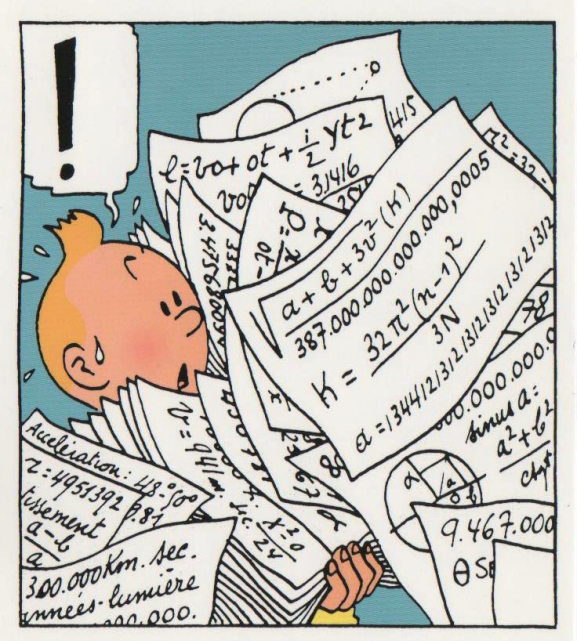
**Université Claude Bernard, Lyon I**
**43, boulevard 11 novembre 1918**
**69622 Villeurbanne cedex, France**

Licence Sciences, Technologies & Santé
Spécialité Mathématiques
L. Pujo-Menjouet
pujo@math.univ-lyon1.fr

# Analyse numérique
## Troisième année de licence




Université Claude Bernard Lyon 1

# Préambule

L'analyse numérique a commencé bien avant la conception des ordinateurs et leur utilisation quotidienne que nous connaissons aujourd'hui. Les premières méthodes ont été développées pour essayer de trouver des moyens rapides et efficaces de s'attaquer à des problèmes soit fastidieux à résoudre à cause de leur grande dimension (systèmes à plusieurs dizaines d'équations par exemple), soit parce qu'il n'existe pas solutions explicites connues même pour certaines équations assez simples en apparence.
Dès que les premiers ordinateurs sont apparus, ce domaine des mathématiques a pris son envol et continue encore à se développer de façon très soutenue.
Les applications extraordinairement nombreuses sont entrées dans notre vie quotidienne directement ou indirectement. Nous les utilisons désormais sans nous en rendre compte mais surtout en ignorant la plupart du temps toute la théorie, l'expertise, le développement des compétences et l'ingéniosité des chercheurs pour en arriver là. Nous pouvons téléphoner, communiquer par satellite, faire des recherches sur internet, regarder des films où plus rien n'est réel sur l'écran, améliorer la sécurité des voitures, des trains, des avions, connaître le temps qu'il fera une semaine à l'avance,...et ce n'est qu'une infime partie de ce que l'on peut faire.
Le but de ce cours et s'initier aux bases de l'analyse numérique en espérant qu'elles éveilleront de l'intérêt, de la curiosité et pourquoi pas une vocation.

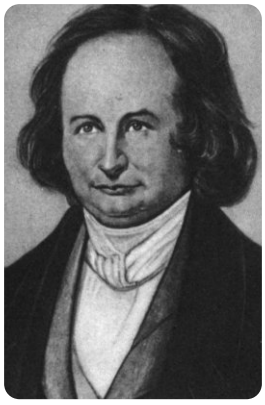
FIGURE 1 – *Entre le Tintin dessiné à la main dans les années 6 par Hergé et celui mis à l'écran par Spielberg, un monde numérique les sépare. Un peu comme les premiers développements à la main des pionniers du numérique et les effets dernier cri des plus puissants des ordinateurs.*

# Table des matières

# Chapitre 1

# Les systèmes linéaires

(a) Johann Carl Friedrich Gauss (1777- 1855), mathématicien, astronome et physicien allemand. Considéré comme l'un des plus grands mathématiciens de tous les temps, nous utiliserons ici une méthode numérique qui porte son nom.



(b) Charles Gustave Jacob Jacobi (1804-1851), un mathématicien allemand surtout connu pour ses travaux sur les intégrales elliptiques, les équations aux dérivées partielles et leur application à la mécanique analytique. On lui doit une méthode que l'on présentera ici

(c) André-Louis Cholesky (1875-1918 ), polytechnicien et officier français, ingénieur topographe et géodésien. On lui doit une méthode que l'on présentera ici

FIGURE 1.1 – Quelques mathématiciens célèbres liés à l'étude des nombres entiers, rationnels et réels.

## 1.1 Introduction

### 1.1.1 Gestion des erreurs

Il arrive souvent lors de calcul que nous soyons obligés de donner une valeur approchée de la solution (0,333 pour 1/3 par exemple), c'est ce que l'on appelle la représentation des nombres

en virgule flottante (8 ou 16 chiffres significatifs après la virgule). Il se peut également qu'une méthode employée pour résoudre un problème ne nous amène pas exactement à la solution mais à côté, on dit alors que l'on a une approximation de la solution (c'est souvent le cas pour les problèmes non linéaires que nous verrons dans le chapitre suivant et moins fréquent pour les problèmes linéaires). Le but est de pouvoir maîtriser ces erreurs de sorte qu'elles ne s'accumulent pas au cours des opérations successives. C'est un des chevaux de bataille des numériciens qui ne souhaitent pas voir leurs approximation s'éloigner des solutions exactes, ce qui rendrait leur travail complètement inefficace.
L'objectif est donc de chercher à élaborer des méthodes numériques qui n'amplifient pas trop les erreurs d'arrondi au cours des calculs. Intuitivement, nous imaginons bien que cette amplification est d'autant plus grande que le nombre d'opérations est important (additions, multiplications, etc.).Nous cherchons donc à réduire le nombre d'opérations au maximum et par voie de conséquence le temps de calcul.
A titre d'exemple, pour calculer le déterminant d'une matrice carrée $24 \times 24$ par la méthode classique de Cramer il faudrait plus de 20 ans à un ordinateur capable de faire $10^{15}$ opérations à la seconde (un Peta-flops) ! On voit donc qu'il est plus que nécessaire de trouver des méthodes beaucoup plus efficaces que celle-là.
Nous introduirons donc les notions de stabilité et d'efficacité. Et suivant les méthodes utilisées, nous aurons affaire à tel ou tel type d'erreur. Pour les systèmes linéaires par exemple nous étudierons deux types de méthodes bien distinctes :

1. **les méthodes directes** : qui consistent en des algorithmes qui permettent de calculer la solution exacte (mis à part les erreurs d'arrondis en virgule flottante) en un nombre fini d'opération,
2. **les méthodes itératives** : qui consistent en l'approche de la solution exacte $x$ par une suite $(x_n)_{n\in\mathbb{N}}$ de solutions approchées. Nous arrêtons alors les calculs lorsque nous atteignons un certain $x_{n_0}$. Nous avons alors une erreur de troncature due à l'approximation de la solution exacte que nous mesurons à l'aide d'une norme appropriée (suivant la dimension de l'espace dans lequel $x$ se trouve) $\|x - x_{n_0}\|$.

### 1.1.2 Exemple de problème menant à la résolution d'un système linéaire

On considère l'équation différentielle du second ordre posée sur l'intervalle $[0, 1]$,

$$-u''(x) + c(x)u(x) = f(x) \text{ pour tout } x \in ]0, 1[,$$

définie en $0$ et $1$ par

$$u(0) = \alpha \text{ et } u(1) = \beta,$$

où $\alpha$ et $\beta$ sont des réels donnés, $f : [0, 1] \to \mathbb{R}$ est continue et donnée, et $c : [0, 1] \to \mathbb{R}^+$ est une fonction continue positive donnée.
L'inconnue est ici la fonction $u : [0, 1] \to \mathbb{R}$ qui est de classe $\mathscr{C}^2$.
On discrétise cette équation par ce que l'on appelle la méthode des différences finies (voir chapitre sur les équations différentielles). On se donne un nombre fini de points dans $[0, 1]$ espacés à égale distance $h$. Nous avons alors pour tout $n \in \mathbb{N}^*$, $h = \dfrac{1}{n}$ que l'on appelle le "pas du maillage".

On note pour tout $j = 0, ..., n$, $x_j = \dfrac{j}{n}$.
Nous allons essayer alors de calculer une valeur approchée de la solution $u$ aux points $x_0$,...$x_n$. Nous passons donc d'un problème continu à un problème discret.
Pour tout $i = 0, ..., n$ on pose $c_i = c(x_i)$, $f_i = f(x_i)$ et $u_i$ la valeur approchée des $u(x_i)$ calculée par le schéma numérique.
Le problème est le suivant : comment approcher $u''(x_i)$ ?
Une des méthodes consiste d'abord à estimer :

1. **la dérivée première** :
   on approche $u'(x_i)$ par un taux d'accroissement. Et on a le choix :
   -soit
   $$u'(x_i) \simeq \frac{u(x_i) - u(x_{i-1})}{x_i - x_{i-1}} \simeq \frac{u_i - u_{i-1}}{h},$$
   -soit
   $$u'(x_i) \simeq \frac{u(x_{i+1}) - u(x_i)}{x_{i+1} - x_i} \simeq \frac{u_{i+1} - u_i}{h}.$$
   Tout le problème quand nous construisons une telle méthode est de savoir comment contrôler le "$\simeq$". Nous voulons que la solution approchée $u_i$ soit proche de la solution exacte $u(x_i)$. Une notion que nous utiliserons pour cela sera la consistance (voir chapitre sur les équations différentielles).
   Si on considère la seconde méthode, comme $u \in \mathscr{C}^2$, nous utilisons une développement de Taylor à l'ordre 1 entre $x_i$ et $x_{i+1}$,
   $$u(x_{i+1}) = u(x_i) + hu'(x_i) + \frac{h^2}{2}u''(\theta_i), \text{ où } \theta_i \in ]x_i, x_{i+1}[.$$
   Ce qui nous donne par un calcul simple :
   $$\frac{u(x_{i+1}) - u(x_i)}{h} - u'(x_i) = \frac{h}{2}u''(\theta_i).$$
   Nous voyons donc que la différence entre la valeur exacte de la dérivée et son approximation par le taux d'accroissement est égal à $\dfrac{h}{2}u''(\theta_i)$, c'est à dire de l'ordre de $h$. Nous choisissons alors $h$ très petit (c'est à dire le nombre $n$ de points entre 0 et 1 très grand).
2. **la dérivée seconde** :
   on considère la seconde méthode ci-dessus, et on fait applque Taylor à l'ordre 3 (en supposant que $u$ soit dérivable jusqu'à l'ordre 4 sur [0,1])
   -entre $x_i$ et $x_{i+1}$ :
   $$u(x_{i+1}) = u(x_i) + hu'(x_i) + \frac{h^2}{2}u''(x_i) + \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\theta_{1i}),$$
   -entre $x_{i-1}$ et $x_i$ :
   $$u(x_{i-1}) = u(x_i) - hu'(x_i) + \frac{h^2}{2}u''(x_i) - \frac{h^3}{6}u'''(x_i) + \frac{h^4}{24}u^{(4)}(\theta_{2i}),$$
   où $\theta_{1i} \in ]x_i, x_{i+1}[$ et $\theta_{2i} \in ]x_{i-1}, x_i[$.
   En faisant la somme nous avons

$$u''(x_i) = \frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{h^2} - \frac{h^4}{24}(u^{(4)}(\theta_{1i}) + u^{(4)}(\theta_{2i})),$$

ce qui nous permet d'écrire

$$u''(x_i) \simeq \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2}$$

et l'erreur locale est donc de l'ordre de $h^2$.

Si l'on pose $u(0) = u_0$ et $u(1) = u_n$, le schéma ainsi obtenu est finalement :

$$\begin{cases} \dfrac{-u_{i+1} + 2u_i - u_{i-1}}{h^2} + c_i u_i &= f_i, \quad \text{pour tout } i = 1, ..., n-1, \\ u_0 &= \alpha, \\ u_n &= \beta. \end{cases}$$

Nous avons bien $n+1$ équations à $n+1$ inconnues, que nous pouvons écrire sous forme matricielle

$$(A_n + C_n)U_n = b_n, \tag{1.1}$$

où

$$A_n = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}, \quad C_n = \begin{pmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_{n-1} \end{pmatrix}, \quad b_n \begin{pmatrix} f_1 + \dfrac{\alpha}{h^2} \\ f_2 \\ \vdots \\ f_{n-2} \\ f_{n_1} + \dfrac{\beta}{h^2} \end{pmatrix}$$

$$\text{et} \quad U_n = \begin{pmatrix} u_1 \\ \vdots \\ u_{n-1} \end{pmatrix}.$$

Notons qu'étant donné que la matrice $A_n + C_n$ est symétrique définie positive, alors elle est inversible et le système (1.1) possède une unique solution.

