

Thème d'analyse numérique
Prépa agreg Maths

Thomas Ourmières-Bonafos

Table des matières

1	Cas des système linéaires	5
1.1	Brefs rappels d'algèbre linéaire	5
1.1.1	Décomposition en valeurs singulières	5
1.1.2	Matrices à diagonale dominante	6
1.1.3	Normes subordonnées et rayon spectral	7
1.1.4	Conditionnement d'une matrice inversible	7
1.2	Résolutions de systèmes linéaires par des méthodes itératives	8
1.3	Recherche de valeurs et vecteurs propres	9
1.4	Méthode des moindres carrés	10
1.5	Discrétisation du Laplacien en dimension 1	11
2	Recherche de zéro	13
2.1	L'algorithme de dichotomie	13
2.2	Méthodes de point fixe	13
2.2.1	Méthode de Banach-Picard	14
2.2.2	Méthode de Newton dans $\mathbb{R}^d$	16
2.2.3	Optimisation et algorithme du gradient à pas constant	16
3	Intégration Numérique	19
3.1	Méthode des rectangles	19
4	Équations différentielles : la méthode d'Euler explicite	23
4.1	Cadre d'étude	23
4.2	La méthode d'Euler explicite	23
4.3	Consistance	24
4.4	Stabilité	24
4.5	Convergence	24
A	Corrections des exercices	27

Chapitre 1

Cas des système linéaires

Le but de ce chapitre est de chercher les solutions $x \in \mathbb{C}^d$ de systèmes linéaires de la forme

$$Ax = b$$

où $A \in \mathcal{M}_{q,d}(\mathbb{C})$ et $b \in \mathbb{C}^q$. Attention, il est possible que $q \neq d$! On cherche à développer des méthodes numériques, c'est-à-dire qui peuvent être implémentées dans un ordinateur.

1.1 Brefs rappels d'algèbre linéaire

1.1.1 Décomposition en valeurs singulières

Le but de ce paragraphe est de démontrer le résultat suivant.

Proposition 1. Soit $M \in \mathcal{M}_{q,d}(\mathbb{C})$, alors il existe deux matrices unitaires¹ $U \in \mathcal{M}_{q,q}(\mathbb{C})$, $V \in \mathcal{M}_{d,d}(\mathbb{C})$ et $\Sigma \in \mathcal{M}_{q,d}(\mathbb{R})$ dont les seuls coefficients non-nuls sont sur la diagonale et telles que

$$M = U\Sigma V^*.$$

De plus, on a $\text{rang}(M) = \text{rang}(\Sigma)$.

Exercice 1. Le but de cet exercice est de démontrer la Proposition 1. Soit $M \in \mathcal{M}_{q,d}(\mathbb{C})$.

- (i) Montrer qu'il existe une base orthonormée $(e_1, \dots, e_d)$ de $\mathbb{C}^d$ et des réels $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \geq 0$ tels que pour tout $x \in \mathbb{C}^d$ on a

$$M^* M x = \sum_{j=1}^d \lambda_j \langle x, e_j \rangle_{\mathbb{C}^d} e_j$$

- (ii) On note $r = \#\{j \in \{1, \dots, d\} : \lambda_j > 0\}$. Montrer que si $r \geq 1$ alors la famille définie pour tout $j \in \{1, \dots, r\}$ par

$$e_j^* = \frac{1}{\sqrt{\lambda_j}} M e_j$$

est une famille orthonormée de $\mathbb{C}^q$.

- (iii) Montrer que pour tout $j \in \{r+1, \dots, d\}$, on a $M e_j = 0$.

- (iv) On considère $(\varepsilon_1, \dots, \varepsilon_d)$ et $(\varepsilon_1^*, \dots, \varepsilon_q^*)$ les bases canoniques de $\mathbb{C}^d$ et $\mathbb{C}^q$ respectivement. Montrer que pour tout $x \in \mathbb{C}^d$ on a

$$M x = \sum_{j=1}^r \sum_{p=1}^d \sqrt{\lambda_j} \langle x, \varepsilon_p \rangle_{\mathbb{C}^d} \langle \varepsilon_p, e_j \rangle_{\mathbb{C}^d} e_j^*$$

et en déduire que pour tout $j \in \{1, \dots, q\}$, $k \in \{1, \dots, d\}$ on a

$$M_{j,k} = \sum_{s=1}^r \sqrt{\lambda_s} \langle \varepsilon_k, e_s \rangle_{\mathbb{C}^d} \langle e_s^*, \varepsilon_j^* \rangle_{\mathbb{C}^q}.$$

1. Une matrice carrée $U \in \mathcal{M}_{d,d}(\mathbb{C})$ est dite unitaire lorsque $U^* U = U U^* = Id$.

- (v) Si $r < q$, on complète la famille $(e_1^*, \dots, e_r^*)$ en $(e_1^*, \dots, e_r^*, e_{r+1}^*, \dots, e_q^*)$, une base orthonormée de $\mathbb{C}^q$. On définit pour tout $j \in \{1, \dots, d\}$ et $k \in \{1, \dots, r\}$

$$\Sigma_{j,k} = \begin{cases} \sqrt{\lambda_j} & \text{si } j = k \text{ et } j \leq r \\ 0 & \text{sinon} \end{cases}$$

et U est la matrice dont les colonnes sont les vecteurs $(e_1^*, \dots, e_q^*)$ et V celle dont les colonnes sont les vecteurs $(e_1, \dots, e_d)$. Montrer que

$$M = U\Sigma V^*.$$

Exercice 2. On considère la matrice

$$M = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \end{pmatrix} \in \mathcal{M}_{2,3}(\mathbb{R}).$$

Donner une décomposition en valeurs singulières de M et l'utiliser pour donner l'ensemble des solutions $X \in \mathbb{R}^3$ de

$$MX = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

1.1.2 Matrices à diagonale dominante

Proposition 2 (Théorème de Gershgoring). Soit $M = (m_{j,k})_{j,k \in \{1, \dots, d\}} \in \mathcal{M}_d(\mathbb{C})$. Si λ est une valeur propre de M alors

$$\lambda \in \bigcup_{j=1}^d D_j$$

où D_j est le j -ème disque de Gershgoring défini comme

$$D_j := \{z \in \mathbb{C} : |m_{j,j} - z| \leq \sum_{\substack{k=1 \\ k \neq j}}^d |m_{j,k}|\}$$

Un corollaire immédiat de la Proposition 2 est le suivant.

Corollaire 1. Soit $M = (m_{j,k})_{j,k \in \{1, \dots, d\}} \in \mathcal{M}_d(\mathbb{C})$. Si λ est une valeur propre de M alors

$$|\lambda| \leq \sup_{j \in \{1, \dots, d\}} \sum_{k=1}^d |m_{j,k}|.$$

Exercice 3 (Matrices à diagonale dominante et disques de Gershgoring). Le but de cet exercice est démontrer la Proposition 2 qui permet de localiser dans le plan complexe les possibles valeurs propres d'une matrice $M \in \mathcal{M}_d(\mathbb{C})$. On dit qu'une matrice $A = (a_{j,k})_{j,k \in \{1, \dots, d\}} \in \mathcal{M}_d(\mathbb{C})$ est à diagonale dominante si pour tout $j \in \{1, \dots, d\}$ on a

$$\sum_{\substack{k=1 \\ k \neq j}}^d |a_{j,k}| \leq |a_{j,j}|.$$

Elle est dite strictement dominante lorsque l'inégalité est stricte.

- (i) Démontrer que si $A = (a_{j,k})_{j,k \in \{1, \dots, d\}}$ est une matrice à diagonale strictement dominante alors A est inversible.
- (ii) En déduire la Proposition 2.

Exercice 4. On considère la matrice

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 1 & 0 & 4 \end{pmatrix}$$

- (i) Montrer, sans calculer les valeurs propres de A , que si $\lambda \in Sp(A)$ alors

$$\lambda \in [0, 5].$$

- (ii) Le vérifier en montrant que A est diagonalisable et en calculant ses valeurs propres.

1.1.3 Normes subordonnées et rayon spectral

On se donne $\|\cdot\|_{\mathbb{C}^d}$ et $\|\cdot\|_{\mathbb{C}^q}$ deux normes sur $\mathbb{C}^d$ et $\mathbb{C}^q$ respectivement. Pour une matrice $M \in \mathcal{M}_{q,d}(\mathbb{C})$, on définit

$$\|M\| = \sup_{\substack{x \in \mathbb{C}^d \\ x \neq 0}} \frac{\|Mx\|_{\mathbb{C}^q}}{\|x\|_{\mathbb{C}^d}}.$$

On rappelle qu'il s'agit d'une norme sur $\mathcal{M}_{q,d}(\mathbb{C})$ et que pour $M \in \mathcal{M}_{q,d}(\mathbb{C})$ et $N \in \mathcal{M}_{d,p}(\mathbb{C})$ on a

$$\|MN\| \leq \|M\| \|N\|.$$

Proposition 3. On peut exprimer ces normes triples explicitement dans certains cas particuliers.

(i) Si les normes sur $\mathbb{C}^d$ et $\mathbb{C}^q$ sont la norme 1, on a

$$\|M\| = \max_{k \in \{1, \dots, d\}} \sum_{j=1}^q |m_{j,k}|.$$

(ii) Si les normes sur $\mathbb{C}^d$ et $\mathbb{C}^q$ sont la norme 2, on a

$$\|M\| = \sqrt{\max(\text{Sp}(M^*M))}.$$

(iii) Si les normes sur $\mathbb{C}^d$ et $\mathbb{C}^q$ sont la norme ∞ , on a

$$\|M\| = \max_{j \in \{1, \dots, q\}} \sum_{k=1}^d |m_{j,k}|.$$

Remarque 1. Souvent, pour une matrice hermitienne $A \in \mathcal{M}_{d,d}(\mathbb{C})$ (donc diagonalisable), on note

$$\rho(A) = \max \text{Sp}(A) = \max\{\lambda_j : j \in \{1, \dots, d\}\}$$

où pour $j \in \{1, \dots, d\}$, les $\lambda_j \in \mathbb{R}$ sont les valeurs propres de A . $\rho(A)$ est le rayon spectral de A et pour les normes 2, le point (ii) de la Proposition 3 devient

$$\|M\| = \sqrt{\rho(M^*M)}.$$

Exercice 5. Démontrer les trois points de la Proposition 3.

1.1.4 Conditionnement d'une matrice inversible

On choisit $M \in GL_d(\mathbb{C})$ et, théoriquement, on sait que si $b \in \mathbb{C}^d$, une solution $x \in \mathbb{C}^d$ du système

$$Mx = b$$

est donnée par $x = M^{-1}b$. Numériquement, on cherche à comprendre la stabilité des solutions par rapport à de petites perturbations de la donnée b . C'est le but du théorème suivant.

Théorème 1. Soit $M \in GL_d(\mathbb{C})$ et soit $b, \delta b \in \mathbb{C}^d$. On pose $x := M^{-1}b$ et $x + \delta x = M^{-1}(b + \delta b)$. Si $b \neq 0$, on a

$$\frac{\|\delta x\|}{\|x\|} \leq (\|M\| \|M^{-1}\|) \frac{\|\delta b\|}{\|b\|}.$$

On peut interpréter le Théorème 1 comme un contrôle de l'erreur relative faite sur la solution par rapport à l'erreur relative de la donnée initiale.

Remarque 2. Une quantité importante apparaît dans le Théorème 1. Il s'agit du conditionnement κ de la matrice M défini comme :

$$\kappa(M) := \|M\| \|M^{-1}\|.$$

Intuitivement, plus κ est petit, meilleur est le contrôle de l'erreur relative de la solution par rapport à l'erreur relative sur les données.

Exercice 6. Démontrer le Théorème 1.

La notion de conditionnement est cruciale lorsqu'il s'agit, en pratique, d'inverser une matrice. Par définition, cette notion dépend de la norme choisie ainsi, pour $p \in \{1, 2, \infty\}$, et $M \in GL_d(\mathbb{C})$, on définit

$$\kappa_p(M) = \|M\|_p \|M^{-1}\|_p.$$

On a

Proposition 4. Soit $M \in GL_d(\mathbb{C})$. On a

- (i) $\kappa(M) \geq 1$, $\kappa(M) = \kappa(M^{-1})$ et pour tout $\alpha \in \mathbb{C}$, $\kappa(\alpha M) = \kappa(M)$.
- (ii) On a

$$\kappa_2(M) = \frac{\sigma_1(M)}{\sigma_d(M)}$$

où $\sigma_1(M) > 0$ est la plus grande valeur singulière de M et $\sigma_d(M) > 0$ sa plus petite valeur singulière.

- (iii) Si $M^*M = MM^* = Id$ alors $\kappa_2(M) = 1$.
- (iv) Si $U \in \mathcal{M}_{d,d}(\mathbb{C})$ est telle que $U^*U = Id$ alors $\kappa_2(M) = \kappa_2(UM) = \kappa_2(MU) = \kappa_2(U^*MU)$.

Exercice 7. Démontrer la Proposition 4.

Exercice 8. Pour $\varepsilon > 0$, on considère la matrice

$$M_\varepsilon = \begin{pmatrix} 1 & 1 + \varepsilon \\ 1 & 1 \end{pmatrix}.$$

- (i) Montrer que M_ε est inversible et donner son inverse.
- (ii) Donner la solution x_ε de $M_\varepsilon x_\varepsilon = \begin{pmatrix} \pi \\ e \end{pmatrix}$.
- (iii) On suppose que l'on a une précision machine de $\eta > 0$ et que l'on a un vecteur $b \in \mathbb{C}^2$ qui est à une distance relative au plus $\eta > 0$ du vecteur $\begin{pmatrix} \pi \\ e \end{pmatrix}$, c'est-à-dire

$$\frac{\left\| b - \begin{pmatrix} \pi \\ e \end{pmatrix} \right\|_{\ell^2(\mathbb{C}^2)}}{\left\| \begin{pmatrix} \pi \\ e \end{pmatrix} \right\|_{\ell^2(\mathbb{C}^2)}} < \eta.$$

Donner une estimation de l'erreur relative commise en résolvant $M_\varepsilon x = b$ par rapport à la solution de référence de $M_\varepsilon x = \begin{pmatrix} \pi \\ e \end{pmatrix}$.

- (iv) Si notre précision machine vaut $\eta = 10^{-6}$, est-il possible pour de petites valeurs de $\varepsilon > 0$ d'obtenir une erreur relative de l'ordre de 10^{-6} sur la solution de $M_\varepsilon x = b$?

1.2 Résolutions de systèmes linéaires par des méthodes itératives

Nous avons vu qu'une résolution "exacte" d'un système linéaire par l'inversion d'une matrice peut être une mauvaise idée si la matrice en question est mal conditionnée. On considère alors des méthodes dites itératives : on fabrique une suite de vecteurs qui, si la méthode est bien pensée, converge vers la solution de notre système linéaire.

Soit $A \in GL_d(\mathbb{C})$, $b \in \mathbb{C}^d$ et $x \in \mathbb{C}^d$ solution de $Ax = b$. On suppose qu'il existe $M \in GL_d(\mathbb{C})$ et $N \in \mathcal{M}_{d,d}(\mathbb{C})$ telle que $A = M - N$. On a alors

$$x = M^{-1}Nx + M^{-1}b. \quad (1.1)$$

Si on pose

$$f : \begin{cases} \mathbb{R}^d & \rightarrow \mathbb{R}^d \\ x & \mapsto M^{-1}Nx + M^{-1}b \end{cases}$$

alors (1.1) se ré-écrit $f(x) = x$ et trouver les solutions de $Ax = b$ revient à trouver les points fixes de f . On est alors tentés de fabriquer une méthode itérative, c'est à dire de se donner une suite $(x_n)_{n \in \mathbb{N}}$ définie pour $x_{in} \in \mathbb{C}^d$ par

$$\begin{cases} x_{n+1} &= f(x_n) \\ x_0 &= x_{in}. \end{cases}$$

En effet, comme la fonction f est continue sur $\mathbb{R}^d$ (pourquoi?), si la suite $(x_n)_{n \in \mathbb{N}}$ converge vers $x \in \mathbb{C}^d$, on a nécessairement $x = f(x)$ et $f(x)$ est donc la solution de $Ax = b$.

Définition 1. On dit que les méthodes itératives convergent si pour tout vecteur initial $x_{in} \in \mathbb{C}^d$, on a

$$\lim_{n \rightarrow +\infty} x_n = x$$

où x est l'unique solution de $Ax = b$.

Proposition 5. Les propositions suivantes sont équivalentes.

- (i) La méthode itérative est convergente,
- (ii) $\rho(M^{-1}N) < 1$,
- (iii) $\|M^{-1}N\| < 1$ pour (au moins) une norme matricielle.

Exercice 9. Démontrer la Proposition 5.

Exercice 10 (Méthode de Jacobi). La méthode de Jacobi consiste en un choix de "découpage" de $A \in GL_d(\mathbb{C})$ sous la forme $A = M + N$. Si $A = (a_{j,k})_{j,k \in \{1, \dots, d\}}$, on suppose que A est à diagonale strictement dominante. On définit alors les matrices $M = (m_{j,k})_{j,k \in \{1, \dots, d\}}$ et $N = (n_{j,k})_{j,k \in \{1, \dots, d\}}$ par

$$m_{j,k} = a_{j,k} \delta_{j,k}, \quad n_{j,k} = \begin{cases} -a_{j,k} & \text{si } j \neq k \\ 0 & \text{sinon} \end{cases}$$

- (i) Montrer que la méthode itérative de Jacobi peut se mettre sous la forme

$$x_{n+1} = Jx_n + c$$

avec $J \in \mathcal{M}_{d,d}(\mathbb{C})$ et $c \in \mathbb{C}^d$ une matrice et un vecteur que l'on exprimera en fonction de M, N et b .

- (ii) Montrer que la méthode de Jacobi converge.

Exercice 11 (Méthode de Gauss-Seidel). La méthode de Gauss-Seidel consiste en un choix de "découpage" de $A \in GL_d(\mathbb{C})$ sous la forme $A = M - N$ où $M = (m_{j,k})_{j,k \in \{1, \dots, d\}}$ et $N = (n_{j,k})_{j,k \in \{1, \dots, d\}}$ vérifient pour tout $j, k \in \{1, \dots, d\}$

$$m_{j,k} = \begin{cases} a_{j,k} & \text{si } j \geq k \\ 0 & \text{sinon} \end{cases}, \quad n_{j,k} = \begin{cases} -a_{j,k} & \text{si } j > k \\ 0 & \text{sinon} \end{cases}.$$

On suppose que les coefficients diagonaux de A sont non-nuls.

- (i) Montrer que la méthode itérative de Gauss-Seidel peut se mettre sous la forme

$$x_{n+1} = Gx_n + c$$

avec $G \in \mathcal{M}_{d,d}(\mathbb{C})$ et $c \in \mathbb{C}^d$ une matrice et un vecteur que l'on exprimera en fonction de M, N et b .

- (ii) Montrer que si $M^* + N$ est hermitienne et définie positive, alors la méthode de Gauss-Seidel converge.

1.3 Recherche de valeurs et vecteurs propres

On cherche à produire une méthode itérative pour calculer les valeurs et vecteurs propres d'une matrice donnée. C'est le but de la méthode dite de la puissance : on se donne $M \in \mathcal{M}_{d,d}(\mathbb{C})$ une matrice diagonalisable et on note $\lambda_1, \dots, \lambda_d$ ses valeurs propres et on suppose que

$$|\lambda_1| > \max_{j \in \{2, \dots, d\}} (|\lambda_j|).$$

On note e_j un vecteur propre associé à λ_j (pour $j \in \{1, \dots, d\}$). Pour $x_{in} \in \mathbb{C}^d$, on considère alors la suite

$$\begin{cases} x_{n+1} &= \frac{Mx_n}{\|Mx_n\|_{\ell^2(\mathbb{C}^d)}} \\ x_0 &= x_{in}. \end{cases}$$

On a alors la proposition suivante.

Proposition 6. Si $x_{in} \notin \text{vect}(e_2, \dots, e_d)$, alors la suite est bien définie et il existe $\alpha \in \mathbb{C}$ tel que

$$\lim_{n \rightarrow +\infty} x_n = \alpha e_1.$$

De plus, on a

$$\lim_{n \rightarrow +\infty} \langle x_{n+1}, x_n \rangle_{\mathbb{C}^d} = \lambda_1.$$

Exercice 12 (Méthode de la puissance).

- (i) Montrer que pour tout $n \in \mathbb{N}$, on a $Mx_n \neq 0_{\mathbb{C}^d}$.
- (ii) Montrer que pour tout $n \in \mathbb{N}^*$, on a $x_n = \frac{M^n x_0}{\|M^n x_0\|}$.
- (iii) Montrer que la suite de terme général $\lim_{n \rightarrow +\infty} \lambda_1^{-n} M^n x_0 = x_{0,1} e_1$.
- (iv) En déduire la limite de la suite de terme général $\left(\frac{\bar{\lambda}_1}{|\lambda_1|}\right)^n x_n$ lorsque $n \rightarrow +\infty$.
- (v) Montrer que

$$\lim_{n \rightarrow +\infty} \langle x_{n+1}, x_n \rangle_{\mathbb{C}^d} = \lambda_1.$$

1.4 Méthode des moindres carrés

Soit $N \in \mathbb{N}^*$ et $(x_j, y_j)_{1 \leq j \leq N}$ un nuage de points de $\mathbb{R}^2$. Pour $d \in \mathbb{N}^*$, on considère le problème de minimisation suivant

$$\inf \left\{ \sum_{j=1}^N |y_j - P(x_j)|^2 : P \in \mathbb{R}_d[X] \right\}.$$

Autrement dit, on cherche un polynôme P_* de degré au plus d qui approche le « mieux possible » le nuage de point. L'expression le « mieux possible » voulant dire ici qu'au sens de la norme 2 dans $\mathbb{R}^N$ les vecteurs $y = (y_1, \dots, y_N)$ et $P_* = (P_*(x_1), \dots, P_*(x_N))$ sont les plus proches possibles. Ce problème est en fait un problème d'optimisation, comme nous le verrons plus tard dans un cadre plus général.

Si on identifie $\mathbb{R}_d[X]$ avec $\mathbb{R}^{d+1}$ grâce à l'isomorphisme canonique

$$\varphi := \begin{cases} \mathbb{R}_d[X] & \longrightarrow \mathbb{R}^{d+1} \\ P(X) = \sum_{j=0}^d a_j X^j & \mapsto (a_0, \dots, a_d). \end{cases}$$

On cherche à minimiser la fonction

$$f := \begin{cases} \mathbb{R}^d & \longrightarrow \mathbb{R} \\ a = (a_0, \dots, a_d) & \mapsto \sum_{j=1}^N \left| y_j - \sum_{k=0}^d a_k x_j^k \right|^2. \end{cases}$$

Exercice 13. Montrer que le problème des moindres carrés se ré-écrit sous la forme

$$\inf_{a \in \mathbb{R}^{d+1}} \|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2$$

où on explicitera la $M \in \mathcal{M}_{N,d+1}(\mathbb{R})$ et $b \in \mathbb{R}^N$.

