ANALYSE NUMÉRIQUE I

Exercices corrigés

Dr. Hafidha SEBBAGH
Dr. Ibtissem DIDI

Ecole supérieure en sciences appliquées Tlemcen
2018

الجمـهـوريـة الجـزائـريـة الديمـقـراطـيـة الـشـعـبـيـة
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التـعـلـيـم العـالي و البـحـث العـلـمـي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للعلوم التطبيقية - تلمسان
Ecole Supérieure en Sciences Appliquées-Tlemcen

Manuscrit d'analyse numérique I
Exercices corrigés

Hafidha DALI YOUCEF née SEBBAGH
Docteur en mathématiques option analyse numérique
Ibtissem DIDI
Docteur en automatique option contrôle des processus

Préface

Ce manuscrit est destiné aux élèves inscrit en deuxième année cycle préparatoire, à l'école supérieure en sciences appliquées. Son contenu, correspond au programme officiel de la matière «Analyse numérique» du premier semestre. Il a été rédigé dans le but de permettre d'avoir un outil de travail et de référence recouvrant les connaissances qui leur sont demandés.

L'objet de ce manuscrit est de proposer une explication succincte des concepts vu en cours, en plus des exercices corrigés permettant d'orienter les raisonnements vers d'autres domaines (physique, économies, etc.), cela afin d'exhiber l'intérêt et l'omniprésence de l'analyse numérique au sens large.

Table des matières

Introduction générale

En mathématiques, il existe des problèmes qui ne peuvent pas être résolu analytiquement, par exemple, les équations algébriques de degré ≥ 5. Pour la résolution de ce type d'équations il a été prouver qu'il n'existe pas de formule par radicaux.

Aussi, dans certains cas les solutions analytiques existent mais sont complètement inefficaces à mettre en oeuvre en pratique, comme la résolution d'un système linéaire $Ax = b$ dont la solution est $x = A^{-1}b$ où le calcul de A^{-1} devient rapidement intolérable lorsque la taille de la matrice augmente. Donc souvent pour résoudre un problème mathématique on fait appel à l'analyse numérique.

L'*analyse numérique* comprend deux mots : l'analyse qui fait référence aux mathématiques et le mot numérique qui fait référence au traitement informatique.

En d'autres termes c'est l'élaboration des méthodes de calcul mathématiques adaptées aux traitement par ordinateur qui ont pour but la résolution des problèmes concrets qui se posent dans différentes disciplines : physique, economie, ingénierie, etc. Ces problèmes doivent être bien posés c'est-à-dire que la solution existe, unique et dépend continûment des données du système.

Il existe souvent plusieurs méthodes numériques pour résoudre un tel système qui doivent être stables et bien conditionné . Cela signifie que ces méthodes donnent des résultats satisfaisants quelques soit la nature des données initiales traités par l'ordinateur et qu'elles soient insensibles aux petites variations de ces données.

Développer une méthode numérique revient a créer un algorithme qui est un processus constitué par un ensemble d'opérations et de règles opératoires données pour un calcul.

Exemple d'Algorithme : Résolution de l'équation réelle : $ax + bx + c = 0$
Algorithme :

1. Lire a,b, et c
2. **Si** $a = 0$

    Si $b = 0$

        Si $c = 0$ **alors** il existe une infinité de solutions

        Sinon il n'y a pas de solution

        FinSi

    Sinon la solution est $x = \dfrac{-c}{b}$

    FinSi

Sinon calculer le discriminent $\Delta = b^2 - 4ac$

Si $\Delta < 0$ **alors** il n'y a pas de solutions réelles

FinSi

Si $\Delta = 0$ **alors** il existe une solution double $x = \dfrac{-b}{2a}$

FinSi

Si $\Delta > 0$ **alors** il existe deux solutions $x = \dfrac{-b-\sqrt{\Delta}}{2a}$ et $x = \dfrac{-b+\sqrt{\Delta}}{2a}$

FinSi

FinSi

3. Afficher la solution x

La résolution par des algorithmes numériques entraîne inévitablement l'introduction et la propagation d'erreurs d'arrondi. De plus, il est souvent nécessaire d'introduire d'autres erreurs liées au fait qu'un ordinateur ne peut effectuer que de manière approximative des calculs impliquant un nombre infini d'opérations arithmétiques. Par exemple, le calcul de la somme d'une série ne pourra être accompli qu'en procédant à une troncature convenable. On doit donc définir un problème numérique, dont la solution diffère de la solution mathématique exacte d'une erreur, appelée erreur de troncature [1].

L'objet de ce manuscrit est d'introduire les méthodes numériques pour la résolution des équations non linéaires (*Dichotomie*, *point fixe*, *Newton* et *sécante*) et des systèmes linéaires (méthodes directes : *élimination de Gauss*, *factorisation LU* et *Cholesky* ; méthodes itératives : *Jacobi*, *Gauss Seidel* et *relaxation*). On va rappeler dans un premier temps le principe de chaque méthode et par la suite nous allons traiter des exercices pour mieux comprendre ces méthodes.

Chapitre 1

Résolution des équations non linéaires

L'objet de ce chapitre est la résolution de l'équation $f(x) = 0$, où f est une fonction de $\mathbb{R}$ dans $\mathbb{R}$. Comme exemple d'équation non linéaire on a l'équation d'état d'un gaz. On veut déterminer le volume V occupé par un gaz de température T et de pression p. L'équation d'état (c'est-à-dire l'équation qui lie p, V et T) est

$$\left[p + a\left(\frac{N}{V}\right)^2\right](V - Nb) = kNT,$$

où a et b sont deux coefficients dépendants de la nature du Gaz, N est le nombre de molécules contenues dans le volume V et k est la constante de Boltzmann. Il faut donc résoudre une équation non linéaire d'inconnue V. Pour la résolution de ce type d'équation, les méthodes analytiques sont limitées à certaines formes algébriques $(a_n x^n + a_{n-1}x^{n-1} + ... + a_0)$, $n < 5$ ou formes particulières. Par conséquent pour les autres formes il faut utiliser les méthodes numériques pour trouver ou approcher les racines.

Ces méthodes là consistent d'abord à :

1. Localiser le (ou les) zéro(s) de f en procédant à l'étude du graphe de f, puis utiliser le théorème des valeurs intermédiaires (voir annexe) afin de trouver un intervalle qui contient une unique racine.
2. Trouver la solution en utilisant l'une des méthodes suivantes.

1.1 Méthode de dichotomie

La dichotomie est un mot grec qui signifie :

tomie : vient de couper
dicho : $\cdots$ en deux.

1.1.1 Principe de la méthode

On considère un intervalle $[a, b]$ et une fonction f de $[a, b]$ dans $\mathbb{R}$. On suppose que $f(a) \times f(b) < 0$ et que l'équation $f(x) = 0$, admet une unique solution $\alpha \in [a, b]$.

La méthode de dichotomie consiste à construire une suite (x_n) qui converge vers α de la manière suivante :

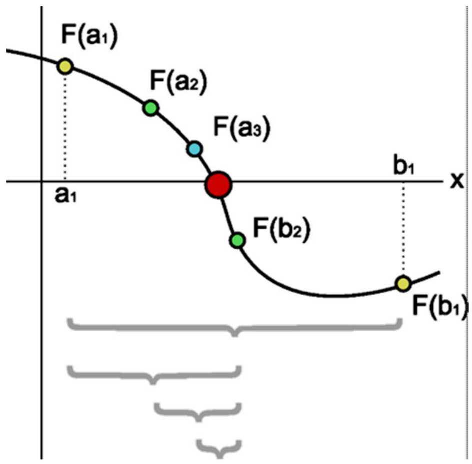
Soit c_0 le milieu de $[a, b]$. La racine se trouve alors dans l'un des deux intervalles $[a, c_0]$ ou $[c_0, b]$ où bien elle est égale à c_0.

- Si $f(a) \times f(c_0) < 0$, alors $\alpha \in \,]\,a, c_0\,[$. On pose $a_1 = a$, $b_1 = c_0$.
- Si $f(a) \times f(c_0) = 0$, alors $\alpha = c_0$.
- Si $f(a) \times f(c_0) > 0$, alors $\alpha \in \,]\,c_0, b\,[$. On pose $a_1 = c_0$, $b_1 = b$.

On prend alors c_1 le milieu de $[a_1, b_1]$. On construit ainsi une suite

$$c_0 = \frac{a+b}{2}, c_1 = \frac{a_1+b_1}{2}, \cdots, c_n = \frac{a_n+b_n}{2}$$

telle que $|a_n - b_n| < \epsilon$ avec $\epsilon > 0$.



Remarque 1 *Cet algorithme réduit à chaque pas l'amplitude de la localisation d'un facteur* 2. *L'erreur est donc réduite d'un facteur 2 à chaque itération. En* 20 *itérations par exemple, l'erreur sera* 10^{-6} *fois l'erreur initiale. Cette méthode est relativement lente. Par contre elle converge dans tous les cas où la fonction change de signe au voisinage de la racine. C'est une méthode que nous qualifierons tout-terrain, lente mais quasiment infaillible.*

1.1.2 Convergence de la méthode

Théorème 1 *Soit f une fonction continue sur $[a, b]$, vérifiant $f(a) \times f(b) < 0$ et soit $\alpha \in [a, b]$ l'unique solution de $f(x) = 0$. Si l'algorithme de dichotomie arrive jusqu'à l'étape n alors on a l'estimation :*

$$|\alpha - c_n| \leq \frac{b-a}{2^{n+1}}$$

Par conséquent la suite $(c_n)_{n\in\mathbb{N}}$ converge vers α. C'est ainsi vrai si $c_n = \alpha$ [2].

1.1.3 Test d'arrêt

Pour que la valeur de c_n de la suite de la $n^{\text{ième}}$ itération soit une valeur approchée de α à $\epsilon > 0$ près. Il suffit que n vérifie :

$$\frac{b-a}{2^{n+1}} \leq \epsilon$$

On a alors

$$|\alpha - c_n| \leq \frac{b-a}{2^{n+1}} \leq \epsilon$$

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

$$\frac{b-a}{2^{n+1}} \leq \epsilon \Longleftrightarrow \frac{b-a}{\epsilon} \leq \ 2^{n+1} \Longleftrightarrow n \geq \frac{\ln\left(\dfrac{b-a}{\epsilon}\right)}{\ln(2)} - 1$$

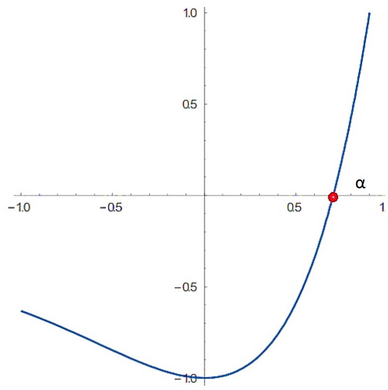
1.1.4 Exercices corrigés

Exercice 1 *On considère l'équation* $f(x) = x^2 \exp(x) - 1 = 0$.

1. *Trouver un intervalle* $I = [a, b]$ *de longueur 1 contenant la racine* α *de* $f(x)$.
2. *Calculer le nombre d'itérations nécessaires pour avoir la solution avec une précision* $\epsilon = 10^{-4}$.
3. *Calculer les cinq premiers itérés en utilisant la méthode de dichotomie.*

Solution

1. D'après le graphe de f, on remarque qu'il existe un $\alpha \in [0, 1]$.
   Vérifions le en utilisant le théorème des valeurs intermédiaires :
   $*f$ est continue sur $[0, 1]$.
   $*f(0) \times f(1) = (-1) \times (1.71) < 0$.
   $*f'(x) = x \exp(x)(2 + x) > 0$ sur $[0, 1]$.
   Donc d'après le théorème des valeurs intermédiaires il existe un unique $\alpha \in [0, 1]$.



2. On a
$$\frac{|b-a|}{2^{n+1}} < \epsilon$$
on a $a = 0$ et $b = 1$, donc pour $\epsilon = 10^{-3}$, on obtient :
$$n \geq \frac{\ln(10^3)}{\ln(2)} - 1$$
$$\Rightarrow n \geq 8,96$$

3. On va utiliser le tableau suivant pour calculer α :

n	a_n	b_n	$f(a_n)$	$f(b_n)$	$c_n = \dfrac{a_n + b_n}{2}$	$f(c_n)$
0	0	1	$\ominus$	$\oplus$	0.5	$\ominus$
1	0.5	1	$\ominus$	$\oplus$	0.75	$\oplus$
2	0.5	0.75	$\ominus$	$\oplus$	0.625	$\ominus$
3	0.625	0.75	$\ominus$	$\oplus$	0.6875	$\ominus$
4	0.6875	0.75	$\ominus$	$\oplus$	0.71875	$\oplus$
5	0.6875	0.71875	$\ominus$	$\oplus$	0.703125	$\ominus$
6	0.703125	0.71875	$\ominus$	$\oplus$	0.7109375	$\ominus$
7	0.703125	0.7109375	$\ominus$	$\oplus$	0.70703125	$\ominus$
8	0.703125	0.70703125	$\ominus$	$\oplus$	0.705078125	$\ominus$
9	0.703125	0.705078125	$\ominus$	$\oplus$	0.704101562	$\ominus$
10	0.703125	0.704101562	$\ominus$	$\oplus$	0.70387031	$\ominus$
11	0.703125	0.70387031	$\ominus$	$\oplus$	0.703497655	$\ominus$

À partir de ce tableau, on a :

$$|b_{11} - a_{11}| = 7.45 \times 10^{-4} < 10^{-3}$$

Donc

$$\alpha \simeq 0.703497655$$

Exercice 2 *Soit* $f(x) = x^{12} - 1.10$.

1. *Donner une approximation de* $(1.10)^{\frac{1}{12}}$ *dans l'intervalle* $[1, 1.1]$, *en utilisant la méthode de dichotomie avec une tolérance de* 10^{-3}.
2. *Calculer le nombre d'itérations nécessaires pour avoir la solution avec une précision de* 10^{-3}.

Solution

1. On a :
   *$f(1) \times f(1.1) = -0.10 \times 2.038 < 0$
   *$f'(x) = 12x^{11} > 0$ sur $[1, 1.1]$
   *f continue sur $[1, 1.1]$
   Donc d'après le théorème des valeurs intermédiaires, il existe une racine dans $[1, 1.1]$.
   On va utiliser le tableau suivant pour estimer la solution.

n	a_n	b_n	$f(a_n)$	$f(b_n)$	$c_n = \dfrac{a_n + b_n}{2}$	$f(c_n)$
0	1	1.10	$\ominus$	$\oplus$	1.05	$\oplus$
1	1	11.05	$\ominus$	$\oplus$	1.025	$\oplus$
2	1	1.025	$\ominus$	$\oplus$	1.0125	$\oplus$
3	1	1.0125	$\ominus$	$\oplus$	1.00625	$\ominus$
4	1.00625	1.0125	$\ominus$	$\oplus$	1.00937	$\oplus$
5	1.00625	1.00937	$\ominus$	$\oplus$	1.00781	$\ominus$
6	1.00781	1.00937	$\ominus$	$\oplus$	1.00859	$\oplus$
7	1.00781	1.00859	$\ominus$	$\oplus$	1.00820	$\oplus$

À partir de ce tableau, on a :

$$|b_7 - a_7| = 7.8 \times 10^{-4} < 10^{-3}$$

Donc

$$(1.10)^{\frac{1}{12}} \simeq 1.00820$$

2. on a :

$$n \geq \frac{1}{2} \ln\left(\frac{0.1}{\epsilon}\right) - 1$$

$$n \geq 5.64$$

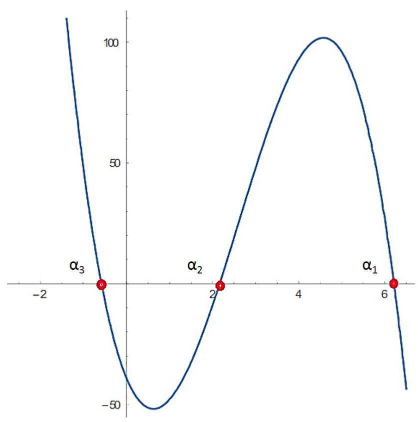
Cela veut dire qu'on a au minimum 6 itérations pour estimer la solution à 10^{-3} près.