Une question que nous pouvons nous poser est la suivante :

-Comment évaluer l'erreur ? Si pour tout $i = 1, ..., n$, nous avons $u(x_i)$ la solution exacte et $u_i$ la solution approchée, nous pouvons montrer que l'erreur notée

$$\max_{0 \leq i \leq n} |u(x_i) - u_i|,$$

est de l'ordre de $h^2$.

**Remarque** *Nous pouvons faire plusieurs remarques à ce stade du cours :*

1. *le système linéaire (1.1) peut devenir très grand, et donc difficile voire impossible à résoudre à la main,*
2. *pour tracer une courbe régulière reliant les valeurs discrètes approchées $u_0$, $u_1$, ..., $u_n$ des solutions nous devrons utiliser une méthode appelée méthode d'interpolation,*

3. *enfin, pour montrer l'existence et l'unicité de la solution de ce système nous avons utilisé un résultat de cours d'algèbre sur les matrices (matrice symétrique définie positive). Etant donné que nous souhaitons que ce cours soit le plus autonome possible, nous allons faire un rappel (non exhaustif) de plusieurs résultats sur les matrices. Pour plus de détails, nous vous conseillons de vous replonger dans votre cours d'algèbre linéaire.*

## 1.2 Quelques rappels sur les matrices

Dans toute la suite, $\mathbb{K}$ désignera le corps des scalaires qui vaudra soit $\mathbb{R}$ pour les réels, soit $\mathbb{C}$ pour les complexes.

### 1.2.1 Notations

Soient $m, n \in \mathbb{N}^*$. Une matrice de type $(m, n)$ sur $\mathbb{K}$ est un tableau de scalaires (réels ou complexes) à à $m$ lignes et $n$ colonnes

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Nous notons $\mathscr{M}_{mn}(\mathbb{K})$ l'ensemble des matrices $(m, n)$ sur $\mathbb{K}$.
Si $m = n$, nous notons juste $\mathscr{M}_n(\mathbb{K})$ l'ensemble des matrices carrées $(n, n)$ sur $\mathbb{K}$.
De plus, quand aucune autre notation ne sera précisée, nous noterons $A_{ij}$ le coefficient de la matrice $A$ à l'intersection de la ligne $i$ et de la colonne $j$.

### 1.2.2 Lien avec les applications linéaires

#### Vecteurs

Soit $E$ un espace vectoriel sur $\mathbb{K}$ de dimension $n \in \mathbb{N}^*$ (appelé également $K$-espace vectoriel). Soit $\mathscr{B} = (e_j)_{1 \leq j \leq n}$ une base de $E$. Tout vecteur $x$ de $E$ admet une décomposition unique

$$x = \sum_{j=1}^{n} x_j e_j,$$

où les scalaires $x_j$ sont les composantes de $x$ dans la base $\mathscr{B}$.
Lorsqu'une base est fixée, on peut identifier $E$ à $\mathbb{K}^n$. On notera alors $x = (x_j)_{1 \leq j \leq n}$.
En notation matricielle, le vecteur $x$ sera toujours représenté par le vecteur colonne

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Avec nos notations, $x$ est une matrice de type $(n, 1)$.

**Représentation matricielle d'une application linéaire**

Soient $E$ et $F$ deux $\mathbb{K}$-espaces vectoriels de dimension finie avec $\dim E = n$ et $\dim F = m$. Nous fixons deux bases :

$$\mathscr{B} = (e_j)_{1\leq j\leq n} \text{ une base de } E \text{ et } \mathscr{B}' = (f_j)_{1\leq j\leq m} \text{ une base de } F.$$

Soit $u : E \to F$ une application linéaire. Tout $x \in E$ s'écrit alors

$$x = \sum_{j=1}^{n} x_j e_j,$$

et donc par linéarité,

$$u(x) = \sum_{j=1}^{n} x_j u(e_j).$$

Pour chaque indice $j$, $u(e_j)$ est un vecteur de $F$ qui se décompose dans la base $\mathscr{B}'$, autrement dit il existe des scalaires $(a_{ij})_{1\leq i,j\leq m}$ tels que

$$u(e_j) = \sum_{i=1}^{m} a_{ij} f_i.$$

On a alors

$$u(x) = \sum_{j=1}^{n}\sum_{i=1}^{m} x_j a_{ij} f_i.$$

Ainsi, relativement aux bases $\mathscr{B}$ et $\mathscr{B}'$, l'application linéaire $u$ est représentée par la matrice

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

De telle sorte que les composantes de $u(x)$ dans $\mathscr{B}'$ s'obtiennent en faisant le produit de la matrice $A$ par le vecteur colonne des composantes de $x$ dans $\mathscr{B}$ (voir la section (1.2.3) pour le calcul du produit de deux matrices). On a alors

$$\begin{pmatrix} (u(x))_1 \\ (u(x))_2 \\ \vdots \\ (u(x))_m \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Le $j$ème vecteur colonne de $A$ que l'on écrit

$$\begin{pmatrix} a_{1j} \\ a_{3j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

représente le vecteur $u(e_j)$ dans la base de $\mathscr{B}'$.
inversement, à une matrice donnée $A \in \mathscr{M}_{mn}(\mathbb{K})$, on peut associer une application linéaire canonique $u : \mathbb{K}^n \to \mathbb{K}^m$ de matrice $A$ associée dans les bases canoniques de $\mathbb{K}^n$ et $\mathbb{K}^m$.

**Noyau, image et rang**

Soit $A \in \mathscr{M}_{mn}(\mathbb{K})$. Le noyau, l'image et le rang sont définis comme étant le noyau, l'image et le rang de l'application linéaire $u$ canonique associée à $A$.
Autrement dit l'image de $A$ est définie par

$$\text{Im}(A) = \{y \in \mathbb{K}^m, \text{ il existe } x \in \mathbb{K}^n, Ax = y\},$$

et le noyau de $A$ par

$$\ker(A) = \{x \in \mathbb{K}^n, Ax = 0\}.$$

Enfin, le rang de $A$ noté $\text{rg}(A)$ est défini par $\dim(\text{Im}(A))$.
On rappelle que

$$u \text{ est surjective si et seulement si } \text{rg}(A) = m,$$

$$u \text{ injective si et seulement si } \ker(A) = \{0\},$$

$$\text{rg}(A) + \dim(\ker(A)) = n.$$

### 1.2.3 Opérations

**Somme**

Nous pouvons ajouter deux matrices de mêmes dimensions $(m, n)$, et nous avons naturellement

$$(A + B)_{ij} = A_{ij} + B_{ij} \text{ pour tous } i = 1, ..., m \text{ et } j = 1, ..., n.$$

**Multiplication par un scalaire**

Pour $\alpha \in \mathbb{K}$ et $A$ une matrice $(m, n)$, le produit $\alpha A$ est une matrice $(m, n)$ et

$$(\alpha A)_{ij} = \alpha A_{ij} \text{ pour tous } i = 1, ..., m \text{ et } j = 1, ..., n.$$

**Produit de deux matrices**

Si $A \in \mathscr{M}_{mn}(\mathbb{K})$ et $B \in \mathscr{M}_{np}(\mathbb{K})$, nous pouvons faire le produit $AB$, il s'agit d'une matrice de $\mathscr{M}_{mp}(\mathbb{K})$, et

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} \text{ pour tous } i = 1, ..., m \text{ et } j = 1, ..., p.$$

**Remarque** *Si $u : \mathbb{K}^n \to \mathbb{K}^m$ (respectivement Si $u : \mathbb{K}^p \to \mathbb{K}^n$) est l'application linéaire canonique associée à $A$ (respectivement $B$), alors $AB$ est la matrice dans les bases canoniques de $u \circ v : \mathbb{K}^p \to \mathbb{K}^m$.*

### Transposée

Si $A \in \mathscr{M}_{mn}(\mathbb{K})$, la transposée de $A$ notée $A^T$, est la matrice de $\mathscr{M}_{nm}(\mathbb{K})$ définie par

$$(A^T)_{ij} = A_{ji} \text{ pour tous } i = 1, ..., n \text{ et } j = 1, ..., m.$$

Si $A \in \mathscr{M}_{mn}(\mathbb{R})$, nous avons les résultats suivants :

1. $(A^T)^T = A$,
2. $(A + B)^T = A^T + B^T$,
3. $(\lambda A)^T = \lambda A^T$, $\lambda \in \mathbb{R}$,
4. $(AB)^T = B^T A^T$,
5. $(A^{-1})^T = (A^T)^{-1}$.

### Matrice adjointe

Si $A \in \mathscr{M}_{mn}(\mathbb{K})$, la matrice adjointe de $A$, notée $A^*$, est la matrice de $\mathscr{M}_{nm}(\mathbb{K})$ définie par

$$(A^*)_{ij} = \overline{A_{ji}} \text{ pour tous } i = 1, ..., n \text{ et } j = 1, ..., m.$$

**Remarque** *Notons que la matrice adjointe d'une matrice $A$ à coefficients complexes est la matrice transposée de la matrice conjuguée de $A$, autrement dit*

$$A^* = \overline{A^T}.$$

Si $A \in \mathscr{M}_{mn}(\mathbb{R})$, nous avons les résultats suivants :

1. $(A^*)^* = A$,
2. $(A + B)^* = A^* + B^*$,
3. $(\lambda A)^* = \overline{\lambda} A^*$, $\lambda \in \mathbb{C}$,
4. $(AB)^* = B^* A^*$,
5. $(A^{-1})^* = (A^*)^{-1}$.

### Matrices particulières

Une matrice carrée $A \in \mathscr{M}_n(\mathbb{K})$ de coefficients $(a_{ij}) \in \mathbb{K}$ pour tous $i, j = 1, ..., n$. est

1. diagonale : si $a_i j = 0$ pour tout $i \neq j$. On note en général $A = diag(a_{11}, a_{22}, ..., a_{nn})$,
2. une matrice diagonale de taille $n$ particulière est la matrice identité $I_n$ dont tous les coefficients valent 1, autrement dit $I_n = diag(1, 1, ..., 1)$,
3. triangulaire supérieure si $a_{ij} = 0$ pour tout $i > j$,
4. triangulaire inférieure si $a_{ij} = 0$ pour tout $i < j$,
5. symétrique si $A$ est réelle et si $A = A^T$,
6. hermitienne si $A$ est complexe et si $A = A^*$,
7. orthogonale si $A$ est réelle et si $AA^T = A^T A = I_n$,
8. unitaire si $A$ est complexe et si $AA^* = A^* A = I_n$,
9. normale si $AA^* = A^* A$.

**Remarque** *Une matrice à éléments réels est hermitienne si et seulement si elle est symétrique*

### 1.2.4 Trace et déterminant

#### Trace

Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice carrée de coefficients $(a_{ij}) \in \mathbb{K}$ pour tous $i, j = 1, ..., n$, sa trace est la somme de ses termes diagonaux

$$\text{Tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

On a les propriétés suivantes

$$\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B), \text{Tr}(AB) = \text{Tr}(BA).$$

#### Déterminant

Le déterminant d'une matrice carrée $(n, n)$ est noté $\det(A)$ ou encore

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{vmatrix}.$$

Il est défini par la formule suivante

$$\det A = \sum_{\sigma \in \mathscr{S}_n} \varepsilon(\sigma) a_{1\sigma(1)} ... a_{n\sigma(n)}$$

où $\mathscr{S}_n$ est le groupe des permutations, c'est à dire des bijections de $\{1, ..., n\}$ dans lui-même, et pour tout $\sigma \in \mathscr{S}_n$, $\varepsilon(\sigma)$ est la signature de $\sigma$.
Les déterminants possèdent les propriétés suivantes :

1. $\det I_n = 1$,
2. $\det A^T = \det A$,
3. $\det A^* = \overline{\det A}$,
4. $\det(AB) = \det A \det B$,
5. pour tout scalaire $\alpha \in \mathbb{K}$, $\det(\alpha A) = \alpha^n \det A$,
6. le déterminant d'une matrice triangulaire (*a fortiori* diagonale) est égal au produit de ses termes diagonaux.

