

Contents

Préface	3
1 Introduction à l'analyse numérique	5
1.1 Sources d'erreurs	5
1.2 Erreur absolue / Erreur relative	6
1.3 Erreurs de troncature.	7
1.3.1 Formule de Taylor et erreurs de troncature	7
1.4 Propagation d'erreurs.	9
1.4.1 Propagation d'erreurs dans une fonction à une variable.	10
1.5 Erreur d'affectation.	10
1.5.1 L'arrondis en notation flottante	10
1.6 Opérations risquées en calcul.	12
1.6.1 Erreur de cancellation.	12
1.6.2 Erreur de cumul	12
2 Résolution d'équation non linéaire	15
2.1 Racine de l'équation $f(x) = 0$	15
2.2 Séparation des racines	16
2.2.1 Méthode graphique	16
2.2.2 méthode de balayage	17
2.3 Méthodes de dichotomie	17
2.4 Méthode de point fixe	20
2.5 Méthode de Newton	24
3 Résolution de systèmes linéaires	29
3.1 Présentation du problème	29
3.2 Méthodes directes	30
3.2.1 Systèmes équivalents et systèmes particuliers	30

3.2.2	Méthode d'élimination de Gauss	31
3.2.3	Factorisation LU	33
3.2.4	Décomposition de Cholsky	34
3.3	Méthodes itératives	35
3.3.1	Rayon spectrale	35
3.3.2	Suites consistantes	36
3.3.3	Méthodes de Jacobi	37
3.3.4	Méthode de Gauss-Seidel	39

Préface

L'analyse numérique est une discipline qui consiste à développer, analyser et appliquer des méthodes issues de divers domaines mathématiques tels que l'analyse, l'algèbre linéaire, la géométrie, la théorie de l'approximation, les équations fonctionnelles, l'optimisation et le calcul différentiel. Les méthodes numériques ont des applications naturelles dans de nombreux problèmes de physique, de sciences biologiques, de sciences de l'ingénieur, d'économie et de finance.

Cette discipline est au carrefour de plusieurs domaines scientifiques et elle ne cesse de creuser de nouveaux horizons et de trouver de nouveaux champs d'application. Elle se développe aussi rapidement que l'outil phare de cette discipline, qui est l'ordinateur.

Ce cours est constitué de trois chapitres. Le premier chapitre représente une introduction aux notions d'erreurs générées par le calcul approché. Le chapitre deux est consacré à la résolution d'équation non linéaire. Le chapitre trois est la résolution de système linéaires.

Un annexe consacré à la présentation et l'apprentissage du logiciel MATLAB. Nous avons choisi ce logiciel puisqu'il est l'un des plus faciles à manipuler et à apprendre. Comme il est l'un des plus rapides et précis.

Dr. Khellaf

Introduction à l'analyse numérique

1.1 Sources d'erreurs

Il existe plusieurs types d'erreurs introduites dans les calculs :

Erreurs de modèles : On est souvent amené lors de la modélisation de phénomènes complexes à ne pas prendre en compte des paramètres qui sont considérés comme "négligeables". Ceci a pour conséquence d'introduire des erreurs dites de modèles ou de modélisation. Par exemple ne pas prendre en compte la résistance de l'air et les frottements du sol en étudiant le mouvement d'un véhicule.

Erreurs de mesures : L'utilisation d'instruments de mesures pour la collecte de données entraîne également des erreurs. Ces derniers sont soumis à un intervalle dit de confiance ou de précision, ainsi une bascule peut avoir une précision de quelques grammes.

Erreurs d'affectation : L'utilisation des ordinateurs comme outil de calcul introduit également des erreurs. L'ensemble des nombres qu'on peut représenter en machine étant fini car lié au système de représentation (nombres de bits utilisés), ceci implique l'existence d'un système d'affectation des nombres non représentés. Par exemple $\frac{1}{3}$ n'a pas de représentation binaire ou décimale finie, ainsi il est remplacé en machine par une partie finie de son développement soit décimale (0.3333...) soit binaire.

Erreurs de troncature : Le nombre d'opérations qu'on peut réaliser sur ordinateur étant fini, les algorithmes développés doivent tous s'arrêter à un moment donné. Ainsi on est amené à introduire des erreurs de troncature, par exemple pour évaluer la valeur de la fonction e^x on doit utiliser la formule
$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$$
 or comme on ne peut calculer indéfiniment sur machine on doit

tronquer le processus pour obtenir seulement une valeur approchée :

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \simeq \sum_{n=0}^{9999} \frac{x^n}{n!}$$

Dans ce chapitre on s'intéressera particulièrement aux erreurs d'affectation et aux erreurs de troncature.

1.2 Erreur absolue / Erreur relative

Les quantités dans le calcul numérique sont divisées en deux catégories:

- **Les quantités exactes** telles $1, 1/4, \sqrt{2}, \pi$
- **Les quantités approximatives** utilisées en machine (ordinateur, calculatrice...) et exprimée en virgule flottante qui sont en général symbolisés par les écritures $\sqrt{2} \simeq 1.41, \pi \simeq 3.14, e \simeq 2.718$.

On notera la valeur approchée stockée en mémoire d'une quantité exacte $fl()$.

Exemple 1. $fl(\pi) = 3.14$ et $fl(\frac{1}{6}) = 0.667$

Définition 1. On appelle *erreur absolue* de x , la quantité

$$E = |x - fl(x)|.$$

Définition 2. On appelle *erreur relative* à x , la quantité

$$E_r = \frac{|x - fl(x)|}{|x|} = \frac{E}{|x|}.$$

Cette quantité est souvent exprimée sous forme de pourcentage.

Exemple 2. Soit $fl\left(\frac{1}{3}\right) = 0.333$ l'erreur absolue est donnée par :

$$E = \left| \frac{1}{3} - 0.333 \right| = \left| \frac{1 - 0.999}{3} \right| = \left| \frac{0.001}{3} \right|$$

et l'erreur relative est donnée par

$$E_r = \frac{\left| \frac{0.001}{3} \right|}{\frac{1}{3}} = 0.001 = 0.1\%.$$

Généralement on ne connaît pas la valeur exacte, par conséquent on ne peut calculer les erreurs absolues et relatives, on introduit alors la notion de majorant de l'erreur.

Définition 3. On appelle majorant de l'erreur absolue d'une valeur approchée $fl(x)$ tout nombre réel positif Δx vérifiant:

$$E = |x - fl(x)| \leq \Delta x.$$

Définition 4. On note par $\delta x = \frac{\Delta x}{x}$ le majorant de l'erreur relative.

Exemple 3. Supposons qu'on utilise comme valeur approchée de $\sqrt{123}$ la valeur $\sqrt{121} = 11$ en appliquant la formule des accroissements finis à la fonction \sqrt{x} on obtient un majorant de l'erreur commise

$$|\sqrt{123} - \sqrt{121}| = \frac{2}{2\sqrt{123}} < \frac{1}{\sqrt{121}} < \frac{1}{10} = 0.1$$

1.3 Erreurs de troncature.

Pour approximer $\pi = 3.141592653589 \dots$, on peut considérer la valeur approchée 3.14 ou encore 3.1415, etc...et cela selon le besoin.

Dans le premier cas on a tronqué (couper en éliminant une partie) le nombre π après 2 décimales

Dans le second cas, on l'a tronqué après 5 décimale.

1.3.1 Formule de Taylor et erreurs de troncature

Brook Taylor 18 août 1685 - 29 décembre 1731. Angleterre.

Joseph-Louis Lagrange (Giuseppe Lodovico Lagrangia) 25 janvier 1736 - 10 avril 1813 Italie.

Théorème 1. Soit a un réel, x dans le voisinage de a et f une fonction n fois dérivables sur l'intervalle $[a, b]$ admettant une dérivée d'ordre $(n+1)$ sur l'intervalle ouvert $]a, b[$ alors il existe un élément ξ de $]a, b[$ tel que :

$$f(x) = f(a) + \frac{(x-a)}{1!} f'(a) + \frac{(x-a)^2}{2!} f''(a) + \dots + \frac{(x-a)^n}{n!} f^{(n)}(a) + \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(\xi).$$

Cette formule est dite formule de Taylor avec reste de Lagrange.

On peut obtenir une formule d'approximation à partir de la formule de Taylor en ne retenant que les n premiers termes.

$$f(x) \approx f(a) + \frac{(x-a)}{1!} f'(a) + \frac{(x-a)^2}{2!} f''(a) + \dots + \frac{(x-a)^n}{n!} f^{(n)}(a).$$

Dans l'approximation ainsi obtenue l'erreur absolue commise en arrêtant le calcul à l'étape n est donnée par :

$$E_n = \left| \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(\xi) \right|.$$

1.3. ERREURS DE TRONCATURE.

On constate que pour des valeurs de x proches de a l'expression $\frac{(x-a)^{n+1}}{(n+1)!}$ est décroissante et tend vers 0 quand n tend vers l'infini et comme $f^{(n+1)}(\xi)$ est une constante alors l'erreur E_n tend vers 0.