Exercice 3 *Soit $f(x) = -39 - 43x + 39x^2 - 5x^3$.*

1. *Montrer que l'équation $f(x) = 0$ admet des racines et isoler chacune d'elles dans un intervalle. .*
2. *Calculer la solution $x \in [2, 3]$ par la méthode de dichotomie pour un $\epsilon = 10^{-3}$.*

Solution

1. $D_f = \mathbb{R}$
$f'(x) = -43 + 78x - 15x^2$
Calculons le Δ pour résoudre l'équation $f'(x) = 0$.
$\Delta = (-78)^2 - 4(-43)(-15) = 3504$
donc, on a :



$$\begin{cases} x_1 = \dfrac{-78 - \sqrt{3504}}{-30} = 4.57 \\ x_2 = \dfrac{-78 + \sqrt{3504}}{-30} = 0.62 \end{cases}$$

Alors on a le tableau de variations suivant :

x	$-\infty$		0.62		4.57		$+\infty$
$f'(x)$		$-$	0	$+$	0	$-$	
$f(x)$	$+\infty$	$\searrow$	-51.86	$\nearrow$	101.7	$\searrow$	$-\infty$

D'après tableau de variations de f et le graphe de $f(x)$, on remarque qu'il existe deux solutions positives et une solution négative.

$$\alpha_1 \in [2,3],\ \alpha_2 \in [5,7] \text{ et } \alpha_3 \in [-1,0]$$

2. On a :
*$f(2) \times f(3) = -9 \times 48 < 0$
*$f'(x) > 0$ sur $[2,3]$
*f continue sur $[2,3]$
Donc d'après le théorème des valeurs intermédiaires, il existe une unique solution $\alpha_1 \in [2,3]$.

n	a_n	b_n	$f(a_n)$	$f(b_n)$	$c_n = \dfrac{a_n + b_n}{2}$	$f(c_n)$
0	2	3	$\ominus$	$\oplus$	2.5	$\oplus$
1	2	2.5	$\ominus$	$\oplus$	2.25	$\oplus$
2	2	2.25	$\ominus$	$\oplus$	2.125	$\ominus$
3	2.125	2.25	$\ominus$	$\oplus$	2.1875	$\oplus$
4	2.125	2.1875	$\ominus$	$\oplus$	2.15625	$\ominus$
5	2.15625	2.1875	$\ominus$	$\oplus$	2.171875	$\oplus$
6	2.15625	2.171875	$\ominus$	$\oplus$	2.1640625	$\ominus$
7	2.1640625	2.171875	$\ominus$	$\oplus$	2.16796875	$\oplus$
8	2.164625	2.16796875	$\ominus$	$\oplus$	2.166015625	$\oplus$
9	2.164625	2.166015625	$\ominus$	$\oplus$	2.165038813	$\ominus$
10	2.165038813	2.166015625	$\ominus$	$\oplus$	2.165527219	$\ominus$

A partir de ce tableau, on voit que

$$|b_{10} - a_{10}| = 9.768 \times 10^{-4} < 10^{-3}$$

donc

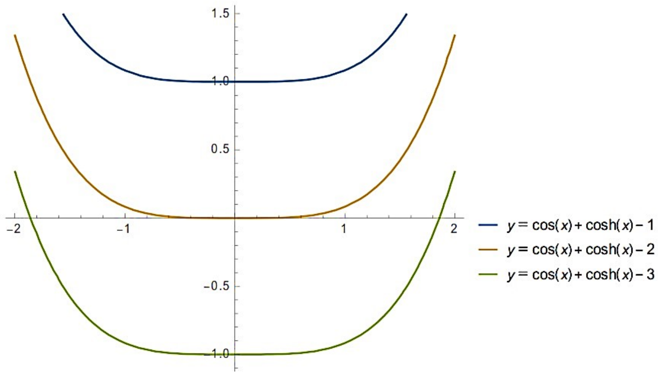
$$\alpha_1 \simeq 2.165527219$$

Exercice 4 *Soit la fonction* $f_\gamma(x) = cosh(x) + cos(x) - \gamma$.

1. *Pour* $\gamma = 1, 2, 3$, *étudier la fonction* f *et trouver graphiquement un intervalle qui contient le zéro de* f_γ.
2. *Calculer les zéros positifs de*f_γ *par la méthode de dichotomie avec une tolérance de* 10^{-3}.

Solution

Étude de f_γ :
On a
$\cosh(x) = \dfrac{\exp(x) + \exp(-x)}{2}$
$\sinh(x) = \dfrac{\exp(x) - \exp(-x)}{2}$
alors,
* $\lim_{x \to \pm\infty} f_\gamma(x) = +\infty$
*$f'_\gamma(x) = \sinh(x) - \sin(x) = 0 \Longleftrightarrow x = 0$



Rappellons nous que :
Pour $x > 0$, on a $\sin(x) < x < \sinh(x)$
et pour $x < 0$, on a $\sinh(x) < x < \sin(x)$.

*$f''_\gamma(x) = \cosh(x) - \cos(x) > 0$ pour tout $x \neq 0$.

Donc d'après le graphe de f_γ, on a :
— Pour $\gamma = 1$, la fonction n'a pas de zéro réel,
— Pour $\gamma = 2$, la fonction a un unique zéro $\alpha = 0$ qui est de multiplicité quatre (c'est-à-dire : $f_2(\alpha) = f'_2(\alpha) = f''_2(\alpha) = f_2^{(3)}(\alpha) = 0$ et $f_2^{(4)}(\alpha) \neq 0$),
— Pour $\gamma = 3$, la fonction a deux zéros distincts, un dans l'intervalle $[-2, -1]$ et l'autre dans $[1, 2]$.

Méthode de dichotomie :
— Pour $\gamma = 1$, la fonction n'a pas de zéro réel,
— Pour $\gamma = 2$, on ne peut pas utiliser la méthode de dichotomie pour trouver un intervalle $[a, b]$ sur lequel $f_2(a) \times f_2(b) < 0$.
— Pour le cas $\gamma = 3$, on va calculer le zéro qui se trouve dans l'intervalle $[1, 2]$, pour cela utilisons le tableau suivant :

n	a_n	b_n	$f(a_n)$	$f(b_n)$	$c_n = \dfrac{a_n + b_n}{2}$	$f(c_n)$
0	1	2	$\ominus$	$\oplus$	1.5	$\ominus$
1	1.5	2	$\ominus$	$\oplus$	1.75	$\oplus$
2	1.5	1.75	$\ominus$	$\oplus$	1.625	$\oplus$
3	1.5	1.625	$\ominus$	$\oplus$	1.5625	$\ominus$
4	1.5625	1.625	$\ominus$	$\oplus$	1.59375	$\ominus$
5	1.59375	1.625	$\ominus$	$\oplus$	1.609375	$\ominus$
6	1.609375	1.625	$\ominus$	$\oplus$	1.6171875	$\ominus$
7	1.6171875	1.625	$\ominus$	$\oplus$	1.62109375	$\ominus$
8	1.62109375	1.625	$\ominus$	$\oplus$	1.623046875	$\ominus$
9	1.623046875	1.625	$\ominus$	$\oplus$	1.624023438	$\ominus$
10	1.624023438	1.625	$\ominus$	$\oplus$	1.624511719	$\ominus$

À partir de ce tableau, on a

$$|b_{10} - a_{10}| = 9.765 \times 10^{-4} < 10^{-3}$$

donc

$$\alpha \simeq 1.624511719$$

Exercice 5 *On considère le problème du calcul de $l \in [0, \pi]$ tel que $l = 1 - \frac{1}{4}\cos(l)$.*

1. *Montrer qu'on peut utiliser la méthode de dichotomie pour approcher l.*
2. *Que vaut l'approximation de l après 3 itérations ? Quelle est l'erreur maximale qu'on obtient après 3 itérations ?.*

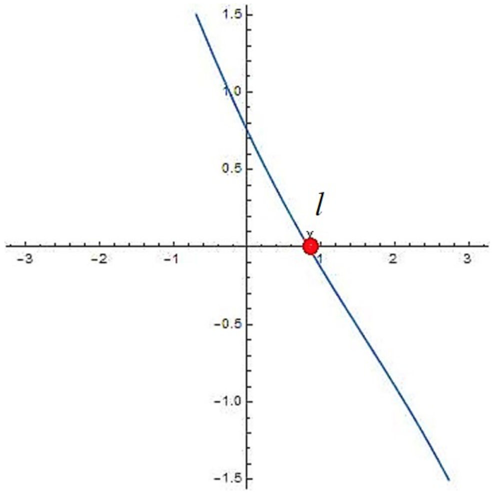
Solution

1. Soit $f : [0, \pi] \to \mathbb{R}$ la fonction définie par

$$f(x) = 1 - \frac{1}{4}\cos(x) - x$$

on a :
* $f \in C^{\infty}$.
* $f(0) \times f(\pi) = 0.75 \times (-1.89) < 0$.
* $f'(x) = \frac{1}{4}\sin(x) - 1 < 0$.

Donc d'après le théorème des valeurs intermédiaires, il existe un unique l tel que $f(l) = 0$. Alors on peut utiliser la méthode de dichotomie pour approcher l.
On remarque aussi que le graphe de la fonction f nous montre qu'il existe un unique zéro dans l'intervalle $[0, \pi]$.



2. Utilisons le tableau suivant pour calculer les trois itérations :

n	a_n	b_n	$f(a_n)$	$f(b_n)$	$c_n = \frac{a_n + b_n}{2}$	$f(c_n)$
0	0	π	$\oplus$	$\ominus$	$\pi/2$	$\ominus$
1	0	$\pi/2$	$\oplus$	$\ominus$	$\pi/4$	$\oplus$
2	$\pi/4$	$\pi/2$	$\oplus$	$\ominus$	$3\pi/8$	$\ominus$
3	$\pi/4$	$3\pi/8$	$\oplus$	$\ominus$	$5\pi/16$	

À partir de ce tableau, on a $l \approx \frac{5\pi}{16}$. L'erreur qu'ont obtient après trois itérations est au plus égale à la largeur de l'intervalle $[a_3, b_3]$, c'est-à-dire inférieure à $\frac{b-a}{2^3} = \frac{\pi}{8}$.

1.1.5 Exercices supplémentaires

Exercice 6 *On veut calculer les racines de l'équation*

$$f(x) = \frac{x}{2} - \sin(x) + \frac{\pi}{6} + \frac{\sqrt{3}}{2} = 0$$

dans l'intervalle $\left[-\frac{\pi}{2}, \pi\right]$. Peut-on appliquer la méthode de dichotomie pour calculer les racines de cette équation ? Pourquoi ? Dans le cas où c'est possible, estimer *le nombre minimal d'itérations nécessaires pour calculer le(s) zéro(s) avec une tolérance de $\epsilon = 10^{-10}$, après avoir choisi un intervalle convenable.*

Exercice 7 *(Fond d'investissement)*
Un compte d'épargne donne un taux $T \in [0,1]$ d'intérêt par an avec un virement annuel d'intérêts sur le compte. Cela signifie que si le premier janvier 2017 on met v euros sur ce compte, à la fin de la $n^{\text{ème}}$ année (i.e. au 31 décembre 2017+n) on en retire $v\,(1+T)^n$ euros. On décide alors d'ajouter au début de chaque année encore v euros. Cela signifie qui si on verse ces v euros le premier janvier 2017+m avec $0 < m < n$, au 31 décembre 2017+n ajoutent $v\,(1+T)^{n-m}$ euros. Si à la fin de la $n^{\text{ième}}$ année on en retire un capital de $M > v$ euros. Quel est le taux d'intérêts annuel de cet investissement ?

Indication :
À la fin de la $n^{\text{ième}}$ année, le capital versé au premier janvier 2017 est devenu $v\,(1+T)^n$, celui versé au premier janvier 2018 est devenu $v\,(1+T)^{n-1}$, par conséquent, le capital final M est relié aux taux d'intérêts annuel T par la relation

$$M = v\sum_{k=1}^{n}(1+T)^k = v\frac{(1+T)^n - 1}{(1+T)-1} = v\frac{1+T}{T}\left((1+T)^n - 1\right).$$

On en déduit que T est racine de l'équation algébrique non linéaire $f(T) = 0$ où

$$f(T) = v\frac{1+T}{T}\left((1+T)^n - 1\right).$$

Donc il reste à étudier la fonction $f(T)$.
Supposons que v=1000 euros et qu'après 5 ans M=6000 euros. On calcul la solution en utilisant la méthode de dichotomie à 10^{-12} près.

1.2 Méthode de point fixe

1.2.1 Principe de la méthode

Le principe de cette méthode consiste à transformer l'équation $f(x) = 0$ en une équation équivalente $g(x) = x$ où g est une fonction bien choisie. Ceci est toujours possible en posant $g(x) = x - f(x)$. Le point α est alors un point fixe de g.
Approcher les zéros de f revient à approcher les points fixe de g. Le choix de la fonction g est motivé par les exigences du théorème du point fixe.

1.2.2 Convergence de la méthode

Théorème 2 *(**Théorème du point fixe**) (condition suffisante de convergence globale)*
Soit g une fonction de classe C^1 sur un intervalle $[a,b] \subset \mathbb{R}$ telle que :

1. *$g(x) \in [a,b], \forall x \in [a,b]$ (on dit que l'intervalle $[a,b]$ est stable par $g(x)$. Si on écrit $I = [a,b]$, alors $g(I) \subset I$.*

2. $\exists k \in \mathbb{R}, 0 < k < 1$ *tel que* $|g'(x)| \leq k < 1, k = \max_{x\in[a,b]} |g'(x)|$. *On dit que g est strictement contractante*

Alors

— *g admet un unique point fixe dans* $I = [a, b]$.
— *Pour tout* $x_0 \in [a, b]$, *la suite* $(x_n)_{n\in\mathbb{N}}$ *définie par la récurrence :*

$$x_{n+1} = g(x_n)$$

Vérifie :

$$\forall n \in \mathbb{N}, x_n \in [a, b] \text{ et } \lim_{n\to+\infty} x_n = \alpha$$

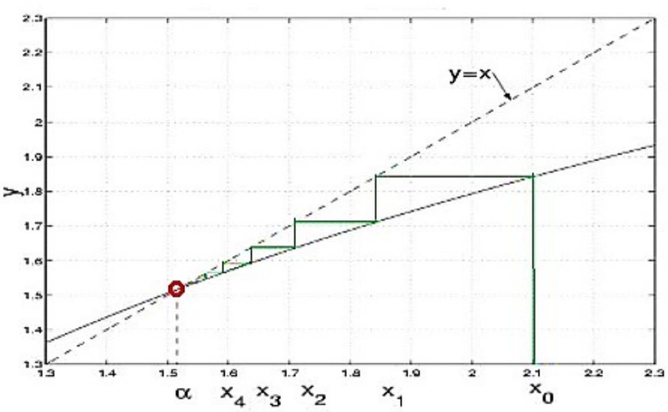
Théorème 3 *(Condition de convergence locale)*
Soient $\alpha \in \mathbb{R}$ *et* $g(x)$ *une fonction de classe* C^1 *définie sur un voisinage de* α.
Pour $x_0 \in \mathbb{R}$ *donné. On définit la suite* $(x_n)_n$ *par la récurrence* $x_{n+1} = g(x_n)$. *Alors si* $g(\alpha) = \alpha$ *et* $|g'(x)| < 1$, *il existe un* $h > 0$ *tel que pour tout* x_0 *appartenant à l'intervalle* $I = [\alpha - h, \alpha + h]$, *on a :*

$$\forall n \in \mathbb{N}, x_n \in I \text{ et } \lim_{n\to+\infty} x_n = \alpha$$

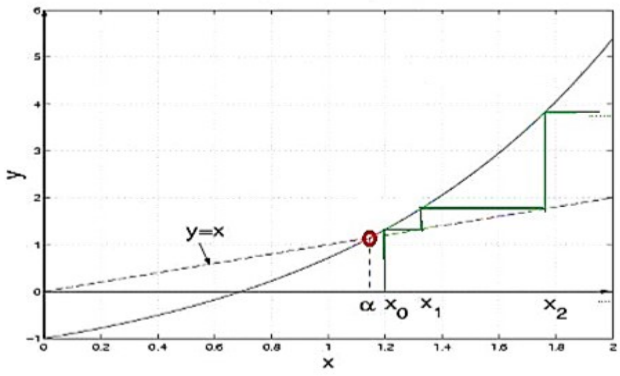
Enfin, α *est l'unique solution de l'équation* $g(x) = x$ *sur* I.

Corollaire 1 *Soit* α *une solution de l'équation* $g(\alpha) = \alpha$ *et* g' *continue au voisinage de* α, *alors on a les trois cas suivants :*

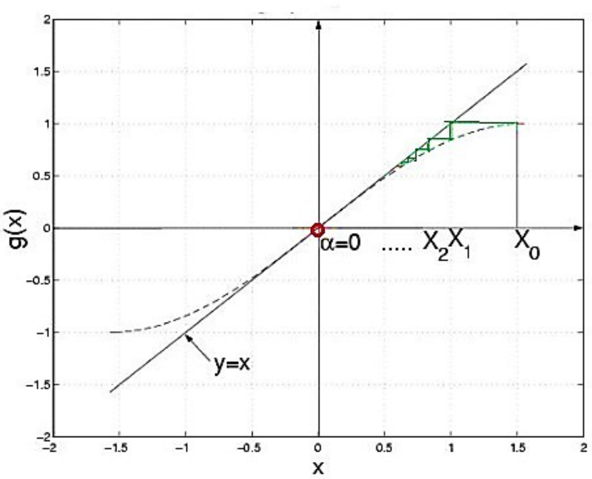
— ***Point fixe attractif :*** *Si* $|g'(x)| < 1$ *alors il existe un intervalle* $[a, b]$ *contenant* α *pour lequel* $\forall x_0 \in [a, b]$ *la suite* $(x_n)_{n\in\mathbb{N}}$ *définie par* $x_{n+1} = g(x_n)$ *converge vers* α.