#### Inverse

On dit que la matrice carrée $A$ de taille $n$ est inversible s'il existe une matrice $B$ de taille $n$ telle que $AB = BA = I_n$.
La matrice $B$ est appelée inverse de $A$ et notée $A^{-1}$.

**Inverse et opérations**

Nous avons les résultats suivants sur les inverses :

$$(AB)^{-1} = B^{-1}A^{-1},\ (A^T)^{-1} = (A^{-1})^T,\ (A^*)^{-1} = (A^{-1})^*.$$

**Caractérisation d'une matrice inversible**

**Théorème 1** (MATRICE INVERSIBLE)

Soit $A \in \mathscr{M}_n(\mathbb{K})$. Les propositions suivantes sont équivalentes

1. $A$ est inversible,
2. $\det A \neq 0$,
3. $Ax = 0$ a pour seule solution $x = 0$,
4. pour tout $b \in \mathbb{K}^n$, $Ax = b$ possède une unique solution.

### 1.2.5 Matrice et produit scalaire

Dans cette section, nous travaillerons dans $\mathbb{K}^n$, mais toutes les définitions se généralisent à un espace vectoriel sur $\mathbb{K}$ de dimension $n$, une fois fixée une base.
L'application $< .,. >: \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ définie par

$$\begin{aligned} < x, y > &= y^T x = x^T y = \sum_{j=1}^{n} x_j y_j \quad \text{si } \mathbb{K} = \mathbb{R}, \\ < x, y > &= y^* x = \overline{x^* y} = \sum_{j=1}^{n} x_j \overline{y}_j \quad \text{si } \mathbb{K} = \mathbb{C}, \end{aligned}$$

est appelé produit scalaire euclidien sur $\mathbb{R}^n$ si $\mathbb{K} = \mathbb{R}$, hermitien sur $\mathbb{C}^n$ si $\mathbb{K} = C$, ou encore produit scalaire canonique si l'on ne précise pas le corps.
Deux vecteurs $x$ et $y$ de $\mathbb{K}^n$ sont orthogonaux si

$$< x, y >= 0.$$

Une famille de vecteurs de $\mathbb{K}^n$, $(x^1, ..., x^p)$ est orthonormale si pour tous $1 \leq j, k \leq p$,

$$< x^j, x^k >= \delta_{jk} = \begin{cases} 1 & si \quad j = k, \\ 0 & si \quad j \neq k. \end{cases}$$

**Produit scalaire et transposée ou adjointe**

1. Cas réel :
   pour toute matrice $A \in \mathscr{M}_n(\mathbb{R})$, tous vecteurs $x, y \in \mathbb{R}^n$,

$$< Ax, y >=< x, A^T y >.$$

Si $A$ est orthogonale, elle "conserve" le produit scalaire euclidien : autrement dit pour tous vecteurs $x, y \in \mathbb{R}^n$,

$$< Ax, y >=< x, A^T Ay >=< x, y >.$$

2. Cas complexe :
de même, pour toute matrice $A \in \mathscr{M}_n(\mathbb{C})$, tous vecteurs $x, y \in \mathbb{C}^n$,

$$< Ax, y >=< x, A^* y >.$$

Si $A$ est unitaire, elle "conserve" le produit scalaire hermitien : autrement dit pour tous vecteurs $x, y \in \mathbb{C}^n$,

$$< Ax, y >=< x, A^* Ay >=< x, y >.$$

**Matrices positives, définitions**

1. Cas réel :
soit $A$ une matrice symétrique réelle dans $\mathscr{M}_n(\mathbb{R})$. On peut lui associer une forme quadratique sur $\mathbb{R}^n$

$$x \mapsto < Ax, x >= x^T Ax = \sum_{i,j=1}^{n} a_{ij} x_i x_j.$$

Nous dirons que $A$ est une matrice définie positive si cette forme quadratique vérifie

$$< Ax, x >> 0 \text{ pour tout } x \in \mathbb{R}^n,\ x \neq 0,$$

et nous dirons que $A$ est une matrice positive si elle vérifie "seulement"

$$< Ax, x >\geq 0 \text{ pour tout } x \in \mathbb{R}^n.$$

2. Cas complexe :
Soit $A$ une matrice hermitienne dans $\mathscr{M}_n(\mathbb{C})$. On peut lui associer une forme hermitienne sur $\mathbb{C}^n$

$$x \mapsto < Ax, x >= \overline{x}^T Ax = \sum_{i,j=1}^{n} a_{ij} \overline{x}_i x_j.$$

Nous remarquerons que cette quantité est toujours réelle. Il est assez facile de montrer en effet, en utilisant la définition du produit scalaire et en l'appliquant à $< Ax, x >$ que l'on a l'égalité :

$$< Ax, x >= \overline{< Ax, x >}.$$

Nous dirons que $A$ est définie positive si cette forme hermitienne vérifie

$$< Ax, x >> 0 \text{ pour tout } x \in \mathbb{C}^n,\ x \neq 0,$$

et nous dirons que $A$ est une matrice positive si elle vérifie "seulement"

$$< Ax, x >\geq 0 \text{ pour tout } x \in \mathbb{C}^n.$$

### 1.2.6 Valeurs propres, vecteurs propres et réduction de matrices

**Définitions**

(a) Cas des matrices complexes :
soit $A \in \mathscr{M}_n(\mathbb{C})$ une matrice carrée de taille $n$ à coefficients complexes. Les valeurs propres de $A$ sont les $n$ racines complexes (distinctes ou confondues) du polynôme caractéristique de $A$ défini par

$$P_A(\lambda) = \det(A - \lambda I_n),$$

et sont données par $\lambda_1$, ..., $\lambda_n$ ou $\lambda_1(A), ..., \lambda_n(A)$.
La multiplicité d'une valeur propre est sa multiplicité en tant que racine de $P_A$. Une valeur propre de multiplicité égale à un est dite simple.
Le spectre de $A$ est l'ensemble des valeurs propres de $A$ (sous ensemble de $\mathbb{C}$), il est noté $\text{sp}(A)$.
Le rayon spectral de la matrice $A$ est le nombre réel positif défini par

$$\rho(A) = \max\{|\lambda_i(A)|, \;\; 1 \leq i \leq n\},$$

où $|.|$ désigne le module d'un nombre complexe.

A toute valeur propre $\lambda$ de $A$ est associé (au-moins) un vecteur $x \in \mathbb{C}^n$ non nul tel que $Ax = \lambda x$. On dit alors que $x$ est un vecteur propre de $A$ associé à la valeur propre $\lambda$.
On appelle sous-espace propre associé à la valeur propre $\lambda$ le sous-espace vectoriel de $\mathbb{C}^n$ :

$$E_\lambda = \ker(A - \lambda I_n) = \{x \in \mathbb{C}^n, Ax = \lambda x\}.$$

Rappelons les expressions de la trace et le déterminant en fonction des valeurs propres

$$\text{Tr}(A) = \sum_{i=1}^{n} \lambda_i(A), \det(A) = \prod_{i=1}^{n} \lambda_i(A).$$

3. Cas des matrices réelles :
pour les matrices à coefficients réels, deux points de vue sont possibles :

   (a) nous pouvons voir $A \in \mathscr{M}_n(\mathbb{R})$ comme une matrice à coefficients complexes, et nous reprenons alors la définition des valeurs propres et vecteurs propres donnés plus haut. Ces quantités sont alors a priori respectivement dans $\mathbb{C}$ et $\mathbb{C}^n$.

   (b) nous ne sortons pas du corps des réels : les valeurs propres de $A \in \mathscr{M}_n(\mathbb{R})$ sont les racines réelles du polynôme caractéristique. Un vecteur propre associé à la valeur propre $\lambda \in \mathbb{R}$ est alors un vecteur $x$ non nul de $\mathbb{R}^n$ tel que $Ax = \lambda x$.
   Le sous-espace propre associé à la valeur propre $\lambda \in \mathbb{R}$ est alors le sous-espace vectoriel de $\mathbb{R}^n$ :

$$E_\lambda = \ker(A - \lambda I_n) = \{x \in \mathbb{R}^n, Ax = \lambda x\}.$$

   Notons qu'avec cette dernière définition, une matrice peut ne pas avoir de valeurs propres.

**Remarque**

1. *Attention : dans la suite de ce cours, la définition de rayon spectral fait intervenir les valeurs propres dans $\mathbb{C}$ même pour une matrice réelle : que $A$ soit dans $\mathscr{M}_n(\mathbb{C})$ ou dans $\mathscr{M}_n(\mathbb{R})$ son rayon spectral est défini comme*

$$\rho(A) = \max\{|\lambda|, \lambda \text{ racines dans } \mathbb{C} \text{ de } P_A\}.$$

2. *Pour la réduction (diagonalisation,...) des matrices réelles, il est important de préciser dans quel corps nous raisonnons : $\mathbb{R}$ ou $\mathbb{C}$.*

### Quelques résultats sur la réduction des matrices

Soit $E$ un $\mathbb{K}$-espace vectoriel de dimension $n$ et $u : E \to E$ un application linéaire représentée par la matrice $A = (a_{ij})$ relativement à une base $\mathscr{B} = (e_j)$.
Relativement à une autre base de $E$, $\mathscr{B}' = (e'_j)$, la même application linéaire $u$ est représentée par la matrice

$$B = P^{-1}AP,$$

où $P$ est la matrice de passage de la base $\mathscr{B} = (e_j)$ à la base $\mathscr{B}' = (e'_j)$. $P$ est la matrice inversible dont le $j$ème vecteur colonne est formé des composantes du vecteur $e'_j$ dans la base $(e_j)$.
Une même application linéaire est donc représentée par différentes matrices selon la base choisie. Le problème de réduction de matrices est de trouver une base pour laquelle la matrice est la plus simple possible, le cas d'une matrice diagonale étant le plus favorable.

### Diagonalisation

Une matrice $A$ est dite diagonalisable sur $\mathbb{K}$ s'il existe une matrice $P \in \mathscr{M}_n(\mathbb{K})$ inversible telle que $P^{-1}AP$ soit diagonale.
Cette diagonale contient alors exactement les valeurs propres de $A$, et les vecteurs colonnes de $P$ sont des vecteurs propres de $A$.
Autrement dit, une matrice est diagonalisable si et seulement s'il existe une base de vecteurs propres.

**Théorème 2** (MATRICE DIAGONALISABLE)

La matrice $A \in \mathscr{M}_n(\mathbb{K})$ est diagonalisable sur $\mathbb{R}$ (respectivement sur $\mathbb{C}$) si et seulement si

1. ses valeurs propres sont toutes dans $\mathbb{R}$ (respectivement dans $\mathbb{C}$),
2. pour chacune de ses valeurs propres $\lambda$, la dimension du sous-espace propre $E_\lambda$ est égale à la multiplicité de $\lambda$ (en tant que racine du polynôme caractéristique).

De ce qui précède nous pouvons déduire le corollaire suivant :

**Corollaire 1** (VALEURS PROPRES SIMPLES)

Une matrice dont toutes les valeurs propres sont simples est diagonalisable.

**Théorème 3** (MATRICE SYMETRIQUE DIAGONALISABLE)

1. Matrices normales :
   soit $A \in \mathscr{M}_n(\mathbb{C})$ une matrice normale (c'est à dire que $AA^* = A^*A$). Alors il existe une matrice unitaire $U$ (c'est à dire $UU^* = U^*U = I_n$) telle que $U^*AU$ soit diagonale, autrement dit $U^*AU = \text{diag}(\lambda_1, ..., \lambda_n)$ où $\lambda_1, ..., \lambda_n$ sont les valeurs propres de $A$.
2. Matrices symétriques réelles :
   une matrice symétrique réelle est diagonalisable dans une base orthonormale, autrement dit, il existe une matrice orthogonale $\mathscr{O}$ ( $\mathscr{O}^T\mathscr{O} = \mathscr{O}\mathscr{O}^T = I_n$) telle que $\mathscr{O}^T A\mathscr{O}$ soit diagonale.

**Triangularisation**

Il existe des matrices réelles ou complexes qui ne sont pas diagonalisables. Mais dans $\mathbb{C}$, nous pouvons toujours trianguliariser une matrice, autrement dit, pour toute matrice $A \in \mathscr{M}_n(\mathbb{C})$, il existe une matrice inversible $P \in \mathscr{M}_n(\mathbb{C})$ telle que $P^{-1}AP$ soit une matrice triangulaire.
La réduction sous forme de Jordan, vue en deuxième année de licence donne une forme réduite simple. Mais nous n'utiliserons pas ce résultat dans ce cours-ci. Voici un résultat plus facile à montrer qui nous suffira.

