Chapitre 1

Introduction

1.1 Qu'est-ce que l'optimisation ?

Un problème d'optimisation est un problème où une fonction est maximisée ou minimisée par rapport à un ensemble donné. La fonction à minimiser ou à maximiser est appelée fonction objectif, et l'ensemble est appelé domaine réalisable (ou domaine des contraintes). Dans ce cours, un domaine réalisable est toujours considéré comme un sous-ensemble de $\mathbb{R}^n$ (espace vectoriel réel à n dimensions) et la fonction objectif est une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$.

Dans ce cours, nous avons divisé les problèmes d'optimisation en deux catégories : les problèmes linéaires et les problèmes non linéaires.

On commence par introduire la minimisation sans contrainte. Soit $f : \mathbb{R}^n \to \mathbb{R}$, un problème de minimisation sans contrainte est un problème de la forme

$$(\mathcal{P}) \qquad \min_{x\in\mathbb{R}^n} f(x)$$

L'étude de ces problèmes est importante pour diverses raisons. Beaucoup de problèmes d'optimisation avec contraintes peuvent être transformés en problèmes d'optimisation sans contrainte (multiplicateur de Lagrange, méthodes de pénalité, ...). L'étude des problèmes d'optimisation sans contrainte trouve aussi des applications dans la résolution des systèmes non linéaires. Ensuite, le problème le plus typique, et celui sur lequel nous nous concentrons le plus ici, est celui de l'optimisation avec des contraintes, et nous en ferons une introduction simple dans ce premier chapitre.

1.1.1 Optimisation linéaire

Parmi eux, le problème de la planification linéaire (*linear programming problem*) est un problème d'optimisation sur $\mathbb{R}^n$, où la fonction objectif est une fonction linéaire, c'est-à-dire qu'elle est de la forme

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

avec $c_i \in \mathbb{R}$, $i = 1, \ldots, n$, et un domaine réalisable est un ensemble fini d'inégalités linéaires et d'égalités de contraintes

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leqslant b_i \qquad i = 1, \ldots, m$$

et

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \qquad i = m+1, \ldots, r$$

La planification linéaire est un outil extrêmement puissant pour résoudre les problèmes d'optimisation dans un large éventail d'applications. Une courte liste de domaines d'application comprend l'allocation des ressources, la planification de la production, le stockage, l'aménagement, la planification du transport, l'emplacement des installations, la planification de l'équipage, l'optimisation d'un portefeuille d'actions, l'estimation des paramètres, etc.

Définition 1.1 – Optimisation linéaire (avec contraintes)/*Linear programming*/线性规划

Un problème d'optimisation linéaire (OL) consiste à minimiser une fonction linéaire sur un polyèdre convexe. Un problème d'OL est sous forme standard/*standard form*/标准形式, s'il est de la forme

$$\min\{c^\top \cdot x,\ A \cdot x = b,\ x \geqslant 0\} \tag{OL}$$

où $A \in \mathcal{M}_{m,n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$.

Remarque 1.1

Soit $(\mathcal{P})$ un problème d'optimisation linéaire, nous pouvons toujours le mettre sous forme standard. Voici quelques techniques usuelles.

1. **Problème de maximisation** : un problème de maximisation est équivalent à un problème de minimisation sous la forme standard comme
$$\max\left(c^\top \cdot x\right) = -\min\left(-c^\top \cdot x\right)$$
2. **Contraintes d'inégalité** : $A \cdot x \leqslant b$ ou $A \cdot x \geqslant b$.
On peut introduire une variable $z \in \mathbb{R}^m$, $z \geqslant 0$ telle que
$$A \cdot x + z = b \text{ ou } A \cdot x - z = b$$
Ceci pourra s'écrire en matrice par blocs comme
$$\begin{bmatrix} A | I \end{bmatrix} \cdot \begin{bmatrix} x \\ z \end{bmatrix} = b \text{ ou } \begin{bmatrix} A | -I \end{bmatrix} \cdot \begin{bmatrix} x \\ z \end{bmatrix} = b$$
Donc les contraintes d'inégalité deviennent des contraintes d'égalité sous la forme standard avec les variables $\begin{bmatrix} x \\ z \end{bmatrix} \geqslant 0$.
Une telle variable z est appelée une variable d'écart/*slack variable*/松弛变量.
3. **Variables négatives** : $x \leqslant 0$.
Il est facile de se ramener à la forme standard. Il suffit de remplacer la variable x par $x' = -x$. Par exemple : la forme standard du problème
$$\min\{c^\top \cdot x,\ A \cdot x = b,\ x \leqslant 0\}$$
est
$$\min\{(-c)^\top \cdot x',\ A \cdot x' = -b,\ x' \geqslant 0\}$$
4. **Variables libres** : $x \in \mathbb{R}^n$.
On pourra ajouter des variables $y, z \in \mathbb{R}^n_+$, telles que
$$x = y - z$$
En remplaçant la variable libre x par $y - z$, on va rendre le problème d'optimisation linéaire sous la forme standard. Par exemple :
$$\min\{c^\top \cdot x,\ A \cdot x = b,\ x \in \mathbb{R}^n\}$$
En remplaçant x par $y - z$, on obtient sa forme standard
$$(\mathcal{P}) \qquad \min\left\{\begin{bmatrix} c^\top | -c^\top \end{bmatrix} \cdot \begin{bmatrix} y \\ z \end{bmatrix} : \begin{bmatrix} A | -A \end{bmatrix} \cdot \begin{bmatrix} y \\ z \end{bmatrix} = b,\ \begin{bmatrix} y \\ z \end{bmatrix} \geqslant 0\right\}$$
où la variable du problème $(\mathcal{P})$ sous forme standard est $\begin{bmatrix} y \\ z \end{bmatrix}$.

