

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA
RECHERCHE SCIENTIFIQUE
UNIVERSITÉ MOULOUD MAMMERI DE TIZI-OUZOU
FACULTÉ DES SCIENCES - DEPARTEMENT DE MATHÉMATIQUES

En vue d'obtention du diplôme de master en MATHÉMATIQUE option
MATHÉMATIQUES APPLIQUÉES A LA GESTION

Thème

Optimisation Et Applications

sous la Direction de

Mr Ouanes

présenté le 15/10/2017 par :

GHERAB Mekioussa

devant le Jury :

Mr CHEBBAH.M , MAA, UMMTO, Président

Mr OUANES.M , MC, UMMTO, Rapporteur

Mme LESLOUS.F , MAA, UMMTO, Examinatrice

Remerciements

Ce travail a été réalisé au niveau du laboratoire de Mathématiques de la Faculté des Sciences et Technologies de l'Université Mouloud Mammeri de Tizi-Ouzou.

Je tiens à remercier mon promoteur, Monsieur Ouanes pour son soutien et orientations durant la préparation de ce travail.

Mes vifs remerciements vont aussi plus particulièrement aux membres du jury pour m'avoir fait l'honneur d'accepter d'examiner et juger mon travail.

Mes meilleurs remerciements vont aux enseignants de l'Université de Mouloud Mammeri de Tizi-Ouzou qui ont contribué à notre formation.

Je tiens spécialement à remercier mes parents pour leurs soutien et encouragements durant nos études.

Je remercis vivement tous ceux qui m'ont porté aide, encouragée et soutenue de près ou de loin, ainsi que tous mes amis pour leurs encouragements.

Dédicaces

*A celle qui m'a été comme une ombre, à celle qui n'est jamais loin de moi,
à celle qui m'a fait voir la lumière, à celle qui fait toujours de son mieux
rien que pour me satisfaire et très chaleureusement à celle qui est chère à
mes yeux*

A ma très chère Maman.

*A celui qui m'a appris un tas de choses sur la vie, à celui qui est toujours là
pour me protéger, à celui qui ne m'a jamais manqué de rien, à celui que
j'estime beaucoup*

A mon très cher Papa.

*A la personne qui m'a rendu heureuse, à celle qui est tout le temps auprès
de moi, à celle qui m'a toujours soutenu et encouragé, à celle qui m'a
beaucoup aidé, à celle qui est toujours dans mon coeur et dans mes pensées*

A mon très cher et adorable frère Abderrahmane.

*A celles qui sont toujours à mes côtés, à celles qui ne m'ont jamais laissé
tomber, à celles qui me font toujours rigoler, à celle auxquels je me suis
habituée*

A mes chères soeurs Sissih et Goucha.

A Damon qui est plein de tendresse et d'affection.

*A toute ma famille sans exception : Grand mère, Doudouche, Dada, Lounis,
Karima, Lynda, Rachid, Sofiane et notre adorable ange Ilyana.*

A mes très chers amis Kaissa et Rabah et Moh.

Et à toutes les personnes que je connais et dont je n'ai pas cité le nom.

Table des matières

Introduction	6
1 Notations et définitions dans $\mathbb{R}^n$	9
1.1 Notations	9
1.2 Fonctions de plusieurs variables	11
1.3 Quelques propriétés importants	12
1.4 Formules de Taylor	14
1.5 Rappels sur les matrices carrées réelles	14
1.6 Convexité	16
1.6.1 Fonction convexes, strictement convexes, fortement convexes	16
1.6.2 Fonctions coercives	20
1.7 Conditions nécessaires et suffisantes pour les extrémums	21
1.8 Existence et unicité d'un point minimum	24
2 Optimisation quadratique convexe avec contraintes et sans contraintes	27
2.1 Applications quadratiques	27
2.2 Optimisation sans contraintes	29
2.3 Optimisation avec contraintes d'égalité	32
2.4 Optimisation avec contraintes d'inégalité	37
3 Régression linéaire et méthode de Markowitz	41
3.1 Introduction à la régression linéaire	41
3.1.1 Critère des moindres carrés	41
3.1.2 Forme standard	42
3.1.3 Recherche d'une solution	43
3.1.4 Série statistique à deux variables	44
3.1.5 Ajustement par la méthode des moindres carrés	45
3.1.6 Ajustement affine par la méthode des moindres carrés de y en x	46
3.1.7 Droite des moindres carrés de x en y	50

3.1.8	Evaluation de la qualité de la régression	50
3.1.9	Position relative de $D_{y/x}$ et $D_{x/y}$	52
3.2	Méthode de Markowitz	53
3.2.1	Etude de l'efficacité ou efficience de Markowitz	53
3.2.2	Le modèle de Markowitz	54
3.2.3	Programmation quadratique	56
3.2.4	Mesures alternatives de risque	57
4	Simulation numérique avec R	60
4.1	Méthode des moindres carrés	60
4.2	Méthode de markowitz	69
	Conclusion	73
	Bibliographie	74

Introduction

L'optimisation est une discipline mathématique et un sujet ancien qui connaît un nouvel essor depuis l'apparition des ordinateurs dont les méthodes s'appliquent dans de très nombreux domaines : économie, gestion, planification, logistique, automatique, robotique, conception optimale, science de l'ingénierie, traitement du signal, etc.

L'optimisation est aussi un sujet très vaste qui touche aussi bien au calcul de variations qu'à la recherche opérationnelle (domaine de l'optimisation des processus de gestion ou de décision)

La Recherche opérationnelle, reste un domaine fécond de la recherche en mathématiques qui bénéficie d'importants nuancements provenant aussi bien du domaine public que du domaine privé, et dont les retombées s'appliquent dans tous les domaines d'activité humaine se prêtant à la modélisation mathématique. De nos jours son champ d'application est vaste : optimisation des ressources, des gains, des coûts dans l'industrie, optimisation du trafic aérien, ferroviaire, routier, dans le transport, optimisation de la couverture radar, de la réactivité d'intervention, de la gestion des stocks et des troupes dans le domaine militaire, etc..., sans parler des sciences dures, physique, chimie, informatique, automatique, traitement du signal, etc..., pour lesquels un nombre de problèmes se ramènent et se résolvent par l'optimisation. On peut séparer l'optimisation en deux grandes branches aux méthodes fortes différentes selon que les variables sont continues ou discrètes. Typiquement si l'on minimise une fonction $f(x)$ avec $x \in \mathbb{R}^n$ il s'agit d'optimisation en variables continues, tandis que si $x \in \mathbb{Z}^n$ on a affaire à l'optimisation combinatoire ou variables discrètes.

L'optimisation en variables continues est souvent plus facile que l'optimisation en variables discrètes car on peut utiliser la notion de dérivée qui est fort utile tant du point de vue théorique qu'algorithmique.

Les problèmes d'optimisation occupent actuellement une place importante dans la communauté scientifique. Les problèmes peuvent être :

Combinatoires (discrets) ou à variables continues, Avec un seul ou plusieurs

objectifs (optimisation multi objectif), Statiques ou dynamiques.

Cette liste n'est pas exhaustive et un problème peut être à la fois continu et dynamique, etc.

Un problème d'optimisation est défini par :

- Un ensemble de variables $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.
- Une fonction objectif (fonction de coût) $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.
- Un ensemble $C \subset \mathbb{R}^n$ est défini par les contraintes du problème.

La formulation mathématique est :

$$\begin{cases} \min f(x) \\ x \in C \end{cases}$$

La fonction objectif définit le but à atteindre, on cherche à minimiser ou à maximiser celle-ci. L'ensemble des contraintes est en général un ensemble d'égalités ou d'inégalités que les variables de l'espace d'état doivent satisfaire. Ces contraintes limitent l'espace de recherche. Le but des méthodes d'optimisation est de rechercher un point ou un ensemble de points dans l'espace de recherche qui satisfont l'ensemble des contraintes, et qui maximisent ou minimisent la fonction objectif. Et parmi les méthodes d'optimisation 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.

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 linéaire entre les variables que nous aurons choisies notamment la variable expliquée à caractère aléatoire et les variables explicatives mesurées sans erreurs. Cette méthode se justifie pour la simple raison qu'elle a servie à plusieurs chercheurs dans le cadre des tests d'efficience telle que la gestion de portefeuille.

La gestion de portefeuille a subi une profonde mutation. Il est loin, en effet, le temps où le gestionnaire pouvait se contenter d'appliquer quelques règles de bon sens et de bien connaître les sociétés cotées. Ce sont les travaux de Markowitz qui, au cours des années 1950, ont marqué le point de départ des développements théoriques modernes relatifs à la gestion des investissements en actifs financiers et au fonctionnement des marchés financiers. Si la notion de diversification était connue bien avant Markowitz, c'est ce dernier

qui l'a conceptualisée et quantifiée, rendant ainsi possible la détermination des proportions optimales à investir dans les différents actifs financiers pris en considération par l'investisseur ou le gestionnaire de fortune.

Pour aborder les stratégies de gestion de portefeuille, il est important d'analyser la place du risque dans la gestion du portefeuille notamment dans l'évaluation des actifs financiers. Ce sont donc les tests sur le risque et le rendement qui ont permis de définir les stratégies modernes de gestion de portefeuille. Le choix d'une stratégie en particulier dépend du degré d'aversion au risque du client et de la conception de l'efficience du marché.

Chaque investisseur sait qu'il y a un compromis entre le risque et le retour sur investissement. Afin d'augmenter le retour espéré, un investisseur doit être prêt à tolérer des risques plus importants. La théorie de gestion de portefeuilles concerne la modélisation de ce compromis pour une collection de n investissements possibles avec des retours $r_i, i = 1; 2; \dots; n$.

Les retours ne sont pas connus d'avance et sont supposés être des variables aléatoires gaussiennes, caractérisés par leur espérance $\mu_i = \mathbb{E}[r_i]$ et leurs variance $\sigma_i^2 = \mathbb{E}[(r_i - \mu_i)^2]$.

Une variance plus grande implique un investissement plus risqué. Un investisseur compose un portefeuille en mettant une fraction x_i de capitaux disponibles dans l'investissement i .

Ce mémoire comporte quatre chapitres. Dans le premier, on présente quelques résultats fondamentaux qu'on utilise dans les chapitres qui suivent. Dans le second chapitre, on a étudié les problèmes d'optimisation quadratique convexe sans contraintes puis avec contraintes d'égalité et d'inégalité. Le troisième chapitre est consacré à la méthode des moindres carrés et le modèle de Markowitz. En dernier, dans la partie pratique on a fait une simulation numérique avec le logiciel R de la méthode des moindres carrés et le modèle de Markowitz.

Chapitre 1

Notations et définitions dans $\mathbb{R}^n$

1.1 Notations

Définition 1.1. Pour tout $n \in \mathbb{N}^*$, $\mathbb{R}^n$ désigne l'espace **euclidien**, $\mathbb{R} * \mathbb{R} * \dots * \mathbb{R}$ ('produit n fois'). En général un vecteur $x \in \mathbb{R}^n$ sera noté $x = (x_1, x_2, \dots, x_n)^T$ (vecteur colonne).

Définition 1.2. On note $e_1, e_2, \dots, e_n$ les vecteurs de la **base canonique** de $\mathbb{R}^n$, où les composantes $e_{ij}, j = 1, \dots, n$ de $e_i, i = 1, \dots, n$ sont données par :

$$e_{ij} = \delta_{ij} = \begin{cases} 0 & \text{si } j \neq i \\ 1 & \text{si } j = i \end{cases}$$

(δ_{ij} est le symbole de **Kronecker**)

Définition 1.3. Pour tout $x, y \in \mathbb{R}^n$, on note par $\langle x, y \rangle \in \mathbb{R}$ le **produit scalaire** de x et y , qui est donné par :

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i.$$

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

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

Rappels des **propriétés d'une norme** (donc aussi de la norme euclidienne) :

- i) $\|\lambda x\| = |\lambda| \|x\|; \forall x \in \mathbb{R}^n \forall \lambda \in \mathbb{R}$
- ii) $\|x + y\| \leq \|x\| + \|y\|; \forall x, y \in \mathbb{R}^n$
- iii) $\|x\| = 0 \Leftrightarrow x = 0$ et $\|x\| > 0$ si $x \in \mathbb{R}^n - \{0\}$.

Définition 1.5. Pour tout $x \in \mathbb{R}^n$ et $r > 0$ on notera par $B(x, r)$ la **boule ouverte** de centre x et de rayon r , donnée par :

$$B(x, r) = \{y \in \mathbb{R}^n, \|y - x\| < r\}.$$

Définition 1.6. Soit $x^{(k)}$ une suite dans $\mathbb{R}^n$ et x un élément de $\mathbb{R}^n$ on dit que $x^{(k)}$ converge vers x (notée $x^{(k)} \rightarrow x$) si $\|x^{(k)} - x\| \rightarrow 0$ quand $k \rightarrow \infty$.

Rappelons que nous avons : $x^{(k)} \rightarrow x$ si seulement si $x_i^{(k)} \rightarrow x_i$, où $x_i^{(k)}$ (respectivement x_i) est la i -ème composante de $x^{(k)}$ (respectivement de x).

Définition 1.7. Soit $U \subset \mathbb{R}^n$.

-On définit **l'intérieur** de U comme l'ensemble des éléments $x \in U$ pour lesquels il existe $r > 0$ tel que $B(x, r) \subset U$.

-On dit que U est **ouvert** si $\forall x \in U; \exists r > 0$ tel que $B(x, r) \subset U$.

-On dit que U est **fermé** si pour toute suite $\{x^{(k)}\} \in U$ tel que $x^{(k)} \rightarrow x \in \mathbb{R}^n$ on a $x \in U$.

Définition 1.8. Si $a, b \in \mathbb{R}^n$ on note $[a, b]$ le sous-ensemble de $\mathbb{R}^n$ donné par :

$$[a, b] = \{a + t(b - a) = (1 - t)a + tb, t \in [0, 1]\}.$$

L'ensemble $[a, b]$ est aussi appelé **le segment** reliant a à b .

Remarque 1.1. Si a et $b \in \mathbb{R}$ avec $a < b$ on retrouve la notation $[a, b]$ pour l'intervalle des nombres $x \in \mathbb{R}$ tel que $a \leq x \leq b$.

Proposition 1.1. l'inégalité de Cauchy-Schwarz est définie par :

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|; \forall x, y \in \mathbb{R}^n.$$

1.2 Fonctions de plusieurs variables

On considère dans cette partie m et n deux nombres entiers naturels non nuls.

Définition 1.9. Soit U un sous-ensemble de $\mathbb{R}^n$ et $f : U \rightarrow \mathbb{R}^m$.

On dit que f est **continue** en $x \in U$ si $f(x^{(k)}) \rightarrow f(x)$ pour toute suite $x^{(k)} \subset U$ telle que $x^{(k)} \rightarrow x$.

On dit que f est continue sur U si f est continue en tout point $x \in U$.

Remarque 1.2. Si $f = (f_1, f_2, \dots, f_m)$ avec $f_1, f_2, \dots, f_m : U \rightarrow \mathbb{R}$ alors f est continue en $x \in U$ si et seulement si $f_1, f_2, \dots, f_m$ sont continues en x .

Définition 1.10. Pour tout $x \in \Omega$ et $f = (f_1, f_2, \dots, f_m)^T$ on note $J_f(x)$ la **matrice jacobienne** de f en x qui est un élément de $M_{m,n} \in \mathbb{R}^{m \times n}$ définie par :

$$(J_f(x))_{ij} = \frac{\partial f_i}{\partial x_j}(x); \forall i = 1, \dots, m, \forall j = 1, \dots, n.$$

Pour ce qui suit on va supposer que Ω est un ouvert de $\mathbb{R}^n$ et f est une fonction $f : \Omega \rightarrow \mathbb{R}$.

Définition 1.11. Pour tout $x \in \Omega$ et $h \in \mathbb{R}^n$ on note (quand elle existe) :

$$f'(x; h) = \lim_{t \rightarrow 0} \frac{1}{t} [f(x + th) - f(x)].$$

La dérivée de f en x dans la direction h .

Définition 1.12. Pour tout $x \in \Omega$ et $i \in \{1, 2, \dots, n\}$ on note (quand elle existe) :

$$f'(x; h) = \frac{\partial f}{\partial x_i}(x) = \lim_{t \rightarrow 0} \frac{1}{t} [f(x + th_i) - f(x)].$$

La dérivée partielle de f en x par rapport à la variable x_i .

On définit le **gradient** de f par :

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right)^T \in \mathbb{R}^n.$$

Rappelons la formule :

$$f'(x; h) = \langle \nabla f(x), h \rangle; \forall x \in \Omega; \forall h \in \mathbb{R}^n.$$

Définition 1.13. On dit que $x \in \Omega$ est un **point critique** pour la fonction f si $\nabla f(x) = 0$.

Définition 1.14. On dit que f est de classe C^p ; $p \in \mathbb{N}$ sur Ω (on notera $f \in C^p(\Omega)$) si les dérivées partielles de f jusqu'à l'ordre p existent et sont continues sur Ω .

Théorème 1.1. On a le **Théorème de Schwarz** : si $f \in C^2(\Omega)$ alors :

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x), \forall x \in \Omega, \forall i, j = 1, 2, \dots, n.$$

Définition 1.15. On note pour tout $x \in \Omega$ (quand il existe), $\nabla^2 f(x)$ la matrice carrée $\in M_n(\mathbb{R})$ donnée par :

$$(\nabla^2 f(x))_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x), \forall i, j = 1, 2, \dots, n$$

la **matrice hessienne** de f en x .

Remarque 1.3. Si $f \in C^2(\Omega)$ alors $\nabla^2 f(x)$ est une matrice **symétrique** $\forall x \in \Omega$ (Théorème de schwarz).

Proposition 1.2 (lien entre ∇ et ∇^2). Nous avons :

$$\nabla^2 f(x)h = \nabla \langle \nabla f(x), h \rangle \forall x \in \Omega, \forall h \in \mathbb{R}^n.$$

où le premier gradient dans le membre de droite de l'égalité est considéré par rapport à la variable x .

Démonstration. On a :

$$\frac{\partial}{\partial x_i} \langle \nabla f(x), h \rangle = \frac{\partial}{\partial x_i} \left(\sum_{j=1}^n \frac{\partial f}{\partial x_j}(x) h_j \right) = \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x) h_j = (\nabla^2 f(x)h)_i.$$

□

1.3 Quelques propriétés importants

1. Si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction **constante** alors $\nabla f = 0$, $\nabla^2 f = 0$.
2. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définit par :

$$f(x) = \langle a, x \rangle \forall x \in \mathbb{R}^n$$

où $a \in \mathbb{R}^n$ est un vecteur donné (c'est à dire, f est une fonction **linéaire**).
Alors on calcule facilement : $\frac{\partial f(x)}{\partial x_k} = a_k$, tel que $k = 1, \dots, n$. donc $\nabla f = a$
Le gradient est constant.

Ceci nous donne :

$$\nabla^2 f = 0.$$

3. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ donnée par :

$$f(x) = \langle Ax, x \rangle \forall x \in \mathbb{R}^n,$$

où $A \in M_n(\mathbb{R})$ est une matrice carrée, réelle, de taille n (c'est à dire, f est la fonction **quadratique** associée à la matrice A). Alors pour un $p \in \{1, 2, \dots, n\}$ fixé, on peut écrire :

$$f(x) = \sum_{i,j=1}^n A_{ij}x_i x_j = A_{pp}x_p^2 + \sum_{j=1, j \neq p}^n A_{pj}x_p x_j + \sum_{j=1, j \neq p}^n A_{ip}x_p x_i + \sum_{i,j=1, i \neq p, j \neq p}^n A_{ij}x_i x_j$$

ce qui nous donne :

$$\frac{\partial f}{\partial x_p} = 2A_{pp}x_p + \sum_{j=1, j \neq p}^n A_{pj}x_j + \sum_{j=1, j \neq p}^n A_{ip}x_i = \sum_{j=1}^n A_{pj}x_j + \sum_{j=1}^n A_{ip}x_i = (Ax)_p + (A^T x)_p.$$

Nous avons donc obtenu :

$$\nabla f(x) = (A + A^T)x, \forall x \in \mathbb{R}^n$$

On peut aussi écrire :

$$\frac{\partial f}{\partial x_i}(x) = \sum_{k=1}^n (A + A^T)_{ik}x_k, \forall i = 1, \dots, n.$$

On a alors immédiatement :

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = (A + A^T)_{ij}, \forall i, j = 1, \dots, n.$$

c'est à dire :

$$\nabla^2 f(x) = A + A^T, \forall x \in \mathbb{R}^n.$$

(donc la hessienne de f est constante).