Proposition 7. La fonction f admet un minimum sur $\mathbb{R}^d$ en un point a qui vérifie l'équation dite équation normale suivante

$$M^\top Ma = M^\top b.$$

Toute autre solution est de la forme $a + v$ où $v \in \ker(M^\top M)$.

Exercice 14. Nous allons démontrer la Proposition 7.

- (i) En utilisant la décomposition en valeurs singulières de M , montrer l'existence d'un minimiseur $a \in \mathbb{R}^{d+1}$.
- (ii) Soit $a, h \in \mathbb{R}^{d+1}$, montrer que l'on a

$$\|M(a+h) - b\|_{\ell^2(\mathbb{R}^N)}^2 = \|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2 + \|Mh\|_{\ell^2(\mathbb{R}^N)}^2 + 2\langle M^\top Ma - M^\top b, h \rangle_{\mathbb{R}^{d+1}}.$$

- (iii) En déduire que si $a \in \mathbb{R}^{d+1}$ vérifie l'équation normale alors a est un minimiseur de f .
- (iv) On suppose maintenant que a est un minimiseur de f , montrer que a vérifie les équations normales.

1.5 Discrétisation du Laplacien en dimension 1

Soit $g \in \mathcal{C}([0, 1])$, on cherche une fonction $f \in \mathcal{C}^2([0, 1])$ solution de l'équation

$$\begin{cases} -f''(t) = g(t) & \text{pour } t \in]0, 1[\\ f(0) = f(1) = 0. \end{cases} \quad (1.2)$$

Exercice 15. On cherche une solution de la forme d'une série

$$f(t) = \sum_{k=1}^{+\infty} f_k \sin(k\pi t).$$

On va raisonner par analyse et synthèse.

(i) Supposons qu'une telle solution existe, montrer que pour tout $k \in \mathbb{N}^*$, on a nécessairement

$$f_k = \frac{g_k}{k^2\pi^2}$$

où on précisera le coefficient g_k .

(ii) Montrer que la fonction f définie avec ce choix de f_k vérifie $f \in \mathcal{C}^2([0, 1])$, $f(0) = f(1) = 0$ et $-f''(t) = g(t)$.

(iii) Montrer que f est l'unique solution de ce problème.

D'un point de vue numérique, il est parfois difficile d'obtenir une telle solution f : il faut calculer les coefficients g_k et tronquer la série à un certain ordre. Il y a aussi des phénomènes numériques, notamment le phénomène de Gibbs qui traduit l'absence de convergence uniforme de la série tronquée vers sa limite.

Nous allons adopter une autre stratégie en discrétisant directement le problème 1.2. Pour ce faire, on se donne $N \in \mathbb{N}^*$ et pour $j \in \{1, \dots, N\}$, on pose $h = \frac{1}{N+1}$ et $t_j := \frac{j}{h}$. Pour $j \in \{1, \dots, N\}$, on va approcher la dérivée seconde de f en un point t_j par sa différence finie d'ordre 2

$$f''(t_j) \simeq \frac{f(t_{j+1}) + f(t_{j-1}) - 2f(t_j)}{h^2}$$

et on va ainsi chercher un vecteur $(f_j)_{j \in \{1, \dots, N\}} \in \mathbb{R}^N$ qui vérifie

$$\begin{cases} -\frac{f_{j+1} + f_{j-1} - 2f_j}{h^2} = g(t_j) & \text{pour } j \in \{1, \dots, N\} \\ f_0 = f_{N+1} = 0. \end{cases} \quad (1.3)$$

Ainsi construit, on espère que pour tout $j \in \{1, \dots, N\}$, f_j soit une bonne approximation de $f(t_j)$.

Exercice 16. Montrer que le système 1.3 se met sous la forme

$$Mx = b$$

où $M \in \mathcal{M}_{N,N}(\mathbb{R})$, $x, b \in \mathbb{R}^N$ sont à préciser.

La matrice M trouvée à l'exercice précédent est la matrice du Laplacien. On va démontrer certaines de ses propriétés dans l'exercice suivant.

Exercice 17.

(i) Montrer que M est diagonalisable et que ses valeurs propres sont strictement positives. On pourra regarder le produit scalaire $\langle Mx, x \rangle$.

(ii) On pose $x_0, x_{N+1} = 0$. Montrer que si $x = (x_1, \dots, x_N)^\top$ est un vecteur propre de M associé à λ alors

$$x_{k+2} + (\lambda - 2)x_{k+1} + x_k = 0$$

pour tout $k \in \{0, \dots, N-1\}$.

(iii) En utilisant les disques de Gershgorin, démontrer que $\lambda \in [0, 4]$ et en déduire que nécessairement les solutions sont de la forme

$$x_k = 2\alpha \sin(k \frac{p\pi}{N+1}), \quad \alpha \in \mathbb{C}, \quad p \in \{1, \dots, N\}.$$

(iv) En déduire que les valeurs propres de M sont données pour $p \in \{1, \dots, N\}$ par

$$\lambda_p := 2 - 2 \cos \left(\frac{p\pi}{N+1} \right).$$

(v) Montrer que $\kappa_2(M) \underset{N \rightarrow +\infty}{\sim} \frac{4N^2}{\pi^2}$.

(vi) Soit x la solution de $Mx = b$. Montrer que si f est solution de (1.2) et que de plus $f \in \mathcal{C}^4([0, 1])$ alors, si on pose $F = (f(t_1), \dots, f(t_N))^T \in \mathbb{R}^N$ on a

$$\|M(x - F)\|_{\ell^\infty(\mathbb{R}^d)} \leq Ch^4$$

(vii) Montrer que $\|x - F\|_{\ell^\infty(\mathbb{R}^N)} \rightarrow 0$ lorsque $N \rightarrow +\infty$.

Chapitre 2

Recherche de zéro

2.1 L'algorithme de dichotomie

Soit $a < b$ et $f : [a, b] \rightarrow \mathbb{R}$ une application continue telle que $f(a)f(b) < 0$. Alors, d'après le théorème des valeurs intermédiaires, il existe (au moins) un $c \in [a, b]$ tel que $f(c) = 0$. On considère alors les suites $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ et $(m_n)_{n \in \mathbb{N}}$ définies par $a_0 = a$, $b_0 = b$ et pour tout $n \in \mathbb{N}$:

$$m_n = \frac{a_n + b_n}{2}, \quad a_{n+1} = \begin{cases} a_n & \text{si } f(a_n)f(m_n) \leq 0 \\ m_n & \text{sinon} \end{cases}, \quad b_{n+1} = \begin{cases} m_n & \text{si } f(a_n)f(m_n) \leq 0 \\ b_n & \text{sinon} \end{cases}$$

Proposition 8. Les suites $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ et $(m_n)_{n \in \mathbb{N}}$ sont bien définies et convergent toutes vers la même limite $\ell \in]a, b[$ qui vérifie $f(\ell) = 0$. De plus, on pour tout $n \in \mathbb{N}$ on a le contrôle suivant de l'erreur

$$|a_n - \ell| \leq \frac{b-a}{2^n}, \quad |b_n - \ell| \leq \frac{b-a}{2^n}, \quad |m_n - \ell| \leq \frac{b-a}{2^n}.$$

Exercice 18. Le but de cet exercice est de démontrer la Proposition 8.

- (i) Montrer que $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ sont des suites adjacentes et en déduire qu'elles convergent vers une même limite $\ell \in]a, b[$. Justifier que $(m_n)_{n \in \mathbb{N}}$ admet également ℓ comme limite en $+\infty$.
- (ii) Montrer que $f(\ell) = 0$.
- (iii) Démontrer le contrôle de l'erreur.

2.2 Méthodes de point fixe

Soit $(E, \|\cdot\|)$ un $\mathbb{K}$ espace vectoriel normé ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$). On considère une application

$$f : \begin{cases} (E, \|\cdot\|) & \rightarrow (E, \|\cdot\|) \\ x & \mapsto f(x) \end{cases}$$

On observe simplement que trouver les zéros de f est équivalent à trouver les points fixes de $g = f + id$. En effet, pour $x \in E$, on a

$$f(x) = 0_E \iff f(x) + x = x \iff g(x) = x.$$

On se concentre alors sur des méthodes dites de point fixe : on se donne $x_{in} \in E$ et on considère la suite $(x_{n+1})_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ définie pour tout $n \in \mathbb{N}$ par

$$\begin{cases} x_{n+1} & = g(x_n) \\ x_0 & = x_{in}. \end{cases} \quad (2.1)$$

La proposition suivante est une simple conséquence de la continuité de g .

Proposition 9. Si g est continue sur E et $(x_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ définie par (2.1) admet une limite $\ell \in E$ alors $g(\ell) = \ell$.

2.2.1 Méthode de Banach-Picard

Définition 2. Un espace métrique est la donnée d'un ensemble E et d'une application distance

$$d : \begin{cases} E \times E & \rightarrow \mathbb{R}_+ \\ (x, y) & \mapsto d(x, y). \end{cases}$$

qui vérifie

- (i) d est symétrique : pour tout $x, y \in E$ on a $d(x, y) = d(y, x)$,
- (ii) d est séparante : pour tout $x, y \in E$, $d(x, y) = 0$ si et seulement si $x = y$,
- (iii) d vérifie l'inégalité triangulaire : pour tout $x, y, z \in E$, on a $d(x, z) \leq d(x, y) + d(y, z)$.

Remarque 3. Si (E, N) est un $\mathbb{K}$ espace vectoriel normé ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$) alors (E, d) est un espace métrique avec

$$d : \begin{cases} E \times E & \rightarrow \mathbb{R}_+ \\ (x, y) & \mapsto N(x - y). \end{cases}$$

Le cas particulier qui va essentiellement nous intéresser et le cas d'espace métrique (A, d) où A est une partie $A \subset E$ munit de la distance

$$d : \begin{cases} A \times A & \rightarrow \mathbb{R}_+ \\ (x, y) & \mapsto N(x - y). \end{cases}$$

Définition 3. Soit (E, d) un espace métrique. Soit $(x_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$.

- (i) On dit que $(x_n)_{n \in \mathbb{N}}$ converge vers $\ell \in E$ si

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, d(x_n, \ell) < \varepsilon.$$

- (ii) On dit que $(x_n)_{n \in \mathbb{N}}$ est de Cauchy dans E si

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0, \forall p \in \mathbb{N} d(x_{n+p}, x_n) < \varepsilon.$$

Remarque 4. Une suite convergente est de Cauchy.

Définition 4. Soit (E, d) un espace métrique. (E, d) est dit complet si toutes les suites de Cauchy convergent dans E .

Proposition 10 (Théorème du point fixe de Banach-Picard). Soit (E, d) un espace métrique complet et $f : E \rightarrow E$ strictement contractante, c'est à dire pour laquelle il existe $k \in]0, 1[$ telle que pour tout $x, y \in E$ on a

$$d(f(x), f(y)) < kd(x, y).$$

Alors f admet un unique point fixe noté $\ell \in E$ et pour tout $x_0 \in E$, la suite récurrente $(x_n)_{n \in \mathbb{N}}$ définie pour tout $n \in \mathbb{N}$ par $x_{n+1} = f(x_n)$ converge vers $\ell \in E$. De plus, pour tout $n \in \mathbb{N}$, on a l'estimation d'erreur

$$d(x_n, \ell) \leq \frac{k^n}{1 - k} d(x_0, x_1).$$

Exercice 19. Le but de cet exercice est de démontrer la Proposition 10.

- (i) Démontrer que si un point fixe existe, alors il est unique.
- (ii) On définit la suite réelle $(u_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ définie pour tout $n \in \mathbb{N}$ par $u_n = d(x_{n+1}, x_n)$. Démontrer que pour tout $n \in \mathbb{N}$ on a

$$u_n \leq k^n u_0.$$

- (iii) En déduire que pour tout $p, n \in \mathbb{N}$, on a

$$d(x_{n+p}, x_n) \leq \frac{(1 - k^p)k^n}{1 - k} u_0$$

et montrer que $(x_n)_{n \in \mathbb{N}}$ est de Cauchy.

(iv) En déduire qu'il existe un point fixe ℓ pour f et que l'on a

$$d(x_n, \ell) \leq \frac{k^n}{1-k} u_0.$$

Une application classique dans $\mathbb{R}^d$ de la méthode du point fixe de Banach-Picard est la suivante.

Proposition 11. Soit $\Omega \subset \mathbb{R}^d$ un ouvert, $\|\cdot\|$ une norme sur $\mathbb{R}^d$ et $\|\|\cdot\|\|$ la norme subordonnée associée sur $\mathcal{L}(\mathbb{R}^d)$ définie pour tout $\psi \in \mathcal{L}(\mathbb{R}^d)$ par

$$\|\|\psi\|\| = \sup_{\substack{h \in \mathbb{R}^d \\ h \neq 0}} \frac{\|\psi(h)\|}{\|h\|}.$$

Soit $f : \Omega \rightarrow \mathbb{R}^d$ de classe $\mathcal{C}^1(\Omega)$ et $\ell \in \Omega$ un point fixe de f .

1. Si $\|\|Df(\ell)\|\| < 1$, alors il existe $\delta > 0$ tel que pour tout $x_{in} \in B_f(\ell, \delta)$ la suite récurrente $(x_n)_{n \in \mathbb{N}}$ définie pour tout $n \geq 0$ par

$$\begin{cases} x_{n+1} &= f(x_n) \\ x_0 &= x_{in} \end{cases}$$

converge vers ℓ . On dit que ℓ est un point fixe attracteur.

2. Si $d = 1$ et que $|f'(x)| > 1$ alors pour tout $x_{in} \neq \ell$ la suite récurrente $(x_n)_{n \in \mathbb{N}}$ définie pour tout $n \geq 0$ par

$$\begin{cases} x_{n+1} &= f(x_n) \\ x_0 &= x_{in} \end{cases}$$

ne converge jamais vers ℓ . On dit que ℓ est un point répulsif.

Exercice 20. Le but de cet exercice est de démontrer la Proposition 11.

(i) Démontrer que si $\|\|Df(\ell)\|\| < 1$ alors il existe $\delta > 0$ tel que pour tout $x \in B_f(\ell, \delta)$ on ait

$$\sup_{x \in B_f(\ell, \delta)} \|\|Df(x)\|\| < 1.$$

(ii) Montrer que l'application f restreinte à $B_f(\ell, \delta)$ est contractante.

(iii) En déduire le point 1.

(iv) On suppose que $d = 1$. Admettons par l'absurde que $(x_n)_{n \in \mathbb{N}}$ converge vers ℓ .

(a) Montrer qu'il existe $M > 1$ et $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$ on ait

$$|x_{n+1} - \ell| \geq M|x_n - \ell|.$$

(b) En déduire une absurdité.

2.2.2 Méthode de Newton dans $\mathbb{R}^d$

La méthode de Newton est utilisée pour trouver des zéros d'une fonction $f : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^d$ où Ω est un ouvert de $\mathbb{R}^d$. Supposons que f est différentiable sur Ω , de différentielle inversible. Alors on a

$$f(x) = 0 \iff 0 \iff (D_x f)^{-1}(f(x)) = 0 \iff x + (D_x f)^{-1}(f(x)) = x \iff g(x) = x$$

où pour $x \in \Omega$, on a posé $g(x) = x + (D_x f)^{-1}(f(x))$. La méthode de Newton consiste à appliquer une méthode itérative à la fonction g .

Théorème 2. Soit $\Omega \subset \mathbb{R}^d$ un ouvert. Soit $f : \Omega \rightarrow \mathbb{R}^d$ de classe $\mathcal{C}^2(\Omega)$. Soit $\ell \in \Omega$ un zéro de f tel que $D_\ell f \in \mathcal{L}(\mathbb{R}^d)$ soit inversible. Alors, il existe $\delta > 0$ tel que pour tout $x_{in} \in B(\ell, \delta)$ la suite récurrente $(x_n)_{n \in \mathbb{N}} \in (\mathbb{R}^d)^\mathbb{N}$ définie pour tout $n \in \mathbb{N}$ par

$$\begin{cases} x_{n+1} &= g(x_n) \\ x_0 &= x_{in} \end{cases}$$

est bien définie et converge vers ℓ . Par ailleurs, on a la convergence quadratique pour la méthode qui se traduit par l'estimation d'erreur suivante : il existe $M > 0$ tel que pour tout $n \in \mathbb{N}$

$$\|x_{n+1} - \ell\|_{\mathbb{R}^d} \leq M \|x_n - \ell\|_{\mathbb{R}^d}^2.$$

Remarque 5. Dans l'Exercice 21 ci-dessous, il suffit de démontrer le résultat pour la question (iii) pour avoir la convergence. L'intérêt de la question (ii) réside dans l'illustration d'une application de la Proposition 11.

Exercice 21. Le but de cet exercice est de démontrer le Théorème 2.

- (i) Montrer qu'il existe $\eta > 0$ tel que pour tout $x \in B(\ell, \eta)$, la fonction g est bien définie et de classe $\mathcal{C}^1(B(\ell, \eta))$.
- (ii) Montrer que $\|D_\ell g\| < 1$ et en déduire que l'existence de $\delta > 0$ tel que la suite des itérés est bien définie si $x_{in} \in B(\ell, \delta)$ et converge vers ℓ .
- (iii) Montrer que, quitte à diminuer $\delta > 0$, l'erreur $e_n := x_n - \ell$ vérifie l'estimée annoncée.

2.2.3 Optimisation et algorithme du gradient à pas constant

Soit $\Omega \subset \mathbb{R}^d$ un ouvert et

$$f : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$$

une application de classe $\mathcal{C}^1(\Omega)$. On sait que si f atteint son minimum en un point $\ell \in \Omega$ alors

$$(\nabla f)(\ell) = 0.$$

Regardons alors l'application gradient

$$G_f : \begin{cases} \mathbb{R}^d & \rightarrow \mathbb{R}^d \\ x & \mapsto (\nabla f)(x) \end{cases}$$

Pour tout $\alpha \in \mathbb{R}_+^*$, on a

$$G_f(x) = 0 \iff (\nabla f)(x) = 0 \iff -\alpha(\nabla f)(x) = 0 \iff x - \alpha(\nabla f)(x) = x \iff g(x) = x$$

où pour $x \in \Omega$ on a posé $g(x) = x - \alpha(\nabla f)(x)$.

On peut alors fabriquer une méthode itérative pour obtenir un minimiseur ℓ de f en posant $x_{in} \in \Omega$ et en considérant la suite définie pour tout $n \in \mathbb{N}$ par

$$\begin{cases} x_{n+1} = g(x_n) &= x_n - \alpha(\nabla f)(x_n) \\ x_0 &= x_{in}. \end{cases}$$

Proposition 12. Soit $M \in \mathcal{M}_{q,d}(\mathbb{R})$ ($M \neq 0$) et $f : x \in \mathbb{R}^d \mapsto \|Mx - b\|_{\ell^2(\mathbb{R}^q)}^2$ la fonctionnelle des moindres carrés et ℓ un minimiseur de cette fonctionnelle. Alors, si $x_{in} \in \ell + \ker(M)^\perp$, l'algorithme du gradient à pas constant converge vers ℓ pour tout $x_{in} \in \mathbb{R}^d$ et tout $\alpha \in]0, \frac{1}{2\sigma_M^2}[$ où $\sigma_M > 0$ est la plus grande valeur singulière de M . De plus, on a

$$\|x_n - \ell\|_{\ell^2(\mathbb{R}^d)} \leq (1 - \alpha\sigma_m^2)^n \|x_0 - \ell\|_{\ell^2(\mathbb{R}^d)}$$

où $\sigma_m > 0$ est la plus petite valeur singulière strictement positive de M .

Exercice 22. Le but de cet exercice est de démontrer la Proposition 12. On rappelle que si $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est de classe $\mathcal{C}^1(\mathbb{R}^d)$ alors pour tout $x, h \in \mathbb{R}^d$ on a

$$(D_x f)(h) = \langle (\nabla f)(x), h \rangle_{\mathbb{R}^d}.$$

(i) Montrer que pour tout $x \in \mathbb{R}^d$ on a $(\nabla f)(x) = 2(M^\top Mx - M^\top b)$.

(ii) En déduire que

$$x_{n+1} - \ell = (Id - 2\alpha M^\top M)(x_n - \ell).$$

(iii) Justifier que l'on a

$$\|x_{n+1} - \ell\|_{\ell^2(\mathbb{R}^d)} \leq (1 - 2\alpha\sigma_m^2)\|x_n - \ell\|_{\ell^2(\mathbb{R}^d)}$$

et conclure.

Proposition 13. Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ de classe $\mathcal{C}^2(\mathbb{R}^d)$. Pour $x \in \mathbb{R}^d$, on définit la matrice Hessienne de f en x en définie par $(H_x f)_{j,k} := (\frac{\partial^2 f}{\partial x_j \partial x_k})(x)$ et on suppose qu'il existe $m, M > 0$ tels que pour tout y

$$m\|y\|_{\ell^2(\mathbb{R}^d)}^2 \leq \langle (H_x f)y, y \rangle_{\mathbb{R}^d} \leq M\|y\|_{\ell^2(\mathbb{R}^d)}^2.$$

Alors f admet un unique minimiseur en $\ell \in \mathbb{R}^d$ et la méthode du gradient à pas constant converge lorsque $\alpha \in]0, \frac{2m}{M^2}[$.

Remarque 6. L'hypothèse impliquant la constante $m > 0$ peut s'interpréter comme une hypothèse dite de m -convexité.

Exercice 23. Le but de cet exercice est de démontrer la Proposition 13.

(i) En utilisant une formule de Taylor, montrer qu'il existe un minimiseur pour f noté $\ell \in \mathbb{R}^d$ et qu'il est nécessairement unique.

(ii) Montrer que si $e_n = x_n - \ell$ est l'erreur de la méthode de l'algorithme du gradient à pas constant pour $n \in \mathbb{N}$, on a

$$\|e_{n+1}\|_{\ell^2(\mathbb{R}^d)}^2 \leq \|e_n\|_{\ell^2(\mathbb{R}^d)}^2 (1 - 2\alpha m + M^2 \alpha^2)$$

(iii) Conclure.

Chapitre 3

Intégration Numérique

3.1 Méthode des rectangles

Nous allons discuter de méthodes d'intégration numérique : comment faire pour calculer des intégrales numériquement.

Pour $a < b$ deux réels, on considère une fonction $f : [a, b] \rightarrow \mathbb{R}$ continue sur $[a, b]$ et on se demande comment approcher

$$\int_a^b f(t)dt.$$

Une approximation grossière est la méthode dite "des rectangles" : on choisit un point $\tau \in [0, 1]$ et on considère

$$I_\tau(f) := (b - a)f((1 - \tau)a + \tau b).$$

On approche l'intégrale, c'est-à-dire l'aire du domaine $\{(x, y) \in [a, b] \times \mathbb{R} : 0 \leq y \leq f(x)\}$, par l'aire du rectangle de côtés $[a, b]$ et $[0, f(\tau)]$ (voir Figure 3.1).

Définition 5. Lorsque $\tau = 0$ on parle d'approximation de f par un rectangle à gauche, lorsque $\tau = 1$ on parle d'approximation de f par un rectangle à droite et lorsque $\tau = \frac{1}{2}$ on parle d'approximation de f par le point milieu.