Nous commençons par un exemple en deux dimensions. En modélisant cet exemple, nous passerons en revue quatre étapes dans le développement du problème d'OL :

1. Identifier les *variables de décision.*
2. Déterminer la quantité que l'on souhaite minimiser (ou maximiser) et écrire l'expression de la *fonction objectif* en tant que fonction linéaire à l'aide des variables de décision.
3. Identifier les contraintes explicites et écrire l'expression fonctionnelle de chaque *contrainte* sous forme d'équation linéaire ou d'inégalité linéaire.
4. Déterminer les contraintes implicites et écrire chaque *contrainte* comme une équation linéaire ou une inégalité linéaire.

Exemple 1.1 – Usine de verre

Un fabricant de verres souhaite optimiser sa production afin de maximiser son profit. Il produit des chopes à bière et des flûtes à champagne personnalisées. Le profit sur une caisse de chopes à bière est de 25 euros tandis que le profit sur une caisse de flûtes à champagne est de 20 euros. Les gobelets sont fabriqués avec une machine appelée extrudeuse qui se nourrit de sable. Chaque caisse de chopes à bière nécessite 10 kg de sable tandis que les verres à champagne nécessitent 6 kg par caisse. L'approvisionnement quotidien en sable est limité à 900 kg maximum. Environ 15 caisses de l'un ou de l'autre produit peuvent être produites par heure. Pour le moment, la famille veut limiter son travail à 8 heures par jour.
Les variables de décision. On note
B le nombre de caisses de chopes à bière à produire quotidiennement.
C le nombre de caisses de flûtes à champagne à produire quotidiennement.
L'objectif de l'entreprise. Maximiser le profit quotidien.
Le modèle d'optimisation.

$$\text{Maximiser } 25B + 20C$$

Les contraintes explicites.
contrainte de sable : $20B + 12C \leqslant 1800$,
contrainte d'horaires de travail $\frac{1}{15}B + \frac{1}{15}C \leqslant 8$.
Les contraintes implicites.

$$0 \leqslant B,\ 0 \leqslant C$$

L'ensemble du modèle du problème de l'usine de gobelets peut maintenant être énoncé succinctement comme

$$
\begin{array}{rll}
(\mathcal{P}) & \max & 25B + 20C \\
 & s.c. & 20B + 12C \leqslant 1800 \\
 & & \dfrac{1}{15}B + \dfrac{1}{15}C \leqslant 8 \\
 & & B \geqslant 0,\ C \geqslant 0
\end{array}
$$

Exemple 1.2 – Problème de transport

Une entreprise de distribution doit transporter un produit disponible en quantité d_i aux dépôts $i = 1, \ldots, m$ vers des destinations j où la demande est b_j, $j = 1, \ldots, n$. Supposons que

1. La demande totale doit être égale à l'offre totale

$$\sum_{i=1}^{m} d_i = \sum_{j=1}^{n} b_j$$

2. Le coût unitaire de transport de i à j est $c_{i,j}$.

Les variables de décision. On note $x_{i,j}$ la quantité non-négative transportée de i à j.
L'objectif de l'entreprise. Minimiser le coût total de transport.

Le modèle d'optimisation.

$$\text{Minimiser} \sum_{i=1}^{m}\sum_{j=1}^{n} c_{i,j}x_{i,j}$$

Les contraintes

$$\begin{aligned}
\sum_{i=1}^{m} x_{i,j} = b_j \quad & (j = 1, \dots, n)\\
\sum_{j=1}^{n} x_{i,j} = d_i \quad & (i = 1, \dots, m)\\
x_{i,j} \geqslant 0 \quad & (i = 1, \dots, m;\ j = 1, \dots, n)
\end{aligned}$$

1.1.2 Optimisation non linéaire

Définition 1.2 – Optimisation non linéaire/*Linear programming*/非线性规划

Un problème d'optimisation non linéaire (ONL) s'écrit sous la forme générale suivante :

$$\min\{f(x),\ g(x) \leqslant 0,\ h(x) = 0\} \tag{ONL}$$

où les fonctions f, g et h sont typiquement non-linéaires. L'équation $g(x) \leqslant 0$ désigne ce que nous appellerons des contraintes d'inégalité et l'équation $h(x) = 0$ des contraintes d'égalité.

