq 0$.
- A est dite semi-définie négative si $X^T A X \leq 0$, $\forall X \in \mathbb{R}^n$.
- Si aucun des cas n'est vérifié, on dit que A est non définie.

2. **Critère de Sylvester :** Soit A une matrice carrée $(m \times m)$ symétrique.

- A est dite définie positive si :

$$\Delta_k = \det A_k > 0, \quad k = 1, \dots, m, \quad A_k = \begin{pmatrix} a_{1,1} & \dots & a_{1,k} \\ \vdots & \ddots & \vdots \\ a_{k,1} & \dots & a_{k,k} \end{pmatrix}$$

$\Delta_k = \det A_k$ est appelé **mineur principal** d'ordre k .

- A est dite *semi-définie positive* si :

$$\Delta_k = \det A_k \geq 0, \quad k = 1, \dots, m.$$

- A est dite *définie négative* si :

$$\Delta_k = (-1)^k \det A_k > 0, \quad k = 1, \dots, m.$$

- A est dite *semi-définie négative* si :

$$\Delta_k = (-1)^k \det A_k \geq 0, \quad k = 1, \dots, m.$$

Remarque :

Si tous les mineurs principaux sont nuls $\Delta_k = \det A_k = 0, \quad k = 1, \dots, m$ alors il faut appliquer le **critère des valeurs propres**.

3. Critère qui utilise les valeurs propres

Soient λ_i les valeurs propres de A :

- A est *définie positive* si $\lambda_i > 0 \quad \forall i$.
- A est *semi-définie positive* si $\lambda_i \geq 0 \quad \forall i$.
- A est *définie négative* si $\lambda_i < 0 \quad \forall i$.
- A est *semi-définie négative* si $\lambda_i \leq 0 \quad \forall i$.
- A est *non définie* s'il existe des valeurs positives et négatives strictement.

4. Signe d'une matrice sur un sous espace

Soient A une matrice carrée $(n \times n)$ symétrique

$\mathbb{M} = \{X \in \mathbb{R}^n / BX = 0\}$ un sous espace, avec B une matrice $(m \times n)$.

On suppose que $m < n$ et $\text{rg}(B) = m$, on a $\dim(\mathbb{M}) = n - m$.

Soient $p = (p_1, p_2, \dots, p_{n-m})$ les vecteurs d'une base de \mathbb{M} et

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{n-m}) \in \mathbb{R}^{n-m}$$

$$\forall X \in \mathbb{M}, \quad X = \sum_{i=1}^{n-m} \alpha_i p_i = (p_1, p_2, \dots, p_{n-m}) \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{n-m} \end{pmatrix} = p\alpha$$

$$X^T A X = (p\alpha)^T A p\alpha = \alpha^T p^T A p\alpha = \alpha^T \cdot p^T A p \cdot \alpha.$$

Le signe de A sur \mathbb{M} est égal au signe de $p^T A p$ sur \mathbb{R}^n .

1.4 Fonctions de plusieurs variables

Définition 1.11

Soit D un domaine de définition de \mathbb{R}^n , une application f de D dans \mathbb{R} est dite fonction réelle de n variables réelles.

$$\begin{cases} f: D \longrightarrow \mathbb{R} \\ X = (x_1, x_2, \dots, x_n) \longrightarrow f(x_1, x_2, \dots, x_n). \end{cases}$$

Définition 1.12

On dit que f est continue en $x^* \in D$ si :

$$\forall \varepsilon > 0, \exists \delta > 0 / \|x - x^*\| < \delta \Rightarrow |f(x) - f(x^*)| < \varepsilon.$$

Et on écrit :

$$\lim_{x \rightarrow x^*} f(x) = f(x^*).$$

Remarque :

f est continue sur D si elle est continue en chaque point de D .

1.4.1 Dérivées partielles

Définition 1.13

• On définit la dérivée partielle par rapport à x_i de la fonction f au point $X = (x_1, x_2, \dots, x_i, \dots, x_n)^T$ par la limite :

$$\frac{\partial f}{\partial x_i}(X) = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_n)}{h}$$

quand elle existe et elle est finie, on la note $\frac{\partial f}{\partial x_i}(x_1, \dots, x_n)$.

• On définit le gradient de f au point X par :

$$\nabla f(X) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(X) \\ \vdots \\ \frac{\partial f}{\partial x_n}(X) \end{pmatrix}.$$

Remarque :

Si les dérivées partielles sont continues sur D on dit que f est de classe C^1 .

Définition 1.14

Soit f une fonction continue dérivable sur \mathbb{R}^n (de classe C^1), on définit la matrice **Jacobienne** de f par :

$$J(f(X)) = \begin{pmatrix} \nabla f_1(X)^T \\ \vdots \\ \nabla f_m(X)^T \end{pmatrix}_{m \times n}.$$

1.4.2 Dérivées d'ordre supérieurs à 1

Définition 1.15

Si les dérivées partielles de f existent et sont dérivables sur D , leurs dérivées partielles sont pour f des dérivées partielles secondes, et on écrit :

$$\frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_i}(X) \right) = \frac{\partial^2 f}{\partial x_j \partial x_i}(X).$$

Théorème 1.4.1 Théorème de Schwarz[\[4\]](#)

Soient E et F deux espaces vectoriels normés, U un ouvert de E , et $f : U \rightarrow F$ une application deux fois dérivable en un point a de U . alors, l'application bilinéaire $d^2 f_a : E \times E \rightarrow F$ est symétrique.

Remarque :

f est dite de classe C^2 si les dérivées secondes sont continues sur D .

Définition 1.16

On définit la matrice **Hessienne** $\forall X \in \mathbb{R}^n$ par :

$$Hf(X) = J(\nabla f(X)) = \begin{pmatrix} \frac{\partial^2 f(X)}{\partial x_1^2} & \frac{\partial^2 f(X)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(X)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(X)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(X)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(X)}{\partial x_n \partial x_1} & \frac{\partial^2 f(X)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_n^2} \end{pmatrix}$$

Remarque :

Pour une fonction de classe C^2 , la matrice hessienne est **symétrique**.

Théorème 1.4.2 Théorème des accroissements finis :

Soient $f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}$

X et X^* deux points de D

Alors : $f(X) - f(X^*) = (X - X^*)^T \nabla f(\bar{X})$ avec \bar{X} compris entre X et X^* .

1.4.3 Formule de Taylor au point X^* à l'ordre 2

Pour f de classe C^2 .

$$f(X) = f(X^*) + (X - X^*)^T \nabla f(X^*) + \frac{1}{2} (X - X^*)^T Hf(X^*) (X - X^*) + \|X - X^*\| \varepsilon(X) \quad (1.1)$$

avec :

$$\lim_{X \rightarrow X^*} \varepsilon(X) = 0$$

Définition 1.17

On définit la **dérivée directionnelle** de f dans la direction d au point X^* par :

$$\lim_{t \rightarrow 0^+} \frac{f(X^* + td) - f(X^*)}{t} = f'(X^*, d)$$

1.5 Convexité

1.5.1 Ensembles convexes

Définition 1.18

Un ensemble \mathbb{K} est dit convexe si et seulement si :
 $\forall X_1, X_2 \in \mathbb{K}$ alors, $\lambda X_1 + (1 - \lambda)X_2 \in \mathbb{K}$, $\lambda \in [0, 1]$.

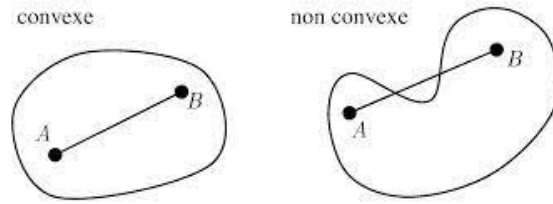


FIGURE 1.1 – Ensemble convexe et non convexe

Définition 1.19

Soient $\lambda_i \geq 0$ $i = 1, \dots, n$, tel que

$$\sum_{i=1}^n \lambda_i = 1 \quad \text{alors} \quad \sum_{i=1}^n \lambda_i X_i$$

est appelée combinaison linéaire convexe des points X_1, \dots, X_n .

On vérifie aisément qu'un ensemble $\mathbb{K} \subset \mathbb{R}^n$ est convexe si et seulement si tout point, combinaison convexe de points de \mathbb{K} est dans \mathbb{K} .

Théorème 1.5.1 [1]

Tout sous ensemble convexe non vide de \mathbb{R}^n a un intérieur relatif non vide.

Définition 1.20

On appelle **enveloppe convexe** d'un ensemble S le plus petit ensemble convexe contenant S , on le note $\text{conv}(S)$.

Remarque : Si S est convexe alors $\text{conv}(S) = S$.

Définition 1.21

1. On définit un **sommet** de S convexe ou bien **point extrême** $X \in S$, si X ne peut pas s'écrire comme combinaison linéaire convexe de deux points de S .
2. Soit $X_0 \notin K$ convexe fermé, on définit la projection de X_0 sur K par $P_{X_0} \in K$ tel que

$$\|P_{X_0} - X_0\| = \min_{X \in K} \|X - X_0\| \leq \|X - X_0\|, \quad \forall X \in K.$$

Théorème 1.5.2 [6]

La projection P_{X_0} de X_0 sur K convexe fermé est unique.

Proposition : [6]

Pour que P_{X_0} soit la **projection** de X_0 sur K convexe fermé, il faut et il suffit que :

$$(X - P_{X_0})(X_0 - P_{X_0}) \leq 0, \quad \forall X \in K.$$

Calcul de la projection de X_0 sur K :

Pour calculer la projection de X_0 sur K , il faut résoudre le problème d'optimisation convexe suivant :

$$\min_{X \in K} \|X - X_0\|$$

1.5.2 Fonctions convexes

Définition 1.22

Soit f une fonction définie sur un convexe $K \subset \mathbb{R}^n$:

- f est dite **convexe** si : $\forall X_1, X_2 \in K, \quad \forall \lambda \in [0, 1]$

$$f(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda f(X_1) + (1 - \lambda)f(X_2).$$

- f est dite **strictement convexe** si : $\forall X_1, X_2 \in K, \quad \forall \lambda \in]0, 1[$

$$f(\lambda X_1 + (1 - \lambda)X_2) < \lambda f(X_1) + (1 - \lambda)f(X_2).$$

• f est dite **fortement convexe** s'il existe $\alpha > 0$ tel que :
 $\forall X_1, X_2 \in \mathbb{K}, \quad \forall \lambda \in [0, 1]$

$$f(\lambda X_1 + (1 - \lambda)X_2) \leq \lambda f(X_1) + (1 - \lambda)f(X_2) - \frac{\alpha}{2}\lambda(1 - \lambda)\|X_1 - X_2\|^2.$$

• f est dite **concave** sur \mathbb{K} si : $-f$ est **convexe**, autrement dit
 si : $\forall X_1, X_2 \in \mathbb{K}, \quad \forall \lambda \in [0, 1]$

$$f(\lambda X_1 + (1 - \lambda)X_2) \geq \lambda f(X_1) + (1 - \lambda)f(X_2).$$

Théorème 1.5.3 [1]

Une combinaison linéaire à coefficients positifs de fonctions convexes est une fonction convexe.

Remarque :

f est dite **concave** (respectivement **strictement concave**, respectivement **fortement concave**) si $-f$ est **convexe** (respectivement **strictement convexe**, respectivement **fortement convexe**) sur \mathbb{K} convexe.

Corollaire [1]

une fonction quadratique dont le hessien est semi-définie positif est une fonction convexe.

1.5.3 Fonctions coercives

Définition 1.23

Soit f une fonction définie sur un ensemble non borné $\mathbb{K} \subset \mathbb{R}^n$, on dit que f est coercive sur \mathbb{K} si on a :

$$\lim_{X \in \mathbb{K}, \|X\| \rightarrow +\infty} f(X) = +\infty$$

On peut aussi écrire : pour toute suite $(X_k)_{k \in \mathbb{N}} \subset \mathbb{K}$, on a $f(X_k) \rightarrow +\infty$, telle que $\|X_k\| \rightarrow +\infty$.

1.6 Optimisation sans contraintes [2]

Soit f une fonction continue $\forall X \in \mathbb{R}^n$ tel que :

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ X &\longrightarrow f(X) \end{aligned}$$

On définit un problème d'optimisation sans contrainte par :

$$(\mathcal{P}) : \left\{ \min_{X \in \mathbb{R}^n} f(X) \right. \quad \text{ou} \quad (\mathcal{P}') \left\{ \max_{X \in \mathbb{R}^n} f(X) \right.$$

Rappel :

$$\max_{X \in \mathbb{R}^n} f(X) = - \min_{X \in \mathbb{R}^n} (-f(X)).$$

Définition 1.24

- On dit que X^* est un minimum local pour f si : $f(X^*) \leq f(X)$, $\forall X \in V(X^*)$.
- On dit que X^* est un minimum global pour f si : $f(X^*) \leq f(X)$, $\forall X \in \mathbb{R}^n$.
- On dit que X^* est un maximum local pour f si : $f(X^*) \geq f(X)$, $\forall X \in V(X^*)$.
- On dit que X^* est un maximum global pour f si : $f(X^*) \geq f(X)$, $\forall X \in \mathbb{R}^n$.

Avec $V(X^*)$ c'est le voisinage de X^* .

1.6.1 Résultat d'existence et d'unicité

Avant d'étudier les propriétés de la solution (ou les solutions) de (\mathcal{P}) il faut s'assurer de leur existence.

Nous donnerons en suite des résultat d'unicité.

Théorème 1.6.1 *Existence* : [2]

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ propre, continue et coercive.