— ***Point répulsif :*** *Si* $|g'(x)| > 1$ *alors* $\forall x_0 \neq 0$ *la suite* $(x_n)_{n\in\mathbb{N}}$ *définie par* $x_{n+1} = g(x_n)$ *ne converge pas vers* α.



— ***Point fixe douteux :*** *Si $|g'(x)| = 1$, on ne peut pas conclure, il peut y avoir convergence ou divergence. Pour cela on doit trouver un bon g qui vérifie $|g'(x)| < 1$.*



Remarque 2 *Lorsque α est un point répulsif de g, celle-ci devient bijective au voisinage de α et $\left|(g^{-1})'(\alpha)\right| = \dfrac{1}{|g'(\alpha)|} < 1$. Par conséquent, le point α devient un point attractif de g^{-1}. En effet, si $|g'(\alpha)| > 1$ alors g est strictement monotone au voisinage de α.*

1.2.3 Test d'arrêt

On a la suite (x_n) converge vers α tel que $g(\alpha) = \alpha$. En fixant la tolérance ϵ on estime qu'on atteint la précision ϵ dès qu'il existe un $n_0 \in \mathbb{N}$ tel que :

$$|x_{n_0+1} - x_{n_0}| < \epsilon$$

Néanmoins, la situation devient plus concrète lorsque g' est négative au voisinage de α. En effet :

Proposition 1 *Soit $g : [a,b] \longrightarrow [a,b]$ de classe C^1. On suppose que g admet un unique point fixe $\alpha \in [a,b]$ vérifiant $-1 < g'(x) < 0$ pour tout x dans l'intervalle de convergence V_α de α. Soit la suite (x_n) définie par :*

$$\begin{cases} x_0 \in V_\alpha \\ x_{n+1} = g(x_n); \quad \forall n \geq 0 \end{cases}$$

Alors

$$\forall n \in \mathbb{N}, |x_{n+1} - \alpha| \leq |x_{n+1} - x_n|$$

Par conséquent, soit n_0 tel que $|x_{n_0} - \alpha| < \epsilon$, alors (x_{n_0}) approche α à ϵ près.

Pour la démonstration (voir [2]).

Estimation de l'erreur :

Le nombre minimal d'itérations pour que la solution soit approchée avec une précision ϵ est :

$$|x_n - \alpha| < \epsilon$$

Sachant que

$$|x_n - \alpha| \leq |x_1 - x_0| \frac{k^n}{1-k}$$

Donc

$$n > \frac{ln\left[\frac{(1-k)\epsilon}{|x_1 - x_0|}\right]}{ln(k)}, k = \max_I |g'(x)|$$

1.2.4 Ordre de convergence

Définition 1 *On dit que la convergence de (x_n) vers α est d'ordre p ($p \in \mathbb{R}_+^*$) s'il existe une constante $c > 0$ telle que*

$$\lim_{n \to +\infty} \frac{e_{n+1}}{e_n^p} = c \text{ avec } e_n = |x_n - \alpha|$$

Définition 2

- *Lorsque $p = 1$, la convergence de x_n vers α est dite linéaire.*
- *Lorsque $p = 2$, la convergence de x_n vers α est dite quadratique.*

Théorème 4 *Si la suite $(x_n)_n$, définie par $x_{n+1} = g(x_n)$ converge vers α et si g est suffisamment dérivable au voisinage de α, alors l'ordre de la méthode est donné par :*

$$\begin{cases} g'(\alpha) = g''(\alpha) = ...g^{(p-1)}(\alpha) = 0 \\ g^{(p)}(\alpha) \neq 0 \end{cases}$$

De plus on :

$$\lim_{n \to +\infty} \frac{e_{n+1}}{e_n^p} = \frac{g^{(p)}(\alpha)}{p!}$$

Donc on remarque que plus p est grand plus la vitesse de la convergence est rapide.

Pour la démonstration (voir [3]).

Concernant la méthode du point fixe on a $p = 1$ donc la convergence est linéaire, car on a :

$$e_{n+1} = x_{n+1} - \alpha = g(x_n) - g(\alpha) = g'(\alpha)e_n + e_n\epsilon(n)$$

où $\epsilon(n) \to 0$ quand $n \longrightarrow +\infty$.

Alors

$$\lim_{n \to +\infty} \frac{e_{n+1}}{e_n} = g'(\alpha)$$

Donc dès que $g'(\alpha) \neq 0$ la convergence est d'ordre 1.

Remarque 3 *La convergence de la méthode du point fixe est de type linéaire donc lente et souvent difficile à réaliser en pratique. De plus certains points fixes dits instables ne peuvent être atteints.*

1.2.5 Exercices corrigés

Exercice 8 *Pour approcher les racines réelles de la fonction $f : \mathbb{R} \to \mathbb{R}$ par*

$$f(x) = \left(x^2 - 3x + 2\right)\exp(x)$$

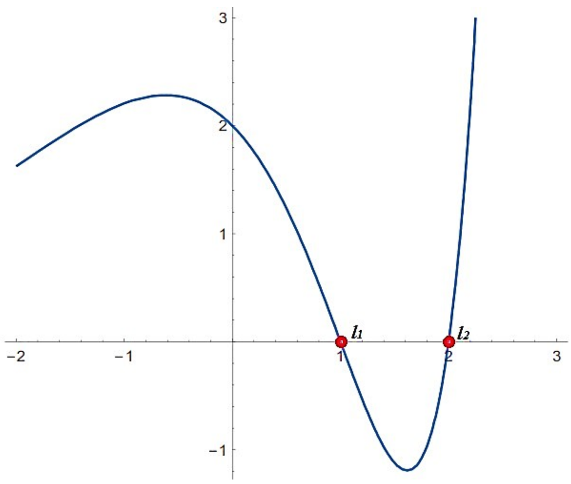
On veut utiliser la méthode du point fixe suivante :

$$\begin{cases} x_0 \text{ donné} \\ x_{n+1} = \varphi(x_n) \text{ pour tout } n \in \mathbb{N} \quad \text{où } \varphi(x) = \dfrac{x^2+2}{3} \end{cases}$$

1. *Montrer qu'il existe deux racines réelles $l_1 < l_2$ de f.*
2. *Faire l'étude graphique de la convergence de la méthode de point fixe et montrer que :*
   - *Si $x_0 \in]-2, 2[$ alors la suite converge vers l_1.*
   - *Si $x_0 = \pm 2$ alors $x_n = l_2$ pour tout $n \in \mathbb{N}$.*
   - *Si $x_0 < -2$ où $x_0 > 2$ alors la suite diverge vers $+\infty$.*
3. *Notons par $[a, b]$ l'intervalle maximal contenant l_1 pour lequel le théorème de point fixe s'applique. Calculer a et b et expliquer pourquoi la suite converge vers l_1 même si $x_0 \in]-2, 2[\setminus[a, b]$.*

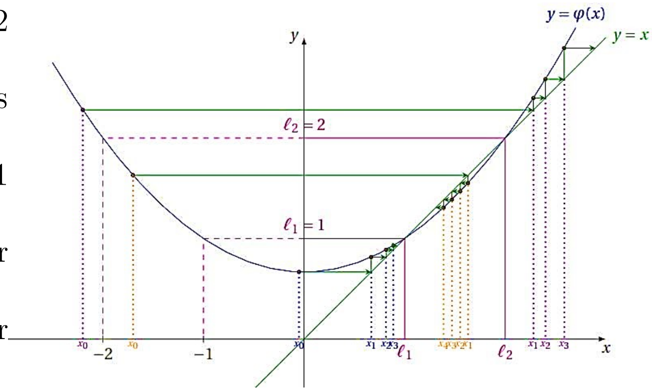
Solution

1. D'après le graphe de f, il existe deux solutions, $l_1 = 1$ et $l_2 = 2$.



2. Comparons le graphe de φ à l'identité :
   L'étude graphique montre que la suite converge quelque soit $x_0 \in [-2, 2]$. Plus précisément, on voit que :

— Si $x_0 = -2$ et $x_0 = 2$ alors $x_2 = 2$ pour tout $n \in \mathbb{N}^*$.
— Si $x_0 \in]-2,-1[$ et $x_0 \in]1,2[$ alors $x_n \searrow 1$ pour tout $n \in \mathbb{N}^*$.
— Si $x_0 = -1$ et $x_0 = 1$ alors $x_2 = 1$ pour tout $n \in \mathbb{N}^*$.
— Si $x_0 \in]-1,1[$ alors $x_n \nearrow 1$ pour tout $n \in \mathbb{N}^*$.
— Si $x_0 > 2$ alors $x_n \nearrow +\infty$ pour tout $n \in \mathbb{N}^*$.



3. L'intervalle maximale pour appliquer le théorème de point fixe est défini comme le plus grand intervalle pour lequel $|\varphi'(x)| < 1$ c'est-à-dire $-1 < \dfrac{2}{3}x < 1$. Donc $x \in \,]-1.5, 1.5\,[$.
On remarque que $]-1.5, 1.5[\subset [-2,2]$. Or si $x_0 \in [-2,-1.5[$ ou $x_0 \in [1.5,2[$, alors $x_1 = \varphi(x_0) < |x_0| \in]1,2[$ car $\varphi(x) < x$ lorsque $x \in]1,2[$ et on montre que la suite $(x_n)_{n\in\mathbb{N}}$ est décroissante et minorée par l_1 donc convergente.
Comme l'unique limite possible est l_1 on conclut que $x_n \searrow l_1$.

Exercice 9 *Pour approcher les racines réelles de la fonction*

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ x &\longmapsto x^3 - x^2 - 1 \end{aligned}$$

On utilise deux méthodes de point fixe :

$$\begin{cases} x_0 \text{ donné} \\ x_{n+1} = \varphi(x_n) \text{ pour tout } n \in \mathbb{N} \end{cases}$$

où

$$\varphi(x) = \varphi_1(x) = x^3 - x^2 + x - 1 \text{ et } \varphi(x) = \varphi_2(x) = \sqrt[3]{x^2+1}$$

1. *Montrer qu'il existe une unique racine réelle l de f. Montrer que $l \in [1,2]$.*
2. *Étudier la convergence des deux méthodes de point fixe et, si elles convergent, donner l'ordre de convergence.*
3. *Calculer l pour un $\epsilon = 10^{-3}$ près.*

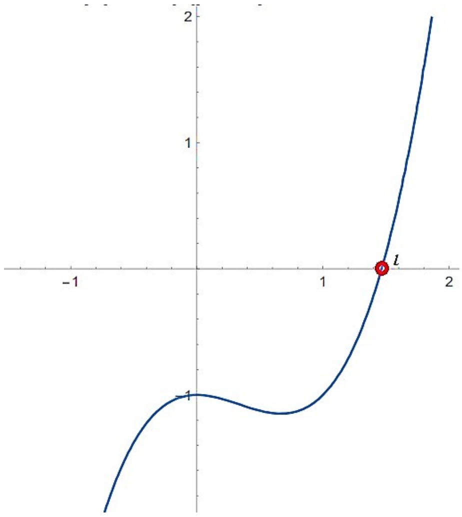
Solution

1. **Étude de f**
$*D_f = \mathbb{R}$ et f est de classe C^∞,
$*\lim\limits_{x\to\pm\infty} f(x) = \pm\infty$,
$*f'(x) = 3x^2 - 2x$, $f'(x) = 0 \Rightarrow (x = 0) \vee \left(x = \dfrac{2}{3}\right)$.

Voici le tableau de variations suivant :

x	$-\infty$		0		$\frac{2}{3}$		$+\infty$
$f'(x)$		$+$	0	$-$	0	$+$	
$f(x)$	$-\infty$	$\nearrow$	-1	$\searrow$	-1.14	$\nearrow$	$+\infty$

*$f(1) \times f(2) = (-1) \times 1 < 0$.
*$f'(x) > 0$ sur $[1, 2]$.
Donc d'après le théorème des valeurs intermédiaires il existe un unique $l \in [1, 2]$ tel que $f(l) = 0$.



2. **La convergence de φ_1 et φ_2**

— **Étude de φ_1**
Soit $\varphi_1(x) = x^3 - x^2 + x - 1$
*$D_{\varphi_1} = \mathbb{R}$ et $\lim\limits_{x \to \pm\infty} \varphi_1(x) = \pm\infty$,
*$\varphi_1'(x) = 3x^2 - 2x + 1$ et $\varphi_1''(x) = 6x - 2$,
*$\varphi_1''(x) = 0 \Leftrightarrow x = \frac{1}{3}$.
Voici le tableau de variations de φ_1 :

x	1		2
$\varphi_1''(x)$		$+$	
$\varphi_1'(x)$	2	$\nearrow$	9
$\varphi_1(x)$	0	$\nearrow$	5

On remarque que le $\max\limits_{x \in [1,2]} \left|\varphi_1'(x)\right| = 9 > 1$. Donc elle n'est pas contractante et que $\forall x \in [1, 2], \varphi_1(x) \notin [1, 2]$. Donc elle n'est pas stable, alors on conclut que $\varphi_1(x_n)$ est divergente.

— **Étude de φ_2**
Soit $\varphi_2 = \sqrt[3]{x^2+1}$,
$*D_{\varphi_2} = \mathbb{R}$ et $\lim\limits_{x \to \pm\infty} \varphi_2(x) = \pm\infty$,

$$*\varphi_2^{'}(x) = \frac{2x}{3}\left(x^2+1\right)^{-\frac{2}{3}} \text{ et } \varphi_2^{''}(x) = -\frac{2}{9}\frac{\left(x^2-3\right)}{\left(x^2+1\right)^{\frac{5}{3}}}.$$

$*\varphi_2^{''}(x) = 0 \Leftrightarrow x = \pm\sqrt{3}$.
Voici le tableau de variations de φ_2 :

x	1		$\sqrt{3}$		2
$\varphi_2''(x)$		$+$	0	$-$	
$\varphi_2'(x)$	0.419	$\nearrow$	0.458	$\searrow$	0.455
$\varphi_2(x)$	1.25		$\nearrow$		1.70

On remarque que le $\max\limits_{x \in [1,2]} \left|\varphi_2^{'}(x)\right| = 0.458 < 1$. Donc elle est contractante et que $\forall x \in [1,2], \varphi_2(x) \in [1,2]$. Donc elle est stable, alors on conclut que $\varphi_2(x_n)$ est convergente.

3. **Le calcul de l**

x_0	1	x_5	1.456889588
x_1	1.25992105	x_6	1.461623151
x_2	1.37284419	x_7	1.463775528
x_3	1.423531047	x_8	1.464754438
x_4	1.446474295		

On remarque que

$$|x_7 - x_8| = 9.78 \times 10^{-4} < 10^{-3}$$

alors

$$l \simeq x_8 = 1.464754438$$

Exercice 10 *On considère la fonction f définie par :*

$$f(x) = \ln(x) - \arctan(x),\ x > 0$$

1. *Montrer que l'équation $f(x) = 0$ admet une unique racine α qu'on localisera dans un intervalle I entre deux entiers consécutifs.*
2. *On considère la suite définie par :*

$$\begin{cases} x_0 = 3 \\ x_{n+1} = g(x_n) \text{ avec } g(x) = x - \ln(x) + \arctan(x) \end{cases}$$

Montrer que sa limite est α et en déduire le nombre d'itérations N qui assure que

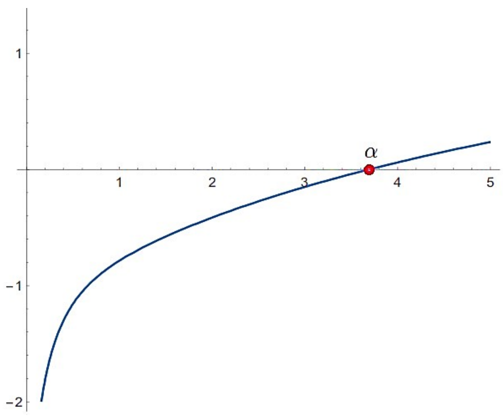
$$|x_N - \alpha| < 10^{-6}$$

3. *Soit la suite*
$$\begin{cases} x_0 = 3 \\ x_{n+1} = h(x_n) \text{ avec } h(x) = \tan(\ln(x)) \end{cases}$$
Cette suite converge t-elle vers α ?
4. *On définit la suite*
$$\begin{cases} x_0 = 3 \\ x_{n+1} = k(x_n) \text{ avec } k(x) = \exp(\arctan(x)) \end{cases}$$
Montrer qu'elle converge vers α et donner le nombre d'itérations P pour que
$$|x_p - \alpha| < 10^{-6}$$
5. *Déterminer la racine α de $f(x) = 0$ par l'une des trois méthodes précédentes en justifiant le choix de la méthode utilisée.*

Solution

1. **Étude de f**

D'après le graphe de f on remarque qu'il existe un unique zéro entre 3 et 4. Vérifions le en utilisant le théorème des valeurs intermédiaires.
*$f(3) \times f(4) = (-0.15) \times 0.06 < 0$,
*f est continue sur $[3, 4]$,
*$f'(x) = \dfrac{1}{x} - \dfrac{1}{x^2+1} > 0$ sur $[3, 4]$.
D'après le théorème des valeurs intermédiaires, il existe un unique $\alpha \in [3, 4]$ tel que $f(\alpha) = 0$.