Exemple 1.3 – Portefeuille d'actions

Soit A, B et C trois titres d'un portefeuille. On note r_A, r_B et r_C leur rentabilité moyenne (espérance) et σ_A, σ_B , σ_C leur volatilité (écart type). Soit x, y, z les poids respectifs investis dans chaque titre. On a donc $x + y + z = 1$ avec $x, y, z \geqslant 0$. On souhaite une rentabilité plus grande en moyenne que α.

$$xr_A + yr_B + zr_C \geqslant \alpha$$

et on souhaite minimiser le risque (variance)

$$f(x, y, z) = x^2\sigma_A^2 + y^2\sigma_B^2 + z^2\sigma_C^2$$

On cherche donc à résoudre le problème

$$(\mathcal{P}) \quad \begin{aligned}
\inf_{(x,y,z)\in K} \quad & f(x, y, z) = x^2\sigma_A^2 + y^2\sigma_B^2 + z^2\sigma_C^2\\
s.c. \quad & xr_A + yr_B + zr_C \geqslant \alpha\\
& x + y + z = 1\\
& (x, y, z) \in (\mathbb{R}_+)^3
\end{aligned}$$

Exemple 1.4 – Un problème en mécanique

On considère une corde horizontale de longueur 1 tendue à ses deux extrémités, avec une tension τ. La déviation éventuelle de la corde par rapport à sa position d'équilibre est désignée par $u(x)$, où $x \in [0, 1]$. Les extrémités étant fixées, on aura toujours $u(0) = u(1) = 0$. On néglige le poids de la corde par rapport à la tension τ , cela permet d'affirmer qu'en l'absence d'action extérieure, la corde est au repos et on a donc $u(x) = 0$, $\forall x \in [0, 1]$.
Supposons maintenant que la corde est écartée de sa position d'origine. Alors on peut montrer que l'énergie

potentielle associée à cette déformation (supposée petite) est

$$E(u) = \frac{1}{2}\int_0^1 \tau \left(\frac{du}{dx}\right)^2 dx$$

En l'absence d'obstacle, la position de repos $u(x) = 0$ minimise cette énergie. Il peut alors être intéressant d'étudier un problème où un obstacle empèche la corde de prendre la position triviale $u(x) = 0$. Intuitivement, on voit bien que la corde va toucher l'obsctale en certains points, mais pas forcément en tous les points de l'intervalle $[0, 1]$ (cela va dépendre de la forme de l'obstacle).
Supposons par exemple que cet obstacle peut être représenté par une fonction $v(x) \geqslant 0$. Alors la présence de l'obstacle se traduit par la condition

$$u(x) \geqslant v(x),\ x \in]0, 1[$$

Si on veut connaître la déformation $u(x)$ de la corde lorsque l'obstacle est présent, on peut donc penser qu'il est raisonnable de considérer le problème

$$(\mathcal{P}) \quad \begin{array}{rl} \min\limits_u & f(u) = \dfrac{1}{2}\displaystyle\int_0^1 \tau \left(\frac{du}{dx}\right)^2 dx \\ s.c. & u(x) \geqslant v(x),\ x \in]0, 1[\\ & u(0) = u(1) = 0 \end{array}$$

Il s'agit, techniquement parlant, d'un problème de calcul des variations, et donc l'inconnue est une fonction (la fonction $u(x)$). Il semble donc pour l'instant impossible de le mettre sous forme standard.

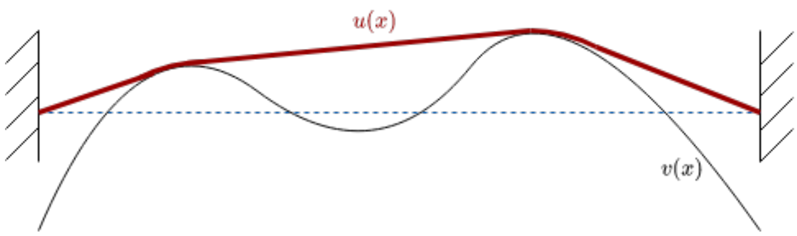

FIGURE 1.1 – Déformation par la force de la corde.

Chapitre 2

Outils techniques

2.1 Algèbre linéaire

Voir le livre [Algèbre linéaire].

Rappel 2.1 – Matrice définie positive/*Positive-definite matrix*/正定矩阵

Soit M une matrice symétrique réelle d'ordre n. Elle est dite *définie positive* si elle est positive et inversible, autrement dit si elle vérifie l'une des quatre propriétés équivalentes suivantes

1. Pour toute matrice colonne non nulle x à n éléments réels, on a $x^\top \cdot M \cdot x > 0$.
   Autrement dit, la forme quadratique définie par M est strictement positive pour $x \neq 0$.
2. Toutes les valeurs propres de M (qui sont nécessairement réelles) sont strictement positives.
3. La forme bilinéaire symétrique

$$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},\ (x, y) \mapsto x^\top \cdot M \cdot y$$

   est un produit scalaire sur $\mathbb{R}^n$.
4. Il existe une matrice $N \in \mathcal{M}_n(R)$ inversible telle que $M = N^\top \cdot N$.
   Autrement dit, M est congruente à la matrice identité.

2.2 Topologie dans $\mathbb{R}^n$

Rappel 2.2 – Boule ouverte, fermée de $\mathbb{R}^n$/*Open, closed ball*/开球，闭球

On se place dans $\mathbb{R}^n$, muni de la norme euclidienne $\|\cdot\|$ et du produit scalaire euclidien $\langle\cdot,\cdot\rangle$. Pour $x \in \mathbb{R}^n$ et $r > 0$, on note $\mathrm{BO}(x, r)$ la boule ouverte de centre x et de rayon r et $\mathrm{BF}(x, r)$ la boule fermée correspondante

$$\mathrm{BO}(x, r) = \left\{y \in \mathbb{R}^n,\ \|y - x\| < r\right\}$$
$$\mathrm{BF}(x, r) = \left\{y \in \mathbb{R}^n,\ \|y - x\| \leqslant r\right\}$$

Définition 2.1 – Ensemble ouvert, fermé de $\mathbb{R}^n$/*Open, closed set*/开集，闭集

— Un sous-ensemble O de $\mathbb{R}^n$ est un *ouvert* de $\mathbb{R}^n$ si

$$\forall x \in O,\ \exists r_x > 0,\ \mathrm{BO}(x, r_x) \subset O$$

On dit aussi que O est ouvert dans $\mathbb{R}^n$.