Ainsi ce qui permet aux calculateurs (de la calculatrice à l'ordinateur) d'évaluer une fonction autre qu'un polynôme ou une fraction rationnelle est le développement de Taylor tronqué (arrêté) à l'étape n , la valeur de n varie en fonction du calculateur, plus n est grand plus la valeur est précise, l'étape n , la valeur de n varie en fonction du calculateur, plus n est grand plus la valeur est précise, mais tous les calculateurs ont en commun que le fait que la valeur qu'ils renvoient est juste une valeur approchée et non exacte à l'exception de quelques situations ($e^0 = 1, \ln(1) = 0 \dots etc.$)

Exemple 4. On peut obtenir une expression permettant le calcul de e par la formule de Taylor de la fonction e^x développé au voisinage de 0. On considérera l'intervalle $[0, 2]$ par exemple

$$e^x = 1 + \frac{x}{1!} + \dots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!}e^\xi$$
$$\Rightarrow e^1 = e = 1 + \frac{1}{1!} + \dots + \frac{1^n}{n!} + \frac{1^{n+1}}{(n+1)!}e^\xi$$

On ne retenant que les n premiers termes du développement on obtient un moyen de calculer e

$$\Rightarrow e \approx fl(e) = 1 + \frac{1}{1!} + \dots + \frac{1}{n!}$$

L'erreur ainsi commise est

$$E_n = \frac{1}{(n+1)!}e^\xi \rightarrow 0$$

Le tableau suivant détaille le nombre de chiffres correctes qu'on peut obtenir en fonction du degré du développement de la formule de Taylor utilisée :

Degré	Résultat obtenu (les chiffres exactes sont en gras)	Chiffres exactes
$n = 10$	2.7182818011 463844797178130511463844797178130511464	(10)
$n = 15$	2.7182818284589944642854695764748674801584854494907	(12)
$n = 20$	2.7182818284590452353397844906664158861464034345402	(20)
$n = 25$	2.7182818284590452353602874687778324515093860784391	(26)

La référence de comparaison utilisée est le nombre

2.7182818284590452353602874713526624977572470937000

qui représente les 50 premiers chiffres exactes de la constante e .

1.4 Propagation d'erreurs.

Dans la pratique, lors d'un processus de calcul (évaluation d'une fonction, calcul des termes successifs d'une suite) on est amené à remplacer une quantité exacte par sa valeur approchée. Le problème est de savoir alors quelle est l'erreur commise à la fin.

Soit par exemple la fonction $f(x) = \frac{x^3 + x^2 + 1}{x^4 + 1}$ si au lieu d'évaluer $f(\frac{1}{3})$ on évalue $f(0.333)$ ainsi on aura introduit une erreur $E = \frac{0.001}{3}$, la question qui se pose est de savoir estimer l'erreur commise sur la fonction $E_f = |f(\frac{1}{3}) - f(0.333)|$ en fonction de l'erreur $E = \frac{0.001}{3}$.

1.4.1 Propagation d'erreurs dans une fonction à une variable.

Supposons que x est une quantité exacte et $fl(x)$ sa valeur approchée avec une erreur absolue Δx , pour estimer l'erreur commise on a recours à la formule de Taylor, en posant :

$$F(x) \simeq F(fl(x)) + \frac{\Delta x}{1!} F'(fl(x)) + \frac{(\Delta x)^2}{2!} F''(fl(x)) + \dots + \frac{(\Delta x)^n}{n!} F^{(n)}(fl(x))$$

La formule la plus utilisée est celle de Taylor du premier ordre :

$$F(x) \simeq F(fl(x)) + \Delta x F'(fl(x)) \Rightarrow \Delta F = \Delta x |F'(fl(x))|$$

Exemple 5. Soit la fonction $F = \sqrt{x} = x^{\frac{1}{2}}$ et $fl(x) = 1.234, \Delta x = 0.0005 \Rightarrow x = 1.234 \pm 0.0005$

$$F(fl(x)) = 1.11086, \quad F'(x) = \frac{1}{2} x^{-\frac{1}{2}}$$

$$\Delta F = \Delta x |F'(fl(x))| = (.0005)(.04501) = 0.0002251$$

$$F(x) = 1.1109 \pm 0.0002251$$

1.5 Erreur d'affectation.

Nous avons souligné lors de l'introduction que l'ensemble des nombres qu'on peut représenter en machine est fini à cause du système de représentation, cette contrainte oblige les concepteurs à introduire un système d'affectation des nombres non représentés, un exemple déjà évoqué est celui de $\frac{1}{3}$ qui n'a pas de représentation binaire ou décimale finie ce qui fait qu'il est remplacé en machine par une partie finie de son développement, cette opération induit une erreur qu'on appelle erreur d'affectation.

Dans cette partie on s'intéresse aux problèmes liés aux erreurs d'affectation en donnant des exemples en calcul flottants, on suppose pour cela que les valeurs sont représentés en écriture flottante c'est à dire sous la forme :

$$fl(x) = \pm .d_1 d_2 d_3 \dots d_s \times 10^e$$

e est l'exposant avec $m \leq e \leq M$ et s est le nombre de chiffres de la mantisse. On dit que les chiffres d_i sont significatifs, de façon générale un chiffre est dit significatif s'il ne peut pas être obtenu à puissance près.

1.5.1 L'arrondis en notation flottante

Supposons que notre machine utilise s chiffres mais qu'on a un nombre dont le développement comporte plus de s chiffres, notre machine effectuera ce qu'on appelle un arrondi pour transformer x en nombre $fl(x)$ à s chiffres.

$$x = \pm.d_1d_2d_3\dots d_sd_{s+1}\dots \times 10^e$$

$$\begin{aligned} \text{On calcule } x' &= \pm.d_1d_2d_3\dots d_sd_{s+1}\dots \times 10^e \\ &\quad \pm.000\dots 0\underbrace{5}_{s}\dots \times 10^e \end{aligned}$$

L'arrondi de x est le nombre dont la mantisse est formée des s premiers chiffres de x' .

Exemple 6. Arrondir les nombres suivants en arithmétique à 3 chiffres.

x	Arrondi à 03 chiffres
1.238	1.24
1.233	1.23
2005	2010
2003	2000

Remarque 1. Dans une suite d'opérations tout résultat intermédiaire doit être arrondi.

Exemple 7. Produit / Division

Soient les nombres $x = 8367200$ et $y = 0.01385$.

$$\begin{aligned} fl((0.83672 \times 10^7) \times (0.1385 \times 10^{-1})) &= fl((0.83672 \times 0.1385) \times 10^6) = fl(0.11588572 \times 10^6) \\ &= 0.11589 \times 10^6 \\ fl((0.83672 \times 10^7) \div (0.1385 \times 10^{-1})) &= fl((0.83672 \times 0.1385) \times 10^8) = fl(6,0412996 \times 10^6) \\ &= fl(0.60412996 \times 10^7) = 0.60413 \times 10^7 \end{aligned}$$

Exemple 8. Addition / Soustraction

Dans ce cas il faut ajouter des zéros à la mantisse du plus petit nombre dans l'opération pour avoir la même puissance.

Soient les nombres $x = 37654$ et $y = 25.784$.

$$\begin{aligned} fl((0.37654 \times 10^5) + (0.25784 \times 10^2)) &= fl((0.37654 \times 10^5) + (0.00025784 \times 10^5)) \\ &= fl(0.37679784 \times 10^5) = (0.37679784 \times 10^5) \\ &= 0.37680 \times 10^5 = 37680.0 \end{aligned}$$

Remarque 2. L'ordre selon lequel sont effectuées les opérations est important et on perd les propriétés élémentaires des opérations arithmétiques. (associativité & distributivité)

Exemple 9. Calculons la somme $37654+25.784-37679 = 0.784$ en arithmétique à 5 chiffres en changeant l'ordre des opérations.

$$O1) (37654 + 25.784) - 37679 : fl(37654 + 25.784) = 37680 \rightarrow fl(37680 - 37679) = 1$$

$$O2) (37654 - 37679) + 25.784 : fl(37654 - 37679) = -25 \rightarrow fl(-25 + 25.784) = 0.784$$

1.6 Opérations risquées en calcul.

1.6.1 Erreur de cancellation.

L'erreur de cancellation intervient quand on soustrait deux nombres proches, si x et y sont donnés en valeurs approchées à s chiffres significatifs et s'ils sont proches alors il possède un nombre de chiffres en commun dans leurs développement lors de la soustraction on perd une partie de ces chiffres d'où une perte de précision.

Exemple 10. Soient $x = 0.83672 \times 10^2$ et $y = 0.83666 \times 10^2$. si on effectue l'opération $x - y$ on obtient 0.6×10^{-4} donc un résultat avec seulement un chiffre significatif qui signifie une perte de précision.

Remarque 3. Dans certaines situations une simple modification algébrique permet d'éviter ce genre de problèmes.

Exemple 11. Supposons que l'on veuille calculer $\sqrt{x+1} - \sqrt{x}$ pour $x = 100000$ dans une machine utilisant une arithmétique à 05 chiffres. ainsi le nombre 100001 ne pouvant pas être représenté dans une arithmétique à 05 chiffres sera convertit en 100000, en d'autres termes nous avons $x+1 = x$ et le résultat de la machine sera 0. Nous pouvons cependant utiliser la simplification

$$\frac{(\sqrt{x+1} - \sqrt{x})(\sqrt{x+1} + \sqrt{x})}{(\sqrt{x+1} + \sqrt{x})} = \frac{1}{\sqrt{x+1} + \sqrt{x}}$$

Grace à cette formule le résultat sera 0.15811×10^{-2}

1.6.2 Erreur de cumul

L'erreur de cumul se produit quand on additionne deux nombres trop différents en ordre de grandeur, si un nombre est trop grand une grande partie de ses chiffres significatifs apparaiteront dans le résultat.

Exemple 12. Soient $x = 10000$ et $y = 4$, effectuons la somme $x + y$ en arithmétique à 4 chiffres.