Proposition 14. Soit $f \in \mathcal{C}^1([a, b])$, il existe $C > 0$ telle que pour tout $\tau \in [0, 1]$, on a

$$\left| \left(\int_a^b f(t)dt \right) - I_\tau(f) \right| \leq C(b - a)^2.$$

Exercice 24. Démontrer la proposition 14 (on pensera à utiliser une formule de Taylor adaptée).

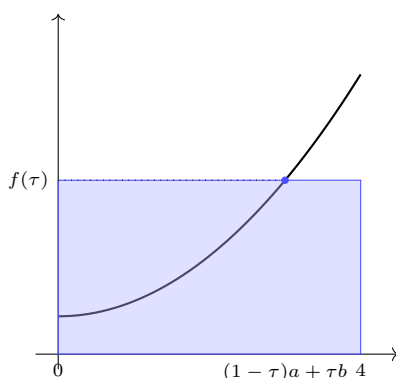


FIGURE 3.1 – Illustration de l'aire donnée par $I_\tau(f)$ dans le cas où $f : [0, 4] \rightarrow \mathbb{R}$ est donnée pour $x \in [0, 4]$ par $f(x) = \frac{1}{5}x^2 + \frac{1}{2}$ et $\tau = \frac{3}{4}$.

La proposition 14 nous permet de remarquer que plus $(b - a)$ est petit, plus l'erreur faite par cette méthode est bonne. Afin de tirer avantage de cette expression, on va considérer une méthode dite composite : on va découper l'intervalle $[a, b]$ en des intervalles plus petits et sur chacun d'entre eux on appliquera la méthode des rectangles. Pour ce faire, on considère une subdivision $a = t_0 < t_1 < t_2 < \dots < t_N < t_{N+1} = b$ de l'intervalle $[a, b]$ et la relation de Chasles donne

$$\int_a^b f(t)dt = \sum_{j=0}^N \int_{t_j}^{t_{j+1}} f(t)dt.$$

Il faut ensuite tâcher d'approcher pour tout $j \in \{0, \dots, N\}$ les intégrales de la forme

$$\int_{t_j}^{t_{j+1}} f(t)dt$$

ce que nous allons faire en considérant pour $\tau \in [0, 1]$ des quantités de la forme $I_\tau(f_j)$ où $f_j = f|_{[t_j, t_{j+1}]}$. On va donc considérer les quantités

$$S_\tau(f) := \sum_{j=0}^N I_\tau(f_j) = \sum_{j=0}^N (t_{j+1} - t_j) f((1 - \tau)t_j + \tau t_{j+1}) \quad (3.1)$$

et on parle de formules de quadrature.

Définition 6. La méthode des rectangles à gauche consiste à choisir $\tau = 0$, la méthode des rectangles à droite consiste à choisir $\tau = 1$ et celle du point milieu consiste à choisir $\tau = \frac{1}{2}$. En particulier, on a

$$S_0(f) = \sum_{j=0}^N (t_{j+1} - t_j) f(t_j), \quad S_1(f) = \sum_{j=0}^N (t_{j+1} - t_j) f(t_{j+1}), \quad S_{\frac{1}{2}}(f) = \sum_{j=0}^N (t_{j+1} - t_j) f\left(\frac{t_j + t_{j+1}}{2}\right).$$

On peut jouer sur le choix de τ pour améliorer la convergence de la méthode. C'est le but de la Proposition suivante.

Proposition 15. Pour $j \in \{0, \dots, N+1\}$ on choisit la subdivision uniforme $t_j = a + j \frac{b-a}{N+1}$. Pour tout $\tau \in \{0, 1\}$, c'est-à-dire pour la méthode des rectangles à gauche ou des rectangles à droite, si $f \in \mathcal{C}^1([a, b])$, il existe $M > 0$ tel que

$$\left| \left(\int_a^b f(t)dt \right) - S_\tau(f) \right| \leq C \frac{(b-a)^2}{N+1}.$$

Pour la méthode du point milieu, si $f \in \mathcal{C}^2([a, b])$, il existe $M > 0$ tel que

$$\left| \left(\int_a^b f(t)dt \right) - S_\tau(f) \right| \leq C \frac{(b-a)^3}{(N+1)^2}.$$

Exercice 25.

- (i) Démontrer la première assertion de la Proposition 15.
- (ii) Démontrer le deuxième point de la Proposition 15.

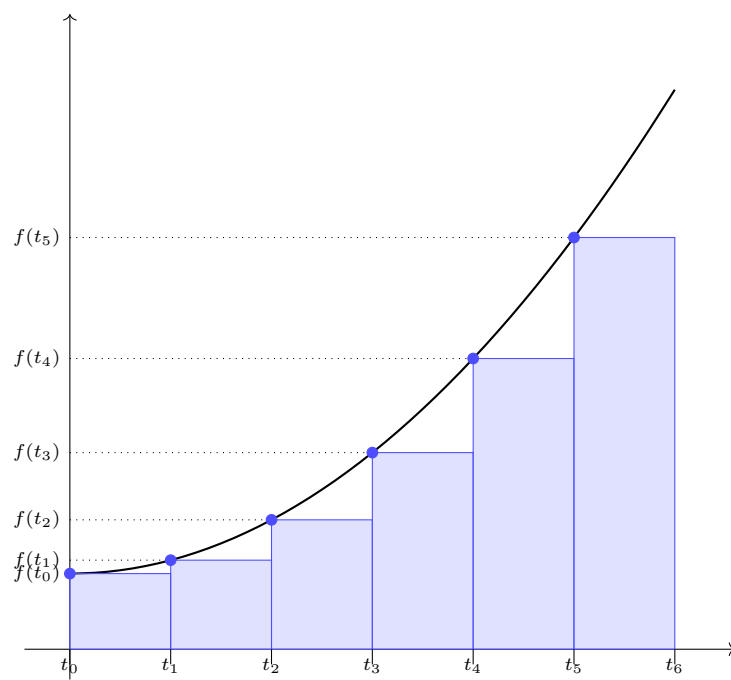


FIGURE 3.2 – Illustration de la méthode des rectangles à gauche pour $f : [0, 4] \rightarrow \mathbb{R}$ donnée pour $x \in [0, 4]$ par $f(x) = \frac{1}{5}x^2 + \frac{1}{2}$.

Chapitre 4

Équations différentielles : la méthode d'Euler explicite

4.1 Cadre d'étude

Soit $T > 0$ et $\Omega \subset \mathbb{R}^d$ un ouvert, on considère

$$f : \begin{cases} [0, T] \times \mathbb{R}^d & \rightarrow \mathbb{R}^d \\ (t, x) & \mapsto f(t, x). \end{cases}$$

On suppose que f est continue sur $[0, T] \times \mathbb{R}^d$ et globalement Lipschitzienne par rapport à sa deuxième variable c'est à dire qu'il existe $L > 0$ telle que pour tout $t \in [0, T]$ et tout $x, y \in \mathbb{R}^d$ on a

$$\|f(t, x) - f(t, y)\|_{\mathbb{R}^d} \leq L\|x - y\|_{\mathbb{R}^d}.$$

Soit $y_0 \in \mathbb{R}^d$, on s'intéresse à des équations différentielles de la forme

$$\begin{cases} y'(t) = f(t, y(t)) & \text{pour tout } t > 0 \\ y(0) = y_0 \end{cases} \quad (4.1)$$

D'après le théorème de Cauchy Lipschitz global il existe une unique solution φ à cette équation différentielle définie sur $[0, T]$. Notre objectif est de fabriquer une méthode numérique qui permette d'approcher la solution φ .

4.2 La méthode d'Euler explicite

Soit $N \in \mathbb{N}$, on construit pour $j \in \{0, \dots, N+1\}$ le point $t_j = jh \in [0, T]$ où $h = \frac{T}{N+1}$. On remarque que si $([0, T], \varphi)$ est solution du problème de Cauchy (4.1) alors on a

$$\varphi(t_{j+1}) = \varphi(t_j) + \int_{t_j}^{t_{j+1}} f(t, \varphi(t)) dt.$$

En approchant cette intégrale avec l'aire du rectangle à gauche, on obtient

$$\varphi(t_{j+1}) \simeq \varphi(t_j) + hf(t_j, \varphi(t_j)).$$

Cela nous incite à construire la suite finie $(\varphi_j)_{j \in \{0, \dots, N\}}$ définie pour $j \in \{0, \dots, N\}$ par

$$\begin{cases} \varphi_{j+1} &= \varphi_j + hf(t_j, \varphi_j) \\ \varphi_0 &= y_0. \end{cases} \quad (4.2)$$

On espère que la suite $(\varphi_j)_{j \in \{0, \dots, N\}}$ ainsi fabriquée permette d'obtenir une approximation de la suite $(\varphi(t_j))_{j \in \{0, \dots, N\}}$. Cette stratégie d'approximation se nomme méthode d'Euler explicite.

Remarque 7. On pourrait approcher l'intégrale par d'autres méthodes d'intégration numérique. Par exemple, la méthode des rectangles à droite donnerait la suite finie $(\varphi_j)_{j \in \{0, \dots, N\}}$ définie pour $j \in \{0, \dots, N\}$ par

$$\begin{cases} \varphi_{j+1} &= \varphi_j + hf(t_{j+1}, \varphi_{j+1}) \\ \varphi_0 &= y_0. \end{cases}$$

On parlerait alors de méthode d'Euler implicite : φ_{j+1} vérifie alors une équation implicite qu'il faut résoudre à chaque étape (avec un algorithme de recherche de zéro par exemple).

4.3 Consistance

L'erreur locale de consistance est l'erreur que l'on commet en appliquant le schéma numérique sur une itération.

Définition 7. L'erreur locale de consistance relative à une solution exacte $([0, T], \varphi)$ de (4.1) est définie pour $j \in \{0, \dots, N\}$ par

$$e_j(\varphi) := \varphi(t_{j+1}) - (\varphi(t_j) + hf(t_j, \varphi(t_j))).$$

L'erreur globale de consistance consiste en la somme de toutes les erreurs locales sur N -étapes du schéma.

Définition 8. L'erreur globale de consistance relative à une solution exacte $([0, T], \varphi)$ de (4.1) pour le schéma d'Euler explicite est donnée par

$$\mathcal{E}_N(\varphi) = \sum_{j=0}^N \|e_j(\varphi)\|_{\mathbb{R}^d}.$$

Le schéma est dit consistant d'ordre $p \in \mathbb{N}^*$ si pour toute solution exacte $([0, T], \varphi)$ de (4.1) il existe $C > 0$ telle que

$$\mathcal{E}_N(\varphi) \leq Ch^p.$$

Proposition 16. Si $([0, T], \varphi)$ est solution de (4.1) de classe $\mathcal{C}^2([0, T])$ alors le schéma d'Euler explicite est consistant d'ordre d'ordre 1.

Exercice 26. Démontrer la Proposition 16 pour une solution $([0, T], \varphi)$ de (4.1) de classe $\mathcal{C}^2([0, T])$.

4.4 Stabilité

On s'intéresse maintenant à la stabilité du schéma par rapport à des erreurs d'approximation numériques qui pourraient intervenir lors du calcul des $(\varphi_j)_{j \in \{0, \dots, N+1\}}$. Cette notion est là pour garantir que l'erreur d'arrondie numérique faite à l'itération j ne s'amplifie pas à cause des itérations successives.

Proposition 17. La méthode d'Euler explicite est stable, c'est-à-dire qu'il existe une constante $S \geq 0$ telle que pour toute suite $(\varphi_j)_{j \in \{0, \dots, N\}}$ et $(\tilde{\varphi}_j)_{j \in \{0, \dots, N\}}$ définies pour tout $j \in \{0, \dots, N\}$ par

$$\varphi_{j+1} = \varphi_j + hf(t_j, \varphi_j), \quad \tilde{\varphi}_{j+1} = \tilde{\varphi}_j + hf(t_j, \tilde{\varphi}_j) + \varepsilon_j$$

on ait

$$\max_{j \in \{0, \dots, N+1\}} \|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} \leq S \left(\|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \sum_{j=0}^N \|\varepsilon_j\|_{\mathbb{R}^d} \right)$$

Exercice 27. Le but de cet exercice est de démontrer la Proposition 17.

(i) Montrer que pour tout $j \in \{0, \dots, N\}$ on a

$$\|\varphi_{j+1} - \tilde{\varphi}_{j+1}\|_{\mathbb{R}^d} \leq (1 + hL) \|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} + \|\varepsilon_j\|_{\mathbb{R}^d}.$$

(ii) Montrer alors par récurrence sur $j \in \{1, \dots, N+1\}$ que l'on a

$$\|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} \leq e^{jLh} \|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \sum_{p=0}^{j-1} e^{Lh(j-1-p)} \|\varepsilon_p\|_{\mathbb{R}^d}.$$

(iii) Conclure.

4.5 Convergence

On se pose maintenant la question de la convergence du schéma d'Euler explicite, c'est-à-dire que l'on cherche à démontrer que si $([0, T], \varphi)$ est solution de (4.1), et que $(\varphi_j)_{j \in \{0, \dots, N+1\}}$ est la suite finie définie en (4.2), on a

$$\lim_{N \rightarrow +\infty} \left(\max_{j \in \{1, \dots, N+1\}} \|\varphi(t_j) - \varphi_j\|_{\mathbb{R}^d} \right) = 0.$$

Si possible, on aimerait avoir une estimée de la vitesse de convergence par rapport à N .

Proposition 18. Soit $([0, T], \varphi)$ solution de (4.1) de classe $\mathcal{C}^2([0, T])$ et $(\varphi_j)_{j \in \{0, \dots, N+1\}}$ est la suite finie définie en (4.2), il existe $C > 0$ tel que pour tout $N \in \mathbb{N}^*$ on a

$$\left(\max_{j \in \{1, \dots, N+1\}} \|\varphi(t_j) - \varphi_j\|_{\mathbb{R}^d} \right) \leq \frac{C}{N+1}$$

Exercice 28. Démontrer la Proposition 18. Pour cela on utilisera la stabilité et la consistance du schéma d'Euler explicite.

Annexe A

Corrections des exercices

Solution (de l'Exercice 1).

(i) On remarque que la matrice $M^*M \in \mathcal{M}_{d,d}(\mathbb{C})$ est hermitienne car on a

$$(M^*M)^* = M^*M.$$

D'après le théorème spectral, elle est diagonalisable en base orthonormée donc il existe $(e_1, \dots, e_d)$ une base de $\mathbb{C}^d$ et $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ des réels tels que pour tout $x \in \mathbb{C}^d$ on a

$$M^*Mx = \sum_{j=1}^d \lambda_j \langle x, e_j \rangle_{\mathbb{C}^d} e_j.$$

Il reste à démontrer que pour tout $j \in \{1, \dots, d\}$, on a $\lambda_j \geq 0$. Pour cela, on remarque que pour tout $j \in \mathbb{C}^d$, on a

$$\lambda_j = \langle M^*Me_j, e_j \rangle_{\mathbb{C}^d} = \langle Me_j, Me_j \rangle_{\mathbb{C}^q} = \|Me_j\|_{\ell^2(\mathbb{C}^q)}^2 \geq 0.$$

(ii) Soit $j, k \in \{1, \dots, r\}$, on a

$$\langle e_j^*, e_k^* \rangle_{\mathbb{C}^q} = \frac{1}{\sqrt{\lambda_j \lambda_k}} \langle Me_j, Me_k \rangle_{\mathbb{C}^q} = \langle M^*Me_j, e_k \rangle_{\mathbb{C}^d} = \frac{\lambda_j}{\sqrt{\lambda_j \lambda_k}} \delta_{j,k} = \delta_{j,k}.$$

On en déduit que $(e_1^*, \dots, e_r^*)$ est une famille orthonormée de $\mathbb{C}^q$.

(iii) Soit $j \in \{r+1, \dots, d\}$, on a

$$\|Me_j\|_{\ell^2(\mathbb{C}^q)}^2 = \langle Me_j, Me_j \rangle_{\mathbb{C}^q} = \langle M^*Me_j, e_j \rangle_{\mathbb{C}^d} = \lambda_j = 0$$

et donc $Me_j = 0$.

(iv) Soit $x \in \mathbb{C}^d$, comme $(e_1, \dots, e_d)$ est une famille orthonormée, on a

$$x = \sum_{j=1}^d \langle x, e_j \rangle_{\mathbb{C}^d} e_j$$

et donc

$$\begin{aligned} Mx &= \sum_{j=1}^d \langle x, e_j \rangle_{\mathbb{C}^d} Me_j = \left(\sum_{j=1}^r \langle x, e_j \rangle_{\mathbb{C}^d} Me_j \right) + \left(\sum_{j=r+1}^d \langle x, e_j \rangle_{\mathbb{C}^d} Me_j \right) \\ &= \sum_{j=1}^r \sqrt{\lambda_j} \langle x, e_j \rangle_{\mathbb{C}^d} e_j^*. \end{aligned}$$

Comme $x = \sum_{p=1}^d \langle x, \varepsilon_p \rangle_{\mathbb{C}^d} \varepsilon_p$, on obtient

$$\begin{aligned} Mx &= \sum_{j=1}^r \sqrt{\lambda_j} \langle x, e_j \rangle_{\mathbb{C}^d} e_j^* = \sum_{j=1}^r \sqrt{\lambda_j} \left\langle \sum_{p=1}^d \langle x, \varepsilon_p \rangle_{\mathbb{C}^d} \varepsilon_p, e_j \right\rangle_{\mathbb{C}^d} e_j^* \\ &= \sum_{j=1}^r \sum_{p=1}^d \sqrt{\lambda_j} \langle x, \varepsilon_p \rangle_{\mathbb{C}^d} \langle \varepsilon_p, e_j \rangle_{\mathbb{C}^d} e_j^*. \end{aligned}$$

Soit $j \in \{1, \dots, d\}$ et $k \in \{1, \dots, q\}$. On remarque que

$$M_{j,k} = \langle M\varepsilon_k, \varepsilon_j^* \rangle_{\mathbb{C}^q}.$$

En particulier, on a

$$M\varepsilon_k = \sum_{s=1}^r \sum_{p=1}^d \sqrt{\lambda_s} \langle \varepsilon_k, \varepsilon_p \rangle_{\mathbb{C}^d} \langle \varepsilon_p, e_s \rangle_{\mathbb{C}^d} e_s^* = \sum_{s=1}^r \sqrt{\lambda_s} \langle \varepsilon_k, e_s \rangle_{\mathbb{C}^d} e_s^*$$

et donc

$$M_{j,k} = \langle M\varepsilon_k, \varepsilon_j^* \rangle_{\mathbb{C}^q} = \sum_{s=1}^r \sqrt{\lambda_s} \langle \varepsilon_k, e_s \rangle_{\mathbb{C}^d} \langle e_s^*, \varepsilon_j^* \rangle_{\mathbb{C}^q}.$$

(v) On remarque que par définition pour tout $j, k \in \{1, \dots, q\}$ on a $U_{j,k} = \langle e_k^*, \varepsilon_j^* \rangle_{\mathbb{C}^q}$ et pour tout $j, k \in \{1, \dots, d\}$ on a $V_{j,k} = \langle e_k, \varepsilon_j \rangle_{\mathbb{C}^d}$.

Soit $j \in \{1, \dots, q\}$ et $k \in \{1, \dots, d\}$, on a

$$\begin{aligned} (U\Sigma V^*)_{j,k} &= \sum_{s=1}^q U_{j,s} (\Sigma V^*)_{s,k} = \sum_{s=1}^q \sum_{p=1}^d U_{j,s} \Sigma_{s,p} V_{p,k}^* \\ &= \sum_{p=1}^q \sum_{s=1}^d \langle e_s^*, \varepsilon_j^* \rangle_{\mathbb{C}^q} \Sigma_{s,p} \langle e_k, e_p \rangle_{\mathbb{C}^d} \\ &= \sum_{s=1}^r \sqrt{\lambda_s} \langle e_s^*, \varepsilon_j^* \rangle_{\mathbb{C}^q} \langle \varepsilon_k, e_s \rangle_{\mathbb{C}^d} \\ &= M_{j,k}. \end{aligned}$$

On en déduit l'existence de la décomposition en valeurs singulières.

Solution (de l'Exercice 2). On remarque que

$$M^\top M = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 9 & 0 \\ 2 & 0 & 4 \end{pmatrix}$$

On remarque que

$$\det(M^\top M - \lambda Id) = \begin{vmatrix} 1 & 0 & 2 \\ 0 & 9 & 0 \\ 2 & 0 & 4 \end{vmatrix} ((9 - \lambda)(1 - \lambda)(4 - \lambda) - 4) = \lambda(9 - \lambda)(\lambda - 5).$$

On pose

$$e_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, e_3 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}$$

et on remarque que

$$M^\top M e_1 = 9e_1, \quad M^\top M e_2 = 5e_2, \quad M^\top M e_3 = 0$$

et (e_1, e_2, e_3) est une base orthonormée de $\mathbb{R}^3$. On pose

$$e_1^* := \frac{1}{3} M e_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad e_2^* = \frac{1}{\sqrt{5}} M e_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

On pose

$$U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & \sqrt{5} & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ 1 & 0 & 0 \\ 0 & \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix}$$

et on remarque que

$$U\Sigma V^* = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & \sqrt{5} & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ \frac{1}{\sqrt{5}} & 0 \\ \frac{2}{\sqrt{5}} & 0 \\ 0 & -\frac{1}{\sqrt{5}} \end{pmatrix} = M.$$

Soit $X \in \mathbb{R}^3$ tel que

$$MX = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

On a alors, en posant $Y = V^*X \in \mathbb{R}^3$:

$$\Sigma Y = U^* \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

On en déduit que si $Y = (y_1, y_2, y_3)^\top$ on a

$$\begin{cases} 3y_1 &= -1 \\ \sqrt{5}y_2 &= 1 \end{cases}$$

et donc que les solutions $Y \in \mathbb{R}^3$ de $\Sigma Y = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ sont les vecteurs de la forme

$$\begin{pmatrix} -\frac{1}{3} \\ \frac{1}{\sqrt{5}} \\ \alpha \end{pmatrix}, \quad \alpha \in \mathbb{R}.$$

On en déduit que les solutions $X \in \mathbb{R}^3$ de $MX = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ sont les vecteurs de la forme

$$X = VY = \begin{pmatrix} 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ 1 & 0 & 0 \\ 0 & \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} -\frac{1}{3} \\ \frac{1}{\sqrt{5}} \\ \alpha \end{pmatrix} = \begin{pmatrix} \frac{1}{5} + \frac{2}{\sqrt{5}}\alpha \\ -\frac{1}{3} \\ \frac{2}{5} - \frac{\alpha}{\sqrt{5}} \end{pmatrix}, \quad \alpha \in \mathbb{R}.$$

Solution (de l'Exercice 3).