— Un sous-ensemble F de $\mathbb{R}^n$ est un *fermé* de $\mathbb{R}^n$ si son complémentaire $\mathbb{R}^n \backslash F$ est ouvert.

Propriété 2.1

1. Les ensembles $\varnothing$ et $\mathbb{R}^n$ sont ouverts dans $\mathbb{R}^n$.
2. Une réunion quelconque d'ouverts est ouverte.
3. Une intersection finie d'ouverts est ouverte. Voici un contre-exemple pour l'intersection infinie

$$\bigcap_{n\in\mathbb{N}^*}\left]-\frac{1}{n},1+\frac{1}{n}\right[=]0,1[$$

Définition 2.2 – Voisinage/*Neighbourhood*/邻域

Soit $a \in \mathbb{R}^n$, on dit que V est un *voisinage* de a dans $\mathbb{R}^n$ si

$$\exists\varepsilon>0,\ \mathrm{BO}(a,\varepsilon)\subset V$$

L'ensemble des voisinages de a sera noté par $\mathcal{V}(a)$.

Propriété 2.2

1. L'ensemble $\mathcal{V}(a)$ est stable par intersections finies (i.e., $\forall(V_1,V_2,\dots,V_n)\in\mathcal{V}(a)^n$, $V_1\cap V_2\cap\cdots\cap V_n\in\mathcal{V}(a)$).
2. L'ensemble $\mathcal{V}(a)$ est stable par croissance (i.e., $\forall V\in\mathcal{V}(a)$, $V\subset W\Rightarrow W\in\mathcal{V}(a)$).

Définition 2.3 – Partie bornée/*Bounded set*/有界集合

Soit A une partie d'un espace métrique $(\mathbb{R}^n, d)$,

1. on dit que A est *bornée* si
$$\exists x\in\mathbb{R}^n,\ \exists r\in\mathbb{R}_+^*,\ A\subset\mathrm{BO}(x,r)$$
2. on dit qu'une fonction $f:\Omega\longrightarrow\mathbb{R}^n$ est *bornée* si
$$f(\Omega)\text{ est borné}$$

Propriété 2.3

Soit A une partie non vide d'un espace métrique $(\mathbb{R}^n, d)$. Les trois assertions suivantes sont équivalentes.

1. A est bornée
2. $\forall x\in\mathbb{R}^n,\ \exists r\in\mathbb{R}_+^*,\ A\subset\mathrm{BO}(x,r)$
3. $\mathrm{diam}(A)=\sup\{d(x,y)\,|\,(x,y)\in A^2\}<+\infty$

Définition 2.4 – Fonction coercive/*Coercive function*/强制泛函

Soit $f:A\to\mathbb{R}$ une fonction numérique continue définie sur une partie $A\subset\mathbb{R}^n$ fermée mais pas bornée. On dit que f est *coercive* si

$$\lim_{\|x\|\to+\infty,\,x\in A} f(x)=+\infty$$

Autrement dit, f est *coercive* si

$$\forall M>0,\ \exists N>0\text{ tel que }\|x\|>N\text{ et }x\in A\ \Rightarrow\ f(x)>M$$

2.3 Calcul différentiel

Rappel 2.3 – Dérivée partielle/*Partial derivative*/偏导数

Soit O un ouvert non vide de $\mathbb{R}^n$ et soit $f : O \to \mathbb{R}^p$.

— On dit que f est *admet une dérivée partielle par rapport à la j-ième variable en $x \in O$* si la fonction

$$t \longmapsto f(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_n)$$

est dérivable en x_j.

— Si c'est le cas, le vecteur dérivé de $\mathbb{R}^p$ obtenu (ou nombre dérivé si $p = 1$) est noté

$$\frac{\partial f}{\partial x_j}(x) \qquad \text{ou} \qquad \partial_j f(x)$$

— Si f admet une dérivée partielle par rapport à la j-ième variable en tout $x \in O$, on définit la *j-ième dérivée partielle de f* par

$$\partial_j f : \left\{ \begin{array}{ccc} O & \longrightarrow & \mathbb{R}^p \\ x & \longmapsto & \partial_j f(x) \end{array} \right.$$

Remarque 2.1 – Dérivée directionnelle/*Directional derivative*/方向导数

Pour un vecteur $v \in \mathbb{R}^n$ non nul, on peut définir plus généralement (sous réserve d'existence) la *dérivée directionnelle de f dans la direction v en x*, notée $\partial_v f(x)$, par [a]

$$\partial_v f(x) = \lim_{t \to 0} \frac{f(x + t\,.v) - f(x)}{t}$$

En particulier, $\partial_j f(x)$ est la dérivée directionnelle de f dans la direction e_j (le j-ième vecteur de la base canonique de $\mathbb{R}^n$) en x

$$\partial_j f(x) = \lim_{t \to 0} \frac{f(x + t\,.e_j) - f(x)}{t}$$

Toujours sous réserve d'existence, la *dérivée directionnelle de f dans la direction v* est alors définie par

$$\partial_v f : \left\{ \begin{array}{ccc} O & \longmapsto & \mathbb{R}^p \\ v & \longmapsto & \partial_v f(x) \end{array} \right.$$

[a] Pour t assez petit, $x + t\,.v \in O$ car O est un ouvert de $\mathbb{R}^n$.