Le nombre 10004 possède 5 chiffres significatifs et ne peut pas être représenté en arithmétique à 4 chiffres, il sera donc arrondi en 10000.

Ainsi c'est comme si le nombre 4 disparaissait ou était absorbé par le nombre 10000 plus grand que lui.

Remarque 4. Dans le cas précédent l'erreur de cumul ne peut pas être évité, dans d'autres cas on peut atténuer l'effet de ces erreurs en observant une règle simple qui consiste à effectuer la somme des éléments par ordre de grandeur croissant, en effectue la somme les plus petits avant de passer au plus grands.

Exemple 13. on désire évaluer en arithmétique à 4 chiffres la somme $10000 + 4 + 4 + 4 + 4 + 4$.

Ici si on effectue le calcul dans l'ordre d'écriture on va éliminer 5 fois le nombre 4.

$$fl_{4CS}(10000 + 4) = 10000 \Rightarrow fl_{4CS}(10000 + 4 + 4) = fl_{4CS}((10000 + 4) + 4) = 10000$$

Par contre en sommant d'abord les plus petits nombres on obtient :

$$4 + 4 + 4 + 4 + 4 = 20 \Rightarrow 10000 + 4 + 4 + 4 + 4 + 4 = 10000 + 20 = 10\,020$$

Dr. Khellaf

Résolution d'équation non linéaire

2.1 Racine de l'équation $f(x) = 0$

Soit f une fonction de \mathbb{R} dans \mathbb{R} dont le domaine de définition est une partie D_f de \mathbb{R} .

Définition 5. On dit que $\alpha \in D_f$ est une racine de l'équation

$$f(x) = 0, \quad (2.1.1)$$

si

$$f(\alpha) = 0. \quad (2.1.2)$$

Résoudre l'équation (2.1.1), c'est de trouver tous les nombres réels α tels que (2.1.2) soit vérifiée. En d'autre termes, on cherche à déterminer l'ensemble

$$\text{Ker}(f) = \{x \in D_f : f(x) = 0\},$$

appelé noyau de f . $\text{Ker}(f)$ est donc l'ensemble des racines de l'équation $f(x) = 0$.

Il n'est pas toujours possibles de résoudre complètement ce problème pour toutes formes de fonctions f . $\text{Ker}(f)$ peut en effet, avoir un grand nombre de structures possibles.

Exemple 14.

- Soit $f(x) = ax^2 + bx + c$, où $a, b, c \in \mathbb{R}$, avec $D_f = \mathbb{R}$. Alors, $\text{Ker}(f)$ contient au plus deux éléments et peut aussi être vide.
- Soit $f(x) = \sin(x)$ et $D_f = \mathbb{R}$, alors les racines de l'équation $f(x) = 0$ sont en nombres fini dénombrable et

$$\text{Ker}(f) = \{x \in \mathbb{R}_+ : x = k\pi, k = 0, 1, 2, 3, \dots\}.$$

2.2. SÉPARATION DES RACINES

- Pour f définie par

$$f(x) = \begin{cases} \sin(\frac{1}{x}) & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Les racines de l'équation $f(x) = 0$ sont dans ce cas le demi droite négative \mathbb{R}_- et les éléments de la suite

$$S = \left\{ \frac{1}{k\pi} : k = 1, 2, 3, \dots \right\}.$$

Définition 6. On dit qu'une racine α d'une équation $f(x) = 0$ est **séparable** si peut trouver un intervalle $[a, b]$ tel que α soit la seule racine de cette équation de $[a, b]$.

Remarque 5. Dans les deux cas 1 et 2 de l'exemple précédent, toutes racines sont séparables. Par contre, dans le cas 3, les seules racines séparables sont les éléments de S .

Nous nous concentrons dans ce chapitre à **la localisation** et à **l'approximation** des racines d'une équation du type $f(x) = 0$.

2.2 Séparation des racines

Il n'y a pas de méthode générale pour séparer les racines d'une équation $f(x) = 0$. Pratiquement, en dehors de l'étude théorique directe de f . Si f est donnée analytiquement, on utilise deux types de méthodes: une méthode graphique et une méthode de balayage.

2.2.1 Méthode graphique

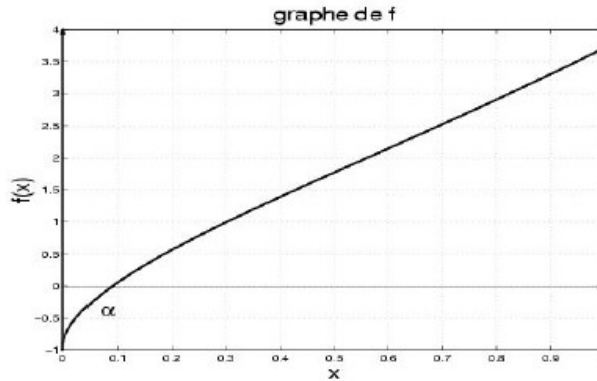
- 1) Soit on trace (expérimentalement ou par étude des variations de f) le graphe de fonction f et on cherche son intersection avec l'axe Ox .
- 2) Soit on décompose f en deux fonctions f_1 et f_2 simples à étudier, telles que : $f = f_1 - f_2$, et on cherche les points d'intersection des graphes de f_1 et f_2 , dont les abscisses sont exactement des racines de l'équation $f(x) = 0$.

Remarque 6. On choisit souvent f_1 et f_2 de façon à ce que leur courbes soit des courbes connues.

Exemple 15. On considère la fonction $f(x) = \exp(x) + 3\sqrt{x} - 2$ sur l'intervalle $[0, 1]$. Le code **Matlab** suivant trace le graphe de f . La figure montre que f admet un unique zéro $\alpha \in [0, 1]$.

Code Matlab

```
x = 0:0.001:1;
f = inline('exp(x)+3*sqrt(x)-2');
plot(x, f(x))
grid on;
ylabel('f(x)');
xlabel('x');
title('graphe de f');
```



2.2.2 méthode de balayage

On considère une suite croissante finie (x_i) (pour $i = 0, 1, 2, \dots, n$) de valeurs réparties sur l'intervalle $[a, b]$ contenu dans le domaine de définition D_f . Si f est continue et si $f(x_i)f(x_{i+1}) < 0$ alors il existe entre x_i et x_{i+1} au moins une racine de l'équation $f(x) = 0$ (c'est le théorème classique des valeurs intermédiaires).

La méthode consiste donc à déterminer parmi les quantités $f(x_i)f(x_{i+1})$ ($i = 0, 1, 2, \dots, n$) celles qui sont négatives.

Remarque 7. La méthode de balayage ne permet de conclure qu'à l'existence d'(au moins) une racine dans l'intervalle $[x_i, x_{i+1}]$.

2.3 Méthodes de dichotomie

Cette méthode est basée sur le théorème suivant:

Théorème 2. Soit $f : [a, b] \rightarrow \mathbb{R}$, une fonction continue. Si $f(a)f(b) < 0$, alors il existe $\alpha \in]a, b[$ tel que $f(\alpha) = 0$.

La méthode de dichotomie consiste à construire les trois suites $\{a_k\}_{k \geq 0}$, $\{b_k\}_{k \geq 0}$ et $\{x_k\}_{k \geq 0}$ définies par:

$$\begin{cases} a_0 = a, b_0 = b, x_0 = \frac{1}{2}(a_0 + b_0), \\ a_{k+1} = a_k, b_{k+1} = x_k, & \text{si } f(a_k)f(x_k) < 0, \\ a_{k+1} = x_k, b_{k+1} = b_k, & \text{si } f(b_k)f(x_k) < 0, \\ x_{k+1} = \frac{1}{2}(a_{k+1} + b_{k+1}), & k \geq 0. \end{cases}$$

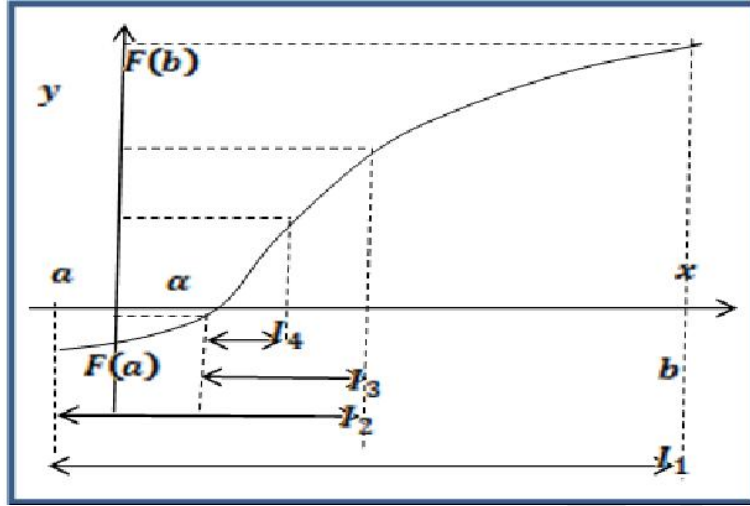


Figure 2.1: Méthode de dichotomie.