Alors (\mathcal{P}) admet au moins une solution.

Preuve : (Voir [2]).

Théorème 1.6.2 *Unicité* :[2]

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ strictement convexe. Alors (\mathcal{P}) admet au plus une solution.

Preuve :(Voir [2]).

Théorème 1.6.3 [2]

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ de classe C^1 .

On suppose qu'il existe $\alpha > 0$ tel que :

$$\forall (X, Y) \in \mathbb{R}^n \times \mathbb{R}^n, \quad (\langle \nabla f(X) - \nabla f(Y), X - Y \rangle) \geq \alpha \|X - Y\|^2. \quad (1.2)$$

Alors f est strictement convexe et coercive, en particulier le problème (\mathcal{P}) admet une solution unique.

Preuve :(Voir [2])

Remarque :

Si la condition (1.2) est vérifiée, alors on dit que $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ est elliptique.

1.6.2 Conditions nécessaires et suffisantes d'optimalité de 1^{er} ordre

conditions nécessaires :

Théorème 1.6.4

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ de classe C^1 différentiable.

Si X^* est un **minimum local** de (\mathcal{P}) , alors $\nabla f(X^*) = 0$.

Preuve :[2]

D'après la formule de Taylor à l'ordre 1, avec $t > 0$, on pose :

$X = (X^* + td), \forall d \in \mathbb{R}^n :$

$$f(X^* + td) = f(X^*) + \nabla f(X^*)^T td + o\|td\|$$

$$\Rightarrow \frac{f(X^* + td) - f(X^*)}{t} = \frac{\nabla f(X^*)^T td + o\|td\|}{t}$$

Par passage à la limite lorsque $t \mapsto 0$ (car f de classe C^1 donc continue)

on aura :

$$\Rightarrow 0 \leq \lim_{t \rightarrow 0^+} \frac{f(X^* + td) - f(X^*)}{t} \leq \nabla f(X^*)^T d$$

Et comme $\nabla f(X^*)^T \geq 0$ et $d = -\nabla f(X^*)$ dans le cas d'un minimum.

Alors :

$$\begin{aligned} \nabla f(X^*)^T (-\nabla f(X^*)) &\geq 0 \Rightarrow -\|\nabla f(X^*)\|^2 \geq 0 \\ \Rightarrow \|\nabla f(X^*)\|^2 &= 0 \\ \Rightarrow \nabla f(X^*) &= 0 \end{aligned}$$

D'où le résultat.

Remarque :

La condition $\nabla f(X^*) = 0$ n'est pas suffisante i.e X^* un point critique, peut être : un minimum, un maximum ou un point selle.

Conditions suffisantes :

Théorème 1.6.5 [3]

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1 différentiable et **convexe**.
 X^* est un **minimum global** si et seulement si $\nabla f(X^*) = 0$.

Remarque :

Dans le cas d'un **maximum global** f doit être différentiable concave.

1.6.3 Conditions nécessaires et suffisantes d'optimalité de 2^{ème} ordre

Conditions nécessaire :

Théorème 1.6.6

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^2 .
 X^* est un **minimum local**, alors :

$$\begin{cases} \nabla f(X^*) = 0 \\ Hf(X^*) \text{ est } \mathbf{semi-définie positive.} \end{cases}$$

X^* est un **maximum local**, alors :

$$\begin{cases} \nabla f(X^*) = 0 \\ Hf(X^*) \text{ est } \mathbf{semi-définie négative.} \end{cases}$$

Preuve : [2]

Il suffit de montrer le cas d'un minimum :

On a pour tout $X \in V(X^*)$, (1.1) est vérifiée car f de classe C^2 .
avec $t > 0$, on pose : $X = (X^* + td)$, $\forall d \in \mathbb{R}^n$, on aura donc :

$$f(X^* + td) = f(X^*) + \nabla f(X^*)td + \frac{1}{2}(td^T)Hf(X^*)(td) + \varepsilon\|td\|^2$$

D'où

$$0 \leq \frac{f(X^* + td) - f(X^*)}{t^2} = \frac{1}{2}(td^T)Hf(X^*)(td) + \frac{\varepsilon\|td\|^2}{t^2}$$

Par passage à la limite lorsque $t \mapsto 0$ on aura :

$$0 \leq \lim_{t \mapsto 0} \frac{f(X^* + td) - f(X^*)}{t^2} = \frac{1}{2}d^T Hf(X^*)d + \lim_{t \mapsto 0} \frac{\varepsilon\|td\|^2}{t^2}$$

D'où le résultat : $d^T Hf(X^*)d \geq 0 \Rightarrow Hf(X^*) \geq 0, \forall d \in \mathbb{R}^n$

De la même manière on montre le cas d'un maximum local.

Remarque :

Si la matrice hessienne en un point critique est non définie, alors il s'agit d'un **point selle**.

Conditions suffisantes :

Théorème 1.6.7 [2]

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ de classe C^2 , si :

$$\begin{cases} \nabla f(X^*) = 0 \\ Hf(X^*) \text{ est } \textbf{définie positive}. \end{cases}$$

Alors X^* est un minimum local strict.

Et si :

$$\begin{cases} \nabla f(X^*) = 0 \\ Hf(X^*) \text{ est } \textbf{définie négative}. \end{cases}$$

Alors X^* est un maximum local strict.

Preuve : (Voir [2])

Remarque :

Dans le cas où f est convexe, alors tout minimum local est aussi global. De plus si f est strictement convexe, alors tout minimum local devient non seulement global mais aussi unique.

Exemple 1.2

Soit le problème suivant à résoudre :

$$\left\{ \min_{X \in \mathbb{R}^3} f(X) = x_1^2 + (x_1 + x_2 + x_3)^2 + (x_2 + x_3)^2 + 2x_3^2 \right.$$

avec $X = (x_1, x_2, x_3) \in \mathbb{R}^3$.

On calcul le gradient $\nabla f(X)$ et $Hf(X)$, $\forall X \in \mathbb{R}^3$:

$$\nabla f(X) = \begin{pmatrix} 4x_1 + 2x_2 + 2x_3 \\ 2x_1 + 4x_2 + 2x_3 \\ 2x_1 + 2x_2 + 8x_3 \end{pmatrix} ; \quad Hf(X) = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 8 \end{pmatrix}$$

Comme f de classe C^2 , et $\nabla f(X^*) = (0, 0, 0)^T$ ce qui donne $X^* = (0, 0, 0)^T$ d'après la résolution d'un système de type $AX^* = (0, 0, 0)^T$ tel que $A = Hf(X)$, avec $rg(A) = 3$.

Et comme $Hf(X)$ est symétrique, car f de classe C^2 , donc d'après le **Critère de Sylvester** :

$$\begin{cases} \Delta_1 = 4 > 0 \\ \Delta_2 = 12 > 0 \\ \Delta_3 = 80 > 0 \end{cases}$$

avec $\Delta_i > 0$, $\forall i = 1, 2, 3$ **les mineurs principal**.

Donc $Hf(X)$ est définie positive $\forall X \in \mathbb{R}^3$.

\Rightarrow la condition suffisante de seconde ordre est vérifiée, alors $X^* = (0, 0, 0)$ est la solution de problème.

1.6.4 Algorithme

Un algorithme est défini par une application $A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ permet de générer une suite d'éléments de \mathbb{R}^n par la formule :

$$\begin{cases} X_0 \in \mathbb{R}^n \text{ donné, } k = 0 \text{ étape d'initialisation.} \\ X_{k+1} = A(X_k), \quad k = k + 1 \text{ itération } k. \end{cases}$$

Algorithme de gradient à pas optimal : [3]

Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.

1. Initialisation : fixer un $\varepsilon > 0$, et choisir un $X_0 \in \mathbb{R}^n$.

2. Itération k tel que $k = 1, 2, \dots, n$.

$$X_{k+1} = X_k - \alpha_k \nabla f(X_k)$$

avec α_k (pas optimal) solution de :

$$\min_{\alpha > 0} f(X_k - \alpha \nabla f(X_k))$$

3. Test d'arrêt :

Si $\|X_{k+1} - X_k\| \leq \varepsilon$ **stop**, X_{k+1} est solution optimal.

Sinon refaire **(2.)** avec $k = k + 1$.

Exemple 1.3 On applique l'algorithme de gradient à pas optimal pour la fonction $f(x_1, x_2) = x_1^2 + x_2^2$.

1. Soit $X_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\varepsilon = 10^{-6} > 0$.

2. Itération 1 :

$$\begin{aligned} X_1 &= X_0 - \alpha_0 \nabla f(X_0) \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \alpha_0 \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\ \text{avec } \alpha_0 &= \min_{\alpha_0 > 0} f(X_0 - \alpha_0 \nabla f(X_0)) \Rightarrow \alpha_0 = \frac{1}{2} \\ \Rightarrow X_1 &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

3. Test d'arrêt : $\|X_1 - X_0\| = \sqrt{2} > \varepsilon$ donc en refaire **2.** avec $k = k + 1$.

Itération 2 :

$$\begin{aligned} X_2 &= X_1 - \alpha_1 \nabla f(X_1) \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \alpha_1 \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\text{ce qui donne } \forall \alpha_1 > 0, X_2 = X_1. \end{aligned}$$

Test d'arrêt : $\|X_2 - X_1\| = 0 < \varepsilon$.

$\Rightarrow X_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ est la solution optimale.

Remarque :

L'algorithme de gradient à pas optimal fait de **zigzag** pour aller vers la solution optimal.

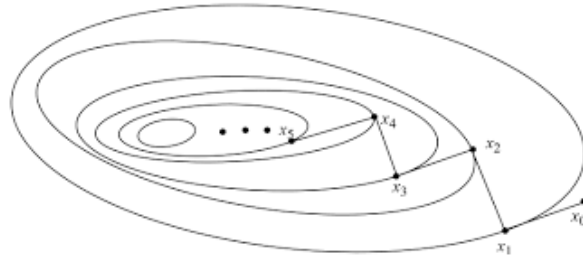


FIGURE 1.2 – Méthode de Gradient

Inconvénients :

- La convergence de l'algorithme de gradient à pas optimal est linéaire donc assez lente.
- Le calcul de α_k d'une façon exacte est parfois compliqué et prend beaucoup de temps.

Méthode d'Armijo : [3]

Pour calculer α_k à chaque itération, on considère la suite $\gamma_i = \frac{1}{2^i}$ avec $i \in \mathbb{N}$, on prend pour α_k le plus grand γ_i , vérifiant la condition d'Armijo :

$$f(X_k - \gamma_i \nabla f(X_k)) \leq f(X_k) - \frac{1}{2} \gamma_i \|\nabla f(X_k)\|^2.$$

Remaque :

Une fois que le α_k est calculer, on utilise les instructions de l'algorithme de gradient à pas optimal on trouve la solution optimale.

Algorithme de gradient à pas fixe : [3]

On calcul $\alpha_k = \alpha$ constant pour toutes les instructions, mais la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ doit vérifier les hypothèses suivantes :

- 1) f de classe C^1 .
- 2) $\|\nabla f(X) - \nabla f(Y)\| \leq L\|X - Y\|$, $\forall X, Y \in \mathbb{R}^n$, L est la constante de Lipschitz.
- 3) $(\nabla f(X) - \nabla f(Y))(X - Y) \geq \rho\|X - Y\|^2$, $\forall X, Y \in \mathbb{R}^n$

Si 1), 2) et 3) sont vérifiées alors on applique l'algorithme :

1. **Initialisation** : fixer un $\varepsilon > 0$, et choisir un $X^0 \in \mathbb{R}^n$.
2. **Itération k** tel que $k = 1, 2, \dots, n$.

$$X_{k+1} = X_k - \alpha \nabla f(X_k), \quad (\alpha = \frac{\rho}{L^2})$$

3. Test d'arrêt :

Si $\|X_{k+1} - X_k\| \leq \varepsilon$ **stop**, X_{k+1} est solution optimale.

Sinon refaire 2. avec $k = k + 1$.

Exemple 1.4 Soit la fonction de l'exemple précédent tel que f vérifie les hypothèses (1), (2), (3) avec $L = 2$ et $\rho = 2 \Rightarrow \alpha = \frac{1}{2}$, on applique l'algorithme :

Itération 1 :

$$X_1 = X_0 - \frac{1}{2} \nabla f(X_0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Test d'arrêt : $\|X_1 - X_0\| = \sqrt{2} > 0$.

Itération 2 :

$$X_2 = X_1 - \frac{1}{2} \nabla f(X_1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Test d'arrêt : $\|X_2 - X_1\| = 0 < \varepsilon$ stop. Donc $X_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ est la solution optimale.

Algorithme de Newton : [3]

Étape initiale : Soit $\varepsilon > 0$: test d'arrêt, choisir X_0 point initiale, poser $k = 0$ et aller à l'étape suivante.

Étape 2 : Si $\|\nabla f(X_k)\| \leq \varepsilon$ stop solution de problème.
Sinon poser $X_{k+1} = X_k - Hf(X_k)^{(-1)}\nabla f(X_k)$, remplacer k par $k+1$ et aller à l'étape 2.

Exemple 1.5 On refaire l'exemple précédent avec l'algorithme de Newton :

1. Soit $X_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\varepsilon = 10^{-6} > 0$.

2. $\|\nabla f(X_0)\| > \varepsilon$, donc on pose :

$$\begin{aligned} X_1 &= X_0 - Hf(X_0)^{-1}\nabla f(X_0) \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

Test d'arrêt : $\|\nabla f(X_1)\| = 0 < \varepsilon$ stop.

