Chapitre 1

Notions Générales

1.1 Introduction

L'optimisation est la recherche des valeurs x^* appartenant à un ensemble X qui minimise [1] une fonction f définie sur X à valeurs réelles. C'est ce qu'on appelle un problème d'optimisation. On notera également ce problème comme suit

$$\min_{x \in X} f(x). \tag{1.1}$$

On dit que X est l'ensemble admissible du problème et un point de X est dit point admissible. La fonction f est appelée fonction objective ou fonction-coût. Nous supposons que X est une partie d'un espace vectoriel de dimension finie sur le corps des réels $\mathbb{R}$ que l'on pourra identifier à $\mathbb{R}^n$ et que la fonction f est assez régulière (au moins continue et si possible différentiable).

En dimension infinie on cherche plutôt une fonction qui représente une trajectoire optimale ou une forme optimale. Lorsque l'ont veut résoudre ce type de problèmes, il est nécessaire de passer par une phase de discrétisation qui, en utilisant une technique adéquat, construit un problème approché en dimension finie, qu'on pourra résoudre en utilisant des algorithmes spécifiques.

1. on peut montrer sans difficulté que c'est aussi équivalent à maximiser

1.2 Notions d'algorithmes

Un algorithmes $\mathcal{A}$ d'obtention de solutions des problèmes d'optimisation est un processus itératif qui, génère une suite de points $\{x_k\}_{k\in\mathbb{N}} \subset X$ [2] en partant d'une valeur initiale $x_0 \in X$, plus précisément

Définition 1.1. *Un algorithme qui construit itérativement une suite*

$$x_0, x_1, ..., x_k, ...$$

est définit par une application multivoque $\mathcal{A}$ de X dans $\mathcal{P}(X)$, et un ensemble de solutions désirées $\Omega \subset X$. De façon abstraite un algorithme est

1)

$$\begin{cases} x_{k+1} \in \mathcal{A}(x_k) & si\ x_k \notin \Omega, \\ x_{k+1} = x_k & si\ x_k \in \Omega. \end{cases}$$

2) *Une fonction test $\alpha : X \to \mathbb{R}$ continue telle que :*

$$x \notin \Omega \Rightarrow \alpha(y) < \alpha(x); \quad \forall y \in \mathcal{A}(x).$$

Remarque 1.1. *voici quelques exemples de l'ensemble des solutions*

- $\Omega = \{x \in X : \nabla f(x) = 0\}$,
- $\Omega = \{x \in X :$ *x est une solution optimale du problème (1.1)* $\}$,
- $\Omega = \{x \in X : f(x) < v + \varepsilon\}$, $\varepsilon > 0$ *et* v *est la valeur minimal de la fonction objective.*

Le choix de la fonction test $\alpha(x)$ est souvent $f(x)$, $\|\nabla f(x)\|$ ou $\|x - x^*\|$ ($x^* \in \Omega$).

Définition 1.2. *Soit $\mathcal{A}$ une fonction multivoque dans X, $\mathcal{A}$ est dite fermée en $x \in X$ si*

$$\begin{cases} x_k \in X, & x_k \to x \\ y_k \in \mathcal{A}(x_k), & y_k \to y \end{cases} \Rightarrow y \in \mathcal{A}(x) \tag{1.2}$$

Théorème 1.1 (Zangwill, [41]). *Si la fonction multivoque $\mathcal{A}$ est localement bornée et fermée alors tout point d'accumulation de $\{x_k\}_{k\in\mathbb{N}}$ est dans Ω.*

2. X est inclue dans un espace au moins topologique

Démonstration. Soit K tel que $\lim_K x_k = x$. Alors $\lim_K \alpha(x_k) = \alpha(x)$ (α continue), $\{\alpha(x_k)\}_{k\in\mathbb{N}^*}$ est une suite décroissante puisque $x_k \in \Omega$ ce qui implique $x_{k+1} = x_k$ et donc $\alpha(x_{k+1}) = \alpha(x_k)$. D'autre part $x_k \notin \Omega$ conduit au fait que $x_{k+1} \in A(x_k)$, donc $\alpha(x_{k+1}) < \alpha(x_k)$. Puisque $\{\alpha(x_k)\}_{k\in\mathbb{N}}$ est décroissante, alors $\lim \alpha(x_k) = \alpha(x)$.