- (i) Supposons par l'absurde que $A = (a_{j,k})_{j,k \in \{1, \dots, d\}}$ soit à diagonale strictement dominante mais ne soit pas inversible. Cela veut dire qu'il existe $x = (x_1, \dots, x_d)^\top \in \mathbb{C}^d \setminus \{0\}$ tel que $Ax = 0$. En particulier, on a pour tout $j \in \{1, \dots, d\}$

$$0 = (Ax)_j = \sum_{k=1}^d a_{j,k}x_k$$

et donc

$$a_{j,j}x_j = - \sum_{\substack{k=1 \\ k \neq j}}^d a_{j,k}x_k.$$

Comme $x \neq 0_{\mathbb{C}^d}$, il existe $j_0 \in \{1, \dots, d\}$ tel que $x_{j_0} \neq 0$ et donc

$$a_{j_0,j_0} = - \frac{1}{x_{j_0}} \sum_{\substack{k=1 \\ k \neq j_0}}^d a_{j_0,k}x_k.$$

À l'aide de l'inégalité triangulaire, on en déduit que

$$|a_{j_0,j_0}| \leq \frac{1}{|x_{j_0}|} \sum_{\substack{k=1 \\ k \neq j_0}} |x_k| |a_{j_0,k}| \leq \frac{\|x\|_{\ell^\infty(\mathbb{C}^d)}}{|x_{j_0}|} \sum_{\substack{k=1 \\ k \neq j_0}} |a_{j_0,k}| \leq \sum_{\substack{k=1 \\ k \neq j_0}} |a_{j_0,k}|.$$

Cela contredit l'hypothèse de diagonale strictement dominante et donc A est inversible.

- (ii) Soit $\lambda \in \mathbb{C}$ une valeur propre de M . Cela signifie que la matrice $M - \lambda Id$ n'est pas inversible. En particulier, elle ne peut pas être à diagonale strictement dominante et donc il existe $j_0 \in \{1, \dots, d\}$ tel que

$$|m_{j_0,j_0} - \lambda| \leq \sum_{\substack{k=1 \\ k \neq j_0}}^d |m_{j_0,k}|.$$

On en déduit que nécessairement $\lambda \in D_{j_0} \subset \bigcup_{j=1}^d D_j$.

Solution (de l'Exercice 4).

(i) Soit $\lambda \in Sp(A)$, on sait alors que

$$|\lambda - 2| \leq 2 \text{ ou } |\lambda - 3| \leq 1 \text{ ou } |\lambda - 4| \leq 1.$$

On en déduit que

$$0 \leq \lambda \leq 4 \text{ ou } 2 \leq \lambda \leq 4 \text{ ou } 0 \leq \lambda \leq 5.$$

et donc

$$\lambda \in [0, 4] \cup [2, 4] \cup [0, 5] = [0, 5].$$

(ii) On remarque en développant par rapport à la première colonne que

$$\begin{aligned} \det(A - \lambda Id) &= \begin{vmatrix} 2 - \lambda & 1 & 1 \\ 0 & 3 - \lambda & 1 \\ 1 & 0 & 4 - \lambda \end{vmatrix} = (2 - \lambda)(3 - \lambda)(4 - \lambda) + 1 - 3 + \lambda \\ &= -\lambda^3 + (4 + 3 + 2)\lambda^2 + (-6 + 8 + 12 + 1)\lambda + 24 - 2 \\ &= -\lambda^3 + 9\lambda^2 - 25\lambda + 22 =: P(\lambda). \end{aligned}$$

On remarque que $P(2) = -8 + 36 - 50 + 22 = 0$ et donc 2 est racine de P . Ainsi, P se factorise comme

$$P(\lambda) = (2 - \lambda)(\lambda^2 - 7\lambda + 11)$$

et donc en calculant le discriminant du polynôme de degré 2 apparaissant dans l'expression de P , on a

$$\Delta = 49 - 44 = 5 > 0$$

et les autres racines de P sont

$$\lambda_{\pm} = \frac{1}{2} (7 \pm \sqrt{5}).$$

Le polynôme caractéristique de A est de degré trois et a trois racines distinctes : A est diagonalisable et $Sp(A) = \{2, \frac{1}{2} (7 + \sqrt{5}), \frac{1}{2} (7 - \sqrt{5})\}$. On remarque que l'on a

$$0 \leq \frac{1}{2} (7 \pm \sqrt{5}) \leq 5 \iff -7 \leq \pm \sqrt{5} \leq 3$$

Or on a bien $5 \leq 9$ et $5 \leq 49$ donc $\sqrt{5} \leq 3$ et $-\sqrt{5} \geq -7$. On en déduit que

$$Sp(A) \subset [0, 5].$$

Solution (de l'Exercice 5).

(i) Soit $x \in \mathbb{C}^d \setminus \{0\}$. On a

$$\begin{aligned} \|Mx\|_{\mathbb{C}^q} &= \sum_{k=1}^q |(Mx)_k| = \sum_{k=1}^q \left| \sum_{j=1}^d m_{k,j} x_j \right| \leq \sum_{k=1}^q \sum_{j=1}^d |m_{k,j}| |x_j| \\ &= \sum_{j=1}^d |x_j| \left(\sum_{k=1}^q |m_{k,j}| \right) \\ &\leq \sum_{j=1}^d \left(|x_j| \max_{j \in \{1, \dots, d\}} \sum_{k=1}^q |m_{k,j}| \right) \\ &= \left(\max_{j \in \{1, \dots, d\}} \sum_{k=1}^q |m_{k,j}| \right) \|x\|_{\mathbb{C}^d}. \end{aligned}$$

On en déduit que

$$\frac{\|Mx\|_{\mathbb{C}^q}}{\|x\|_{\mathbb{C}^d}} \leq \max_{j \in \{1, \dots, d\}} \sum_{k=1}^q |m_{k,j}|.$$

Par passage au sup sur $x \in \mathbb{C}^d \setminus \{0\}$ on a

$$\|M\| \leq \max_{j \in \{1, \dots, d\}} \sum_{k=1}^q |m_{k,j}|.$$

Soit $j_0 \in \{1, \dots, d\}$ tel que

$$\sum_{k=1}^d |m_{k,j_0}| = \max_{j \in \{1, \dots, d\}} \sum_{k=1}^q |m_{k,j}|.$$

On considère le vecteur $x = (x_1, \dots, x_d)^\top \in \mathbb{C}^d$ définit pour tout $j \in \{1, \dots, d\}$ par $x_j = \delta_{j,j_0}$. On a

$$\|x\|_{\mathbb{C}^d} = 1, \quad \|Mx\|_{\mathbb{C}^q} = \sum_{k=1}^q |(Mx)_k| = \sum_{k=1}^q \left| \sum_{j=1}^d m_{k,j} x_j \right| = \sum_{k=1}^q |m_{k,j_0}| = \max_{j \in \{1, \dots, d\}} \sum_{k=1}^q |m_{k,j}|.$$

On en déduit que

$$\|M\| \geq \frac{\|Mx\|_{\mathbb{C}^q}}{\|x\|_{\mathbb{C}^d}} = \max_{j \in \{1, \dots, d\}} \sum_{k=1}^q |m_{k,j}|.$$

On en déduit que

$$\|M\| = \max_{j \in \{1, \dots, d\}} \sum_{k=1}^q |m_{k,j}|.$$

- (ii) Comme $M^*M \in \mathcal{M}_{d,d}(\mathbb{C})$ est une matrice hermitienne, d'après le théorème spectral, il existe une matrice unitaire $U \in \mathcal{M}_{d,d}(\mathbb{C})$ telle que

$$M^*M = U^*DU$$

où D est une matrice diagonale dont les coefficients diagonaux sont réels que l'on note $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$. De plus, si $(e_1, \dots, e_d)$ est la base canonique de $\mathbb{C}^d$, pour tout $j \in \{1, \dots, d\}$, on a

$$\begin{aligned} \lambda_j &= \langle De_j, e_j \rangle_{\mathbb{C}^d} = \langle U(M^*M)U^*e_j, e_j \rangle_{\mathbb{C}^d} \\ &= \langle U(M^*M)U^*e_j, U^*e_j \rangle_{\mathbb{C}^d} \\ &= \langle MU^*e_j, MU^*e_j \rangle_{\mathbb{C}^q} \\ &= \|MU^*e_j\|_{\mathbb{C}^q}^2 \geq 0. \end{aligned}$$

En particulier, les éléments de D sont positifs et la quantité $\sqrt{\max(M^*M)}$ est bien définie.

Soit $x \in \mathbb{C}^d \setminus \{0\}$, on a

$$\begin{aligned} \|Mx\|_{\mathbb{C}^q}^2 &= \langle Mx, Mx \rangle_{\mathbb{C}^q} = \langle M^*Mx, x \rangle_{\mathbb{C}^d} \\ &= \langle U^*DUx, x \rangle_{\mathbb{C}^d} \\ &= \langle D(Ux), Ux \rangle_{\mathbb{C}^d} \\ &= \sum_{j=1}^d \lambda_j |(Ux)_j|^2 \\ &\leq \lambda_1 \|Ux\|_{\mathbb{C}^d}^2 \\ &= \lambda_1 \|x\|_{\mathbb{C}^d}^2. \end{aligned}$$

On en déduit que

$$\frac{\|Mx\|_{\mathbb{C}^q}}{\|x\|_{\mathbb{C}^d}} \leq \sqrt{\lambda_1} = \sqrt{\max(\text{Sp}(M^*M))}.$$

Soit $x = U^*e_1$, on a

$$\begin{aligned} \|Mx\|_{\mathbb{C}^q}^2 &= \langle M^*Mx, x \rangle_{\mathbb{C}^d} = \langle M^*MU^*e_1, U^*e_1 \rangle_{\mathbb{C}^d} \\ &= \langle UM^*MU^*e_1, e_1 \rangle_{\mathbb{C}^d} \\ &= \langle \Sigma e_1, e_1 \rangle_{\mathbb{C}^d} \\ &= \lambda_1. \end{aligned}$$

De plus, on a $\|x\|_{\mathbb{C}^d} = \|U^* e_1\|_{\mathbb{C}^d} = \|e_1\|_{\mathbb{C}^d} = 1$ et donc

$$\|M\| \geq \frac{\|Mx\|_{\mathbb{C}^q}}{\|x\|_{\mathbb{C}^d}} = \sqrt{\lambda_1} = \sqrt{\max(\text{Sp}(M^*M))}.$$

On en déduit que

$$\|M\| = \sqrt{\max(\text{Sp}(M^*M))}$$

(iii) Soit $x \in \mathbb{C}^d \setminus \{0\}$. Pour tout $k \in \{1, \dots, q\}$, on a

$$\begin{aligned} |(Mx)_k| &= \left| \sum_{j=1}^d m_{k,j} x_j \right| \leq \sum_{j=1}^d |m_{k,j}| |x_j| \\ &\leq \|x\|_{\mathbb{C}^d} \max_{k \in \{1, \dots, q\}} \left(\sum_{j=1}^d |m_{k,j}| \right) \end{aligned}$$

et par passage au max sur $k \in \{1, \dots, q\}$, on a

$$\|Mx\|_{\mathbb{C}^q} \leq \|x\|_{\mathbb{C}^d} \max_{k \in \{1, \dots, q\}} \left(\sum_{j=1}^d |m_{k,j}| \right)$$

ce qui nous donne

$$\frac{\|Mx\|_{\mathbb{C}^q}}{\|x\|_{\mathbb{C}^d}} \leq \max_{k \in \{1, \dots, q\}} \left(\sum_{j=1}^d |m_{k,j}| \right).$$

Soit $k_0 \in \{1, \dots, q\}$ tel que

$$\left(\sum_{j=1}^d |m_{k_0,j}| \right) = \max_{k \in \{1, \dots, q\}} \left(\sum_{j=1}^d |m_{k,j}| \right)$$

On construit le vecteur $x = (x_1, \dots, x_d)^T \in \mathbb{C}^d$ définit pour tout $j \in \{1, \dots, d\}$ par

$$x_j = \begin{cases} \frac{\overline{m_{k_0,j}}}{|m_{k_0,j}|} & \text{si } m_{k_0,j} \neq 0 \\ 0 & \text{sinon.} \end{cases}$$

On a pour tout $k \in \{1, \dots, q\}$:

$$|(Mx)_k| = \left| \sum_{\substack{j=1 \\ m_{k_0,j} \neq 0}}^d m_{k,j} \frac{\overline{m_{k_0,j}}}{|m_{k_0,j}|} \right| \leq \sum_{\substack{j=1 \\ m_{k_0,j} \neq 0}}^d |m_{j,k}| \leq \sum_{j=1}^d |m_{j,k}| \leq \sum_{j=1}^d |m_{j,k_0}|$$

et donc

$$\|Mx\|_{\mathbb{C}^q} \leq \sum_{j=1}^d |m_{j,k_0}|.$$

Comme

$$(Mx)_{k_0} = \sum_{\substack{j=1 \\ m_{k_0,j} \neq 0}}^d m_{j,k_0} \frac{\overline{m_{k_0,j}}}{|m_{k_0,j}|} = \sum_{\substack{j=1 \\ m_{k_0,j} \neq 0}}^d |m_{j,k_0}| = \sum_{j=1}^d |m_{j,k_0}| \geq \|Mx\|_{\mathbb{C}^q}$$

on en déduit que

$$\|Mx\|_{\mathbb{C}^q} = \sum_{j=1}^d |m_{j,k_0}|.$$

De plus, si $M \neq 0^1$, nécessairement il existe j_0 tel que $m_{j_0,k_0} \neq 0$ et donc

$$\|x\|_{\mathbb{C}^d} = 1.$$

1. Si M est la matrice nulle, le résultat est bien évidemment vrai puisque tous les termes sont nuls.

On en déduit que

$$\|M\| \geq \frac{\|Mx\|_{\mathbb{C}^q}}{\|x\|_{\mathbb{C}^d}} = \sum_{j=1}^d |m_{j,k_0}| = \max_{k \in \{1, \dots, q\}} \left(\sum_{j=1}^d |m_{k,j}| \right).$$

En particulier, on a bien obtenu que

$$\|M\| = \max_{k \in \{1, \dots, q\}} \left(\sum_{j=1}^d |m_{k,j}| \right).$$

Solution (de l'Exercice 6). On remarque que l'on a

$$\|b\| = \|Mx\| \leq \|M\| \|x\|, \quad \|\delta x\| = \|M^{-1}(b + \delta b) - x\| = \|M^{-1}(b + \delta b) - M^{-1}b\| = \|M^{-1}\delta b\| \leq \|M^{-1}\| \|\delta b\|$$

Comme M est inversible, M est différente de la matrice nulle et donc $\|M\|$ est non-nulle. Par ailleurs, si $b \neq 0$ alors nécessairement, $x \neq 0$ car $x = M^{-1}b$. Donc on a

$$\frac{1}{\|x\|} \leq \frac{\|M\|}{\|b\|}, \quad \|\delta x\| \leq \|M^{-1}\| \|\delta b\|.$$

On en déduit que

$$\frac{\|\delta x\|}{\|x\|} \leq (\|M\| \|M^{-1}\|) \frac{\|\delta b\|}{\|b\|}.$$

Solution (de l'Exercice 7). Soit $M \in GL_d(\mathbb{C})$.

(i) On a $Id = MM^{-1}$ et donc

$$1 = \|Id\| = \|MM^{-1}\| \leq \|M\| \|M^{-1}\| = \kappa(M).$$

Comme $(M^{-1})^{-1} = M$ on a

$$\kappa(M^{-1}) = \|M^{-1}\| \|(M^{-1})^{-1}\| = \|M^{-1}\| \|M\| = \kappa(M).$$

De même si $\alpha \in \mathbb{C}$, on a

$$\kappa(\alpha M) = \|\alpha M\| \|\alpha^{-1} M^{-1}\| = \|M\| \|M^{-1}\| = \kappa(M).$$

(ii) M^*M est une matrice hermitienne positive. Elle est diagonalisable et son spectre est constitué de valeurs propres réelles telles que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$. En effet, M^*M est inversible est positive car pour tout $x \in \mathbb{C}^d$ on a

$$\langle M^*Mx, x \rangle_{\mathbb{C}^d} = \|Mx\|_{\ell^2(\mathbb{C}^d)}^2 \geq 0$$

et donc si λ est une valeur propre de M^*M et $x \in \mathbb{C}^d \setminus \{0\}$ un vecteur propre associé, on a

$$\lambda \|x\|_{\ell^2(\mathbb{C}^d)}^2 = \langle M^*Mx, x \rangle_{\mathbb{C}^d} = \|Mx\|_{\ell^2(\mathbb{C}^d)}^2 \geq 0.$$

De plus, comme M est inversible, 0 n'est pas une valeur propre.

Les valeurs singulières de M sont définies pour tout $j \in \{1, \dots, d\}$ par $\sigma_j(M) = \sqrt{\lambda_j}$. Par ailleurs, pour $x \in \mathbb{R}^d \setminus \{0\}$, on a

$$\|Mx\|_{\ell^2(\mathbb{C}^d)}^2 = \langle Mx, Mx \rangle_{\mathbb{C}^d} = \langle M^*Mx, x \rangle_{\mathbb{C}^d}.$$

Si $(e_1, \dots, e_d)$ est une base orthonormée avec pour tout $j \in \{1, \dots, d\}$, e_j un vecteur propre associé à λ_j ,

on a $x = \sum_{j=1}^d x_j e_j$ et $\|x\|_{\ell^2(\mathbb{C}^d)}^2 = \sum_{j=1}^d |x_j|^2$ donc

$$\|Mx\|_{\ell^2(\mathbb{C}^d)}^2 = \sum_{j=1}^d \lambda_j |x_j|^2 \leq \lambda_1 \|x\|_{\ell^2(\mathbb{C}^d)}^2$$

et on en déduit en divisant par $\|x\|_{\ell^2(\mathbb{C}^d)}^2$ et en prenant la racine carrée que

$$\|M\| \leq \sqrt{\lambda_1}.$$

On remarque que l'on a $\|e_1\|_{\ell^2(\mathbb{C}^d)} = 1$ et

$$\|Me_1\|_{\ell^2(\mathbb{C}^d)}^2 = \langle M^*Me_1, e_1 \rangle_{\mathbb{C}^d} = \lambda_1$$

et donc

$$\|M\| \geq \frac{\|Me_1\|_{\ell^2(\mathbb{C}^d)}}{\|e_1\|_{\ell^2(\mathbb{C}^d)}} = \sqrt{\lambda_1}.$$

On en déduit que $\|M\| = \sqrt{\lambda_1}$. De la même façon, on démontrerait que $\|M^{-1}\| = \frac{1}{\sqrt{\lambda_d}}$ car $\frac{1}{\lambda_d}$ est la plus grande valeur propre de $(M^*M)^{-1}$. On en déduit que

$$\kappa_2(M) = \|M\| \|M^{-1}\| = \frac{\sqrt{\lambda_1}}{\sqrt{\lambda_d}}.$$

(iii) C'est une conséquence du point précédent : si $M^*M = Id$ alors ses valeurs singulières sont toutes égales à 1.

(iv) Soit U une matrice unitaire, il suffit de démontrer que pour toute matrice $A \in \mathcal{M}_{d,d}(\mathbb{C})$, on a

$$\|UA\| = \|A\|.$$

Soit $x \in \mathbb{C}^d \setminus \{0\}$, on a

$$\|UAx\|_{\ell^2(\mathbb{C}^d)} = \|Ax\|_{\ell^2(\mathbb{C}^d)}.$$

En particulier, on obtient

$$\|UA\| = \sup_{\substack{x \in \mathbb{C}^d \\ x \neq 0}} \left(\frac{\|UAx\|_{\ell^2(\mathbb{C}^d)}}{\|x\|_{\ell^2(\mathbb{C}^d)}} \right) = \sup_{\substack{x \in \mathbb{C}^d \\ x \neq 0}} \left(\frac{\|Ax\|_{\ell^2(\mathbb{C}^d)}}{\|x\|_{\ell^2(\mathbb{C}^d)}} \right) = \|A\|.$$

Solution (de l'Exercice 8).

(i) On remarque que

$$\det(M_\varepsilon) = \begin{vmatrix} 1 & 1+\varepsilon \\ 1 & 1 \end{vmatrix} = 1 - (1+\varepsilon) = -\varepsilon < 0$$

donc M_ε est inversible et son inverse est

$$M_\varepsilon^{-1} = \frac{1}{\varepsilon} \begin{pmatrix} -1 & 1+\varepsilon \\ 1 & -1 \end{pmatrix}.$$

(ii) On a

$$x_\varepsilon = \frac{1}{\varepsilon} \begin{pmatrix} -1 & 1+\varepsilon \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \pi \\ e \end{pmatrix} = \begin{pmatrix} \frac{e-\pi}{\varepsilon} + e \\ \frac{\pi-e}{\varepsilon} \end{pmatrix}.$$

(iii) On pose $x = M_\varepsilon^{-1}b$, $\delta b = \begin{pmatrix} \pi \\ e \end{pmatrix} - b$ et $\delta x = x - x_\varepsilon$. On a

$$M_\varepsilon(x_\varepsilon) = \begin{pmatrix} \pi \\ e \end{pmatrix}, \quad M_\varepsilon x = M_\varepsilon(x_\varepsilon + \delta x).$$

On a donc

$$\frac{\|\delta x\|_{\ell^2(\mathbb{C}^2)}}{\|(\frac{e-\pi}{\varepsilon} + e, \frac{\pi-e}{\varepsilon})^\top\|_{\ell^2(\mathbb{C}^2)}} \leq \kappa_2(M) \frac{\|\delta b\|_{\ell^2(\mathbb{C}^2)}}{\|(\pi, e)^\top\|_{\ell^2(\mathbb{C}^2)}}$$

On a

$$\|(\frac{e-\pi}{\varepsilon} + e, \frac{\pi-e}{\varepsilon})^\top\|_{\ell^2(\mathbb{C}^2)} = \frac{\pi-e}{\varepsilon} \sqrt{1 + (1 + \varepsilon \frac{e}{e-\pi})^2}, \quad \|(\pi, e)^\top\|_{\ell^2(\mathbb{C}^2)} = \sqrt{e^2 + \pi^2}.$$

Par ailleurs, on remarque que

$$M^*M = \begin{pmatrix} 1 & 1 \\ 1+\varepsilon & 1 \end{pmatrix} \begin{pmatrix} 1 & 1+\varepsilon \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2+\varepsilon \\ 2+\varepsilon & (1+\varepsilon)^2 + 1 \end{pmatrix}$$

et donc pour $\lambda \in \mathbb{R}$, on a

$$\begin{aligned}\det(M^*M - \lambda Id) &= \begin{vmatrix} 2 - \lambda & 2 + \varepsilon \\ 2 + \varepsilon & (1 + \varepsilon)^2 + 1 - \lambda \end{vmatrix} = (2 - \lambda)((1 + \varepsilon)^2 + 1 - \lambda) - (2 + \varepsilon)^2 \\ &= \lambda^2 - (2 + (1 + \varepsilon)^2 + 1)\lambda + 2(1 + \varepsilon)^2 + 2 - (2 + \varepsilon)^2 \\ &= \lambda^2 - (4 + 2\varepsilon + \varepsilon^2)\lambda + \varepsilon^2.\end{aligned}$$

Et donc, on obtient

$$\Delta = (4 + 2\varepsilon + \varepsilon^2)^2 - 4\varepsilon^2 = (4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2) > 0.$$

Les valeurs propres de M^*M sont donc

$$\lambda_1 = \frac{1}{2} \left(4 + 2\varepsilon + \varepsilon^2 + \sqrt{(4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2)} \right), \quad \lambda_2 = \frac{1}{2} \left(4 + 2\varepsilon + \varepsilon^2 - \sqrt{(4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2)} \right)$$

et donc

$$\begin{aligned}\kappa_2(M)^2 &= \frac{4 + 2\varepsilon + \varepsilon^2 + \sqrt{(4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2)}}{4 + 2\varepsilon + \varepsilon^2 - \sqrt{(4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2)}} \\ &= \frac{\left(4 + 2\varepsilon + \varepsilon^2 + \sqrt{(4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2)} \right)^2}{4\varepsilon^2}.\end{aligned}$$

Ainsi, on obtient

$$\kappa_2(M) = \frac{4 + 2\varepsilon + \varepsilon^2 + \sqrt{(4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2)}}{2\varepsilon}$$

ce qui donne

$$\frac{\|\delta x\|_{\ell^2(\mathbb{C}^d)}}{\|x_\varepsilon\|_{\ell^2(\mathbb{C}^d)}} \leq \frac{4 + 2\varepsilon + \varepsilon^2 + \sqrt{(4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2)}}{2\varepsilon} \eta.$$

En particulier, une condition suffisante pour avoir une erreur relative sur la solution x_ε de 10^{-6} est

$$\frac{4 + 2\varepsilon + \varepsilon^2 + \sqrt{(4 + \varepsilon^2)(4 + 4\varepsilon + \varepsilon^2)}}{2\varepsilon} \leq 10^{-6}$$

Le terme dans le membre de gauche converge vers $+\infty$ lorsque $\varepsilon \rightarrow 0$: il n'est pas possible d'avoir une précision suffisante sur l'erreur relative commise en résolvant le système $M_\varepsilon x = b$ pour ε petit.