$\Rightarrow X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ est la solution optimale.

1.7 Optimisation avec contraintes [7]

On s'intéresse dans cette partie à l'étude d'un type de problèmes :

$$(P) : \min_{X \in \mathbb{R}^n} \{f(X) : X \in \mathbb{K}\} \text{ avec } \mathbb{K} \subseteq \mathbb{R}^n. \quad (1.3)$$

$$(P') : \max_{X \in \mathbb{R}^n} \{f(X) : X \in \mathbb{K}\} \text{ avec } \mathbb{K} \subseteq \mathbb{R}^n. \quad (1.4)$$

1.7.1 Optimisation avec contraintes d'égalités

Soit le problème (1.3) avec $\mathbb{K} = \{X \in \mathbb{R}^n : k_i(X) = 0, i = 1, \dots, m\}$.

Définition 1.25

On définit le problème d'optimisation avec contraintes d'égalités par :

$$(P) : \begin{cases} \min f(X) \\ k_i(X) = 0 \quad (i = 1, \dots, m) \\ X \in \mathbb{R}^n \end{cases}$$

Définition 1.26

On pose : $K(X) = \begin{pmatrix} k_1(X) \\ \vdots \\ k_i(X) \\ \vdots \\ k_m(X) \end{pmatrix}.$

- On dit que $JK(X)$ est de rang plein si

$$rg(JK(X)) = m$$

- On dit qu'un point X^* est un point régulier ou bien vérifie la condition de qualification des contraintes si $JK(X^*)$ est de rang plein.

Théorème 1.7.1 théorème de Lagrange (Condition nécessaire de 1^{er} ordre)[\[3\]](#)

Supposons que X^* est un minimum local régulier, pour le problème (P), alors il existe $\lambda_1^*, \lambda_2^*, \dots, \lambda_n^* \in \mathbb{R}$ tel que

$$\begin{aligned} \nabla f(X^*) + \sum_{i=1}^m \lambda_i^* \nabla k_i(X^*) &= 0. \\ k(X^*) &= 0 \quad (i = 1, \dots, m). \end{aligned}$$

Remarque :

- Cette condition n'est pas suffisante c'est-à-dire les points qui vérifient cette condition peuvent être des maximaux, des minimaux ou des points cols (points selles).

- On pratique, on résout le système

$$\begin{cases} \nabla f(X) + \sum_{i=1}^m \lambda_i \nabla k_i(X) = 0 \Leftrightarrow \nabla_X(L(X^*, \lambda^*)) = 0. \\ k_i(X) = 0 \quad (i = 1, \dots, m) \quad \Leftrightarrow \nabla_\lambda(L(X^*, \lambda^*)) = 0. \end{cases}$$

de $(n + m)$ équations à $(n + m)$ inconnus.

Définition 1.27

On appelle **Lagrangien** une fonction de **Lagrange** associée au problème (P) de la fonction

$$L(X, \lambda) = f(X) + \sum_{i=1}^m \lambda_i k_i(X)$$

Proposition

Soit X^* un minimum local régulier pour (P) et λ^* le vecteur des multiplicateurs de Lagrange associé à X^* alors $L(X^*, \lambda^*)$ est un point stationnaire de Lagrangien c'est-à-dire $\nabla L(X^*, \lambda^*) = 0$.

Preuve :

D'après le théorème de Lagrange, on a

$$\begin{cases} \nabla f(X) + \sum_{i=1}^m \lambda_i \nabla k_i(X) = 0 \Leftrightarrow \nabla_X(L(X^*, \lambda^*)) = 0, \dots \textbf{(1)} \\ k_i(X) = 0 \quad (i = 1, \dots, m) \quad \Leftrightarrow \nabla_\lambda(L(X^*, \lambda^*)) = 0, \dots \textbf{(2)} \end{cases}$$

De **(1)** et **(2)** $\Rightarrow \nabla L(X^*, \lambda^*) = 0 \Rightarrow (X^*, \lambda^*)$ est un point stationnaire de Lagrangien.

Remarque

- Si la qualification des contraintes est vérifiée partout :
 - (a) alors parmi les points stationnaires de Lagrangien se trouve la solution à confirmer par la condition suffisante.
 - (b) S'il n'existe pas de multiplicateurs de Lagrange alors le problème n'admet pas de solutions.
- Si la qualification des contraintes n'est pas vérifiée partout, alors les

extremums qui ne sont pas réguliers ne peuvent pas être trouvés par la méthode de Lagrange .

- En général pour les points pratiques la qualification des contraintes est vérifié partout.

Définition 1.28

- On appelle sous espace tangent aux contraintes au point X^* noté $T(X^*) = \{d \in \mathbb{R}^n / JK(X^*)d = 0\}$
- d est dite direction admissible en X^* si $X^* + td$ est admissible, ($t > 0$).

Proposition

soit X^* un point admissible, alors $T(X^*)$ contient l'ensemble des directions admissibles.

Preuve

Soit d une direction admissible au point X^* c'est-à-dire $X^* + td$ avec ($t > 0$) est admissible ,on a la formule de Taylor à l'ordre 1 au point X^* :

$$K(X^* + td) = K(X^*) + JK(X^*)td + \varepsilon(\|td\|)$$

$$\Rightarrow \frac{K(X^* + td) - K(X^*)}{t} = \frac{JK(X^*)td + \varepsilon(\|td\|)}{t}$$

Par passage à la limite :

$$\lim_{t \rightarrow 0^+} \frac{K(X^* + td) - K(X^*)}{t} = JK(X^*)d + \lim_{t \rightarrow 0^+} \frac{\varepsilon(t)}{t} \|d\| = JK(X^*)d.$$

Donc $d \in T(X^*)$.

Théorème 1.7.2 (Condition nécessaire de 2^{ème} ordre)[\[3\]](#)

On suppose que f et k_i ($i = 1, \dots, m$) sont de classe C^2 . Soit X^* un minimum local régulier alors :

1. X^* vérifie les conditions nécessaires de 1^{er} ordre .
2. $H_X L(X^*, \lambda^*)$ est semi-définie positive sur $T(X^*)$.

Théorème 1.7.3 (Condition suffisante de 2^{ème} ordre)[3]

On suppose que f et k_i ($i = 1, \dots, m$) sont de classe C^2 . Soit X^* un point régulier vérifie :

1. X^* vérifie les conditions nécessaires de 1^{er} ordre .
2. $H_X L(X^*, \lambda^*)$ est définie positive sur $T(X^*)$.

Alors X^* est un minimum local strict.

Exemple 1.6

Soit une unité de fabrication dans une usine, fabrique deux types de produit (a et b), tel que f la fonction de fabrication :

$$\begin{aligned} f : (\mathbb{R}_+^*)^2 &\longrightarrow \mathbb{R} \\ (a, b) &\longrightarrow 5a^2 + 5b^2 - 2ab - 2a - 1000. \end{aligned}$$

Et vente : $\begin{cases} a \text{ avec 1 unité monétaire.} \\ b \text{ avec 3 unités monétaires.} \end{cases}$

Ce qui donne la fonction des gains représentée par :

$$\begin{aligned} g : (\mathbb{R}_+^*)^2 &\longrightarrow \mathbb{R} \\ (a, b) &\longrightarrow a + 3b - f(a, b) = -5a^2 - 5b^2 + 2ab + 3a + 3b + 1000. \end{aligned}$$

le but c'est de savoir s'il est envisageable de travailler à plein régime.

On calcule le gradient et la hessienne de $g(a, b)$:

$$\nabla g(a, b) = \begin{pmatrix} -10a + 2b + 3 \\ 2a - 10b + 3 \end{pmatrix} ; \quad A = Hg(a, b) = \begin{pmatrix} -10 & 2 \\ 2 & -10 \end{pmatrix}$$

On a d'après le critère de Sylvester :

$$\begin{cases} \Delta_1 = (-1)^1 \det A_1 = 10 > 0 \\ \text{et} \\ \Delta_2 = (-1)^2 \det A_2 = 96 > 0 \end{cases} \Rightarrow Hg(a, b) \text{ est définie négative.}$$

Donc $g(a, b)$ est strictement concave.

à plein régime : On veut maximiser les gains de l'usine sur :

$$\omega \subseteq (\mathbb{R}_+^*)^2 = \{(a, b) \in (\mathbb{R}_+^*)^2 / a + b = 20\}$$

Tel que :

$$(P) : \max_{(a,b) \in \omega} g(a, b)$$

Comme $g(a, b)$ est quadratique car $(A = Hg(a, b))$ est symétrique $A = A^T$, et strictement concave, donc g est coersive.

Et comme ω est fermé est de dimension finie, donc (P) admet une unique solution qui vérifie les conditions de Lagrange tel que :

$$L((a, b), \lambda) = g(a, b) + \lambda(a + b - 20).$$

$$\begin{cases} \nabla_a L(a^*, b^*, \lambda^*) = -10a^* + 2b^* + 3 + \lambda^* = 0. \\ \nabla_b L(a^*, b^*, \lambda^*) = 2a^* + -10b^* + 3 + \lambda^* = 0 \\ \nabla_\lambda L(a^*, b^*, \lambda^*) = a^* + b^* - 20 = 0. \end{cases} \Rightarrow a^* = 10; \quad b^* = 10; \quad \lambda^* = 77.$$

$$g(a^*, b^*) = 260$$

1.7.2 Optimisation avec contraintes d'inégalités [1]

Soit le problème (1.3) avec $\mathbb{K} = \{X \in \mathbb{R}^n : g_i(X) \leq 0, i = 1, \dots, m\}$.

Définition 1.29

On définit le problème d'optimisation avec contraintes d'inégalités par :

$$(P) : \begin{cases} \min f(X) \\ g_i(X) \leq 0 \quad (i = 1, \dots, m) \\ X \in \mathbb{R}^n \end{cases}$$

Définition 1.30

• Les contraintes $g_i(X) \leq 0$ sont dites **actives** (ou saturées) en X^* si $g_i(X^*) = 0$, et on note l'ensemble des indices des contraintes actives en X^* par : $E(X^*) = \{i \in \{1, \dots, m\} / g_i(X^*) = 0\}$.

• On dit qu'elles sont **inactives** si $g_i(X^*) < 0$.

• Un point X^* admissible est dit régulier si la condition de **FIACCO-**

Mc CORMICK est vérifiée :

$$rgJ(G_{E(X^*)}(X^*)) = \text{card}E(X^*) = |E(X^*)|.$$

Avec $G_{E(X^*)}$ la matrice associée pour les contraintes actives.

Propositions :

1. Si les fonctions $g_i(X)$ $i = 1, \dots, m$ sont linéaires alors la qualification des contraintes est vérifiée par tout (**KARLIN 1959**).
2. Si les fonctions $g_i(X)$ $i = 1, \dots, m$ sont convexes et il existe X^* tel que : $g_i(X^*) < 0$, $i = 1, \dots, m$, alors la qualification des contraintes est vérifiée par tout **SLATER 1950**.

Théorème 1.7.4 Karush-Kuhn-Tucker K.K.T.

(Conditions nécessaire de 1^{er} ordre) :[1]

On suppose que les fonctions f, g_i avec $(i = 1, \dots, m)$ sont de classe C^1 et que l'hypothèse de la qualification des contraintes est vérifiée en $X^* \in \mathbb{K}$. Alors une condition nécessaire pour que X^* soit un optimum local de (P) et qu'il existe des nombres $\mu_i \geq 0$ ($i = 1, \dots, m$) appelés multiplicateurs de K.K.T, tels que

$$\begin{cases} \nabla f(X^*) + \sum_{i=1}^m \mu_i^* \nabla g_i(X^*) = 0. \\ \mu_i^* g_i(X^*) = 0, \quad (i = 1, \dots, m) \end{cases} \quad (1.5)$$

Preuve :[1]

Remarque :

Si les fonctions f et g_i sont convexes, alors le problème (\mathcal{P}) est convexe \Rightarrow la condition nécessaire de K.K.T est aussi suffisante pour un minimum global.

Théorème 1.7.5 Condition nécessaire de 2^{ème} ordre :[5]

On suppose que les fonctions f et g_i sont de classe C^2 , et soit X^* un minimum régulier de problème (\mathcal{P}) , alors :

1. Les conditions de K.K.T (1.5).

2. $H_X L(X^*, \mu^*)$ est semi-définie positive (S.D.P) sur $T(X^*)$.

Remarque :

$T(X^*) = \{d \in \mathbb{R}^n / JG(X^*)d = 0\}$ est appelé le sous espace tangent des contraintes au point X^* .

Théorème 1.7.6 Condition suffisante de 2^{ème} ordre :

On suppose que les fonctions f et g_i sont de classe C^2 , et soit X^* un point régulier de problème (\mathcal{P}) vérifie :

1. Les conditions de K.K.T (1.5).

2. $H_X L(X^*, \mu^*)$ est définie positive (D.P) sur $T(X^*)$.

Alors X^* est un minimum local strict de problème (\mathcal{P}) .