$\{x_k\}_{k\in K}$ bornée et $\mathcal{A}$ localement bornée conduit au fait que $\{x_{k+1}\}_{k\in K}$ est bornée.

Soit $K' \subset K$ tel que $\lim_{K'} x_{k+1} = x'$, alors $\alpha(x') = \alpha(x)$.

Puisque $\mathcal{A}$ est fermée, alors $x' \in \mathcal{A}(x)$. Si $x \notin \Omega$, alors $x' \in \mathcal{A}(x)$ donc $\alpha(x') < \alpha(x)$, qui est une contradiction □

Remarque 1.2. *Ce théorème a une valeur théorique, on verra plus loin que dans des cas spécifiques, une démonstration directe de convergence est plus facile en général.*

1.3 Convergence

On dira qu'un algorithme $\mathcal{A}$ converge globalement si une sous-suite des suites $\{x_k\}_{k\in\mathbb{N}}$ de valeurs qu'il engendre tendent vers un x^* appartenant à l'ensemble de solutions Ω. Il se pose alors le problème de comparer l'efficacité des algorithmes par l'examen des suites qu'il génèrent. La notion de vitesse de convergence permet de qualifier le comportement asymptotique d'une suite.

1.3.1 Vitesse de convergence en quotient

On s'intéresse à l'étude du quotient

$$q_k = \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|\,.} \tag{1.3}$$

- Si $\limsup q_k = 1$, on dit que la convergence est sous-linéaire.
- Si $\limsup q_k = \alpha < 1$, on dit que la convergence est linéaire et α est le taux de convergence associé.
- Si $\lim q_k = 0$, on dit que la convergence est super-linéaire.
- Si $\exists \gamma > 1$ tel que *lim* sup $\dfrac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^\gamma} = M < +\infty$, on dit que la convergence est

super-linéaire d'ordre γ. En particulier pour $\gamma = 2$ on parle de vitesse de convergence quadratique.

Remarque 1.3.

- *Une suite qui converge sous-linéairement vers sa limite en pratique est considérée comme ne convergeant pas du tout.*
- *Une suite convergeant super-linéairement est convergente linéairement.*

1.3.2 Vitesse de convergence en racine

On s'intéresse à l'étude du taux

$$r_k = \|x_k - x^*\|^{1/k}. \tag{1.4}$$

- Si $\limsup r_k = \alpha < 1$, on dit que la convergence est r-linéaire et α est le taux de convergence associé.
- Si $\limsup r_k = 1$, on dit que la convergence est r-sous-linéaire.
- Si $\lim\ r_k = 0$, on dit que la convergence est r-superelinéaire.

Remarque 1.4. *Une suite convergente linéairement en quotient converge linéairement en racine.*

1.4 Notations et définitions

On note $u^\top v$ le produit scalaire Euclidien des vecteurs u et v de $\mathbb{R}^n$.

1.4.1 Définitions et résultats classiques

Définition 1.3. *pour tout $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$,*

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|\,, \quad \|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2},$$

et

$$\|x\|_\infty = \max_{1\le i\le n} |x_i|\,, \quad \|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}},$$

sont des normes vectorielles équivalentes sur $\mathbb{R}^n$.

Définition 1.4. *Soit* $\mathcal{M}_n$ *l'anneau des matrices carrées* $n \times n$ *dans* $\mathbb{R}$. *Une matrice* $A \in \mathcal{M}_n$ *est dite semi définie positive si*

$$\forall x \in \mathbb{R}^n : x^\top Ax \ge 0,$$

elle est dite définie positive si

$$\forall x \in \mathbb{R}^n_* : x^\top Ax > 0.$$

Définition 1.5.