2. **La convergence de g**

Soit $g(x) = x - \ln(x) + \arctan(x)$.
Pour montrer que la suite x_{n+1} converge vers α, il suffit de montrer que g est contractante et stable sur $[3, 4]$.
$g'(x) = \dfrac{x-1}{x} + \dfrac{1}{x^2+1}$ et $g''(x) = \dfrac{1}{x^2} - \dfrac{2x}{(x^2+1)^2}$.
Voici le tableau de variations de g :

x	3	4
$g''(x)$	+	
$g'(x)$	0.76 ↗	0.80
$g(x)$	3.15 ↗	3.93

On remarque que le $\max_{x\in[3,4]} |g'(x)| = k_1 = 0.80 < 1$, donc g est contractante et que $\forall x \in [3,4], g(x) \in [3,4]$, donc g est stable. Alors $g(x_n)$ converge vers α.

Nombre d'itérations

$$|x_n - \alpha| \leq \frac{k_1^N}{1-k_1} |x_1 - x_0| < 10^{-6}$$

$$N > \frac{\ln\left(\frac{(1-k_1)10^{-6}}{|x_1-x_0|}\right)}{\ln(k_1)}$$

$$N > 60.63$$

donc on prend

$$N = 61$$

3. **Étude de la convergence de h**
   Soit $h(x) = \tan(\ln(x))$,
   $h'(x) = \frac{1}{x}\left[1+\tan^2(\ln(x))\right]$ et $h''(x) = \frac{1}{x^2}\left[1+\tan^2(\ln(x))\right]$
   Voici le tableau de variations de h :

x	3	4
$h''(x)$	+	
$h'(x)$	1.61 $\nearrow$	7.42
$h(x)$	1.95 $\nearrow$	5.35

   On remarque que le $\max_{x\in[3,4]} |h'(x)| = k_2 = 7.42 > 1$, donc h n'est pas contractante et que $\forall x \in [3,4], h(x) \notin [3,4]$, donc h n'est pas stable. Alors $h(x_n)$ ne converge pas vers α.

4. **Étude de la fonction k**
   Soit $k(x) = \exp(\arctan(x))$,
   $k'(x) = \frac{1}{x^2+1}\exp(\arctan(x))$ et $k''(x) = \frac{-2x+1}{x^2+1}\exp(\arctan(x))$,

Voici le tableau de variations de k :

x	3 → 4
$k''(x)$	$-$
$k'(x)$	$0.34 \searrow 0.22$
$k(x)$	$3.48 \nearrow 3.76$

On remarque que le $\max_{x\in[3,4]} |K'(x)| = k_3 = 0.34 < 1$, donc k est contractante et que $\forall x \in [3,4], k(x) \in [3,4]$, donc k est stable. Alors $k(x_n)$ converge vers α.

Nombre d'itérations

$$|x_P - \alpha| \leq \frac{k_3^P}{1-k_3} |x_1 - x_0| < 10^{-6}$$

$$P > \frac{\ln\left(\frac{(1-k_3)10^{-6}}{|x_1-x_0|}\right)}{\ln(k_3)}$$

$$P > 12.52$$

donc on prend

$$P = 13$$

5. **Le calcul de α**

Pour estimer α on utilise la troisième méthode car $P < N$, donc $x_{n+1} = k(x_n)$ avec $k(x) = \exp(\arctan(x))$.

x_0	3	x_6	3.692362115
x_1	3.487013964	x_7	3.692529274
x_2	3.638287909	x_8	3.69252771452
x_3	3.678721749	x_9	3.692584779
x_4	3.689077084	x_{10}	3.692584779
x_5	3.691699757	x_{11}	3.692585457

On remarque que

$$|x_{11} - x_{10}| = 6.7 \times 10^{-7} < 10^{-6}$$

donc

$$\alpha \simeq x_{11} = 3.692585457$$

Exercice 11 *Soit la fonction* $f(x) = \dfrac{x-1}{x} - \exp(-x)$

1. *Tracer l'allure du graphe de f, puis en déduire que l'équation $f(x) = 0$ possède une racine α_1 négative et une racine α_2 positive. Localiser chacun de ces deux entiers successifs.*
2. *Pour déterminer ces racines, on utilise dans un premier temps la méthode du point fixe. Pour cela on propose les deux algorithmes suivants :*

$$x^{(k+1)} = g_1(x^k), \text{ avec } g_1(x) = 1 + x\exp(-x)$$

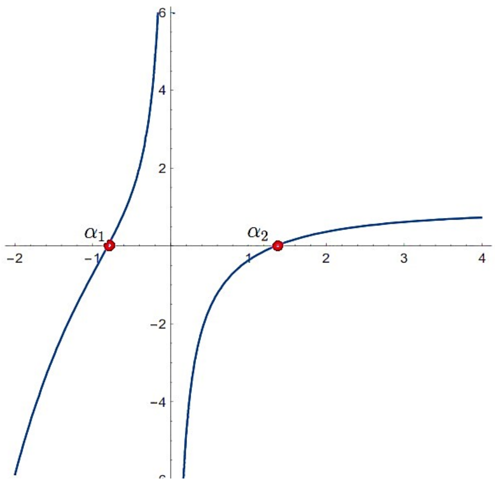
$$x^{(k+1)} = g_2(x^k), \text{ avec } g_2(x) = (x-1)\exp(x)$$

a- *Vérifier que les points fixes de ces algorithmes sont des racines de l'équation $f(x) = 0$ puis étudier leur convergence. S'il y a convergence, dire vers qu'elle racine.*
b- *Donner une approximation à 10^{-5} près, de la racine α_2 en initialisant les itérations par $x^{(0)} = 2$.*

Solution

1. **Le graphe de f**
D'après le graphe de f, on remarque qu'il existe deux racines, $\alpha_1 \in [-1,0]$ et $\alpha_2 \in [1,2]$, on a :

— Pour α_1 :
$*f(-1) \times f(0) < 0$,
$*f$ continue sur $[-1,0]$,
$*f'(x) = \frac{1}{x^2} + \exp(-x) > 0$ sur $[-1,0]$.
Donc d'après le théorème des valeurs intermédiaires il existe un unique $\alpha_1 \in [-1,0]$ tel que $f(\alpha_1) = 0$.



— Pour α_2 :
$*f(1) \times f(2) = -0.36 \times 0.36 < 0$,
$*f$ continue sur $[1,2]$,
$*f'(x) = \frac{1}{x^2} + \exp(-x) > 0$ sur $[1,2]$.
Donc d'après le théorème des valeurs intermédiaires il existe un unique $\alpha_2 \in [1,2]$ tel que $f(\alpha_2) = 0$.

a-
(i) Soit $x^{(k+1)} = g_1(x^k)$, avec $g_1(x) = 1 + x\exp(-x)$, vérifions que les points fixes de cet algorithme sont des racines de l'équation $f(x) = 0$.

Pour α_1 :
Montrons que $g_1(\alpha_1) = \alpha_1 \iff f(\alpha_1) = 0$.

$$
\begin{aligned}
g_1(\alpha_1) = \alpha_1 \iff & \quad 1 + \alpha_1 \exp(-\alpha_1) = \alpha_1 \\
\iff & \quad \exp(-\alpha_1) = \frac{\alpha_1 - 1}{\alpha_1} \\
\iff & \quad \frac{\alpha_1 - 1}{\alpha_1} - \exp(-\alpha_1) = 0 \\
\iff & \quad f(\alpha_1) = 0
\end{aligned}
$$

Pour α_2 :
Montrons que $g_1(\alpha_2) = \alpha_2 \iff f(\alpha_2) = 0$.

$$
\begin{aligned}
g_1(\alpha_2) = \alpha_2 \iff & \quad 1 + \alpha_2 \exp(-\alpha_2) = \alpha_2 \\
\iff & \quad \exp(-\alpha_2) = \frac{\alpha_2 - 1}{\alpha_2} \\
\iff & \quad \frac{\alpha_2 - 1}{\alpha_2} - \exp(-\alpha_2) = 0 \\
\iff & \quad f(\alpha_2) = 0
\end{aligned}
$$

Étude de la convergence de g_1

Pour $\alpha_1 \in [-1, 0]$:
$g_1'(x) = (1 - x)\exp(-x)$ et $g_1''(x) = (x - 2)\exp(-x)$

Voici le tableau de variations de g_1 :

x	-1	0
$g_1''(x)$	$-$	
$g_1'(x)$	$5.43 \searrow$	1
$g_1(x)$	$-1.71 \nearrow$	1

D'après ce tableau, on remarque que le $\max\limits_{x\in[-1,0]} |g_1'(x)| = 5.43 > 1$, donc g_1 n'est pas contractante et que $\forall x \in [-1, 0], g_1(x) \notin [-1, 0]$, donc g_1 n'est pas stable. Alors $g_1(x_n)$ ne converge pas vers α_1.

Pour $\alpha_2 \in [1,2]$:
$g_1'(x) = (1-x)\exp(-x)$ et $g_1''(x) = (x-2)\exp(-x)$

Voici le tableau de variations de g_1 :

x	1	2
$g_1''(x)$	$-$	
$g_1'(x)$	$0 \searrow$	-0.13
$g_1(x)$	$1.36 \searrow$	1.27

D'après ce tableau, on remarque que le $\max\limits_{x\in[-1,0]} |g_1'(x)| = 0.13 < 1$, alors g_1 est contractante et que $\forall x \in [1,2], g_1(x) \in [1,2]$, donc g_1 est stable. Alors que $g_1(x_n)$ converge vers α_2.

(ii) Soit $x^{(k+1)} = g_2(x^k)$, avec $g_2(x) = (x-1)\exp(x)$, vérifions que les points fixes de cet algorithme sont des racines de l'équation $f(x) = 0$.

Pour α_1 :
Montrons que $g_2(\alpha_1) = \alpha_1 \Longleftrightarrow f(\alpha_1) = 0$.

$$\begin{aligned}
g_2(\alpha_1) = \alpha_1 \Longleftrightarrow &\quad (\alpha_1 - 1)\exp(\alpha_1) = \alpha_1 \\
\Longleftrightarrow &\quad \frac{\alpha_1 - 1}{\alpha_1}\exp(\alpha_1) = 1 \\
\Longleftrightarrow &\quad \frac{\alpha_1 - 1}{\alpha_1} = \exp(\alpha_1 - 1) \\
\Longleftrightarrow &\quad \frac{\alpha_1 - 1}{\alpha_1} - \exp(\alpha_1 - 1) = 0 \\
\Longleftrightarrow &\quad f(\alpha_1) = 0
\end{aligned}$$

Pour α_2 :
Montrons que $g_2(\alpha_2) = \alpha_2 \Longleftrightarrow f(\alpha_2) = 0$.

$$\begin{aligned}
g_2(\alpha_2) = \alpha_2 \Longleftrightarrow &\quad (\alpha_2 - 1)\exp(\alpha_2) = \alpha_2 \\
\Longleftrightarrow &\quad \frac{\alpha_2 - 1}{\alpha_2}\exp(\alpha_2) = 1 \\
\Longleftrightarrow &\quad \frac{\alpha_2 - 1}{\alpha_2} = \exp(\alpha_2 - 1) \\
\Longleftrightarrow &\quad \frac{\alpha_2 - 1}{\alpha_2} - \exp(\alpha_2 - 1) = 0 \\
\Longleftrightarrow &\quad f(\alpha_2) = 0
\end{aligned}$$

Etude de la convergence de g_2

Pour $\alpha_1 \in [-1, 0]$:

$g_2'(x) = x\exp(x)$ et $g_2''(x) = (x+1)\exp(x)$.

Voici le tableau de variations de g_2 :

x	-1	0
$g_2''(x)$	$+$	
$g_2'(x)$	-0.36 ↗	0
$g_2(x)$	-0.73 ↘	-1

D'après ce tableau, on remarque que le $\max\limits_{x\in[-1,0]} |g_2'(x)| = 0.36 < 1$, donc g_2 est contractante, et que $\forall x \in [-1, 0], g_2(x) \in [-1, 0]$, donc g_2 est stable. Alors $g_2(x_n)$ converge vers α_1.
Pour $\alpha_2 \in [1, 2]$:

$g_2'(x) = x\exp(x)$ et $g_2''(x) = (x+1)\exp(x)$.
Voici le tableau de variations de g_2 :

x	1	2
$g_2''(x)$	$+$	
$g_2'(x)$	2.71 ↗	14.77
$g_2(x)$	0 ↗	7.38

D'après ce tableau, on remarque que le $\max\limits_{x\in[1,2]} |g_2'(x)| = 14.77 > 1$, donc g_2 n'est pas contractante, et que $\forall x \in [1, 2], g_2(x) \notin [1, 2]$, donc g_2 n'est pas stable. Alors $g_2(x_n)$ ne converge pas vers α_2.
b-
Pour calculer $\alpha_2 \in [1, 2]$, on va utiliser l'algorithme suivant :

$$x^{(k+1)} = g_1(x^k), \text{ avec } g_1(x) = 1 + x\exp(-x)$$

$x^{(0)}$	2	$x^{(4)}$	1.350031356
$x^{(1)}$	1.270670566	$x^{(5)}$	1.349971507
$x^{(2)}$	1.356605268	$x^{(6)}$	1.349976937
$x^{(3)}$	1.349371373		

On remarque que

$$\left|x^{(6)} - x^{(5)}\right| = 5.43 \times 10^{-6} < 10^{-5}$$

donc

$$\alpha_2 \simeq x^{(6)} = 1.349976937$$

Exercice 12 *On propose de résoudre une équation $f(x) = 0$ en utilisant la méthode de point fixe. Dans chaque cas déterminer les points fixes et leur nature. Tracer le graphe de φ et quelques points de la suite.*

1. $\begin{aligned} \varphi : [0,1] &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2}x(1-x) \end{aligned}$

2. $\begin{aligned} \varphi : [0,1] &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2}x(1+x) \end{aligned}$

3. $\begin{aligned} \varphi : [0,1] &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2}x(1-x^2) \end{aligned}$

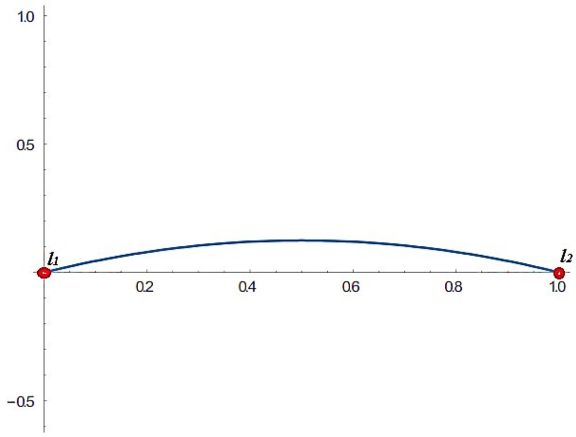
Solution

1. $\begin{aligned} \varphi : [0,1] &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{2}x(1-x) \end{aligned}$

$\varphi'(x) = \dfrac{1}{2} - x$ et $\varphi''(x) = -1$.

x	0		$\frac{1}{2}$		1
$\varphi''(x)$		$-$	0	$-$	
$\varphi'(x)$	$\frac{1}{2}$	$\searrow$	0	$\searrow$	$-\frac{1}{2}$
$\varphi(x)$	0	$\nearrow$	$\frac{1}{8}$	$\searrow$	0

D'après le tableau de variations de φ, on remarque que le $\max_{x\in[0,1]} |\varphi'(x)| = 0.5 < 1$ et $\varphi(0) = 0$, donc $l_1 = 0$ est un point fixe attractif.

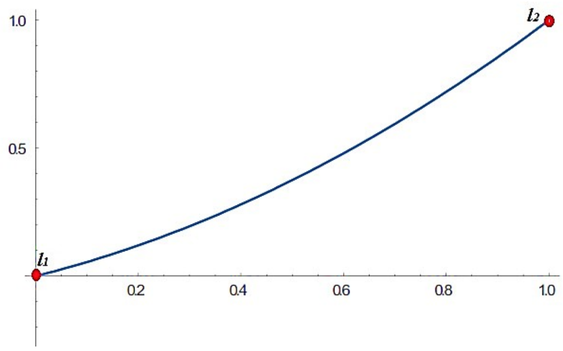


2. $\varphi : [0,1] \longrightarrow \mathbb{R}$
$$x \longmapsto \frac{1}{2}x(1+x)$$
$\varphi'(x) = \dfrac{1}{2} + x$ et $\varphi''(x) = 1$.