Exemple 1.7 Dans **Exemple 1.6**, on modifie l'ensemble des contraintes par :

$$\omega \subseteq (\mathbb{R}_+^*)^2 = \{(a, b) \in (\mathbb{R}_+^*)^2 / a + b \leq 20\}$$

On aura :

$$L((a, b), \mu) = g(a, b) + \mu(a + b - 20).$$

Les conditions de K.K.T :

$$\begin{cases} \nabla_a L(a^*, b^*, \mu^*) = -10a^* + 2b^* + 3 + \mu^* = 0. \\ \nabla_b L(a^*, b^*, \mu^*) = 2a^* + -10b^* + 3 + \mu^* = 0 \\ \nabla_\mu L(a^*, b^*, \mu^*) = a^* + b^* \leq 20 \\ \mu(a^* + b^* - 20) = 0 \end{cases}$$

Pour la contrainte active : $\mu > 0$

$$(a^*, b^*) = (10, 10), \mu = 77 \Rightarrow g(a^*, b^*) = 260$$

Pour la contrainte inactive : $\mu = 0$

$$(a^*, b^*) = \left(\frac{3}{8}, \frac{3}{8}\right), \mu = 0 \Rightarrow g(a^*, b^*) = 1001, 125$$

il vaut mieux de travailler avec ce cas pour maximiser le bénéfice.

1.7.3 Optimisation avec contraintes mixtes

Soit le problème (1.3) avec $\mathbb{K} = \left\{ X \in \mathbb{R}^n : \begin{array}{l} g_i(X) \leq 0, \ i = 1, \dots, m \\ k_j(X) = 0, \ j = 1, \dots, p \end{array} \right\}$

Définition 1.31

On définit un problème mixte par :

$$(P) \quad \begin{cases} \min f(X) \\ g_i(X) \leq 0 \quad i = 1, \dots, m \\ k_j(X) = 0 \quad j = 1, \dots, p \\ X \in \mathbb{R}^n. \end{cases}$$

Définition 1.32

Un point X^* est dit régulier si la condition de **FIACCO-Mc COR-MICK** est vérifiée, les vecteurs $\nabla_{E(X^*)} g_i(X^*)$ et $\nabla k_j(X^*)$ sont linéairement indépendants :

$$rg \begin{pmatrix} JG_{E(X^*)}(X^*) \\ JK(X^*) \end{pmatrix} = |E(X^*)| + p$$

Exemple 1.8

Soit les contraintes suivantes :

$$\begin{cases} x_1^2 + x_2 \leq 2 \Rightarrow g_1(X) = x_1^2 + x_2 - 2 \leq 0 \\ x_1^2 + x_2^2 = 2 \Rightarrow k_1(X) = x_1^2 + x_2^2 - 2 = 0 \\ X = (x_1, x_2) \in \mathbb{R}^2 \end{cases}$$

Le point $X^* = (1, 1)$ est admissible, car la contrainte $g_1(X)$ est active en X^* .

$E(X^*) = \{1\}$ car $g_1(X^*) = 0$.

Et comme :

$$rg \begin{pmatrix} JG_{E(X^*)}(X^*) \\ JK(X^*) \end{pmatrix} = rg \begin{pmatrix} 2 & 1 \\ 2 & 2 \end{pmatrix} = 1 + 1 = 2$$

Donc X^* est régulier.

Propositions :

- 1) Si les fonctions $g_i(i = 1, \dots, m)$, $k_j(j = 1, \dots, p)$ sont linéaires, alors tous les points sont réguliers (**KARLIN**).
- 2) Si les fonctions $g_i(i = 1, \dots, m)$ sont convexes, les $k_j(j = 1, \dots, p)$ sont linéaires et $\exists X^* / g_i(X^*) < 0 (i = 1, \dots, m)$, $k_j(X^*) = 0 (j = 1, \dots, p)$ alors tous les points sont réguliers (**SLATER**).

Théorème 1.7.7 Karush-Kuhn-Tucker K.K.T.

(Conditions nécessaire de 1^{er} ordre) : [4] On suppose que les fonctions f , $g_i(i = 1, \dots, m)$, $k_j(j = 1, \dots, p)$ sont de classe C^1 .

Soit X^* un minimum local régulier.

Alors il $\exists (\lambda_j^* \in \mathbb{R} (j = 1, \dots, p), \mu_i^* \in \mathbb{R}^+ (i = 1, \dots, m))$ tel que :

$$\begin{cases} \nabla f(X^*) + \sum_{j=1}^p \lambda_j^* \nabla k_j(X^*) + \sum_{i=1}^m \mu_i^* \nabla g_i(X^*) = 0. \\ k_j(X^*) = 0 (j = 1, \dots, p). \\ \mu_i^* g_i(X^*) = 0 (i = 1, \dots, m). \end{cases}$$

Remarque :

Si f est convexe, les fonctions $g_i, i = 1, \dots, m$ sont convexes et les fonctions $k_j, j = 1, \dots, p$ sont linéaires, et si la condition de Slater est vérifiée, alors la condition nécessaire du premier ordre de K.K.T devient suffisante et on obtient un minimum global .

Définition 1.33

On dit qu'une contrainte active est fortement active si le multiplicateur associe est strictement positif.

Théorème 1.7.8 Condition nécessaire de 2^{ème} ordre : [1]

Soient les fonctions f , $g_i(i = 1, \dots, m)$, $k_j(j = 1, \dots, p)$ sont de classe C^2 .

X^* un minimum local régulier, alors :

- X^* vérifie la condition nécessaire de 1^{er} ordre.
- $H_X L(X^*, \lambda^*, \mu^*)$ est semi-définie positive (S.D.P) sur $T(X^*)$, avec :

$$T(X^*) = \left\{ d \in \mathbb{R}^n / JG_{E(X^*)} d = 0; JK(X^*) d = 0 \right\}$$

Théorème 1.7.9 Condition suffisante de 2^{ème} ordre[1]

Soient les fonctions $f, g_i (i = 1, \dots, m), k_j (j = 1, \dots, p)$ sont de classe C^2 .

Soit X^* un point régulier, tel que :

- X^* vérifie la condition nécessaire de 1^{er} ordre.
- $H_X L(X^*, \lambda^*, \mu^*)$ est définie positive (D.P) sur $\bar{T}(X^*)$, avec :

$$\bar{T}(X^*) = \left\{ d \in \mathbb{R}^n / JG_{\bar{E}(X^*)} d = 0; JK(X^*)d = 0 \right\}$$

Alors X^* un minimum local.

$\bar{E}(X^*)$: l'ensemble des indices correspond à les contraintes fortement actives.

Application des moindres carrés

Introduction :

La méthode des moindres carrés, indépendamment élaborée par Legendre et Gauss au début de XIX^e siècle, permet de comparer des données expérimentales, généralement entachées d'erreurs de mesure, à un modèle mathématique peut prendre diverses formes. En effet cette méthode c'est un nuage de points $M_i(x_i, y_i)$ que l'on désire ajuster au mieux par une courbe mathématique (C) de type $y = f(x_1, x_2, \dots, x_n)$ dont on devra choisir le type de façon pertinente eu égard au phénomène étudié, en utilisant la régression linéaire ou non linéaire.

2.1 Régression linéaire

L'étude d'un phénomène peut, le plus souvent, être schématisé de la manière suivante : on s'intéresse à une grandeur b , que nous appellerons par la suite variable expliquée, qui dépend d'un certain nombre de variables v_1, v_2, \dots, v_n que nous appellerons variables explicatives. On cherche à mettre en évidence la liaison (relation fonctionnelle) pouvant exister entre la variable expliquée b et les variables explicatives v_1, v_2, \dots, v_n . On s'intéresse aux modèles dits linéaire, i.e. aux modèles du type :

$$b = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = \sum_{i=1}^n \alpha_i v_i$$

Les α_i sont des réels appelés coefficients du modèle.

2.1.1 Critère des moindres carrés :

On cherche donc un modèle qui nous permet d'obtenir un \hat{b} le plus proche possible de b . Pour cela on effectue m mesures ($m < n$) des variables v_1, v_2, \dots, v_n et de b . On cherche alors $\alpha_1, \alpha_2, \dots, \alpha_n$ tel que, $j = 1, \dots, m$:

$$\hat{b}_j = \alpha_1 v_{j,1} + \alpha_2 v_{j,2} + \dots + \alpha_n v_{j,n} = \sum_{i=1}^n \alpha_i v_{j,i}$$

soit le plus proche possible de b_j .

On utilise les notations matricielles, le système :

$$\begin{cases} \hat{b}_1 = \alpha_1 v_{1,1} + \alpha_2 v_{1,2} + \dots + \alpha_n v_{1,n} \\ \hat{b}_2 = \alpha_1 v_{2,1} + \alpha_2 v_{2,2} + \dots + \alpha_n v_{2,n} \\ \vdots \\ \hat{b}_m = \alpha_1 v_{m,1} + \alpha_2 v_{m,2} + \dots + \alpha_n v_{m,n} \end{cases}$$

s'écrit :

$$\begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_m \end{pmatrix} = \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,n} \\ v_{2,1} & v_{2,2} & \dots & v_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m,1} & v_{m,2} & \dots & v_{m,n} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

Ainsi on cherche $x = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$ tel que Ax soit le plus proche possible de b . On comprend alors que la notion de distance apparaît. On souhaite que la distance $d(\hat{b} = Ax, b)$ soit minimale, on rappelle que la distance euclidienne usuelle est définie comme suit :

$$\forall x, y \in \mathbb{R}^m, d(x, y) = \sqrt{\|x - y\|^2}$$

Où $\|\cdot\|$ est la norme euclidienne.

Ce qui s'écrit :

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 \quad (2.1)$$

2.1.2 Forme standard :

Définition 2.1 On appelle forme standard d'un problème de moindres carrés donnée par (2.1) tel que :

$$A = \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,n} \\ v_{2,1} & v_{2,2} & \dots & v_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m,1} & v_{m,2} & \dots & v_{m,n} \end{pmatrix} \in \mathbf{M}_{m,n}$$

A : c'est la matrice des données.

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$$

b : vecteur réponse.

$$x = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{R}^n$$

x : l'expression de critère.

2.1.3 Recherche d'une solution :

On fait l'hypothèse que les variables explicatives sont linéairement indépendantes (i.e. $rg(A) = n$). Soit $f(x) = \|Ax - b\|^2$ la fonction erreur. Selon la théorie d'optimisation, la valeur minimale se trouve par la résolution de l'équation $\nabla f(x) = 0$.

Théorème 2.1.1 [8]

Soit $A \in \mathbf{M}_{m,n}$ avec $m > n$ et $b \in \mathbb{R}^n$. Une condition nécessaire et suffisante pour que $x \in \mathbb{R}^n$ réalise (2.1) est que :

$$A^T A x = A^T b \quad (2.2)$$

Les équations (2.2) sont appelées équation normale elles correspondent aux conditions de 1^{er} ordre de $f(x)$. Elles sont obtenues en dérivant $f(x)$ par rapport au vecteur de paramètre x .

Ce système admet toujours au moins une solution. Si la matrice $A^T A$ est inversible i.e si $\text{rg}(A) = n$, alors la solution est unique.

2.2 Ajustement [9]

Définition 2.2 Nuage de points :

Dans un repère orthogonal, l'ensemble des points M_i de coordonnées (x_i, y_i) constitue le nuage de points associé à la série statistique à deux variables, la forme du nuage obtenu peut indiquer le type de dépendance possible entre X et Y .

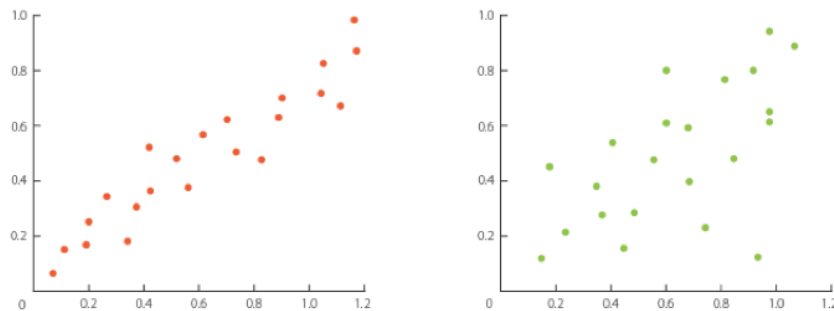


FIGURE 2.1 – Exemple de nuage de points

Définition 2.3 Point moyen :

Le point moyen d'un nuage de points est le point G de coordonnées (\bar{x}, \bar{y}) ou :

\bar{x} représente la moyenne des x_i : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

\bar{y} représente la moyenne des y_i : $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Définition 2.4

On appelle la covariance de X et de Y le nombre :

$$COV(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

On appelle la variance de X : $Var(X) = COV(X, X)$ tel que elle est utilisé pour le calcul de l'écart type :

$$\sigma(X) = \sqrt{Var(X)}.$$

2.2.1 Position de problème

On se donne deux séries $X = (x_i)_{i \in \mathbb{N}}$ et $Y = (y_i)_{i \in \mathbb{N}}$ ayant le même nombre d'élément $n \in \mathbb{N}$.

Le but est d'établir un lien si il existe entre X et Y .

Selon l'allure de nuage, on a envie de remplacer ce nuage par le graphe d'une fonction f . (la courbe de $Y = f(X)$) cette opération s'appelle un **ajustement**, La nature de l'ajustement dépend de la forme du nuage de points, on peut penser à beaucoup d'ajustement :

- **Ajustement linéaire (affine)** : $Y = aX + b$, a et $b \in \mathbb{R}$.

Proposition :

- **Ajustement exponentiel** : $Y = b \exp(aX)$, a et $b \in \mathbb{R}$, celui ci peut être transformé à un ajustement affine par un changement de variables $\ln(Y) = Y'$.

Preuve :

$$Y = b \exp(aX) \Rightarrow \ln(Y) = \ln(b \exp(aX)) \Rightarrow \ln(Y) = aX + \ln(b)$$

On posons $\ln(Y) = Y'$ et $\ln(b) = b' \in \mathbb{R}^+$

Donc $Y = b \exp(aX) \Leftrightarrow Y' = aX + b'$.