Remarque 1.4. En particulier, si A est **symétrique** (c'est à dire $A = A^T$) alors :

$$\nabla \langle Ax, x \rangle = 2Ax, \forall x \in \mathbb{R}^n.$$

$$\nabla^2 \langle Ax, x \rangle = 2A, \forall x \in \mathbb{R}^n.$$

1.4 Formules de Taylor

Proposition 1.3. Soit $\Omega \subset \mathbb{R}^n$ ouvert, $f : \Omega \mapsto \mathbb{R}$, $a \in \Omega$ et $h \in \mathbb{R}^n$ tels que $[a, a+h] \subset \Omega$. Alors :

1. Si $f \in C^1(\Omega)$ alors :

$$i) f(a+h) = f(a) + \int_0^1 \langle \nabla f(a+th), h \rangle dt$$

(formule de Taylor à l'ordre 1 avec reste intégral).

$$ii) f(a+h) = f(a) + \langle \nabla f(a+\theta h), h \rangle \text{ avec } 0 < \theta < 1$$

(formule de Taylor - Young à l'ordre 1).

$$iii) f(a+h) = f(a) + \langle \nabla f(a), h \rangle o(\|h\|)$$

2. Si $f \in C^2(\Omega)$ alors :

$$i) f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \int_0^1 (1-t) \langle \nabla^2 f(a+th)h, h \rangle dt$$

(Formule de Taylor à l'ordre 2 avec reste intégral).

$$ii) f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \langle \nabla^2 f(a+\theta h)h, h \rangle \text{ avec } 0 < \theta < 1$$

(Formule de Taylor - Maclaurin à l'ordre 2).

$$iii) f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \langle \nabla^2 f(a)h, h \rangle + o(\|h\|^2)$$

(Formule de Taylor - Young à l'ordre 2).

Remarque 1.5. Dans la proposition précédente la notation $o(\|h\|^k)$ pour $k \in \mathbb{N}^*$ signifie une expression qui tend vers 0 plus vite que $\|h\|^k$ (c'est à dire, si on la divise par $\|h\|^k$, le résultat tend vers 0 quand $\|h\|$ tend vers 0).

1.5 Rappels sur les matrices carrées réelles

Soit $n \in \mathbb{N}^*$ et $A \in M_n(\mathbb{R})$ une matrice carrée réelle.

Définition 1.16. Une matrice qui a le même nombre de lignes et de colonnes est appelée matrice carrée.

Le nombre n est entier, on dit que la matrice est d'ordre n et l'on note M_n .

Définition 1.17. Si A est une matrice, alors A^T est sa **transposée** :

$$(A^T)_{ij} = a_{ji}.$$

Définition 1.18. Une matrice A est **régulière** (invertible) s'il existe une matrice B appelée inverse de A et notée :

$$B = A^{-1},$$

telle que :

$$AB = BA = I_n.$$

Définition 1.19. Une matrice A est *symétrique* si $A^T = A$.

Définition 1.20. Soit $\mathbb{C}$ l'ensemble des nombres complexes. On dit que $\lambda \in \mathbb{C}$ est une **valeur propre** de A s'il existe $x \in \mathbb{C}^n$ avec $x \neq 0$ tel que $Ax = \lambda x$; on appelle x **vecteur propre** de A associé à la valeur propre λ .

Définition 1.21. la matrice A est **semi-définie positive** si $\langle Ax, x \rangle \geq 0$, $\forall x \in \mathbb{R}^n$.

A est **définie positive** si $\langle Ax, x \rangle > 0$, $\forall x \in \mathbb{R}^n$ avec $x \neq 0$.

Définition 1.22. A est *symétrique* si toutes ces valeurs propres sont réelles; en plus il existe n vecteurs propres de A appartenant à $\mathbb{R}^n$ formant une base orthonormée en $\mathbb{R}^n$.

Définition 1.23. Si la matrice A est *symétrique* alors :

$$\langle Ah, h \rangle \geq \lambda_{\min} \|h\|^2, \forall h \in \mathbb{R}^n$$

où $\lambda_{\min} \in \mathbb{R}$ est la plus petite valeur propre de A .

L'inégalité précédente devient égalité si h est un vecteur propre associé à la valeur propre $\lambda_{\min}$.

A est semi-définie positive si et seulement si $\lambda_{\min} \geq 0$ et A est définie positive si et seulement si $\lambda_{\min} > 0$.

1.6 Convexité

1.6.1 Fonction convexes, strictement convexes, fortement convexes

Définition 1.24. Un ensemble $U \subset \mathbb{R}^n$ est dit **convexe** si $\forall x, y \in U$ on a $[x, y] \subset U$

$$\lambda x + (1 - \lambda)y \in U, \forall t \in [0, 1]$$

(quelque soit deux points dans U , tout le segment qui les unit est dans U).

Définition 1.25. Soit $U \subset \mathbb{R}^n$ un ensemble convexe et $f : U \rightarrow \mathbb{R}$ une fonction.

f est **convexe** sur U si :

$$f(ty + (1 - t)x) \leq tf(y) + (1 - t)f(x), \forall x, y \in U, \forall t \in [0, 1].$$

Définition 1.26. f est **strictement convexe** sur U si :

$$f(ty + (1 - t)x) < tf(y) + (1 - t)f(x), \forall x, y \in U, \text{ avec } x \neq y \forall t \in]0, 1[.$$

Définition 1.27. f est **fortement convexe** sur U s'il existe $\alpha > 0$ tel que :

$$f(ty + (1 - t)x) \leq tf(y) + (1 - t)f(x) - \frac{\alpha}{2}t(1 - t) \|y - x\|^2, \\ \forall x, y \in U, \forall t \in [0, 1].$$

Définition 1.28. f est **concave** (respectivement **strictement concave**, respectivement **fortement concave**) si $-f$ est convexe (respectivement **strictement convexe**, respectivement **fortement convexe**).

Remarque 1.6. On a : fortement convexe $\Rightarrow$ strictement convexe $\Rightarrow$ convexe.

Les réciproques ne sont pas vraies en général.

Exemple 1.1. l'application affine $f(x) = Ax + b$ est convexe (et aussi concave) mais elle n'est pas strictement convexe (ni strictement concave) donc elle n'est pas fortement convexe (ni fortement concave).

Proposition 1.4. (*caractérisation de la convexité*)

Soit $U \subset \mathbb{R}$ une fonction de classe C^1 . Alors :

-Les trois propositions suivantes sont équivalentes :

1. f est convexe sur U .
2. $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \forall x, y \in U$.
3. ∇f est **monotone** sur U , c'est à dire :

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0 \quad \forall x, y \in U.$$

-Si de plus f est de classe C^2 sur Ω alors f est convexe sur U si et seulement si :

$$\langle \nabla^2 f(x)(y - x), y - x \rangle \geq 0, \quad \forall x, y \in U. \quad (1.1)$$

Démonstration.

a) On montre l'équivalence entre 1), 2) et 3).

1) $\implies$ 2) : Supposons f convexe : la définition de la convexité peut s'écrire :

$$f(x + t(y - x)) - f(x) \leq t[f(y) - f(x)]$$

En fixant x, y en divisant par t et en faisant tendre t vers 0 car $t \in [0, 1]$ on obtient 2).

2) $\implies$ 3) : De 2) on déduit :

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in U$$

et aussi (en inversant x et y) :

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle, \quad \forall x, y \in U$$

En faisant la somme de ces 2 inégalités on obtient 3).

3) $\implies$ 1) : Soient $x, y \in U$ fixés. On introduit la fonction $g : I \longrightarrow \mathbb{R}$ définie par :

$$t \in I \longrightarrow g(t) = f(ty + (1 - t)x) \in \mathbb{R}$$

où I est un intervalle ouvert qui contient $[0, 1]$. on remarque que g est de classe C^1 et on a :

$$g'(t) = \langle \nabla f(ty + (1 - t)x), y - x \rangle, \quad \forall t \in I.$$

Soient $t_1, t_2 \in [0, 1]$ avec $t_1 < t_2$. Alors :

$$\begin{aligned} g'(t_2) - g'(t_1) &= \langle \nabla f(x + t_2(y - x)), y - x \rangle - \langle \nabla f(x + t_1(y - x)), y - x \rangle = \\ &= \langle \nabla f(x + t_2(y - x)) - \nabla f(x + t_1(y - x)), (t_2 - t_1)(y - x) \rangle = \frac{1}{t_2 - t_1} \langle \nabla f(x + t_2(y - x)) - \nabla f(x + t_1(y - x)), (t_2 - t_1)(y - x) \rangle. \end{aligned}$$

Par hypothèse 3) le dernier terme de l'égalité précédente est ≥ 0 , ce qui montre que la fonction g' est une fonction croissante.

On déduit alors que g est une fonction convexe sur $[0, 1]$, ce qui nous donne pour tout $t \in [0, 1]$:

$$g(t.1 + (1 - t).0) \leq tg(1) + (1 - t)g(0)$$

c'est à dire :

$$f(ty + (1 - t)x) \leq tf(1) + (1 - t)f(x)$$

donc f est convexe.

b) On suppose ici $f \in C^2(\Omega)$.

" $\implies$ " Supposons que f est convexe et montrons. Soit $h \in \mathbb{R}^n$ fixé et notons $g : \Omega \mapsto \mathbb{R}$ la fonction donnée par $g(x) = \langle \nabla f(x), h \rangle, \forall x \in \Omega$. Nous avons aussi en utilisant la proposition 1.1 :

$$\langle \nabla^2 f(x)h, h \rangle = \langle \nabla g(x), h \rangle = \frac{\partial g}{\partial h}(x) = \lim_{t \rightarrow 0} \frac{\langle \nabla f(x + th), h \rangle - \langle \nabla f(x), h \rangle}{t}$$

ce qui nous donne :

$$\langle \nabla^2 f(x)h, h \rangle = \lim_{t \rightarrow 0} \frac{\langle \nabla f(x + th) - \nabla f(x), th \rangle}{t^2}.$$

Considérons maintenant $x, y \in U$ arbitraires et $h = y - x$. Comme $x + t(y - x) \in U \forall t \in [0, 1]$, de l'égalité précédente on déduit à l'aide de la monotonie de $\langle \nabla^2 f(x)h, h \rangle \geq 0$, c'est à dire (1.1).

" $\impliedby$ " Supposons maintenant que (1.1) est satisfaite et montrons que f est convexe. Soient $x, y \in U$ fixées arbitraires, et considérons la fonction $g_1 : \Omega \mapsto \mathbb{R}$ donnée par :

$$g_1(z) = \langle \nabla f(z), x - y \rangle \quad \forall z \in \Omega. \text{ Alors :}$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = g_1(x) - g_1(y) = \langle \nabla g_1(y + \theta(x - y)), x - y \rangle$$

avec $\theta \in]0, 1[$ (on utilisant la formule de Taylor).

D'autre part, nous avons :

$$\nabla g_1(z) = \nabla^2 f(z)(x - y)$$

et ceci permet de déduire, en utilisant (1.1) :

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = \langle \nabla^2 f(y + \theta(x - y))(x - y) \rangle \geq 0$$

Ceci donne la monotonie de ∇f donc la convexité de f . $\square$

Proposition 1.5. (caractérisation de la convexité stricte) [8]

Soit $U \subset \mathbb{R}^n$ ouvert avec U convexe et $f : \Omega \mapsto \mathbb{R}$ une fonction de classe C^1 . Alors :

- Les trois propositions suivantes sont équivalentes :

1. f est strictement convexe sur U .
2. $f(y) > f(x) + \langle \nabla f(x), y - x \rangle$, $\forall x, y \in U$ avec $x \neq y$
3. ∇f est **monotone** sur U , c'est à dire :

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle > 0, \forall x, y \in U \text{ avec } x \neq y$$

- Si de plus f est de classe C^2 sur Ω alors f est convexe sur U si et seulement si :

$$\langle \nabla^2 f(x)(y - x), y - x \rangle > 0, \forall x, y \in U \text{ avec } x \neq y$$

Proposition 1.6. (caractérisation de la forte convexité)[8]

Soit $\Omega \in \mathbb{R}^n$ ouvert, $U \subset \Omega$ avec U convexe et $f : \Omega \mapsto \mathbb{R}$ une fonction de classe C^1 . Alors :

- Les trois propositions suivantes sont équivalentes :

1. f est fortement convexe sur U .
2. Il existe $\beta > 0$ tel que :

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \beta \|y - x\|^2, \forall x, y \in U.$$
3. Il existe $\gamma > 0$ tel que :

∇f est **monotone** sur U , c'est à dire :

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \gamma \|y - x\|^2, \forall x, y \in U.$$

- Si de plus f est de classe C^2 sur Ω alors f est fortement convexe sur U si et seulement si il existe $\beta > 0$ tel que :

$$\langle \nabla^2 f(x)(y - x), y - x \rangle \geq \beta \|y - x\|^2, \forall x, y \in U.$$

1.6.2 Fonctions coercives

Définition 1.29. Soit $\Omega \subset \mathbb{R}^n$ un ensemble **non borné** et $f : \Omega \rightarrow \mathbb{R}$. On dit que f est **coercive** sur Ω si on a :

$$\lim_{x \in \Omega, \|x\| \rightarrow +\infty} f(x) = +\infty$$

Ceci peut aussi s'écrire :

$f(x^k) \rightarrow +\infty$ pour toute suite $\{x^k\}_{k \in \mathbb{N}} \subset \Omega$ telle que $\|x^k\| \rightarrow +\infty$.

Remarque 1.7. Supposons qu'on a $\Omega_0 \subset \Omega \subset \mathbb{R}^n$ avec Ω_0 non borné et $f : \Omega \rightarrow \mathbb{R}$. Si f est coercive sur Ω alors f est coercive sur Ω_0 , la réciproque n'étant pas vraie en général.

Proposition 1.7. Soit $\Omega \subset \mathbb{R}^n$ un ouvert et $U \subset \Omega$ un ensemble convexe et non borné.

Soit $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 sur Ω et fortement convexe sur U . Alors f est coercive sur U .

Démonstration.

Soit l'inégalité de la proposition 1.5. avec $x = a$ où a est un élément fixé de U :

$$f(y) = f(a) + \langle \nabla f(a), y - a \rangle + \beta \|y - a\|^2, \forall y \in U$$

où $\beta > 0$ est un nombre fixé. D'autre part, nous avons :

$$\langle \nabla f(a), y - a \rangle \geq -\|\nabla f(a)\| \cdot \|y - a\|$$

(conséquence de l'inégalité de Cauchy-Schwarz). Ceci donne :

$$f(y) \geq f(a) - \|\nabla f(a)\| \cdot \|y - a\| + \beta \|y - a\|^2, \forall y \in U$$

En faisant $\|y\| \rightarrow +\infty$ (ce qui donne $\|y - a\| \rightarrow +\infty$) on obtient le résultat. $\square$

1.7 Conditions nécessaires et suffisantes pour les extrémums

Définition 1.30. Soit $U \subset \mathbb{R}^n$, $u^* \in U$ et $f : U \rightarrow \mathbb{R}$.

- On dit que u^* est un point minimum absolu (ou global) de f sur U si :

$$f(u) \geq f(u^*), \forall u \in U.$$

- On dit que u^* est un point maximum relatif (ou local) de f sur U si $\exists V$ voisinage de u^* dans U , tel que :

$$f(u) \leq f(u^*), \forall u \in U \cap V.$$

- On dit que u^* est un point maximum absolu (respectivement relatif) de f sur U si u^* est un point minimum absolu (respectivement relatif) de $-f$ sur U .

- On dit que u^* est un extrémum absolu (respectivement relatif) de f sur U si u^* est : soit un point minimum absolu (respectivement relatif) de f sur U , soit un point maximum absolu (respectivement relatif) de f sur U .

Remarque 1.8. Un point minimum absolu est clairement un point minimum relatif.

Définition 1.31. Soit $U \subset \mathbb{R}^n$ et $u^* \in U$. On dit que le vecteur $w \in \mathbb{R}^n$ est une **direction admissible** pour u^* en U s'il existe $t_0 > 0$ tel que $u^* + tw \in U, \forall t \in [0; t_0]$.

On a le résultat suivant :

Lemme 1.1. [3] Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $U \subset \Omega$, $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 et un point de minimum relatif de f sur U . Soit $w \in \mathbb{R}^n$ une direction admissible pour u^* en U . Alors :

$$\langle \nabla f(u^*), w \rangle \geq 0$$

Nous pouvons alors énoncer :

Théorème 1.2. Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $U \subset \Omega$ un ensemble convexe et $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 . Soit $u^* \in U$ un point minimum relatif de f sur U . Alors :

$$\langle \nabla f(u^*), u - u^* \rangle \geq 0, \forall u \in U \text{ (l'inéquation d'Euler).}$$

Démonstration. Pour tout $u \in U$ le vecteur $u - u^*$ est une direction admissible pour u^* en U car la convexité de U nous donne :

$$u^* + t(u - u^*) \in U, \forall t \in [0, 1]$$

Alors, $\langle \nabla f(u^*), u - u^* \rangle \geq 0, \forall u \in U$ est une conséquence immédiate de Lemme 1.1 avec $w = u - u^*$.

On considère deux cas particuliers : le cas u^* à l'intérieur de U et le cas U sous espace affine. $\square$

Lemme 1.2. Soit $U \subset \mathbb{R}^n$, $b \in \mathbb{R}^n$ et u^* un élément appartenant à l'intérieur de U ($u^* \in \dot{U}$). Alors les deux assertions suivantes sont équivalentes :

1. $\langle b, u - u^* \rangle \geq 0, \forall u \in U$ (1.3)
2. $b = 0$. (1.4)

Démonstration.

L'implication (1.4) $\implies$ (1.3) est évidente.

Pour montrer l'implication inverse, soit $w \in \mathbb{R}^n$, avec $w \neq 0$. Alors il existe $\theta_0 > 0$ telle que :

$$u^* + \theta w \in U, \forall \theta \in [-\theta_0, \theta_0]$$

(Il suffit de prendre $\theta_0 = \frac{r}{\|w\|}$ où $r > 0$ est tel que $B(u^*, r) \subset U$). Alors en prenant $u^* + \theta w$ à la place de u dans (1.3) on déduit :

$$\langle a, \theta w \rangle \geq 0, \forall \theta \in [-\theta_0, \theta_0].$$

En prenant d'abord $\theta = \theta_0$ et ensuite $\theta = -\theta_0$ dans l'inégalité précédente, on déduit $\langle a, w \rangle = 0$ et ceci quelque soit $w \in \mathbb{R}^n$, $w \neq 0$.

Comme l'égalité précédente est évidemment valable aussi pour $w = 0$ alors elle est valable pour tout $w \in \mathbb{R}^n$. Ceci donne immédiatement (1.4) (il suffit de prendre $w = b$ dans cette égalité). $\square$

Lemme 1.3. Soit U_0 un sous espace vectoriel de $\mathbb{R}^n$, $a \in \mathbb{R}^n$ et $U = a + U_0$ (on dit que U est un **sous-espace affine** de $\mathbb{R}$). Soit $b \in \mathbb{R}^n$ et $u^* \in U$. Alors les deux assertions suivantes sont équivalentes :

1. $\langle b, u - u^* \rangle \geq 0, \forall u \in U$ (1.5)
2. $\langle b, h \rangle = 0 \forall h \in U_0$. (1.6)

Démonstration.

L'implication (1.6) $\implies$ (1.5) est immédiate du fait que pour tout $u \in U$ on a $u - u^* \in U_0$.