Voici le tableau de variations de φ :

x	0 $\quad$ 1
$\varphi''(x)$	$+$
$\varphi'(x)$	$0.5 \nearrow 1.5$
$\varphi(x)$	$0 \nearrow 1$

D'après le tableau de variations de φ, on remarque que $\varphi(0) = 0$ et $\varphi(1) = 1$, donc $l_1 = 0$ et $l_2 = 1$ sont deux points fixes, et on a :
$\varphi'(l_1) = 0.5 \in]-1, 1[$, donc l_1 est attractif.
$\varphi'(l_2) = 1.5 > 1$, donc l_2 est répulsif.

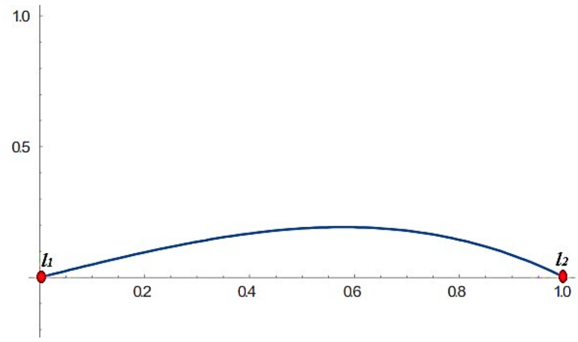


3. $\varphi : [0,1] \longrightarrow \mathbb{R}$
$$x \longmapsto \frac{1}{2}x(1+x^2)$$
$\varphi'(x) = \dfrac{1}{2} + \dfrac{3}{2}x^2$ et $\varphi''(x) = 3x$.

Voici le tableau de variations de φ :

x	0 1
$\varphi''(x)$	$+$
$\varphi'(x)$	$0.5 \nearrow 2$
$\varphi(x)$	$0 \nearrow 1$

D'après le tableau de variations de φ, on remarque que : $\varphi(0) = 0$ et $\varphi(1) = 1$, donc $l_1 = 0$ et $l_2 = 1$ sont deux points fixes, et on a :
$\varphi'(l_1) = 0.5 \in]-1, 1[$, donc l_1 est attractif.
$\varphi'(l_2) = 2 > 1$, donc l_2 est répulsif.



Exercice 13 *(Évolution d'un capital)*
On investit un capital C_0. Le placement a un taux de 5% par an et des frais de gestion fixes de 50 UM (unité monétaire) qui sont prélevés chaque année.

1. *Décrire la suite récurrente qui décrit l'évolution du placement.*
2. *Donner les points fixes du système et indiquer s'ils sont attractifs ou répulsifs.*
3. *Étudier l'évolution du capital au fil des ans selon la valeur de C_0.*

Solution

1. Notons u_n le capital au début de la $\text{n}^{\text{ième}}$ année, alors on a :
$$\begin{cases} u_n = 0 \\ u_{n+1} = (1 + 5\%)u_n - 50 = 1.05u_n - 50, \forall n \in \mathbb{N} \end{cases}$$
2. Soit $u_{n+1} = \varphi(u_n)$ avec $\varphi(x) = 1.05x - 50$.
On a
$$\varphi(\alpha) = \alpha \Longleftrightarrow 1.05\alpha - 50 = \alpha$$
$$\alpha = 1000$$
Donc l'unique point fixe du système est $\alpha = 1000$.
De plus $\varphi'(x) = 1.05 > 1$, il s'agit donc d'un point fixe répulsif.
3. — Si $C_0 > 1000$ alors $u_n \to +\infty$,
— Si $C_0 = 1000$ alors $u_n = 1000$, pour tout $n \in \mathbb{N}$,
— Si $C_0 < 1000$ alors $u_n \to -\infty$.

1.2.6 Exercices supplémentaires

Exercice 14 *Soit* $h(x) = 2\tan(x) - x, x \in [0, \frac{\pi}{2}[$.
Calculer $h^{-1}(1)$ *par la méthode du point fixe, avec la précision* $\epsilon = 10^{-6}$.

Exercice 15 *On se propose de calculer* $\sqrt[4]{\frac{1}{3}}$ *en trouvant les racines réelles de l'application* f *de* $\mathbb{R}$ *dans* $\mathbb{R}$ *définie par* $f(x) = x^4 - \frac{1}{3}$.

1. *Situer les deux racines de* f *(c'est-à-dire indiquer deux intervalles qui contiennent chacun une et une seule racine).*
2. *Soit* g *la fonction définie sur* $[0, 1]$ *par*

$$g(x) = \frac{x(9x^4 + 5)}{3(5x^4 + 1)}$$

   (a) *Faire l'étude complète de la fonction* g *et la comparer à l'identité.*

   (b) *Soit la suite* $(x_n)_{n\in\mathbb{N}}$ *définie par*

$$x_{n+1} = g(x_n), \quad x_0 \in]0, 1[$$

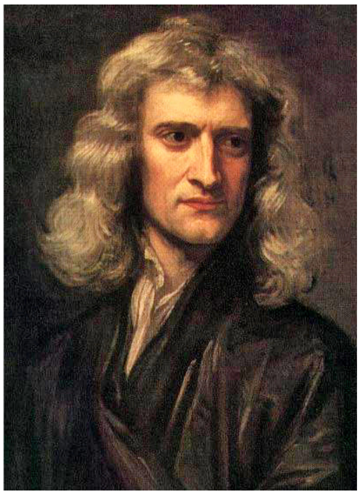
   À l'aide des graphes de g *et de l'identité sur* $[0, 1]$, *dessiner la suite* $(x_n)_{n\in\mathbb{N}}$ *sur l'axe des abscisses. Observer graphiquement la convergence.*

   (c) *Justifier mathématiquement la convergence observée graphiquement.*

   (d) *Calculer l'ordre de convergence de la méthode.*

1.3 Méthode de Newton

Isaac Newton (25 décembre 1642 J - 20 mars 1727 J, ou 4 janvier 1643 G - 31 mars 1727 G) est un physicien, mathématicien, philosophe, alchimiste, astronome et théologien anglais, puis britannique. Figure emblématique des sciences, il est surtout reconnu pour avoir fondé la mécanique classique, pour sa théorie de la gravitation universelle et la création, en concurrence avec Gottfried Wilhelm Leibniz, du calcul infinitésimal. Il est aussi connu pour la généralisation du théorème du binôme et l'invention dite de la méthode de Newton permettant de trouver des approximations d'un zéro (ou racine) d'une fonction d'une variable réelle à valeurs réelles.



1.3.1 Principe de la méthode

On suppose que la fonction est dérivable sur l'intervalle $[a, b]$. Le principe de la méthode de Newton qu'on appelle aussi la méthode de la tangente, consiste à choisir un point x_0 de l'intervalle de définition de f, et on considère la tangente à la courbe représentative de f en $(x_0, f(x_0))$. Soit x_1 l'abscisse de l'intervalle de la tangente avec l'axe des abscisses. Puisque la tangente est proche de la courbe, on peut espérer que x_1 donne une meilleure estimation d'une solution de l'équation $f(x) = 0$ que x_0. On recommence alors le procédé à partir de x_1 et on construit par récurrence une suite (x_n) définie par :

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

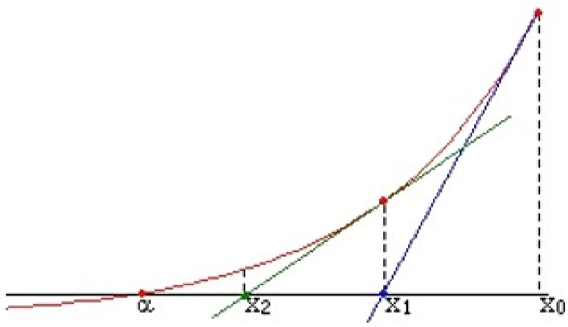
— L'équation de la tangente est donnée par :

$$\frac{[y - f(x_0)]}{(x - x_0)} = f'(x_0).$$

— La racine de la tangente est donnée par :

$$x_1 = \frac{x_0 - f(x_0)}{f'(x_0)}$$

.



1.3.2 Convergence de la méthode

Théorème 5 *(**Théorème de Newton**) (condition suffisante de convergence globale)*
Soit $f : [a, b] \to \mathbb{R}$ de classe C^2 vérifiant :

1. $f(a) \times f(b) < 0$
2. $f'(x) \neq 0, \forall x \in [a, b]$
3. $f''(x) \neq 0, \forall x \in [a, b]$

4. Partant d'un point x_0 qui satisfait l'inégalité

$$f(x_0).f''(x_0) > 0$$

(vérifié par un certain choix de $x_0 \in [a,b]$)

Si les conditions énoncées, ci-dessus, sont satisfaites, alors le processus de Newton :

$$\begin{cases} x_0 \text{ choisi} \\ x_{n+1} = g(x_n) = x_n - \dfrac{f(x_n)}{f'(x_n)} \end{cases}$$

Converge pour ce choix de x_0 vers l'unique solution α de $f(x)$.

De plus, nous avons l'estimation suivante :

$$|x_n - \alpha| \leq \frac{M}{2m} |x_{n-1} - \alpha|^2, \forall n > 0$$

avec

$$M = \max_{x \in [a,b]} |f'(x)|, m = \min_{x \in [a,b]} |f''(x)|$$

Cette inégalité peut s'écrire en fonction de x_0

$$|x_n - \alpha| \leq \left(\frac{M}{2m}\right)^{2^n - 1} |x_0 - \alpha|^{2n}, \forall n \in \mathbb{N}$$

Théorème 6 *(Condition de convergence locale)*
Soit f une fonction de classe C^2 de l'intervalle $[a,b]$ dans $\mathbb{R}$, avec $a < b$. Supposons qu'il existe un $\alpha \in [a,b]$ tel que $f(\alpha) = 0$ et $f'(\alpha) \neq 0$. Alors il existe $\delta > 0$ tel que pour tout $x_0 \in [\alpha - \delta, \alpha + \delta]$, la suite des itérés de Newton converge vers α quand $n \to +\infty$ (voir chap2).

1.3.3 Test d'arrêt

Les itérations s'achèvent dés que

$$|x_{n+1} - x_n| < \epsilon \qquad (*)$$

où ϵ est une tolérance fixée. En fait, si on note $e_n = \alpha - x_n$ l'erreur à l'itération n, on a

$$e_{n+1} = \alpha - x_{n+1} = g(\alpha) - g(x_n) = g'(\xi_n)e_n$$

avec ξ_n entre x_n et α, et

$$x_{n+1} - x_n = \alpha - x_n - \alpha + x_{n+1} = e_n - e_{n+1} = (1 - g'(\xi_n))e_n$$

Or, si n est suffisamment grand, $g'(\xi_n) \approx g'(\alpha) = 0$ et donc

$$e_n \cong x_{n+1} - x_n$$

L'erreur qu'on commet lorsque l'on adopte le critère $(*)$ est donc plus petite que la tolérance fixée.

1.3.4 Ordre de convergence

Théorème 7 *On suppose que $f'(\alpha) \neq 0$, alors si la suite $(x_n)_{n \in \mathbb{N}}$ des itérés de Newton converge, sa vitesse de convergence est (au moins) quadratique.*

Remarque 4 *Dans ce qui précède, nous avons supposé que la fonction f dont nous cherchons le zéro α vérifie $f(\alpha) = 0$ et $f'(\alpha) \neq 0$. Autrement dit, nous avons supposé que α est une racine simple de f. Maintenant la question qui se pose est : que se passe-t-il quand α est une racine de f de multiplicité $m \geq 2$?*
Si on garde la même fonction g que précédemment, la méthode de Newton perd son caractère de convergence quadratique. En effet on peut écrire

$$f(x) = (x-\alpha)^m h(x) \text{ avec } h(\alpha) \neq 0$$

Donc

$$g(x) = x - \frac{(x-\alpha)^m h(x)}{m(x-\alpha)^{m-1}h(x) + (x-\alpha)^m h'(x)} = x - \frac{(x-\alpha)h(x)}{mh(x) + (x-\alpha)h'(x)}$$

et

$$\lim_{x\to+\infty} \frac{g(x)-g(\alpha)}{x-\alpha} = 1 - \frac{1}{m} \neq 0,$$

ce qui implique en terme de suite que

$$\frac{|x_{n+1}-\alpha|}{|x_n-\alpha|} \to 1 - \frac{1}{m} \in]0, +\infty[$$

Ceci se traduit par une convergence linéaire et pas du tout quadratique. Pour récupérer cette dernière, on fait appel à la méthode de Newton modifiée.

1.3.5 Méthode de Newton modifiée

On suppose que α est une racine de multiplicité $m \geq 2$, c'est-à-dire :

$$f(x) = (x-\alpha)^m h(x) \text{ avec } h(\alpha) \neq 0$$

On suppose que $f \in C^2([a,b])$ et par conséquent h l'est aussi, on définit alors la fonction g par :

$$g(x) = x - m\frac{f(x)}{f'(x)}$$

Dans ce cas,

$$\lim_{x\to\alpha} g(x) = \frac{g(x)-g(\alpha)}{x-\alpha} = 0,$$

ce qui implique

$$\lim_{x\to+\infty} \frac{|x_{n+1}-\alpha|}{|x_n-\alpha|} = 0$$

et

$$\lim_{x\to+\infty} \frac{|x_{n+1}-\alpha|}{|x_n-\alpha|^2} = \frac{h'(\alpha)}{mh(\alpha)}$$

Remarque 5 *Les méthodes quadratiques sont plus intéressantes que les méthodes linéaires car leur vitesse de convergence est plus rapide. Des méthodes d'ordre plus élevé seraient encore plus rapides mais plus lourdes à mettre en place et sont plus utilisées dans la pratique.*
Il n'est pas toujours possible de choisir avec certitude un x_0 de départ dans la zone de convergence. Il pourra alors y avoir, même si la suite converge, une phase de recherche dans laquelle il n'y a pas réduction progressive de l'erreur, il pourra y avoir divergence.
La méthode de Newton nécessite le calcul des dérivées $f'(x)$, c'est un inconvénient dans la pratique où l'on ne dispose pas toujours d'expression analytique pour la fonction f.

1.3.6 Exercices corrigés

Exercice 16 *Soit la fonction* $f(x) = (x+2)^{\frac{2}{5}}$.

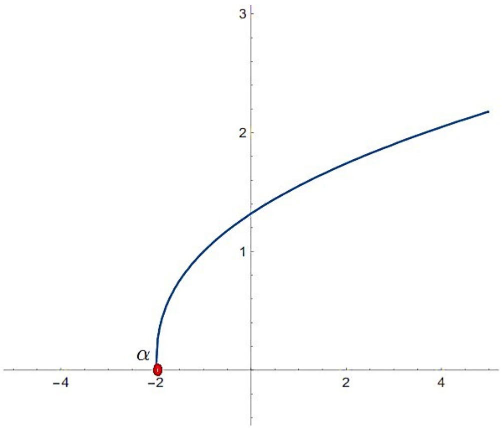
1. *Déterminer de manière analytique, l'unique racine de* f.
2. *En appliquant la méthode de Newton, peut-on déterminer cette racine ? Justifier.*

Solution

Soit la fonction $f(x) = (x+2)^{\frac{2}{5}}$.

1. $f(x) = 0 \Leftrightarrow x = -2$.
   Donc l'unique racine de f est -2.

2. D'après le graphe de f on remarque qu'il existe un unique $\alpha \in [-3, 0]$.
   Vérifions le par le théorème des valeurs intermédiaires.
   $*f(-3) \times f(0) = -1 \times 1.31 < 0$,
   $*f$ est continue sur $[-3, 0]$,
   $*f'(x) = \dfrac{2}{5}(x+2)^{-\frac{3}{5}} < 0$ sur $[-3, 0]$.
   Donc d'après le théorème des valeurs intermédiaires il existe un unique $\alpha \in [-3, 0]$.



   — **Procédure de Newton** :
   $*f \in C^2([-3, 0])$,
   $*f'(x) = \dfrac{2}{5}(x+2)^{-\frac{3}{5}} < 0$. Pour $x \neq -2$
   On a $f'(-2)$ n'est pas définie donc elle diverge. Alors on ne peut pas appliquer la méthode de Newton.

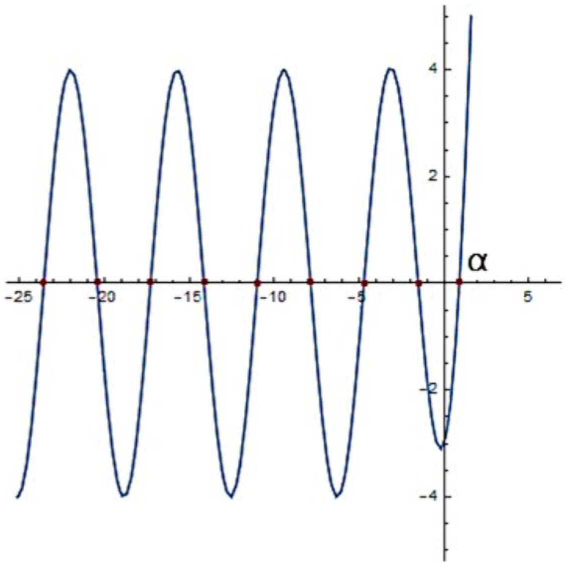
Exercice 17 *Soit la fonction définie par :*

$$f(x) = \exp(x) - 4\cos(x) = 0$$

1. *Séparer graphiquement les racines de* $f(x)$ *et déduire le nombre de racines.*
2. *Chercher à une précision* $\epsilon = 10^{-5}$ *près, une racine de* $f(x)$ *par la méthode de Newton dans l'intervalle* $\left[\dfrac{\pi}{4}, \dfrac{\pi}{2}\right]$. *Justifier le choix de* x_0 *et tester* $|x_{n+1} - x_n| < 10^{-5}$.