- **Ajustement logarithmique** : $Y = a \ln(X) + b$, a et $b \in \mathbb{R}$, ce que est peut transformé à un ajustement affine par un changement de variable $X' = \ln(X)$.

Proposition :

• **Ajustement puissance :** $Y = bX^a$, a et $b \in \mathbb{R}$, celui ci peut être transformé à un ajustement affine par un changement de variables $Y' = \ln(Y)$ et $X' = \ln(X)$.

Preuve :

$$Y = bX^a \Rightarrow \ln(Y) = \ln(bX^a) \Rightarrow \ln(Y) = \ln(b) + a \ln(X).$$

On posons $Y' = \ln(Y)$ et $X' = \ln(X)$ et $b' = \ln(b) \in \mathbb{R}^+$
donc $Y = bX^a \Leftrightarrow Y' = aX' + b'$.

2.2.2 Ajustement par la méthode des moindres carrés

Principe de la méthode :

Historiquement, l'idée est manuelle. On traçait le nuage de points et on faisait "pivoter" une règle qui passait au "mieux" entre les points du nuage. Il faut avoir conscience que l'on pourrait ajuster de diverses manières, en fonction du choix de la distance qu'on veut minimiser.

Soient $X = (x_i)_{i \in \mathbb{N}}$ et $Y = (y_i)_{i \in \mathbb{N}}$ deux séries. On note M_i les points du plan de coordonnées (x_i, y_i) .

Soient D une droite quelconque, d'équation $Y = aX + b$, $a, b \in \mathbb{R}$ et P_i la projection de M_i sur D parallèlement à (Oy) . Donc les coordonnées de P_i sont de la forme (x_i, \hat{y}_i) . Notez que $P_i \in D$ donc $\hat{y}_i = ax_i + b$.

Définition 2.5

Ajuster par la méthode des moindres carrés consiste à rechercher parmi toutes les droites D , celle qui rend minimale la somme des carrés des longueurs $P_i M_i$, ($i = 1, \dots, n$), i.e la quantité :

$$f(a, b) = \sum_{i=1}^n (P_i M_i)^2 = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

tel que $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^+$.

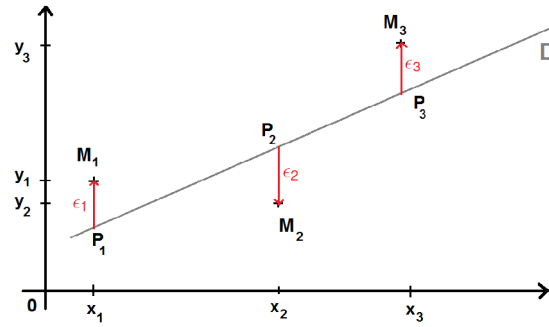


FIGURE 2.2 – Ajustement linéaire

Ainsi on cherche à déterminer la droite qui rend f minimale (**problème de minimisation sans contrainte**) :

$$\min_{(a,b) \in \mathbb{R}^2} f(a,b) \quad (2.3)$$

Détermination de minimum :

Soit $(M_1, \dots, M_n) = ((x_1, y_1), \dots, (x_n, y_n))$ points donnés.

On cherche une droite $Y = aX + b$ telle que $\sum_{i=1}^n (y_i - (ax_i + b))^2$ est minimale. On cherche donc (2.3).

Pour déterminer les points critiques de f on applique la condition nécessaire de 1^{er} ordre :

$$\nabla f(a,b) = 0 \Rightarrow \begin{cases} \frac{\partial f}{\partial a} = -2 \sum_{i=1}^n x_i (y_i - (ax_i + b)) = 0 \\ \frac{\partial f}{\partial b} = -2 \sum_{i=1}^n (y_i - (ax_i + b)) = 0 \end{cases}$$

D'où

$$\begin{cases} \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - b \sum_{i=1}^n x_i = 0 & (1) \\ \sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - nb = 0 & (2) \end{cases}$$

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{COV(X,Y)}{Var(X)}.$$

\Rightarrow

$$b = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}.$$

Et de (2), on tire que $b = \bar{y} - a\bar{x}$.

On a donc un point critique (a, b) , reste à vérifier c'est bien un extrémum, on applique la condition nécessaire de 2^{ème} ordre :

$$Hf(a, b) = \begin{pmatrix} 2 \sum_{i=1}^n x_i^2 & 2 \sum_{i=1}^n x_i \\ 2 \sum_{i=1}^n x_i & 2n \end{pmatrix}$$

On aura :

$$\begin{cases} tr(Hf(a, b)) = 2n + 2 \sum_{i=1}^n x_i^2 > 0 \\ \text{et} \\ \det(Hf(a, b)) = 4n \sum_{i=1}^n x_i^2 - 4 \left(\sum_{i=1}^n x_i \right)^2 = 4n^2 (Var(X)) > 0 \end{cases}$$

$\Rightarrow Hf(a, b)$ est définie positive, donc le (a, b) trouver est bien un minimum de f , et la droite $Y = aX + b$ minimise $\sum_{i=1}^n \varepsilon_i^2$.

Remarque :

- $\frac{\partial f}{\partial b} = 0$ donne que $\sum_{i=1}^n \varepsilon_i = 0$ la droite passe bien au milieu de nuage de points $M_i(x_i, y_i)$.

- Géométriquement toute ces manipulation, correspondant à une projection.

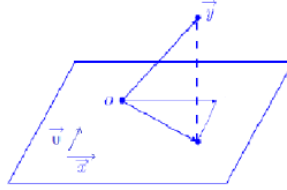


FIGURE 2.3 – projection de b sur $Im(a)$

Exemple 2.1

Soit le tableau suivant d'un échantillon d'individus en fonction de leurs tailles.

individu	1	2	3	4	5	6	7	8	9	10	11	12
taille X (cm)	145	150	153	158	160	163	165	171	175	177	183	187
poids Y (Kg)	50	52	58	60	61	65	62	68	76	75	80	90

TABLE 2.1 – Le poids d'individus en fonction de leurs tailles

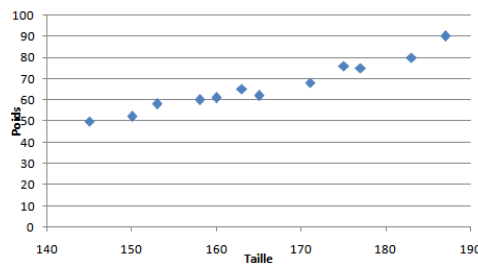


FIGURE 2.4 – Le nuage de points

On cherche à trouver une droite de régression qui permet d'ajuster ce nuage de points.

- Calcule de la moyenne \bar{x} et de \bar{y}

$$\bar{x} = \frac{145+150+153+158+160+163+165+171+175+177+183+187}{12} = 165.58,$$

$$\bar{y} = \frac{50+52+58+60+61+65+62+68+76+75+80+90}{12} = 66.42,$$

-Calculer la variance et la covariance à l'aide du tableau suivant :

x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
145	50	-20,583333	423,673611	-16,416667	337,909722
150	52	-15,583333	242,840278	-14,416667	224,659722
153	58	-12,583333	158,340278	-8,416667	105,909722
158	60	-7,583333	57,5069444	-6,416667	48,6597222
160	61	-5,583333	31,1736111	-5,416667	30,2430556
163	65	-2,583333	6,67361111	-1,416667	3,65972222
165	62	-0,583333	0,34027778	-4,416667	2,57638889
171	68	5,41666667	29,3402778	1,58333333	8,57638889
175	76	9,41666667	88,6736111	9,58333333	90,2430556
177	75	11,4166667	130,340278	8,58333333	97,9930556
183	80	17,4166667	303,340278	13,5833333	236,576389
187	90	21,4166667	458,673611	23,5833333	505,076389

TABLE 2.2 – Calcul de la covariance et de la variance

$$\Rightarrow \left\{ \begin{array}{l} COV(X, Y) = \frac{\sum_{i=1}^{12} (x_i - \bar{x})(y_i - \bar{y})}{12} = 141,0069444. \\ Var(X) = \frac{\sum_{i=1}^{12} (x_i - \bar{x})^2}{12} = 160,9097222. \end{array} \right.$$

Donc les coefficients a et b de la droite de régression (D) sont :

$$a = \frac{COV(X, Y)}{Var(X)} = \frac{141,0069444}{160,9097222} = 0,876310906.$$

$$b = \bar{y} - a\bar{x} = -78,68581416.$$

$$\Rightarrow (D) : Y = 0.876310906X - 78.68581416.$$

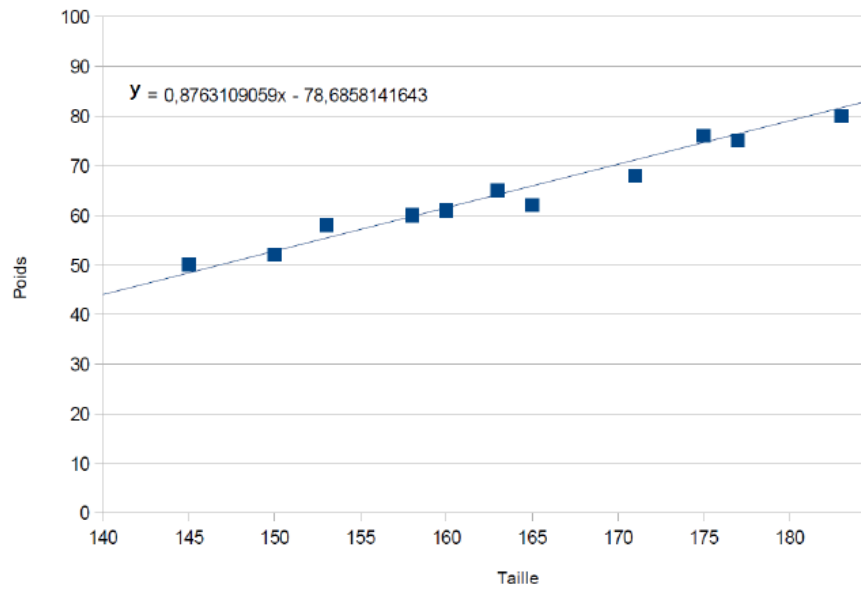


FIGURE 2.5 – Le poids d’individus en fonction de taille

2.3 Corrélation[9]

2.3.1 position de problème

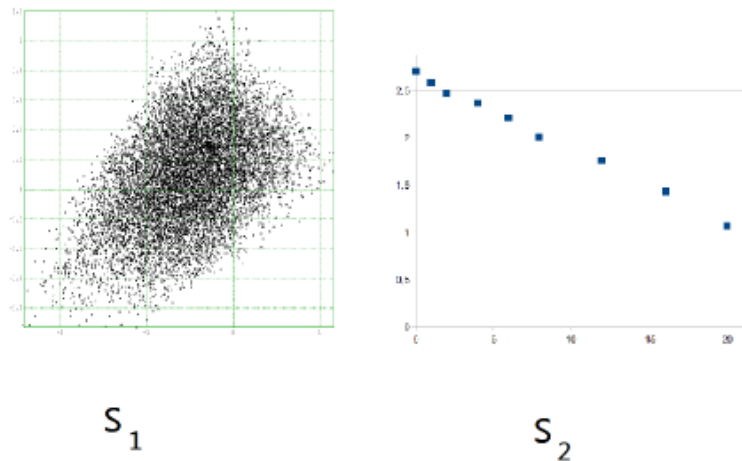


FIGURE 2.6 – S_1 , S_2 : Nuage de points

Ajuster linéairement S_2 semble réaliste alors que ajuster linéairement S_1 semble illusoire. pourtant tout les calculs peuvent être pratiqués... les résultats seront de mauvaise "qualité".

2.3.2 Droites de régression

Reprenons l'exemple 1.1.1 : On avait exprimé Y en fonction de X et on avait trouvé : $Y = 0.876X - 78.7$

On dit qu'on a effectué **une régression de Y en fonction de X** .

On aurait très bien pu effectué une régression de X en fonction de Y .

Intervertissons les rôles de X et Y . On refait l'ajustement linéaire. On obtient un a' et un b' donné par :

$$\begin{cases} a' = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n y_i^2 - n\bar{y}^2} \\ b' = \bar{x} - a'\bar{y} \end{cases}$$

La droite (D') de régression de X par rapport à Y a donc pour équation $x = a'y + b'$ c'est-à-dire : $Y = \frac{1}{a'}X - \frac{b'}{a'}$.

Dans notre exemple 1.1.1 , on trouve :

$$\begin{cases} (D) : Y = 0.876X - 78.7 \\ (D') : X = 1.092Y + 93.1, \text{ ie } : Y = 0.916X - 85.2 \end{cases}$$

Telle que (D) minimise la somme des écarts au carrés verticaux alors que (D') minimise la somme des écarts au carrés horizontaux.