Rappel 2.4 – Fonction différentiable et différentielle en un point/*Differentiable function*/可导方程

Soit O un ouvert non vide de $\mathbb{R}^n$ et soit $x_0 \in O$. Une fonction $f : O \to \mathbb{R}^p$ est *différentiable en $x_0 \in O$* s'il existe une application linéaire $\mathrm{d}f_{x_0} \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^p)$ telle que, pour tout $h \in \mathbb{R}^n$ tel que $x + h \in O$,

$$f(x_0 + h) = f(x_0) + \mathrm{d}f_{x_0}(h) + \underset{h \to 0_{\mathbb{R}^n}}{o}(\|h\|)$$

Si c'est le cas, $\mathrm{d}f_{x_0}$ s'appelle la *différentielle de f en x_0*.

Remarque 2.2 – Lien entre différentielle et dérivée (dimension 1)

Soit I un intervalle ouvert non vide de $\mathbb{R}$, soit $x_0 \in I$ et soit $f : I \to \mathbb{R}$. Alors f est différentiable en x_0 si et seulement si elle est dérivable en x_0. Si c'est le cas, pour tout $h \in \mathbb{R}$ tel que $x_0 + h \in I$,

$$f(x_0 + h) = f(x_0) + f'(x_0)\,h + \underset{h \to 0}{o}(h)$$

On en déduit que $\mathrm{d}f_{x_0}(h) = f'(x_0)\,h$ pour tout $h \in \mathbb{R}$. En particulier, $f'(x_0) = \mathrm{d}f_{x_0}(1)$.

Rappel 2.5 – Différentielle d'une fonction composée

Soit O un ouvert non vide de $\mathbb{R}^n$, soit $x_0 \in O$ et soit $f : O \to \mathbb{R}^p$. Soit également O' un ouvert non vide de $\mathbb{R}^p$ contenant $f(O)$ et soit $g : O' \to \mathbb{R}^q$. On peut alors considérer

$$g \circ f : O \subset \mathbb{R}^n \xrightarrow{f} f(O) \subset O' \subset \mathbb{R}^p \xrightarrow{g} \mathbb{R}^q$$

Si f est différentiable en x_0 et si g est différentiable en $f(x_0)$, alors $g \circ f$ est différentiable en x_0 et

$$\mathrm{d}(g \circ f)_{x_0} = \left(\mathrm{d}g_{f(x_0)}\right) \circ \mathrm{d}f_{x_0} \tag{2.1}$$

Rappel 2.6 – Gradient/*Gradient*/梯度

Soit f une fonction définie sur un ouvert O de $\mathbb{R}^n$ à valeurs dans $\mathbb{R}$ et soit $a \in O$. Lorsque f est différentiable en a, on appelle *gradient de f en a* le vecteur

$$\nabla f(a) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{bmatrix} \in \mathbb{R}^n$$

Remarque 2.3 – Lien entre différentielle et gradient

Si $p = 1$ (donc $f : O \to \mathbb{R}$), on peut représenter la forme linéaire $\mathrm{d}f_{x_0} \in \mathscr{L}(\mathbb{R}^n, \mathbb{R})$ à l'aide du produit scalaire (théorème de représentation, voir le cours « algèbre linéaire avancé ») : il existe un unique vecteur $\nabla f(x_0) \in \mathbb{R}^n$, appelé *gradient de f en x_0*, tel que

$$\forall h \in \mathbb{R}^n, \quad \mathrm{d}f_{x_0}(h) = \langle \nabla f(x_0), h \rangle$$

De plus, si $\mathscr{E} = (e_1, \ldots, e_n)$ est une base orthonormée de $\mathbb{R}^n$ (par exemple la base canonique), alors

$$\mathrm{Mat}_{\mathscr{E}}\left(\nabla f(x_0)\right) = \begin{bmatrix} \partial_{e_1} f(x_0) \\ \vdots \\ \partial_{e_n} f(x_0) \end{bmatrix}$$

Rappel 2.7 – Matrice jacobienne/*Jacobian matrix*/雅可比矩阵

Soit O un ouvert non vide de $\mathbb{R}^n$, soit $x_0 \in O$ et soit $f : O \to \mathbb{R}^p$ différentiable en x_0. La matrice de la différentielle $\mathrm{d}f_{x_0} \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^p)$ de f en x_0 dans les bases canoniques de $\mathbb{R}^n$ et de $\mathbb{R}^p$ est appelée la *matrice jacobienne de f en x_0*, notée $J_f(x_0)$.