1. *Le rayon spectral d'une matrice* $A \in \mathcal{M}_n$ *est le réel positif* $\rho(A)$ *défini par*

$$\rho(A) = \max_{\lambda\in\sigma(A)} |\lambda|\,,$$

où $\sigma(A)$*est l'ensemble des valeurs propres de* A *(spectre de* A*).*

2. *La trace d'une matrice* $A \in \mathcal{M}_n$ *(notée* $Tr(A)$*) est la quantité*

$$Tr(A) = \sum_{\lambda_i\in\sigma(A)} \lambda_i.$$

Théorème 1.2 (de Shur). *Si* A *est symétrique, il existe une matrice orthogonale* Q *telle que* $Q^{-1}AQ = Q^\top AQ$ *soit diagonale ce qui veux dire que toute matrice symétrique est diagonalisable.*

Remarque 1.5.

- *Une matrice symétrique est dite définie positive si et seulement si ses valeurs propres sont toutes positives, elle est dite semi définie positive si ses valeurs propres sont toutes non négatives, avec au moins l'une d'entre elles nulle.*
- *Soit* $A \in \mathcal{M}_n$, *si* A *est semi définie positive alors* $Tr(A) \ge 0$.

Définition 1.6. *On appelle norme matricielle subordonnée à une norme définie sur $\mathbb{R}^n$, la norme matricielle (également notée$\|.\|$) définie par*

$$\|A\| = \sup_{x\in\mathbb{R}^n_*} \frac{\|Ax\|}{\|x\|} = \sup_{x\in\mathbb{R}^n, \|x\|\leq 1} \|Ax\| = \sup_{x\in\mathbb{R}^n, \|x\|=1} \|Ax\| .$$

Par définition, on a pour toute norme matricielle subordonnée, la propriété très utile $\|Ax\| \leq \|A\| \|x\|$, l'égalité étant toujours possible pour au moins un vecteur x.

$$\|A\|_1 = \max_j \sum_i |a_{ij}| , \quad \|A\|_2 = \sqrt{\rho(A^\top A)} = \sqrt{\rho(AA^\top)} = \left\|A^\top\right\|_2 ,$$

$$\|A\|_\infty = \max_i \sum_j |a_{ij}| , \quad \|A\|_S = \sqrt{Tr(A^\top A)} = \left(\sum_{i=1,j=1}^{n} |a_{ij}|^2\right)^{1/2} \text{(norme de Schur)}.$$

Sont des normes matricielles subordonnées aux normes vectorielles de même nom.

Remarque 1.6.

- *On remarquera que $\|A\|_2$ est la plus grande valeur propre de $A^\top A$. D'autre part si A est symétrique, on a*

$$\|A\|_2 = \rho(A). \tag{1.5}$$

- *Si Q est une matrice symétrique définie positive alors elle engendre une norme sur $\mathbb{R}^n$(appelée norme de Froebinus) définie par*

$$\|x\|_Q^2 = x^\top Q\, x.$$

Définition 1.7. *Soit $A \in \mathcal{M}_n$, une matrice symétrique, on appelle quotient de Rayleigh de la matrice A en $x \in \mathbb{R}^n_*$, la quantité*

$$\mathfrak{R}(x) = \frac{x^\top Ax}{x^\top x}.$$

Ce quotient à une grande utilité dans le calcul des valeurs propres.

Théorème 1.3. *Soit $A \in \mathcal{M}_n$, une matrice symétrique, et $\lambda_1 = \min \sigma(A)$, $\max \sigma(A)$, alors*

$$\lambda_1 \leq \mathfrak{R}(x) \leq \lambda_n,$$

$\forall x \in \mathbb{R}^n_*$.