Pour montrer (1.5) $\implies$ (1.6) soit $h \in U_0$ fixé arbitraire et remarquons que nous pouvons prendre en (1.5) $u = u^* \pm h$ car c'est un élément de U . Ceci nous donne :

$$\langle b, \pm h \rangle \geq 0$$

Ce qui nous donne (1.6). $\square$

En appliquant ces deux lemmes avec $b = \nabla f(u^*)$ on obtient immédiatement le corollaire suivant :

Corollaire 1.1. *Soit $\Omega \in \mathbb{R}^n$ un ouvert, $U \subset \Omega$ un ensemble convexe et $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 . Soit $u^* \in U$ un point minimum relatif de f sur U . Alors :*

1. *Si u^* est dans l'intérieur de U ($u^* \in U_0$) alors :*

$$\nabla f(u^*) = 0 \text{ (c'est l'équation d'Euler).}$$

2. *Si U est un sous-espace affine de $\mathbb{R}^n$ (donc $U = a + U_0$ avec $a \in \mathbb{R}^n$ et U_0 un sous espace vectoriel de $\mathbb{R}^n$) alors :*

$$\langle \nabla f(u^*), h \rangle = 0 \quad \forall h \in U_0.$$

Remarque 1.9. - *l'inégalité et l'égalités d'Euler et si*

$\langle \nabla f(u^*), h \rangle = 0 \quad \forall h \in U_0$ *donnent seulement des conditions **nécessaire** de minimum relatif, qui ne sont pas en général suffisantes.*

- *Si U est un ensemble ouvert alors u^* est forcément dans l'intérieur de U et donc l'équation d'Euler peut être utilisé comme condition nécessaire de minimum relatif.*

Le resultat suivant donne aussi des conditions suffisantes et nécessaire de minimum.

Théorème 1.3. *Soit $\Omega \subset \mathbb{R}^n$ ouvert, $U \subset \Omega$ un ensemble convexe, $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 et convexe et $u^* \in U$. Alors les trois assertions suivantes sont équivalentes :*

1. *u^* est un point minimum absolu de f sur U .*
2. *u^* est un point minimum relatif de f sur U .*
3. *$\langle \nabla f(u^*), u - u^* \rangle \geq 0, \quad \forall u \in U$.*

Démonstration. 1) $\Rightarrow$ 2) est évidente et 2) $\Rightarrow$ 3) est une conséquence immédiate du Théorème 1.1.

Montrons 3) $\Rightarrow$ 1). Grâce à la convexité de f et à la partie I de la proposition 1.3. nous avons

$$f(u) - f(u^*) \geq \langle \nabla f(u^*), u - u^* \rangle, \forall u \in U.$$

De 3) nous obtenons alors 1) $\square$

Les conditions d'optimalités de second ordre :

Théorème 1.4. [4]

Soit $\Omega \subset \mathbb{R}^n$ ouvert, $U \subset \Omega$, $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^2 et $u^* \in U^0$. Supposons que u^* est un point critique de f (c'est à dire $\nabla f(u^*) = 0$). On a les résultats suivants :

1. Si u^* est point minimum local de f sur U alors la matrice $\nabla^2 f(u^*)$ est semi définie positive (**condition nécessaire**).
2. Si $\nabla^2 f(u^*)$ est définie positive alors u^* est un minimum local de f sur U (**condition suffisante**).

Remarque 1.10. Pour la maximisation il suffit de faire la minimisation de $-f$.

1.8 Existence et unicité d'un point minimum

Théorème 1.5. (Existence) Soit $U \subset \mathbb{R}^n$ un ensemble non vide et fermé et $f : U \rightarrow \mathbb{R}$ une fonction continue. On suppose :

- Soit U est bornée.
- Soit U est non bornée et f est une fonction coercive.

Alors il existe au moins un point minimum de f sur U (c'est à dire, $\exists u^* \in U$ tel que $f(u^*) \leq f(u), \forall u \in U$)

Démonstration.

On distingue 2 cas :

Cas 1) L'ensemble U est borné.

Alors comme U est aussi fermé, U est compact. Comme f est continue, le théorème de Weierstrass nous assure que f est bornée sur U et elle atteint ses bornes. Donc il existe au moins un point minimum absolu de f sur U .

Cas 2) L'ensemble U est non borné.

Soit $a \in U$ et considérons l'ensemble :

$$E = \{x \in U, f(x) \leq f(a)\} \text{ tel que } a \in E$$

il est facile de montrer :

1. E est fermé (car $E = f^{-1}(] \infty, f(a)])$, donc E est l'image inverse d'un intervalle fermé par une fonction continue).

2. E est borné (supposons le contraire : alors il existe une suite $x_k \in E$ avec $\|x_k\| \rightarrow +\infty$ pour $k \rightarrow +\infty$. Comme f est coercive, ceci entraîne $f(x_k) \mapsto +\infty$ ce qui est absurde, car $f(x_k) \leq f(a), \forall k \in \mathbb{N}$.)

On déduit alors que E est un ensemble compact dans $\mathbb{R}^n$. Du Théorème de Weierstrass, $\exists u^* \in E$ tel que :

$$f(u^*) \leq f(u), \forall u \in U - E.$$

(car $f(u^*) \leq f(a) < f(u), \forall u \in U - E$).

Ceci prouve que u^* est un point minimum absolu de f sur U . $\square$

Théorème 1.6. (*Unicité*)

Soit $U \subset \mathbb{R}^n$ un ensemble convexe et $f : U \rightarrow \mathbb{R}$ une fonction strictement convexe. Alors il existe au plus un point minimum de f sur U .

Démonstration.

On va raisonner par l'absurde. Soient $u_1, u_2 \in U$ avec $u_1 \neq u_2$ deux points minimum de f sur U . Nous avons donc :

$$f(u_1) = f(u_2) \leq f(u), \forall u \in U. \quad (1)$$

Comme f est strictement convexe, on a :

$$f(\frac{1}{2}u_1 + \frac{1}{2}u_2) < \frac{1}{2}f(u_1) + \frac{1}{2}f(u_2) = f(u_1)$$

(car $f(u_1) = f(u_2)$) et ceci contredit (1).

$\square$

Corollaire 1.2. (*Existence et unicité*)

Soit $\Omega \subset \mathbb{R}^n$ un ouvert, $U \subset \Omega$ un ensemble fermé et convexe et $f : \Omega \rightarrow \mathbb{R}$ une fonction de classe C^1 sur Ω et fortement convexe sur U . Alors il existe un unique point minimum de f sur U .

Démonstration.

Existence : il y a deux possibilités :

1. Soit U est borné et alors l'existence est immédiate du théorème 1.4.
2. Soit U est non borné et alors comme conséquence de la proposition 1.6 f est coercive sur U et donc l'existence résulte encore du théorème 1.4.

Unicité : Comme f est strictement convexe sur U , l'unicité est une conséquence immédiate du théorème 1.5. $\square$

Remarque 1.11. *On utilisera très souvent ce corollaire avec $\Omega = \mathbb{R}^n$ et $f : \mathbb{R}^n \rightarrow R$ une fonction de classe C^1 et fortement convexe sur $\mathbb{R}^n$.*

Alors pour tout ensemble $U \subset \mathbb{R}^n$ convexe et fermé on a l'existence et l'unicité d'un point minimum de f sur U .

Chapitre 2

Optimisation quadratique convexe avec contraintes et sans contraintes

2.1 Applications quadratiques

Définition 2.1. Une application $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est quadratique lorsque c'est un polynôme de degré 2.

Une application quadratique est de la forme :

$$f(x_1, x_2, \dots, x_n) = \underbrace{\frac{1}{2} \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i < j} a_{ij} x_i x_j}_{\text{forme quadratique}} - \underbrace{\sum_{i=1}^n b_i x_i}_{\text{forme linéaire}} + \underbrace{c}_{\text{constante}}$$

En posant :

$$A = (a_{ij})_{i=1, \dots, n, j=1, \dots, n} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \in M_n(\mathbb{R})$$

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^n$$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

on écrit l'application quadratique sous forme matricielle :

$$f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c$$

ou encore :

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c$$

Théorème 2.1. Soit $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c$ une application quadratique. Alors f vérifie les 3 points suivant :

1. f est infiniment différentiable,
2. $\nabla f(x) = Ax - b$,
3. $\nabla^2 f(x) = A$.

Démonstration.

f est polynomiale donc infiniment différentiable.

Pour tout $i = 1, 2, \dots, n$ on a $\frac{\partial f}{\partial x_i}(x) = \sum_{j=1}^n a_{ij}x_j - b_i$ donc $\nabla f(x) = Ax - b$.

Pour tout $i, j = 1, 2, \dots, n$, $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = a_{ij}$, et l'on obtient $\nabla^2 f(x) = A$.

□

Théorème 2.2. [3]

Soit $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c$.

avec $A \in M_n(\mathbb{R})$ une matrice carrée réelle et symétrique de taille n , avec $b \in \mathbb{R}$ un vecteur et $c \in \mathbb{R}$ un scalaire. On calcule facilement :

$$\nabla f(x) = Ax - b$$

$$\nabla^2 f(x) = A$$

(donc la hessienne de f est constante).

En utilisant les rappels de la section 1.5. et les propositions 1.3., 1.4. et 1.5. on déduit :

1. f est une fonction convexe $\iff A$ est une matrice semi définie positive
2. f fortement convexe $\iff f$ est strictement convexe $\iff A$ est une matrice définie positive.

Théorème 2.3. Soit $f(x) = \frac{1}{2} \langle Ax, x \rangle - \langle b, x \rangle + c$ une application quadratique sur $\mathbb{R}^n$ et $u \in \mathbb{R}^n$.

Si A est semi-définie positive (resp. négative), alors les propositions suivantes sont équivalentes :

1. u est un minimum (resp. maximum) local de f .
2. u est un minimum (resp. maximum) global de f .
3. $Au = b$, i.e. u est solution du système d'équation linéaires $Ax = b$

Si A est définie positive f admet un unique minimum (resp. maximum) global.

Si A n'est pas semi-définie positive (resp. négative) alors f n'admet aucun minimum (resp. maximum) local ou global.

Démonstration.

Puisque A est semi-définie positive (resp. négative), f (resp. $-f$) est convexe et l'équivalence des 3 assertions découle des théorèmes 1.5, 1.2 et 2.1.

Si A est définie positive, $\det(A) \neq 0$, le système $Au = b$ est de Cramer et f admet donc un unique extremum.

Si A n'est pas semi-définie positive (resp. négative) la condition nécessaire du second ordre théorème 1.3. et le théorème 2.1 impliquent que f n'a aucun minimum (resp. maximum) local ni donc aussi global. $\square$

2.2 Optimisation sans contraintes

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction continue. On définit un problème d'optimisation sans contraintes par :

$$\min_{x \in \mathbb{R}^n} f(x).$$

où bien

$$\max_{x \in \mathbb{R}^n} f(x).$$

(Rappelons que $\min f(x) = -\max -f(x)$).

Définition 2.2. On dit que x^* est un minimum local si $f(x) \geq f(x^*)$ pour tout $x \in V(x^*)$.

On dit que x^* est un minimum global si $f(x) \geq f(x^*)$ pour tout $x \in \mathbb{R}^n$.

Théorème 2.4. (Condition nécessaire du 1^{er} ordre)
 x^* est un minimum local $\Rightarrow \nabla f(x^*) = 0$.

Démonstration. On a $f(x) = f(x^*) + \nabla f(x^*)(x - x^*) + o(\|x\|)$ pour tout $x \in V(x^*)$.

On pose $x = x^* + t.d$, avec $t > 0$ et $d \in \mathbb{R}^n$. On aura donc :

$$f(x^* + t.d) = f(x^*) + \nabla f(x^*)t.d + o(\|t.d\|).$$

i.e

$$0 \leq \frac{f(x^* + t.d) - f(x^*)}{t} = \nabla f(x^*).d + \frac{o(\|t.d\|)}{t}$$

Par passage à la limite lorsque $t \rightarrow 0$ (car f est continue) on aura :

$$0 \leq \lim_{t \rightarrow 0} \frac{f(x^* + t.d) - f(x^*)}{t} = \nabla f(x^*).d + \lim_{t \rightarrow 0} \frac{o(\|t.d\|)}{t}.$$

D'où le résultat. $\square$

Remarque 2.1. L'implication inverse est fausse i.e $\nabla f(x^*) = 0 \not\Rightarrow x^*$ minimum local.

Remarque 2.2. Les points vérifiant $\nabla f(x) = 0$ sont appelés points critiques (stationnaires).

Parmis les points critiques : minimum, maximum, point selle.

Théorème 2.5. (Condition suffisante du 1^{er} ordre)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexe et de classe C^1 alors on a l'équivalence suivante : x^* est un minimum global $\Leftrightarrow \nabla f(x^*) = 0$.

Démonstration. $\Rightarrow$) est déjà faite.

$\Leftarrow$) On va montrer que $\nabla f(x^*) = 0^*$ est un minimum global. En effet, comme f est convexe, $f(x) \geq f(x^*) + \nabla f(x^*) \cdot (x - x^*)$ pour tout $x \in \mathbb{R}^n$.

Donc x^* est un minimum global. $\square$

Théorème 2.6. (Condition nécessaire du 2nd ordre)

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de classe C^2 .

Si x^* est un minimum local alors $\begin{cases} \nabla f(x^*) = 0 \\ H_{f(x^*)} S.D.P \end{cases}$

Démonstration. On a pour tout $x \in V(x^*)$;

$$f(x) = f(x^*) + \nabla f(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^T H_{f(x^*)}(x - x^*) + o(\|x - x^*\|).$$

On pose $x = x^* + t.d$, avec $t > 0$ et $d \in \mathbb{R}^n$. On aura donc :

$$f(x^* + t.d) = f(x^*) + \nabla f(x^*)t.d + \frac{dt}{2}t.d^T H_{f(x^*)}t.d + o(\|t.d\|^2).$$

D'où :

$$0 \leq \frac{f(x^* + t.d) - f(x^*)}{t^2} = \frac{1}{2}d^T H_{f(x^*)}d + \frac{o(\|t.d\|^2)}{t^2}.$$

Par passage à la limite lorsque t tend vers 0 on aura :

$$0 \leq \lim_{t \rightarrow 0} \frac{f(x^* + t.d) - f(x^*)}{t^2} = \frac{1}{2}d^T H_{f(x^*)}d + \lim_{t \rightarrow 0} \frac{o(\|t.d\|^2)}{t^2}.$$

D'où le résultat. $\square$

Théorème 2.7. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de classe C^2 .

Si x^* est un maximum local alors $\begin{cases} \nabla f(x^*) = 0 \\ H_{f(x^*)} S.D.N \end{cases}$

Exemple 2.1. Considérons la fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par

$$f(x, y) = x^2 - y^2.$$

On a :
 $f \in C^2(\mathbb{R}^2)$.

$$\nabla f(x^*, y^*) = 0 \Leftrightarrow \begin{pmatrix} 2x^* \\ -2y^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} x^* \\ y^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Et comme : $H_{f(0;0)} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$ n'est pas définie positive ce qui nous confirme que le point $(0;0)$ n'est pas un minimum local de la fonction f .
De même comme la hessienne n'est pas définie négative donc le point $(0;0)$ n'est pas un maximum local; et on conclut que le point $(0;0)$ est un point selle de la fonction f .

Théorème 2.8. (Condition suffisante du 2nd ordre) Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de classe C^2 .

(1) Soit x^* vérifiant $\begin{cases} \nabla f(x^*) = 0 \\ H_{f(x^*)} D.P \end{cases} \Rightarrow x^*$ est un minimum local.

(2) Soit x^* vérifiant $\begin{cases} \nabla f(x^*) = 0 \\ H_{f(x^*)} D.N \end{cases} \Rightarrow x^*$ est un maximum local.

Démonstration. Il suffit de montrer la (1); en effet grâce au développement de Taylor au voisinage du point x^* on obtient :

$$f(x^* + td) = f(x^*) + \nabla f(x^*)t.d + \frac{1}{2}t.d^T H_{f(x^*)}t.d + o(\|t.d\|^2).$$

Comme $\nabla f(x^*) = 0$ par hypothèse et comme $o(\|t.d\|^2) \longrightarrow 0$ lorsque $t \rightarrow 0$ on aura :

$$\text{sgn}\{f(x^* + td) - f(x^*)\} = \text{sgn}\left\{\frac{1}{2}t.d^T H_{f(x^*)}t.d\right\} = \text{positif}.$$

Ce qui nous donne : $f(x^* + td) \geq f(x^*)$ pour tout $d \in V(0)$ qui est équivalent à :

$f(x) \geq f(x^*)$ pour tout $x \in V(x^*)$. Et cela implique que x^* est un minimum local.

De la même manière on montre la (2). $\square$

Exemple 2.2. Considérons la fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par

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

On a :

$f \in C^2(\mathbb{R}^2)$.

Le point $(0;0)$ est un point critique car :

$$\nabla f(x^*, y^*) = 0 \Leftrightarrow \begin{pmatrix} 2x^* \\ 2y^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} x^* \\ y^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Et comme $H_{f(0;0)} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ est définie positive ; on conclut que le point $(0;0)$ est un minimum local (on peut montrer qu'il est minimum global).

2.3 Optimisation avec contraintes d'égalité

On définit un problème d'optimisation avec contraintes d'égalité par :

$$(P) \begin{cases} \min f(x) \\ g_i(x) = 0 \text{ avec } i = 1; 2; \dots; m \\ x \in \mathbb{R}^n \end{cases}$$

Exemple 2.3. Soit le problème suivant :

$$\begin{cases} \min \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i = 1 \\ x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \end{cases} \Leftrightarrow \begin{cases} \min \sum_{i=1}^n x_i^2 \\ g(x) = \sum_{i=1}^n x_i - 1 = 0 \\ x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \end{cases}$$

Ce problème revient à minimiser la distance d'un point

$$x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

se trouvant sur l'hyperplan par rapport à l'origine.

$$\text{Rappelons que } G : \mathbb{R}^n \longrightarrow \mathbb{R}^m \text{ telle que } G(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{pmatrix} \in \mathbb{R}^m$$

La matrice Jacobienne de G est :

$$JG(x) = \begin{pmatrix} \nabla g_1(x)^T \\ \nabla g_2(x)^T \\ \vdots \\ \nabla g_m(x)^T \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(x) & \cdots & \frac{\partial g_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1}(x) & \cdots & \frac{\partial g_m}{\partial x_n}(x) \end{pmatrix}$$

Et supposons que $m \leq n$.

Définition 2.3. (Point régulier) Le point x^* est régulier admissible pour le problème (P) i.e vérifie la qualification de contrainte si $\text{rg } JG(x^*) = m$.

Exemple 2.4. Pour le problème de l'exemple précédent le point $x^* = (1, 0, \dots, 0)$ est un point admissible. et aussi régulier car $JG(x^*) = \nabla g(x^*) = (1, 1, \dots, 1)$. $\text{rg } JG(x^*) = 1$ donc x^* est régulier.

Pour tout x point admissible tel que $\text{rg } JG(x^*) = 1$, on dira dans ce cas la qualification de contrainte est vérifié partout.

Exemple 2.5. Pour le problème :

$$\begin{cases} \min \sum_{i=1}^n x_i^2 \\ x_1^2 + x_2^2 - x_3^2 = 0 \\ x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \end{cases}$$

La Jacobienne est : $JG(x) = (2x_1, 2x_2, -2x_3)$.

Soit le point $x^* = (0, 0, 0)$, on a : $JG(x^*) = (0, 0, 0)$ donc $\text{rg } JG(x^*) = 0$ d'où le point x^* n'est pas régulier.

Proposition 2.1. [6] Soit x^* un minimum local régulier du problème (P) et $G(x) = 0$ alors les m premières composante de x s'exprime en fonction des $n - m$ composante restante.

Théorème 2.9. Théorème de Lagrange (condition nécessaire du 1^{er} ordre)[6] Soit $f, g \in C^1$, soit x^* un minimum régulier du problème (P) alors :

$$\exists \lambda_1^*, \dots, \lambda_m^* \in \mathbb{R} \text{ tels que } \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0$$

$$\lambda_i^* \nabla g_i(x^*) = 0.$$

Les nombres réels λ_i^* ; $i = 1, 2, \dots, n$ sont appelés multiplicateurs de Lagrange.