Solution (de l'Exercice 9). On pose $B = M^{-1}N$.

(i) $\implies$ (ii) : Supposons que la méthode itérative converge. Pour tout $n \in \mathbb{N}$, on définit l'erreur à l'ordre n par $e_n = x_n - x$. Supposons par l'absurde que $\rho(B) \geq 1$, cela veut dire qu'il existe une valeur propre de B^*B qui est plus grande que 1. Notons λ cette valeur propre et ε un vecteur propre normalisé associé. On a

$$x_{n+1} = Bx_n + M^{-1}b$$

et donc

$$e_{n+1} = B(x_n - x) + Bx + M^{-1}b - x = e_{n+1} = Be_n + Bx + M^{-1}b - x$$

Mais on a $f(x) = x$ c'est-à-dire

$$x = M^{-1}Nx + M^{-1}b$$

donc

$$e_{n+1} = Be_n$$

et par une récurrence immédiate (à rédiger) on obtient pour tout $n \in \mathbb{N}$

$$e_n = B^n e_0.$$

Si on choisit $x_{in} = \varepsilon + x$, on a alors $e_0 = \varepsilon$ et

$$e_n = \lambda_1^n \varepsilon.$$

Or, comme la méthode converge on a $\|e_n\| \rightarrow 0$ lorsque $n \rightarrow +\infty$. Par ailleurs, on a

$$\|e_n\| = \lambda_1^n \xrightarrow{n \rightarrow +\infty} \begin{cases} +\infty & \text{si } \lambda_1 > 1 \\ 1 & \text{sinon} \end{cases}$$

c'est absurde et on en déduit que nécessairement $\lambda_1 < 1$ et donc que $\rho(B) < 1$.

(ii) $\implies$ (iii) : Supposons que $\rho(B) < 1$. Alors, on sait que pour la norme subordonnée à la norme $\ell^2(\mathbb{C}^d)$ on a $\|B\| = \sqrt{\rho(B)} < 1$.

(iii) $\implies$ (i) : Supposons qu'il existe une norme matricielle telle que $\|B\| < 1$. Alors, comme pour tout $n \in \mathbb{N}$ on a $e_n = B^n e_0$ on obtient

$$\|e_n\| = \|B^n e_0\| \leq \|B^n\| \|e_0\| \leq \|B\|^n \|e_0\|$$

où on a utilisé la norme $\|\cdot\|$ sur $\mathbb{C}^d$ associée à la norme subordonnée $\|\cdot\|$. Par encadrement, comme $\|B\| < 1$, lorsque $n \rightarrow +\infty$ on a $\|e_n\| = 0$ et donc la méthode itérative converge.

Solution (de l'Exercice 10).

(i) On a

$$Ax = b \iff (M - N)x = b \iff Mx = Nx + b \iff x = M^{-1}Nx + M^{-1}b$$

où on a utilisé le fait que M est inversible car les coefficients diagonaux de A sont non nuls parce que la matrice A est à diagonale strictement dominante. Si on pose

$$J = M^{-1}N, \quad c = M^{-1}b$$

on obtient le résultat demandé, à savoir

$$x_{n+1} = Jx_n + c.$$

(ii) Pour tout $n \in \mathbb{N}$, on pose $e_n = x_n - x$ et en utilisant le fait que $x = M^{-1}Nx + M^{-1}b$, on remarque que

$$\begin{aligned} e_{n+1} &= Jx_n + c - x = Je_n + Jx + c - M^{-1}Nx - M^{-1}b = Je_n + M^{-1}Nx + M^{-1}b - M^{-1}Nx - M^{-1}b \\ &= Je_n. \end{aligned}$$

On en déduit que pour tout $n \in \mathbb{N}$, on a $e_n = J^n e_0$ et donc

$$\|e_n\|_{\ell^\infty(\mathbb{C}^d)} \leq \|J\|^n \|e_0\|_{\ell^\infty(\mathbb{C}^d)}$$

où $\|\cdot\|$ est la norme subordonnée à $\|\cdot\|_{\ell^\infty(\mathbb{C}^d)}$. Pour tout $j, k \in \{1, \dots, d\}$, on a

$$J_{j,k} = \sum_{p=1}^d (M^{-1})_{j,p} n_{p,k} = \frac{1}{a_{j,j}} n_{j,k} = \begin{cases} -\frac{a_{j,k}}{a_{j,j}} & \text{si } j \neq k \\ 0 & \text{sinon} \end{cases}.$$

En particulier, pour tout $j \in \{1, \dots, d\}$, on a

$$\sum_{k=1}^d |J_{j,k}| = \sum_{\substack{k=1 \\ k \neq j}}^d \frac{|a_{j,k}|}{|a_{j,j}|} = \frac{1}{|a_{j,j}|} \left(\sum_{\substack{k=1 \\ k \neq j}}^d |a_{j,k}| \right) < 1$$

car A est à diagonale strictement dominante. On en déduit que

$$\|J\| = \max_{k=1}^d \left(\sum_{j=1}^d |J_{j,k}| \right) < 1$$

et donc, comme $\|e_n\|_{\ell^\infty(\mathbb{C}^d)} \leq \|J\|^n \|e_0\|_{\ell^\infty(\mathbb{C}^d)}$, en passant à la limite lorsque $n \rightarrow +\infty$ on obtient que $\lim_{n \rightarrow +\infty} x_n = x$ et donc que la méthode converge.

Solution (de l'Exercice 11).

(i) C'est exactement pareil qu'à l'exercice précédent :

$$G = M^{-1}N, \quad c = M^{-1}b.$$

(ii) On définit le produit hermitien sur $\mathbb{C}^d$ par

$$\varphi : (y, z) \in \mathbb{C}^d \times \mathbb{C}^d \mapsto \langle Ay, z \rangle_{\mathbb{C}^d}.$$

On note $\|\cdot\|$ la norme associée à ce produit hermitien dans $\mathbb{C}^d$ et $\|\cdot\|$ la norme subordonnée associée. On a

$$\|M^{-1}N\| = \|M^{-1}(M - A)\| = \|Id - M^{-1}A\|.$$

Pour $v \in \mathbb{C}^d$ tel que $\|v\| = 1$, on a

$$\begin{aligned}
\|v - M^{-1}Av\|^2 &= \varphi(v - M^{-1}Av, v - M^{-1}Av) = \varphi(v, v) + \varphi(M^{-1}Av, M^{-1}Av) - 2\Re(\varphi(v, M^{-1}Av)) \\
&= \|v\| + \langle AM^{-1}Av, M^{-1}Av \rangle_{\mathbb{C}^d} - 2\Re(\langle Av, M^{-1}Av \rangle_{\mathbb{C}^d}) \\
&= 1 + \langle Aw, w \rangle_{\mathbb{C}^d} - 2\Re(\langle Av, w \rangle_{\mathbb{C}^d}) \\
&= 1 + \langle Aw, w \rangle_{\mathbb{C}^d} - 2\Re(\langle Mw, w \rangle_{\mathbb{C}^d}) \\
&= 1 + \langle Mw, w \rangle_{\mathbb{C}^d} - \langle Nw, w \rangle_{\mathbb{C}^d} - \langle Mw, w \rangle_{\mathbb{C}^d} - \langle w, Mw \rangle_{\mathbb{C}^d} \\
&= 1 - \langle (M^* + N)w, w \rangle_{\mathbb{C}^d} < 1
\end{aligned}$$

car $M^* + N$ étant définie positive on a nécessairement $\langle (M^* + N)w, w \rangle_{\mathbb{C}^d} > 0$.

On en déduit que pour tout $v \in \mathbb{C}^d$ tel que $\|v\| = 1$, on a

$$\|v - M^{-1}Av\| < 1.$$

Soit $K = \{v \in \mathbb{C}^d : \|v\| = 1\}$. K est un ensemble fermé borné dans $\mathbb{C}^d$ donc il est compact. De plus, l'application

$$g : v \in \mathbb{C}^d \mapsto \|v - M^{-1}Av\|^2$$

est continue sur $\mathbb{C}^d$ car c'est une application polynomiale (à plusieurs variables). Donc $g|_K$ est bornée et atteint sa borne en un certain $v_* \in \mathbb{C}^d$ vérifiant

$$\sqrt{g(v_*)} = \|I - M^{-1}A\| < 1$$

et donc $\rho(M^{-1}N) < 1$ et la méthode converge.

Solution (de l'Exercice 12).

(i) Soit $n \in \mathbb{N}^*$, on décompose x_0 dans la base de diagonalisation $(e_1, \dots, e_d)$, on a

$$x_0 = \sum_{j=1}^d x_{0,j} e_j$$

et donc

$$M^n x_0 = \sum_{j=1}^d \lambda_j^n x_{0,j} e_j.$$

En particulier, comme par hypothèse $x_0 = x_{in} \notin \text{vect}(e_2, \dots, e_d)$ alors $x_{0,1} \neq 0$ et comme $|\lambda_1| > \max_{j \in \{2, \dots, d\}} |\lambda_j| \geq 0$ alors $\lambda_1 \neq 0$. On en déduit que la composante selon e_1 de $M^n x_0$ est $\lambda_1^n x_{0,1} \neq 0$ et donc $M^n x_0 \neq 0_{\mathbb{C}^d}$.

(ii) Démontrer par récurrence que pour tout $n \in \mathbb{N}^*$, on a $x_n = \frac{M^n x_0}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}}$.

Initialisation : Pour $n = 1$ c'est évident par définition.

Hypothèse de récurrence : Soit $n \geq 1$, on suppose que $x_n = \frac{M^n x_0}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}}$.

Hérédité : On a

$$x_{n+1} = \frac{M x_n}{\|x_n\|_{\ell^2(\mathbb{C}^d)}} = \frac{M \frac{M^n x_0}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}}}{\left\| \frac{M^n M x_0}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}} \right\|_{\ell^2(\mathbb{C}^d)}} = \frac{M^{n+1} x_0}{\|M^{n+1} x_0\|_{\ell^2(\mathbb{C}^d)}}$$

ce qui est précisément la propriété à vérifier au rang $n + 1$.

D'après le principe de raisonnement par récurrence, on en déduit que pour tout $n \in \mathbb{N}^*$, on a $x_n = \frac{M^n x_0}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}}$.

(iii) Soit $n \in \mathbb{N}^*$, on a

$$\begin{aligned}
\lambda_1^{-n} M^n x_0 &= \lambda_1^{-n} \sum_{j=1}^d \lambda_j^n x_{0,j} e_j = x_{0,j} e_1 + \underbrace{\sum_{j=2}^d \left(\frac{\lambda_j}{\lambda_1}\right)^n x_{0,j} e_j}_{\varepsilon_n}.
\end{aligned}$$

On remarque que l'on a

$$\left\| \sum_{j=2}^d \left(\frac{\lambda_j}{\lambda_1} \right)^n x_{0,j} e_j \right\|_{\ell^2(\mathbb{C}^d)} \leq \sum_{j=2}^d \left| \frac{\lambda_j}{\lambda_1} \right|^n |x_{0,j}| \|e_j\|_{\ell^2(\mathbb{C}^d)}.$$

Comme pour tout $j \in \{2, \dots, d\}$, on a $\left| \frac{\lambda_j}{\lambda_1} \right| < 1$, on en déduit que $\lim_{n \rightarrow +\infty} \varepsilon_n = 0$ et donc que

$$\lim_{n \rightarrow +\infty} (\lambda_1^{-n} M^n x_0) = x_{0,j} e_1.$$

(iv) Soit $n \in \mathbb{N}^*$, on a

$$\begin{aligned} \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n x_n &= \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n \frac{M^n x_0}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}} = \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n \left(\frac{\lambda_1^n}{\lambda_1^n} \right) \frac{M^n x_0}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}} \\ &= \frac{|\lambda_1|^n}{\lambda_1^n} \frac{M^n x_0}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}}. \end{aligned}$$

D'après la question précédente, on a

$$\lim_{n \rightarrow +\infty} (\lambda_1^{-n} M^n x_0) = x_{0,1} e_1, \quad \lim_{n \rightarrow +\infty} \left(|\lambda_1|^n \frac{1}{\|M^n x_0\|_{\ell^2(\mathbb{C}^d)}} \right) = \frac{1}{|x_{0,1}| \|e_1\|_{\ell^2(\mathbb{C}^d)}}.$$

et donc

$$\lim_{n \rightarrow +\infty} \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n x_n = \frac{x_{0,1}}{|x_{0,1}| \|e_1\|_{\ell^2(\mathbb{C}^d)}} e_1.$$

(v) Soit $n \in \mathbb{N}^*$, on a

$$\begin{aligned} \langle x_{n+1}, x_n \rangle_{\mathbb{C}^d} &= \left\langle \left(\frac{|\lambda_1|}{\lambda_1} \right)^{n+1} \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^{n+1} x_{n+1}, \left(\frac{|\lambda_1|}{\lambda_1} \right)^n \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n x_n \right\rangle_{\mathbb{C}^d} \\ &= \left(\frac{|\lambda_1|}{\lambda_1} \right)^{n+1} \left(\frac{|\lambda_1|}{\lambda_1} \right)^n \left\langle \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^{n+1} x_{n+1}, \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n x_n \right\rangle_{\mathbb{C}^d} \\ &= \frac{|\lambda_1|^{2n+1}}{\lambda_1 |\lambda_1|^{2n}} \left\langle \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^{n+1} x_{n+1}, \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n x_n \right\rangle_{\mathbb{C}^d} \\ &= \lambda_1 \left\langle \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^{n+1} x_{n+1}, \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n x_n \right\rangle_{\mathbb{C}^d}. \end{aligned}$$

Maintenant, comme le produit scalaire canonique dans $\mathbb{C}^d$ est une forme bilinéaire continue, on peut passer à la limite dans le produit scalaire et grâce à la question précédente, on obtient

$$\lim_{n \rightarrow +\infty} \left\langle \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^{n+1} x_{n+1}, \left(\frac{\overline{\lambda_1}}{|\lambda_1|} \right)^n x_n \right\rangle_{\mathbb{C}^d} = \left\langle \frac{x_{0,1}}{|x_{0,1}| \|e_1\|_{\ell^2(\mathbb{C}^d)}} e_1, \frac{x_{0,1}}{|x_{0,1}| \|e_1\|_{\ell^2(\mathbb{C}^d)}} e_1 \right\rangle = 1$$

et donc

$$\lim_{n \rightarrow +\infty} \langle x_{n+1}, x_n \rangle_{\mathbb{C}^d} = \lambda_1.$$

Solution (de l'Exercice 13).

Posons $M = (m_{j,k})_{\substack{j \in \{1, \dots, N\} \\ k \in \{1, \dots, d+1\}}}$ définie pour tout $j \in \{1, \dots, N\}$ et tout $k \in \{1, \dots, d+1\}$ par

$$m_{j,k} = x_j^{k-1}.$$

et b le vecteur dont les composantes sont les coordonnées y . C'est-à-dire que l'on a

$$M = \begin{pmatrix} 1 & x_1 & \cdots & x_1^d \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_N & \cdots & x_N^d \end{pmatrix}, \quad b = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

Pour tout $a = (a_0, \dots, a_d) \in \mathbb{R}^{d+1}$ et tout $j \in \{1, \dots, N\}$, on a

$$(Ma)_j = \sum_{k=1}^{d+1} m_{j,k} a_{k-1} = \sum_{k=0}^d m_{j,k-1} a_{k-1} = \sum_{k=0}^d a_k x_j^k.$$

On en déduit que

$$\|Ma - b\|_{\ell^2(\mathbb{R}^d)}^2 = \sum_{j=1}^d ((Ma)_j - b_j)^2 = \sum_{j=1}^d \left(\left(\sum_{k=0}^d a_k x_j^k \right) - y_j \right)^2$$

ce qui est précisément la reformulation demandée pour le problème des moindres carrés.

Solution (de l'Exercice 14).

- (i) On sait qu'il existe $\Sigma \in \mathcal{M}_{N,d+1}(\mathbb{R})$ contenant les valeurs singulières de M sur la diagonale, notées $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ et $U \in O_N(\mathbb{R})$ et $V \in O_{d+1}(\mathbb{R})$ telles que

$$M = U\Sigma V^\top.$$

En particulier, pour tout $a \in \mathbb{R}^{d+1}$, on a

$$\|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2 = \|U\Sigma V^\top a - b\|_{\ell^2(\mathbb{R}^N)}^2 = \|\Sigma V^\top a - U^\top b\|_{\ell^2(\mathbb{R}^N)}^2.$$

En posant $\alpha = V^\top a \in \mathbb{R}^{d+1}$ et $\beta = U^\top b \in \mathbb{R}^N$ on obtient

$$\|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2 = \left(\sum_{j=1}^r (\sigma_j \alpha_j - \beta_j)^2 \right) + \sum_{j=r+1}^N \beta_j^2 \geq \sum_{j=r+1}^N \beta_j^2.$$

On en déduit que

$$\inf_{a \in \mathbb{R}^{d+1}} \|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2 \geq \sum_{j=r+1}^N \beta_j^2.$$

et en choisissant $\alpha_j = \frac{1}{\sigma_j} \beta_j$ pour $j \in \{1, \dots, r\}$ et $\alpha_j = 0$ si $j \geq r+1$ on obtient

$$\|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2 = \sum_{j=r+1}^N \beta_j^2$$

et on en déduit que l'infimum est un minimum.

- (ii) Soient $a, h \in \mathbb{R}^{d+1}$, il suffit de développer le produit scalaire

$$\begin{aligned} \|M(a+h) - b\|_{\ell^2(\mathbb{R}^N)}^2 &= \langle M(a+h) - b, M(a+h) - b \rangle_{\mathbb{R}^N} = \langle Ma - b + Mh, Ma - b + Mh \rangle_{\mathbb{R}^N} \\ &= \langle Ma - b, Ma - b \rangle_{\mathbb{R}^N} + \langle Mh, Mh \rangle_{\mathbb{R}^N} \\ &\quad + \langle Ma - b, Mh \rangle_{\mathbb{R}^N} + \langle Mh, Ma - b \rangle_{\mathbb{R}^N} \\ &= \|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2 + \|Mh\|_{\ell^2(\mathbb{R}^N)}^2 \\ &\quad + 2\langle Ma - b, Mh \rangle_{\mathbb{R}^N} \\ &= \|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2 + \|Mh\|_{\ell^2(\mathbb{R}^N)}^2 \\ &\quad + 2\langle M^\top Ma - M^\top b, h \rangle_{\mathbb{R}^{d+1}}. \end{aligned}$$

- (iii) Si a vérifie l'équation normale alors pour tout $h \in \mathbb{R}^d$, on a

$$\|M(a+h) - b\|_{\ell^2(\mathbb{R}^N)}^2 = \|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2 + \|Mh\|_{\ell^2(\mathbb{R}^N)}^2 \geq \|Ma - b\|_{\ell^2(\mathbb{R}^N)}^2.$$

On en déduit que a est donc un minimiseur pour $x \mapsto \|Mx - b\|_{\ell^2(\mathbb{R}^N)}^2$.

- (iv) On remarque que par le développement précédent, on a obtenu une expression que f est différentiable et de différentielle f :

$$(D_a f)(h) = 2\langle M^\top Ma - M^\top b, h \rangle_{\mathbb{R}^{d+1}}.$$

Si a est un minimiseur alors nécessairement pour tout $h \in \mathbb{R}^{d+1}$, on a

$$2\langle M^\top Ma - M^\top b, h \rangle_{\mathbb{R}^{d+1}} = 0.$$

En choisissant $h = M^\top Ma - M^\top b$ on aboutit à l'équation normale.

Solution (de l'Exercice 16). On définit le vecteur $x = (f_1, \dots, f_N)^\top \in \mathbb{C}^N$ et la matrice $M = (m_{j,k})_{j,k \in \{1, \dots, N\}}$ définie pour $j, k \in \{1, \dots, N\}$ par

$$m_{j,k} = \begin{cases} 2 & \text{si } j = k \\ -1 & \text{si } j = k + 1 \\ -1 & \text{si } j = k - 1 \\ 0 & \text{sinon} \end{cases}$$

et on pose $b = h^2(g(t_1), \dots, g(t_N))^\top$. Pour $j \in \{2, \dots, N-1\}$, on a

$$(Mx)_j = \sum_{k=1}^N m_{j,k} x_k = 2x_j - x_{j-1} - x_{j+1} = -2f_j + f_{j-1} + f_{j+1} = h^2 g(t_j) = b_j$$

et aussi

$$(Mx)_1 = \sum_{k=1}^N m_{1,k} x_k = 2x_1 - x_2 = 2f_1 - f_2 = 2f_1 - f_2 - f_0 = h^2 g(t_1)$$

et

$$(Mx)_N = \sum_{k=1}^N m_{N,k} x_k = 2x_N - x_{N-1} = 2f_N - f_{N-1} = 2f_N - f_{N-1} - f_{N+1} = h^2 g(t_N)$$

où on a utilisé le fait que $f_0 = f_{N+1} = 0$. On en déduit que $Mx = b$.

Solution (de l'Exercice 17).