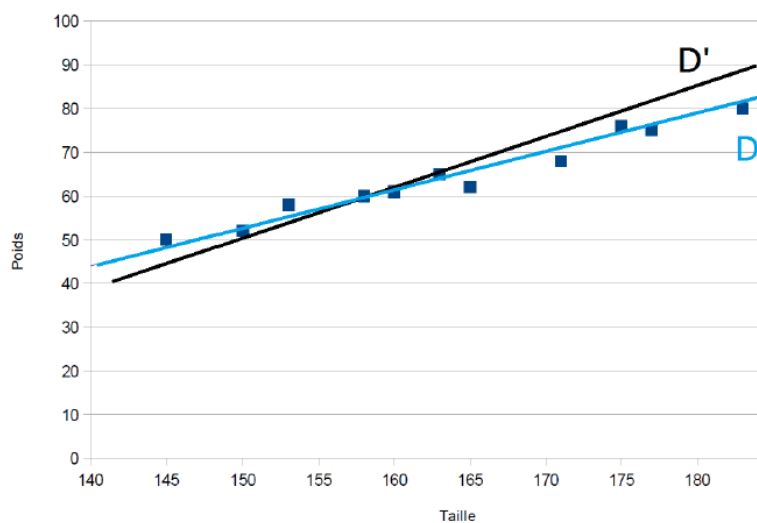


FIGURE 2.7 – Les droites de régressions (D) et (D')

Proposition :[9]

- (D) et (D') sont d'autant plus distinctes que le nuage de point est dispersé.

- Plus le nuage de points est "aligné", plus les deux droites de régression sont "proches" (les variables aléatoires X et Y sont dépendantes).

Remarque :

Si le nuage est parfaitement aligné, on a (D) et (D') sont confondues. Ainsi, les deux droites ont les mêmes équations, ce qui se traduit par :

$$a = \frac{1}{a'}, \text{ et } b = -\frac{b'}{a'} \Rightarrow aa' = 1.$$

2.3.3 Coefficient de corrélation

Ce qui précède, incite donc à considérer le nombre aa' , puis à le comparer à 1.

On dit que a et a' sont de même signe car :

$$a = \frac{COV(X,Y)}{Var(X)} \text{ et } a' = \frac{COV(X,Y)}{Var(Y)}.$$

Définition 2.6

On appelle coefficient de corrélation le nombre R tel que $R^2 = aa'$ et que R soit du signe de a et a' .

- Si a et a' positive, alors $R = \sqrt{aa'}$.
- Si a et a' négative, alors $R = -\sqrt{aa'}$.

Donc :

$$\text{On peut retenir } R(X,Y) = \frac{COV(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{COV(X,Y)}{\sigma(X)\sigma(Y)} \in [-1,1].$$

Avantage de coefficient de corrélation

On utilise le coefficient de corrélation pour prendre le meilleur ajustement correspondant a notre série de deux variable X et Y , tel que on calcul :

- $R(X,Y)$ dans le cas linéaire.
- $R(X,\ln(Y))$ dans le cas exponentiel.

- $R(\ln(X), \ln(Y))$ dans le cas puissance.
- $R(\ln(X), Y)$ dans le cas logarithme.

puis en choisir le meilleur ajustement i.e l'ajustement qui a un $R \approx |1|$.

Exemple 2.2

On a relevé le bénéfice en milliers d'euros sur 24 mois :

Mois X	0	1	2	4	6	8	12	16	20	24
Bénéfice Y	15	13.3	11.8	10.7	9.1	7.4	5.8	4.2	2.9	2

- 1) lorsque on fait un ajustement linéaire pour cette série, on remarque qu'il semble peu adapté.

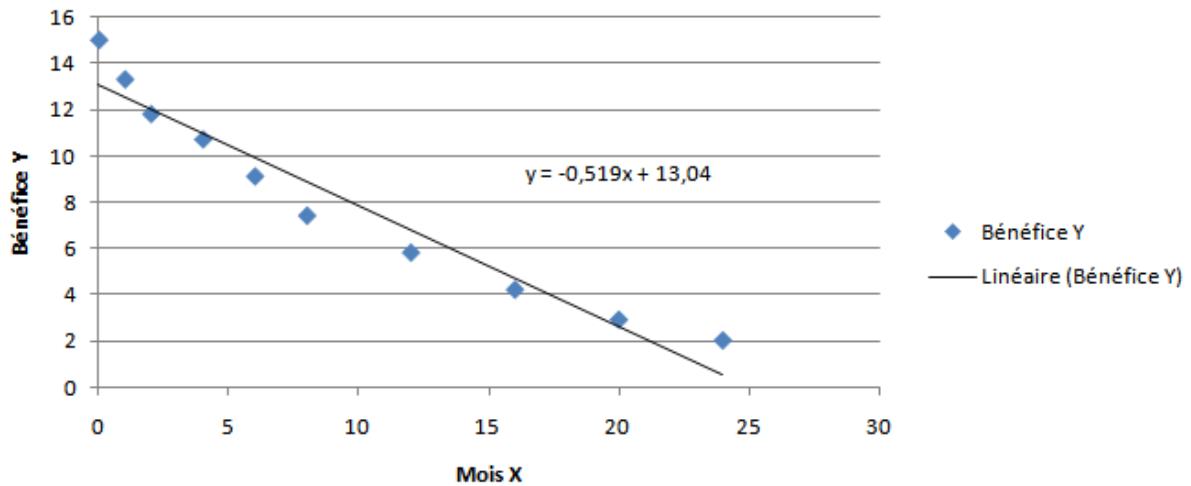


FIGURE 2.8 – Ajustement affine

- 2) Si on fait un ajustement exponentiel de type :

$$Y = b \exp(aX), \quad a, b \in \mathbb{R}. \quad (2.4)$$

Tel que pour déterminer les coefficients a, b , on fait un changement de variable :

$\ln(Y) = Y', \ln(b) = b' \Rightarrow (2.4)$ devenir sous forme d'un ajustement linéaire :

$$Y' = aX + b', \quad a \in \mathbb{R}, b \in \mathbb{R}_+ \quad (2.5)$$

On refaire les mêmes étapes de l'exemple précédent on aura :

$$Y' = -0,081X + 2,683.$$

Donc :

$$Y = 14.62 \exp(-0.081X).$$

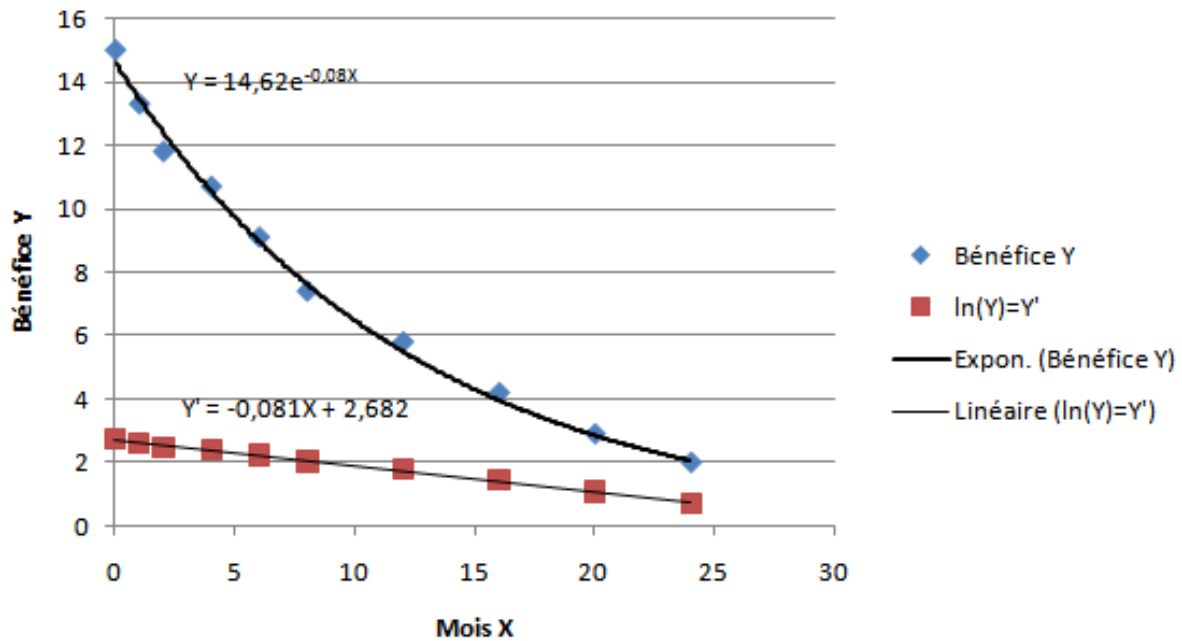


FIGURE 2.9 – L'ajustement affine et l'ajustement exponentielle

-Calcul de coefficient de corrélation :

Dans 1) :

$$R = \frac{COV(X, Y)}{\sigma(X)\sigma(Y)} = \frac{-32.816}{(7,950471684)(4,256712346)} = -0,969657687$$

Dans 2) :

$$R' = \frac{COV(X, Y')}{\sigma(X)\sigma(Y')} = \frac{-5,11890052}{(7,950471684)(0,644750275)} = -0,998601602$$

Donc $-1 \leq R' \leq R$, on conclue que le meilleur ajustement pour cette série c'est l'ajustement exponentiel.

Conclusion :

Dans ce chapitre, on s'est intéressé à l'étude des problèmes de dépendances qui entrent dans la théorie de corrélation, cette notion permet de fabriquer un nombre R qui mesure cette dépendance, et trouve la meilleure relation qui lie les deux grandeurs X et Y .

Simulation numérique avec MATLAB

Dans ce chapitre, on va traiter quelques exemples de l'ajustement par la méthode des moindres carrés sous logiciel MATLAB version 7.8.0.347(R2009a).

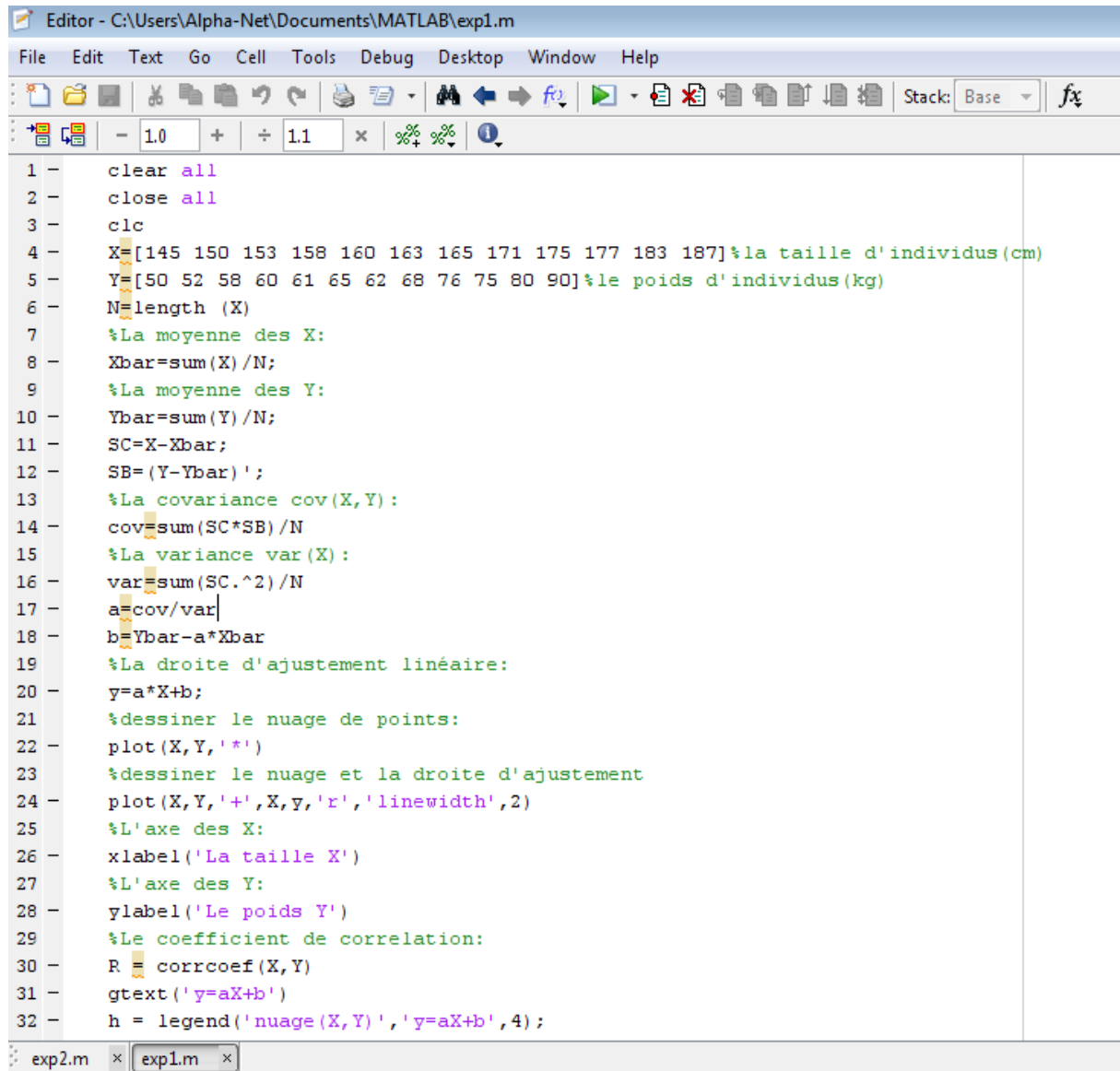
Exemple 3.1

Soit le tableau suivant d'un échantillon d'individus en fonction de leurs tailles.