On a donc

$$J_f(x_0) = \left[\partial_j f_i(x_0)\right]_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant n}} = \begin{bmatrix} \partial_1 f_1(x_0) & \partial_2 f_1(x_0) & \cdots & \partial_n f_1(x_0) \\ \partial_1 f_2(x_0) & \partial_2 f_2(x_0) & \cdots & \partial_n f_2(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_p(x_0) & \partial_2 f_p(x_0) & \cdots & \partial_n f_p(x_0) \end{bmatrix} \in \mathrm{M}_{p,n}(\mathbb{R})$$

Définition 2.5 – Matrice hessienne/*Hessian matrix*/海森矩阵

Soit $x_0 \in \mathbb{R}^n$ et $f : \mathbb{R}^n \to \mathbb{R}$ dont toutes les dérivées paritelles secondes existent en x_0, alors la *matrice hessienne (ou simplement la hessienne) de f en x_0*, notée $H_f(x_0)$, est la matrice carrée définie par

$$H_f(x_0) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)\right]_{(i,j) \in [\![1,n]\!]^2} \in \mathrm{M}_{n,n}(\mathbb{R})$$

Remarque 2.4

D'après le théorème de Schwarz (voir le Poly de Y1S2), si f est une fonction de classe $\mathscr{C}^2$ définie sur un ouvert O de $\mathbb{R}^n$ et $\forall x_0 \in O$, on a $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_0)$ pour $\forall (i,j) \in [\![1,n]\!]^2$, et donc $H_f(x_0)$ est symétrique.

Rappel 2.8 – Développement limité/*Taylor expansion*/泰勒展开

Soit $f : \mathbb{R}^n \to \mathbb{R}$ différentiable en $x \in \mathbb{R}^n$, alors f admet un *développement limité à l'ordre* 1 *en* x, donné par

$$f(x+h) = f(x) + df_x(h) + o(\|h\|)$$

ou bien sous la forme matricielle

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

Proposition 2.1 – Développement limité d'ordre 2

Soit $f : \mathbb{R}^n \to \mathbb{R}$ deux fois différentiable en $x \in \mathbb{R}^n$, alors f admet un développement limité à l'ordre 2 en *x, donné par*

$$f(x+h) = f(x) + df_x(h) + \frac{1}{2} d^2 f_x(h,h) + o(\|h\|^2)$$

ou bien sous la forme matricielle

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

2.4 Analyse convexe

Rappel 2.9 – Ensemble convexe/*Convex set*/凸集

Soit $S \subset \mathbb{R}^n$, S est dit *convexe* si

$$\forall (x,y) \in S^2,\ \forall \lambda \in [0,1],\ \lambda.x + (1-\lambda).y \in S$$

Rappel 2.10 – Fonction convexe/*Convex function*/凸函数

Soit $S \subset \mathbb{R}^n$ un ensemble convexe non vide. Une fonction $f : S \to \mathbb{R}$ est dite *convexe* si

$$\forall (x,y) \in S^2, \forall \theta \in [0,1], f(\theta.x + (1-\theta).y) \leqslant \theta f(x) + (1-\theta) f(y)$$

f est dite strictement convexe si l'inégalité est stricte dès que $x \neq y$ et $\theta \in]0,1[$.

Rappel 2.11 – Fonction strictement convexe/*Strictly convex function*/严格凸函数

Soit E un espace vectoriel (ou affine) réel. On dit qu'une fonction $f : E \to \mathbb{R} \cup \{+\infty\}$ est *strictement convexe* si, pour tous x_1 et x_2 distincts dans E et pour tout $t \in]0,1[$, on a

$$f(t.x_1 + (1-t).x_2) < t\, f(x_1) + (1-t)\, f(x_2)$$

Définition 2.6 – Fonction fortement convexe/*Strongly convex function*/强凸函数

Soit E un espace vectoriel (ou affine) réel. On dit qu'une fonction $f : E \to \mathbb{R} \cup \{+\infty\}$ est *fortement convexe* de module $m > 0$ si, pour tous x_1 et x_2 distincts dans E et pour tout $t \in]0,1[$, on a

$$f(t.x_1 + (1-t).x_2) \leqslant t\, f(x_1) + (1-t)\, f(x_2) - \frac{m}{2} t(1-t)\, \|x_1 - x_2\|^2$$

On dit également que f est m-fortement convexe.

Proposition 2.2 – Fonction fortement convexe

*Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable. Alors f est m-*fortement convexe *s'il existe une constante $m > 0$ telle que*

$$\forall (x,h) \in \mathbb{R}^n, \langle H_f(x) \cdot h, h\rangle \geqslant m\|h\|^2$$

On dit également que f est m-fortement convexe.

Théorème 2.1 – Caractérisation de la convexité par le gradient

Soit $C \subset \mathbb{R}^n$ un ouvert convexe, $f : C \subset \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable, alors f est convexe si et seulement si

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle, \; \forall (x,y) \in K^2$$

La fonction f est strictement convexe si et seulement si

$$f(y) > f(x) + \langle \nabla f(x), y - x\rangle, \; \forall (x,y) \in K^2, \; x \neq y$$

Démonstration 1

On voit bien l'interprétation géométrique de ce dernier resultat quand $n = 1$: le graphe d'une fonction convexe f se trouve toujours au-dessus de la tangente en un point donné.
Soit f convexe i.e.

$$\forall (x,y) \in C^2, \; \forall \lambda \in [0,1] : f(x + \lambda.(y-x)) \leqslant f(x) + \lambda(f(y) - f(x))$$

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

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

En faisant tendre λ vers 0, on a bien

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

Soit à présent $(x,y) \in C^2$ et $\lambda \in [0,1]$. Posons $x_\lambda = x + \lambda.(y-x)$. En retranchant respectivement x et y de x_λ on obtient

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

En appliquant la relation à la fonction f (dans le théorème) respectivement aux points (x, x_λ) et (y, x_λ), on obtient

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

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

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

En sommant les deux inégalités on obtient la relation assurant la convexité de f.