- (i) On remarque que $M^\top = M$ donc $M \in \mathcal{S}_N(\mathbb{R})$ et elle est diagonalisable en base orthonormée et ses valeurs propres sont réelles. De plus, pour $x = (x_1, \dots, x_N)^\top \in \mathbb{R}^N$, on a

$$\begin{aligned} \langle Mx, x \rangle_{\mathbb{R}^d} &= \sum_{j=1}^N (Mx)_j x_j \\ &= \sum_{j=1}^N \sum_{k=1}^N m_{j,k} x_k x_j \\ &= \left(\sum_{k=1}^N m_{1,k} x_k x_1 \right) + \left(\sum_{k=1}^N m_{N,k} x_k x_N \right) + \left(\sum_{j=2}^{N-1} \sum_{k=1}^N m_{j,k} x_k x_j \right) \\ &= 2x_1^2 - x_2 x_1 - x_{N-1} x_N + 2x_N^2 + \left(\sum_{j=2}^{N-1} -x_{j-1} x_j - x_{j+1} x_j + 2x_j^2 \right) \\ &= 2x_1^2 - x_2 x_1 - x_{N-1} x_N + 2x_N^2 + \left(\sum_{j=2}^{N-1} x_j^2 - x_{j-1} x_j \right) + \left(\sum_{j=2}^{N-1} x_j^2 - x_{j+1} x_j \right) \\ &= 2x_1^2 - x_2 x_1 - x_{N-1} x_N + 2x_N^2 + \left(\sum_{j=1}^{N-2} x_{j+1}^2 - x_j x_{j+1} \right) + \left(\sum_{j=2}^{N-1} x_j^2 - x_{j+1} x_j \right) \\ &= 2x_1^2 - x_2 x_1 - x_{N-1} x_N + 2x_N^2 + x_2^2 - x_1 x_2 + x_{N-1}^2 - x_N x_{N-1} + \left(\sum_{j=2}^{N-2} x_{j+1}^2 + x_j^2 - 2x_j x_{j+1} \right) \\ &= x_1^2 + (x_1 - x_2)^2 + x_N^2 + (x_N - x_{N-1})^2 + \sum_{j=2}^{N-2} (x_j - x_{j+1})^2 \\ &\geq 0. \end{aligned}$$

On en déduit que M est positive. De plus si $x = (x_1, \dots, x_N)^\top \in \mathbb{R}^N$ est tel que $\langle Mx, x \rangle_{\mathbb{R}^d} = 0$ alors on aboutit à

$$0 = x_1^2 + (x_1 - x_2)^2 + x_N^2 + (x_N - x_{N-1})^2 + \sum_{j=2}^{N-2} (x_j - x_{j+1})^2.$$

Il s'agit d'une somme de termes positifs qui est nulle et donc chacun des termes est nul. On en déduit que $x_1 = 0$ puis que $x_2 = 0$ et enfin que $x_j = 0$ pour tout $j \in \{1, \dots, N\}$. Ainsi, si λ est valeur propre

2. Il faudrait démontrer rigoureusement par récurrence finie que pour $j \in \{1, \dots, N\}$ on a $x_j = 0$.

associée à $x \neq 0$ on a

$$\lambda \|x\|_{\ell^2(\mathbb{R}^d)}^2 = \langle Mx, x \rangle_{\mathbb{R}^d} > 0.$$

Comme $\|x\|_{\ell^2(\mathbb{R}^d)}^2 > 0$, on en déduit que $\lambda > 0$.

- (ii) Soit $x = (x_1, \dots, x_N) \in \mathbb{R}^d \setminus \{0_{\mathbb{R}^d}\}$ un vecteur propre de M associé à la valeur propre $\lambda > 0$. Pour tout $j \in \{1, \dots, N\}$, on a

$$0 = ((M - \lambda Id)x)_j = \sum_{k=1}^d m_{j,k} x_k = \begin{cases} (2 - \lambda)x_1 - x_2 & \text{si } j = 1 \\ (2 - \lambda)x_N - x_{N-1} & \text{si } j = N \\ (2 - \lambda)x_j - x_{j-1} - x_{j+1} & \text{sinon.} \end{cases}$$

Donc pour tout $j \in \{1, \dots, N\}$, on a

$$x_{j+1} - (2 - \lambda)x_j + x_{j-1} = 0,$$

où on a utilisé le fait que $x_0 = x_{N+1} = 0$. En posant $k = j - 1$, on obtient pour tout $k \in \{0, \dots, N\}$

$$x_{k+2} + (\lambda - 2)x_{k+1} + x_k = 0.$$

- (iii) On sait qu'une valeur propre de M se trouve dans le disque de centre 2 et de rayon 2. On en déduit, comme les valeurs propres sont réelles, qu'elles se trouvent dans l'intervalle $[0, 4]$. On cherche la forme des solutions de cette suite récurrente d'ordre 2, pour cela, on considère l'équation caractéristique

$$r^2 + (\lambda - 2)r + 1 = 0.$$

Son discriminant est

$$\Delta = (\lambda - 2)^2 - 4 = (\lambda - 2 - 2)(\lambda - 2 + 2) = \lambda(\lambda - 4) \leq 0.$$

Si $\Delta = 0$, on a une seule racine

$$r = \frac{2 - \lambda}{2}$$

et alors pour tout $k \in \{0, N + 1\}$, on a

$$x_k = (\alpha k + \beta)r^k$$

avec $\alpha, \beta \in \mathbb{R}$. Nécessairement, $r \neq 0$ car sinon $x = 0_{\mathbb{R}^d}$ ce qui est exclu. Par ailleurs, on a $x_0 = \beta = 0$ donc $\beta = 0$ et $0 = x_{N+1} = \alpha(N + 1)r^{N+1}$ donc $\alpha = 0$ et $x = 0_{\mathbb{R}^d}$ ce qui est exclu : $\Delta > 0$.

Dans ce cas, les racines sont de la forme

$$r_{\pm} := \frac{2 - \lambda \pm i\sqrt{\lambda(4 - \lambda)}}{2}.$$

On remarque que

$$|r_{\pm}| = \frac{1}{4} ((2 - \lambda)^2 - (\lambda - 2)^2 + 4) = 1.$$

En particulier, on sait qu'il existe $\theta \in [0, 2\pi[$ tel que $r_+ = e^{i\theta}$ et $r_- = \overline{r_+} = e^{-i\theta}$. Ainsi, pour tout $k \in \{0, \dots, N + 1\}$, on a

$$x_k = \alpha e^{ik\theta} + \beta e^{-ik\theta}, \quad \alpha, \beta \in \mathbb{C}.$$

On se souvient que l'on cherche des solutions à valeurs dans $\mathbb{R}$. On regarde d'abord la condition initiale $0 = x_0$ qui donne $\alpha + \beta = 0$ et donc

$$x_k = \alpha (e^{ik\theta} - e^{-ik\theta}) = 2i\alpha \sin(k\theta).$$

Par ailleurs, on a

$$0 = x_{N+1} = 2i\alpha \sin((N + 1)\theta)$$

et donc $(N + 1)\theta \equiv 0[\pi]$ et donc $\theta = \frac{p\pi}{N+1}$ pour $p \in \mathbb{Z}$. On remarque que l'on a

$$\cos(\theta) = \frac{2 - \lambda}{2}$$

et donc

$$\lambda = 2 - 2 \cos\left(\frac{p\pi}{N+1}\right).$$

On remarque que pour $p \in \{1, \dots, N\}$ on obtient N valeurs distinctes pour λ et que dans ce cas $x \neq 0_{\mathbb{R}^d}$ car pour tout $k \in \{1, \dots, N\}$ on peut choisir

$$x_k = \sin(kt_p\pi).$$

(iv) On a

$$\kappa_2(M) = \frac{2 - 2\cos(\frac{N\pi}{N+1})}{2 - 2\cos(\frac{\pi}{N+1})} = \frac{1 - \cos(\frac{N\pi}{N+1})}{1 - \cos(\frac{\pi}{N+1})}.$$

En utilisant que lorsque $N \rightarrow +\infty$ on a

$$\cos\left(\frac{\pi}{N+1}\right) - 1 = 1 - \frac{\pi^2}{2(N+1)^2} + o\left(\frac{1}{N^2}\right) = 1 - \frac{\pi^2}{2N^2} + o\left(\frac{1}{N^2}\right)$$

on obtient

$$1 - \cos\left(\frac{\pi}{N+1}\right) \underset{N \rightarrow +\infty}{\sim} \frac{\pi^2}{2N^2}.$$

Par ailleurs

$$\lim_{N \rightarrow +\infty} 1 - \cos\left(\frac{N\pi}{N+1}\right) = 2$$

On en déduit que

$$\kappa_2(M) \underset{N \rightarrow +\infty}{\sim} \frac{4N^2}{\pi^2}.$$

(v) Soit $j \in \{1, \dots, N\}$, on a

$$(M(x-F))_j = 2(x_j - f(t_j)) - (x_{j-1} - f(t_{j-1})) - (x_{j+1} - f(t_{j+1})) = h^2 g(t_j) - (2f(t_j) - f(t_{j-1}) - f(t_{j+1})).$$

où pour $j \in \{1, N\}$ on utilise le fait que $0 = x_0 = x_{N+1} = f(t_0) = f(t_{N+1})$. Comme f est trois fois dérivable sur $[0, 1]$ et $f^{(3)}$ est dérivable sur $]0, 1[$, d'après la formule de Taylor-Lagrange, pour tout $j \in \{1, \dots, N\}$, il existe $c_j^+ \in]t_j, t_{j+1}[$ et $c_j^- \in]t_{j-1}, t_j[$ tels que

$$\begin{aligned} f(t_{j+1}) &= f(t_j) + f'(t_j)h + \frac{h^2}{2}f''(t_j) + \frac{f^{(3)}(t_j)}{6}h^3 + \frac{f^{(4)}(c_j^+)}{24}h^4 \\ f(t_{j-1}) &= f(t_j) - f'(t_j)h + \frac{h^2}{2}f''(t_j) - \frac{f^{(3)}(t_j)}{6}h^3 + \frac{f^{(4)}(c_j^-)}{24}h^4. \end{aligned}$$

En particulier, on obtient

$$h^2 g(t_j) = -h^2 f''(t_j) = (2f(t_j) - f(t_{j-1}) - f(t_{j+1})) + \frac{f^{(4)}(c_j^-) - f^{(4)}(c_j^+)}{24}h^4.$$

On en déduit que

$$|h^2 g(t_j) - (2f(t_j) - f(t_{j-1}) - f(t_{j+1}))| \leq \frac{\|f^{(4)}\|_{L^\infty([0,1])}}{12} h^4$$

et donc comme le membre de droite ne dépend pas de j , on a

$$\|M(x-F)\|_\infty \leq Ch^4$$

pour une constante $C > 0$.

(vi) Pour tout $y \in \mathbb{R}^N$, on sait que l'on a

$$\|y\|_{\ell^\infty(\mathbb{R}^N)} \leq \|y\|_{\ell^2(\mathbb{R}^N)} \leq \sqrt{N}\|y\|_{\ell^\infty(\mathbb{R}^N)}.$$

En particulier, on a

$$\|M(x-F)\|_{\ell^2(\mathbb{R}^N)} \leq \sqrt{N}\|M(x-F)\|_{\ell^\infty(\mathbb{R}^N)}.$$

Or on sait que si $y \in \mathbb{R}^d$, comme M est diagonalisable, il existe O matrice orthogonale telle que $M = O^T D O$ où

$$D = \text{diag}(\lambda_1, \dots, \lambda_N).$$

On en déduit que

$$\begin{aligned} \|My\|_{\ell^2(\mathbb{R}^d)} &= \|O^T D O y\|_{\ell^2(\mathbb{R}^N)} = \|D O y\|_{\ell^2(\mathbb{R}^N)} = \sqrt{\sum_{j=1}^N \lambda_j^2 |(Oy)_j|^2} \geq \left(\inf_{j \in \{1, \dots, N\}} (\lambda_j) \right) \|Oy\|_{\ell^2(\mathbb{R}^d)} \\ &= \inf_{j \in \{1, \dots, N\}} (\lambda_j) \|y\|_{\ell^2(\mathbb{R}^d)}. \end{aligned}$$

En particulier, on obtient

$$\begin{aligned}
 \left(\inf_{j \in \{1, \dots, N\}} (\lambda_j) \right) \|x - F\|_{\ell^\infty(\mathbb{R}^N)} &\leq \left(\inf_{j \in \{1, \dots, N\}} (\lambda_j) \right) \|x - F\|_{\ell^2(\mathbb{R}^N)} \leq \|M(x - F)\|_{\ell^2(\mathbb{R}^N)} \\
 &\leq \sqrt{N} \|M(x - F)\|_{\ell^\infty(\mathbb{R}^N)} \\
 &\leq C \frac{\sqrt{N}}{(N+1)^4} \\
 &\leq \frac{C}{N^{\frac{7}{2}}}.
 \end{aligned}$$

Par ailleurs, on a il existe $c > 0$ et $N_0 \in \mathbb{N}$ tel que pour tout $N \geq N_0$ on ait

$$\inf_{j \in \{1, \dots, N\}} (\lambda_j) = 2 - 2 \cos\left(\frac{\pi}{N+1}\right) \geq \frac{c}{N^2}.$$

On en déduit que pour tout $N \geq N_0$, on a

$$\|x - F\|_{\ell^\infty(\mathbb{R}^N)} \leq \frac{cCN^2}{N^{\frac{7}{2}}} = \frac{K}{\sqrt{N}}.$$

Solution (de l'Exercice 18).

1. Nous allons démontrer par récurrence que pour tout $n \in \mathbb{N}$ on a

$$a_{n+1} \geq a_n, \quad b_{n+1} \leq b_n, \quad a_{n+1} < b_{n+1}.$$

Initialisation : pour $n = 0$, si $f(a_0)f(m_0) \leq 0$ on sait que

$$a_1 = a_0, \quad b_1 = m_0 = \frac{a_0 + b_0}{2} = b_0 - \frac{b_0 - a_0}{2} = b_0 - \frac{b - a}{2}.$$

et donc $a_1 \geq a_0$, $b_1 < b_0$ (car $b - a > 0$) et on a

$$a_1 = a_0 < b_1 = \frac{a_0 + b_0}{2} = a_0 + \underbrace{\frac{b - a}{2}}_{>0}.$$

De même, si $f(a_0)f(m_0) > 0$, on sait que

$$a_1 = \frac{a_0 + b_0}{2} = a_0 + \frac{b_0 - a_0}{2}, \quad b_1 = b_0$$

et donc $a_1 > a_0$ car $b_0 - a_0 = b - a > 0$ et $b_1 \leq b_0$. De plus, on a

$$a_1 = b_0 - \frac{b_0 - a_0}{2} = b_0 - \underbrace{\frac{b - a}{2}}_{>0}.$$

donc $a_1 < b_0 = b_1$.

Hypothèse de récurrence : Soit $n \in \mathbb{N}$, on suppose que

$$a_{n+1} \geq a_n, \quad b_{n+1} \leq b_n, \quad a_{n+1} < b_{n+1}.$$

Hérédité : Si $f(a_{n+1})f(m_{n+1}) \leq 0$, on a

$$a_{n+2} = a_{n+1}, \quad b_{n+2} = m_{n+1} = \frac{b_{n+1} + a_{n+1}}{2} = b_{n+1} - \frac{b_{n+1} - a_{n+1}}{2}.$$

Donc $a_{n+2} \geq a_{n+1}$ et d'après l'hypothèse de récurrence $b_{n+1} - a_{n+1} < 0$ donc $b_{n+2} \leq b_{n+1}$ et de plus

$$a_{n+2} = a_{n+1} = b_{n+2} - (b_{n+2} - a_{n+1}) = b_{n+2} - \frac{b_{n+1} - a_{n+1}}{2}.$$

D'après l'hypothèse de récurrence $b_{n+1} - a_{n+1} > 0$ et donc $a_{n+2} < b_{n+2}$.

Si $f(a_{n+1})f(m_{n+1}) > 0$ alors on a

$$a_{n+2} = m_{n+1} = \frac{a_{n+1} + b_{n+1}}{2} = a_{n+1} + \frac{b_{n+1} - a_{n+1}}{2}, \quad b_{n+2} = b_{n+1}.$$

Comme $b_{n+1} - a_{n+1} > 0$ d'après l'hypothèse de récurrence, on obtient $a_{n+2} \geq a_{n+1}$ et $b_{n+2} \leq b_{n+1}$. De plus, on a

$$a_{n+2} = \frac{a_{n+1} + b_{n+1}}{2} = b_{n+2} - b_{n+1} + \frac{a_{n+1} + b_{n+1}}{2} = b_{n+2} - b_{n+1} + \frac{a_{n+1} + b_{n+1}}{2} = b_{n+2} - \frac{b_{n+1} - a_{n+1}}{2}.$$

D'après l'hypothèse de récurrence, $b_{n+1} - a_{n+1} > 0$ et donc $a_{n+2} < b_{n+2}$, ce qui termine la récurrence. On en déduit que $(a_n)_{n \in \mathbb{N}}$ est croissante et que $(b_n)_{n \in \mathbb{N}}$ est décroissante. De plus, on remarque que pour tout $n \in \mathbb{N}$, on a

$$\begin{aligned} b_{n+1} - a_{n+1} &= \begin{cases} m_n - a_n & \text{si } f(a_n)f(m_n) \leq 0 \\ b_n - m_n & \text{sinon} \end{cases} \\ &= \begin{cases} \frac{a_n + b_n}{2} - a_n & \text{si } f(a_n)f(m_n) \leq 0 \\ b_n - \frac{a_n + b_n}{2} & \text{sinon} \end{cases} \\ &= \frac{b_n - a_n}{2}. \end{aligned}$$

On remarque alors que $(b_n - a_n)_{n \in \mathbb{N}}$ est une suite géométrique de raison $\frac{1}{2}$ et alors on obtient pour tout $n \in \mathbb{N}$ que

$$b_n - a_n = \frac{b_0 - a_0}{2^n} = \frac{b - a}{2^n}.$$

Comme pour tout $n \in \mathbb{N}$, $b_n - a_n > 0$, en passant à la limite lorsque $n \rightarrow +\infty$ on obtient par encadrement

$$\lim_{n \rightarrow +\infty} b_n - a_n = 0.$$

$(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ sont des suites adjacentes et donc converge vers une limite commune notée ℓ . Par ailleurs on a pour tout $n \in \mathbb{N}$:

$$a_n \leq m_n \leq b_n$$

car $m_n = a_n + \frac{b_n - a_n}{2} = b_n - \frac{b_n - a_n}{2}$ et $b_n > a_n$. On en déduit par encadrement que

$$\lim_{n \rightarrow +\infty} m_n = \ell.$$

2. Par construction, pour tout $n \in \mathbb{N}$, on a

$$f(a_n)f(b_n) \leq 0.$$

Par ailleurs, on sait que la fonction f est continue sur $[a, b]$ donc en passant à la limite $n \rightarrow +\infty$ on obtient

$$f(\ell)^2 \leq 0$$

et donc $f(\ell) = 0$.

3. On a déjà vu que

$$b_n - a_n = \frac{b - a}{2^n}.$$

Or, comme $(a_n)_{n \in \mathbb{N}}$ est croissante et $(b_n)_{n \in \mathbb{N}}$ est décroissante, on a

$$\left. \begin{array}{l} b_n - \ell \\ \ell - a_n \end{array} \right\} \leq b_n - \ell + \ell - a_n = b_n - a_n = \frac{b - a}{2^n}.$$

Or $b_n - \ell \geq 0$ donc $|b_n - \ell| = b_n - \ell$ et $\ell - a_n \geq 0$ donc $|a_n - \ell| = \ell - a_n$. On a également

$$|m_n - \ell| = \left| \frac{a_n + b_n}{2} - \ell \right| = \left| \frac{(a_n - \ell) + (b_n - \ell)}{2} \right| \leq \frac{1}{2} (|a_n - \ell| + |b_n - \ell|) \leq \frac{1}{2} \left(\frac{b - a}{2^n} + \frac{b - a}{2^n} \right) = \frac{b - a}{2^n}.$$

Solution (de l'Exercice 19).

- (i) Soient ℓ_1, ℓ_2 deux points fixes pour f . Comme f est contractante, on a

$$d(\ell_1, \ell_2) = d(f(\ell_1), f(\ell_2)) \leq kd(\ell_1, \ell_2)$$

et donc $0 \leq (k-1)d(\ell_1, \ell_2)$ or $k < 1$ donc ceci est possible uniquement si $d(\ell_1, \ell_2) = 0$ c'est-à-dire $\ell_1 = \ell_2$.

- (ii) Pour $n \in \mathbb{N}$, comme f est k -lipschitzienne on a

$$u_{n+1} = d(x_{n+2}, x_{n+1}) = d(f(x_{n+1}), f(x_n)) \leq kd(x_{n+1}, x_n) = ku_n.$$

Montrons par récurrence que pour tout $n \in \mathbb{N}$ $u_n \leq k^n u_0$.

Initialisation : Pour $n = 0$, on a toujours $u_0 \leq u_0 = k^0 u_0$.

Hypothèse de récurrence : Soit $n \in \mathbb{N}$. On suppose que $u_n = k^n u_0$.

Hérédité : On a, d'après l'inégalité et l'hypothèse de récurrence,

$$u_{n+1} \leq ku_n = k(k^n u_0) = k^{n+1} u_0.$$

On a donc démontré la propriété au rang $n+1$ et d'après le principe du raisonnement par récurrence on vient de démontrer que pour tout $n \in \mathbb{N}$ on a

$$u_n \leq k^n u_0.$$

- (iii) Soient $p, n \in \mathbb{N}$. On a

$$d(x_{n+p}, x_n) \leq \sum_{j=0}^{p-1} d(x_{n+j+1}, x_{n+j}) = \sum_{j=0}^{p-1} u_{n+j} \leq \sum_{j=0}^{p-1} k^{n+j} u_0 = u_0 k^n \frac{1-k^p}{1-k}$$

où on a reconnu la somme des termes d'une suite géométrique de raison k . Le membre de droite tend vers 0 lorsque $n, p \rightarrow +\infty$ et on en déduit donc que la suite $(x_n)_{n \in \mathbb{N}}$ est de Cauchy.

- (iv) La suite $(x_n)_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ étant de Cauchy dans l'espace métrique complet (E, d) on sait qu'elle converge vers un certain $\ell \in E$. Comme f est contractante, elle est Lipschitzienne donc continue et on en déduit en passant à la limite dans l'équation $x_{n+1} = f(x_n)$ que

$$\ell = f(\ell).$$

Soit $n \in \mathbb{N}$ fixé, on a pour tout $p \in \mathbb{N}$

$$d(x_{n+p}, x_n) \leq u_0 k^n \frac{1-k^p}{1-k}$$

et en passant à la limite lorsque $p \rightarrow +\infty$ on obtient

$$d(\ell, x_n) \leq \frac{k^n}{1-k} u_0 = \frac{k^n}{1-k^p} d(x_1, x_0)$$

ce qui est précisément l'estimée d'erreur annoncée.

Solution (de l'Exercice 20).