Définition 1.8. *Soit x, y deux vecteurs de $\mathbb{R}^n$, on définie le produit $xy^\top$ par*

$$\begin{aligned} xy^\top : \mathbb{R}^n &\to \mathbb{R}^n \\ d &\to (y^\top d) \cdot x \end{aligned}$$

appelé aussi produit tensorielle ou produit dyadique.

Proposition 1.1. *Soit $x \in \mathbb{R}^n$ alors la matrice $A = xx^\top$ vérifie les propriétés suivantes*

1. *A est symétrique.*
2. *$\sigma(A) = \{0, x^\top x\}$.*
3. *x est un vecteur propre qui correspond à la valeur propre $x^\top x$.*
4. *$Tr(A) = x^\top x$.*

1.4.2 Quelques résultats importants

On notera dans toute la suite, f la fonction objective, et s'il y a lieu g son gradient et H son hessien.

Formule de Taylor : Soit $f : \mathbb{R}^n \to \mathbb{R}$, une application de classe C^2. Alors

$$f(x+d) = f(x) + g(x)^\top d + \frac{1}{2} d^\top H(x) d + o(\|d\|^2).$$

Formule de Taylor avec reste intégral : Soit $f : \mathbb{R}^n \to \mathbb{R}$, une application de classe C^1, alors

$$f(x+d) = f(x) + \int_0^1 (g(x+td)^\top d) dt = f(x) + \int_x^{x+d} g(z)^\top dz. \tag{1.6}$$

Formule de la moyenne : Soit $f : \mathbb{R}^n \to \mathbb{R}$, une application de classe C^1 sur le segment $[x_1, x_2]$, alors il existe $\hat{x} \in [x_1, x_2]$ tel que

$$f(x_2) = f(x_1) + g(\hat{x})^\top (x_2 - x_1).$$

Il faut noter que les résultats précédents restent valables, si f est une fonction vectorielle.

1.5 Les méthodes numériques de résolution des systèmes d'équations non linéaires

On verra plus loin qu'un problème d'optimisation puisse être transformé en une recherche de solutions d'un système d'équations non linéaires, donc la première idée naturelle est d'appliquer une méthode numérique de résolution de ces systèmes. Ce qui nous ramènes à étudier deux de telles méthodes : Gauss-Seidel et approximation successives, qui sont des méthodes de premier ordre.

1.5.1 Méthode de Gauss-Seidel

On veut résoudre le système non linéaire

$$\begin{cases} F(x) = 0, \\ F(x) = (F_1(x), F_2(x), .., F_i(x), .., F_n(x)); 1 \leq i \leq n, \\ x = (x_1, ..., x_j, ..., x_n); \ 1 \leq j \leq n, \end{cases} \tag{1.7}$$

où, $F : \mathbb{R}^n \to \mathbb{R}^n$ est une fonction vectorielle dans $\mathbb{R}^n$.

Algorithme 1.1.

Initialisation : $i = k = 0$, $(x_1^0, ..., x_n^0) = (0, ..., 0)$

(1) Test d'arrêt : Si $\left\| F(x_1^k, ..., x_n^k) \right\| \simeq 0$, *stop*

Sinon $k \leftarrow k + 1$

(2) $i \leftarrow i + 1$

On résolve l'équation $F_i(x_1^{k+1}, ..., x_{i-1}^{k+1}, x_i, x_{i+1}^k, ...x_n^k) = 0$ *par rapport à* x_i

$x_i^{k+1} \leftarrow x_i$

si $i < n$ *aller à (2)*

sinon

aller à (1)

Remarque 1.7.

- *Cette méthode consiste à résoudre une équation (non linéaire) à une inconnue à chaque itération.*
- *La convergence est mal connue, donc la méthode est peut intéressante.*

1.5.2 La méthode des approximations successives

Pour résoudre le système (1.7), on utilise le schéma itératif

$$x_{k+1} = x_k + t \cdot F(x_k); \; t < 0, \tag{1.8}$$

t est un coefficient fixé.