Solution
Soit la fonction définie par : $f(x) = \exp(x) - 4\cos(x) = 0$.

1. **Le graphe de f**



2. On a $\alpha \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right]$.

— **Procédure de Newton**

$*f \in C^{\infty}\left(\left[\frac{\pi}{4}, \frac{\pi}{2}\right]\right)$,

$*f\left(\frac{\pi}{4}\right) \times f\left(\frac{\pi}{2}\right) = -0.63 \times 4.81 < 0 \Rightarrow \alpha \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right]$,

$*f'(x) = \exp(x) - 4\sin(x) > 0$ sur $\left]\frac{\pi}{4}, \frac{\pi}{2}\right[$

$*f''(x) = \exp(x) + 4\cos(x) > 0$ sur $\left]\frac{\pi}{4}, \frac{\pi}{2}\right[$

Donc on peut appliquer la méthode de Newton.

— **Le choix de x_0**

$f\left(\frac{\pi}{2}\right) \times f''\left(\frac{\pi}{2}\right) = 4.81 \times 4.81 > 0.$

Donc on prend $x_0 = \frac{\pi}{2}$.

Alors le processus de Newton est :

$$\begin{cases} x_0 = \dfrac{\pi}{2} \\ x_{n+1} = g(x_n) = x_n - \dfrac{f(x_n)}{f'(x_n)} = x_n - \dfrac{\exp(x_n) - 4\cos(x_n)}{\exp(x_n) + 4\sin(x_n)}, \end{cases}$$

converge vers la solution α.

Exercice 18 *On considère l'équation* $f(x) = x^2 - A \qquad (1)$

1. *Montrer que la méthode de Newton appliquée à (1) conduit à l'algorithme :*

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{A}{x_n}\right), n = 0, 1, ...; x_0 \text{ donné}$$

2. *Montrer que cette suite converge quadratiquement vers* $\sqrt{A}$.

Solution

1. L'algorithme de la méthode de Newton s'écrit :

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, n = 0, 1....; x_0 \text{ donné.}$$

Appliqué à l'équation (1), il devient :

$$x_{n+1} = x_n - \frac{x_n^2 - A}{2x_n} = \frac{x_n^2 + A}{2x_n} = \frac{1}{2}\left(x_n + \frac{A}{x_n}\right)$$

2. L'ordre de convergence d'une méthode est tel que :

$\varphi'(\alpha) = \varphi''(\alpha) = ... = \varphi^{(k-1)}(\alpha) = 0$ et $\varphi^{(k)}(\alpha)$, où α est solution de l'équation $\varphi(\alpha) = \alpha$.

Dans le cas présent, on a : $\varphi(x) = \dfrac{1}{2}\left(x + \dfrac{A}{x}\right)$ et $\alpha = \sqrt{A}$, d'où :

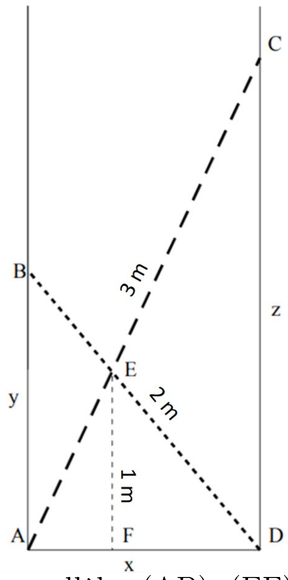
$$\varphi'(x) = \frac{1}{2}\left(1 - \frac{A}{x^2}\right) \text{ et } \varphi'(\alpha) = \frac{1}{2}\left(1 - \frac{A}{A}\right) = 0$$

$$\varphi''(x) = \frac{1}{2}\left(\frac{2Ax}{x^4}\right) = \frac{A}{x^3} \text{ et } \varphi''(\alpha) = \frac{1}{\sqrt{A}}$$

La méthode est donc d'ordre 2.

Exercice 19 *Entre deux murs (verticaux) parallèles, on place deux échelles en les croisant. La première fait 3 mètres de long, la seconde 2 mètres. On constate qu'elles se croisent à une hauteur de 1 mètre.*
En utilisant la méthode de Newton, calculer approximativement la largeur du couloir avec une tolérance de 10^{-3}.

Solution



On pose :
AD=x
AB=y
CD=z

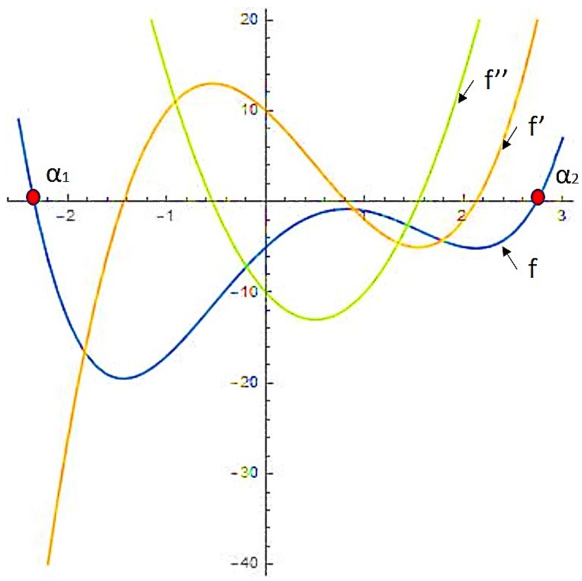
En utilisant le théorème de Pythagore (voir Annexe), on a :
Pour le triangle DAC : $x^2 + z^2 = 9$ (1)
Pour le triangle ABD : $x^2 + y^2 = 4$ (2)
On en déduit par soustraction de (1) par (2) :
$z^2 - y^2 = 5$ (3)

En utilisant le théorème de Thalès (voir Annexe), avec les parallèles (AB), (EF) et (CD), on a :
$\frac{EF}{AB} = \frac{AF}{AD}$ et $\frac{EF}{CD} = \frac{FD}{AD}$, soit $\frac{EF}{y} = \frac{AF}{AD}$ et $\frac{EF}{z} = \frac{FD}{AD}$
comme $EF = 1$: $\frac{1}{y} = \frac{AF}{AD}$ (4) et $\frac{1}{z} = \frac{FD}{AD}$ (5)
On en déduit par addition de (4) et (5) :
$\frac{1}{y} + \frac{1}{z} = \frac{AF}{AD} + \frac{FD}{AD} = \frac{AF + FD}{AD}$
On peut noter que $AF + FD = AD$
Donc : $\frac{1}{y} + \frac{1}{z} = \frac{AD}{AD}$ ou encore $\frac{1}{y} + \frac{1}{z} = 1$
On en déduit $\frac{1}{y} = 1 - \frac{1}{z}$ soit $\frac{1}{y} = \frac{z}{z} - \frac{1}{z}$ ou $y = \frac{z}{z-1}$

En remplaçant l'expression de y dans (3), on obtient : $z^2 - \frac{z^2}{(z-1)^2} = 5$
ou $z^2(z-1)^2 - z^2 = 5(z-1)^2$

puis par simplifications successives, on obtient : $z^4 - 2z^3 - 5z^2 + 10z - 5 = 0$

Tout d'abord, posons $f(z) = z^4 - 2z^3 - 5z^2 + 10z - 5$, représentée par le graphe suivant



D'après ce graphe, on remarque qu'il existe deux solutions $\alpha_1 \in [-2.5, -2]$ et $\alpha_2 \in [2.5, 3]$. On a
$*f(-3) \times f(-2) = 55 \times -13 \leq 0$
$*f$ est continue sur $[2.5, 3]$
$*f'(z) = 4z^3 - 6z^2 - 10z + 10 > 0$

D'après le théorème des valeurs intermédiaires il existe un unique $\alpha = \alpha_2 \in [2.5, 3]$ tel que $f(\alpha) = 0$.

Procédure de Newton

$*f$ est de classe $C^2([2.5, 3])$,
$*f'(z) = 4z^3 - 6z^2 - 10z + 10 > 0$ sur $[2.5, 3]$,
$*f''(z) = 12z^2 - 12z - 10 > 0$ sur $[2.5, 3]$,
donc le procédé de Newton est le suivant :

$$\begin{cases} z_0 & \text{donné} \\ z_{n+1} = & g(z_n) = z_n - \dfrac{f(z_n)}{f'(z_n)} = z_n - \dfrac{z^4 - 2z^3 - 5z^2 + 10z - 5}{4z^3 - 6z^2 - 10z + 10} \end{cases}$$

Le choix de z_0

On a $f(3) \times f''(3) = 7 \times 62 > 0$. Donc on a le tableau suivant :

z_0	3
z_1	2.794117647
z_2	2.739447775
z_3	2.735739747
z_4	2.735723253

À partir de ce tableau, on a

$$|z_4 - z_3| = 1.64 \times 10^{-5} < 10^{-3}$$

alors

$$\alpha \simeq 2.735723253$$

Donc pour $z = 2.735723253$ en remplaçant z dans (1) : $x^2 + z^2 = 9$ on obtient

$$x = 1.231185722$$

Alors la largeur du couloir est **1.231 mètres**.

Exercice 20 *Soit f une application de $\mathbb{R}$ dans $\mathbb{R}$ définie par*

$$f(x) = \exp(x^2) - 4x^2$$

On se propose de trouver les racines réelles de f.

1. *Localiser les 4 racines de f.*
2. *Montrer qu'il y a une racine $\hat{x}$ comprise entre 0 et 1.*
3. *Écrire la méthode de Newton pour la recherche des zéros de la fonction f.*

4. *Soit la méthode de point fixe*

$$\begin{cases} x_{k+1} = \varphi(x_k) \\ x_0 \in]0,1[\end{cases} \quad (1)$$

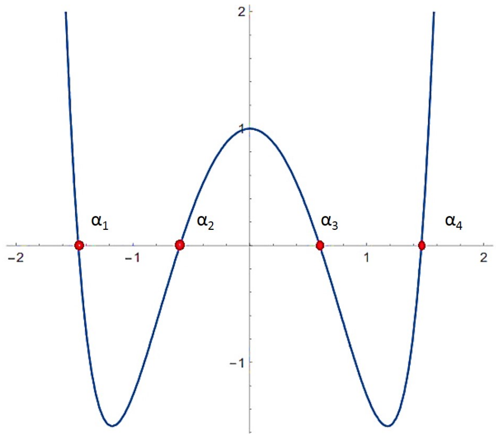
avec φ l'application de $\mathbb{R}$ dans $\mathbb{R}$ définie par

$$\varphi(x) = \frac{\sqrt{\exp(x^2)}}{2}$$

Examiner la convergence de cette méthode et préciser l'ordre de convergence.

5. *Entre la méthode de Newton et la méthode de point fixe (1), quelle est la plus efficace ? Justifier la réponse.*

Solution



1. D'après le graphe de f, on remarque qu'il y a quatre racines : $\alpha_1 \in [-2,-1]$, $\alpha_2 \in [-1,0]$, $\alpha_3 \in [0,1]$ et $\alpha_4 \in [1,2]$.
2. Montrons qu'il existe un unique $\hat{x} \in [0,1]$, pour cela utilisons le théorème des valeurs intermédiaires.
   $*f(0) \times f(1) = 1 \times -1.28 < 0$,
   $*f$ continue sur $[0,1]$,
   $*f'(x) = 2x[\exp(x) - 4] < 0$.
   Donc il existe un unique $\hat{x} \in [0,1]$.
3. Soit la méthode de Newton :

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{\exp(x_k^2) - 4x_k^2}{2x_k(\exp(x_k^2) - 4)}$$

4. Étude de la convergence de la méthode (1)
   Soit $\varphi(x) = \dfrac{\sqrt{\exp(x^2)}}{2}$
   On a : $0 < \sqrt{\dfrac{\exp(x^2)}{4}} < \sqrt{\dfrac{e}{4}} < 1$
   Donc $\varphi :]0,1[\longrightarrow]0,1[$. De plus on a :
   $\varphi'(x) = \dfrac{1}{2}x\sqrt{\exp(x^2)}$,
   $\varphi''(x) = \dfrac{1}{2}\sqrt{\exp(x^2)}(1+x)$,

Voici le tableau de variations de φ :

x	0 $\qquad\qquad$ 1
$\varphi''(x)$	$+$
$\varphi'(x)$	$0 \nearrow 0.82$
$\varphi(x)$	$0.5 \nearrow 0.82$

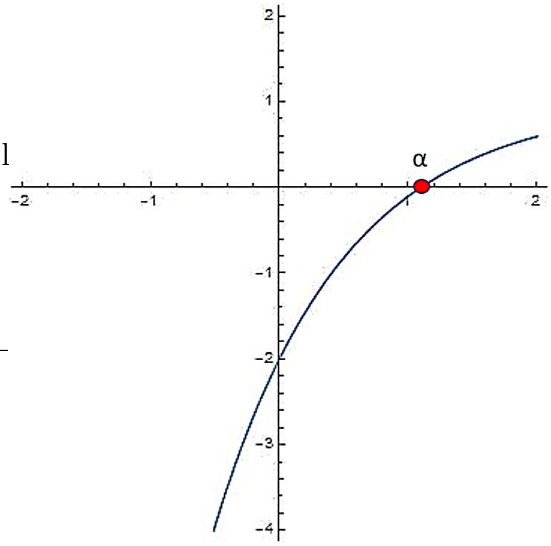
D'après le tableau de variations de φ, on remarque que :
$\forall x \in [0,1], |\varphi'(x)| = 0.82 < 1$, alors elle est contractante.
$\forall x \in [0,1], \varphi(x) \in [0,1]$, alors elle est stable. Donc on conclut que φ est convergente.
De plus on a : $\varphi'(\hat{x}) = \frac{1}{2}\hat{x}\sqrt{\exp(\hat{x}^2)} = \hat{x}\varphi(\hat{x}) = \hat{x}^2 \neq 0$ (car $\varphi(\hat{x}) = \hat{x}$).
Donc la méthode (1) converge seulement à l'ordre 1.

5. Puisque $\hat{x}$ est une racine simple de f, la méthode de Newton converge à l'ordre 2 tandis que la méthode de point fixe (1) converge à l'ordre 1, donc la méthode de Newton est plus efficace.

Exercice 21 *Soit la fonction $f(x) = 1 - 3\exp(-x), x \in \mathbb{R}$*

1. *Montrer que l'équation $f(x) = 0$ a une unique racine α qu'on localisera dans un intervalle I entre deux entiers consécutifs.*
2. *Appliquer à $f(x)$ la méthode de Newton pour trouver α.*
3. *Sur quelle fonction définit-on les itérations du point fixe ?*
4. *Calculer α avec la méthode de point fixe.*
5. *Calculer α avec la méthode de dichotomie.*
6. *Calculer α avec la méthode de Newton.*
   Tous les calculs se font à la précision $\epsilon = 10^{-3}$.

Solution



1. D'après le graphe de f, on remarque qu'il existe une racine dans $[1, 2]$. On a :
   $*f(1) \times f(2) = -0.10 * 0.59 < 0$.
   $*f$ est continue sur $[1, 2]$.
   $*f'(x) = 3\exp(-x) > 0$.
   Donc d'après le théorème des valeurs intermédiaires, il existe un unique $\alpha \in [1, 2]$.

2. **Méthode de Newton**
   $*f \in C^2([1, 2])$,
   $*f'(x) = 3\exp(-x)$ sur $[1, 2]$
   $*f''(x) = -3\exp(-x)$ sur $[1, 2]$
   donc le procédé de Newton est le suivant :

$$\begin{cases} x_0 & \text{donné} \\ x_{n+1} = & g(x_n) = x_n - \dfrac{f(x_n)}{f'(x_n)} = (x_n + 1) - \dfrac{1}{3\exp(-x_n)} \end{cases}$$

   Le choix de x_0
   On a $f(1) \times f''(1) = -0.10 \times -1.10 > 0$. Donc on prend $x_0 = 1$.