Proposition 1. Pour tout $k \geq 0$,

$$b_k - a_k = \frac{b - a}{2^k}.$$

Proof. On procède par récurrence. Le cas $k = 0$ est triviale à obtenir. Supposons que

$$b_k - a_k = \frac{b - a}{2^k}.$$

On distingue deux cas: Si $f(a_k)f(x_k) < 0$, alors $a_{k+1} = a_k$, $b_{k+1} = x_k = \frac{1}{2}(a_k + b_k)$, donc,

$$b_{k+1} - a_{k+1} = \frac{1}{2}(a_k + b_k) - a_k = \frac{1}{2}(b_k - a_k) = \frac{b - a}{2^{k+1}}.$$

De la même façon on obtient le résultat pour $f(b_k)f(x_k) < 0$. □

Théorème 3.

$$\lim_{k \rightarrow +\infty} x_k = \alpha.$$

Proof. D'après le théorème 9, $\alpha \in [a_k, b_k]$, $k \geq 0$ et $x_k = \frac{1}{2}(a_k + b_k)$, donc

$$|x_k - \alpha| \leq \frac{1}{2}(b_k - a_k) = \frac{b - a}{2^{k+1}} \rightarrow 0.$$

□

Test d'arrêt. Pour que la valeur de x_k de la suite à la n -ième itération soit une valeur approchée de α à $\varepsilon > 0$ près, il suffit que k vérifié:

$$\frac{b-a}{2^{k+1}} \leq \varepsilon.$$

On a alors:

$$|\alpha - x_k| \leq \frac{b-a}{2^{k+1}} \leq \varepsilon.$$

ce qui permet de calculer à l'avance le nombre maximal $k_0 \in \mathbb{N}$ d'itérations assurant la précision ε .

$$\frac{b-a}{2^{k+1}} \leq \varepsilon \Rightarrow \frac{b-a}{\varepsilon} \leq 2^{k+1} \Rightarrow k \geq \frac{\log(\frac{b-a}{\varepsilon})}{\log(2)} - 1.$$

Exemple 16. Soit la fonction $f(t) = \exp(t) + 3\sqrt{t} - 2$. **Le code Matlab** suivant pour approcher la racine de f sur l'intervalle $[0, 1]$ et il permet aussi de calculer la valeur de k nécessaire (le nombre d'itérations) pour atteindre la précision $\varepsilon = 10^{-10}$.

Code Matlab.

```
g = inline('exp(t) + 3*sqrt(t)-2');
Nit = 0;
epsilon = 1e-10;
borneinf = 0;
bornesup = 1;
pmilieu = (borneinf + bornesup)/2;

while and(g(pmilieu) ~= 0, (bornesup-borneinf) >= epsilon )
    Nit = Nit+1;
    if g(pmilieu)*g(borneinf) < 0
        bornesup = pmilieu;
    else
        borneinf = pmilieu;
    end
    pmilieu = (borneinf + bornesup)/2;
end
pmilieu
g(pmilieu)
Nit - 1
n_theorique = 10*log(10)/log(2) - 1
```

2.4. MÉTHODE DE POINT FIXE

Le résultat sur machine:

$$\alpha = 0.0910$$

$$f(\alpha) = -8.9593e - 12$$

$$k = n_{\text{numérique}} = 33 \quad n_{\text{théorique}} = 10 \log(10) / \log(2) - 1 = 32.2193$$

2.4 Méthode de point fixe

Nous abordons dans cette section la résolution d'équations non linéaires de type

$$x = g(x),$$

où g est une fonction donnée de D_g (domaine de définition) dans \mathbb{R} .

Définition 7. Soit $G : D_g \rightarrow \mathbb{R}$. g est dite **Lipschitzienne** s'il existe $l > 0$ tel que $\forall x, y \in D_g$;

$$|g(x) - g(y)| \leq l |x - y|.$$

g est dite **contractante** si $l < 1$.

Exemple 17.

1. La fonction $g(x) = \exp(x)$ définie sur $[-1, 1]$, est une fonction Lipschitzienne; En effet (d'après théorème des accroissements finis) pour tout $x, y \in [-1, 0]$:

$$|\exp(x) - \exp(y)| \leq \max_{x \in [-1, 1]} |\exp(x)| |x - y|,$$

alors

$$|\exp(x) - \exp(y)| \leq e |x - y|.$$

2. La fonction $g(x) = \sqrt{x}$ définie sur $[1, 2]$ est une fonction contractante; En effet (d'après théorème des accroissements finis) pour tout $x, y \in [1, 2]$:

$$|\sqrt{x} - \sqrt{y}| \leq \max_{x \in [1, 2]} \left| \frac{1}{2\sqrt{x}} \right| |x - y|,$$

alors

$$|\sqrt{x} - \sqrt{y}| \leq \frac{1}{2} |x - y|.$$

Définition 8. Si $x^* \in D_g$ tel que $x^* = g(x^*)$, x^* est dit **point fixe** de g .

Exemple 18.

1. $g : \mathbb{R}_+ \rightarrow \mathbb{R}$, $g(x) = \ln x$. Trouver les points fixes de g . Il est clair qu'il existe un seul point fixe $x^* = e$.
2. $g : \mathbb{R}_+ \rightarrow \mathbb{R}$, $g(x) = -\sqrt{x}$. Trouver les points fixes de g . Il n'existe aucun point fixe.

Il s'agit toujours de rechercher les racines d'équation $f(x) = 0$. Cette équation est supposée mise sous la forme $x = g(x)$ et ceci est toujours possible en posant par exemple $g(x) = x - f(x)$. La racine x^* de la fonction $f(x)$ est alors un point fixe de la fonction $g(x)$

$$f(x^*) = 0 \iff g(x^*) = x^*.$$

Exemple 19. l'équation $x^3 - 5 = 0$, devient sous la forme $x^3 + x - 5 = x$.

Théorème 4. Soit g une fonction réelle, définie et continue sur l'intervalle $[a, b]$ telle que:

1. $g(x) \in [a, b]$ pour $x \in [a, b]$ (on dit que l'intervalle $[a, b]$ est stable par g).
Si on écrit $I = [a, b]$ alors $g(I) \subset I$.
2. Supposons que g est une fonction contractante.

Alors il existe un unique $x^* \in [a, b]$ point fixe de g . De plus, la suite

$$\begin{cases} x_0 \in [a, b], \\ x_{m+1} = g(x_m), \quad m \geq 0, \end{cases} \quad (1)$$

converge vers x^* .

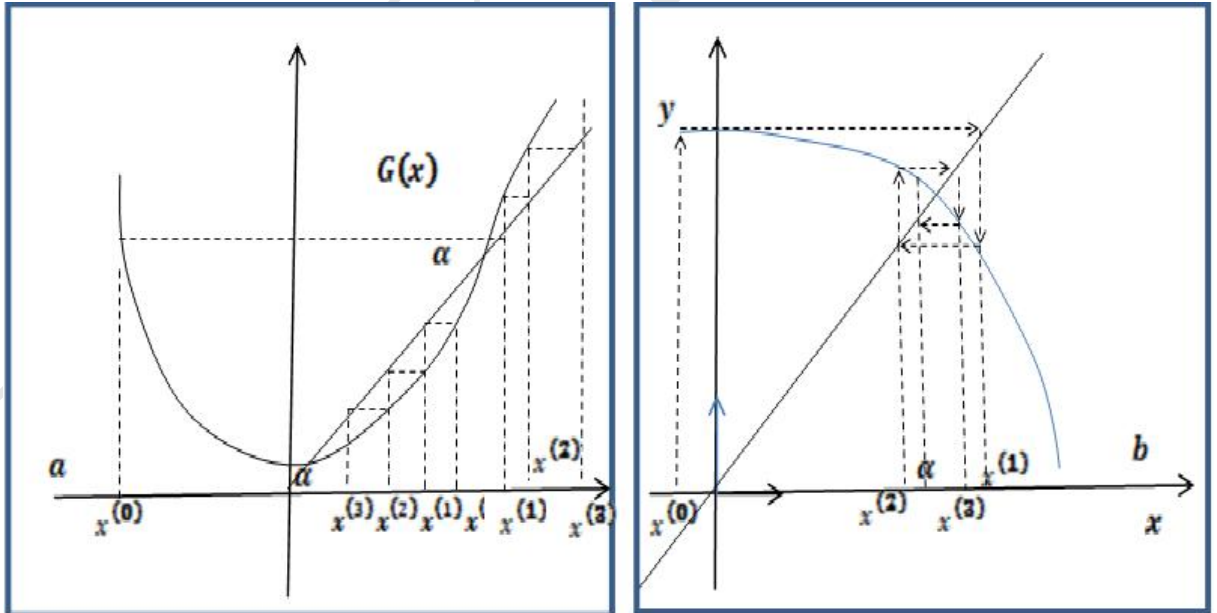


Figure 2.2: Méthode de point fixe.

Interprétation géométrique de la méthode du point fixe

Trouver l'intersection de la courbe de $g(x)$ avec la première bissectrice revient à résoudre l'équation : $x - g(x) = 0$.

Le principe du cheminement consiste à démarrer à partir d'un point x_0 , le segment portée par la droite $y = g(x_0)$ coupe la première bissectrice au point (x_1, x_1) , il suffit ensuite de faire une projection verticale pour obtenir $g(x_1)$ et ainsi de suite, sous les bonnes conditions ce procédé converge vers le point fixe.

Proof. 1-Unicité :

Si $x^*, y^* \in [a, b]$ sont deux points fixes d'une fonction g qui est contractante de rapport $l < 1$ alors :

$|x^* - y^*| = |g(x^*) - g(y^*)| \leq l |x^* - y^*|$. Ceci n'est possible que pour $|x^* - y^*| = 0$, donc $x^* = y^*$.