Définition 1.9. *Une fonction F est dite localement lipschitzienne sur un borné B si $\exists M > 0$, $\forall x, y \in B$*

$$\|F(x) - F(y)\| \leq M \|x - y\| . \tag{1.9}$$

Elle est dite localement coercive si, Il existe $m > 0$ tels que, $\forall x, y \in B$

$$m \|x - y\|^2 \leq (F(x) - F(y))^\top (x - y). \tag{1.10}$$

Théorème 1.4. *On suppose que F est localement lipschitzienne et coercive et soit x^* une solution de (1.7). Alors le shéma (1.8), converge si $t < 0$ suffisamment petit.*

Démonstration.

Soit x_0 la première itération et $x^* \in \mathbb{R}^n$ une solution (i.e. $F(x^*) = 0$), et prenant $B = \overline{B}(x_0, \|x_0 - x^*\|)$ (la boule fermé de centre x_0 et de rayon $\|x_0 - x^*\|$). Puisque $x_1 = x_0 +$

$tF(x_0)$ et $F(x^*) = 0$, on a

$$\begin{aligned}
\|x_1 - x^*\|^2 &= (x_1 - x^*)^\top (x_1 - x^*) \\
&= x_1^T x_1 + x^{*\top} x^* - 2x_1^\top x^* \\
&= (x_0 + tF(x_0))^\top (x_0 + tF(x_0)) + (x_0 + tF(x_0))^\top x^* \\
&= x_0^\top x_0 + x^{*\top} x^* - 2x_0^\top x^* + t^2 F(x_0)^\top F(x_0) + 2t(x_0 - x^*)^\top F(x_0) \\
&= \|x_0 - x^*\| + 2t(x_0 - x^*)^\top (F(x_0) - F(x^*)) + \\
&\quad + t^2 \|F(x_0) - F(x^*)\|^2 \|x_0 - x^*\|^2 .
\end{aligned}$$

Prenons $t < 0$, alors d'après (1.9) et (1.10)

$$\|x_2 - x^*\|^2 \leq (1 + 2mt + M^2 t^2) \|x_1 - x^*\|^2 . \tag{1.11}$$

Il suffit de prendre $t > -2\frac{m}{M^2}$ pour avoir

$$0 \leq K = 1 + 2mt + M^2 t^2 < 1.$$

On itère la formule (1.11) pour avoir

$$\|x_k - x^*\|^2 \leq K^k \|x_1 - x^*\|^2 . \tag{1.12}$$

□

Remarque 1.8.

- *On peut déduire de (1.11) que*

$$\|x_{k+1} - x^*\| \leq K \|x_k - x^*\| ,$$

  avec $0 \leq K < 1$, *donc* $\{x_k\}$ *converge linéairement vers* x^*.

- *La solution* x^* *est unique, car si* $\bar{x}^*$ *est une autre solution on a une contradiction avec (1.10)*

Exercices

Exercice 1.1. *Calculer s'il y a lieux* $g(x)$, $H(x)$, *et* $Jf(x)$ *(le jacobien de la fonction* f *en* x*) pour*

1. $f : \mathbb{R}^n \to \mathbb{R},\ x \mapsto f(x) = c^t x + \gamma.$
2. $f : \mathbb{R}^n \to \mathbb{R}^n,\ x \mapsto F(x) = Lx + b$, *où* $L \in M_{m,n}(\mathbb{R})$.
3. $f : \mathbb{R}^n \to \mathbb{R},\ x \mapsto f(x) = \dfrac{1}{2} x^t Ax + d^t x + \delta$, *où* $L \in M_n(\mathbb{R})$.
4. $f : \mathbb{R}^n \to \mathbb{R},\ x \mapsto f(x) = \sum_{i=1}^{m} [r_i(x)]^2$, *où les* $r_i : \mathbb{R}^n \to \mathbb{R}$ *sont deux fois différentiables.*

