

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abdelhafid Boussouf - Mila
Institut des Sciences et de la Technologie

Département des Mathématiques et Informatique

Troisième année Mathématiques

SUPPORT DU COURS

Module : Optimisation sans contraintes

Présenté par :
M^r AZI Mourad

Année Universitaire 2020 – 2021

CHAPITRE 1

QUELQUES RAPPELS ET GÉNÉRALITÉS

Introduction

L'optimisation est une branche des mathématiques cherchant à modéliser, à analyser et à résoudre analytiquement ou numériquement les problèmes de décision qui consistent, généralement, à déterminer le meilleur élément d'un ensemble afin de minimiser ou maximiser un critère donné qui mesure la qualité de la décision. Autrement dit, l'optimisation consiste à chercher la meilleure configuration parmi tous les configurations possibles du système par rapport à un critère donné.

Le mot optimisation vient du latin optimum qui signifie le meilleur. Historiquement les premiers problèmes d'optimisation auraient été formulés par Euclide, au troisième siècle avant notre ère, dans son ouvrage historique *Éléments*. Trois cents ans plus tard, Héron d'Alexandrie énonce le « principe du plus court chemin » dans le contexte de l'optique.

1.1 Formulation mathématique d'un problème d'optimisation classique

Un problème d'optimisation classique peut se formuler de la manière suivante : Trouvez un $x^* \in X$, tel que

$$f(x^*) \leq f(x), \forall x \in X.$$

Ce que on note

$$f(x^*) = \min_{x \in X} f(x) \tag{1.1}$$

L'ensemble X peut être défini de la manière suivante :

$$X = \left\{ \begin{array}{l} x \in \mathbb{R}^n \text{ tel que} \\ g_i(x) = 0, i = 1, \dots, k, \\ g_i(x) \leq 0, i = k + 1, \dots, m. \end{array} \right. \tag{1.2}$$

$f(\cdot)$ est appelée fonction objectif et $g(\cdot)$ est appelée fonction contrainte.

- Si f et $g_i, i = 1, \dots, m$ sont linéaires alors le problème (1.1)-(1.2) est appelé programme linéaire ;
- Si l'une des fonctions est non linéaire alors le problème est dit programme non linéaire avec contrainte (ou problème d'optimisation non linéaire avec contrainte).
- Si $X = \mathbb{R}^n$ le problème est dit programme non linéaire sans contraintes (problème d'optimisation non linéaire sans contraintes).

Pour la formulation mathématique d'un problème d'optimisation concret il faut commencer par le mettre en equation (effectuer la modélisation), qui nécessite un effort de conception qui doit aboutir à la détermination des trois éléments suivants :

- a) Les variables de décision pour lesquelles on agit sur le problème et décider du niveau à atteindre ;
- b) La fonction objectif qui décrit la relation à optimiser ;
- c) Les contraintes du modèle qui décrivent les relations entre les variables représentant les restrictions à ne pas violer.

1.2 Concepts de base de la topologie

Soit $S \subset \mathbb{R}^n$ un sous-ensemble de \mathbb{R}^n .

Définition 1.2.1. "Boule ouverte"

Soient $x^0 \in \mathbb{R}^n$ et un nombre réel $r > 0$. L'ensemble

$$B_r(x^0) = \{x \in \mathbb{R}^n : \|x - x^0\| < r\},$$

est appelé boule ouverte de centre x^0 et de rayon r .

Définition 1.2.2. "Point intérieur"

Un vecteur $x \in S \subset \mathbb{R}^n$ est appelé point intérieur de S , s'il existe un nombre réel $\varepsilon > 0$, tel que : $B_\varepsilon(x) \subset S$.

L'ensemble des points intérieurs de S est appelé l'intérieur de S , noté par $\text{int}(S)$.

Définition 1.2.3. "Frontière d'un ensemble"

Un vecteur $x \in S \subset \mathbb{R}^n$ est appelé point frontière de S , si pour tout nombre réel $\varepsilon > 0$, on a :

$$B_\varepsilon(x) \cap S \neq \emptyset \quad \text{et} \quad B_\varepsilon(x) \cap C_{\mathbb{R}^n} S \neq \emptyset.$$

L'ensemble des points frontières de S , est appelé frontière de l'ensemble S .