**Théorème 4** (FACTORISATION DE SCHUR)

Pour toute matrice $A \in \mathscr{M}_n(\mathbb{C})$, il existe une matrice unitaire $U$ telle que $U^*AU$ soit triangulaire.

## 1.3 Normes vectorielles et matricielles

Etant donné que nous allons travailler sur les matrices : étudier la convergence de suites de matrices, mesurer des erreurs, etc., il est important de donner quelques résultats utiles qui nous serviront d'outils pour les méthodes de résolution de systèmes linéaires.

### 1.3.1 Rappels sur les normes vectorielles

Ici, $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.

**Définition 1** (NORME DE VECTEUR)

Une norme sur $\mathbb{K}^n$ est une application $\|.\| : \mathbb{K}^n \to \mathbb{R}^+$ telle que

1. pour tout $x \in \mathbb{K}^n$, $\|x\| = 0_{\mathbb{K}}$ implique $x = 0_n$,
2. pour tout $x \in \mathbb{K}^n$, pour tout $\lambda \in \mathbb{K}$, $\|\lambda x\| = |\lambda|\|x\|$,
3. pour tous $x, y \in \mathbb{K}^n$, $\|x + y\| \leq \|x\| + \|y\|$ (inégalité triangulaire).

Les normes usuelles que l'on utilisera le plus souvent sont :

1. la norme
$$\|x\|_1 = \sum_{j=1}^{n} |x_j|$$
2. la norme associée au produit scalaire (appelée aussi norme euclidienne sur $\mathbb{R}^n$) :
$$\|x\|_2 = \sqrt{< x, x >} = \left( \sum_{j=1}^{n} |x_j|^2 \right)^{1/2}$$
3. la norme $l^\infty$ (appelée également norme du sup) :
$$\|x\|_\infty = \max_{1 \leq j \leq n} |x_j|,$$
4. la norme $l^p$, pour $1 \leq p < +\infty$ :
$$\|x\|_p = \left( \sum_{j=1}^{n} |x_j|^p \right)^{1/p}.$$

### 1.3.2 Boules

Définissons maintenant les boules ouvertes et fermées.
-Pour les boules ouvertes centrées en $a \in \mathbb{K}^n$ de rayon $r > 0$ :
$$\mathscr{B}(a, r) = \{x \in \mathbb{K}^n \text{ tel que } \|x - a\| < r\}$$
-Pour les boules fermées centrées en $a \in \mathbb{K}^n$ de rayon $r > 0$ :
$$\overline{\mathscr{B}}(a, r) = \{x \in \mathbb{K}^n \text{ tel que } \|x - a\|| \leq r\}$$

**Théorème 5** (NORMES ÉQUIVALENTES)

Sur $\mathbb{K}^n$, toutes les normes sont équivalentes.
Autrement dit, si $N_1$ et $N_2$ sont deux normes, il existe deux réels $c$ et $C$ strictement positifs, tels que pour tout $x \in \mathbb{K}^n$,
$$cN_2(x) \leq N_1(x) \leq CN_2(x).$$

On choisit la norme qui convient le mieux suivant le contexte du problème considéré.

### 1.3.3 Normes matricielles

1. $\mathscr{M}_n(K)$ vu comme $\mathbb{K}^{n^2}$ :
   Nous considérons $\mathscr{M}_n(\mathbb{K})$ comme un $K$-espace vectoriel de dimension $n^2$. Nous pouvons définir toutes les normes $l^p$, avec $1 \leq p < \infty$ de la façon suivante :

$$\|A\|_{l^p} = \left( \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^p \right)^{1/p}.$$

   Pour $p = 2$, nous notons

$$\|A\|_F = \|A\|_{l^2} = \left( \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}, \text{ (norme de Frobenius)}$$

   et

$$\|A\|_{l^\infty} = \max_{1 \leq i,j \leq n} |a_{ij}|.$$

   Mais en général, ces normes n'ont pas de bonnes propriétés par rapport au produit matriciel. D'où la section suivante.

2. Normes matricielles :

**Définition 2** (NORME MATRICIELLE)

Une norme matricielle est une norme sur $\mathscr{M}_n(\mathbb{K})$ qui vérifie de plus

$$|||AB||| \leq |||A|||\,|||B|||,$$

pour tous $A, B \in \mathscr{M}_n(K)$.

**Exemple**

*(a) La norme de Frobenius est une norme matricielle,*

*(b) la norme $|||A|||_{l^\infty}$ n'est pas une norme matricielle comme nous l'avons définie dans la section précédente.*

3. Normes subordonnées : énonçons une manière de construire des normes matricielles.

**Définition 3** (NORME MATRICIELLE SUBORDONNÉE)

Nous nous donnons une norme vectorielle $\|.\|$ sur $\mathbb{K}$. Nous lui associons ensuite une norme sur $\mathscr{M}_n(\mathbb{K})$, notée encore $|||.|||$ que nous définissons pour tout $A \in \mathscr{M}_n(\mathbb{K})$ par

$$|||A||| = \sup_{x \in \mathbb{K}^n \setminus \{0_n\}} \frac{\|Ax\|}{\|x\|}$$

Nous l'appellerons norme subordonnée à la norme vectorielle $\|.\|$.

**Propriété 1** (PROPRIÉTÉS DES NORMES SUBORDONNÉES)

Soient $A \in \mathscr{M}_n(\mathbb{K})$ et $x \in \mathbb{K}^n$, pour toute norme matricielle subordonnée,

$$\|Ax\| \leq |||A|||\|x\|.$$

**Propriété 2** (PROPRIÉTÉS DES NORMES SUBORDONNÉES)

Soit $|||.|||$ une norme subordonnée. Nous avons alors les propriétés suivantes :

(a) $|||A||| = \sup_{x \in \mathbb{K}^n \setminus \{0_n\}} \dfrac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\|$,
$|||A|||$ est bien définie, autrement dit le sup est bien $< +\infty$,

(b) le sup est atteint, c'est à dire qu'il existe un $x \in \mathbb{K}^n \setminus \{0_n\}$ tel que $\|Ax\| = |||A|||\|x\|$,

(c) $|||.|||$ est une norme matricielle,

(d) $|||I_n||| = 1$.

Nous pouvons alors définir les normes subordonnées aux normes vectorielles $\|.\|_1$, $\|.\|_2$, $\|.\|_\infty$ de la façon suivante :

**Propriété 3** (NORMES SUBORDONNÉES CLASSIQUES)

(a) $|||A|||_1 = \max_{j=1,\dots,n} \sum_{i=1}^{n} |a_{ij}|$,

(b) $|||A|||_\infty = \max_{i=1,\dots,n} \sum_{j=1}^{n} |a_{ij}|$,

(c) $|||A|||_2 = \sqrt{\rho(A^*A)} = \sqrt{\rho(AA^*)} = |||A^*|||_2$,

(d) si en plus $A$ est normale ($A^*A = AA^*$), alors $|||A|||_2 = \rho(A)$.

### 1.3.4 Conditionnement

L'objectif de cette section est de répondre à la question suivante : peut-on dire *a priori* qu'une matrice est sensible aux erreurs et à des petites perturbations en général ?
Dans toute la section, nous allons considérer $A \in \mathscr{M}_n(\mathbb{K})$ une matrice inversible. Et nous considé-

rons le système

$$Ax = b, \tag{1.2}$$

où $b$ et $x$ sont des vecteurs de $\mathbb{K}^n$.
Plusieurs cas se présentent à nous.

1. Cas 1 : si nous perturbons $b$
   Nous considérons ici le système suivant

   $$A(x + \delta x) = b + \delta b, \tag{1.3}$$

   où $\delta b$ et $\delta x$ sont des vecteurs de $\mathbb{K}^n$ considérés comme des "petites" perturbations de $b$ et $x$ respectivement.
   Le système $(1.3)$ s'écrit alors

   $$Ax + A\delta x = b + \delta b,$$

   et comme $x$ est solution de $(1.2)$ il vient

   $$\delta x = A^{-1}\delta b.$$

   Avec une norme subordonnée, nous obtenons

   $$\|\delta x\| = \|A^{-1}\delta b\| \leq |||A^{-1}|||\|\delta b\|.$$

   D'autre part,

   $$\|b\| = \|Ax\| \leq |||A|||\|x\|,$$

   et donc

   $$\frac{1}{\|x\|} \leq \frac{|||A|||}{\|b\|},$$

   en supposant bien entendu que $x \neq 0_n$ (ce que nous supposons pour avoir la norme subordonnée) et que $b \neq 0_n$, mais ce cas là serait trivial puisqu'étant donné que la matrice $A$ est inversible la seule solution serait alors $x = 0_n$).
   Ainsi, on obtient

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

2. Cas 2 : si nous perturbons $A$ :
   Nous considérons maintenant le système suivant

   $$(A + \Delta A)(x + \Delta x) = b. \tag{1.4}$$

   Dans ce cas là, en développant, nous obtenons

   $$Ax + \Delta A(x + \Delta x) + A\Delta x = b,$$

   soit encore

   $$\Delta x = -A^{-1}(\Delta A(x + \Delta x)).$$

   Ainsi, en considérant les normes,

$$\begin{aligned} \|\Delta x\| &= \|A^{-1}\Delta A(x+\Delta x)\|, \\ &\leq |||A^{-1}\Delta A|||\|x+\Delta x\|, \\ &\leq |||A^{-1}|||\,|||\Delta A|||\|x+\Delta x\|. \end{aligned}$$

Et finalement

$$\frac{\|\Delta x\|}{\|x+\Delta x\|} \leq (|||A^{-1}|||\,|||A|||)\,\frac{|||\Delta A|||}{|||A|||},$$

en supposant bien entendu que la perturbation $\Delta x$ n'annule pas le vecteur $x$ et $A$ soit une matrice non nulle.

**Définition 4** (CONDITIONNEMENT DE $A$)

La quantité $\text{cond}(A) = |||A^{-1}|||\,|||A|||$ est appelée conditionnement de $A$ relativement à la norme matricielle $|||.|||$.

Nous pouvons alors énoncer un résultat plus précis *via* le théorème suivant.

**Théorème 6** (CONDITIONNEMENT OPTIMAL)

Soient $A \in \mathscr{M}_n(\mathbb{K})$ une matrice inversible et $|||.|||$ sa norme subordonnée.

1. Pour $b \in \mathbb{K}^n$, $b \neq 0_n$ et $\delta b \in \mathbb{K}^n$, si $Ax = b$ et $A(x+\delta x) = b + \delta b$, alors
$$\frac{\|\delta x\|}{\|x\|} \leq \text{cond}(A)\frac{\|\delta b\|}{\|b\|}.$$
Cette majoration est optimale. C'est à dire qu'il existe $b \neq 0_n$ et $\delta b$ tels que
$$\frac{\|\delta x\|}{\|x\|} = \text{cond}(A)\frac{\|\delta b\|}{\|b\|}.$$
2. Pour tout $\Delta A \in \mathscr{M}_n(\mathbb{K})$ et pour tout $b \in \mathscr{M}_n(\mathbb{K})$, si si $Ax = b$ et $(A+\Delta A)(x+\delta x) = b$, alors
$$\frac{\|\Delta x\|}{\|x+\Delta x\|} \leq \text{cond}(A)\frac{|||\Delta A|||}{|||A|||}.$$
Cette majoration est optimale. C'est à dire qu'il existe $A$ matrice non nulle et $\Delta A$ telles que
$$\frac{\|\Delta x\|}{\|x+\Delta x\|} = \text{cond}(A)\frac{|||\Delta A|||}{|||A|||}.$$

Il arrive que certains systèmes linéaires soient très sensibles aux erreurs, c'est à dire mal conditionnés, indépendamment de la méthode de résolution utilisée. Nous allons donc énoncer quelques propriétés sur les conditionnements de matrices qui pourront nous être fortement utiles dans des cas pratiques.

**Propriété 4** (PROPRIÉTÉS DES CONDITIONNEMENTS)

Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice inversible, nous avons les propriétés suivantes :

1. $\text{cond}(A) \geq 1$,
2. $\text{cond}(A) = \text{cond}(A^{-1})$,
3. $\text{cond}(\alpha A) = \text{cond}(A)$, pour tout $\alpha \in \mathbb{K}/\{0\}$,
4. $\text{cond}_2(A) = |||A|||_2 |||A^{-1}|||_2 = \dfrac{\mu_n(A)}{\mu_1(A)}$ où $0 < \mu_1(A) < ... < \mu_n(A)$ sont les racines carrées des valeurs propres de $A^*A$ (valeurs singulières de $A$),
5. si $A$ est normale alors $\text{cond}_2(A) = \dfrac{|\lambda_{max}(A)|}{|\lambda_{min}(A)|}$, où $\lambda_{max}(A)$ et $\lambda_{min}(A)$ sont respectivement la plus grande et la plus petite valeur propre en module de $A$,
6. si $U$ est unitaire, $\text{cond}_2(U) = 1$.