- (i) Posons $r = \|Df(\ell)\| < 1$. Comme $x \mapsto Df(x)$ est continue de Ω dans $\mathcal{L}(\mathbb{R}^d)$, on en déduit que pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que si $x \in B_f(\ell, \delta)$ alors

$$Df(x) \in B_{\|\cdot\|}(Df(\ell), \varepsilon).$$

En particulier, pour $\varepsilon = \frac{1-r}{2}$, il existe $\delta > 0$ tel que pour tout $x \in B_f(\ell, \delta)$ on a

$$\|Df(x)\| \leq r + \varepsilon = r + \frac{1-r}{2} = \frac{r+1}{2} < 1.$$

- (ii) On remarque que pour tout $x \in B_f(\ell, \delta)$, comme $\ell = f(\ell)$, on a

$$\|f(x) - \ell\| = \|f(x) - f(\ell)\| \leq \left(\sup_{y \in B_f(\ell, \delta)} \|Df(y)\| \right) \|x - \ell\|$$

où on a utilisé l'inégalité des accroissements finis pour $f|_{B_f(\ell, \delta)}$ (f est de classe $\mathcal{C}^1$ sur $B_f(\ell, \delta) \subset \Omega$ et $B_f(\ell, \delta)$ est convexe). On a donc

$$\|f(x) - \ell\| < \|x - \ell\| \leq \delta.$$

En particulier $f(B_f(\ell, \delta)) \subset B_f(\ell, \delta)$. Par ailleurs, pour tout $x, y \in B_f(\ell, \delta)$, l'inégalité des accroissements finis livre

$$\|f(x) - f(y)\| \leq \left(\sup_{z \in B_f(\ell, \delta)} \|Df(z)\| \right) \|x - y\|.$$

Or, $B_f(\ell, \delta)$ est un ensemble compact car fermé et borné dans $\mathbb{R}^d$ et l'application $z \mapsto \|Df(z)\|$ est continue de $B_f(\ell, \delta)$ vers $\mathcal{L}(\mathbb{R}^d)$ (car f est de classe $\mathcal{C}^1$ sur Ω à valeurs dans $\mathcal{L}(\mathbb{R}^d)$ et que $\|\cdot\|$ est une norme donc continue sur $\mathcal{L}(\mathbb{R}^d)$). $z \mapsto \|Df(z)\|$ est donc bornée et atteint ses bornes sur $B_f(\ell, \delta)$ et donc il existe $z_* \in B_f(\ell, \delta)$ tel que

$$k := \|Df(z_*)\| = \sup_{z \in B_f(\ell, \delta)} \|Df(z)\| < 1.$$

On en déduit que $f : B_f(\ell, \delta) \rightarrow B_f(\ell, \delta)$ est contractante.

- (iii) $B_f(\ell, \delta)$ étant fermé dans $\mathbb{R}^d$ qui est complet, il y est complet et $B_f(\ell, \delta)$ munit de la distance $d : (x, y) \in B_f(\ell, \delta) \times B_f(\ell, \delta) \mapsto \|x - y\|$ est un espace métrique. D'après le Théorème de Banach-Picard, la suite $(x_n)_{n \in \mathbb{N}}$ converge vers l'unique point fixe de f dans $B_f(\ell, \delta)$, à savoir ℓ .
- (iv) (a) Soit $n \in \mathbb{N}$, la formule de Taylor-Young nous livre

$$f(x_n) = f(\ell) + f'(\ell)(x_n - \ell) + \varepsilon(|x_n - \ell|)$$

où $\rho^{-1}\varepsilon(\rho) \rightarrow 0$ lorsque $\rho \rightarrow 0$. Comme $f(x_n) = x_{n+1}$ et $f(\ell) = \ell$, on a

$$x_{n+1} - \ell = f'(\ell)(x_n - \ell) + \varepsilon(|x_n - \ell|).$$

On en déduit que

$$|x_{n+1} - \ell| \geq |f'(\ell)||x_n - \ell| - |\varepsilon(|x_n - \ell|)|$$

Comme $\rho^{-1}\varepsilon(\rho) \rightarrow 0$ lorsque $\rho \rightarrow 0$, on sait que pour tout $\delta > 0$, il existe $\eta > 0$ tel que si $\rho \in (0, \eta)$ on a $\varepsilon(\rho) < \delta\rho$. Ainsi, en choisissant $\delta = \frac{|f'(\ell)|-1}{2} > 0$, on sait qu'il existe $\eta > 0$ tel que si $\rho \in (0, \eta)$ alors $\varepsilon(\rho) < \frac{|f'(\ell)|-1}{2}\rho$. Par ailleurs comme par hypothèse $x_n - \ell \rightarrow 0$, on sait qu'il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$ on ait $|x_n - \ell| < \eta$. En particulier pour tout $n \geq n_0$ on a $|\varepsilon(|x_n - \ell|)| < \frac{|f'(\ell)|-1}{2}|x_n - \ell|$. On en déduit que pour tout $n \geq n_0$ on a

$$|x_{n+1} - \ell| \geq |f'(\ell)||x_n - \ell| - \left(\frac{|f'(\ell)|-1}{2} \right) |x_n - \ell| = \left(\frac{|f'(\ell)|+1}{2} \right) |x_n - \ell|.$$

En posant $M = \left(\frac{|f'(\ell)|+1}{2} \right)$ on a bien $M > 1$ car $|f'(\ell)| > 1$ et on obtient le résultat demandé.

- (b) Par une récurrence (à écrire proprement le jour de l'écrit), on obtient que pour tout $n \in \mathbb{N}$, on a

$$|x_n - \ell| \geq M^n |x_0 - \ell|.$$

Comme $x_0 \neq \ell$ le membre de droite est strictement positif et comme $M > 1$ il tend vers $+\infty$ lorsque $n \rightarrow +\infty$. On en déduit que $\lim_{n \rightarrow +\infty} |x_n - \ell| = +\infty$ ce qui est absurde puisque $\lim_{n \rightarrow +\infty} x_n = \ell$. On en déduit que $(x_n)_{n \in \mathbb{N}}$ ne converge jamais vers ℓ .

Solution (de l'Exercice 21).

- (i) On considère l'application $\det : \mathcal{M}_{d,d}(\mathbb{R}) \rightarrow \mathbb{R}$ qui est de classe $\mathcal{C}^\infty(\mathcal{M}_{d,d}(\mathbb{R}))$ car on peut le voir comme une application polynomiale (en d^2 variables). Si $(e_1, \dots, e_d)$ est la base canonique de $\mathbb{R}^d$, on considère l'isomorphisme

$$\varphi : \begin{cases} \mathcal{L}(\mathbb{R}^d) & \rightarrow \mathcal{M}_{d,d}(\mathbb{R}) \\ \theta & \mapsto (\langle \theta(e_k), e_j \rangle_{\mathbb{R}^d})_{j,k \in \{1, \dots, d\}} \end{cases}$$

. On remarque que $\varphi \in \mathcal{C}^\infty(\mathcal{L}(\mathbb{R}^d), \mathcal{M}_{d,d}(\mathbb{R}))$ car c'est une application linéaire et par conséquent est différentiable à tout ordre $k \in \mathbb{N}$ avec $D_\theta^{(n)}\varphi = \varphi$. On en déduit que $\alpha : x \mapsto \det \circ \varphi \circ D_x f$ est en particulier continue sur $\mathbb{R}^d$ par composition (par hypothèse $x \mapsto D_x f$ est de classe $\mathcal{C}^1(\Omega)$ donc continue sur Ω). Or, comme on a $\alpha(\ell) \neq 0$, par continuité de α en ℓ , on sait qu'il existe $\eta > 0$ tel que pour tout $x \in B(\ell, \eta)$ $\alpha(\ell) \neq 0$. Autrement dit, $D_x f$ est inversible pour tout $x \in B(\ell, \eta)$.

- (ii) Soit $x \in B(\ell, \eta)$. Notons $A(x) = (D_x f)^{-1} \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)$. On note que comme $x \mapsto D_x f$ est de classe $\mathcal{C}^1(B(\ell, \eta))$ et que $D_x f$ est inversible pour tout $x \in B(\ell, \eta)$ alors $A : x \mapsto (D_x f)^{-1}$ est de classe $\mathcal{C}^1(B(\ell, \eta))$. En effet, on a

$$D_{x+h} f = D_x f + (D_x^2 f)(h, \cdot) + \varepsilon(h)$$

où $D_x^2 f \in \mathcal{L}(\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d), \mathbb{R}^d) \simeq \mathcal{L}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)$ est la différentielle seconde de f en x et avec $\frac{\|\varepsilon(h)\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)}}{\|h\|_{\mathbb{R}^d}} \rightarrow 0$.
On a alors

$$D_{x+h} f = D_x f (Id + (D_x f)^{-1} \circ (D_x^2 f)(h, \cdot) + (D_x f)^{-1} \circ \varepsilon(h))$$

et on remarque que

$$\|(D_x f)^{-1} \circ (D_x^2 f)(h, \cdot) + (D_x f)^{-1} \circ \varepsilon(h)\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)} \xrightarrow{h \rightarrow 0} 0.$$

On en déduit que pour h assez petit, on a

$$(D_{x+h} f)^{-1} = (D_x f)^{-1} - (D_x f)^{-1} \circ (D_x^2 f)(h, D_x f^{-1} \cdot) + (D_x f)^{-1} \circ \varepsilon(h) (D_x f)^{-1}. \quad (\text{A.1})$$

Comme

$$\begin{cases} \mathbb{R}^d & \mapsto \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d) \\ h & \mapsto (D_x f)^{-1} \circ (D_x^2 f)(h, D_x f^{-1} \cdot) \end{cases}$$

est une application linéaire et

$$\frac{\|(D_x f)^{-1} \circ \varepsilon(h) (D_x f)^{-1}\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)}}{\|h\|_{\mathbb{R}^d}} \leq \|(D_x f)^{-1}\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)}^2 \frac{\|\varepsilon(h)\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)}}{\|h\|_{\mathbb{R}^d}} \xrightarrow{h \rightarrow 0} 0$$

On aboutit au fait que $x \mapsto (D_x f)^{-1} = A(x)$ est différentiable sur $B(\ell, \eta)$ et que

$$D_x A(h) = (D_x f)^{-1}(h) = (D_x f)^{-1} \circ (D_x^2 f)(h, D_x f^{-1} \cdot).$$

On a

$$g(x) = x - A(x)(f(x))$$

et donc on obtient

$$\begin{aligned} g(x+h) - g(x) &= h - (A(x+h) - A(x))(f(x+h)) - A(x)(f(x+h) - f(x)) \\ &= h - (D_x A(h) + \varepsilon_1(h))(f(x+h)) - A(x)(D_x f(h) + \varepsilon_2(h)) \end{aligned}$$

où

$$\lim_{h \rightarrow 0} \frac{\|\varepsilon_1(h)\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)}}{\|h\|_{\mathbb{R}^d}} = 0, \quad \lim_{h \rightarrow 0} \frac{\|\varepsilon_2(h)\|_{\mathbb{R}^d}}{\|h\|_{\mathbb{R}^d}} = 0.$$

On en déduit que

$$g(x+h) - g(x) = h - D_x A(h)(f(x)) - A(x)(D_x f(h)) - D_x A(h)(D_x f(h) + \varepsilon_2(h)) - \varepsilon_1(h)(f(x+h)) - A(x)(\varepsilon_2(h)).$$

On on a

$$\begin{aligned} \frac{\| - D_x A(h)(D_x f(h) + \varepsilon_2(h)) - \varepsilon_1(h)(f(x+h)) \|}{\|h\|_{\mathbb{R}^d}} &\leq \frac{1}{\|h\|_{\mathbb{R}^d}} (\|D_x A\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)} \|D_x f\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)} \|h\|_{\mathbb{R}^d}^2 \\ &\quad + \|D_x A\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)} \|h\|_{\mathbb{R}^d} \|\varepsilon_2(h)\|_{\mathbb{R}^d} \\ &\quad + \|\varepsilon_1(h)\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)} \|f(x+h)\|_{\mathbb{R}^d}) \end{aligned}$$

et le membre de droite tend vers 0 quand $h \rightarrow 0$ d'après les hypothèses sur $\varepsilon_1, \varepsilon_2$ et le fait que f soit continue en x .

Par conséquent, on a

$$D_x g(h) = h - D_x A(h)(f(x)) - A(x)(D_x f(h))$$

et en particulier comme $f(\ell) = 0$ on obtient

$$D_\ell g(h) = h - (D_\ell A)(h)(f(\ell)) + A(\ell)(D_\ell f(h)) = h - A(\ell)(D_\ell f(h)) = h - (D_\ell f)^{-1}(D_\ell f(h)) = h - h = 0.$$

On en déduit que

$$\|D_\ell g\|_{\mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)} = \|D_\ell\| < 1$$

et donc qu'il existe $\delta > 0$ tel que pour tout $x_{in} \in B(\ell, \delta)$ on ait la convergence de la suite récurrente $(x_n)_{n \in \mathbb{N}}$.

(iii) Il nous reste à estimer l'erreur $e_n = x_n - \ell$. Soit $x \in B(\ell, \delta)$ (qui est convexe), comme $f \in \mathcal{C}^2(B(\ell, \delta))$, on a la formule de Taylor suivante :

$$\begin{aligned} f(x) &= f(\ell)(x - \ell) + (D_\ell f)(x - \ell) + \int_0^1 (1-t)(D_{(1-t)\ell+tx}^2 f)(x - \ell, x - \ell) dt \\ &= (D_\ell f)(x - \ell) + \int_0^1 (1-t)(D_{(1-t)\ell+tx}^2 f)(x - \ell, x - \ell) dt \end{aligned}$$

car $f(\ell) = 0$. D'après A.1, on a

$$f(x) - \ell = A + B + C + D$$

où

$$\begin{aligned} A &= (D_\ell f)^{-1} \circ (D_\ell^2 f)(x - \ell, ((D_\ell f)^{-1})^2 x - \ell), \\ B &= \int_0^1 (1-t)(D_\ell f)^{-1} \circ (D_\ell f)^2(x - \ell, (D_\ell f)^{-1}(D_{\ell+t(x-\ell)}^2 f)(x - \ell, x - \ell)) dt, \\ C &= (D_\ell f)^{-1} \circ \varepsilon(x - \ell)((D_\ell f)^{-1})^2(x - \ell), \quad D = \int_0^1 (1-t)(D_\ell f)^{-1} \circ \varepsilon(x - \ell)(D_\ell f)^{-1}(D_{\ell+t(x-\ell)}^2 f)(x - \ell, x - \ell) dt. \end{aligned}$$

En particulier, en utilisant le fait que les applications linéaires et bilinéaires sont continues sur $\mathbb{R}^d$, il existe $c > 0$ tel que

$$\|A\|_{\mathbb{R}^d} \leq c\|x - \ell\|_{\mathbb{R}^d}^2, \quad \|C\|_{\mathbb{R}^d} \leq c\|\varepsilon(x - \ell)\| \|x - \ell\|_{\mathbb{R}^d}$$

et

$$\begin{aligned} \|B\|_{\mathbb{R}^d} &\leq c \left(\sup_{t \in [0,1]} \|D_{\ell+t(x-\ell)}^2 f\|_{\mathcal{L}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)} \right) \|x - \ell\|_{\mathbb{R}^d}^3, \\ \|D\|_{\mathbb{R}^d} &\leq c\|\varepsilon(x - \ell)\| \left(\sup_{t \in [0,1]} \|D_{\ell+t(x-\ell)}^2 f\|_{\mathcal{L}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)} \right) \|x - \ell\|_{\mathbb{R}^d}^2. \end{aligned}$$

Si on choisit $x \in \overline{B(\ell, \frac{\delta}{2})}$ on a $M := \left(\sup_{y \in B(\ell, \frac{\delta}{2})} \|D_y^2 f\|_{\mathcal{L}(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)} \right) < +\infty$ où on a utilisé le fait que sur $\overline{B(\ell, \frac{\delta}{2})}$ est fermé borné donc compact et que donc $D^2 f$ y atteint son maximum. Par conséquent, on a

$$\|B\|_{\mathbb{R}^d} \leq cM \frac{\delta}{2} \|x - \ell\|_{\mathbb{R}^d}^2, \quad \|C\|_{\mathbb{R}^d} \leq c\|\varepsilon(x - \ell)\| M \|x - \ell\|_{\mathbb{R}^d}^2. \quad (\text{A.2})$$

Comme $\|\rho\|_{\mathbb{R}^d}^{-1} \varepsilon(\rho) \rightarrow 0$ lorsque $\rho \rightarrow 0_{\mathbb{R}^d}$, quitte à diminuer δ , si $\|x - \ell\| \leq \frac{\delta}{2}$, on a en utilisant la définition de la limite avec $\varepsilon = 1$:

$$\|\varepsilon(x - \ell)\| \leq \|x - \ell\|_{\mathbb{R}^d}$$

et donc

$$\|C\|_{\mathbb{R}^d} \leq cM \|x_n - \ell\|_{\mathbb{R}^d}^2, \quad \|D\|_{\mathbb{R}^d} \leq cM \|x_n - \ell\|_{\mathbb{R}^d}^3 \leq cM \frac{\delta}{2} \|x_n - \ell\|_{\mathbb{R}^d}^2.$$

En particulier, si $x \in \overline{B(x, \frac{\delta}{2})}$ il existe $K > 0$ tel que

$$\|f(x) - \ell\|_{\mathbb{R}^d} \leq K \|x - \ell\|_{\mathbb{R}^d}^2. \quad (\text{A.3})$$

Maintenant, on démontre par récurrence que pour tout $n \geq 0$ on a

$$\|e_n\|_{\mathbb{R}^d} \leq K^{2^n-1} (\|e_0\|_{\mathbb{R}^d})^{2^n}$$

Initialisation : Pour $n = 0$, on a le résultat car $K^{2^n-1} (\|e_0\|_{\mathbb{R}^d})^{2^n} = K^0 \|e_0\|_{\mathbb{R}^d} = \|e_0\|_{\mathbb{R}^d} \geq \|e_0\|_{\mathbb{R}^d}$.

Hypothèse de récurrence : Soit $n \geq 1$, supposons que $\|e_n\|_{\mathbb{R}^d} \leq M^{2^n-1} (\|e_0\|_{\mathbb{R}^d})^{2^n}$. Hérédité : D'après (A.3) et l'hypothèse de récurrence, on a

$$\|e_{n+1}\|_{\mathbb{R}^d} \leq K \|e_n\|_{\mathbb{R}^d}^2 \leq K (K^{2^n-1} \|e_0\|_{\mathbb{R}^d}^{2^n})^2 = K K^{2^{n+1}-2} \|e_0\|_{\mathbb{R}^d}^{2^{n+1}} = K^{2^{n+1}-1} \|e_0\|_{\mathbb{R}^d}^{2^{n+1}}.$$

En particulier, quitte à diminuer la taille de $\delta > 0$, on peut choisir $K \|e_0\|_{\mathbb{R}^d} < 1$ et alors la convergence est quadratique.

Solution (de l'Exercice 22).

(i) Tout d'abord, on remarque que f est de classe $\mathcal{C}^\infty(\mathbb{R}^d)$ car pour tout $x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ on a

$$f(x) = \|Mx - b\|_{\ell^2(\mathbb{R}^q)}^2 = \sum_{j=1}^q ((Mx)_j - b_j)^2 = \sum_{j=1}^q \left(b_j - \sum_{k=1}^d m_{j,k} x_k \right)^2$$

et donc f est une fonction polynomiale à d -variables et est donc de classe $\mathcal{C}^\infty(\mathbb{R}^d)$. De plus, on a

$$\begin{aligned} f(x+h) - f(x) &= \langle M(x+h) - b, M(x+h) - b \rangle_{\mathbb{R}^q} - \langle Mx - b, Mx - b \rangle_{\mathbb{R}^q} \\ &= 2\langle Mx - b, Mh \rangle_{\mathbb{R}^q} + \|Mh\|_{\ell^2(\mathbb{R}^q)}^2 \\ &= 2\langle M^\top Mx - M^\top b, h \rangle_{\mathbb{R}^q} + \|Mh\|_{\ell^2(\mathbb{R}^q)}^2. \end{aligned}$$

Comme on a pour tout $h \in \mathbb{R}^d$

$$\|Mh\|_{\ell^2(\mathbb{R}^q)}^2 \leq \|M\|^2 \|h\|_{\ell^2(\mathbb{R}^d)}$$

et donc $D_x f(h) = 2\langle M^\top Mx - M^\top b, h \rangle_{\mathbb{R}^q}$ ce qui livre

$$\nabla f(x) = 2(M^\top Mx - M^\top b).$$

(ii) On a

$$x_{n+1} = x_n - \alpha(\nabla f)(x_n) = x_n - 2(M^\top Mx_n - M^\top b).$$

Comme un minimiseur $\ell \in \mathbb{R}^d$ vérifie l'équation normale $M^\top M\ell - M^\top b = 0$, on a

$$\begin{aligned} x_{n+1} - \ell &= x_n - \ell - 2\alpha(M^\top Mx_n - M^\top b) \\ &= x_n - \ell - 2\alpha(M^\top Mx_n - M^\top b) - 2\alpha(M^\top M\ell - M^\top b) \\ &= x_n - \ell - 2\alpha(M^\top M(x_n - \ell)) \\ &= (Id - 2\alpha M^\top M)(x_n - \ell). \end{aligned}$$

(iii) Comme $M^\top M$ est symétrique réelle et positive, elle est diagonalisable en base orthonormée, que l'on note $(e_1, \dots, e_d)$, et si $r = \text{rang}(M)$, on note ses valeurs propres strictement positive dans l'ordre décroissant par $\sigma_1^2 \geq \dots \geq \sigma_r^2$. On a donc qu'il existe $O \in \mathcal{O}_d(\mathbb{R})$ (le groupe orthogonal : $O^\top O = OO^\top = Id$) telle que

$$M^\top M = ODO^\top, \quad D := \text{diag}(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0) \in \mathcal{M}_{d,d}(\mathbb{R}).$$

Par ailleurs, on a

$$x_n - \ell = (Id - 2\alpha M^\top M)^n(x_{in} - \ell).$$

et comme

$$x_{in} - \ell = \sum_{j=1}^d \langle x_{in} - \ell, e_j \rangle_{\mathbb{R}^d} e_j$$

on obtient

$$x_n - \ell = \sum_{j=1}^r (1 - 2\alpha\sigma_j^2)^n \langle x_{in} - \ell, e_j \rangle_{\mathbb{R}^d} e_j + \sum_{j=r+1}^d \langle x_{in} - \ell, e_j \rangle_{\mathbb{R}^d} e_j$$

et la deuxième somme est nulle car $x_{in} - \ell \in \ker(M)^\top$ puisque $(e_{r+1}, \dots, e_d)$ est une base de $\ker(M^\top M) = \ker(M)$ et que $\text{vect}(e_1, \dots, e_r)$ est orthogonal à $\text{vect}(e_{r+1}, \dots, e_d)$. Par conséquent, on a

$$\begin{aligned} \|x_n - \ell\|_{\ell^2(\mathbb{R}^d)}^2 &= \sum_{j=1}^r (1 - 2\alpha\sigma_j^2)^{2n} |\langle x_{in} - \ell, e_j \rangle_{\mathbb{R}^d}|^2 \\ &\leq \left(\max_{j \in \{1, \dots, r\}} |1 - 2\alpha\sigma_j^2| \right)^2 \sum_{j=1}^r |\langle x_{in} - \ell, e_j \rangle_{\mathbb{R}^d}|^2 \\ &\leq \left(\max_{j \in \{1, \dots, r\}} |1 - 2\alpha\sigma_j^2| \right)^2 \sum_{j=1}^d |\langle x_{in} - \ell, e_j \rangle_{\mathbb{R}^d}|^2 \\ &= \left(\max_{j \in \{1, \dots, r\}} |1 - 2\alpha\sigma_j^2| \right)^2 \|x_0 - \ell\|_{\ell^2(\mathbb{R}^d)}^2 \end{aligned}$$

Maintenant, si $\alpha \in (0, \frac{1}{2\sigma_1^2})$ on a pour tout $j \in \{1, \dots, r\}$

$$0 < 1 - 2\alpha\sigma_1^2 \leq 1 - 2\alpha\sigma_j^2 \leq 1 - 2\alpha\sigma_r^2.$$

On en déduit que $\max_{j \in \{1, \dots, r\}} |1 - 2\alpha\sigma_j^2| = 1 - 2\alpha\sigma_r^2$ et donc que

$$\|x_{n+1} - \ell\|_{\ell^2(\mathbb{R}^d)} \leq 1 - 2\alpha\sigma_r^2 \|x_n - \ell\|_{\ell^2(\mathbb{R}^d)}$$

et une récurrence livre que pour tout $n \in \mathbb{N}$ $e_n \leq |1 - 2\alpha\sigma_r^2|^n e_0$ où on a posé l'erreur $e_n := \|x_n - \ell\|_{\ell^2(\mathbb{R}^d)}$.