Définition 2.4. On appelle fonction de Lagrange ou bien le Lagrangien associé au problème (P) la fonction :

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x).$$

Proposition 2.2. [6] Soit x^* un minimum régulier du problème (P) alors (x^*, λ^*) est un point critique du Lagrangien.

Exemple 2.6. Soit le problème :

$$\begin{cases} \min y \\ y^3 = x^2 \\ (x, y) \in \mathbb{R}^2 \end{cases}$$

On a :

$$L(x, y, \lambda) = y + \lambda(y^3 - x^2).$$

On résoud le système : $\nabla L(x, y, \lambda) = 0$ i.e on résoud :

$$\begin{cases} -2\lambda x = 0 \\ 1 + 3\lambda y^2 = 0 \\ y^3 - x^2 = 0 \end{cases}$$

Il n'existe pas un multiplicateur de Lagrange qui vérifie le système ce qui implique que le problème ne possède pas de minimum régulier.

Remarquons que le problème possède un minimum $(0, 0)$ mais il n'est pas régulier car : $JG(x^*, y^*) = (-2x^*, 3y^{*2}) = (0, 0)$.

et $\text{rg } JG(x^*, y^*) = 0$. D'où le point $(0, 0)$ n'est pas régulier, il est dit singulier.

Théorème 2.10. (condition suffisante du 1^{er} ordre)[6]

Soit $f, g \in C^1$, Supposons que le problème (P) soit convexe et :

$$\exists \lambda_1^*, \dots, \lambda_m^* \in \mathbb{R} \text{ tels que } \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0.$$

Alors x^* un minimum global de (P)

Définition 2.5. On appelle sous espace tangent aux contraintes du problème (P) au point x^0 l'espace :

$$T(x^0) = \{d \in \mathbb{R}^n : JG(x^0).d = 0\}.$$

Exemple 2.7. Soit la contrainte $x_1^2 + x_2^2 = 1$ alors on a :

$$T(x^0) = \{d = (d_1, d_2) \in \mathbb{R}^2 : (2x_1^0, 2x_2^0) \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = 0\}.$$

i.e

$$T(x^0) = \{d_1, d_2) \in \mathbb{R}^2 : x_1^0 d_1 + x_2^0 d_2 = 0\}.$$

On aura :

$$T(x^0) = \{(d_1, d_2) \in \mathbb{R}^2 : d_2 = -\frac{x_1^0}{x_2^0} d_1\}.$$

D'où :

$$T(x^0) = \{d_1(1, -\frac{x_1^0}{x_2^0}) , d_1 \in \mathbb{R}\}.$$

Théorème 2.11. (Condition nécessaire du 2nd ordre)[7]

On suppose que f et les g_i sont de classe C^2 et soit x^* un minimum régulier du problème (P) , alors :

- (1) $\exists \lambda_1^*, \dots, \lambda_m^* \in \mathbb{R}$ tels que $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0$.
- (2) $H_x L(x^*, \lambda^*)$ est S.D.P sur $T(x^*)$.

Remarque 2.3. Pour un problème de maximisation la (2) devient $H_x L(x^*, \lambda^*)$ est S.D.N sur $T(x^*)$.

Théorème 2.12. (Condition suffisante du 2nd ordre)[7]

Soit f et les g_i sont de classe C^2 et x^* un point régulier vérifie :

- (1) $\exists \lambda_1^*, \dots, \lambda_m^* \in \mathbb{R}$ tels que $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0$.
- (2) $H_x L(x^*, \lambda^*)$ est D.P sur $T(x^*)$.

Alors : x^* est un minimum local du problème (P).

Exemple 2.8. Soit un parallélépipède d'arrêtes qui ont des longueurs $x; y; z$. on note par V son volume et par S sa surface. On veut résoudre le problème d'optimisation suivant :

$$(\Theta) \begin{cases} V(x, y, z) = xyz \longrightarrow \max \\ S(x, y, z) = 2xy + 2yz + 2xz \\ x > 0; y > 0; z > 0 \end{cases}$$

Ce problème est équivalent à :

$$(\Theta^*) \begin{cases} f(x, y, z) = -xyz \longrightarrow \min \\ h_1(x, y, z) = 2(xy + yz + xz) - S \\ x > 0; y > 0; z > 0 \end{cases}$$

La fonction de Lagrange est :

$$L(x, y, z; \lambda) = -xyz + \lambda[2(xy + xz + yz) - S].$$

On résout le système

$$\nabla L(x, y, z; \lambda) = 0 \Leftrightarrow \begin{pmatrix} -yz + 2\lambda(y + z) \\ -xz + 2\lambda(x + z) \\ -xy + 2\lambda(x + y) \\ 2(xy + xz + yz) - S \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

On aura après la résolution : $x = y = z = \sqrt{\frac{S}{6}}$ et $\lambda = \frac{x}{4}$. D'où le point critique admissible est :

$$(x^*, y^*, z^*; \lambda^*) = (\sqrt{\frac{S}{6}}, \sqrt{\frac{S}{6}}, \sqrt{\frac{S}{6}}; \frac{1}{4}\sqrt{\frac{S}{6}}).$$

On a :

$$HL(x^*, y^*, z^*; \lambda^*) = \begin{pmatrix} 0 & -z^* + 2\lambda^* & -y^* + 2\lambda^* \\ -z^* + 2\lambda^* & 0 & -x^* + 2\lambda^* \\ -y^* + 2\lambda^* & -x^* + 2\lambda^* & 0 \end{pmatrix} = -\frac{1}{2}\sqrt{\frac{S}{6}} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

On a les valeurs propres de la matrice $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ sont -1 et 2 car le polynôme caractéristique

$$P_\lambda = -\lambda(\lambda^2 - 1) - (-\lambda - 1) + (1 - \lambda) = -(\lambda + 1)^2(\lambda - 2).$$

Le plan tangent :

$$T((x^*, y^*, z^*)) = \{d = (d_1, d_2, d_3) \in \mathbb{R}^3 : JG(x^*, y^*, z^*).d = 0\}.$$

$$T((x^*, y^*, z^*)) = \{d = (d_1, d_2, d_3) \in \mathbb{R}^3 : JG(2y^* + 2z^*, 2z^* + 2x^*, 2x^* + 2y^*). \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = 0\}.$$

$$T((x^*, y^*, z^*)) = \{d = (d_1, d_2, d_3) \in \mathbb{R}^3 : 4\sqrt{\frac{S}{6}}(1, 1, 1). \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = 0\}.$$

$$T((x^*, y^*, z^*)) = \{d = (d_1, d_2, d_3) \in \mathbb{R}^3 : d_1 + d_2 + d_3 = 0\} = \{(-d_2 - d_3, d_2, d_3) \in \mathbb{R}^3\}.$$

Donc on aura :

$$T((x^*, y^*, z^*) = \{d_2(-1, 1, 0) + d_3(-1, 0, 1) \text{ avec } d_2, d_3 \in \mathbb{R}\}.$$

D'où : $T(x^*, y^*, z^*)$ est un sous espace vectoriel de $\dim = 2$ et de base $\{(-1, 1, 0); (-1, 0, 1)\}$.

$$\text{La matrice du passage est : } \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

On aura :

$$P^T (-HL(x^*, y^*, z^*)) P = \begin{pmatrix} -2 & -1 \\ -1 & -1 \end{pmatrix}.$$

Comme

$$\begin{cases} A_1 = -2 < 0 \\ A_2 = (-2) \cdot (-1) - (-1)(-1) = 1 > 0 \end{cases}$$

On aura : Donc d'après le critère de Sylvester on conclut que la matrice $A = \begin{pmatrix} -2 & -1 \\ -1 & -1 \end{pmatrix}$ est D.N

Ce qui implique que $HL(x^*, y^*, z^*)$ est D.P sur $T(x^*, y^*, z^*)$ donc le point (x^*, y^*, z^*) est un minimim de la fonction f définie par $f(x, y, z) = -xyz$ et cela prouve que (x^*, y^*, z^*) soit un maximum du volume $V = xyz$.

2.4 Optimisation avec contraintes d'inégalité

On définit un problème d'optimisation avec contraintes d'inégalité par :

$$(P_1) \begin{cases} \min f(x) \\ g_i(x) \leq 0 \text{ avec } i = 1; 2; \dots m \\ x \in \mathbb{R}^n \end{cases}$$

Définition 2.6. La contrainte $g_i(x^0)$ est dite active (saturée) en x^0 si $g_i(x^0) = 0$. On note l'ensemble des contraintes active en x^0 l'ensemble

$$E(x^0) = \{i \text{ tels que } g_i(x^0) = 0\}.$$

Définition 2.7. Un point x^0 admissible est dit régulier si

$$rg JG(x^0) = \text{card } E(x^0) = |E(x^0)|$$

avec G formée des contraintes saturées.

Exemple 2.9. Soit le problème suivant :

$$\begin{cases} \min x_1^2 + x_2^2 \\ x_1 + x_2 \leq -1 \leftarrow g_1(x) = x_1 + x_2 + 1 \\ x_1 - x_2 \leq 1 \leftarrow g_2(x) = x_1 - x_2 - 1 \\ (x_1, x_2) \in \mathbb{R}^2 \end{cases}$$

Le point $x^0 = (0, -1)$ est admissible. Et aussi le point x^0 est régulier car : $E(x^0) = \{1, 2\}$.

$$rg JG(x^0) = rg \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = rg \begin{pmatrix} g_1(x) \\ g_2(x) \end{pmatrix} = 2 = |E(x^0)|.$$

D'où x^0 est régulier vérifie Q.C (la qualification des contraintes).

Résultats : 1) Si les fonctions g_i sont linéaires alors Q.C (qualification de contrainte) est vérifié partout. (Karlin 1959)

2) Si les g_i sont convexes et s'il existe un $\bar{x}$ tel que $g_i(\bar{x}) < 0$ quelque soit i ; alors Q.C est vérifiée partout sur le domaine admissible. (Slater 1950)

Théorème 2.13. *Krauch-Kuhn-Tucker (K.K.T) (Condition nécessaire du 1^{er} ordre)[2]*

On suppose que $f; g_i$ avec $i = 1, 2, \dots, m$ sont de classe C^1 et soit x^* un minimum local régulier du problème (P_1) alors :

$$\exists \lambda_i^* (i = 1; 2; \dots; m) \text{ tels que : } \begin{cases} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0 \\ \lambda_i^* g_i(x^*) = 0 ; i = 1; \dots; m \end{cases}$$

Démonstration. Il suffit de se ramener à un problème avec contrainte égalité

$$g_i(x) + x_{n+i} = 0 \text{ avec } i = 1; \dots; m.$$

et appliquer le théorème 2.9. $\square$

Théorème 2.14. *(Condition nécessaire du 2nd ordre)[7]*

On suppose que f et les g_i sont de classe C^2 et soit x^* un minimum régulier du problème (P) , alors :

$$(1) \exists \lambda_1^*, \dots, \lambda_m^* \in \mathbb{R} \text{ tels que } \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0.$$

$$\lambda_i^* g_i(x^*) = 0 ; i = 1; \dots; m.$$

$$(2) H_x L(x^*, \lambda^*) \text{ est S.D.P sur } T(x^*).$$

Théorème 2.15. (Condition suffisante du 2nd ordre)[7]

Soit f et les g_i sont de classe C^2 et x^* un point régulier vérifie :

$$(1) \exists \lambda_1^*, \dots, \lambda_m^* \in \mathbb{R} \text{ tels que } \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) = 0.$$

$$\lambda_i^* g_i(x^*) = 0 ; i = 1; \dots; m.$$

$$(2) H_x L(x^*, \lambda^*) \text{ est D.P sur } T(x^*).$$

Alors : x^* est un minimum local du problème (P).

Définition 2.8. On définit un problème d'optimisation avec contrainte mixte (égalité et inégalité) par :

$$(P_2) \begin{cases} \min f(x) \\ g_i(x) \leq 0 \text{ avec } i = 1; 2; \dots m \\ h_j(x) = 0 \text{ avec } j = 1; 2; \dots p \\ x \in \mathbb{R}^n \end{cases}$$

Définition 2.9. On définit le gradient des contraintes actives pour les fonctions g_i au point x^0 par :

$$\nabla_{E(x^0)} g_i(x^0).$$

Définition 2.10. Le point admissible x^0 est dit régulier si les vecteurs $\nabla_{E(x^0)} g_i(x^0)$ et $\nabla h_j(x^0)$ avec $i \in E$ et $j = 1; \dots, p$ sont linéairement indépendants c'est à dire :

$$rg \begin{pmatrix} JG_E(x^0) \\ JH(x^0) \end{pmatrix} = p + |E|$$

Exemple 2.10. Soit le problème :

$$\begin{cases} \min f(x) \\ x_1 + x_2 \leq 1 \\ x_1 + 2x_2 = 2 \\ x_1; x_2 \in \mathbb{R} \end{cases}$$

On a : $g_1(x) = x_1 + x_2 - 1$ et $h_1(x) = x_1 + 2x_2 - 2$.

Le point $x^0 = (0, 1)$ est admissible.

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

$\nabla g_1(x) = (1, 1)$ et $\nabla h_1(x) = (1, 2)$, donc on aura :

$$rg \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = |E| + p = 1 + 1$$

Théorème 2.16. (*Condition nécessaire du 1^{er} ordre pour un problème mixte*)[2]

*On suppose que $f; g_i, h_j$ sont de classe C^1 , ($i = 1; \dots; m$ et $j = 1; \dots; p$)
et soit x^* un minimum local régulier du problème (P_2) alors :*

*$\exists \lambda_i^*$ avec $i = 1; \dots; m \in \mathbb{R}^+$ (appelés multiplicateur K.K.T) et
 $\exists \mu_j; j = 1; \dots; p \in \mathbb{R}$ tels que :*

$$\left\{ \begin{array}{l} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^p \mu_j \nabla h_j(x^*) = 0 \\ \lambda_i^* g_i(x^*) = 0 \text{ avec } i = 1; \dots; m \end{array} \right.$$

Chapitre 3

Régression linéaire et méthode de Markowitz

3.1 Introduction à la 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 :

$$b = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n = \sum_{j=1}^n \alpha_j v_j$$

Les α_j sont des réels appelés coefficients du modèle.

3.1.1 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, $i = 1, \dots, m$:

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

soit le plus proche possible de b_i .

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 :

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

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 $d(\hat{b} = Ax, b)$ soit minimale, ce qui s'écrit :

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

3.1.2 Forme standard

Définition 3.1. On appelle forme standard d'un problème de moindres carrés la donnée :

$$\text{— la matrice } A = \begin{pmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,n} \\ v_{2,1} & v_{2,2} & \dots & v_{2,n} \\ \vdots & \vdots & & \vdots \\ v_{m,1} & v_{m,2} & \dots & v_{m,n} \end{pmatrix}$$

appelée matrice des données.

$$\text{— le vecteur réponse } b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$$

— l'expression du critère :

$$\text{on cherche } x = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} \in \mathbb{R}^m$$

réalisant :

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

Exemple 3.1. On s'intéresse à une quantité physique z qui varie selon un paramètre physique t . On pose le modèle :

$$z = f(t) = a + bt + ct^2 + dt^3$$

La variable expliquée est z , les variables explicatives sont t, t^2, t^3 et «1» et les paramètres du modèle, que l'on cherche, sont a, b, c et d . On réalise une série de m mesure $(t_i, z_i)_{1 \leq i \leq m}$.

La forme standard du problème de moindres carrés correspondant est :

$$\text{On cherche } x = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \text{ réalisant :}$$

$$\min_{x \in \mathbb{R}^4} \|Ax - b\|^2$$

où

$$\begin{pmatrix} 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 & t_m^3 \end{pmatrix} \in M_{m,n} \text{ et } b = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} \in \mathbb{R}^m$$

3.1.3 Recherche d'une solution

On fait l'hypothèse que les variables explicatives sont linéairement indépendantes (i.e. $rg(A) = n$)

Soit $E(x) = \|Ax - b\|^2$ la fonction erreur. Selon la théorie d'optimisation, la valeur minimale se trouve par la résolution de l'équation $\nabla E(x) = 0$.

Théorème 3.1. [2] Soit $A \in M_{m,n}$ avec $m > n$ et $b \in \mathbb{R}^n$. Une condition nécessaire et suffisante pour que $x \in \mathbb{R}^n$ réalise le minimum de $E(x) = \|Ax - b\|^2$ est que :

$$A^T Ax = A^T b \quad (1)$$

Les équations (1) sont appelées équations normales elles correspondent aux conditions du premier ordre de $E(x)$. Elles sont obtenues en dérivant $E(x)$ 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.

3.1.4 Série statistique à deux variables

Définition 3.2. On appelle **série statistique à deux variables** (ou série statistique doubles) une série statistique où deux caractères sont étudiés **simultanément**.

Définition 3.3 (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

Si les points sont plutôt alignés, on peut envisager une relation de type $Y = aX + b$ (équation de droite)

Si le nuage forme une parabole, on peut envisager une relation de type $Y = aX^2 + bX + c$.

Définition 3.4 (Point moyen). Le point moyen d'un nuage de points est le point G de coordonnées $(\bar{x}; \bar{y})$ où :

1. $\bar{x}$ représente la moyenne des x_i :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

2. $\bar{y}$ représente la moyenne des y_i :

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{n} \sum_{i=1}^n y_i$$

Définition 3.5 (Ajustement d'une série statistique à deux variables). Effectuer un ajustement de y en x d'un nuage de points consiste à trouver une fonction f telle que la courbe d'équation $y = f(x)$ passe au plus près des points du nuage.

3.1.5 Ajustement par la méthode des moindres carrés

Définition 3.6 (Principe de la méthode). *Effectuer un ajustement de y en x d'un nuage de points par la méthode des moindres carrés consiste à trouver la fonction f du modèle retenu qui minimise la somme des carrés des écarts entre les valeurs observées y_i à celles données $\hat{y} = ax_i + b$ par la droite $y = ax + b$.*

Si e_i désigne cet écarts des valeurs observées, appelé aussi résidu, pour le i -ème point, on cherche les valeurs de a et b qui minimise la somme des carrés des écarts :

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - ax_i + b)^2 = \sum_{i=1}^n (y_i - f(x_i))^2 .$$

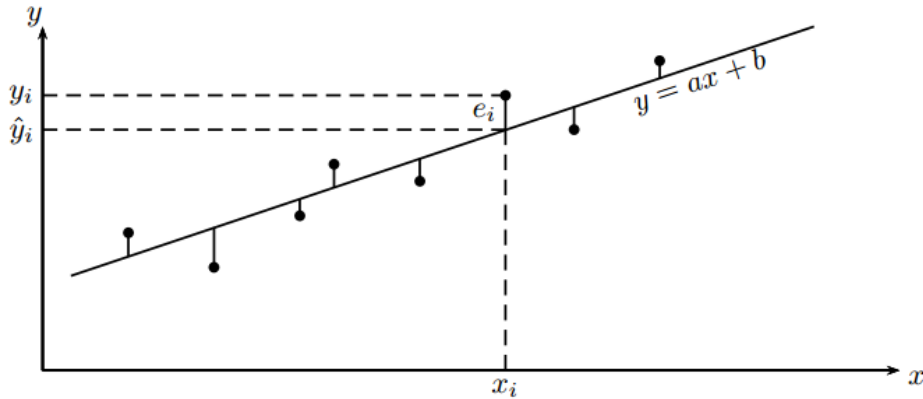


FIG.1 - Méthode des moindres carrés.

Remarque 3.1. 1. *Pour une valeur x_0 donnée du caractère x , la fonction f permet donc de prévoir le résultat correspondant de la variable y . On supposera que $y_0 = f(x_0)$.*

2. *Si x_0 est compris entre x_1 et x_n , on parle d'interpolation.*

3. *Si x_0 est en dehors de l'intervalle d'observation du caractère x , on parle d'extrapolation.*

3.1.6 Ajustement affine par la méthode des moindres carrés de y en x

Définition 3.7. On appelle covariance de x et de y le nombre :

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

Définition 3.8. La variance du caractère x est :

$$V(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2$$

$$V(y) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - (\bar{y})^2$$

Elle est utilisée pour le calcul de l'écart type : $\sigma(x) = \sqrt{V(x)}$.