Définition 1.2.4. "Point adhérent"

Un vecteur $x \in S \subset \mathbb{R}^n$ est appelé point adhérent de S , si pour tout nombre réel $\varepsilon > 0$, on a : $B_\varepsilon(x) \cap S \neq \emptyset$.

L'ensemble de tous les points d'adhérence de S est appelé l'adhérence de S^1 , notée $\text{cl}(S)$. On a évidemment $S \subset \text{cl}(S)$.

Définition 1.2.5. "Ensemble ouvert"

Un ensemble S est dit ouvert s'il coïncide avec son intérieur, c'est-à-dire si $S = \text{int}(S)$.

Propriétés 1. L'intersection d'un nombre fini d'ensembles ouverts est un ouvert.

La réunion d'un nombre fini ou infini d'ensembles ouverts est un ouvert.

1. On dit aussi fermeture ou clôture

Définition 1.2.6. "Voisinage"

Étant donné $x \in \mathbb{R}^n$, on appelle voisinage de x , noté $V(x)$ tout sous-ensemble ouvert contenant x . De façon équivalente, $V(x)$ est un voisinage de x s'il contient une boule de centre x .

Définition 1.2.7. "Point d'adhérence, adhérence d'un sous-ensemble"

On dit que $x \in \mathbb{R}^n$ est un point d'adhérence d'un sous-ensemble S , si tout voisinage de x rencontre S . Autrement dit, $\forall \varepsilon > 0, \exists y \in S$ tel que $\|x - y\| < \varepsilon$. L'ensemble de tous les points d'adhérence de S est appelé l'adhérence (fermeture ou clôture) de S , notée $cl(S)$. On a évidemment $S \subset cl(S)$.

Définition 1.2.8. "Ensemble fermé"

Un ensemble $S \subset \mathbb{R}^n$ est dit fermé s'il coïncide avec son adhérence, c'est-à-dire si $cl(S) = S$.

Proposition 1.2.1. Le complémentaire (dans \mathbb{R}^n) d'un ouvert est un fermé. Le complémentaire (dans \mathbb{R}^n) d'un fermé est un ouvert.

Propriétés 2. L'union d'un nombre fini d'ensembles fermés est un fermé. L'intersection d'un nombre fini ou infini d'ensembles fermés est un fermé.

Définition 1.2.9. "Ensemble borné"

Un ensemble $S \subset \mathbb{R}^n$ est borné, s'il existe un réel $M > 0$ tel que pour tout $x \in S$, on ait $\|x\| \leq M$.

La propriété suivante donne une caractérisation simple des ensembles compacts dans \mathbb{R}^n .

Propriétés 3. Dans \mathbb{R}^n , un sous-ensemble S est compact si et seulement s'il est fermé et borné.

1.3 Quelques Rappels du calcul différentiel

Soit $f : X \rightarrow \mathbb{R}$, avec $X \subset \mathbb{R}^n$ ouvert.

Définition 1.3.1. On dit que f est de classe C^m sur X si toutes les dérivées partielles jusqu'à l'ordre m existent et sont continues.

Définition 1.3.2. Pour tout $x \in X$ et tout $i \in \{1, 2, \dots, n\}$ on note (quand \exists)

$$\frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial e_i}(x) = \lim_{t \rightarrow 0} \frac{1}{t} [f(x + te_i) - f(x)]$$

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

Avec e_i est le vecteur de la base canonique de \mathbb{R}^n , donné par :

$$(e_i)_j = \begin{cases} 0 & \forall j = 1, 2, \dots, n, j \neq i, \\ 1 & \text{si } j = i. \end{cases}$$

Définition 1.3.3. Le gradient de la fonction f au point x est défini par :

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}. \tag{1.3}$$

Définition 1.3.4. La dérivée directionnelle de f dans la direction d au point x est :

$$\begin{aligned} \frac{\partial f}{\partial d} &= \lim_{t \rightarrow 0} \frac{f(x + td) - f(x)}{t} \\ &= \frac{\partial f(x + td)}{\partial x_1} \Big|_{t=0} d_1 + \dots + \frac{\partial f(x + td)}{\partial x_n} \Big|_{t=0} d_n \\ &= \nabla f(x)^T d. \end{aligned}$$

Proposition 1.3.1. Soit $f : X \rightarrow \mathbb{R}^n$, avec $X \subset \mathbb{R}^n$ ouvert. Supposons que f est continûment différentiable sur X , alors on :

$$\lim_{d \rightarrow 0} \frac{f(x + d) - f(x) - \nabla f(x)^T d}{\|d\|} = 0, \quad \forall x \in X.$$

En effet

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \theta(\|y - x\|), \quad (1.4)$$

où $\theta(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction qui vérifie $\lim_{t \rightarrow 0} \frac{\theta(t)}{t} = 0$.