2-Existence :

Une récurrence facile montre que $|x_{m+1} - x_m| \leq l^m |x_1 - x_0|$ pour tout $m \in \mathbb{N}$, puis

$$\begin{aligned} |x_{m+p} - x_m| &\leq |x_{m+p} - x_{m+p-1}| + \dots + |x_{m+1} - x_m| \\ &\leq (l^{p-1} + \dots + l^0) |x_{m+1} - x_m| = \frac{1 - l^p}{1 - l} |x_{m+1} - x_m| \\ &\leq \frac{1}{1 - l} |x_{m+1} - x_m| \leq \frac{l^m}{1 - l} |x_1 - x_0| \end{aligned}$$

Pour tout $m, p \in \mathbb{N}$. La suite (x_m) est donc de Cauchy et d'où elle est convergente vers un point μ . Notons $\mu = \lim x_m$ sa limite et vérifions qu'il s'agit d'un point fixe. Puisque g est contractante, elle est continue. L'équation de récurrence $x_{m+1} = g(x_m)$ donne donc

$$\mu = \lim x_{m+1} = \lim g(x_m) = g(\lim x_m) = g(\mu).$$

□

Test d'arrêt: Le nombre minimum d'itérations pour que la solution soit approchée avec une précision ε si nous avons la relation suivante:

$$|x^* - x_m| < \varepsilon.$$

Sachant que

$$|x^* - x_m| \leq \frac{l^m}{1 - l} |x_1 - x_0|$$

donc,

$$m \geq \frac{\log \left[\frac{(1-l)\varepsilon}{|x_1 - x_0|} \right]}{\log(l)}.$$

Exemple 20. Utiliser le code Matlab pour calculer la solution du problème $g(x) = x$, où

$$g(x) = x^2 + \frac{1}{4}(e^{-x} + x^2), \quad x \in [0, 1].$$

Code Matlab

```
>> g=inline('(x^2)+ 1/4*(exp(-x)+(x^2))');
a=0;
b=1;
k=0;
E=1e-10; % epsilon
#####
x0=0;% Point initiale
x1=g(x0); % La première itération
while and(g(x0)~=0 , abs(x1-x0)>=E) %
x0=x1;
x1=g(x0);
k=k+1;
end
y=[x0,k]
```

y =

```
0.294808743210598 33.000000000000000
```

2.5 Méthode de Newton

On suppose que $f \in C^1(a, b)$. Écrivons le développement de TAYLOR de f en α au premier ordre : $f(\alpha) = f'(\xi)(\alpha - x) + f(x)$ où ξ est entre α et x . Le problème “chercher α tel que $f(\alpha) = 0$ ”, devient alors “chercher α tel que $f'(\xi)(\alpha - x) + f(x) = 0$ ”.

Cette équation conduit à la méthode itérative suivante : On appelle suite de **Newton** associée à f , la suite $\{x_k\}_{k \in \mathbb{N}}$ définit par:

$$\begin{cases} x_0 \in]a, b[, \\ x_{k+1} = x_k - (q_k)^{-1} f(x_k), \end{cases}$$

où q_k est une approximation de $f'(x_k)$.

Si on pose $q_k = f'(x_k)$ on obtient la méthode de **Newton**

$$\begin{cases} x_0 \in]a, b[, \\ x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}. \end{cases}$$

Interprétation géométrique de la méthode de NEWTON:

Pour calculer x_{k+1} par la méthode de **NEWTON** on prend l'intersection de l'axe des abscisses avec la droite tangente au graphe de f passant par le point $(x_k, f(x_k))$, *i.e.* on cherche x solution du système linéaire

$$\begin{cases} y = f'(x_k)(x - x_k) + f(x_k), \\ y = 0. \end{cases}$$

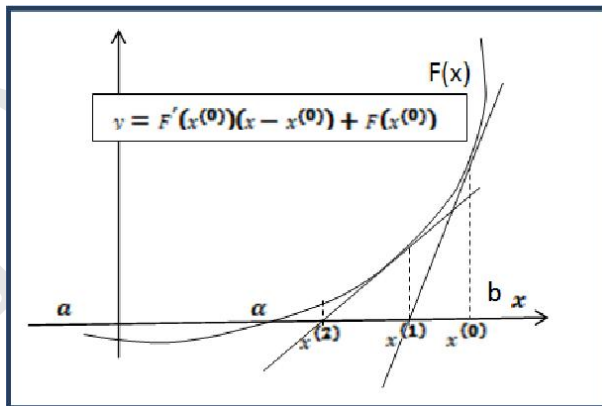


Figure 2.3: Méthode de Newton.

Théorème 5. Supposons que

1) $\alpha \in [a, b]$, où α est la racine cherchée.

2) $f \in C^2([a, b])$,

3) $f' \neq 0$ et $M = \frac{\max_{x \in [a, b]} |f''(x)|}{2 \min_{x \in [a, b]} |f'(x)|} < +\infty$, alors

$$|x_{k+1} - \alpha| \leq M (x_k - \alpha)^2, \quad k \geq 0$$

et

$$|x_k - \alpha| \leq M^{2^k - 1} |x_0 - \alpha|^{2^k} \leq k \geq 1.$$

Proof. Reprenons la formule de Newton:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k = 0, 1, 2, \dots$$

On a donc $x_{k+1} = g(x_k)$, où $g(x) = x - \frac{f(x)}{f'(x)}$.

Et comme

$$g'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} = \frac{f(x)f''(x)}{f'(x)^2}.$$

Alors, on a $f(\alpha) = 0$ cela implique que $g'(\alpha) = 0$. Donc on peut trouver un voisinage V de α tel que $|g'(x)| < 1$ pour $x \in V$, en plus $g(V) \subset V$. Conséquence: les conditions *i)* et *ii)* du théorème de point fixe sont alors réalisées, et par suite, on a convergence de la suite $(x_k)_k$ vers α pour un point initiale $x_0 \in V$.

Faisons maintenant un développement de Taylor de f à l'ordre 1 au voisinage de x_k . On obtient:

$$f(\alpha) = f(x_k) - f'(x_k)(\alpha - x_k) + \frac{1}{2}f''(\xi_k)(\alpha - x_k)^2,$$

en divisant par $f'(x_k) \neq 0$ et en translatant, on aura:

$$x_k - \frac{f(x_k)}{f'(x_k)} - \alpha = \frac{f''(\xi_k)(\alpha - x_k)^2}{2f'(x_k)}.$$

D'où, après simplification:

$$x_{k+1} - \alpha = \frac{f''(\xi_k)(\alpha - x_k)^2}{2f'(x_k)}.$$

Et comme $M = \frac{\max_{x \in [a, b]} |f''(x)|}{2 \min_{x \in [a, b]} |f'(x)|} < +\infty$ on obtient alors

$$|x_{k+1} - \alpha| \leq M (x_k - \alpha)^2, \leq M^3 (x_{k-1} - \alpha)^4 k \geq 0$$

2.5. MÉTHODE DE NEWTON

Maintenant en utilisant la dernière inégalité (par récurrence), on a :

$$|x_k - \alpha| \leq M^{2^k-1} |x_0 - \alpha|^{2^k} \leq \frac{C^{2^k}}{M}, \quad k \geq 1.$$

où $C = M(b - a)$. Donc, on peut voir que il y a une convergence vers α , si $C = M(b - a) < 1$, et c'est pour cette raison que l'on choisit $[a, b]$ de longueur $b - a < \frac{1}{M}$. \square

Remarque 8. *La méthode de Newton possède de grands avantages (convergence rapide), mais elle nécessite le calcul de la dérivée $f'(x)$ qui n'est pas toujours facile.*

On contourne cette difficulté en remplaçant $f'(x_n)$ dans l'algorithme de Newton par :

1-La méthode de la corde. Dans la méthode de la corde on pose $q_k = \frac{f(b) - f(a)}{b - a}$, on déduit la relation de récurrence suivante :

$$\begin{cases} x_0 \in]a, b[, \\ x_{k+1} = x_k - \frac{b - a}{f(b) - f(a)} f(x_k). \end{cases}$$

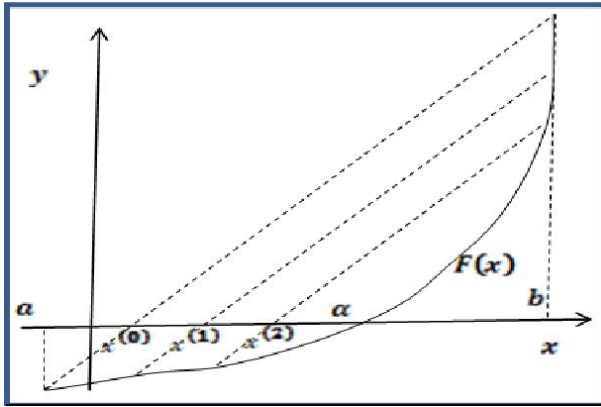


Figure 2.4: Méthode de la Corde.

2-La méthode de la sécante. Dans la méthode de la sécante on pose $q_k = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$, on déduit la relation de récurrence suivante:

$$\begin{cases} x_0, x_1 \in]a, b[, \\ x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} f(x_k). \end{cases}$$

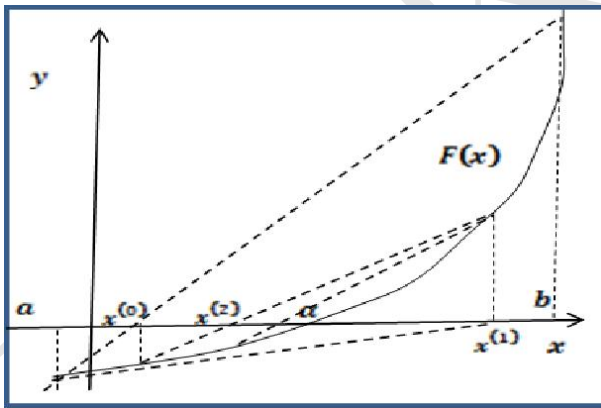


Figure 2.5: Méthode de la secante.