Exercice 1.2. *Montrer les propositions suivantes*

1. *Soit* $f : \mathbb{R}^n \to \mathbb{R}$, *différentiable en* $\bar{x}$. *Si* $\bar{x}$ *est un minimum local, alors* $g(\bar{x}) = 0$.
2. *Soit* $f : \mathbb{R}^n \to \mathbb{R}$, *deux fois différentiable en* $\bar{x}$ *et supposons que* $\bar{x}$. *soit un minimum local, alors* $g(\bar{x}) = 0$ *et* $H(\bar{x})$ *est semi défini positif.*
3. *Soit* $f : \mathbb{R}^n \to \mathbb{R}$, *deux fois différentiable en* $\bar{x}$. *Si* $g(\bar{x}) = 0$ *et* $H(\bar{x})$ *est défini positif, alors* $\bar{x}$ *est un minimum local strict.*
4. *Si* f *est convexe sur* $\mathbb{R}^n$ *et si* $g(\bar{x}) = 0$ *alors* $\bar{x}$ *est un minimum global.*

Exercice 1.3. *Soit* $\{x_k\}$ *une suite de vecteurs dans* $\mathbb{R}^n$. *Montrez que si* x *tend Q-linéairement vers* x^*, *alors pour tout* $q > \limsup q_k$, *il existe* k_0 *et* $C > 0$ *tels que*

$$\|x_k - x^*\| \leq Cq^k \text{ pour tout } k \geq k_0.$$

où $q_k = \dfrac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}$.

Exercice 1.4. *Montrer que la suite* $x_k = \dfrac{1}{k}$ *n'est pas convergente Q-linéairement vers* 0.

Exercice 1.5. *Montrer que la suite* $x_k = 1 + (0.5)^{2^k}$ *convergent Q-qudratiquement vers* 1.

Exercice 1.6. *Est-ce que la suite* $\dfrac{1}{k!}$ *converge Q-superlinairement ? Q-quadratiquement ?*

Exercice 1.7. *Considérons la suite* $\{x_k\}$ *définie*

$$x_k = \begin{cases} (\frac{1}{4})^{2^k}, & k \text{ est pair} \\ \dfrac{x_{k-1}}{k}, & k \text{ est impaire} \end{cases}$$

Est-ce que la suite est Q-superlineairement convergente ? Q-quadratiquement convergente ? R-quadratiquement convergente ?

Exercice 1.8. *Soit* $f : \mathbb{R}^n \to \mathbb{R}$*, une fonction continument différentiable. Soit* x_0 *tel que* $f(x_0) = 0$*.*

1. *Que représente* $g(x_0)$ *pour la surface de niveau*

$$S := \{x \in \mathbb{R}^n \, : \, f(x) = f(x_0)\} \, .$$

2. *Donner l'équation de l'hyperplan affine (de* $\mathbb{R}^{n+1}$*) tangent au graphe de* f *en* $(x_0 \, , \, f(x_0))$*.*

3. *Donner à l'aide de* $g(x_0)$ *un vecteur normal à cet hyperplan.*

Exercice 1.9. *Soit* $f : \mathbb{R}^n \to \mathbb{R}$*, une fonction continument différentiable.*

1. *On suppose qu'il existe* $L > 0$ *tel que*

$$\left\|g(x) - g(x')\right\| \leq L \, \|x - x_0\| \text{ pour tout } (x \, , \, x') \in \mathbb{R}^n \times \mathbb{R}^n.$$

2. *Montrer qu'alors*

$$\left|f(x+d) - f(x) - g(x)^t d\right| \leq \frac{L}{2} \, \|d\|^2 \text{ pour tout } (x \, , \, d) \in \mathbb{R}^n \times \mathbb{R}^n.$$