Remarque 3.2. $V(x) = \text{cov}(x, x)$

Théorème 3.2. Lors d'un ajustement affine par la méthode des moindres carrés, la droite $D_{y/x} : y = ax + b$ servant à l'ajustement de y en x a pour coefficients :

$$a = \frac{\text{cov}(x, y)}{V(x)} \text{ et } b = \bar{y} - a\bar{x}$$

Démonstration.

soit $y = ax + b$ une droite $D_{x/y}$ à ajuster entre les points du nuage $M_i(x_i, y_i)$ et les points $P_i(x_i, ax_i + b)$ correspondant sur la droite d'ajustement.

On cherche donc le couple $(\hat{a}, \hat{b})$ qui minimise $f(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2$.

Conditions nécessaires de minimum :

$$\frac{\partial f}{\partial a}(\hat{a}, \hat{b}) = 0 \text{ et } \frac{\partial f}{\partial b}(\hat{a}, \hat{b}) = 0$$

$$\text{Soit : } \sum_{i=1}^n 2(y_i - \hat{a}x_i - \hat{b})(-x_i) = 0 \text{ et } \sum_{i=1}^n 2(y_i - \hat{a}x_i - \hat{b})(-1) = 0,$$

$$\sum_{i=1}^n (y_i - \hat{a}x_i - \hat{b})x_i = 0 \text{ et } \sum_{i=1}^n (y_i - \hat{a}x_i - \hat{b}) = 0,$$

$$\sum_{i=1}^n y_i x_i - \hat{a} \sum_{i=1}^n x_i^2 - \hat{b} \sum_{i=1}^n x_i = 0 \text{ et } \sum_{i=1}^n y_i - \hat{a} \sum_{i=1}^n x_i - n\hat{b} = 0,$$

En divisant par n , on obtient :

$$\frac{1}{n} \sum_{i=1}^n y_i x_i - \hat{a} \frac{1}{n} \sum_{i=1}^n x_i^2 - \hat{b} \frac{1}{n} \sum_{i=1}^n x_i = 0,$$

$$\text{et } \frac{1}{n} \sum_{i=1}^n y_i - \hat{a} \frac{1}{n} \sum_{i=1}^n x_i - \hat{b} = 0,$$

On a alors :

$$\text{cov}(x, y) + \bar{x}\bar{y} - \hat{a}(V(x) + \bar{x}^2) - \hat{b}\bar{x} = 0 \text{ et } \bar{y} - \hat{a}\bar{x} - \hat{b} = 0$$

De $\hat{b} = \bar{y} - \hat{a}\bar{x}$ on déduit que :

$$\text{cov}(x, y) + \bar{x}\bar{y} - \hat{a}(V(x) + \bar{x}^2) - \bar{y} + (\hat{a}\bar{x})\bar{x} = \text{cov}(x, y) - \hat{a}V(x) = 0$$

$$\text{donc } \hat{a} = \frac{\text{cov}(x, y)}{V(x)}$$

Pour vérifier qu'il s'agit bien d'un minimum, on doit montrer que la matrice hessienne des dérivées secondes est définie positive. Cette matrice vaut :

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 b}(\hat{a}, \hat{b}) & \frac{\partial^2 f}{\partial a \partial b}(\hat{a}, \hat{b}) \\ \frac{\partial^2 f}{\partial b \partial a}(\hat{a}, \hat{b}) & \frac{\partial^2 f}{\partial^2 a}(\hat{a}, \hat{b}) \end{pmatrix}$$

On a :

$$\begin{aligned} \frac{\partial^2 f}{\partial^2 b}(\hat{a}, \hat{b}) &= 2n, \quad \frac{\partial^2 f}{\partial^2 a}(\hat{a}, \hat{b}) = 2 \sum_{i=1}^n x_i^2 \\ \frac{\partial^2 f}{\partial a \partial b}(\hat{a}, \hat{b}) &= 2n\bar{x} \end{aligned}$$

La matrice hessienne vaut donc :

$$H = 2 \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

et peut s'écrire $H = 2A^T A$ où :

$$A = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

On admet que ce couple $(\hat{a}, \hat{b})$ minimise $f(a, b)$.

Pour tout vecteur $x \in \mathbb{R}^2$, les formes quadratiques $x^T H x$ peuvent s'écrire $2v^T v$ en posant $v = Ax$. Comme $v^T v$ est toujours positive, la matrice H est définie positive.

La droite des moindres carrés de y en x a donc pour équation :

$$D_{x/y} : y = \hat{a}x + \hat{b}, \text{ avec } \hat{a} = \frac{\text{cov}(x,y)}{V(x)} \text{ et } \hat{b} = \bar{y} - \hat{a}\bar{x}.$$

□

Exemple 3.2. Soit le tableau qui donne la part en % consacrée au logement dans le budget d'un foyer.

Année (x_i)	1991	1997	2005	2007	2013	2017
Part en % (y_i)	4.4	5.2	4.3	3.2	3.3	2.8

TAB. 1 - Part consacré au logement dans un foyer

Le nuage de points associé à cette série statistique est représentée sur la figure 2 :

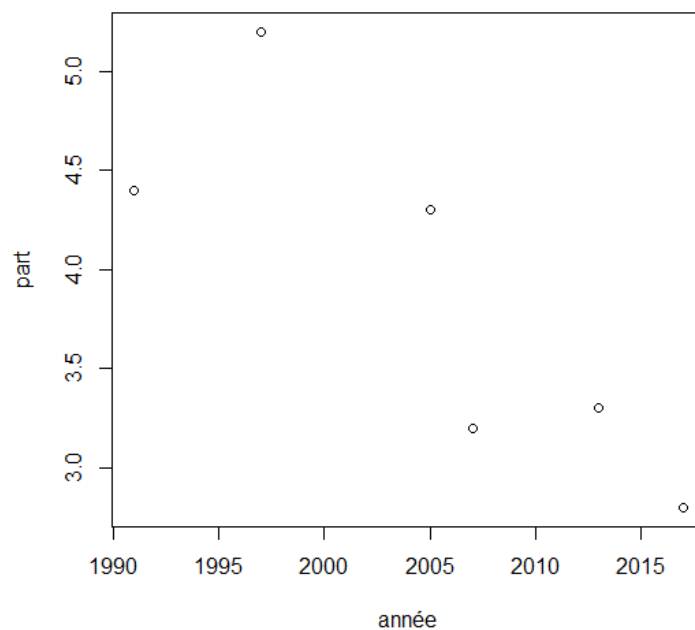


FIG.2 - nuage de points

- Calculer $\bar{x}$ et $\bar{y}$:

$$\bar{x} = \frac{1991+1997+2005+2007+2013+2017}{6} = 2005$$

$$\bar{y} = \frac{4.4+5.2+4.3+3.2+3.3+2.8}{6} \approx 3.87$$

Le point moyen est donc $G(2005; 3.9)$.

- On remplace x par $\hat{x}_i = x - 1991$ ce qui revient à prendre le rang des années :

Année (x_i)	1991	1997	2005	2007	2013	2017
Rang des années (x_i)	0	6	14	16	22	26
Part en % (y_i)	4.4	5.2	4.3	3.2	3.3	2.8

TAB. 2 - Part consacré au logement dans un foyer

Dans ce cas, $\bar{x} = \frac{0+6+14+16+22+26}{6} = 14$ et $\bar{y} = 3.87$

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

le rang(x_i)	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
0	4.4	-14	196	0.53	-7.47
6	5.2	-8	64	1.33	-10.67
14	4.3	0	0	0.43	0
16	3.2	2	4	-0.67	-1.33
22	3.3	8	64	-0.57	-4.53
26	2.8	12	144	-1.07	-12.8

TAB .3 - Calcul de la covariance et de la variance

- La covariance et la variance s'obtient alors par le calcul suivant :

$$\text{cov}(x, y) = \frac{\sum_{i=1}^6 (x_i - \bar{x})(y_i - \bar{y})}{6} \approx -6.13$$

$$V(x) = \frac{(x_i - \bar{x})^2}{6} \approx 68.67$$

- Le coefficient directeur de la droite de régression est donc :

$$a = \frac{\text{cov}(x, y)}{V(x)} \approx \frac{-6.13}{68.67} \approx -0.08$$

- La droite de régression a donc une équation de la forme $y = -0.08x + b$
De plus, elle passe par $G(14; 3.87)$ donc :

$$-0.08 \times 14 + b = 3.87$$

$$-1.12 + b = 3.87$$

$$b = 3.87 + 1.12 = 4.99$$

- L'équation de la droite de régression est donc : $y = -0.08x + 4.99$

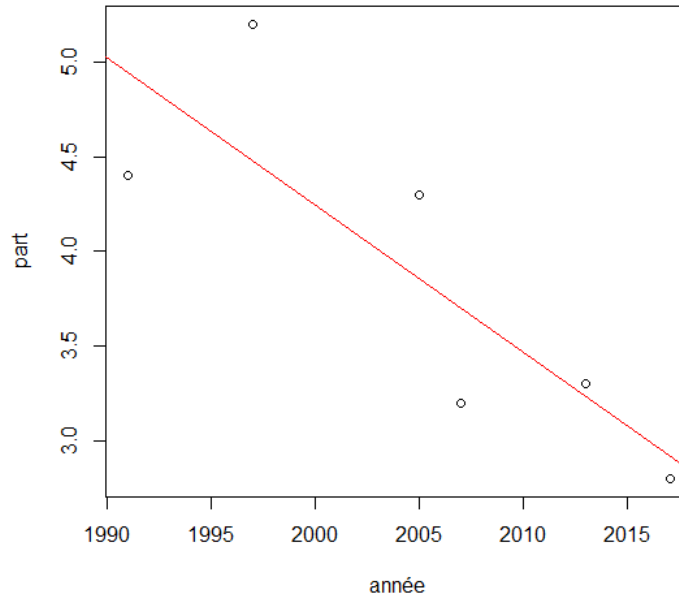


FIG.3 - droite de régression de y en x

3.1.7 Droite des moindres carrés de x en y

On a $D_{y/x} : y = \hat{a}x + \hat{b}$, avec $\hat{a} = \frac{\text{cov}(x,y)}{V(x)}$ et $\hat{b} = \bar{y} - \hat{a}\bar{x}$.

On cherche à ajuster une droite $D_{x/y}$ d'équation $x = a'y + b'$ au nuage de points.

la droite des moindres carrés de x en y est :

$$D_{x/y} : x = \hat{a}'y + \hat{b}', \text{ avec } \hat{a}' = \frac{\text{cov}(x,y)}{V(y)} \text{ et } \hat{b}' = \bar{x} - \hat{a}'\bar{y}.$$

Remarque 3.3. $D_{y/x} : y - \bar{y} = \hat{a}(x - \bar{x})$,

$$D_{x/y} : x - \bar{x} = \hat{a}'(y - \bar{y}).$$

les droites $D_{y/x}$ et $D_{x/y}$ se coupent alors au point $G(\bar{x}, \bar{y})$

3.1.8 Evaluation de la qualité de la régression

Le calcul des deux coefficients $\hat{a}$ et $\hat{b}$ de la droite de régression ne détermine pas si le modèle est acceptable ou ce n'est pas le cas. Pour mesurer la qualité de l'approximation d'un nuage $M_i(x_i, y_i)$ par sa droite des moindres carrés, il est donc indispensable de calculer son coefficient de corrélation linéaire r .

Définition 3.9. On nomme coefficient de corrélation linéaire des variables x et y , le nombre réel r :

$$r = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}, r \in [-1; 1]$$

Interprétation

- Quand $|r| = 1$, tout les points sont alignés.
- Quand $|r|$ est proche de 1, les variables x et y sont fortement corrélées.
- Quand $r < 0$, la droite de régression a une pente négative.
- Quand $r > 0$, la droite de régression a une pente positive.

Définition 3.10 (La formule de décomposition). La notion de liaison entre X et Y signifie qu'une variation de X entraîne une variation de Y .

La formule de décomposition permet de préciser la part de variation de Y expliquée par la variation de X :

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2, \hat{y}_i = \hat{a}x_i + \hat{b}$$

La démonstration repose sur le fait que le double produit s'annule :

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \bar{y}) = \hat{a} \sum_{i=1}^n (x_i - \bar{x})e_i = 0, \text{ avec } e_i = y_i - \hat{y}_i \text{ (erreur observée),}$$

et grâce aux équations définissant $\hat{a}$ et $\hat{b}$.

La somme des carrés totale : $\sum_{i=1}^n (y_i - \bar{y})^2$ mesure la variation globale des y_i autour de leur moyenne $\bar{y}$.

La somme des carrés expliquée par la variable X : $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{a}^2 \sum_{i=1}^n (x_i - \bar{x})^2$ mesure la variation de Y expliquée par la variable X . Ce terme n'est d'ailleurs fonction que de la pente de la droite des moindres carrés et des valeurs de X .

La somme des carrés résiduelle : $\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$ mesure de la variation de Y non expliquée par la variable X .

Définition 3.11 (Coefficient de détermination). Le coefficient de détermination $R^2 \in [0; 1]$ qui est aussi le carré du coefficient de corrélation r permet de mesurer la force entre les variables X et Y qui est donné par :

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\text{somme des carrés expliquée}}{\text{somme des carrés totale}}$$

3.1.9 Position relative de $D_{y/x}$ et $D_{x/y}$

- Si $r_{x,y}^2 = 0$, alors $cov(x, y) = 0$, et donc $\hat{a} = \hat{a}' = 0$. Ainsi, $D_{y/x} : y = \bar{y}$ et $D_{x/y} : x = \bar{x}$

- Si $r_{x,y}^2 \neq 0$, alors $\hat{a} \neq 0$ et $\hat{a}' \neq 0$. On a alors : $D_{y/x} : y - \bar{y} = \hat{a}(x - \bar{x})$ et $D_{x/y} : y - \bar{y} = \frac{1}{\hat{a}'}(x - \bar{x})$

On a : $\hat{a}\hat{a}' = \frac{cov(x,y)}{V(x)} \frac{cov(x,y)}{V(y)} = \left(\frac{cov(x,y)}{V(x)V(y)} \right)^2 = r_{x/y}^2$

- Si $r_{x/y}^2 = 1$, alors $\hat{a} = \frac{1}{\hat{a}'}$ et donc $D_{x/y}$ a pour équation :

$(y - \hat{y}) = \frac{1}{\hat{a}'}(x - \bar{x}) = \hat{a}(x - \bar{x})$, soit l'équation de $D_{y/x}$. Ainsi, $D_{x/y} = D_{y/x}$.

- Si $0 < r_{x,y} < 1$, alors $0 < \hat{a}\hat{a}' < 1$ et deux cas sont possibles :

1. soit $0 < r_{x,y} < 1$ et alors $\hat{a} > 0$, $\hat{a}' > \frac{1}{\hat{a}}$;
2. soit $-1 < r_{x,y} < 1$ et alors $\hat{a} < 0$, $\hat{a}' < \frac{1}{\hat{a}}$.

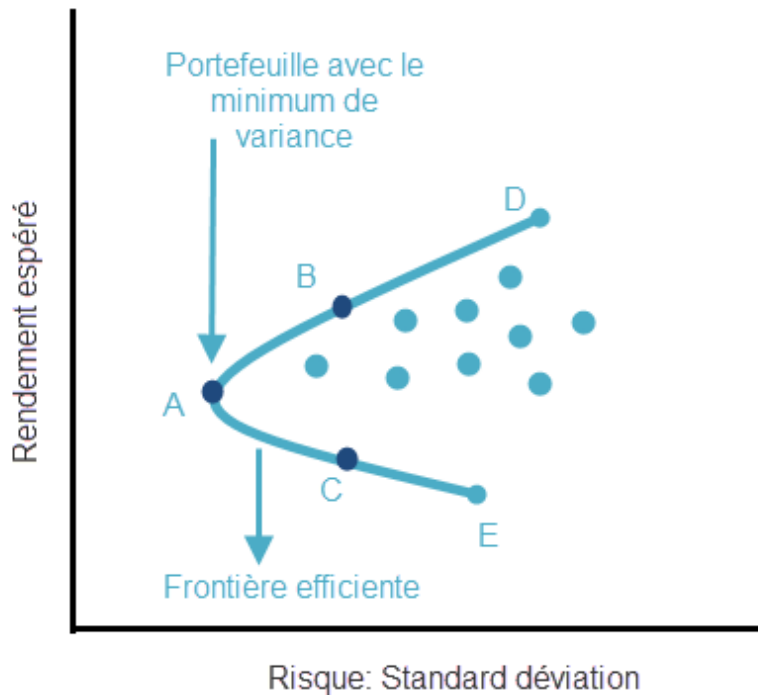
3.2 Méthode de Markowitz

3.2.1 Etude de l'efficacité ou efficience de Markowitz

Définition 3.12. *Un portefeuille est efficient au sens de Markowitz s'il permet d'atteindre un niveau de rendement anticipé fixé avec le minimum de risque (c'est à dire le minimum de variance).*

Ou de manière équivalente, si, pour un niveau de risque donné, il permet d'atteindre le maximum de rendement moyen (espéré).

On cherche la combinaison optimale de n titres risqués permettant d'atteindre un niveau maximal de rendement espéré pour un niveau de variance de rendement donné.



Le schéma ci-dessus représente l'ensemble des portefeuilles disponibles pour un nombre x d'actifs. L'investisseur doit choisir son portefeuille dans ceux compris sur la ligne $A-D$ qui forme la frontière efficiente de Markowitz, les autres étant "dominés" par ces derniers, c'est à dire qu'ils sont moins efficaces. Par exemple, le portefeuille B présente le même niveau de risque que le portefeuille C , mais possède un rendement espéré plus élevé que ce dernier. On dit qu'il le domine. Le portefeuille A représente le portefeuille boursier efficient avec le plus faible risque.

C'est Markowitz qui fut le premier à définir le concept de portefeuille efficient. Le portefeuille efficient est celui avec le plus faible risque pour un certain niveau de rendement espéré.

Le portefeuille A représente le portefeuille avec le minimum de variance, calculé à partir de la variance de chaque actif puis de la covariance de chacun des actifs entre eux.

3.2.2 Le modèle de Markowitz

Nous supposons que nous avons un montant d'argent que nous souhaitons investir pour certaines période de temps, disons un an. Notre modèle est que nous investissons maintenant, et nous ne sommes pas autorisé à modifier notre investissement avant la fin de la période, à quel moment nous vendons tout et observons combien nous avons fait. Nous investissons dans un univers des actifs $A_i, i \in I$. Nous modélisons le retour d'un an de ces actifs au variables aléatoires $\tilde{r}_i$. Le ième atout a prévu le retour r_i . Nous sommes autorisés à investir dans chaque actif n'importe quel montant. Cela nous permet de normaliser le montant qu'on doit investir dans un seul actif.

Si x_i représente le montant que nous investissons dans l'actif i , dans le cas le plus simple nous avons $x_i \geq 0, i = 1, \dots, n$, (pas de vente à découvert), et $\sum_{i=1}^n x_i \leq 1$.

La collection globale de nos placements (non nulle x_i) est appelée notre portefeuille P .

Le retour (une variable aléatoire) est $\tilde{r} = \sum_{i=1}^n x_i \tilde{r}_i$, et son attente est $r = \sum_{i=1}^n x_i r_i$. Le risque de notre portefeuille est mesuré par la variance du retour du portefeuille total. Il peut être écrit comme :

$$V = \sum_{i=1}^n x_i E[(\tilde{r}_i - r_i)(\tilde{r}_j - r_j)] x_j$$

où E représente les attentes, et $E[(\tilde{r}_i - r_i)(\tilde{r}_j - r_j)]$ la covariance des variables aléatoires $\tilde{r}_i$ et $\tilde{r}_j$.