**Remarque** ***Remèdes à un mauvais conditionnement.***
*Malheureusement il n'y a pas de méthode universelle. Une solution possible est l'équilibrage.*
*Il s'agit de remplacer $Ax = b$ par un système équivalent $BAx = Bb$, où $B$ est inversible.*
*S'il n'y a pas de $0$ sur la diagonale, $D = diag(a_{11}, ..., a_{nn})$ est inversible et on résout le système $D^{-1}Ax = D^{-1}b$.*
*Nous avons parfois $cond(D^{-1}A \ll cond(A)$.*

## 1.4 Méthodes directes de résolution de systèmes linéaires

un système linéaire est la donnée de $n$ équations à $m$ inconnues du type

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1m}x_m = b_1, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nm}x_m = b_n. \end{cases}$$

Les inconnues ici sont $x_1$, $x_2$,...$x_m$.
Sous forme matricielle, ce système s'écrit

$$Ax = b$$

où

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \in \mathbb{K}^m, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{K}^n \text{ et, } \quad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \in \mathscr{M}_{nm}(\mathbb{K}).$$

Trois cas se présentent :

1. si $n < m$ : il y a moins d'équations que d'inconnues. Le système est sous-déterminé. L'ensemble des solutions est un sous espace affine,
2. si $n > m$ : il y a plus d'équations que d'inconnues. Le système est sur-déterminé,
3. si $n = m$ = il y a exactement le même nombre d'équations que d'inconnues. Dans la suite nous considérerons (sauf exceptions) ces systèmes carrés.
   Dans ce cas, il existe une unique solution $x$ à $Ax = b$ pour tout vecteur $b \in \mathbb{K}^n$ si et seulement si $A \in \mathscr{M}_{nn}(K)$ est inversible.

Il existe plusieurs méthodes pour résoudre ces systèmes : les méthodes directes et les méthodes itératives. Commençons par étudier les méthodes directes comme l'indique l'intitulé de cette section.

### 1.4.1 Principe des méthodes directes

Nous nous ramenons en un nombre fini d'opérations à un système simple à résoudre en général triangulaire (parfois une matrice orthogonale). Ces méthodes reviennent souvent à écrire une décomposition de la matrice $A = BC$, où $B$ et $C$ ont des propriétés particulières.
Ainsi résoudre $Ax = b$ équivaut à résoudre $B(Cx) = b$ ce qui est équivalent à résoudre

$$By = b,$$

puis

$$Cx = y.$$

La plus célèbre des méthodes est appelée pivot de Gauss.

### 1.4.2 Pivot de Gauss - Décomposition $LU$

Le pivot de Gauss (ou méthode d'élimination de Gauss), permet de décomposer la matrice $A$ en produit de deux matrices $LU$ où :
-$L$ est une matrice triangulaire inférieure (Lower en anglais),
-$U$ est une matrice triangulaire supérieure (Upper en anglais).
Nous avons le résultat suivant.

#### L'algorithme de Gauss

**Théorème 7** (PIVOT DE GAUSS)

Pour toute matrice $A \in \mathscr{M}_n(\mathbb{K})$, il existe une matrice inversible $M$ telle que $T = MA$ soit triangulaire.

#### Remarque

1. *Ce résultat reste vrai pour $A$ non inversible, mais alors $T$ n'est pas inversible non plus et on ne peut résoudre le système triangulaire obtenu.*
2. *Ce résultat n'est pas un théorème de réduction de matrice.*

### La factorisation $LU$

Nous allons montrer ici que l'algorithme permettant de prouver le résultat du théorème de la section précédente permet d'écrire (dans certains cas)

$$A = LU,$$

où $L$ est une matrice triangulaire inférieure et $U$ est une matrice triangulaire supérieure.
Par conséquent, résoudre le système $Ax = b$ revient à résoudre les deux systèmes triangulaires

$$\begin{cases} Ly &= b, \\ Ux &= y. \end{cases}$$

Cette méthode est intéressante quand on doit résoudre beaucoup de systèmes du type

$$Ax_j = b_j,$$

parce que $L$ et $U$ sont calculées une fois pour toute et nous n'avons plus qu'à les utiliser pour chacun des cas.
**Notation :** on notera dans la suite de ce chapitre, la matrice

$$\Delta_k(A) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix} \in \mathscr{M}_k(\mathbb{K}).$$

**Théorème 8** (DÉCOMPOSITION $LU$)

Soit $A \in \mathscr{M}_n(\mathbb{K})$. Si pour tout $1 \leq k \leq n$, la matrice $\Delta_k(A)$ est inversible (autrement dit $\det \Delta_k(A) \neq 0$), alors il existe une unique décomposition de $A$ de la forme $A = LU$ avec $L$ une matrice triangulaire inférieure ne possédant que des $1$ sur la diagonale, et $U$ une matrice triangulaire inférieure.

### La méthode de Crout

Cette méthode permet de calculer la décomposition $LU = A$
(voir détails en cours).

### Complexité de la méthode de Crout

Il faut environ $2n^3/3$ opérations pour calculer les matrices $LU$. Pour résoudre un système triangulaire il faut $n^2$ opérations. Rien à voir avec le nombre d'opérations de l'ordre de $(n+1)!$ nécessaires.

**Application au calcul du déterminant et de l'inverse**

-pour le déterminant :
une fois obtenue la décomposition $A = LU$, nous avons

$$\det A = \det L \det U.$$

Il est facile de remarquer que $\det L = 1$. Par conséquent,

$$\det A = \det U = \prod_{i=1}^{n} u_{ii}.$$

-pour l'inverse :
on note

$$A^{-}1 = (x_1|...|x_n),$$

où chacun des vecteurs $x_i$ est solution de $Ax_i = e_i$, avec $e_i$ le ième vecteur de la base canonique (composé de $0$ partout sauf la ième composante qui vaut $1$). Nous procédons alors de la façon suivante :
- nous calculons la décomposition $LU$,
- nous résolvons

$$\begin{cases} Ly_i &= e_i, \\ Ux_i &= y_i, \end{cases} \text{ pour tout } i = 1, ..., n.$$

Au total le nombre d'opérations sera de l'ordre de $8n^3/3$.

**Recherche de pivot**

Si nous tombons sur un pivot nul, nous procédons à une permutation. Si nous tombons sur un pivot très petit, nous l'utilisons alors au risque de commettre de grosses erreurs de calculs d'arrondis.
Comment y remédier ?
Nous procédons à la recherche de pivot.
-pivot partiel : nous recherchons le plus grand coefficient de la ligne (ou de la colonne) considérée et nous faisons une permutation avant de continuer.
-pivot total : nous cherchons le plus grand coefficient dans toute la matrice.
Complexité des algorithmes de tri :
-$n^2 \log_2(n)$ pour le pivot partiel,
-$n^3 \log_2(n)$ pour le pivot total.
Or, $n^2 \log_2(n) = o(n^3)$ donc en général on utilise le pivot partiel.

### 1.4.3 Cas des matrices symétriques définies positives :la factorisation de Cholesky

**Théorème 9** (CHOLESKY)

Si $A \in \mathscr{M}_n(\mathbb{K})$ est une matrice symétrique réelle définie positive, il existe au-moins une matrice $B$ triangulaire inférieure telle que

$$A = BB^T.$$

De plus, on peut imposer que les coefficients diagonaux de $B$ soient strictement positifs et alors la décomposition est unique.

Complexité : nous calculons $B$ par identification $BB^T = A$. Nous obtenons alors un ordre en $n^3/6$ additions, $n^3/6$ multiplications et $n$ extractions de racines, ce qui est moins que la méthode $LU$.

### 1.4.4 Factorisation $QR$

**Théorème 10** (DÉCOMPOSITION $QR$)

Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice inversible. Il existe un unique couple $(Q, R)$ tel que $Q$ est orthogonale et $R$ est une matrice triangulaire supérieure à diagonale strictement positive telles que $A = QR$.

**Remarque** *Résoudre $Ax = b$ est alors équivalent à résoudre le système*

$$\begin{cases} Qy &= b, \\ Rx &= y. \end{cases}$$

*Sachant que $Qy = b$ s'inverse facilement en donnant car $Q^{-1} = Q^T$ et donc $y = Q^T b$.*

La preuve s'inspire du procédé d'orthogonalisation de Gram-Schmidt.

## 1.5 Méthodes itératives de résolution de systèmes linéaires

Dans la section précédente, nous avons vu les méthodes appelées méthodes directes car elles fournissent des solutions exactes (aux erreurs d'arrondis près bien entendu, ce qui n'est pas rien).Cependant, ces calculs peuvent être lourds pour de très grandes matrices.
Une alternative consiste à faire converger des suites vers la solutions. Nous perdrons alors l'exactitude de la solution, mais nous gagnerons en rapidité d'exécution dans l'approximation de la solution (suivant le degré de précision voulu).

### 1.5.1 Principe des méthodes itératives

Nous décomposons $A$ inversible sous la forme $A = M - N$. Nous avons alors

$$Ax = b \text{ équivalent à } Mx = Nx + b.$$

Sous cette forme, nous cherchons à approcher la solution du problème par la méthode itérative suivante :

$$\begin{cases} x_0 \in K^n & \text{donné,} \\ Mx_{k+1} & = Nx_k + b, \quad \text{pour tout entier } k \geq 0. \end{cases}$$

A chaque itération, nous devons résoudre un système linéaire de matrice $M$ pour calculer $x_{k+1}$ en fonction de $x_k$.
La méthode est 'intéressante" quand $M$ est "facile" à inverser (ou plutôt le système facile à résoudre).
Montrons que cette méthode, si elle converge, approche bien la solution de $Ax = b$.
Soit $x \in \mathbb{K}^n$ solution de $Ax = b$.
Supposons que la suite $(x_k)_{k\in\mathbb{N}}$ converge vers $x_0 \in \mathbb{K}^n$ nous avons

$$\lim_{k\to+\infty} \|x_k - x_0\| = 0$$

pour une norme quelconque.
Comme $M$ et $N$ sont linéaires, elles sont continues (nous sommes en dimension finie), donc à la limite $k \to +\infty$ dans $Mx_{k+1} = Nx_k + b$. Nous obtenons

$$Mx_\infty = Nx_\infty + b, \text{ soit encore } Ax_\infty = b.$$

Et comme $A$ est inversible, par unicité de la solution nous avons $x_\infty = x$
-Étude de la convergence :
Notons $e_k = x_k - x$, l'erreur commise au rang $k$. La méthode converge si et seulement si $\lim_{k\to+\infty} e_k = 0$ ou encore $\lim_{k\to+\infty} \|e_k\| = 0$.
Pour tout $k \in \mathbb{N}$, on a

$$e_{k+1} = x_{k+1} - x.$$

Nous choisirons $M$ inversible dans cette décomposition. De telle sorte que

$$Mx_{k+1} = Nx_k + b,$$

équivalent à

$$x_{k+1} = M^{-1}Nx_k + M^{-1}b.$$

De même

$$Ax = b = Mx - Nx,$$

qui est équivalent à

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

Nous pouvons en déduire que

$$e_{k+1} = M^{-1}Ne_k.$$

Par une récurrence immédiate nous obtenons

$$e_k = (M^{-1}N)^k e_0.$$

Nous démontrerons dans les sections suivantes les résultats permettant d'énoncer le théorème de convergence.

**Théorème 11** (CONVERGENCE DE LA MÉTHODE ITÉRATIVE)

Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice inversible. La méthode itérative associée à la décomposition $A = M - N$, avec $M$ inversible, converge si et seulement si le rayon spectral $\rho(M^{-1}N) < 1$.

Tout le problème consiste dès lors à bien choisir la décomposition $M-N$. Les trois décompositions proposées ci-dessous sont les trois méthodes les plus connues.