3. **Calcul de α par la méthode de point fixe**
   Soit $g(x) = x - 1 + 3\exp(-x)$
   α est un point fixe de g c'est-à-dire $g(\alpha) = \alpha \Longleftrightarrow f(\alpha) = 0$
   $\alpha - 1 + 3\exp(-\alpha) = \alpha \Longleftrightarrow -1 + 3\exp(-\alpha) = 0 \Longleftrightarrow 1 - 3\exp(-\alpha) = 0 \Longleftrightarrow$ $f(\alpha) = 0$
   Étudions la convergence de la méthode :
   $g'(x) = 1 - 3\exp(-x)$
   $g''(x) = 3\exp(-x)$

x	1		$\ln(3)$		2
$g''(x)$		$+$	0	$+$	
$g'(x)$	-0.1	$\nearrow$	0	$\nearrow$	0.59
$g(x)$	1.10	$\searrow$	1.09	$\nearrow$	1.40

   D'après le tableau de variations de g, on remarque que :
   $\max\limits_{x \in [1,2]} |g'(x)| = k = 0.59 < 1$, donc g est contractante.

x_0	1
x_1	1.103638324
x_2	1.098624898
x_3	1.098612289

$\forall x \in [1,2], g(x) \in [1,2]$, donc g est stable. Alors g converge vers α.

On a

$$|x_3 - x_2| = 1.26 \times 10^{-5} < 10^{-3}$$

alors

$$\alpha \simeq 1.09861228$$

4. **Calcul de α par la méthode de Dichotomie**

n	a_n	b_n	$f(a_n)$	$f(b_n)$	$c_n = \dfrac{a_n + b_n}{2}$	$f(c_n)$
0	1	2	$\ominus$	$\oplus$	1.5	$\oplus$
1	1	1.5	$\ominus$	$\oplus$	1.25	$\oplus$
2	1	1.25	$\ominus$	$\oplus$	1.125	$\oplus$
3	1	1.125	$\ominus$	$\oplus$	1.0625	$\ominus$
4	1.0625	1.125	$\ominus$	$\oplus$	1.09375	$\ominus$
5	1.09375	1.125	$\ominus$	$\oplus$	1.109375	$\oplus$
6	1.09375	1.109375	$\ominus$	$\oplus$	1.1015625	$\oplus$
7	1.09375	1.1015625	$\ominus$	$\oplus$	1.09765625	$\ominus$
8	1.09765625	1.1015625	$\ominus$	$\oplus$	1.099609375	$\oplus$
9	1.09765625	1.099609375	$\ominus$	$\oplus$	1.098632813	$\oplus$
10	1.09765625	1.098632813	$\ominus$	$\oplus$	1.098144532	

On a

$$|x_{10} - x_9| = 9.76 \times 10^{-4} < 10^{-3}$$

Donc

$$\alpha \simeq 1.098144532$$

5. **Calcul de α par la méthode de Newton**

Soit $\varphi(x) = x + 1 - \dfrac{1}{3\exp(-x)}$

x_0	1
x_1	1.093906057
x_2	1.098601232
x_3	1.098612289

On a

$$|x_3 - x_2| = 1.10 \times 10^{-5} < 10^{-3}$$

alors

$$\alpha \simeq 1.09861228$$

1.3.7 Exercices supplémentaires

Exercice 22 *Le but de cet exercice est de calculer la racine cubique d'un nombre positif a.*
Soit g la fonction définie sur $\mathbb{R}_+^*$ par

$$g(x) = \frac{2}{3}x + \frac{a}{3x^2} \quad (a > 0 \textit{ fixé})$$

1. *Faire l'étude complète de la fonction g.*
2. *Soit la suite $(x_n)_{n \in N}$ définie par :*

   $$x_{n+1} = g(x_n)$$

   Étudier la convergence de la méthode de point fixe.
3. *Calculer l'ordre de la méthode.*
4. *Expliciter la méthode de Newton pour la recherche du zéro de la fonction f définie par $f(x) = x^3 - a$. Que remarque-t-on ?*

Exercice 23 *On veut résoudre l'équation $\exp(-\alpha x) = x$ avec $0 < \alpha < 1$.*

1. *Vérifier que cette équation admet une unique solution notée l_α, dans $\mathbb{R}$.*
2. *Écrire la méthode de Newton pour résoudre l'équation $\exp(-\alpha x) = x$ avec $0 < \alpha < 1$.*
3. *Soit $g : \mathbb{R} \to \mathbb{R}$ la fonction définie par $\exp(\alpha x)$. On définit la suite récurrente :*

   $$\begin{cases} u_0 \in \mathbb{R} \\ u_{n+1} = g(u_n) \end{cases} \quad (1)$$

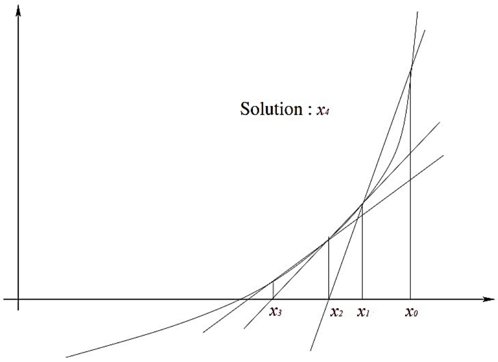
   — *On veut montrer que u_n converge vers l_α. Pour cela comparer d'abord le graphe de g à l'identité et observer graphiquement la convergence.*
   — *Justifier mathématiquement la convergence observée graphiquement.*
4. *Parmi la méthode de Newton et la méthode de point fixe (1), la quelle faut-il préférer vis-à-vis de la vitesse de convergence.*

1.4 Méthode de la sécante

1.4.1 Principe de la méthode

C'est une méthode du type :

$$x_n + 1 = F(x_n, x_{n-1}, ..., x_{n-N})$$



La méthode de Newton est rapide mais nécessite le calcul de la dérivée de f en tout point x_n, ce qu'on a pas toujours. La plus simple et la plus ancienne est la méthode de la sécante. Elle consiste à se donner deux points x_0 et x_1, tracer la droite qui passe par les points $(x_0, f(x_0))$ et $(x_1, f(x_1))$, elle coupe l'axe des x en x_2 et on recommence avec les points x_1 et x_2.

L'algorithme s'écrit :

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

1.4.2 Convergence de la méthode

Théorème 8 *Soit f une fonction de classe C^2 dans un voisinage $]\alpha - r, \alpha + r[$ d'une racine α de f avec $f'(\alpha) \neq 0$. Alors il existe un $\delta < r$ tel que si x_0 et $x_1 \in]\alpha - \delta, \alpha + \delta[$, alors x_n converge vers α, et*

$$|x_{n+1} - \alpha| \leq C(\delta) |x_n - \alpha| |x_{n-1} - \alpha|$$

avec

$$C(\delta) |x_n - \alpha| \leq \delta (C(\delta))^{\lambda_n}$$

où λ_n est la suite de Fibonacci définie par :

$$\lambda_{n+1} = \lambda_n + \lambda_{n-1}, \lambda_0 = \lambda_1 = 1$$

Remarque
La méthode de la sécante ne converge aussi que si l'on part suffisamment près de la racine,

$$|e_{n+1}| \leq c |e_n|^{\lambda}, \lambda \approx 1.6 \text{ nombre d'or}$$

C'est une convergence superlinéaire qui a pour ordre de convergence $\frac{1+\sqrt{5}}{2}$.

1.4.3 Critère d'arrêt

L'algorithme de la sécante n'est évidemment pas complet tant qu'on n'a pas précisé un critère d'arrêt. Généralement, x_n converge vers la solution x_∞ cherchée ce qui signifie que pour n grand, x_n est voisin de x_∞. Sans plus d'informations il est difficile de savoir ce que signifie numériquement "n grand", et c'est là où il y a une difficulté fréquente en analyse numérique.
Un critère d'arrêt souvent utilisé consiste à choisir à priori une tolérance ϵ et à terminer l'algorithme lorsque

$$|x_{n+1} - x_n| \leq \epsilon \qquad (1)$$

Ce n'est pas complètement satisfaisant ; pour s'en convaincre, on peut penser à la suite :

$$x_n = 1 + \frac{1}{2} + ... + \frac{1}{n}$$

qui vérifie $\lim\limits_{n\to+\infty} |x_{n+1} - x_n| = 0$.

Puisqu'on veut résoudre $f(x) = 0$, un autre critère d'arrêt possible consiste à s'arrêter lorsque

$$|f(x_n)| < \epsilon \qquad (2)$$

Un autre critère d'arrêt possible consiste à s'arrêter lorsqu'un numéricien rigoureux décide de s'arrêter que lorsque les deux critères (1) et (2) sont simultanément vérifiés.

Dans le cas de la méthode de la sécante, on peut utiliser

$$|f(x_n) - f(x_{n-1})| < \epsilon$$

Ce critère d'arrêt a l'avantage d'éviter une possible division par 0.

Pour éviter à l'ordinateur de tourner sans s'arrêter lorsqu'il n'y a pas convergence, il est évidemment indispensable de toujours mettre un critère limitant le nombre total d'itérations.

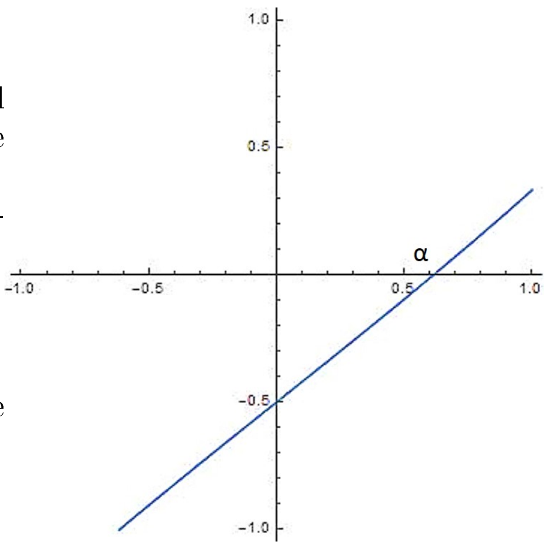
1.4.4 Exercices corrigés

Exercice 24 *Soit* $f(x) = x - 0.2\sin(x) - 0.5$

1. *Montrer que l'équation* $f(x) = 0$ *admet une unique racine dans* I.
2. *Calculer cette racine en utilisant la méthode de la sécante avec une tolérance de* 10^{-3}.

Solution

1. D'après le graphe de f, on remarque qu'il existe une unique racine dans l'intervalle $[0.5, 1]$.
   Utilisant le théorème des valeurs intermédiaires pour le montrer. On a
   *$f(0.5) \times f(1) = -0.09 \times 0.33 < 0$,
   *f continue sur $[0.5, 1]$,
   *$f'(x) = 1 - 0.2\cos(x) > 0$ sur $[0.5, 1]$
   Alors il existe un unique $\alpha \in [0.5, 1]$ tel que $f(\alpha) = 0$.



2. L'algorithme de la sécante est :

$$\begin{cases} x_{-1} = 0.5, x_0 = 1 \\ x_{n+1} = x_n = \dfrac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n) \end{cases}$$

On a le tableau suivant :

x_0	1
x_1	0.61212248
x_2	0.61549349
x_3	0.61546816

On a

$$|x_3 - x_2| = 2.533 \times 10^{-5} < 10^{-3}$$

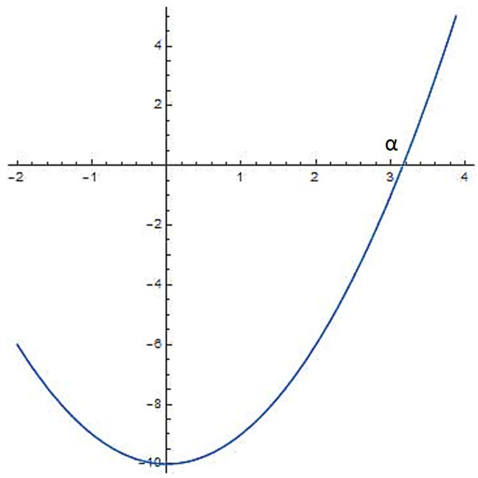
alors

$$\alpha \simeq 0.61546816$$

Exercice 25 *Estimer* $\sqrt{10}$, *en utilisant la méthode de la sécante avec une tolérance de* 10^{-4}.

Solution

Soit $f(x) = x^2 - 10$.



D'après le graphe de f on remarque qu'il existe une racine dans l'intervalle $[3, 4]$

Vérifions le, en utilisant le théorème des valeurs intermédiaires.

$f(3) \times f(4) = -1 \times 6 < 0$,

$*f$ continue sur $[3, 4]$,

$*f'(x) = 2x > 0$ sur $[3, 4]$

Alors il existe un unique $\alpha \in [3, 4]$ tel que $f(\alpha) = 0$.

L'algorithme de la sécante est :

$$\begin{cases} x_{-1} = 3, x_0 = 4 \\ x_{n+1} = x_n = \dfrac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n) \end{cases}$$

On a le tableau suivant :

x_0	4
x_1	3.142857143
x_2	3.16
x_3	3.16201117318
x_4	3.16224648985
x_5	3.16227401437

On a

$$|x_5 - x_4| = 2.75 \times 10^{-5} < 10^{-4}$$

alors

$$\alpha \simeq 3.16227401437$$

La méthode	Avantages	Inconvénients
Dichotomie	- La convergence est assurée ; - On a un encadrement de la solution ; - Un seul calcul de la fonction à chaque itération.	- Vitesse de convergence linéaire donc lente ; - Sensible aux erreurs d'arrondi ; Comme par exemple, la fonction $f(x) = e^x - 1 - x - \frac{x^2}{2}$ qui s'annule théoriquement en $x = 0$ seulement. Par contre numériquement, elle change de signe un grand nombre de fois autour de $x = 0$.
Point fixe	- Son analyse mathématique est relativement aisée et permet d'en déduire celle de la méthode de Newton ; - Joue un rôle considérable pour adapter des algorithmes linéaires à des situations non linéaires.	- Convergence souvent difficile à réaliser en pratique ; - Convergence lente de type linéaire ; - Certains points fixes dits instables, ne peuvent être atteints.
Méthode de Newton	- Convergence rapide ; - Relativement stable et peu sensible aux erreurs d'arrondis si $f'(x_\infty)$ n'est pas trop petit.	- Peut diverger ou converger vers un autre zéro que celui cherché si la donnée initiale est mal choisie ; - Nécessite le calcul de la dérivée de la fonction, ce qui est numériquement difficile si on le connais pas explicitement ; - Chaque étape nécessite deux évaluations de la fonction.
Méthode de la sécante	- Nécessite une seule évaluation de la fonction à chaque étape ; - Convergence relativement rapide mais moins que celle de Newton.	- Peut diverger si les données initiales sont mal choisis ; - Le calcul de $f(x_n) - f(x_{n-1})$ peut produire des erreurs de chute.

Chapitre 2

Résolution des systèmes linéaires

La résolution des systèmes linéaires est considérée comme l'un des deux problèmes fondamentaux de l'analyse numérique matricielle et cette résolution intervient à divers domaines, en particulier lorsqu'il s'agit de modéliser puis résoudre numériquement des problèmes comme par exemple en :

— **Électricité** : recherche des intensités du courant dans un circuit électronique en utilisant les lois des noeuds et des mailles ;
— **Chimie** : l'équilibre des équations des réactions chimiques ;
— **Algèbre linéaire** : recherche des vecteurs propres d'une matrice carrée associés à une valeur propre donnée, déterminer le noyau d'un endomorphisme en dimension finie (système homogène) ;
— **Géométrie analytique** : étudier la position relative des droites et des plans dans l'espace ;
— **Optimisation** : minimisation ou maximisation d'une fonction quadratique ;
— **Analyse numérique** : interpolation polynomiale (matrice de Vandermonde), résolution numérique des équations aux dérivées partielles (méthodes des différences finies où les matrices rencontrées sont tridiagonales).

Position du problème. Le problème consiste à trouver la solution du système d'équations

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

On cherche à résoudre le système $Ax = b$ avec A est une matrice à coefficients réels de taille $n \times n$, b un vecteur de $\mathbb{R}^n$ et x l'inconnu à chercher de $\mathbb{R}^n$.

Remarque 6 *Si $b = 0$ le système est dit homogène, donc il admet au moins une solution $x = 0$ qui est la solution nulle ou triviale.*

Remarque 7 *Si $det(A) \neq 0$, alors le système admet une unique solution. Donc pour résoudre un système d'équations linéaires on cherche à remplacer le système initial par un autre plus simple, ayant la même solution. Pour cela, il existe deux grandes familles de méthode de résolution :* **les méthodes directes** *et* **les méthodes itératives**.

2.1 Méthodes directes

Les méthodes directes de résolution des systèmes linéaires sont des méthodes dans lesquelles la solution est obtenue de façon exacte en un nombre fini d'opérations soit par triangularisation soit par décomposition de la matrice A.

Remarque 8 *On n'utilise jamais les formules de Cramer car elles nécessitent le calcul du déterminant qui requièrent un nombre trop important d'opérations élémentaires.*

Remarque 9 *On ne calcule pas* A^{-1} *pour ensuite faire* $x = A^{-1}b$ *avec* $x = (u_1 \cdots u_n)^T$. *En effet, le calcul de* $A^{-1} = (u_1, u_2, \cdots, u_n)$, *décrit par ces vecteurs colonnes, revient déjà à résoudre* n *systèmes linéaires :* $Ax_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = e_i$, *ce qui est bien-sûre trop coûteux pour résoudre un seul système.*

2.1.1 Méthode de Gauss

Johann Carl Friedrich Gauss (30 avril 1777-23 février 1855) est un mathématicien, astronome et physicien allemand. Doté d'un grand génie, il a apporté de très importantes contributions à ces trois sciences, surnommé "***Le prince des mathématiciens***", il est considéré comme l'un des plus grand mathématiciens de tous les temps.