Comme $E[(\tilde{r}_i - r_i)(\tilde{r}_j - r_j)] = E[(\tilde{r}_j - r_j)(\tilde{r}_i - r_i)]$, V est une forme quadratique symétrique ; i.e. il peut être écrit sous la forme :

$$V = x^T Q x$$

tel que Q est une matrice symétrique. Puisque la variance n'est pas négative, Q est semi définie positive.

En supposant qu'il n'y ait aucun portefeuille sans risque est essentiellement équivalent de supposer que Q est défini positive. Il est pratique d'insérer le facteur $\frac{1}{2}$ à la forme quadratique. Ensuite, on résume, le plus simple modèle de Markowitz :

$$\min -rx + \mu \frac{1}{2} x^T Q x$$

sous

$$ex = 1, x \geq 0$$

Où μ est un facteur non négatif qui représente un compromis entre risque et retour.

Remarque 3.4. *Si μ est important, cela implique que le risque est considéré comme plus important, Alors que si μ est petit, le modèle accorde plus d'importance au retour.*

Le modèle équivalent compromis de risque et récompense est :

$$\min \frac{1}{2} x^T Q x$$

sous

$$rx = r_0$$

$$ex = 1$$

$$x \geq 0$$

Lorsque nous exigeons un rendement minimum r_0 . Plus nous mettons cela, plus le risque à prendre en charge est élevé. Si nous autorisons la vente à découvert, alors les x_i ne sont plus nécessairement non négatif dans les deux modèles ci-dessus. À ce stade, nous devons considérer plusieurs choses :

1. Comment obtenons-nous les données estimés, les rendements attendus, et la covariances des retours ?
2. Comment résoudre le problème d'optimisation ?
3. Comment faisons-nous des généralisations plus sophistiquées et plus précises du modèle ?

3.2.3 Programmation quadratique

Le modèle de Markowitz est une instance très simple d'un programme quadratique convexe. C'est l'optimisation nécessaire pour résoudre le problème de Markowitz, la programmation quadratique mérite d'être étudiée pour d'autres raisons. Tout d'abord, il s'agit du problème d'optimisation non linéaire le plus simple à résoudre. Deuxièmement, c'est un outil utilisé dans de nombreux optimiseurs plus généraux pour les programmes non linéaires.

Le programme quadratique général est :

$$\min g(x) = px + \frac{1}{2}x^T Qx$$

sous

$$\begin{aligned} A_1 x &= b_1 \text{ (QP)} \\ A_2 x &\geq b_2 \end{aligned}$$

Où Q est une matrice semi-définie positive. (Nous pouvons également assumer sans perte de généralité que Q est symétrique parce que :

$$x^T Qx = x^T \left(\frac{Q^T + Q}{2} \right) x$$

On note que le gradient de $g(x)$ est $\nabla g(x) = p + Qx$, et la Hessienne à x est $H(x) = Q$. Nous supposons que (QP) est réalisable et a une solution optimale.

Exemple 3.3. *Le modèle Markowitz avec vente à découvert :*

On considère le modèle Markowitz sous la forme suivante :

$$\min \frac{1}{2}x^T Qx$$

sous :

$$\begin{aligned} rx &= r_0 \\ ex &= 1 \end{aligned}$$

Nous formons le lagrangien, $L(x; \lambda, \mu) = \frac{1}{2}x^T Qx - \lambda(rx - r_0) - \mu(ex - 1)$.

Puis on obtient les conditions d'optimalité nécessaires et suffisantes pour un portefeuille efficace (avec vente à découvert autorisée) avec le retour r_0 :

$$\begin{aligned} (***) \quad Qx - \lambda r - \mu e &= 0 \\ rx &= r_0 \\ ex &= 1 \end{aligned}$$

Ceux-ci sont exclusivement résolus (en supposant que les contraintes sont réalisables) et une frontière efficace peut être générée en paramétrant sur r_0 . Mais en fait, les choses sont encore plus simples. Supposons que nous ayons deux portefeuilles efficaces, qui satisfaisaient (***) et fournissent des retours r_0^1 et r_0^2 respectivement. Soient leurs valeurs optimales x_j , λ_j , et μ_j pour $j = 1, 2$.

Soit :

$$\begin{aligned}x_\alpha &= (1 - \alpha)x_1 + \alpha x_2, \\ \lambda_\alpha &= (1 - \alpha)\lambda_1 + \alpha\lambda_2, \\ \mu_\alpha &= (1 - \alpha)\mu_1 + \alpha\mu_2\end{aligned}$$

Notez que α n'est pas limité par un, ni même nécessairement non négatif. Ensuite, par substitution directe, ces valeurs satisfont également (***) avec le retour $r_0^\alpha = (1 - \alpha)r_0^1 + \alpha r_0^2$. Cette observation simple a d'importantes conséquences financières.

Le théorème des deux fonds : Soient deux portefeuilles efficaces, tout autre portefeuille efficace peut être adapté (même moyenne et variance) comme combinaison linéaire des deux autres.

Ce résultat dépend fortement de la possibilité de vente à découvert. Aussi, les portefeuilles avec vente à découvert choisissent de nombreux atouts pour leur portefeuilles efficient, mais lorsque la vente à découvert n'est pas autorisée, moins ils sont choisis.

Remarque 3.5. *Le modèle permettant la vente à découvert est une détente de celui où la vente est interdite, de sorte que la vente à découvert augmente le retour.*

3.2.4 Mesures alternatives de risque

Comment mesurons-nous le risque ? Il existe plusieurs façons. Markowitz a mesuré le risque d'un portefeuille en traitant les portefeuilles comme variables aléatoires et en utilisant les variances du rendement en tant que mesure de risque. La variance mesure la fluctuation de la valeur de la moyenne, supérieure et inférieure à la moyenne. Toutefois, pour nos investissements, nous sommes habituellement beaucoup plus concernés par les fluctuations inférieures à la moyenne (le risque de baisse) ; Nous sommes satisfait si les fluctuations sont supérieures à la moyenne.

Si les fluctuations aléatoires du retour sont symétrique sur le retour prévu ; Par exemple, si nous supposons qu'ils sont normaux alors ce n'est pas un problème car les fluctuations au dessus de la moyenne est statistiquement identique aux fluctuations ci-dessous, donc les deux mesures sont liées par un facteur constant.

Nous pouvons également mesurer le risque, en fonction des scénarios, tout comme nous l'avons fait dans notre premier Exemple. Ces techniques peuvent donner "un côté" des mesures de risque de manière directe. Une façon de mesurer la variation est faite par la somme des valeurs absolues des différences entre le rendement réel et sa valeur attendue [7].

Nous recherchons le portefeuille avec le rendement prévu qui minimise le risque mesuré par la somme sur les périodes des écarts absolus du total des portefeuilles sur la valeur de la moyenne. cela conduit au programme d'optimisation suivant pour déterminer les portefeuilles optimaux en utilisant la mesure de risque L_1 .

Soit $a_{ij} = r_{ij} - r_j$ la différence entre le rendement observé (le rapport du prix de l'action au moment où j'ai divisé par son prix au moment $i - 1$) au moment i pour la sécurité j et sa valeur attendue calculée comme décrite ci-dessus :

$$\min \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij} x_j \right|$$

sous :

$$\sum_{j=1}^n P_j x_j = M_0,$$

$$\sum_{j=1}^n R_j P_j x_j \geq R_0 M_0$$

où $x_j \geq 0, j = 1, \dots, n$
Ce qui est équivalent à :

$$\min \sum_{j=1}^n y_j$$

sous :

$$y_i + \sum_{j=1}^n a_{ij} x_j \geq 0, i = 1, \dots, m$$

$$y_i - \sum_{j=1}^n a_{ij} x_j \geq 0, i = 1, \dots, m$$

$$\sum_{j=1}^n P_j x_j = M_0,$$

$$\sum_{j=1}^n R_j P_j x_j \geq R_0 M_0$$

où $x_j \geq 0$, $j = 1, \dots, n$

Si on souhaite considérer uniquement le risque de baisse, on peut lancer le deuxième ensemble des inégalités impliquant le y_j . Donc, en utilisant cette approche alternative, nous pouvons remplacer un programme quadratique par un programme linéaire et utiliser un risque unilatéral dans la négociation.

On peut effectuer une analyse similaire lorsque on mesure le risque en tant que valeur minimale de notre portefeuille au cours de la période de collection de données. Ça aussi donne naissance à un modèle de programmation linéaire [14].

Chapitre 4

Simulation numérique avec R

4.1 Méthode des moindres carrès

Exemple 4.1. *On cherche à modéliser la relation entre le prix d'un appartement et sa surface.*

On pose :

y = prix en euros/1000,

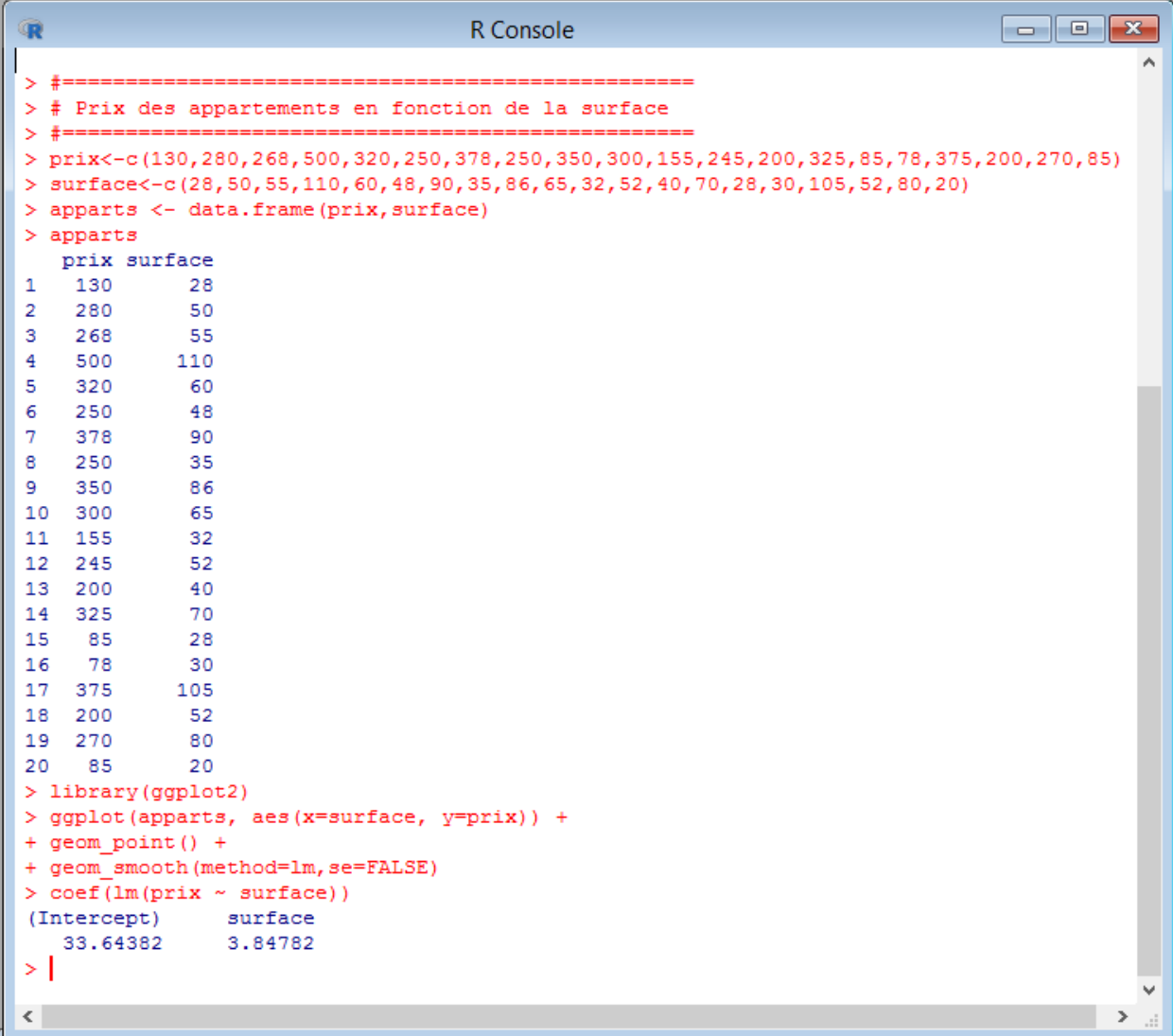
x = surface en m^2 .

On suppose que cette relation est linéaire de la forme :

$$y = \beta_0 + \beta_1 x$$

On veut estimer cette relation appelée droite de régression théorique.

On utilise un échantillon de n appartements dont on connaît le prix et la surface.



```
> #=====
> # Prix des appartements en fonction de la surface
> #=====
> prix<-c(130,280,268,500,320,250,378,250,350,300,155,245,200,325,85,78,375,200,270,85)
> surface<-c(28,50,55,110,60,48,90,35,86,65,32,52,40,70,28,30,105,52,80,20)
> apparts <- data.frame(prix,surface)
> apparts
  prix surface
1  130      28
2  280      50
3  268      55
4  500     110
5  320      60
6  250      48
7  378      90
8  250      35
9  350      86
10 300      65
11 155      32
12 245      52
13 200      40
14 325      70
15  85      28
16  78      30
17 375     105
18 200      52
19 270      80
20  85      20
> library(ggplot2)
> ggplot(apparts, aes(x=surface, y=prix)) +
+ geom_point() +
+ geom_smooth(method=lm,se=FALSE)
> coef(lm(prix ~ surface))
(Intercept)      surface
   33.64382    3.84782
> |
```

FIG.5 - Simulation des prix des appartements en fonction de la surface.

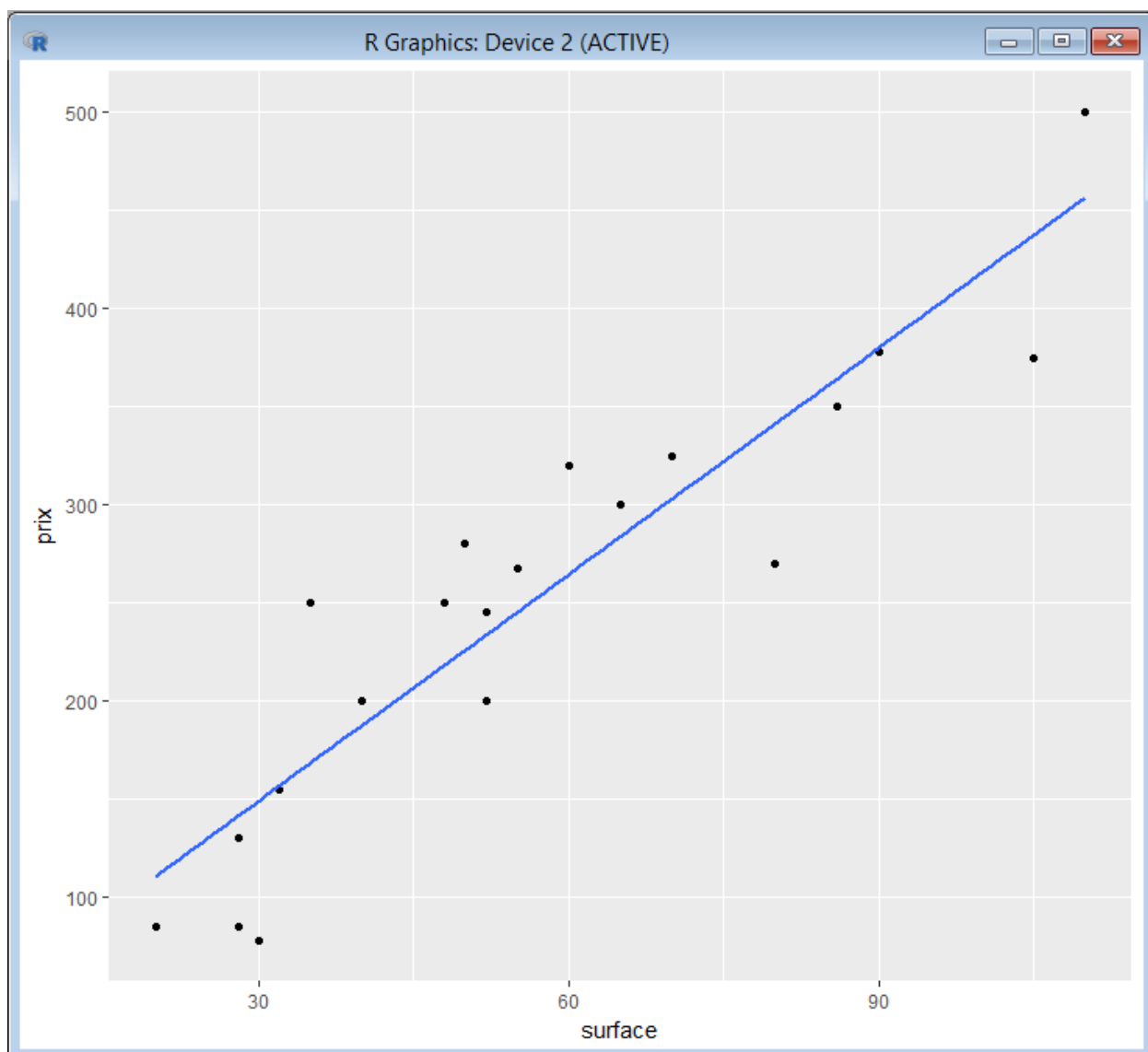
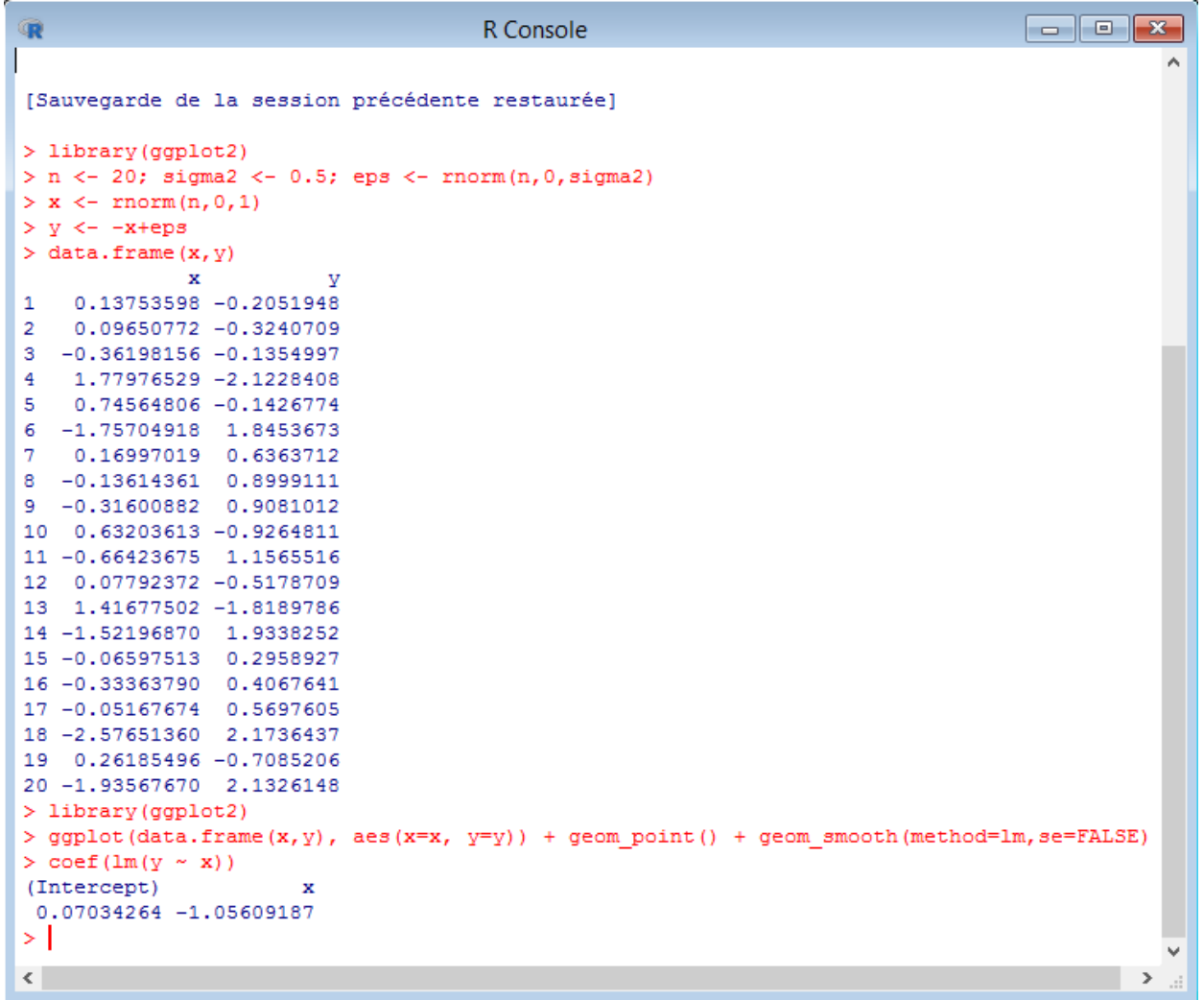


FIG.6 - Nuage de points et droite de régression pour les prix des appartements en fonction de la surface.