### 1.5.2 Trois méthodes classiques

Nous écrivons

$$A = D - E - F,$$

de la façon suivante

$$\begin{pmatrix} & & -F \\ & D & \\ -E & & \end{pmatrix}$$

où $D = \text{diag}(a_{11}, ..., a_{nn})$, et $E$ et $F$ sont des matrices triangulaires définies par

$$-(E)_{ij} = \begin{cases} a_{ij} & \text{si } i > j, \\ 0 & \text{sinon,} \end{cases} \quad \text{et} \quad -(F)_{ij} = \begin{cases} a_{ij} & \text{si } i < j, \\ 0 & \text{sinon,} \end{cases}$$

Nous supposerons que $A$ ne possède pas de zéros sur sa diagonale.

#### Méthode de Jacobi

Dans cette méthode, nous prenons

$$M = D, \text{ et } \quad N = E + F,$$

l'avantage de cette méthode est que dans ce cas là, $M$ est très facile à inverser. La méthode de Jacobi s'exprime alors par la suite

$$\begin{cases} x_0 \in K^n & \text{donné,} \\ Dx_{k+1} & = (E+F)x_k + b, \quad \text{pour tout entier } k \geq 0. \end{cases}$$

**Méthode de Gauss-Seidel**

Dans cette méthode, nous prenons

$$M = D - E, \text{ et } \quad N = F.$$

La méthode de Gauss-Seidel s'exprime alors par la suite

$$\begin{cases} x_0 \in K^n & \text{donné,} \\ (D-E)x_{k+1} & = Fx_k + b, \quad \text{pour tout entier } k \geq 0. \end{cases}$$

**Méthode de relaxation**

C'est une méthode intermédiaire entre les deux méthodes précédentes. Nous mettons un peu de diagonale de chaque côté en posant Dans cette méthode, nous prenons

$$M = \frac{1}{\omega}D - E, \text{ et } \quad N = \frac{1-\omega}{\omega}D + F,$$

où $\omega \neq 0$ est un paramètre de relaxation. La méthode de relaxation s'exprime alors par la suite

$$\begin{cases} x_0 \in K^n & \text{donné,} \\ (\frac{1}{\omega}D - E)x_{k+1} & = (\frac{1-\omega}{\omega}D + F)x_k + b, \quad \text{pour tout entier } k \geq 0. \end{cases}$$

### 1.5.3 Critère général de convergence, étude des suites d'itérées de matrices

**Lien entre rayon spectral et normes matricielles**

**Théorème 12** (CONVERGENCE DE LA MÉTHODE ITÉRATIVE)

1. Pour toute norme matricielle sur $\mathscr{M}_n(\mathbb{C})$ et pour toute matrice $A \in \mathscr{M}_n(\mathbb{C})$, $\rho(A) \leq \|A\|$.
2. Pour toute matrice $A \in \mathscr{M}_n(\mathbb{C})$ et pour tout $\varepsilon > 0$, il existe une norme subordonnée telle que $\|A\| \leq \rho(A) + \varepsilon$. Attention : dans ce cas là la norme $\|.\|$ dépend de $A$ et de $\varepsilon$.

**Suites d'itérés de matrices**

**Théorème 13** (CONVERGENCE ET ÉQUIVALENCES)

Soit $B \in \mathscr{M}_n(\mathbb{C})$ les assertions suivantes sont équivalentes :

1. $\lim_{k\to+\infty} B^k = 0$ (autrement dit $\lim_{k\to+\infty} |||B^k||| = 0$ pour $|||.|||$ norme quelconque sur $\mathscr{M}_n(\mathbb{C})$,
2. pour tout $x \in \mathbb{C}^n$, $\lim_{k\to+\infty} B^k x = 0$ (autrement dit $\lim_{k\to+\infty} \|B^k x\| = 0$ pour $\|.\|$ norme quelconque sur $\mathscr{M}_n(\mathbb{C})$,
3. $\rho(B) < 1$,
4. il existe une norme matricielle telle que $|||B||| < 1$.

## 1.5.4 Quelques cas particuliers de convergence

**Cas des matrices symétriques (ou hermitiennes)**

**Théorème 14** (CONVERGENCE ET MATRICE SYMÉTRIQUE)

Soit $A \in \mathscr{M}_n(\mathbb{R})$ une matrice symétrique réelle, définie, positive. Nous la décomposons sous la forme $A = M - N$, où $M$ est inversible. Alors

1. $M^T + N$ est symétrique,
2. si de plus $M^T + N$ est définie positive, alors $\rho(M^{-1}N) < 1$.

**Applications**

1. Si $A$ est symétrique, si $A$ et $2D - A$ sont définies positives alors la méthode de Jacobi converge.

2. Si $A$ est symétrique définie positive, alors la méthode de Gauss-Seidel converge.

3. Si $A$ est symétrique définie positive, et si $0 < \omega < 2$, alors la méthode de relaxation converge.

**Matrice à diagonale strictement dominante**

**Définition 5** (MATRICE A DIAGONALE STRICTEMENT DOMINANTE)

On dit que $A \in \mathscr{M}_n(\mathbb{C})$ est à diagonales strictement dominantes si pour tout $1 \leq i \leq n$,

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|.$$

**Théorème 15** (CONVERGENCE MATRICE DIAGONALE STRICT. DOM.)

Si $A \in \mathscr{M}_n(\mathbb{C})$ est à diagonales strictement dominante, alors la méthode de Jacobi converge.

**Remarque** *Une matrice à diagonale strictement dominante est inversible.*

## 1.6 Méthodes numériques de calcul de valeurs propres et vecteurs propres

### 1.6.1 Motivation : modes propres

**Exemple** *L'étude des valeurs est par exemple un outil fondamental pour l'étude des vibrations de structures mécaniques : comme pour la résistance aux séismes dans la construction d'immeubles. Un modèle très simplifié alternant les murs de masse $m_1$ avec des ressorts de raideur $k_i$ où $i = 1, .., 3$. Le principe fondamental de la dynamique nous permet d'écrire le système d'équations :*

$$\begin{cases} m_1 y_1'' + k_1 y_1 + k_2(y_1 - y_2) & = & 0, \\ m_2 y_2'' + k_2(y_2 - y_1) + k_3(y_2 - y_3) & = & 0, \\ m_3 y_3'' + k_3(y_3 - y_2) & = & 0. \end{cases}$$

Notons

$$y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix}, \quad M = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix} \quad \text{et} \quad M = \begin{pmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{pmatrix}.$$

Le système se réécrit alors

$$My''(t) + Ky(t) = 0.$$

Nous cherchons des solutions périodiques en temps $s$ sous la forme

$$y(t) = y(0)e^{i\omega t},$$

où $y(0)$ est le vecteur condition initiale.
Alors

$$y''(t) = -\omega^2 e^{i\omega t} y(0).$$

Nous obtenons alors

$$(-\omega^2 My(0) + Ky(0))e^{i\omega t} = 0, \text{ pour tout } t > 0.$$

Soit encore,

$$(M^{-1}K)y(0) = \omega^2 y(0).$$

Les modes qui peuvent se propager sont telles que $\omega^2$ est valeur propre de $M^{-1}K$.

### 1.6.2 Difficultés

La recherche des éléments propres est un problème beaucoup plus difficile que celui de la résolution de systèmes linéaires. c'est un problème d'algèbre linéaire mais ce n'est pas un problème linéaire. Les valeurs propres de $A + B$ ne sont en général pas la somme de celles de $A$ et de celles de $B$.
Il n'existe pas de méthode directe (calcul exacte en un nombre fini d'étapes). Les valeurs propres sont les racines d'un polynôme. Depuis Galois et Abel, nous savons que ce n'est pas possible au-delà du degré $5$.
Par conséquent, il n'y a que des méthodes itératives pour résoudre ce problème.
Il n'y a pas de méthodes pour laquelle nous savons prouver la convergence pour toute matrice.
Nous ne calculerons pas le polynôme caractéristique puis ses racines pour calculer les valeurs propres (sauf éventuellement pour des petites matrices (inférieur ou égal à 3).

### 1.6.3 Conditionnement spectral

Nous nous intéressons à l'amplification des erreurs, ici pour le problème de recherche des valeurs propres.

**Exemple** *Considérons le bloc de Jordan suivant*

$$A = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathscr{M}_k(\mathbb{C}) \text{ , et pour } \varepsilon > 0, A(\varepsilon) = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ \varepsilon & 0 & \cdots & 0 \end{pmatrix}.$$

*Le polynôme caractéristique de $A(\varepsilon)$ est $X_{\varepsilon(x)} = x^n - \varepsilon$ (matrice compagne) dont les racines sont $\varepsilon^{1/n} e^{i2k\pi/n}$, pour $k = 0, ..., n-1$.*
*Une erreur d'ordre $\varepsilon$ (par exemple $10^{-20}$), entraînera une erreur d'ordre $\varepsilon^{1/n}$ sur les valeurs propres. Par exemple, pour $n = 20$, cela donnerait $\varepsilon^{1/n} = 10^{-1}$.*
*Cette matrice est mal conditionnée pour le problème de recherche de valeurs propres.*

-Pour les matrices diagonalisables :
nous définissons le conditionnement spectral par

$$\Gamma(A) = \inf\{\text{cond}(P), \quad \text{inversible} \quad \text{telle} \quad \text{que} \quad P^{-1}AP \quad \text{diagonale}\}.$$

**Propriété 5** (PROPRIÉTÉS CONDITIONNEMENT SPECTRAL)

Soit $A \in \mathscr{M}_n(\mathbb{K})$, nous avons les propriétés suivantes :

1. $\Gamma(A) \geq 1$,
2. $\Gamma(A) = 1$ pour $A$ normale (symétrique orthogonale, unitaire)

**Théorème 16** (LOCALISATION DES VALEURS PROPRES AVEC PERTURBATION)

Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice diagonalisable, de valeurs propres $\lambda_1, ..., \lambda_n$. Soit $|||.|||$ une norme subordonnée telle que

$$|||\text{diag}(\lambda_1, ..., \lambda_n)||| = \max_{1\leq i\leq n} |\lambda_i|.$$

C'est bon pour $|||.|||_1$, $|||.|||_2$, $|||.|||_\infty$.
Le spectre de $A + \delta A$ (où $\delta A$ est une matrice quelconque) est inclus dans $\bigcup_{1\leq i\leq n} D_i$,
où

$$D_i = \{z \in \mathbb{C} \quad \text{tel} \quad \text{que} \quad |z - \lambda_i| \leq \Gamma(A)|||\delta A|||\}.$$

**Remarque**

1. *Nous obtenons une majoration de l'erreur et non de l'erreur relative.*

2. *Ne pas confondre le conditionnement pour la résolution de systèmes linéaires et le conditionnement spectral.*
   *Par exemple :*

$$A_n = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \in \mathscr{M}_n(\mathbb{R}),$$

*La matrice $A_n$ est symétrique donc $\Gamma(A_n) = 1$ pour tout $n \in \mathbb{N}^*$.*
*D'un autre côté* $\text{cond}_2(A_n) \underset{n\to+\infty}{\sim} \frac{4}{\pi^2}n^2 \underset{n\to+\infty}{\to} +\infty$.

### 1.6.4 Méthode de la puissance

Soient $A \in \mathscr{M}_n(\mathbb{C})$ et $y_0 \in \mathbb{C}^n$. La méthode de la puissance est basée sur l'itération suivante.
Pour tout $k \geq 0$,

$$y_{k+1} = Ay_k.$$

Alors pour tout $k \geq 0$,

$$y_k = A^k y_0.$$

Et comme on a

$$\lim_{k \to +\infty} \left|\left|\left| A^k \right|\right|\right|^{1/k} = \rho(A).$$

Autrement dit, $A^k$ se comporte pour $k$ suffisamment grand comme $\rho(A)^k I_n$.
La méthode de la puissance permet de calculer une valeur propre approchée de la plus grande valeur propre en module et un vecteur propre associé.

**Théorème 17** (CALCUL DES VECTEURS PROPRES ET VALEURS PROPRES)

Soit $A \in \mathscr{M}_n(\mathbb{K})$ une matrice diagonalisable, de valeurs propres $\lambda_1$, ..., $\lambda_n$ telles que

$$|\lambda_1| > |\lambda_2| \geq ... \geq |\lambda_n|.$$

Notons $(v_1, v_2, ..., v_n)$ une base de vecteurs propres tels que

$$\|v_i\|_2 = 1 \text{ pour tout } i = 1, ..., n.$$

Soient $y_0 \in \mathbb{C}^n$ et $(y_k)_{k \in \mathbb{N}}$ définie par

$$y_{k+1} = Ay_k \text{ pour tout } k \in \mathbb{N}.$$

Nous écrivons $y_0 = \sum_{i=1}^{n} a_i v_i$. Nous supposons que $a_i \neq 0$.