<i>individu</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>
<i>taille X (cm)</i>	<i>145</i>	<i>150</i>	<i>153</i>	<i>158</i>	<i>160</i>	<i>163</i>	<i>165</i>	<i>171</i>	<i>175</i>	<i>177</i>	<i>183</i>	<i>187</i>
<i>poids Y (Kg)</i>	<i>50</i>	<i>52</i>	<i>58</i>	<i>60</i>	<i>61</i>	<i>65</i>	<i>62</i>	<i>68</i>	<i>76</i>	<i>75</i>	<i>80</i>	<i>90</i>

TABLE 3.1 – Le poids d'individus en fonction de leurs tailles

On cherche à trouver une droite de régression qui permet d'ajuster ce nuage de points.



```

1 - clear all
2 - close all
3 - clc
4 - X=[145 150 153 158 160 163 165 171 175 177 183 187]%la taille d'individus(cm)
5 - Y=[50 52 58 60 61 65 62 68 76 75 80 90]%le poids d'individus(kg)
6 - N=length (X)
7 - %La moyenne des X:
8 - Xbar=sum(X)/N;
9 - %La moyenne des Y:
10 - Ybar=sum(Y)/N;
11 - SC=X-Xbar;
12 - SB=(Y-Ybar)';
13 - %La covariance cov(X,Y):
14 - cov=sum(SC*SB)/N
15 - %La variance var(X):
16 - var=sum(SC.^2)/N
17 - a=cov/var
18 - b=Ybar-a*Xbar
19 - %La droite d'ajustement linéaire:
20 - y=a*X+b;
21 - %dessiner le nuage de points:
22 - plot(X,Y,'*')
23 - %dessiner le nuage et la droite d'ajustement
24 - plot(X,Y,'+',X,y,'r','linewidth',2)
25 - %L'axe des X:
26 - xlabel('La taille X')
27 - %L'axe des Y:
28 - ylabel('Le poids Y')
29 - %Le coefficient de correlation:
30 - R = corrcoef(X,Y)
31 - gtext('y=aX+b')
32 - h = legend('nuage (X,Y)', 'y=aX+b',4);

```

FIGURE 3.1 – Programme matlab 01

```

MATLAB 7.8.0 (R2009a)
File Edit Debug Parallel Desktop Window Help
Current Directory: C:\Users\Alpha-Net\Documents\MATLAB
Shortcuts How to Add What's New

X =
    145    150    153    158    160    163    165    171    175    177    183    187

Y =
    50    52    58    60    61    65    62    68    76    75    80    90

N =
    12

cov =
    141.0069

var =
    160.9097

a =
    0.8763

b =

b =
   -78.6858

R =
    1.0000    0.9784
    0.9784    1.0000

fx >>
Start
  
```

FIGURE 3.2 – Résultat 01

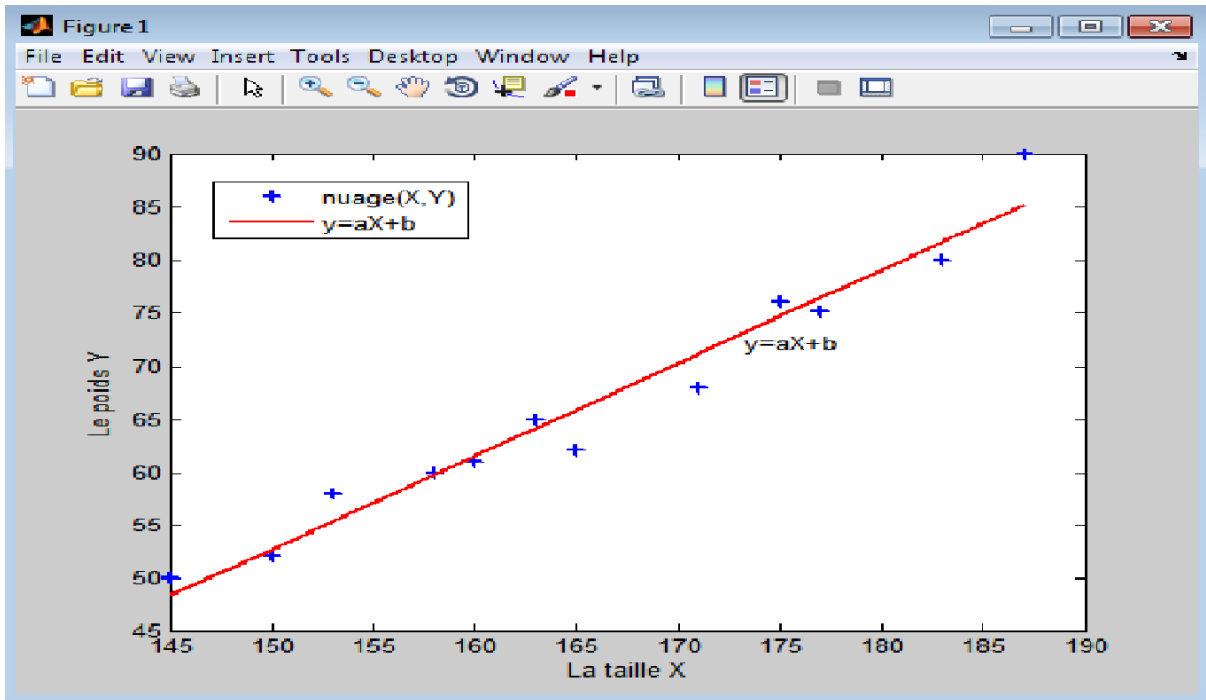


FIGURE 3.3 – L'ajustement de programme 01

La droite d'ajustement est : $y = 0.8763x - 78.6858$ tel que le coefficient de corrélation $R = 0.9784$ est très proche de 1.

Exemple 3.2 soit le tableau suivant de rendement de maïs Y (en quintal) en fonction de la quantité d'engrais X (en Kg) sur des parcelles de terrains similaires.

La quantité d'engrais(Kg) X	20	24	28	22	32	28	32	36	41	41
Le rendement de maïs (Qtl) Y	16	18	23	24	28	29	26	31	32	34

Dans cet exemple on cherche à trouver une dépendance linéaire pour ce nuage de points $M(X_i, Y_i)$.

```

1 - clear all
2 - close all
3 - clc
4 - X=[20 24 28 22 32 28 32 36 41 41]%la quantité d'engrais(Kg)
5 - Y=[16 18 23 24 28 29 26 31 32 34]%le rendement de maïs (Qt1)
6 - N=length(X)
7 - %La moyenne des X:
8 - Xbar=sum(X)/N;
9 - %La moyenne des Y:
10 - Ybar=sum(Y)/N;
11 - SC=X-Xbar;
12 - SB=(Y-Ybar)';
13 - %La covariance cov(X,Y):
14 - cov=sum(SC*SB)/N
15 - %La variance var(X):
16 - var=sum(SC.^2)/N
17 - a=cov/var
18 - b=Ybar-a*Xbar
19 - %La droite d'ajustement linéaire:
20 - y=a*X+b;
21 - %dessiner le nuage de points:
22 - plot(X,Y,'*')
23 - %dessiner le nuage et la droite d'ajustement
24 - plot(X,Y,'+',X,y,'r','linewidth',2)
25 - %L'axe des X:
26 - xlabel('La quantité d'engrais(Kg)')
27 - %L'axe des Y:
28 - ylabel('Le rendement de maïs (Qt1)')
29 - %Le coefficient de correlation:
30 - R = corrcoef(X,Y)
31 - gtext('y=aX+b')
32 - h = legend('nuage (X,Y)', 'y=aX+b',2);

```

FIGURE 3.4 – Programme matlab 02

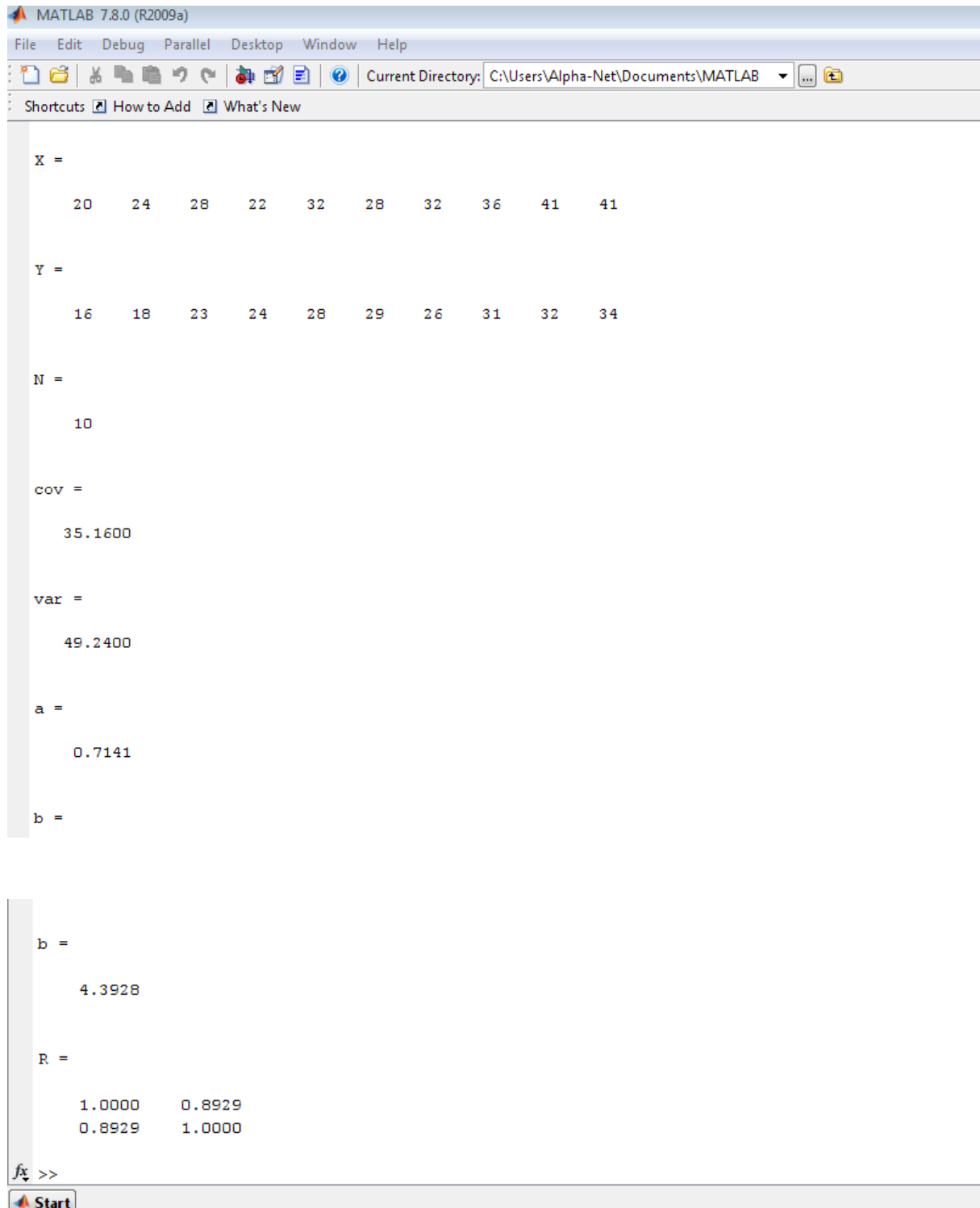


FIGURE 3.5 – Résultat 02

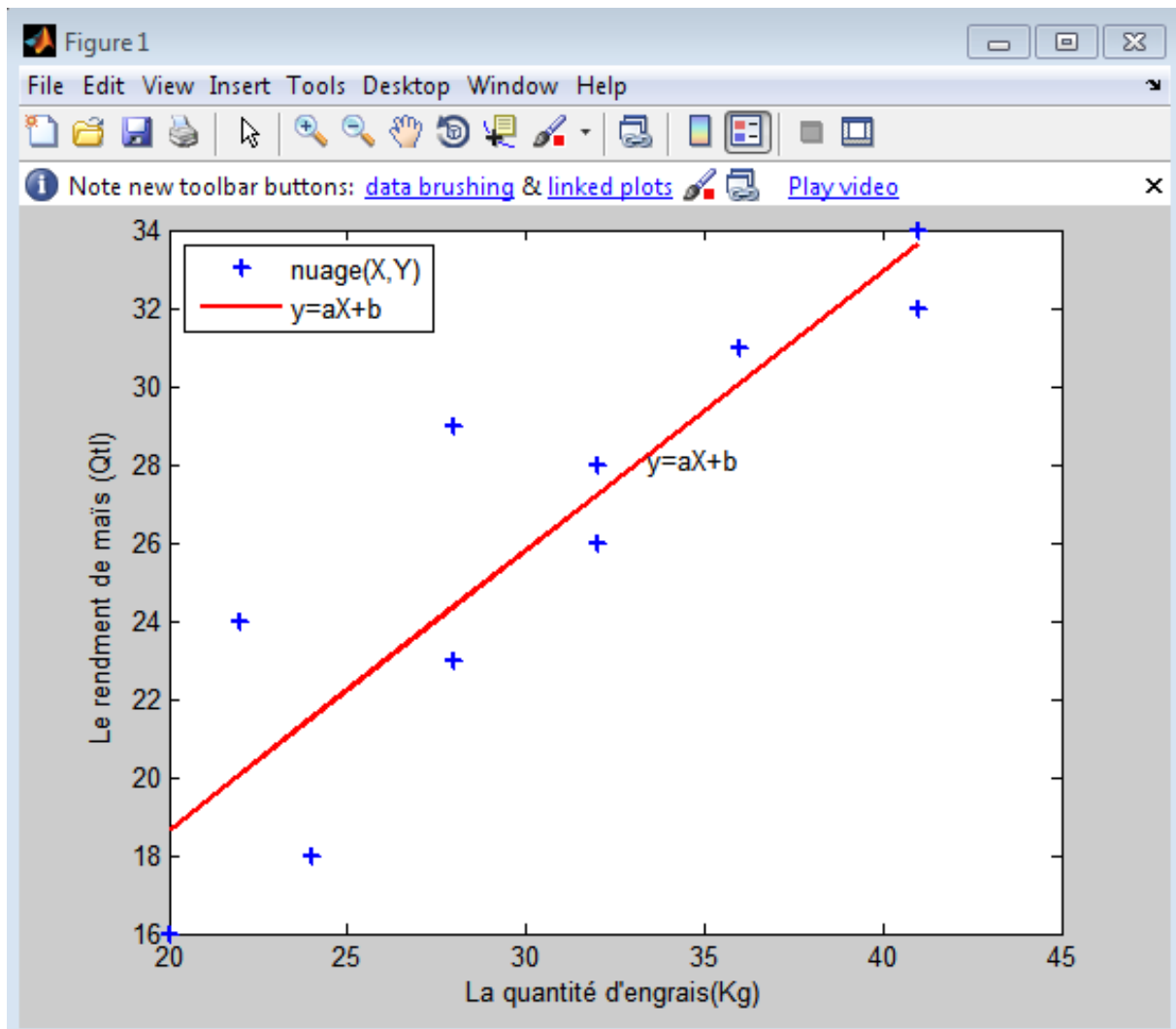


FIGURE 3.6 – L'ajustement linéaire

On conclue que la droite d'ajustement est : $y = 0.7141x + 4.3928$, avec le coefficient de corrélation $R = 0.8929$ n'est pas complètement linéaire.