Exemple 4.2. Données simulées à partir du modèle $y = -x + \epsilon$:



```
[Sauvegarde de la session précédente restaurée]

> library(ggplot2)
> n <- 20; sigma2 <- 0.5; eps <- rnorm(n,0,sigma2)
> x <- rnorm(n,0,1)
> y <- -x+eps
> data.frame(x,y)
      x      y
1  0.13753598 -0.2051948
2  0.09650772 -0.3240709
3 -0.36198156 -0.1354997
4  1.77976529 -2.1228408
5  0.74564806 -0.1426774
6 -1.75704918  1.8453673
7  0.16997019  0.6363712
8 -0.13614361  0.8999111
9 -0.31600882  0.9081012
10 0.63203613 -0.9264811
11 -0.66423675  1.1565516
12 0.07792372 -0.5178709
13 1.41677502 -1.8189786
14 -1.52196870  1.9338252
15 -0.06597513  0.2958927
16 -0.33363790  0.4067641
17 -0.05167674  0.5697605
18 -2.57651360  2.1736437
19 0.26185496 -0.7085206
20 -1.93567670  2.1326148
> library(ggplot2)
> ggplot(data.frame(x,y), aes(x=x, y=y)) + geom_point() + geom_smooth(method=lm,se=FALSE)
> coef(lm(y ~ x))
(Intercept)          x
 0.07034264 -1.05609187
> |
```

FIG. 7 - Simulation des prix des appartements en fonction de la surface.

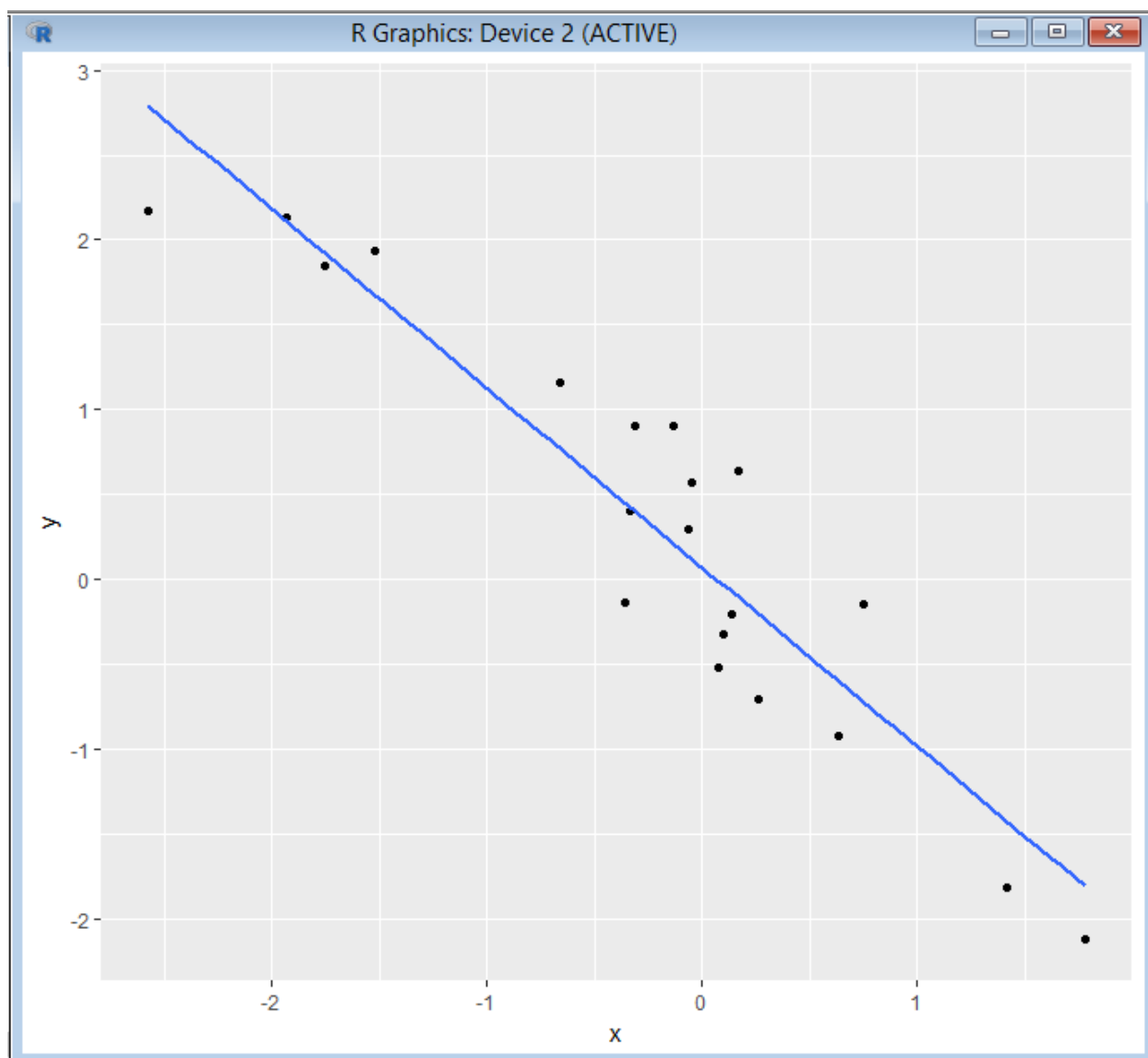


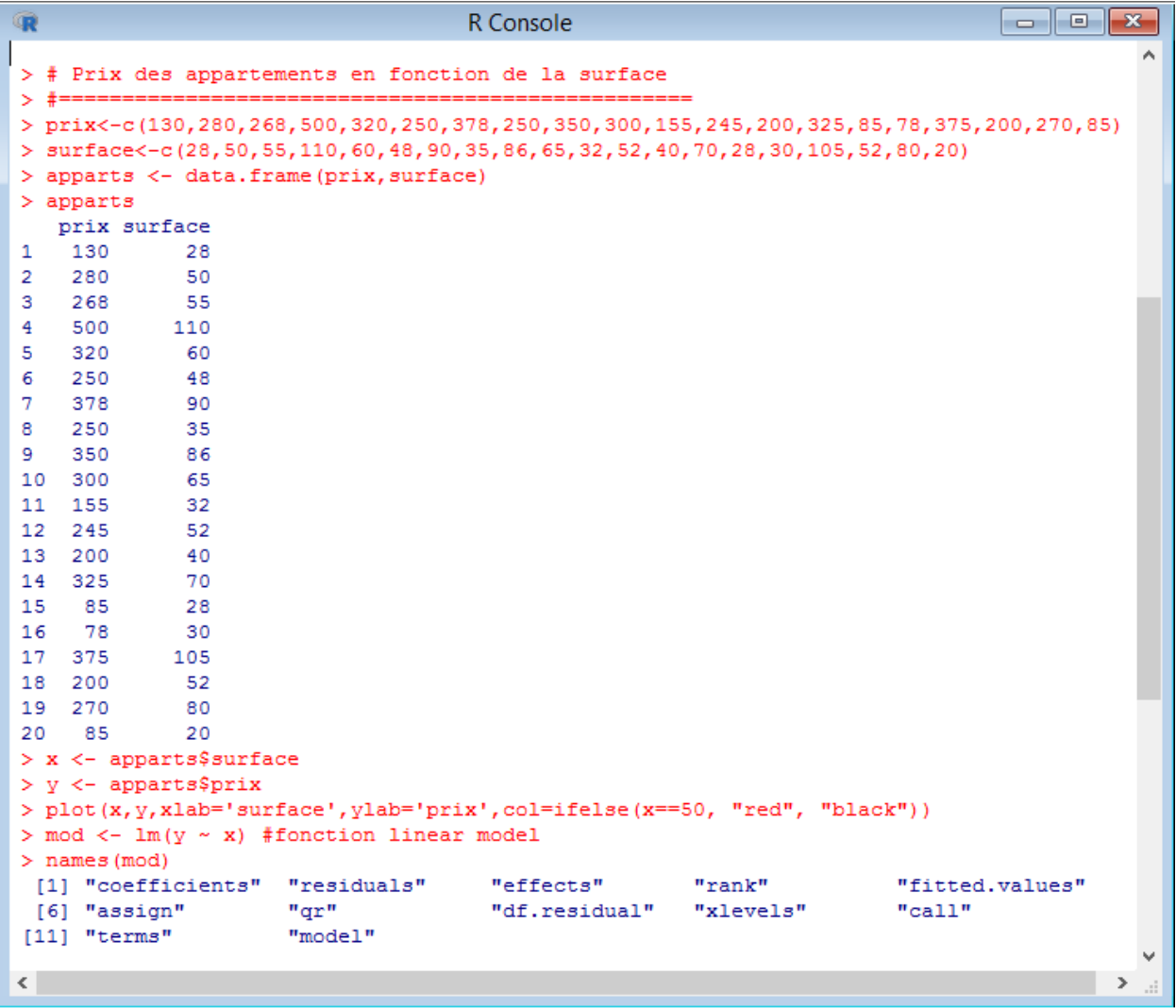
FIG. 8 - Nuage de points et droite de régression pour les prix des appartements en fonction de la surface.

Exemple 4.3. *Données réelles : les appartements Parisiens. Résolution du problème d'optimisation :*

$$\min_{(\beta_0; \beta_1)} F(\beta_0; \beta_1)$$

avec

$$F(\beta_0; \beta_1) = \sum_{i=1}^n \{y_i - (\beta_0 + \beta_1 x_i)\}^2$$



```

> # Prix des appartements en fonction de la surface
> #=====
> prix<-c(130,280,268,500,320,250,378,250,350,300,155,245,200,325,85,78,375,200,270,85)
> surface<-c(28,50,55,110,60,48,90,35,86,65,32,52,40,70,28,30,105,52,80,20)
> apparts <- data.frame(prix,surface)
> apparts
  prix surface
1  130      28
2  280      50
3  268      55
4  500     110
5  320      60
6  250      48
7  378      90
8  250      35
9  350      86
10 300      65
11 155      32
12 245      52
13 200      40
14 325      70
15  85      28
16  78      30
17 375     105
18 200      52
19 270      80
20  85      20
> x <- apparts$surface
> y <- apparts$prix
> plot(x,y,xlab='surface',ylab='prix',col=ifelse(x==50, "red", "black"))
> mod <- lm(y ~ x) #fonction linear model
> names(mod)
[1] "coefficients" "residuals"      "effects"        "rank"          "fitted.values"
[6] "assign"       "qr"            "df.residual"    "xlevels"       "call"
[11] "terms"        "model"

```

FIG.9 - Simulation des prix des appartements en fonction de la surface.

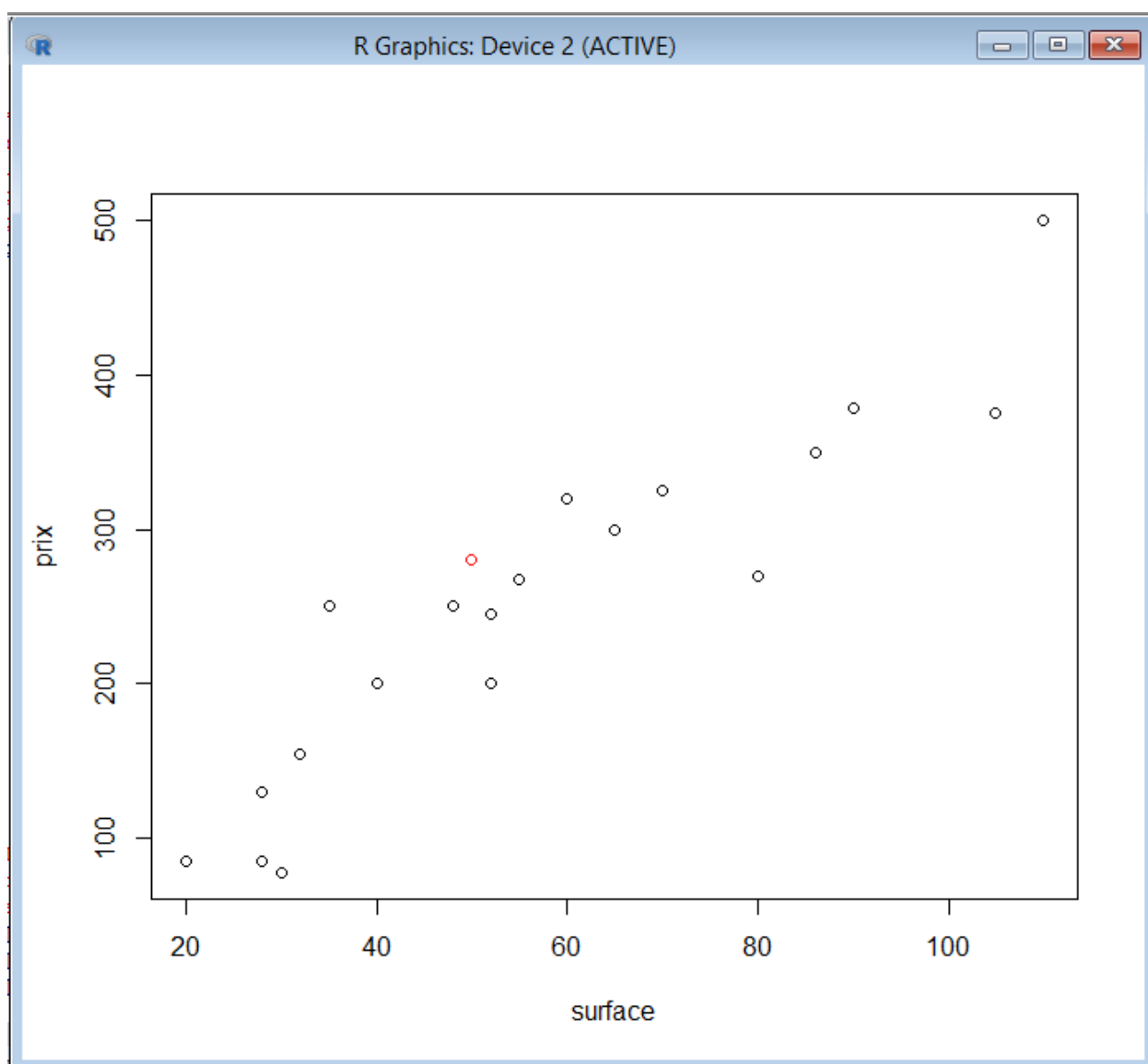


FIG.10 - Nuage de points pour les prix des appartements en fonction de la surface.

```
R Console

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  33.6438    24.4450   1.376   0.186
x             3.8478     0.3922   9.811 1.2e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 45.02 on 18 degrees of freedom
Multiple R-squared:  0.8425,    Adjusted R-squared:  0.8337
F-statistic: 96.26 on 1 and 18 DF,  p-value: 1.197e-08

> lm(y ~ x)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)          x
    33.644         3.848

> plot(x, y)
> abline(33.644 , 3.848, col='red')
> mod <- lm(y~x) #fonction linear model
> round(data.frame(y,val.predites=fitted(mod),residus=resid(mod)) [1:5,],digit=2)
  y val.predites residus
1 130      141.38   -11.38
2 280      226.03    53.97
3 268      245.27    22.73
4 500      456.90    43.10
5 320      264.51    55.49
> n <- 20
> sqrt(sum(resid(mod)^2)/(n-2)) #residual standard error (square root of SCR)
[1] 45.01659
> |
```

FIG.11 - Simulation de la méthode des moindres carrés.