Dr. Khellaf

Résolution de systèmes linéaires

Le but de ce chapitre est l'étude qualitative et quantitative des systèmes linéaires sur \mathbb{R}^n . Nous allons étudier les différentes méthodes construites pour le calcul de la solution du système qu'on suppose qu'elle existe et est unique.

3.1 Présentation du problème

Soit $A = (a_{ij})_{1 \leq i, j \leq n}$ une matrice carrée dans $\mathcal{M}_n(\mathbb{R})$. A est **inversible** si elle vérifie l'une des conditions suivantes:

- $\det(A) \neq 0$,
- $rg(A) = \dim(\mathcal{R}(A)) = n$ la surjectivité,
- $Av = 0 \implies v = 0$, l'injectivité.

Notre **problème** consiste à trouver l'unique $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ tel que, pour

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^n \text{ fixé}$$

$$\mathbf{Ax} = \mathbf{b}.$$

Ce qui est équivalent au système d'équations suivant:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \vdots &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{cases}$$

3.2. MÉTHODES DIRECTES

Dans la plus part des cas, nous traitons des matrices inversibles. Nous ne faisons pas non plus de révision systématique de l'algèbre linéaire élémentaire que nous supposons connue.

La solution de l'équation $Ax = b$ est $x = A^{-1}b$, alors la discussion peut sembler close. Nous savons cependant que le calcul de la matrice inverse A^{-1} est plus difficile et plus long que la résolution du système linéaire de départ.

3.2 Méthodes directes

Dans cette partie nous allons étudier les méthodes directes de résolution des systèmes linéaires. Par le mot **directes**, on cherche à dire qu'elles nous permettent d'obtenir la solution exacte du système (affectée par les erreurs d'arrondi si le calcul s'effectue avec un ordinateur) après un nombre fini des opérations (addition, multiplication, division, soustraction).

3.2.1 Systèmes équivalents et systèmes particuliers

Soient $A, B \in \mathcal{M}(\mathbb{R}^n)$ et $b, c \in \mathbb{R}^n$. Les deux systèmes suivants:

$$(P) : \mathbf{Ax}=\mathbf{b}, \quad (\tilde{P}) : \mathbf{Bx}=\mathbf{c},$$

sont dit **équivalent** s'ils admettent la même solution.

Donc, notre but est de transformer le problème qu'on cherche à résoudre en un système facile à résoudre. Nous pouvons distinguer deux types de problèmes faciles à résoudre:

- Si A est **triangulaire inférieure**, c-à-d $a_{ij} = 0$ pour $j \geq i + 1$.

$$A = \begin{pmatrix} a_{11} & 0 & \dots & \dots & 0 \\ a_{21} & a_{22} & \ddots & & 0 \\ a_{i1} & \ddots & a_{ii} & \ddots & \vdots \\ \vdots & & \vdots & \ddots & 0 \\ a_{n1} & \dots & a_{n3} & 0 & a_{nn} \end{pmatrix}$$

- La solution de (P) est donnée par

$$\begin{aligned} x_1 &= \frac{b_1}{a_{11}}, \\ x_i &= \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j \right), \quad i \geq 2. \end{aligned}$$

- Si A est **triangulaire supérieure**, c-à-d $a_{ij} = 0$ pour $i \leq j - 1$.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ 0 & a_{22} & & & a_{2n} \\ \vdots & \ddots & a_{ii} & \dots & a_{in} \\ \vdots & & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 0 & a_{nn} \end{pmatrix}$$

La solution de (P) est donnée par

$$\begin{aligned} x_n &= \frac{b_n}{a_{nn}}, \\ x_i &= \frac{1}{a_{ii}} \left(b_i - \sum_{j=i+1}^n a_{ij} x_j \right), \quad i \leq n-1. \end{aligned}$$

3.2.2 Méthode d'élimination de Gauss

La méthode des **Pivots de Gauss** ou la méthode d'élimination de Gauss consiste à construire à partir de (P) : $Ax = b$, un problème équivalent à ce dernier mais avec une matrice triangulaire supérieure ou inférieure sachant que l'ensemble des solutions d'un système linéaire ne change pas si on effectue sur les équations les opérations élémentaires suivantes :

- Changer l'ordre des équations,
- Multiplier une équation par une constante non nulle,
- Ajouter à une équation une combinaison linéaire des autres équations.

Considérons une matrice inversible $A \in \mathcal{M}(\mathbb{R}^n)$ dans le terme diagonal a_{11} est supposé non nul. On pose $A^{(1)} = A$ et $b^{(1)} = b$. On introduit les multiplicateurs

$$m_{i1} = \frac{a_{i1}^{(1)}}{a_{11}^{(1)}}, \quad i = 2, 3, \dots, n,$$

où les $a_{ij}^{(1)}$ désignent les éléments de $A^{(1)}$. On peut éliminer l'inconnue x_1 des lignes $i = 2, 3, \dots, n$ en leur retranchant m_{i1} fois la première ligne et en faisant de même pour le membre de droite. on définit alors

$$\begin{aligned} l_i^{(2)} &= l_i^{(1)} - m_{i1} l_1^{(1)}, \quad i = 2, 3, \dots, n \Rightarrow a_{ij}^{(2)} = a_{ij}^{(1)} - m_{i1} a_{1j}^{(1)}, \quad i, j = 2, 3, \dots, n, \\ b_i^{(2)} &= b_i^{(1)} - m_{i1} b_1^{(1)}, \quad i = 2, 3, \dots, n. \end{aligned}$$

On obtient un nouveau système de la forme

$$\begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \dots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \dots & a_{nn}^{(2)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(2)} \\ \vdots \\ b_n^{(2)} \end{pmatrix},$$

3.2. MÉTHODES DIRECTES

qu'on le note $A^{(2)}x = b^{(2)}$ et qui est équivalent au système de départ. On peut à nouveau transformer ce système de façon à éliminer l'inconnue x_2 des lignes 3, ..., n . En poursuivant ainsi, on obtient une suite finie de systèmes

$$A^{(k)}x = b^{(k)}, \quad 1 \leq k \leq n.$$

Pour $k \geq 2$, la matrice $A^{(k)}$ est de la forme suivante

$$A^{(k)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & \dots & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & & & & a_{2n}^{(2)} \\ \vdots & & \ddots & & & \vdots \\ 0 & \dots & 0 & a_{kk}^{(k)} & \dots & a_{kn}^{(k)} \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nk}^{(k)} & \dots & a_{nn}^{(k)} \end{pmatrix}$$

où, on a supposé $a_{ii}^{(i)} \neq 0$ pour $i = 1, 2, \dots, k-1$. Il est clair que pour $k = n$ on obtient alors le système triangulaire supérieur $A^{(n)}x = b^{(n)}$ suivant

$$A^{(k)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \dots & \dots & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & & & & a_{2n}^{(2)} \\ \vdots & & \ddots & & & \vdots \\ 0 & & 0 & a_{kk}^{(k)} & \dots & a_{kn}^{(k)} \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & a_{nn}^{(k)} \end{pmatrix}.$$

Exemple 21. Utilisons la méthode de Gauss pour résoudre le système

$$A = \begin{pmatrix} 2 & 1 & 2 \\ 6 & 4 & 4 \\ 3 & 2 & -1 \end{pmatrix} \quad \text{et} \quad b = \begin{pmatrix} 10 \\ 26 \\ 4 \end{pmatrix}.$$

La matrice augmentée du système est

$$\begin{pmatrix} 2 & 1 & 2 & 10 \\ 6 & 4 & 4 & 26 \\ 3 & 2 & -1 & 4 \end{pmatrix}$$

Etape $n^\circ = 1$:

$$\begin{cases} l_2 - \frac{6}{2}l_1 \rightarrow l_2 \\ l_3 - \frac{3}{2}l_1 \rightarrow l_3 \end{cases}$$

$$\begin{pmatrix} 2 & 1 & 2 & 10 \\ 0 & 1 & -2 & -4 \\ 0 & \frac{1}{2} & -4 & -11 \end{pmatrix}$$

Etape $n^\circ = 2$:

$$l_3 - \frac{1}{2}l_2 \rightarrow l_3$$

$$\begin{pmatrix} 2 & 1 & 2 & 10 \\ 0 & 1 & -2 & -4 \\ 0 & 0 & -3 & -9 \end{pmatrix},$$

la solution du système est

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 3.$$

Remarque 9.

1- Si on veut résoudre plusieurs systèmes d'équations possédant la même matrice A mais à second membres bi-différents, il suffit d'ajouter tout les seconds membres à la matrice pour former une seule matrice augmentée.

2- On peut calculer l'inverse d'une matrice sans recourir à la méthode des cofacteurs de l'algèbre linéaire. La recherche de la matrice inverse revient à résoudre n système d'équations ayant la même matrice des coefficients A et ayant pour second membres les n colonnes de la matrice identité (**TD**) .

3.2.3 Factorisation LU

Soit $A \in \mathcal{M}(\mathbb{R}^n)$ tel qu'ils existent $L \in \mathcal{M}(\mathbb{R}^n)$ triangulaire inférieure et $U \in \mathcal{M}(\mathbb{R}^n)$ triangulaire supérieure qui vérifient: $A = LU$. Alors la résolution du problème $Ax = b$, $b \in \mathbb{R}^n$ revient à résoudre les deux systèmes faciles suivants:

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

Donc, notre problème est de connaitre sous quelle condition l'existence de L et U est assurée? Le théorème suivant nous donne la réponse.