Définition 1.3.5. Soit $f : X \rightarrow \mathbb{R}$ une fonction réelle de classe \mathcal{C}^2 . Pour tout $x \in X$ et $i, j \in \{1, 2, \dots, n\}$ on note

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

la dérivée partielle à l'ordre 2.

On note $\nabla^2 f(x)$ la matrice carrée d'ordre n donnée par

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

$\nabla^2 f(x)$ s'appelle matrice Hessienne de f en x , qui s'écrit aussi :

$$\begin{aligned} \nabla^2 f(x) &= \left(\nabla \frac{\partial f}{\partial x_1}, \nabla \frac{\partial f}{\partial x_2}, \dots, \nabla \frac{\partial f}{\partial x_j}, \dots, \nabla \frac{\partial f}{\partial x_n} \right) \\ &= \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}. \end{aligned} \quad (1.5)$$

Dans ce qui suit, nous présentons deux principaux résultats d'approximation (linéaire et quadratique) qui sont des conséquences directes du théorème d'approximation de Taylor qui seront fréquemment utilisés dans la suite de ce support de cours.

Théorème 1.1. (Approximation linéaire) Soit $f : X \rightarrow \mathbb{R}^n$ une fonction deux fois continûment différentiable sur X -ouvert subset \mathbb{R}^n , et soit $x \in X$ et $r > 0$ d'une manière à avoir $B(x, r) \subseteq X$.

Alors pour tout $y \in B(x, r)$, il existe $z \in [x, y]$ tel que :

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2(z) (y - x). \quad (1.6)$$

Théorème 1.2. (Approximation quadratique) Soit $f : X \rightarrow \mathbb{R}^n$ une fonction deux fois continûment différentiable sur X -ouvert subset \mathbb{R}^n , et soit $x \in X$ et $r > 0$ d'une manière à avoir $B(x, r) \subseteq X$.

Alors pour tout $y \in B(x, r)$, $\exists z \in [x, y]$ tel que :

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2(x) (y - x) + \theta(\|y - x\|^2). \quad (1.7)$$

1.4 Classification de Matrices(nature de matrice)

Soit $F(x) = x^T D x$ une forme quadratique avec D symétrique.

Définition 1.4.1. (*Nature de matrice*)

- ✓ D est dite définie positive (noté $D > 0$) si $x^T D x > 0, \forall x \in \mathbb{R}^n$ et $x \neq 0$. Elle est dite semi-définie positive ou définie non négative (noté $D \geq 0$) si $x^T D x \geq 0, \forall x \in \mathbb{R}^n$.
- ✓ D est dite définie négative (noté $D < 0$) si $x^T D x < 0, \forall x \in \mathbb{R}^n$ et $x \neq 0$. Elle est dite semi-définie négative ou définie non positive si $x^T D x \leq 0, \forall x \in \mathbb{R}^n$.
- ✓ D est dite non définie si la forme quadratique associe prend des valeurs positives pour certains valeurs de x et des valeurs négatives pour d'autre.

1.4.1 Critère de Sylvester pour les matrices définies et semi-définies

L'intérêt du critère du Sylvester est de caractériser une matrice (forme quadratique) définie ou semi-définie. Pour cela, considérons la matrice symétrique suivante :

$$D = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{pmatrix}.$$

Le mineur de la matrice D , formé des lignes i_1, i_2, \dots, i_p et les colonnes j_1, j_2, \dots, j_p , sera noté comme suit :

$$D \begin{pmatrix} i_1, i_2, \dots, i_p \\ j_1, j_2, \dots, j_p \end{pmatrix} = \begin{vmatrix} d_{i_1 j_1} & d_{i_1 j_2} & \cdots & d_{i_1 j_p} \\ d_{i_2 j_1} & d_{i_2 j_2} & \cdots & d_{i_2 j_p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{i_p j_1} & d_{i_p j_2} & \cdots & d_{i_p j_p} \end{vmatrix}.$$

Ce mineur est dit principal si $i_1 = j_1, i_2 = j_2, \dots, i_p = j_p$, c'est-à-dire s'il est formé de lignes et de colonnes portant les mêmes numéros.