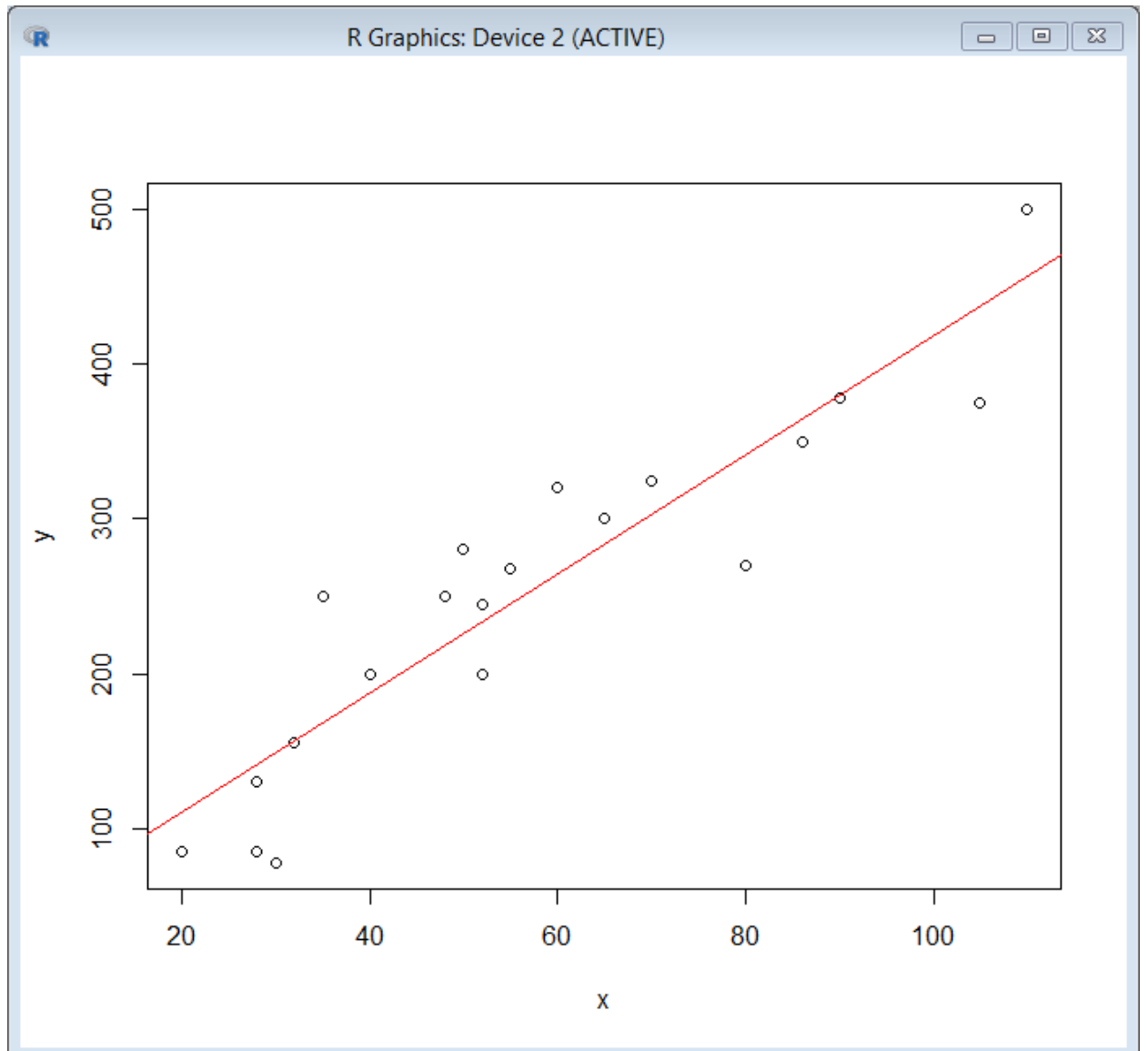
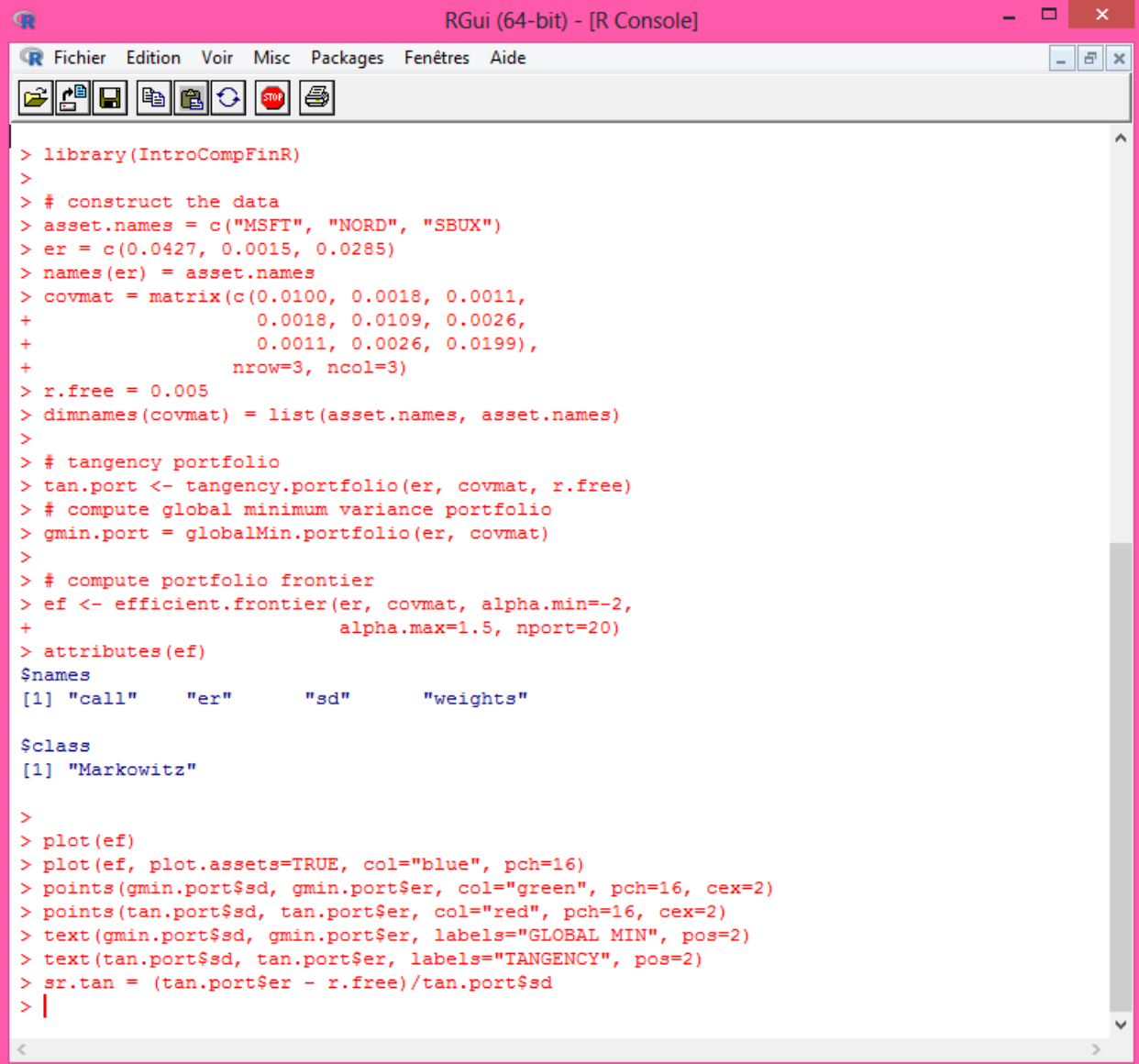


FIG.12 - Nuage de points et droite de régression pour les prix des appartements en fonction de la surface.

4.2 Méthode de markowitz

La frontière efficace est un graphique du rendement attendu du portefeuille et de son l'écart-type pour une collection de portefeuilles efficaces à variance moyenne.



```
> library(IntroCompFinR)
>
> # construct the data
> asset.names = c("MSFT", "NORD", "SBUX")
> er = c(0.0427, 0.0015, 0.0285)
> names(er) = asset.names
> covmat = matrix(c(0.0100, 0.0018, 0.0011,
+                  0.0018, 0.0109, 0.0026,
+                  0.0011, 0.0026, 0.0199),
+                nrow=3, ncol=3)
> r.free = 0.005
> dimnames(covmat) = list(asset.names, asset.names)
>
> # tangency portfolio
> tan.port <- tangency.portfolio(er, covmat, r.free)
> # compute global minimum variance portfolio
> gmin.port = globalMin.portfolio(er, covmat)
>
> # compute portfolio frontier
> ef <- efficient.frontier(er, covmat, alpha.min=-2,
+                         alpha.max=1.5, nport=20)
> attributes(ef)
$names
[1] "call"      "er"        "sd"        "weights"

$class
[1] "Markowitz"

>
> plot(ef)
> plot(ef, plot.assets=TRUE, col="blue", pch=16)
> points(gmin.port$sd, gmin.port$er, col="green", pch=16, cex=2)
> points(tan.port$sd, tan.port$er, col="red", pch=16, cex=2)
> text(gmin.port$sd, gmin.port$er, labels="GLOBAL MIN", pos=2)
> text(tan.port$sd, tan.port$er, labels="TANGENCY", pos=2)
> sr.tan = (tan.port$er - r.free)/tan.port$sd
> |
```

FIG.13 - Simulation du modèle de Markowitz

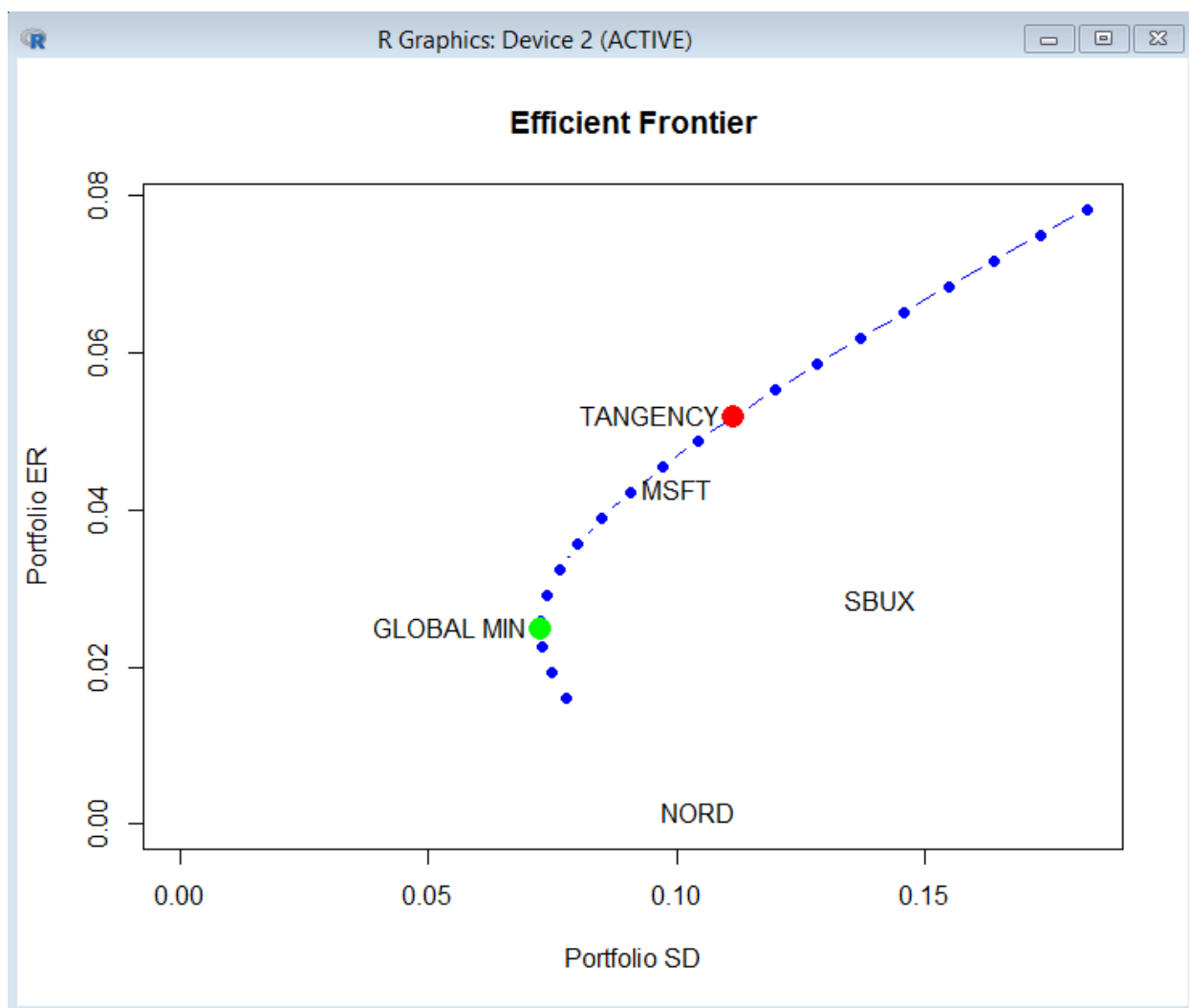


FIG.14 - La frontière efficiente d'un portefeuille.

```
> library(IntroCompFinR)
>
> # construct the data
> asset.names = c("MSFT", "NORD", "SBUX")
> er = c(0.0427, 0.0015, 0.0285)
> names(er) = asset.names
> covmat = matrix(c(0.0100, 0.0018, 0.0011,
+                   0.0018, 0.0109, 0.0026,
+                   0.0011, 0.0026, 0.0199),
+                 nrow=3, ncol=3)
> r.free = 0.005
> dimnames(covmat) = list(asset.names, asset.names)
>
> # tangency portfolio
> tan.port <- tangency.portfolio(er, covmat, r.free)
> # compute global minimum variance portfolio
> gmin.port = globalMin.portfolio(er, covmat)
>
> # compute portfolio frontier
> ef <- efficient.frontier(er, covmat, alpha.min=-2,
+                          alpha.max=1.5, nport=20)
> attributes(ef)
$names
[1] "call"      "er"        "sd"        "weights"

$class
[1] "Markowitz"

>
> plot(ef)
> plot(ef, plot.assets=TRUE, col="blue", pch=16)
> points(gmin.port$sd, gmin.port$er, col="green", pch=16, cex=2)
> points(tan.port$sd, tan.port$er, col="red", pch=16, cex=2)
> text(gmin.port$sd, gmin.port$er, labels="GLOBAL MIN", pos=2)
> text(tan.port$sd, tan.port$er, labels="TANGENCY", pos=2)
> sr.tan = (tan.port$er - r.free)/tan.port$sd
> abline(a=r.free, b=sr.tan, col="green", lwd=2)
```

FIG.15 - Simulation du modèle de Markowitz.

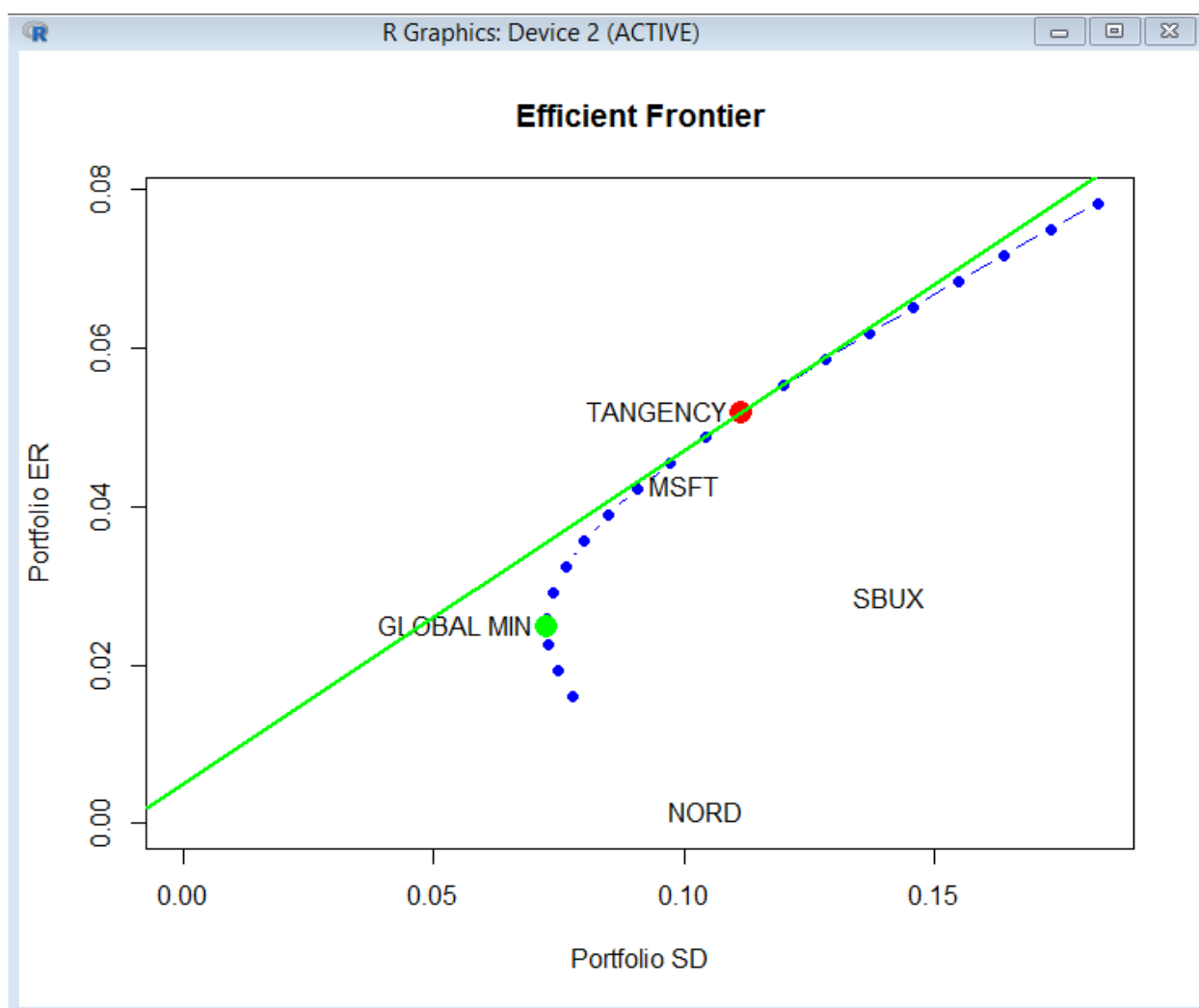


FIG.16 - La frontière efficiente d'un portefeuille et la droite de la tangente.

Conclusion générale et perspectives

Dans ce mémoire, nous avons étudié les problèmes d'optimisation en général et les problèmes d'optimisation quadratiques convexes, afin de résoudre des problèmes d'optimisation spécifiques, souvent à la demande de l'industrie qui désire résoudre les problèmes associés à la modélisation de leur environnement. Nous avons pour ce faire étudié et implémenté différentes approches de résolution.

Nous avons d'abord abordé la méthode des moindres carrés pour résoudre les problèmes d'optimisation quadratique sans contraintes, son principe c'est de comparer des données expérimentales, généralement entachées d'erreurs de mesure, à un modèle mathématique censé décrire ces données, ce qui permet de sélectionner parmi ces fonctions celle qui reproduit le mieux les données expérimentales. Ce nouveau concept d'approche robuste des moindres carrés se généralise aisément à de nombreux problèmes d'optimisation. Il se généralise également à de nombreux problèmes sur les systèmes linéaires. Il nous semble donc que la robustesse pour des problèmes incertains aura des applications très intéressantes, surtout lorsque que l'on cherche à évaluer les dégradations des incertitudes sur un problème donné.

Ensuite le modèle de Markowitz qui a été étudié qui se modélise sous la forme d'un problème d'optimisation quadratique convexe avec contraintes. Il a le grand intérêt de conceptualiser le couple risque-rentabilité, les mérites de la diversification est que le caractère unique de la préférence personnelle par rapport au risque. La conclusion du modèle de Markowitz selon laquelle le portefeuille risque optimal est le portefeuille représentatif du marché dans son ensemble, n'est plus vérifiée dans un marché inefficent. Alors, une information meilleure ou une analyse supérieure permettront à l'investisseur de trouver des portefeuilles ayant une rentabilité supérieure (après pondération du risque) au portefeuille marché.

On peut modéliser le modèle de Markowitz sous la forme d'un problème semi-infini c'est à dire avec un nombre infini de contraintes.

Bibliographie

- [1] D. Azé et J.B. Hiriart-Urruty, *Analyse variationnelle et optimisation*
Cépadués, 2010.
- [2] Eric.Walter, *Méthodes Numériques & optimisation*
Fluctuant Nec Merguntur, Paris, 2015
- [3] G. Allaire, *Analyse numérique et optimisation*
l'école Polytechnique, 2005
- [4] Jean Baptiste Hiriart Urruty *Optimisation et Analyse Convexe*
PUF, 2009.
- [5] J. Cea, *Optimisation : théorie et algorithmes*
Dunod, 1971.
- [6] Jean Christophe Culioli, *Introduction à l'optimisation*
Ellipses Edition Marketing S.A., 2012,32, rue Bague 75740 Paris cedex
15.
- [7] Jacqueline Lelong-Ferrand et Jean-Marie Arnaudiès, *Cours de Mathématiques,tome 2 : Analyse*
Dunod, 1976.
- [8] Leonard,TODJIHOUNDE, *Calcul Différentiel Cours Et Exercices Corrigés, 2ème édition*
Edition Cépadués, Paris, Octobre 2009.
- [9] Leunberger, David, *Investment Science*
Oxford, 1998.
- [10] M. Bergounioux, *Optimisation et contrôle des systèmes linéaires*
Dunod, Paris, 2001.
- [11] Nocedal, Jorge, and Stephen J. Wright, *Numerical Optimization*,
Springer, Chapter 16, 1999.
- [12] P. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*
Masson, Paris, 1982.

- [13] P. Ciarlet, B. Miara, J.M. Thomas, *Exercices l'analyse numérique matricielle et d'optimisation*
Masson, Paris, 1987.
- [14] Young, Martin R, *A Minimax Portfolio Selection Rule with Linear Programming Solution*
Management Science, v. 44, no. 5, May 1998.

Les mots clés

- Optimisation
- Optimisation continue
- Optimisation discrète
- Extrémum
- Minimum local
- Minimum global
- Maximim local
- Maximum global
- Convexité
- Concave
- Optimisation quadratique
- Optimisation quadratique avec contrainte
- Optimisation quadratique sans contrainte
- Optimisation quadratique avec contrainte mixte
- Moindres carrés
- Méthode de Markowitz
- Portefeuille efficient

Résumé

L'optimisation est une discipline mathématique qui touche à différents domaines : économie, gestion, planification logistique, automatique, robotique, traitement du signal.....etc. elle traite la recherche d'un extremum d'une fonction, dont les entrées peuvent être soumises à des contraintes.

L'optimisation en variables continues est plus facile que l'optimisation en variable discrète. Les problèmes d'optimisation occupent actuellement une place importante dans la communauté scientifique. Les différentes méthodes d'optimisation sont la méthode des moindres carrés et le modèle de Markowitz.

La méthode des moindres carrés résout les problèmes d'optimisation quadratique sans contraintes, elle permet de comparer les données expérimentales entachées d'erreurs de mesure.

Le modèle de Markowitz se modélise sous la forme d'un problème d'optimisation quadratique convexe avec contraintes, le portefeuille risque optimal est le portefeuille représentatif du marché.

L'optimisation touche au calcul de variable qu'à la recherche opérationnelle.