Théorème 2.2 – Caractérisation de la convexité par la matrice hessienne

Soit $C \subset \mathbb{R}^n$ un ouvert convexe, $f : C \subset \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable, alors

1. *f est convexe si et seulement si pour tout $x \in C$, $H_f(x)$ est semi-définie positive ($H_f(x) \geqslant 0$)*
2. *f est strictement convexe si $H_f(x)$ est définie positive ($H_f(x) > 0$).*

3. f est fortement convexe si et seulement si $H_f(x)$ est définie positive ($H_f(x) > 0$).

Démonstration 2

Supposons que f est convexe, et soit $x \in C$. Vu que C est ouvert, alors pour tout $y \in \mathbb{R}^n$, il existe $\lambda \in \mathbb{R}^*$ suffisamment petit tel que $x + \lambda.y \in C$. D'après le théorème précédent et la différentiabilté d'ordre 2 de f on a

$$f(x + \lambda.y) \geqslant f(x) + \langle \nabla f(x), \lambda.y\rangle$$

$$f(x + \lambda.y) = f(x) + \langle \nabla f(x), \lambda.y\rangle + \frac{1}{2}\langle H_f(x) \cdot (\lambda.y), \lambda.y\rangle + \lambda^2\|x\|^2\epsilon(\lambda.x)$$

En substituant, la deuxième équation dans la première on obtient

$$\frac{1}{2}\langle H_f(x) \cdot (\lambda.y), \lambda.y\rangle + \lambda^2\|x\|^2\epsilon(\lambda.x) \geqslant 0$$

En divisant cette inégalité par λ^2 et en faisant tendre $\lambda \to 0$, on conclut que la hessienne est semi-définie positive. Inversement, supposons que la matrice hessienne et semi-définie positive en tout point de C. Considérons $(x, y) \in C^2$ et $\lambda \in]0, 1[$ et posons $\hat{y} = \lambda.x + (1 - \lambda).y$. Du développement de Taylor de f, on a

$$f(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle H_f(\hat{y}) \cdot (y - x), y - x\rangle$$

Notons que $\hat{y} \in C$ par convexité. De la semi-définie positivité de la matrice hessienne $H_f(\hat{y})$, on déduit que

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

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

Le corolaire suivant est immédiat.

Propriété 2.4

Soit f une forme quadratique définie par

$$f(x) = \frac{1}{2}\langle A \cdot x, x\rangle - \langle b, x\rangle$$

alors f est convexe si et seulement si $A \geqslant 0$, et strictement convexe si et seulement si $A > 0$.

Cela provient du fait que $H_f(x) = A$. En réalité on a même

$$f \text{ fortement convexe } \iff A > 0$$

Proposition 2.3 – Propriétés des fonctions fortement convexes

Soit $m > 0$ et $f : \mathbb{R}^n \to \mathbb{R}$ une fonction m-fortement convexe. Alors f vérifie les propriétés suivantes

1. *f est strictement convexe.*
2. *Pour tous $(x, y) \in (\mathbb{R}^n)^2$,*

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{m}{2}\|y - x\|^2$$

3. *f est coercive.*
4. *f admet un unique minimum global $x^\star$.*
5. *Pour tout $x \in \mathbb{R}^n$,*

$$f(x^\star) \geqslant f(x) - \frac{1}{2m}\|\nabla f(x)\|^2 \text{ et } \|x - x^\star\| \leqslant \frac{2}{m}\|\nabla f(x)\|$$

Démonstration 3

Soit f une fonction m-fortement convexe.

1. Comme pour tout $x \in \mathbb{R}^n$, la matrice hessienne $H_f(x)$ est définie positive, f est strictement convexe.

2. Soit $(x, y) \in (\mathbb{R}^n)^2$. D'après la formule de Taylor, il existe $z \in]x, y[$ tel que

$$f(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle H_f(z) \cdot (y - x), y - x\rangle$$

En appliquant l'inégalité de m-forte convexité au dernier terme on obtient la minoration annoncée.

3. En prenant $x = 0$ dans l'inégalité précédente on a

$$f(y) \geqslant f(x) + \langle \nabla f(0), y\rangle + \frac{m}{2}\|y\|^2$$

Cette fonction minorante est coercive, donc f est elle aussi coercive.

4. f est strictement convexe et coercive, elle admet donc un unique minimum global.
5. Soit $x \in \mathbb{R}^n$. Alors, pour tout $y \in \mathbb{R}^n$,

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle + \frac{m}{2}\|y - x\|^2$$

Le terme de droite est une fonction quadratique de la variable y qui est minimale en $y^\star = x - \frac{1}{m}\nabla f(x)$ et cette valeur minimale vaut $f(x) - \frac{1}{2m}\|\nabla f(x)\|^2$.
Ainsi, on a pour tout $y \in \mathbb{R}^n$,

$$f(y) \geqslant f(x) - \frac{1}{2m}\|\nabla f(x)\|^2$$

En prenant $y = x^\star$ on a

$$f(x^\star) \geqslant f(x) - \frac{1}{2m}\|\nabla f(x)\|^2$$

2.5 Exercices

2.1 Trouvez une matrice M d'ordre n symétrique non définie positive telle que sur deux sous-espaces vectoriels supplémentaires E_1 et E_2 de $\mathbb{R}^n$, on ait $\langle M \cdot x, x\rangle > 0$, pour tout $x \in E_1\backslash\{0\}$ et pour tout $x \in E_2\backslash\{0\}$.
Remarque. Une matrice peut donc être définie positive dans deux sous-espaces supplémentaires sans être définie positive.