Les mineurs suivants

$$D_1 = d_{11}, \quad D_2 = \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix}, \dots, \quad D_n = \begin{vmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{vmatrix},$$

sont appelés mineurs principaux successifs.

Le critère de Sylvester se formule comme suit :

Théorème 1.3. (*Critère de Sylvester*)

- Pour qu'une matrice symétrique D soit définie positive ($D > 0$), il est nécessaire et suffisant que tous ses mineurs principaux successifs soient positifs :

$$D_p > 0, \forall p = 1, 2, \dots, n \Leftrightarrow D_1 > 0, \quad D_2 > 0, \dots, \quad D_n > 0, \quad (1.8)$$

- Pour que la matrice D soit semi-définie positive ($D \geq 0$), il est nécessaire et suffisant que tous ses mineurs principaux soient non négatifs :

$$D \begin{pmatrix} i_1, i_2, \dots, i_p \\ i_1, i_2, \dots, i_p \end{pmatrix} \geq 0, \quad 1 \leq i_1 < i_2 < \dots < i_p \leq n, \quad p = 1, 2, \dots, n. \quad (1.9)$$

- Pour qu'une matrice symétrique D soit définie négative ($D < 0$), il est nécessaire et suffisant que tous ses mineurs principaux successifs vérifient la relation suivante :

$$(-1)^p D_p > 0, \quad \forall p = 1, 2, \dots, n. \quad (1.10)$$

- Pour que la matrice D soit semi-définie négative ($D \leq 0$), il est nécessaire et suffisant que la relation suivante soit vérifiée :

$$(-1)^p D \begin{pmatrix} i_1, i_2, \dots, i_p \\ i_1, i_2, \dots, i_p \end{pmatrix} \geq 0, \quad 1 \leq i_1 < i_2 < \dots < i_p \leq n, \quad p = 1, 2, \dots, n. \quad (1.11)$$

1.5 Notions sur la convexité

La convexité joue un rôle central dans la théorie classique de l'optimisation. Elle est un outil indispensable pour la recherche des conditions à la fois nécessaires et suffisantes d'optimalité.

1.5.1 Ensemble convexe

Définition 1.5.1. Un ensemble X de \mathbb{R}^n est dit convexe, si pour tout couple (x, y) d'éléments de X , le segment $[x, y]$ tout entier est inclus dans X . Autrement dit, X est convexe lorsque

$$\forall x, y \in X, \lambda \in [0, 1], \text{ le vecteur } z = \lambda x + (1 - \lambda)y \in X. \quad (1.12)$$

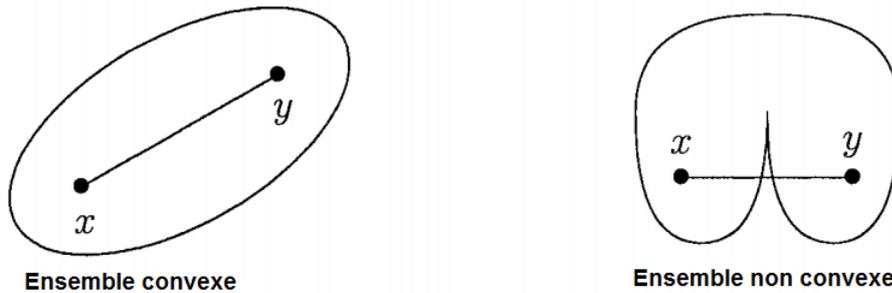


FIGURE 1.1 – Ensembles convexe et non convexe

1.5.1.1 Propriétés des ensembles convexes

Propriétés 4. Soit une famille $\{X_i\}_{i=1, \dots, k}$ d'ensembles convexes, alors on a :

- $X = \bigcap_{i=1}^k X_i$ est un ensemble convexe.
- $X = \prod_{i=1}^k X_i$ est un ensemble convexe.