Alors

$$y_k = \lambda_1^k \left( a_1 v_1 + \underset{k \to +\infty}{O} \left( \left| \frac{\lambda_2}{\lambda_1} \right|^k \right) \right).$$

De plus le quotient de Raleygh satisfait

$$\frac{\langle Ay_k, y_k \rangle}{\|y_k\|_2} = \lambda_1 + \underset{k \to +\infty}{O} \left( \left| \frac{\lambda_2}{\lambda_1} \right|^k \right).$$

Enfin,

$$\lim_{k \to +\infty} \frac{\overline{\lambda_1^k}}{|\lambda_1|^k} \frac{y_k}{\|y_k\|_k} = q$$

un vecteur associé à $\lambda_1$

**Remarque**

1. *Si $A$ est symétrique, il existe une base orthonormée de vecteurs propres et la convergence est d'ordre* $\underset{k\to+\infty}{0}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right)$.
2. *On peut encore montrer un résultat de convergence dans des cas plus généraux où*
$$|\lambda_1| > |\lambda_2| \geq ... \geq |\lambda_n|.$$
*et $A$ est non diagonalisable. La convergence est cependant moins rapide.*
3. *$A$ a une seule valeur propre de plus grand module, mais pas nécessairement simple.*
4. *La méthode de la puissance ne marche plus quand $A$ admet deux valeurs propres de même module.*
5. *En pratique, $\|y_k\|$ croît exponentiellement donc pour éviter un "overflow" on normalise $y_k$ à chaque étape de la façon suivante :*
$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|}, \text{ pour tout } k \geq 0.$$
*Alors pour tout $k \in \mathbb{N}$,*
$$x_k = \frac{A^k x_0}{\|A^k x_0\|}.$$
*Pour trouver la valeur propre de $A$ la plus proche d'un $\mu$ donné, on utilise ce qu'on appelle la méthode de la puissance inverse :*
$$(A - \mu I_n)y_{k+1} = y_k, \text{ pour tout } k \in \mathbb{N}.$$
*Si $\lambda_1$ est proche de $\mu$ on a*
$$\frac{1}{|\lambda_1 - \mu|} \gg \frac{1}{|\lambda_i - \mu|} \text{ pour tout } i \neq 1.$$

### 1.6.5 Généralisation de la méthode de la puissance : la méthode $QR$

Supposons $|\lambda_1| > |\lambda_2| \geq ... \geq |\lambda_n|$. Nous cherchons à calculer $\lambda_1$ et $\lambda_2$. Nous considérons $y_0$, $z_0 \in \mathbb{C}^n$ tels que $< y_0, z_0 >= 0$ (orthogonaux).
Pour $k \geq 0$, nous posons

$$\begin{cases} y_{k+1} &= Ay_k, \\ z_{k+1} &= Az_k - \beta_{k+1}y_{k+1}, \end{cases}$$

où $\beta_{k+1}$ est choisi tel que $< y_{k+1}, z_{k+1} >= 0$.
Alors

$$\begin{cases} y_k &= A^k y_0, \\ z_k &= A^k z_0 - \gamma_k y_k, \end{cases}$$

où $\gamma_k$ est défini est que $< z_k, y_k >= 0$.
Il s'agit ici de combiner la méthode de la puissance avec la projection sur $y_k^{\perp}$.
Nous posons $y_0 = \sum_{i=1}^{n} a_i v_i$ et $z_0 = \sum_{i=1}^{n} b_i v_i$.
Étudions $z^k$ :

$$< y_k, z_k >= 0 = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{\lambda_i}^k \lambda_j^k \overline{a_i}(b_j - \gamma_k a_j) < v_j, v_i >.$$

Le terme prépondérant correspond à $i = j = 1$ donc $\gamma_k \simeq \dfrac{b_1}{a_1}$, et

$$\overline{a_1}\,(b_1 - \gamma_k a_1)\,\lambda_1^k \left(1 + \underset{k\to+\infty}{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\right) =$$
$$-\overline{a_1}\,(b_1 - \gamma_k a_2)\,\lambda_2^k \left(< v_2, v_1 > + \underset{k\to+\infty}{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\right).$$

Donc

$$z_k = \lambda_2^k\,(b_2 - \gamma_k a_2)\left(v_2 - < v_2, v_1 > v_1 + \underset{k\to+\infty}{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)\right),$$

autrement dit $k_k$ s'approche de $v_2 - < v_2, v_1 > v_1$ qui est le projeté orthogonal de $v_2$ sur $v_1^{\perp}$.
Pour récupérer $\lambda_1$ et $\lambda_2$, nous écrivons

$$U_k = \left(\frac{y_k}{\|y_k\|_2}, \frac{z_k}{\|z_k\|_2}\right).$$

Alors

$$\lim_{k\to+\infty} U_k^* A U_k = \begin{pmatrix} \lambda_1 & * \\ 0 & \lambda_2 \end{pmatrix}$$

Ainsi,

1. Le terme $(1,1)$ de la matrice précédente est tel que
$$\lim_{k\to+\infty} \frac{\langle Ay_k, y_k\rangle}{\|y_k\|_2} = \lambda_1$$
2. Le terme $(2,2)$ de la matrice précédente est tel que
$$\lim_{k\to+\infty} \frac{\langle Az_k, z_k\rangle}{\|z_k\|_2} = \lambda_2$$
3. Le terme $(2,1)$ de la matrice précédente est tel que
$$\lim_{k\to+\infty} \frac{\langle Ay_k, z_k\rangle}{\|y_k\|_2\|z_k\|_2} = 0$$

Pour obtenir toutes les valeurs propres, nous partons de $n$ vecteurs orthogonaux, et nous utilisons le même principe.
La réécriture de cette méthode sous la forme suivante constitue la décomposition $QR$ de la recherche des valeurs propres. C'est la méthode la plus utilisée en pratique, elle a été introduite dans les années 60.
L'algorithme décrit précédemment peut se réécrire

$$AU_k = U_{k+1}R_{k+1},$$

où $R$ est triangulaire supérieure.
nous choisissons $U_n$ orthogonale telle que pour tout $k \in \mathbb{N}$,

$$\left\{ \begin{array}{lcl} z_{k+1} & = & AU_k, \\ U_{k+1}R_{k+1} & = & Z_{k+1}, \end{array} \right.$$

c'est la décomposition $QR$.
L'algorithme $QR$ est le suivant

1. $A_0 = A$,
2. pour tout $k \in \mathbb{N}$,
   $\left\{ \begin{array}{lcl} A_k & = & Q_kR_k, \quad (\text{décompostion } QR) \\ A_{k+1} & = & R_kQ_k, \quad (\text{définition de } A_{k+1}). \end{array} \right.$

Nous avons ainsi

$$A_0 = Q_0R_0,\ A_1 = R_0Q_0,\ A_2 = R_2Q_1 = Q_2R_2, \text{ etc.}$$

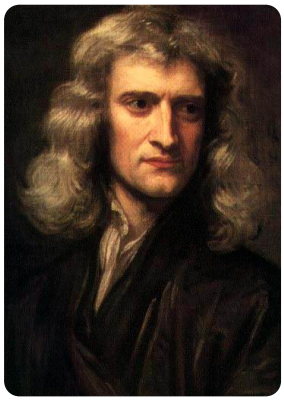
Pour tout $k \in \mathbb{N}$, $Q_R$ est orthogonale (ou unitaire), c'est à dire que $Q_k^TQ_k = In$ ou $Q_k^*Q_k = In$, et donc

$$A_{k+1} = Q_k^TA_kQ_k.$$

Toutes mes matrices $A_k$ sont semblables, c'est à dire qu'elles ont les mêmes valeurs propres.
La méthode $QR$ converge pour beaucoup plus de matrice que celles pour lesquelles nous savons montrer un théorème de convergence.
Qu'entend-on par converger ici ?
Nous voulons que les matrices $A_k$ deviennent simples : triangulaires ou diagonales.

# Chapitre 2

# Résolution approchée d'équations non linaires



(a) Sir Isaac Newton (1642–1727), Newton qui partage avec Gottfried Wilhelm Leibniz la découverte du calcul infinitésimal, a également développé une méthode éponomyme pour trouver un point fixe.

(b) Thomas Simpson (1710-1761), mathématicien anglais généralisa la méthode de Newton au calcul itératif des solutions d'une équation non linéaire, en utilisant les dérivées

(c) Stefan Banach (1892-1945), mathématicien polonais est l'un des fondateurs de l'analyse fonctionnelle, à qui l'on doit entre autre un théorème éponyme du point fixe.

FIGURE 2.1 – Quelques mathématiciens célèbres liés à l'étude des nombres entiers, rationnels et réels.

## 2.1 Introduction

Le problème est ici de résoudre des équations de la forme

$$f(x) = 0,$$

òù $f$ est définie d'un sous-ensemble de $\mathbb{R}$ à valeurs dans $\mathbb{R}$.
Cela paraît simple mais nous verrons plusieurs méthodes pour y arriver. Nous dirons à la fin de ce chapitre un mot sur les systèmes d'équations non-linéaires où cette fois-ci $f : D \subset \mathbb{R}^n \to \mathbb{R}$, où $n \geq 2$.

## 2.2 Dichotomie

Cette méthode repose sur le théorème des valeurs intermédiaires. En particulier :
si $f : [a, b] \to \mathbb{R}$ continue et si $f(a)f(b) < 0$, alors il existe $c \in ]a, b[$ tel que $f(c) = 0$.

## 2.3 Méthode de type point fixe

### 2.3.1 Théorème-énoncé général

Nous nous intéressons ici à la résolution de l'équation

$$\varphi(x) = x.$$

**Théorème 1** (MÉTHODE DU POINT FIXE)

Soit $E$ un $\mathbb{R}-$espace vectoriel normé complet. Soit $\varphi : E \to E$ une fonction contractante. Il existe $K \in ]0, 1[$, tel que pour tous $x, y \in E$

$$\|\varphi(x) - \varphi(y)\| \leq K\|x - y\|.$$

Alors :

1. $\varphi$ admet un unique point fixe $x_* \in E$,
2. pour tout $x_k \in E$, la suite définie par

$$x_{k+1} = \varphi(x_k),$$

converge vers $x_*$.

Ici, nous utiliserons ce théorème avec $E = \mathbb{R}$ (ou $\mathbb{R}^*$) ou $E = [a, b]$.
Nous allons donner caractérisation du caractère contractant à l'aide de la dérivée pour les fonctions $\varphi$ définies sur $[a, b]$.
Supposons $\varphi$ de classe $\mathscr{C}^1$ sur $[a, b]$. Si $\varphi$ est contractante, nous avons pour tous $x, y \in [a, b]$

$$|\varphi(x) - \varphi(y)| \leq K|x - y|,$$

équivalent à

$$|\varphi'(x)| \leq K \text{ pour tout } x \in [a, b],$$

équivalent à

$$|\varphi'(x)| \leq 1 \text{ pour tout } x \in [a,b].$$

Pour montrer cette caractérisation, on utilise d'autre part, le théorème des accroissements finis :

$$|\varphi(x) - \varphi(y)| \leq \sup_{t\in[a,b]} |\varphi'(t)||x-y|.$$

### 2.3.2 Construction de méthodes pour $f(x) = 0$

Le but ici est de trouver $\varphi$ tel que résoudre $f(x) = 0$ soit équivalent à résoudre $\varphi(x) = x$.

1. Un premier choix simple serait de prendre $\varphi(x) = x - f(x)$.
   Supposons $f$ de classe $\mathscr{C}^1$ sur $[a,b]$. Alors $\varphi$ l'est aussi et nous avons

   $$\varphi'(x) = 1 - f'(x), \text{ pour tout } x \in [a,b].$$

   Nous souhaitons nous assurer de l'existence d'un réel positif $K < 1$ tel que

   $$|\varphi'(x)| \leq K,$$

   c'est à dire

   $$|1 - f'(x)| \leq K,$$

   ou encore

   $$1 - K \leq f'(x) \leq 1 + K. \tag{2.1}$$

   En particulier, $f'(x) > 0$ pour tout $x \in [a,b]$. Donc $f$ doit être strictement croissante sur $[a,b]$.
   Par conséquent, la condition (2.1) est assez restrictive.