Théorème 6. Soit $A \in \mathcal{M}(\mathbb{R}^n)$ inversible. La factorisation **LU** de A avec $l_{ii} = 1$, $1 \leq i \leq n$ existe et est unique si et seulement si les sous matrices principales $A_k = (a_{ij})_{1 \leq i, j \leq k}$, $1 \leq k \leq n-1$ de A sont inversibles.

Proof. Supposons que les sous matrices principales $A_k = (a_{ij})_{1 \leq i, j \leq k}$, $1 \leq k \leq n-1$ de A sont inversibles et montrons que la factorisation **LU** de A avec $l_{ii} = 1$, $1 \leq i \leq n$ existe et est unique. Pour cela, on procède par récurrence sur n :

Le résultat est triviale pour $n = 1$. Supposons que la propriété est vraie à l'ordre $n-1$, c-à-d la sous matrice principales A_{n-1} admet la factorisation suivante:

$$A_{n-1} = L_{n-1}U_{n-1},$$

et cherchons la décomposition de A sous la forme:

$$A = \begin{pmatrix} A_{n-1} & c \\ d^t & a_{nn} \end{pmatrix} = LU = \begin{pmatrix} L_{n-1} & 0 \\ l^t & 1 \end{pmatrix} \begin{pmatrix} U_{n-1} & u \\ 0^t & U_{nn} \end{pmatrix},$$

3.2. MÉTHODES DIRECTES

où, $c = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{n-1n} \end{pmatrix}$ et $d = \begin{pmatrix} a_{n1} \\ \vdots \\ a_{nn-1} \end{pmatrix}$. En calculant le produit de ces deux matrices et en identifiant par blocs les éléments de la matrice A , on obtient

$$\begin{aligned} L_{n-1}u &= c, \\ l^t U_{n-1} &= d^t \iff U_{n-1}^t l = d, \\ U_{nn} &= a_{nn} - l^t u. \end{aligned}$$

Nous savons que $0 \neq \det(A_{n-1}) = \det(L_{n-1})\det(U_{n-1})$. Donc l, u et U_{nn} existent et sont uniques. D'où le résultat souhaité. Inversement, supposons que A admet une factorisation LU avec $l_{ii} = 1$, $1 \leq i \leq n$. Alors,

$$\det(A) = \det(L)\det(U) = u_{11}u_{22} \dots u_{nn} \neq 0.$$

De ce fait,

$$\det(A_{n-1}) = \det(L_{n-1})\det(U_{n-1}) = u_{11}u_{22} \dots u_{n-1n-1} \neq 0.$$

De la même façon, en répétant pour $n-2, n-3, \dots, 1$, on obtient le résultat. \square

Pour calculer les matrices L et U il y a deux méthodes:

- La méthode de **Doolittle** fournit d'abord la k -ième ligne de U , puis la k -ième colonne de L , selon les formules : pour $k = 1, \dots, n$

$$\begin{aligned} u_{kj} &= a_{kj} - \sum_{r=1}^{k-1} l_{kr}u_{rj}, \quad j = k, \dots, n, \\ l_{ik} &= \frac{1}{u_{kk}} \left(a_{ik} - \sum_{r=1}^{k-1} l_{ir}u_{rk} \right), \quad i = k+1, \dots, n. \end{aligned}$$

- La méthode de **Crout** fournit d'abord la k -ième colonne de L puis la k -ième ligne de U , selon les formules : pour $k = 1, \dots, n$

$$\begin{aligned} l_{ik} &= a_{ik} - \sum_{r=1}^{k-1} l_{ir}u_{rk}, \quad i = k, \dots, n, \\ u_{kj} &= \frac{1}{l_{kk}} \left(a_{kj} - \sum_{r=1}^{k-1} l_{kr}u_{rj} \right), \quad j = k+1, \dots, n. \end{aligned}$$

3.2.4 Décomposition de Cholsky

Soit $A \in \mathcal{M}(\mathbb{R}^n)$. A est dite **symétrique** si $a_{ij} = a_{ji}$, pour tout $1 \leq i, j \leq n$. Elle est dite **définie positive** si

$$\forall x \in \mathbb{R}^n, \quad x^t A x \geq 0.$$

Soit $A \in \mathcal{M}(\mathbb{R}^n)$ symétrique et définie positive. Alors, il existe une matrice triangulaire supérieure H dont les termes diagonaux sont strictement positifs telle que $A = H^t H$. Les coefficients de H peuvent être calculés comme suit: $h_{11} = \sqrt{a_{11}}$ et pour $i = 2, \dots, n$

$$h_{ij} = \frac{1}{h_{ii}} \left(a_{ij} - \sum_{k=1}^{j-1} h_{ik} h_{jk} \right), \quad j = 1, \dots, i-1,$$

$$h_{ii} = \sqrt{\left(a_{ii} - \sum_{k=1}^{i-1} h_{ik}^2 \right)}.$$

3.3 Méthodes itératives

Les méthodes directes perdent leurs utilités lorsque les systèmes étudiés sont très grands ou lorsque la matrice est proche d'une matrice singulière. Les méthodes itératives consistent à construire une suite $\{x_n\}_{n \in \mathbb{N}}$ de \mathbb{R}^n qui converge vers la solution x du problème concerné.

3.3.1 Rayon spectrale

On appelle **rayon spectrale** de $A \in \mathcal{M}(\mathbb{R}^n)$, la quantité suivante:

$$\rho(A) = \max \{ |\lambda_i| : \lambda_i \text{ valeur propre de } A \}.$$

On dit qu'une norme matricielle $\|\cdot\|$ est **consistante** avec une norme vectorielle $\|\cdot\|_{\mathbb{R}^n}$ si elle vérifie,

$$\forall x \in \mathbb{R}^n, \quad \|Ax\|_{\mathbb{R}^n} \leq \|A\| \|x\|_{\mathbb{R}^n}.$$

Proposition 2. Si $\|\cdot\|$ est consistante, alors pour tout $A \in \mathcal{M}(\mathbb{R}^n)$:

$$\rho(A) \leq \|A\|.$$

Proof. Soit λ une valeur propre de $A \in \mathcal{M}(\mathbb{R}^n)$, alors il existe $v \in \mathbb{R}^n / \{0\}$ vecteur propre tel que : $Av = \lambda v$. On obtient,

$$|\lambda| \|v\|_{\mathbb{R}^n} = \|Av\|_{\mathbb{R}^n} \leq \|A\| \|v\|_{\mathbb{R}^n} \implies |\lambda| \leq \|A\|.$$

□

Proposition 3.

$$\rho(A) = \inf_{\|\cdot\| \text{ consistante}} \|A\|,$$

c-à-d $\forall \varepsilon > 0$, il existe $\|\cdot\|$ une norme matricielle consistante tel que

$$\|A\| \leq \rho(A) + \varepsilon.$$

3.3. MÉTHODES ITÉRATIVES

Proof. Voir □

Théorème 7. Soit $A \in \mathcal{M}(\mathbb{R}^n)$, alors

$$\lim_{n \rightarrow +\infty} \|A^n\| = 0 \iff \rho(A) < 1.$$

Proof. Supposons que $\lim_{n \rightarrow +\infty} \|A^n\| = 0$. Soit λ une valeur propre de A et $x \in \mathbb{R}^n / \{0\}$ son vecteur propre, c-à-d $Ax = \lambda x$ ce qui donne $A^n x = \lambda^n x$. Donc

$$\begin{aligned} |\lambda|^n \|x\|_{\mathbb{R}^n} = \|A^n x\|_{\mathbb{R}^n} &\leq \|A^n\| \|x\|_{\mathbb{R}^n} \implies |\lambda|^n \leq \|A^n\|, \\ &\implies \lim_{n \rightarrow +\infty} |\lambda|^n \leq \lim_{n \rightarrow +\infty} \|A^n\| = 0, \\ &\implies |\lambda| < 1. \end{aligned}$$

Inversement, supposons que $\rho(A) < 1$. Alors, il existe $\varepsilon > 0$ tel que $\rho(A) < 1 - \varepsilon$. Donc, d'après la proposition précédente, il existe une norme consistante $\|\cdot\|$ tel que:

$$\|A\| \leq \rho(A) + \varepsilon < 1.$$

D'où,

$$\lim_{n \rightarrow +\infty} \|A^n\| \leq \lim_{n \rightarrow +\infty} \|A\|^n = 0.$$

□

3.3.2 Suites consistantes

Soit la suite $\{x^{(k)}\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ définit par :

$$\begin{cases} x^0 & \text{donné,} \\ x^{k+1} & = Bx^k + c, \end{cases}$$

où, $B \in \mathcal{M}(\mathbb{R}^n)$. Cette suite est dite **consistante** si le système est équivalent au système $Ax = b$, c-à-d,

$$c = (I - B)A^{-1}b.$$

On met, $e^{(k)} = x^{(k)} - x$, $k \geq 0$. Donc,

$$\lim_{k \rightarrow +\infty} e^{(k)} = 0 \iff \lim_{k \rightarrow +\infty} x^{(k)} = x.$$

Cette propriété de consistance ne suffit pas à assurer la convergence d'une méthode itérative, comme le montre l'exemple suivant.

Exemple 22. On veut résoudre le système linéaire $2Ix = b$ avec la méthode itérative $x^{k+1} = -x^k + b$, qui est clairement consistante.