2.2 a. Si A est une matrice symétrique définie positive, il existe $\alpha > 0$ tel que $\langle A \cdot x, x\rangle \geqslant \alpha\|x\|_2^2$, pour tout $x \in \mathbb{R}^n$. Donnez la meilleure constante α.

b. Si A est une matrice quelconque, il existe $\alpha > 0$ tel que $\|A^\top \cdot A \cdot x\|_2 > m\|A \cdot x\|_2$, pour tout $x \in \mathbb{R}^n$. Donnez la meilleure constante α. En déduire que $x \mapsto \|A^\top \cdot x\|_2$ est une norme sur $\mathrm{Im}(A)$.

2.3 Soit $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction définie par :

$$f(x, y) = x^4 - 4xy + y^4$$

a. Montrer que f est coercive à l'aide de la norme *sup* : $\|(x, y)\|_\infty = \max(|x|, |y|)$.

b. Trouver le minimum global de f sur $\mathbb{R}^2$.

2.4 Soit une fonction $f : \mathbb{R}^3 \to \mathbb{R}$ définie par :

$$f(x, y, z) = e^{x-y} + e^{y-x} + z^2$$

a. Calculer sa hessienne et étudier son signe. Montrer que f est convexe sur $\mathbb{R}^3$.

b. Déterminer les points critiques de f. En déduire le minimum global de f.

2.5 Soit $u : \mathbb{R}^3 \to \mathbb{R}^2$, $f : \mathbb{R}^2 \to \mathbb{R}^2$, et $h = f \circ u$ où

$$f(u) = \begin{bmatrix} u_1^2 - u_2 \\ u_1 + u_2^2 \end{bmatrix}, \; u(x) = \begin{bmatrix} x_1 + x_3 \\ x_2^2 - x_3 \end{bmatrix}$$

Calculer la matrice jacobienne de f.

2.6 Les fonctions suivantes, sont-elles convexes ? Sont-elles strictement convexes ?

a. La fonction linéaire $f(x) = \langle x, a\rangle + \alpha$;

b. $f(x) = \|x\|_2$;

c. $f(x) = \|x\|_\infty$;

d. $f + g$ (si f est strictement convexe et que g est convexe).

2.7 Soit $f : \mathbb{R}^2 \to \mathbb{R}$ la fonction définie par $f(x, y) = 2x^3 + 6xy - 3y^2 + 2$ pour tout $(x, y) \in \mathbb{R}^2$.

a. Déterminer les extréma locaux de la fonction f.

b. La fonction f possède-t-elle des extréma globaux sur $\mathbb{R}^2$?

c. Représenter le segment de droite L défini par

$$L = \{(x, y) \in \mathbb{R}^2 | -2 \leqslant x \leqslant 0,\ y = x + 1\}$$

et déterminer les extréma globaux de la restriction de f à L, en précisant en quels points de L ils sont atteints.

2.8 On considère la fonction définie pour tout $(x, y) \in \mathbb{R}^2$ par

$$f(x, y) = xye^{-\frac{1}{2}(x^2+y^2)}$$

a. Etudier les extréma locaux de f sur $\mathbb{R}^2$. On pourra utiliser les symétries de la fonction f pour réduire le nombre de cas à étudier.

b. Démontrer que $f(x, y) \to 0$ quand $\|(x, y)\| \to \infty$.

c. Déduire de ce qui précède l'existence des extréma globaux de f sur $\mathbb{R}^2$ et les déterminer.

2.9 Une boîte rectangulaire doit être construite. On désigne par x_1, x_2, x_3 les dimensions de la boîte : x_3 est la hauteur de la boîte, x_1 et x_2 sont les dimensions du fond de la boîte.
La boîte n'a pas de couvercle. Les côtés et le fond de la boîte seront fait en un métal coûtant 1 euro par centimètre carré. On veut construire une boîte de coût minimum et de volume au moins 10 centimètres cube.

a. Formuler le problème.

b. On a à résoudre un problème sous contraintes. Montrer que nécessairement toute solution optimale vérifie $x_3 = \frac{10}{x_1 x_2}$. Se ramener alors à un programme linéaire sans contrainte en les variables x_1 et x_2 que l'on formulera.

2.10 Un nouvel entrepôt doit être placé de sorte que la somme des carrés des distances à 4 entrepôts déjà existants soit minimisée.
Les 4 entrepôts sont localisés aux points $(1, 2)$, $(-2, 4)$, $(2, 6)$ et $(-6, -3)$.
On note x_1 et x_2 les coordonnées du nouvel entrepôt.

a. Formuler le problème.

b. On note f la fonction à minimiser. Montrer que f est convexe.

c. Résoudre le problème en cherchant un point critique de f.