2. Un choix moins contraignant réside dans le choix suivant de $\varphi$ :

   $$\varphi(x) = x - \lambda f(x),$$

   où $\lambda$ est un réel non nul à choisir convenablement. Dans ce cas là,

   $$|\varphi'(x)| \leq K \text{ pour tout } x \in [a,b]$$

   est équivalent à

   $$1 - K \leq \lambda f'(x) \leq 1 + K. \tag{2.2}$$

   Nous devons encore avoir $f'$ de signe constant mais la condition (2.3) reste beaucoup moins contraignant que la condition (2.1) pour un $\lambda$ bien choisi.

Ce qui précède permet de déduire ce que nous appellerons la méthode de la corde, décrite par la suite :

$$\begin{cases} x_0 & \in [a,b] \quad \text{quelconque}, \\ x_{k+1} & = \qquad x_k - \lambda f(x_k). \end{cases}$$

### 2.3.3 Vitesse de convergence

Nous reprenons le cadre du théorème général du point fixe :
soient $E$ un espace de Banach, $x_*$ l'unique point fixe de $\varphi$ et $K$ la constante de contraction.
Si nous notons $e_n = \|x_n - x_*\|$ pour tout $n \in \mathbb{N}$, alors

$$e_{n+1} = \|x_n - x_*\| = \|\varphi(x_n) - x_*\| = \|\varphi(x_n) - \varphi(x_*)\| \leq K\|x_n - x_*\|.$$

Autrement dit

$$e_{n+1} \leq Ke_n, \text{ pour tout } n \in \mathbb{N}.$$

Par récurrence immédiate, nous obtenons

$$e_n \leq K^n e_0, \text{ pour tout } n \in \mathbb{N},$$

par conséquent, $\lim_{n\to+\infty} e_n = 0$ au plus à la même vitesse que $K^n$.
Par conséquent, plus $K$ est petit, plus la convergence sera rapide.
Si nous regardons ce que ça donne pour la méthode de la corde citée un peu plus haut, d'après (2.3) nous avons, sous réserve que $f$ soit de classe $\mathscr{C}^1$,

$$1 - K \leq \lambda f'(x) \leq 1 + K. \tag{2.3}$$

Ce qui pour un $x$ fixé, nous donne pour le cas limite où $K = 0$,

$$\lambda = \frac{1}{f'(x)}.$$

C'est l'idée de la méthode de Newton : en chaque point, nous cherchons le meilleur $\lambda$ (qui va dépendre du point $x$).

## 2.4 Méthode de Newton

### 2.4.1 Principe

La dernière remarque ici conduit à l'écriture suivante

$$\begin{cases} x_0 \in [a, b] & \text{donné,} \\ x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)}, & n \in \mathbb{N}. \end{cases}$$

Une autre manière de voir la méthode de Newton est la suivante.
Nous cherchons à résoudre $f(x) = 0$ en se donnant un point de départ $x_0$. Nous ne savons pas en général résoudre une équation non linéaire, mais nous savons résoudre une équation affine.
L'idée ici est de remplacer le problème non linaire $f(x) = 0$ par un problème affine $g(x) = 0$, où $g$ est la "meilleure" approximation affine" de $f$ au voisinage de $x_0$.
De façon naturelle on identifie la représentation de $g$ à la tangente à la courbe de $f$ au point d'abscisse $x_0$ si $f$ est de classe $\mathscr{C}^1$ au voisinage de $x_0$. En effet,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \underset{x \to x_0}{o}(x - x_0).$$

Nous posons donc

$$g(x) = f(x_0) + f'(x_0)(x - x_0),$$

et nous définissons $x_1$ tel que $g(x_1) = 0$. Notons que $x_1$ est bien défini si $f'(x_0) \neq 0$.
En réitérant le procédé, nous obtenons la définition suivante

$$\begin{cases} x_0 \in [a, b] & \text{donné,} \\ x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)}, & n \in \mathbb{N}. \end{cases}$$

### 2.4.2 Théorème de convergence

**Théorème 2** (THÉORÈME DE CONVERGENCE)

Soit $f : [a, b] \to \mathbb{R}$ une fonction de classe $\mathscr{C}^2$ sur $[a, b]$. Nous supposons qu'il existe $x_* \in ]a, b[$ tel que $f(x_*) = 0$ et $f'(x_*) \neq 0$. Alors il existe $\varepsilon > 0$ tel que pour tout $x \in [x_* - \varepsilon, x_* + \varepsilon]$, la suite des itérés de Newton

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \text{ pour tout } n \in \mathbb{N},$$

1. soit bien définie pour tout $n \in \mathbb{N}$,
2. reste dans $[x_* - \varepsilon, x_* + \varepsilon]$,
3. converge vers $x$ quand $n$ tend vers $+\infty$,
4. admette l'existence d'un $C > 0$ tel que

$$|x_{n+1} - x_*| \leq C|x_n - x_*|^2, \text{ pour tout } n \in \mathbb{N}.$$

**Remarque** *La vitesse de convergence est telle que*

$$e_{n+1} \leq e_n^2.$$

*En général au bout de 3 ou 4 itérations, nous obtenons une précision de $10^{-8}$ à $10^{-16}$.*

**Remarque** *Cette méthode possède quand même quelques limites :*

1. *il faut connaître la dérivée $f'$ en chacun des points de la suite, ce qui peut être problématique quand $f$ provient de données expérimentales,*
2. *il faut partir d'un point assez proche de la solution cherchée en général, donc nous avons besoin d'informations a priori précises sur $f$ et $x_*$.*
   *Une des solutions pour palier ce problème est d'utiliser la méthode de dichotomie pour localiser assez grossièrement $x_*$ avant d'appliquer la méthode de Newton.*

## 2.5 Méthode de la sécante

L'idée ici, est de ne pas utiliser $f'$ et donc de remplacer la dérivée $f'(x_n)$ par une différence finie

$$d_n = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}.$$

La méthode s'écrit alors

$$\begin{cases} x_0, x_1 \in [a,b] & \text{donnés,} \\ x_{n+1} = x_n - f(x_n)\dfrac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}, & n \in \mathbb{N}^*. \end{cases}$$

C'est une méthode à deux pas : elle permet de calculer $x_{n+1}$ en fonction des deux valeurs précédentes $x_n$ et $x_{n-1}$.

**Théorème 3** (MÉTHODE DE LA SÉCANTE)

Soit $f : [a,b] \to \mathbb{R}$ une fonction de classe $\mathscr{C}^2$ sur $[a,b]$. Nous supposons qu'il existe $x_* \in ]a,b[$ tel que $f(x_*) = 0$ et $f'(x_*) \neq 0$. Alors il existe $\varepsilon > 0$ tel que pour tout $x \in [x_* - \varepsilon, x_* + \varepsilon]$, la suite de la méthode de la sécante définis par

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}, n \in \mathbb{N}^*,$$

1. soit bien définie pour tout $n \in \mathbb{N}$,
2. reste dans $[x_* - \varepsilon, x_* + \varepsilon]$,
3. converge vers $x$ quand $n$ tend vers $+\infty$,

## 2.6 Ordre d'une méthode itérative

Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'approximation de $x_*$, solution de $f(x_*) = 0$.
Nous notons

$$e_{n+1} = |x_n - x_*| \text{ pour tout } n \in \mathbb{N}.$$

Nous disons que la méthode itérative est d'ordre $\lambda \geq 1$ si $\lambda$ est le sup des réels $\mu$ pour lesquels il existe $C > 0$ tel que

$$e_{n+1} \leq C e_n^{\mu} \text{ pour tout } n \in \mathbb{N}.$$

1. Une méthode d'ordre 1 converge s'il existe un constante $c \leq 1$ telle que
$$e_{n+1} \leq C e_n \text{ pour tout } n \in \mathbb{N}.$$
2. Pour les méthodes d'ordre $\lambda > 1$, il faut que l'erreur initiale soit suffisamment petite.
$$e_n \leq C^{1+\lambda+\ldots+\lambda^{n-1}}(e_0)^{\lambda^n} \leq \frac{1}{\lambda - 1}(Ce_0)^{\lambda^n}.$$

Rappelons en effet que

$$1 + \lambda + ... + \lambda^{n-1} = \frac{\lambda^n - 1}{\lambda - 1} \leq \frac{\lambda^n}{\lambda - 1}.$$

Par conséquent, plus $\lambda$ est grand, plus la convergence est rapide.

(a) Méthode de Newton : ordre 2,

(b) Méthode sécante : ordre $\frac{1+\sqrt{5}}{2}$,

(c) Méthode du point fixe : ordre 1,

(d) Méthode de dichotomie : ordre 1.

## 2.7 Systèmes d'équations non linéaires

Considérons le problème

$$f(x) = 0,$$

où cette fois-ci $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$.

### 2.7.1 Point fixe

Le théorème du point fixe s'applique dans $\mathbb{R}^n$. Nous considérons alors la fonction $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ définie par

$$\varphi(x) = x - f(x),$$

(ou $\varphi(x) = x - \lambda f(x)$), $\lambda \neq 0$ est à choisir).
Pour appliquer le théorème du point fixe, il faut que $\varphi$ soit contractante. Il faut trouver une norme sur $\mathbb{R}^n$ pour laquelle il existe $K < 1$ tel que pour tous $x, y \in \mathbb{R}$,

$$\|\varphi(x) - \varphi(y)\| \leq K\|x - y\|.$$

**Remarque** *Si $\varphi$ est de classe $\mathscr{C}^2$, nous avons encore une inégalité des accroissements finis sur $\mathbb{R}^n$.*
*Si nous fixons $\mathscr{J}_{\varphi(x)}$, la jacobienne de $\varphi$ au point $x$ définie par*

$$\mathscr{J}_\varphi(x) = \left(\frac{\partial \varphi_i}{\partial x_j}\right)_{0 \leq i,j \leq n}(x),$$

*alors*

$$\|\varphi(x) - \varphi(y)\| \leq \sup_{t \in [0,1]} |||\mathscr{J}_\varphi(tx + (1-t)y)|||\|x - y\|.$$

### 2.7.2 Méthode de Newton dans $\mathbb{R}^n$

C'est à peu près la même idée que dans $\mathbb{R}$. Nous remplaçons $f$ au voisinage du point considéré par sa meilleure approximation affine

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + \underset{x \to x_0}{o} \|x - x_0\|,$$

ou encore

$$f(x) = f(x_0) + \mathscr{J}_f(x_0)(x - x_0) + \underset{x \to x_0}{o} \|x - x_0\|.$$

On pose

$$g(x) = f(x_0) + \mathscr{J}_f(x_0)(x - x_0),$$

et $x_1$ est défini par $g(x_1) = 0$, et on obtient, si $\mathscr{J}_f(x_0)$ est inversible

$$x_1 = x_0 - (\mathscr{J}_f(x_0))^{-1} f(x_0).$$

La méthode s'écrit alors

$$\begin{cases} x_0, \in R^d & \text{donné,} \\ x_{n+1} = x_n - (\mathscr{J}_f(x_0))^{-1} f(x_n), & n \in \mathbb{N}^*. \end{cases}$$

On a un résultat de convergence similaire à celui vu pour $d = 1$.

### 2.7.3 Retour sur les systèmes linéaires et aux méthodes itératives

Nous cherchons à résoudre $Ax = b$, où $b$ est un vecteur de $\mathbb{R}^n$ donné et $A \in \mathscr{M}_n(\mathbb{R})$ est une matrice inversible.
Nous écrivons $A = M - N$, $M$ inversible et nous proposons l'algorithme suivant :

$$\begin{cases} x_0, \in R^d & \text{donné,} \\ Mx_{k+1} = Nx_k + b, & k \in \mathbb{N}^*. \end{cases}$$

Ce qui se réécrit pour tout $k \in \mathbb{N}^*$ par

$$x_{k+1} = M^{-1}Nx_k + M^{-1}b.$$

Si nous posons

$$\begin{array}{rccl} f: & \mathbb{R}^n & \to & \mathbb{R}^n, \\ & x & \mapsto & M^{-1}Nx + M^{-1}b, \end{array}$$

alors pour tout $k \in \mathbb{N}^*$

$$x_{k+1} = f(x_k).$$

Si la suite $(x_k)_{k \in \mathbb{N}}$ converge, elle converge vers un point fixe de $f$.
Soit $\|.\|$ une norme de $\mathbb{R}^n$.
Alors pour tous $x, y \in \mathbb{R}^n$,

$$\|f(x) - f(y)\| = \|M^{-1}N(x - y)\| \leq |||M^{-1}N|||\|x - y\|.$$

Si pour une norme $|||M^{-1}N||| < 1$ on a $f$ contractante et le théorème du point fixe s'applique.