Exemple 3.3

On a relevé le bénéfice en milliers d'euros sur 24 mois :

Mois X	0	1	2	4	6	8	12	16	20	24
Bénéfice Y	15	13.3	11.8	10.7	9.1	7.4	5.8	4.2	2.9	2

Dans cet exemple on cherche à trouver le meilleur ajustement correspondant à notre série statistique .

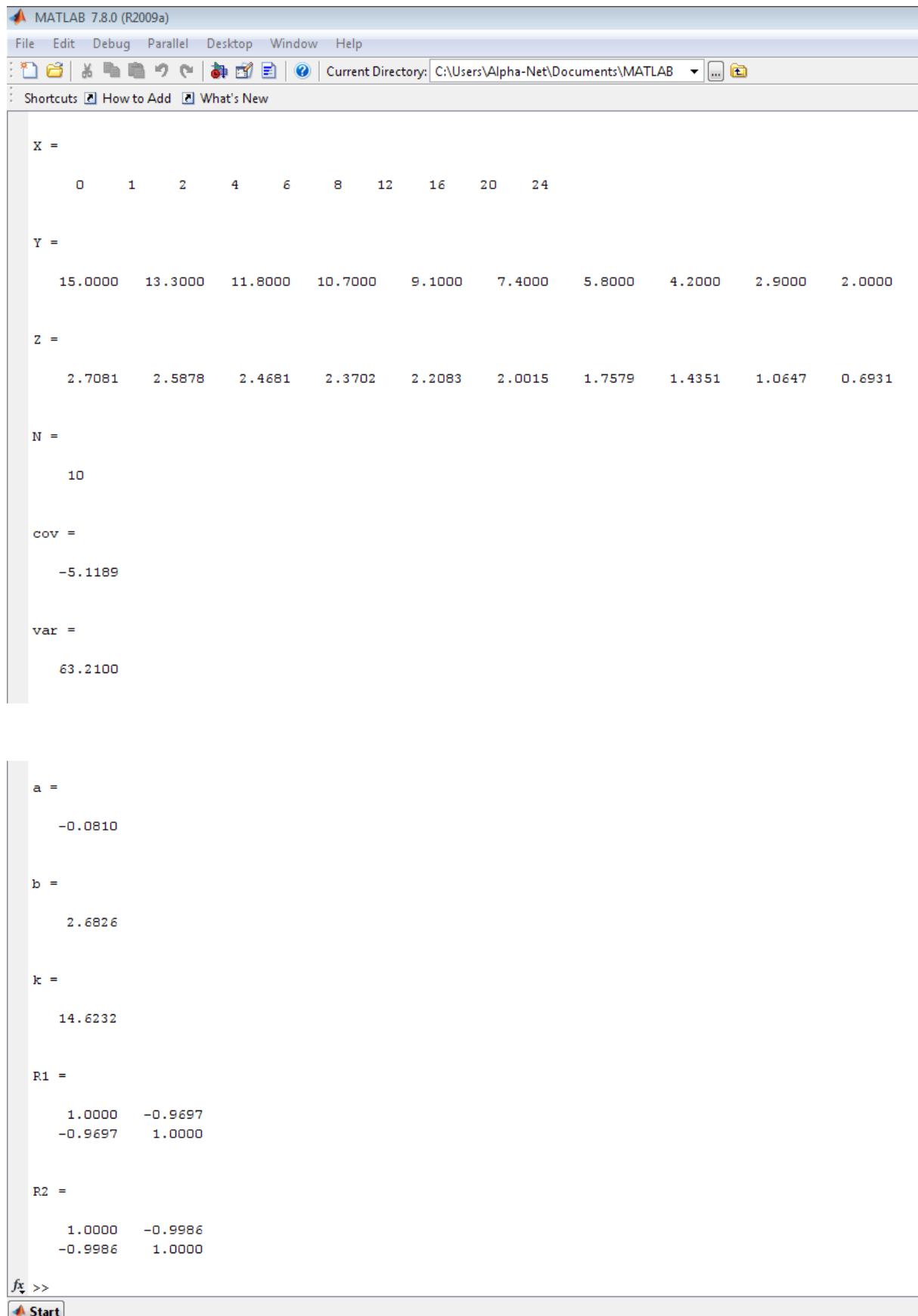


FIGURE 3.8 – Résultat 03

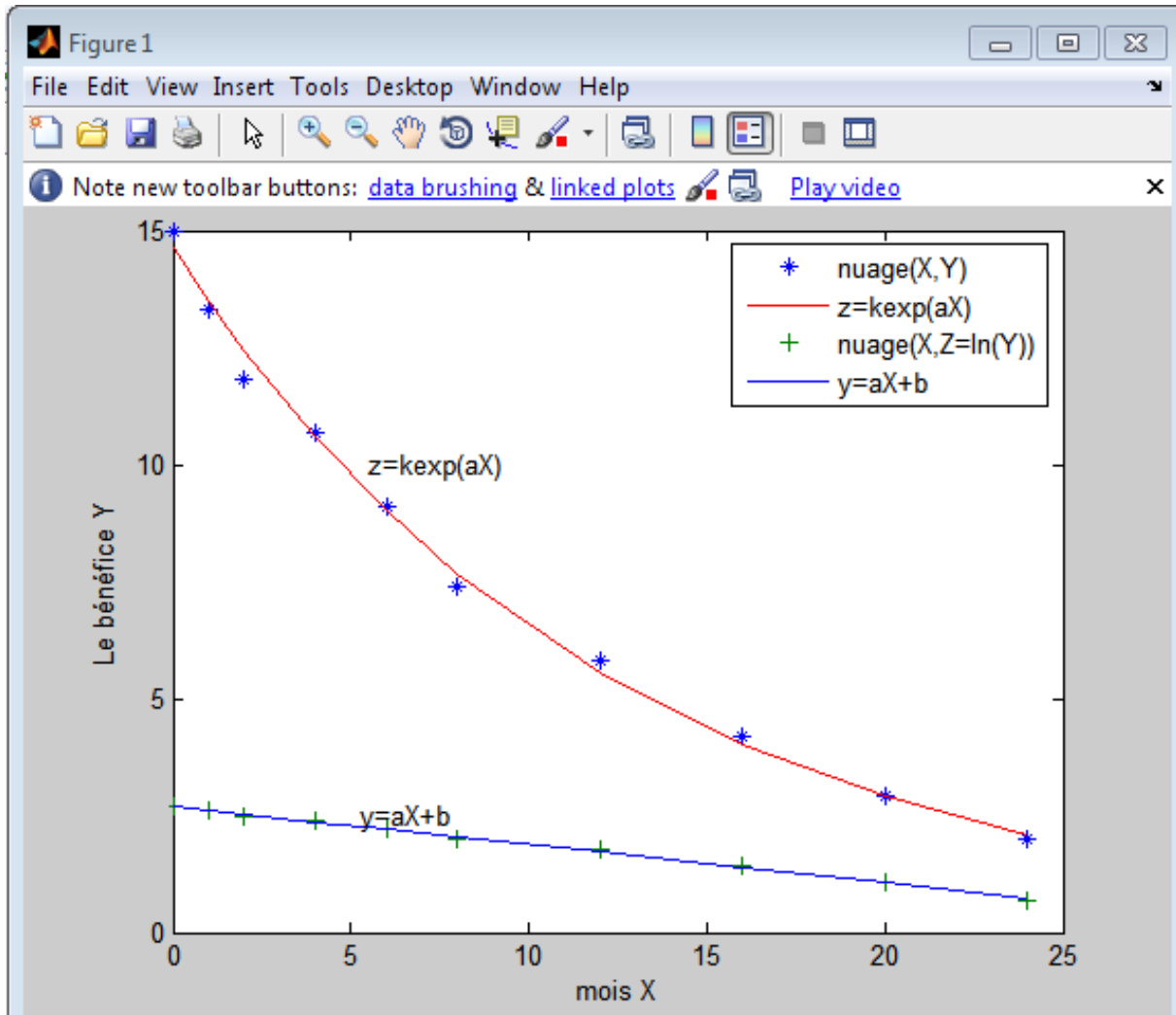


FIGURE 3.9 – L'ajustement linéaire et l'ajustement exponentielle

$$R_2 = -0,9986, \quad R_1 = -0,9697$$

Le meilleur ajustement c'est l'ajustement qui a un $R \approx |1|$.

Donc $-1 \leq R_2 \leq R_1$, on conclue que le meilleur ajustement pour cette série c'est l'ajustement exponentiel avec $z = 14.6232 \exp(-0.0810X)$.

Conclusion générale

Dans ce mémoire, on a étudié quelque problèmes d'optimisations non linéaires, sans contraintes, avec contraintes d'égalités et/ou d'inégalités ; les algorithmes de résolutions de ces problèmes.

Nous avons pour ce faire étudié et implémenté différentes approches de résolution.

Nous avons parlé sur la méthode des moindres carrés, cette méthode est utilisée pour résoudre plusieurs problèmes d'optimisations dans différents domaines (statistique, biologie, gestion, ...), elle permet de sélectionner la meilleure fonction qui reproduit le mieux les données expérimentales, on peut s'intéresser à des problèmes d'optimisations non convexes.

Pour montrer l'efficacité de l'application des moindres carrés, on a traité quelques exemples numériques sur le logiciel MATLAB, et on a conclue que le meilleur ajustement pour une dépendance donnée est celui qui a un coefficient de corrélation $|R| \approx 1$.

Bibliographie

- [1] Michel Minoux Programmation mathématique : Théorie et algorithme, 2^{ème} édition, Editions TEC & DOC, 2008.
- [2] J. Nocedal, and S. J. Wright, Numerical Optimization, Springer Series in Operations Research and Financial Engineering, second ed, 2006.
- [3] Jean-christophe CULIOLI Introduction à l'optimisation, 2^{ème} édition, ellipses 2012.
- [4] Eric Walter Méthodes Numériques & optimisation, Fluctuant Nec Merguntur, Paris. 2015.
- [5] Jacqueline Lelong-Ferrand et Jean-Marie Arnaudiès, Cours de Mathématiques,tome 2 : Analyse Dunod, 1976.
- [6] Laurent Schwartz, ANALYSE, Topologie générale et analyse fonctionnelle , France 1970.
- [7] Orban, D. Numerical Methods for Nonlinear Optimization and Optimal Control, notes du cours MTH8408, 2010.
- [8] David C.Lay, Algèbre linéaire et application.Université du Maryland-College Park,4^{ème} édition, 2004.
- [9] Clément.Rau Ajustement linéaire par les moindres carrés, Laboratoire de Mathématiques de Toulouse Université Paul Sabatier-IUT GEA Ponsan,,Module : Maths approfondies.

Les mots clés

- Optimisation.
- Optimisation non linéaire.
- Convexité.
- Concavité.
- Conditions d'optimalités.
- Karush-Kuhn-Tucker (KKT).
- Extrémum.
- Minimum local.
- Minimum global.
- Maximum local.
- Maximum global.
- Algorithme de gradient à pas optimal.
- Algorithme de gradient à pas fixe.
- Méthode d'Armijo.
- Algorithme de Newton.
- Moindre carré.
- Ajustement.
- Corrélation.
- Simulation de la méthode des moindre carrés.

Résumé

L'optimisation est une discipline mathématique et un outil d'aide à la décision, elle permet de trouver les valeurs des variables entrantes (les données) qui rendent la fonction objectif optimale.

On s'intéresse à l'étude de quelques problèmes d'optimisation non linéaire dont les données sont non linéaires et différentiables, les différents algorithmes de résolutions de ces problèmes et à l'application de la méthode des moindres carrés.

La méthode des moindres carrés consiste à considérer un nuage de points $M_i(x_i, y_i)$ que l'on désire ajuster au mieux par une courbe mathématique (c) de type $(y = f(x))$ dont on devra choisir le type de façon pertinente eu égard au phénomène étudié. On cherche les paramètres de f , fonction affine, polynôme, exponentielle, ... etc., minimisant la somme des carrés des distances entre y_i et $f(x_i)$. On peut toujours se ramener à un ajustement linéaire (affine) en effectuant un changement de variables.

Finalement, on a exécuté des simulations numériques de ces méthodes par le logiciel MATLAB .