Cette suite n'est pas convergente pour une donnée initiale arbitraire.

Si par exemple $x^0 = 0$, la méthode donne $x^{2k} = 0$, $x^{2k+1} = b$, $k = 0, 1, \dots$

Si $x^0 = \frac{1}{2}b$ la méthode converge.

Théorème 8. Soit $\{x^{(k)}\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ une suite consistante avec le système $Ax = b$. Cette suite converge vers la solution x si et seulement si $\rho(B) < 1$.

Proof. Supposons que $\rho(B) < 1$, alors $e^{(k+1)} = Be^{(k)} \implies e^{(k)} = B^k e^{(0)}$. Donc, $\|e^{(k)}\|_{\mathbb{R}^n} \leq \|B^k\| \|e^{(0)}\|_{\mathbb{R}^n}$. On utilise le théorème précédent pour obtenir la convergence.

Pour démontrer le sens inverse, on utilise la contraposée: Supposons que $\rho(B) \geq 1$, alors il existe λ valeur propre de B tel que $|\lambda| \geq 1$ et v son vecteur propre. On met, $x^{(0)} = v + x$, alors

$$Be^{(0)} = \lambda e^{(0)} \implies e^{(k)} = \lambda^k e^{(0)}.$$

Comme $|\lambda| \geq 1$, $\lim_{k \rightarrow +\infty} e^{(k)}$ ne peut pas être 0. □

3.3.3 Méthodes de Jacobi

La méthode de **Jacobi** consiste à calculer la suite consistante $\{x^{(k)}\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ définit par :

$$\begin{cases} x^0 & \text{donné,} \\ x^{k+1} & = B_J x^k + c, \end{cases}$$

où,

$$\begin{aligned} B_J &= D^{-1}(L + U), \\ c &= D^{-1}b, \end{aligned}$$

tels que pour $1 \leq i, j \leq n$

$$D_{ij} = \begin{cases} a_{ii} & i = j, \\ 0 & i \neq j, \end{cases} = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_{nn}^{(2)} \end{pmatrix},$$

$$L_{ij} = \begin{cases} -a_{ij} & i > j, \\ 0 & i \leq j, \end{cases} = \begin{pmatrix} 0 & \dots & \dots & 0 \\ -a_{21} & 0 & \dots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -a_{n1} & \dots & -a_{nn-1} & 0 \end{pmatrix},$$

$$U_{ij} = \begin{cases} -a_{ij} & i < j, \\ 0 & i \geq j. \end{cases} = \begin{pmatrix} 0 & -a_{12} & \dots & -a_{1n} \\ 0 & 0 & \dots & \vdots \\ \vdots & \ddots & \ddots & -a_{n-1n} \\ 0 & \dots & 0 & 0 \end{pmatrix}.$$

On remarque que L et U sont respectivement la partie triangulaire inférieure et supérieure sans la diagonale de la matrice A , et D est la partie diagonale de la

3.3. MÉTHODES ITÉRATIVES

même matrice. On remarque que le calcul de D^{-1} est facile et elle est donnée par:

$$D_{ij} = \begin{cases} \frac{1}{a_{ii}} & i = j, \\ 0 & i \neq j. \end{cases}$$

Donc la formule générale de calcul par la méthode de Jacobi est

$$\begin{cases} x^0 & \text{donné} \\ x_i^{k+1} & = \frac{1}{a_{ii}} (b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^k), \quad i = 1 \dots n. \end{cases}$$

La méthode de Jacobi suppose des pivots a_{ii} non nuls sinon un simple échange de ligne suffit à résoudre le problème (pour avoir D inversible).

Exemple 23. Soit le système d'équations

$$\begin{cases} 4x + y + 2z = 12 \\ x + 8y + z = 20 \\ x + y - 5z = -12 \end{cases}$$

$$\begin{aligned} \begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix} &= \begin{pmatrix} 0 & -\frac{1}{4} & -\frac{1}{2} \\ -\frac{1}{8} & 0 & -\frac{1}{8} \\ \frac{1}{5} & \frac{1}{5} & 0 \end{pmatrix} \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} + \begin{pmatrix} \frac{3}{5} \\ \frac{5}{2} \\ \frac{12}{5} \end{pmatrix} \\ &= \begin{pmatrix} \frac{3}{5} & -\frac{1}{4}y_n & -\frac{1}{2}z_n \\ \frac{5}{2} & -\frac{1}{8}x_n & -\frac{1}{8}z_n \\ \frac{12}{5} & +\frac{1}{5}x_n & +\frac{1}{5}y_n \end{pmatrix} \end{aligned}$$

On peut obtenir ce système sans recourir à l'inversion matricielle avec la technique suivante

$$\begin{cases} 4x + y + 2z = 12 \\ x + 8y + z = 20 \\ x + y - 5z = -12 \end{cases} \Rightarrow \begin{cases} 4x = 12 - y - 2z \\ 8y = 20 - x - z \\ -5z = -12 - x - z \end{cases} \Rightarrow \begin{cases} x = \frac{3}{5} - \frac{1}{4}y - \frac{1}{2}z \\ y = \frac{5}{2} - \frac{1}{8}x - \frac{1}{8}z \\ z = \frac{12}{5} + \frac{1}{5}x + \frac{1}{5}z \end{cases}$$

Ensuite il suffit de considérer la suite

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix} = \begin{pmatrix} \frac{3}{5} & -\frac{1}{4}y_n & -\frac{1}{2}z_n \\ \frac{5}{2} & -\frac{1}{8}x_n & -\frac{1}{8}z_n \\ \frac{12}{5} & +\frac{1}{5}x_n & +\frac{1}{5}y_n \end{pmatrix}.$$

Maintenant, on cherche à construire une condition sur A pour que cette suite converge.

Définition 9. On dit qu'une matrice A est une **matrice diagonale dominante** si pour $1 \leq i \leq n$

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|.$$

Théorème 9. Soit $A \in \mathcal{M}(\mathbb{R}^n)$ une matrice à diagonale dominante. Alors la suite de **Jacobi** construite pour le système $Ax = b$ converge vers la solution.

Proof. Pour montrer que la suite de **Jacobi** converge vers la solution, il suffit de montrer que $\rho(B_J) < 1$. On a pour tout $x \in \mathbb{R}^n$, $y = B_J x$ et pour $1 \leq i \leq n$,

$$\max_{1 \leq i \leq n} |y_i| = \max_{1 \leq i \leq n} \left| \sum_{j=1}^n B_{ij} x_j \right| \leq \max_{1 \leq i \leq n} \sum_{j=1}^n |B_{ij}| \max_{1 \leq j \leq n} |x_j|.$$

Donc,

$$\|B_J\| = \sup_{\|x\|_\infty \neq 0} \frac{\|B_J x\|_\infty}{\|x\|_\infty} \leq \max_{1 \leq i \leq n} \sum_{j=1}^n |B_{ij}|.$$

Mais,

$$B_{ij} = \begin{cases} \frac{a_{ij}}{a_{ii}} & i \neq j, \\ 0 & i = j. \end{cases}$$

Donc, $\rho(B_J) \leq \|B_J\| < 1$. □

3.3.4 Méthode de Gauss-Seidel

La méthode de **Gauss-Seidel** consiste à calculer la suite consistante $\{x^{(k)}\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ définit par :

$$\begin{cases} x^0 & \text{donné,} \\ x^{k+1} & = B_{GS} x^k + c, \end{cases}$$

où,

$$\begin{aligned} B_{GS} &= (D - L)^{-1} U, \\ c &= (D - L)^{-1} b, \end{aligned}$$

avec, D , L et U sont les mêmes matrices construites pour la méthode de **Jacobi**. Donc la formule générale de calcul par la méthode de Jacobi est

$$\begin{cases} x^0 \text{ donné,} \\ x_i^{k+1} = \frac{1}{a_{ii}} (b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{k+1} - \sum_{j=i+1}^n a_{ij} x_j^k), \quad i = 1 \dots n. \end{cases}$$

Théorème 10. Soit $A \in \mathcal{M}(\mathbb{R}^n)$ une matrice à diagonale dominante. Alors la suite de **Gauss-Seidel** construite pour le système $Ax = b$ converge vers la solution.

Proposition 4. Soit $A \in \mathcal{M}(\mathbb{R}^n)$, $\{x^{(k)}\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ est une suite consistante définit par :

$$\begin{cases} x^0 & \text{donné,} \\ x^{k+1} & = Bx^k + c, \quad k \geq 1. \end{cases}$$

3.3. MÉTHODES ITÉRATIVES

Soit $\|\cdot\|$ une norme matricielle telle que :

$$\|B\| < 1.$$

Établir les inégalités suivantes :

$$\|e^{(k)}\| = \|x - x^{(k)}\| \leq \|B\|^k \|e^{(0)}\| = \|B\|^k \|x - x^{(0)}\|.$$

$$\|e^{(k)}\| = \|x - x^{(k)}\| \leq \frac{\|B\|^k}{1 - \|B\|} \|x^{(1)} - x^{(0)}\|.$$

De même pour $x^{(0)} = c$, on a :

$$\|e^{(k)}\| = \|x - x^{(k)}\| \leq \frac{\|B\|^{k+1}}{1 - \|B\|} \|c\|.$$

Cette dernière inégalité nous permet d'estimer à l'avance le nombre d'itérations nécessaires pour approcher x avec une précision donnée ε .

On résoud l'inégalité

$$\frac{\|B\|^{k+1}}{1 - \|B\|} \|c\| \leq \varepsilon.$$