On remarque que pour $\alpha \in]0, \frac{1}{2\sigma_1^2}[$ l'on a

$$0 < 1 - \alpha\sigma_r^2 < 1$$

et donc que la suite $(x_n)_{n \in \mathbb{N}}$ converge vers ℓ lorsque $n \rightarrow +\infty$.

Solution (de l'Exercice 23).

(i) Comme $f \in \mathcal{C}^2(\mathbb{R}^d)$, la formule de Taylor avec reste intégral livre que pour tout $x, y \in \mathbb{R}^d$ on a :

$$f(y) = f(x) + \langle (\nabla f)(x), y - x \rangle_{\mathbb{R}^d} + \int_0^1 (1-t) \langle (H_{(1-t)x+ty} f)(y-x), y-x \rangle_{\mathbb{R}^d} dt.$$

et donc pour tout $x, y \in \mathbb{R}^d$, on a

$$f(y) \geq f(x) + \langle (\nabla f)(x), y - x \rangle_{\mathbb{R}^d} + m \|y - x\|_{\ell^2(\mathbb{R}^d)}^2. \quad (\text{A.4})$$

D'une part, pour tout $x, y \in \mathbb{R}^d$, cette inégalité combinée à l'inégalité de Cauchy-Schwartz livre

$$f(y) \geq f(x) - \|\nabla f(x)\|_{\ell^2(\mathbb{R}^d)} \|y - x\|_{\ell^2(\mathbb{R}^d)} + m \|y - x\|_{\ell^2(\mathbb{R}^d)}^2.$$

En particulier pour $x \in \mathbb{R}^d$ fixé (par exemple $x = 0_{\mathbb{R}^d}$), on obtient que $\lim_{\|y\|_{\ell^2(\mathbb{R}^d)} \rightarrow +\infty} f(y) = +\infty$. En particulier, il existe $R > 0$ tel que pour tout $x \in \mathbb{R}^d \setminus B_f(0_{\mathbb{R}^d}, R)$ on ait $|f(x)| \geq |f(0_{\mathbb{R}^d})|$. On en conclut que

$$\inf_{x \in \mathbb{R}^d} f(x) = \inf_{x \in B_f(0_{\mathbb{R}^d}, R)} f(x).$$

Or f est continue sur le compact $B_f(0_{\mathbb{R}^d}, R)$ (compact car fermé borné) donc f y est bornée et atteint ses bornes. On note $\ell \in \mathbb{R}^d$ un tel minimiseur et $(\nabla f)(\ell) = 0$. Si $\ell' \in \mathbb{R}^d$ est aussi minimiseur de f alors d'après (A.4), on a

$$f(\ell) = f(\ell') \geq f(\ell) + m \|\ell' - \ell\|_{\ell^2(\mathbb{R}^2)}^2$$

et $\|\ell' - \ell\|_{\ell^2(\mathbb{R}^2)}^2 \leq 0$ et donc $\ell' = \ell$.

(ii) Soit $n \in \mathbb{N}$, on a

$$\begin{aligned} \|x_{n+1} - \ell\|_{\ell^2(\mathbb{R}^d)}^2 &= \|x_n - \ell - \alpha(\nabla f)(x_n)\|_{\ell^2(\mathbb{R}^d)}^2 \\ &= \|x_n - \ell - \alpha((\nabla f)(x_n) - (\nabla f)(\ell))\|_{\ell^2(\mathbb{R}^d)}^2 \\ &= \|x_n - \ell\|_{\ell^2(\mathbb{R}^d)}^2 + \alpha^2 \|(\nabla f)(x_n) - (\nabla f)(\ell)\|_{\ell^2(\mathbb{R}^d)}^2 \\ &\quad - 2\alpha \langle (\nabla f)(x_n) - (\nabla f)(\ell), x_n - \ell \rangle_{\mathbb{R}^d}. \end{aligned}$$

Maintenant, on utilise la formule de Taylor avec reste intégral suivante

$$\nabla f(x_n) = \nabla f(\ell) + \int_0^1 (H_{\ell+te_n} f) e_n dt.$$

En particulier, on a

$$\langle (\nabla f)(x_n) - (\nabla f)(\ell), x_n - \ell \rangle_{\mathbb{R}^d} = \int_0^1 \langle (H_{\ell+te_n} f) e_n, e_n \rangle_{\mathbb{R}^d} dt \geq m \|e_n\|_{\ell^2(\mathbb{R}^d)}^2.$$

Maintenant, comme pour tout $x \in \mathbb{R}^d$, $H_x f$ est symétrique par le théorème de Schwartz et définie positive d'après les hypothèses, on en déduit que $H_x f$ est diagonalisable en base orthonormée : il existe $O_x \in \mathcal{O}(\mathbb{R}^d)$ telle que

$$H_x f = O_x^\top D_x O_x$$

avec $D_x = \text{diag}(\lambda_1(x)^2, \dots, \lambda_d(x)^2)$ avec $\lambda_j(x) \in \mathbb{R}_+^*$. En posant

$$D_x^{\frac{1}{2}} = \text{diag}(\lambda_1(x), \dots, \lambda_d(x)), \quad (H_x f)^{\frac{1}{2}} := O_x^\top D_x^{\frac{1}{2}} O_x$$

on obtient que pour tout $x \in \mathbb{R}^d$, on a $((H_x f)^{\frac{1}{2}})^\top = (H_x f)^{\frac{1}{2}}$ et $(H_x f)^{\frac{1}{2}}(H_x f) = (H_x f)(H_x f)^{\frac{1}{2}}$. Ainsi, pour tout $v \in \mathbb{R}^d$ on a

$$\|(H_x f)v\|_{\ell^2(\mathbb{R}^d)}^2 = \langle (H_x f)(H_x f)^{\frac{1}{2}}v, (H_x f)^{\frac{1}{2}}v \rangle_{\mathbb{R}^d} \leq M \|(H_x f)^{\frac{1}{2}}v\|_{\ell^2(\mathbb{R}^d)}^2 = M \langle (H_x f)v, v \rangle_{\mathbb{R}^d} \leq M^2 \|v\|_{\ell^2(\mathbb{R}^d)}^2.$$

Ainsi, on a

$$\|(\nabla f)(x_n) - (\nabla f)(\ell)\|_{\ell^2(\mathbb{R}^d)}^2 \leq \left(\int_0^1 \|(H_{\ell+te_n} f)e_n\|_{\ell^2(\mathbb{R}^d)}^2 dt \right) \leq M^2 \|e_n\|^2.$$

On en déduit que

$$\|e_{n+1}\|_{\ell^2(\mathbb{R}^d)}^2 \leq \|e_n\|_{\ell^2(\mathbb{R}^d)}^2 (1 + M^2 \alpha^2 - 2\alpha m).$$

(iii) Une récurrence classique livre que pour tout $n \in \mathbb{N}$, on a

$$\|e_n\|_{\ell^2(\mathbb{R}^d)}^2 \leq (1 + M^2 \alpha^2 - 2\alpha m)^n \|e_0\|_{\ell^2(\mathbb{R}^d)}^2.$$

Pour obtenir de la convergence lorsque $n \rightarrow +\infty$, il suffit que $|1 + M^2 \alpha^2 - 2\alpha m| < 1$ c'est-à-dire

$$-1 < 1 + M^2 \alpha^2 - 2\alpha m < 1 \iff -2 < M^2 \alpha^2 - 2\alpha m < 0.$$

Comme $\alpha > 0$, l'inégalité de droite se ré-écrit comme $M^2 \alpha - 2m < 0$ c'est à dire $\alpha < \frac{2m}{M^2}$. Maintenant on étudie le signe du polynôme $\alpha \mapsto M^2 \alpha^2 - 2\alpha m + 2$ pour vérifier si l'inégalité de gauche est vérifiée. On a alors le discriminant qui vaut

$$\Delta = 4m^2 - 8M^2 = 4(m^2 - M^2) - 4M^2 < 0 \quad (\text{car } 0 < m < M).$$

On en déduit que $\alpha \mapsto M^2 \alpha^2 - 2\alpha m + 2$ ne change pas de signe sur $\mathbb{R}$ et est donc du signe de $M^2 > 0$ ce qui conclut la démonstration.

Solution (de l'Exercice 24). Soient $\tau \in [0, 1]$ et $t \in [a, b]$ tels que $(1 - \tau)a + \tau b \neq t$. Comme f est dérivable sur l'intervalle $[t, (1 - \tau)a + \tau b]$ si $t < (1 - \tau)a + \tau b$ ou $[(1 - \tau)a + \tau b, t]$ si $t > (1 - \tau)a + \tau b$ et dérivable sur l'intérieur de ces intervalles, la formule de Taylor-Lagrange nous dit qu'il existe $c_{t,\tau} \in]t, (1 - \tau)a + \tau b[$ (ou $](1 - \tau)a + \tau b, t[$) tel que

$$f(t) = f((1 - \tau)a + \tau b) + (t - (1 - \tau)a - \tau b)f'(c_{t,\tau}).$$

En particulier, on a

$$\left(\int_a^b f(t)dt \right) - f((1 - \tau)a + \tau b)(b - a) = \int_a^b (t - (1 - \tau)a - \tau b)f'(c_{t,\tau})dt$$

donc

$$\begin{aligned} \left| \left(\int_a^b f(t)dt \right) - f((1 - \tau)a + \tau b)(b - a) \right| &\leq \left(\sup_{t \in [a,b]} |f'(t)| \right) \left(\int_a^b (t - (1 - \tau)a - \tau b)dt \right) \\ &\leq \left(\sup_{t \in [a,b]} |f'(t)| \right) (b - a) \left(\int_a^b dt \right) \\ &= \left(\sup_{t \in [a,b]} |f'(t)| \right) \frac{(b - a)^2}{2}. \end{aligned}$$

Comme f est de classe $\mathcal{C}^1([a, b])$ alors $f' \in \mathcal{C}^0([a, b])$ et est bornée et atteint ses bornes. On en déduit qu'il existe $M > 0$ tel que

$$\left| \left(\int_a^b f(t)dt \right) - I_\tau(f) \right| = \left| \left(\int_a^b f(t)dt \right) - f((1 - \tau)a + \tau b)(b - a) \right| \leq M(b - a)^2.$$

Solution (de l'Exercice 25).

(i) Si $f \in \mathcal{C}^1([a, b])$ on a

$$\begin{aligned} \left(\int_a^b f(t) dt \right) - S_\tau(f) &= \left(\sum_{j=0}^N \int_{t_j}^{t_{j+1}} f(t) dt \right) - \sum_{j=0}^N (t_{j+1} - t_j) f((1-\tau)t_j + \tau t_{j+1}) \\ &= \sum_{j=0}^N \left(\int_{t_j}^{t_{j+1}} f(t) - f((1-\tau)t_j + \tau t_{j+1}) dt \right). \end{aligned}$$

En particulier, d'après l'inégalité des accroissements finis, on sait que

$$|f(t) - f((1-\tau)t_j + \tau t_{j+1})| \leq \left(\sup_{t \in [a, b]} |f'(t)| \right) |t - (1-\tau)t_j - \tau t_{j+1}|$$

et comme $f \in \mathcal{C}^1([a, b])$ on a $f' \in \mathcal{C}^0([a, b])$ et donc f' est bornée et atteint ses bornes sur le segment $[a, b]$.

En particulier, il existe $M > 0$ tel que

$$\begin{aligned} \left| \left(\int_a^b f(t) dt \right) - S_\tau(f) \right| &\leq \sum_{j=0}^N \left(\int_{t_j}^{t_{j+1}} |f(t) - f((1-\tau)t_j + \tau t_{j+1})| dt \right) \\ &\leq M \sum_{j=0}^N \left(\int_{t_j}^{t_{j+1}} |t - (1-\tau)t_j - \tau t_{j+1}| dt \right). \end{aligned}$$

Si $\tau = 0$, on a

$$\int_{t_j}^{t_{j+1}} |t - (1-\tau)t_j + \tau t_{j+1}| dt = \int_{t_j}^{t_{j+1}} (t - t_j) dt = \frac{(t_{j+1} - t_j)^2}{2}$$

et si $\tau = 1$, on a

$$\int_{t_j}^{t_{j+1}} |t - (1-\tau)t_j + \tau t_{j+1}| dt = \int_{t_j}^{t_{j+1}} (t_{j+1} - t) dt = \frac{(t_{j+1} - t_j)^2}{2}.$$

On en déduit que pour $\tau \in \{0, 1\}$, on a

$$\left| \left(\int_a^b f(t) dt \right) - S_\tau(f) \right| \leq \frac{M}{2} \sum_{j=0}^N (t_{j+1} - t_j)^2 = \frac{M}{2} \sum_{j=0}^N \left(\frac{b-a}{N+1} \right)^2 = \frac{M}{2(N+1)} (b-a)^2.$$

(ii) Soit $j \in \{0, \dots, N\}$ et $t \in]t_j, t_{j+1}[$. On pose $m_j := \frac{t_{j+1}+t_j}{2}$. Si $f \in \mathcal{C}^2([a, b])$ d'après la formule de Taylor-Lagrange, il existe $c_{t,j} \in]t, m_j[$ ou $]m_j, t[$ tel que

$$f(t) = f(m_j) + f'(m_j)(t - m_j) + \frac{1}{2} f''(c_{t,j})(t - m_j)^2$$

et donc

$$\begin{aligned} \left(\int_{t_j}^{t_{j+1}} f(t) dt \right) - f(m_j)(t_{j+1} - t_j) &= f'(m_j) \left(\int_{t_j}^{t_{j+1}} (t - m_j) dt \right) + \frac{1}{2} \left(\int_{t_j}^{t_{j+1}} f''(c_{t,j})(t - m_j)^2 dt \right) \\ &= \frac{1}{2} f'(m_j) ((t_{j+1} - m_j)^2 - (t_j - m_j)^2) + \frac{1}{2} \left(\int_{t_j}^{t_{j+1}} f''(c_{t,j})(t - m_j)^2 dt \right) \\ &= \frac{1}{2} \left(\int_{t_j}^{t_{j+1}} f''(c_{t,j})(t - m_j)^2 dt \right) \end{aligned}$$

car $t_{j+1} - m_j = \frac{t_{j+1}-t_j}{2}$ et $t_j - m_j = \frac{t_j-t_{j+1}}{2}$. On a donc

$$\begin{aligned} \left| \left(\int_{t_j}^{t_{j+1}} f(t) dt \right) - f(m_j)(t_{j+1} - t_j) \right| &\leq \frac{1}{6} \left(\sup_{t \in [a, b]} |f''(t)| \right) ((t_{j+1} - m_j)^3 - (t_j - m_j)^3) \\ &= \frac{1}{24} \left(\sup_{t \in [a, b]} |f''(t)| \right) (t_{j+1} - t_j)^3. \end{aligned}$$

Comme f'' est continue sur $[a, b]$, elle y est bornée et atteint ses bornes. En particulier, il existe $M > 0$ tel que On en déduit alors que

$$\left| \left(\int_a^b f(t) dt \right) - S_{\frac{1}{2}}(f) \right| \leq M \sum_{j=0}^N (t_{j+1} - t_j)^3 = M \frac{(b-a)^3}{(N+1)^2}.$$

Solution (de l'Exercice 26). Si $\varphi \in \mathcal{C}^2([0, T])$ pour tout $j \in \{0, \dots, N\}$, d'après le théorème de Taylor reste intégral, on a

$$\varphi(t_{j+1}) = \varphi(t_j) + \varphi'(t_j)h + \int_{t_j}^{t_{j+1}} (t_{j+1} - t)f''(t)dt = \varphi(t_j) + f(t_j, \varphi(t_j))h + \int_{t_j}^{t_{j+1}} (t_{j+1} - t)f''(t)dt$$

On en déduit que

$$e_j(\varphi) = \varphi(t_{j+1}) - (\varphi(t_j) + f(t_j, \varphi(t_j))h) = \int_{t_j}^{t_{j+1}} (t_{j+1} - t)f''(t)dt.$$

et donc comme $\varphi \in \mathcal{C}^2([0, T])$, φ'' est bornée et atteint ses bornes sur $[0, T]$, en particulier on a

$$\|e_j(\varphi)\|_{\mathbb{R}^d} \leq 2M \int_{t_j}^{t_{j+1}} (t_{j+1} - t)dt = Mh^2, \quad \text{avec} \quad M := \frac{1}{2} \left(\sup_{t \in [0, T]} \|f''(t)\|_{\mathbb{R}^d} \right).$$

On en déduit que

$$\mathcal{E}_N(\varphi) = \sum_{j=0}^N \|e_j(\varphi)\|_{\mathbb{R}^d} \leq Mh^2 \left(\sum_{j=0}^N 1 \right) = Mh^2(N+1) = M \frac{T^2}{N+1} = MTh.$$

On en déduit que la méthode d'Euler explicite est consistante d'ordre 1.

Solution (de l'Exercice 27).

(i) Soit $j \in \{0, \dots, N\}$, on a

$$\begin{aligned} \varphi_{j+1} - \tilde{\varphi}_{j+1} &= \varphi_j + hf(t_j, \varphi_j) - (\tilde{\varphi}_j + hf(t_j, \tilde{\varphi}_j) - \varepsilon_j) \\ &= (\varphi_j - \tilde{\varphi}_j) + h(f(t_j, \varphi_j) - f(t_j, \tilde{\varphi}_j)) - \varepsilon_j. \end{aligned}$$

On en déduit que

$$\begin{aligned} \|\varphi_{j+1} - \tilde{\varphi}_{j+1}\|_{\mathbb{R}^d} &\leq \|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} + h\|f(t_j, \varphi_j) - f(t_j, \tilde{\varphi}_j)\|_{\mathbb{R}^d} + \|\varepsilon_j\|_{\mathbb{R}^d} \\ &\leq \|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} + Lh\|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} + \|\varepsilon_j\|_{\mathbb{R}^d} \\ &= (1 + Lh)\|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} + \|\varepsilon_j\|_{\mathbb{R}^d} \end{aligned}$$

où on a utilisé l'hypothèse que f est globalement Lipschitzienne par rapport à sa seconde variable.

(ii) Démontrons le résultat par récurrence.

Initialisation : Pour $j = 1$, on a

$$\begin{aligned} \|\varphi_1 - \tilde{\varphi}_1\|_{\mathbb{R}^d} &\leq (1 + Lh)\|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \|\varepsilon_0\|_{\mathbb{R}^d} \\ &\leq e^{Lh}\|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \|\varepsilon_0\|_{\mathbb{R}^d} \end{aligned}$$

où on a utilisé l'inégalité³ qui dit que pour tout $x \in \mathbb{R}$, on a $1 + x \leq e^x$. La propriété est donc vraie pour $j = 1$.

Hypothèse de récurrence : Soit $j \in \{1, \dots, N\}$, on suppose que l'on a

$$\|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} \leq e^{jLh}\|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \sum_{p=0}^{j-1} e^{Lh(j-1-p)}\|\varepsilon_p\|_{\mathbb{R}^d}.$$

3. On peut le justifier en étudiant la fonction $(x \mapsto e^x - x - 1)$ ou bien en se souvenant que la fonction exponentielle est convexe donc est au dessus de toute ses tangentes, en particulier celle en 0.

Hérédité : On a

$$\begin{aligned}
\|\varphi_{j+1} - \tilde{\varphi}_{j+1}\|_{\mathbb{R}^d} &\leq (1 + Lh)\|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} + \|\varepsilon_j\|_{\mathbb{R}^d} \\
&\leq (1 + Lh) \left(e^{jLh} \|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \sum_{p=0}^{j-1} e^{Lh(j-1-p)} \|\varepsilon_p\|_{\mathbb{R}^d} \right) + \|\varepsilon_j\|_{\mathbb{R}^d} \\
&\leq e^{Lh} \left(e^{jLh} \|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \sum_{p=0}^{j-1} e^{Lh(j-1-p)} \|\varepsilon_p\|_{\mathbb{R}^d} \right) + \|\varepsilon_j\|_{\mathbb{R}^d} \\
&\leq e^{(j+1)Lh} \|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \left(\sum_{p=0}^{j-1} e^{Lh(j-p)} \|\varepsilon_p\|_{\mathbb{R}^d} \right) + \|\varepsilon_j\|_{\mathbb{R}^d} \\
&= e^{(j+1)Lh} \|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \left(\sum_{p=0}^j e^{Lh(j-p)} \|\varepsilon_p\|_{\mathbb{R}^d} \right)
\end{aligned}$$

ce qui est précisément ce qu'il fallait démontrer.

(iii) On en déduit que pour tout $j \in \{1, \dots, N+1\}$ on a

$$\begin{aligned}
\|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} &\leq e^{(N+1)Lh} \|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + e^{(j-1)Lh} \sum_{p=0}^{j-1} e^{-pLh} \|\varepsilon_p\|_{\mathbb{R}^d} \\
&\leq e^{(N+1)Lh} \left(\|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \sum_{p=0}^N \|\varepsilon_p\|_{\mathbb{R}^d} \right) \\
&= e^{TL} \left(\|\varphi_0 - \tilde{\varphi}_0\|_{\mathbb{R}^d} + \sum_{p=0}^N \|\varepsilon_p\|_{\mathbb{R}^d} \right)
\end{aligned}$$

ce qui démontre que le schéma d'Euler explicite est stable.

Solution (de l'Exercice 28). Soit $N \in N^*$ on remarque que pour tout $j \in \{0, \dots, N\}$, par définition, on a

$$\varphi(t_{j+1}) = \varphi(t_j) + hf(t_j, \varphi(t_j)) + e_j(\varphi)$$

En posant $\tilde{\varphi}_j = \varphi(t_{j+1})$ et en utilisant la stabilité de la méthode d'Euler explicite, on obtient qu'il existe $S > 0$ tel que pour tout $j \in \{1, \dots, N+1\}$

$$\|\varphi_j - \varphi(t_j)\|_{\mathbb{R}^d} = \|\varphi_j - \tilde{\varphi}_j\|_{\mathbb{R}^d} \leq S \left(\|\varphi_0 - \varphi(0)\|_{\mathbb{R}^d} + \sum_{p=0}^N \|e_p(\varphi)\| \right) = S\mathcal{E}_N(\varphi)$$

où on a utilisé le fait que $\varphi_0 = \varphi(0) = y_0$ et la définition de l'erreur globale de consistance. On sait que le schéma d'euler explicite est consistant d'ordre 1 ce qui livre l'existence de $C > 0$ telle que

$$\|\varphi_j - \varphi(t_j)\|_{\mathbb{R}^d} \leq Ch = \frac{CT}{N+1}$$

et qui conclut la démonstration.