Propriétés 5. Si X est convexe, et $\lambda \in \mathbb{R}$, alors l'ensemble $K = \{x | x = \lambda x_1, x_1 \in X\}$ est convexe.

1.5.2 Fonction convexe

Cette section est consacré au concept de fonctions convexes, qui est fondamental dans la théorie de l'optimisation.

Définition 1.5.2. Soit f fonction définie sur un ensemble convexe X de \mathbb{R}^n . f est dite convexe, si pour tous les points $x, y \in X$, et pour tout nombre réel $\lambda \in [0, 1]$ l'inégalité suivante est vérifiée :

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (1.13)$$

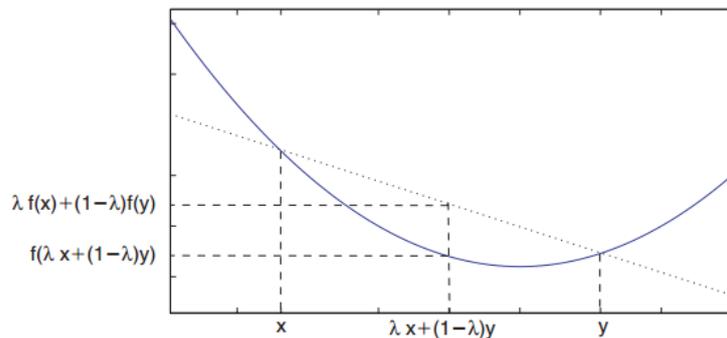


FIGURE 1.2 – Fonction convexe

Définition 1.5.3. On dit que f est strictement convexe sur X -convexe si :

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y), \quad \forall x, y \in X \text{ avec } x \neq y, \forall \lambda \in]0, 1[. \quad (1.14)$$

Définition 1.5.4. On dit que f est concave (respectivement strictement concave) si $-f$ est convexe (respectivement strictement convexe).

1.5.2.1 Propriétés des fonctions convexes

Propriétés 6. Soit f une fonction réelle définie sur un ensemble convexe $X \subset \mathbb{R}^n$. f est convexe si et seulement si son épigraphe

$$\text{epi}(f) = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : x \in X, f(x) \leq \alpha\},$$

est un ensemble convexe.

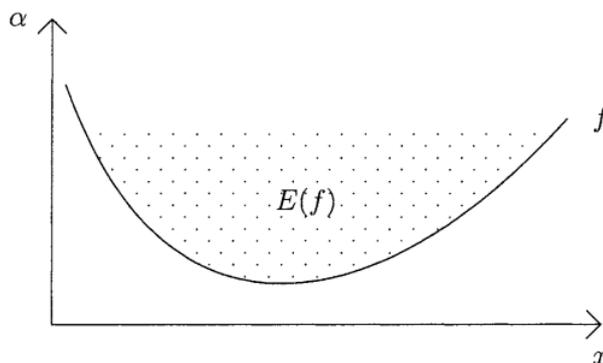


FIGURE 1.3 – Epigraphe

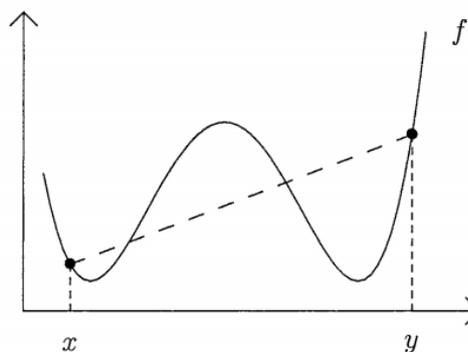


FIGURE 1.4 – Fonction non convexe

Théorème 1.4. Soit f une fonction réelle continûment différentiable, définie sur un ensemble convexe $X \in \mathbb{R}^n$. Alors f est convexe si et seulement si

$$f(y) - f(x) \geq (y - x)^T \nabla f(x), \forall x, y \in X. \quad (1.15)$$

Théorème 1.5. Soit f une fonction réelle deux fois continûment différentiable, définie sur un ensemble convexe $X \in \mathbb{R}^n$.

Alors f est convexe si et seulement si $\forall x \in X$, le Hessien $\nabla^2 f(x)$ est une matrice semi-définie positive.

Les preuves des théorèmes sera en classe.