Principe de la méthode

La méthode consiste à construire une matrice inversible M telle que MA soit une matrice triangulaire supérieure. Le système initial $Ax = b$ est alors équivalent par multiplicité à gauche par M au système triangulaire $MAx = Mb$. Donc la méthode de Gauss se décompose en trois étapes :

1$^{\text{ière}}$ **étape** : Trouver une matrice inversible M telle que la matrice MA soit triangulaire supérieure.

2$^{\text{ième}}$ **étape** : Calcul simultané du vecteur Mb.

3$^{\text{ième}}$ **étape** : Résolution du système triangulaire supérieure $MAx = Mb$ (remontée).

Commençons par décrire étape par étape la méthode dans sa forme de base, en considérant le système linéaire $Ax = b$, A étant une matrice inversible d'ordre n. Supposons de plus que le terme a_{11} de la matrice A est non nul. Nous pouvons alors éliminer l'inconnu x_2 des lignes 2 à n du système en leur retranchant respectivement la première ligne multiplier par le coefficient $\dfrac{a_{i1}}{a_{11}}$, $i = \overline{2, n}$.

En notant $A^{(2)}$ et $b^{(2)}$ la matrice et le vecteur du second membre résultant de ces opérations, on a alors

$$a_{ij}^{(2)} = a_{ij} - \frac{a_{i1}}{a_{11}} a_{1j} \quad \text{et} \quad b_i^{(2)} = b_i - \frac{a_{i1}}{a_{11}} b_1,\ i, j = \overline{2, n}$$

et le système $A^{(2)}x = b^{(2)}$ est équivalent au système de départ. En supposant que le coefficient diagonal $a_{22}^{(2)} \neq 0$ de $A^{(2)}$, on peut procéder à l'élimination de l'inconnu x_3 des lignes 3 à n de ce système, et ainsi de suite.

On obtient sous l'hypothèse $a_{kk}^{(2)} \neq 0$, $k = \overline{1, n-1}$, une suite finie de matrices $A^{(k)}$, $k \geq 2$ de la forme :

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

et telles que le système $A^{(k)}x = b^{(k)}$ est triangulaire supérieure. Les quantités $a_{kk}^{(k)}$, $k = \overline{1, n-1}$ sont appelées pivots et l'on a supposé qu'elles étaient non nulles à chaque étape, les formules permettant de passer du $k^{\text{ième}}$ système linéaire au $(k+1)^{\text{ième}}$ se résumant à :

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} a_{kj}^{(k)} \quad \text{et} \quad b_i^{(k+1)} = b_i^{(k)} - \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} b_k^{(k)},\ i, j = \overline{2, n}$$

En pratique, pour une résolution "à la main" d'un système linéaire $Ax = b$ par cette méthode, il est commode d'appliquer l'élimination à la matrice "augmentée" $(A\ b)$.

On remarque qu'il faut s'assurer que $a_{kk}^{(k)}$ n'est pas nul ni même trop petit avant chaque opérations. Pour cela, on adopte une stratégie du pivot.

1. **Stratégie du pivot partiel** : on détermine l'élément $a_{ik}^{(k)}$ tel que $k \leq i \leq n$ et

$$\left| a_{ik}^{(k)} \right| = \max_{k \leq p \leq n} \left| a_{pk}^{(k)} \right|$$

   on permute alors les lignes d'indices i et k pour amener en position de pivot l'élément $a_{ik}^{(k)}$.

2. **Stratégie du pivot total** : on détermine l'élément $a_{ij}^{(k)}$ tel que $k \leq i, j \leq n$ et

$$\left| a_{ij}^{(k)} \right| = \max_{k \leq l, p \leq n} \left| a_{lp}^{(k)} \right|$$

   on effectue alors des permutations de lignes et de colonnes pour ramener en position de pivot l'élément $a_{ij}^{(k)}$.

Erreur d'arrondi et choix du pivot

Revenons à présent sur le choix du pivot à chaque étape de l'élimination. Si à la $k^{\text{ième}}$ étape l'élément $a_{kk}^{(k)}$ est non nul, il semble naturel de l'utiliser comme pivot. Cependant, à cause des erreurs d'arrondi existant en pratique, cette manière de procéder est en générale à proscrire, comme l'illustre l'exemple suivant :

Exemple numérique : Supposons que les calculs sont effectués en virgule flottante dans le système décimal, considérons le système

$$\begin{pmatrix} 10^{-4} & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

dont la solution est $x_1 = 1.0001$ et $x_2 = 0.9999$. En prenant $a_{11} = 10^{-4}$, on obtient le système triangulaire suivant

$$\begin{pmatrix} 10^{-4} & 1 \\ 0 & 9999 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 9998 \end{pmatrix}$$

mais la version machine prend la version arrondie

$$\begin{pmatrix} 10^{-4} & 1 \\ 0 & 10^4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 10^4 \end{pmatrix}$$

ainsi la solution numérique calculée est

$$x_1 = 0 \quad \text{et} \quad x_2 = 1$$

qui est très éloignée de la véritable solution du système.

Si par contre, on commence par échanger les deux équations du système pour utiliser le nombre 1 comme pivot, on trouve

$$\begin{pmatrix} 1 & 1 \\ 0 & 0.9999 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0.9998 \end{pmatrix}$$

d'où la solution numérique calculée est

$$x_2 = 1 \quad \text{et} \quad x_1 = 2 - x_2 = 1$$

ce qui est correct à 10^{-4} près.

En général, le changement de pivot n'a pas un effet aussi spectaculaire que dans l'exemple ci-dessus, mais il en demeure pas au moins essentiel lorsque les calculs sont effectués en arithmétique à virgule flottante. De fait pour éviter la propagation d'erreurs et obtenir une meilleure stabilité numérique de la méthode, on peut utiliser pour le choix du pivot : *pivot partiel* et *pivot total*.

Calcul du nombre d'opérations

La mise en oeuvre de la méthode du pivot nécessite $\dfrac{n^3 - n}{3}$ additions-soustractions, autant de multiplications et $\dfrac{n(n-1)}{2}$ divisions, soit un total d'opérations de l'ordre de $\dfrac{2n^3}{3}$.

Quantification de l'erreur

Les solutions obtenues numériquement ne sont que des valeurs approchées de la solution exacte. En général, si x est la solution exacte du système $Ax = b$ et $\tilde{x}$ est la solution approchée calculée, nous avons donc les relations suivantes : $Ax - b = 0$ et $A\tilde{x} - b = r$ où r est appelé le *résidu*.

Soit z l'erreur commise sur la solution $x = \tilde{x} + z$. On montre facilement que z vérifie la relation $Az + r = 0$. Ainsi, on peut calculer l'erreur z à partir du résidu r.

Remarque 10 *Malgré les avantages de la méthode de Gauss, elle n'est vraiment utilisée en analyse numérique que pour résoudre des problèmes où la matrice A n'a aucune propriété particulière. Sinon, des algorithmes performants lui seront préférés (factorisation LU et Cholesky).*

2.1.2 Méthode LU

Principe de la méthode

Le principe de la méthode est de se ramener à résoudre deux systèmes triangulaires. Soit le système $Ax = b$, où A est une matrice dont tous les mineurs principaux sont non nuls, alors pour la résolution on procède par les étapes suivantes :

1. **Décomposition de** A : $A = LU$ avec L matrice triangulaire inférieure à diagonale unité (c'est-à-dire : $l_{ii} = 1$) et U triangulaire supérieure.
2. **Résolution du système triangulaire inférieure** : $Ly = b$ (déscente).
3. **Résolution du système triangulaire supérieure** : $Ux = y$ (remontée).

$$Ax = b \overset{A=LU}{\longrightarrow} LUx = b \longrightarrow \begin{cases} Ly = b \\ Ux = y \end{cases}$$

Notons

$$l_{ik} = \frac{a_{ik}^k}{a_{kk}^k}$$

Cette méthode est intéressante lorsqu'on doit résoudre beaucoup de systèmes du types :

$$Ax_j = b_j$$

car L et U sont calculées une fois pour toute et nous n'avons plus qu'à les utiliser pour chacun des cas.

L'existence et l'unicité de la décomposition

Théorème 9 *Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice dont tous les mineurs principaux sont non nuls. Il existe un unique couple de matrices (L, U), avec L triangulaire inférieure à diagonale unité et U triangulaire supérieure tel que $A = LU$.*

Proposition 2 *Si A est une matrice à diagonale strictement dominante alors elle admet une factorisation LU.*

Remarque 11 *La décomposition LU permet de calculer facilement le déterminant de A, qui est égal au produit des éléments diagonaux de la matrice U.*
En effet :

$$det(A) = det(LU) = det(L) \times det(U) = 1 \times det(U) = \prod_{i=1}^{n} u_{ii}$$

Théorème 10 *Soit A une matrice d'ordre n inversible. Alors il existe une matrice P (resp. des matrices P et Q) tenant compte d'une stratégie de pivot partiel (resp. total), une matrice triangulaire inférieure L, dont les éléments sont inférieures ou égaux à 1 en valeur absolue, et une matrice triangulaire supérieure U telle que :*

$$PA = LU \qquad (\textit{resp. } PAQ = LU)$$

Démonstraion. voir [5]

Exemple 1 *Considérons la matrice inversible*

$$A = A^{(1)} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 7 & 8 & 9 \end{pmatrix}$$

On a alors

$$A^{(2)} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & -1 \\ 0 & -6 & -12 \end{pmatrix}$$

et l'élimination s'interrompe à l'issue de la seconde étape, le pivot $a_{22}^{(2)} = 0$. En échangeant la deuxième et la troisième ligne, on arrive à

$$A^{(2)} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -12 \\ 0 & 0 & -1 \end{pmatrix} = U$$

Les matrices d'élimination aux deux étapes effectuées sont respectivement

$$E^{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -7 & 0 & 1 \end{pmatrix} \quad \textit{et } E^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

et la matrice P est la matrice de permutation

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

d'où

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 7 & 0 & 1 \end{pmatrix}$$

Dans le cas d'une factorisation de type $PA = LU$, la résolution du système linéaire $Ax = b$ après factorisation s'effectue en appliquant tout d'abord la matrice de permutation P au vecteur b pour obtenir le second membre Pb et en résolvant ensuite le système $Ly = Pb$ par une méthode de descente, puis le système $Ux = y$ par une méthode de remontée.

Complexité de l'algorithme

Une décomposition LU pour la résolution d'un système linéaire nécessite : $\dfrac{n^3 - n}{3}$ multiplications, $\dfrac{2n^3 - 3n^2 + n}{6}$ additions, à l'étape de décomposition en un produit LU. Soit donc un total de l'ordre de $\dfrac{n^3}{3}$, de plus la remontée et la descente nécessitent un ordre de n^2. On note que pour n grand on néglige les puissances de n inférieures à 3, ce qui permet de dire que le travail essentiel se trouve dans la décomposition elle-même.

2.1.3 Méthode de Cholesky

André-Louis Cholesky (15 octobre 1875-31 août 1918) était un mathématicien et un militaire français. Il a effectué sa carrière dans les services géographiques de l'armée. C'est pour résoudre des problèmes de moindres carrés posés par la géodésie que Cholesky inventa la décomposition qui porte aujourd'hui son nom.

Principe de la méthode

La décomposition de Cholesky n'est qu'une variante de la décomposition LU pour des matrices particulières. Dans toute cette section $A \in \mathcal{M}_{\mathrm{n}}(\mathbb{R})$ est une matrice carrée réelle symétrique définie positive. Pour la résolution du système $Ax = b$, on procède par les étapes suivantes :

1. **Décomposition de** A : $A = LL^T$ avec L matrice triangulaire inférieure à éléments diagonaux positifs.
2. **Résolution du système triangulaire inférieure** : $Ly = b$ (déscente).
3. **Résolution du système triangulaire supérieure** : $L^T x = y$ (remontée).

$$Ax = b \overset{A=LL^T}{\longrightarrow} LL^T x = b \longrightarrow \begin{cases} Ly = b \\ L^T x = y \end{cases}$$

On détermine $\{l_{ij}\}$ à l'aide du système d'équations

$$a_{ij} = \sum_{k=1}^{i} l_{ik} l_{jk}, \qquad 1 \leq i \leq j \leq n$$

où

$$L = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ l_{n-1,1} & & \ddots & 0 \\ l_{n1} & \cdots & \cdots & l_{nn} \end{pmatrix}$$

Donc la résolution peut se faire de proche en proche, colonne par colonne, comme suit :

$$\begin{cases} l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2} & \text{lorsque } i = j \\ \\ l_{ji} = \dfrac{a_{ij} - \sum_{k=1}^{i-1} l_{ik} l_{jk}}{l_{ii}} & \text{lorsque } i > j \end{cases}$$

L'existence et l'unicité de la décomposition

Théorème 11 *Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive. Alors il existe (au moins) une matrice réelle triangulaire inférieure L telle que $A = LL^T$. De plus, on peut imposer que les éléments diagonaux de la matrice L soient strictement positifs et la décomposition $A = LL^T$ correspondante est alors unique.*

Évaluation du nombre d'opérations

La méthode de Cholesky ne requiert que : $\frac{n^3}{6}$ multiplications, $\frac{n^3}{6}$ additions, $\frac{n^2}{2}$ divisions. La méthode de Cholesky est donc fortement conseillée (plutôt que Gauss) quand la matrice est symétrique définie positive. Elle a aussi l'avantage de se révéler plus stable par rapport aux erreurs d'arrondi. Notons également l'estimation a priori $|l_{ik}| < \sqrt{a_{ii}}$, $\forall i = 1, \cdots, n$, $\forall k = 1, \cdots, i$, provenant de la relation

$$\sum_{k=1}^{i} l_{ik}^2 = a_{ii}$$

et qui assure que les l_{ik} ne deviendront jamais trop grand.

2.1.4 Conclusion

Dans ce chapitre on a vu les quatre méthodes directes de résolution des systèmes linéaires qu'on a étudié théoriquement. Voici un tableau récapitulatif résumant notre travail et dans lequel on compare le coût des quatre méthodes en citant quelques avantages et inconvénients.

La méthode	Le coût	Avantages	Inconvénients
L'élimination de Gauss	$O\left(\frac{2n^3}{3}\right)$	- Résolution d'un seul système linéaire (comparaison avec les autres méthodes) ; - Utilisation des opérations élémentaires d'additions et de multiplications.	- On échange la structure de A ; - Blocage dans le cas d'un pivot nul sans pivotement ; - L'implémentation sur machine donne généralement de mauvais résultats à cause des erreurs d'arrondi qui s'accumulent.
Factorisation LU	$O\left(\frac{n^3}{3}\right)$	- Cette méthode est très utile lors de la résolution d'une suite de systèmes de même matrice A où seul le vecteur b change ; - Utilisation des opérations élémentaires d'additions et multiplications.	- La résolution d'un seul système linéaire nécessite deux systèmes linéaires triangulaires de matrices L et U (stockage de deux matrices L et U) ; -Particularité de la matrice carrée A (mineurs principaux non nuls).
Cholesky	$O\left(\frac{4n^3}{3}\right)$	- Cette méthode est très utile lors de la résolution d'une suite de systèmes de même matrice A où seul le vecteur b change ; - Elle nécessite le calcul d'une seule matrice L ; - Elle demande une place mémoire deux fois inférieure à celle de la factorisation LU car on ne doit stocker que L.	- Elle ne permet pas la résolution linéaire d'un système quelconque (particularité de la matrice A carrée symétrique définie positive) ; - La résolution d'un seul système linéaire nécessite la résolution de deux systèmes linéaires triangulaires de matrices L et L^T ;

D'après le tableau on constate que la méthode de Cholesky est la moins coûteuse parmi les autres.

2.1.5 Exercices corrigés

Exercice 26 *Soit le système d'équations suivant*

$$\begin{cases} 3x - y + 2z & = & r \\ -x + 2y - 3z & = & s \\ x + 2y + z & = & t \end{cases}$$

avec, $r,\ \ s, t \in \mathbb{R}$.

1. *Ecrire ce système sous la forme matricielle* $AX = b$.
2. *Résoudre le système par la méthode de Gauss.*

Solution

1. Soit le système linéaire suivant :

$$\underbrace{\begin{pmatrix} 3 & -1 & 2 \\ -1 & 2 & -3 \\ 1 & 2 & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x \\ y \\ z \end{pmatrix}}_{X} = \underbrace{\begin{pmatrix} r \\ s \\ t \end{pmatrix}}_{b}$$

Commençons d'abord par rendre la matrice $(A|b)$ triangulaire supérieure

$$(A|b) = \begin{pmatrix} 3 & -1 & 2 & | & r \\ -1 & 2 & -3 & | & s \\ 1 & 2 & 1 & | & t \end{pmatrix} \begin{array}{l} \to L_1 \\ \to L_2 \\ \to L_3 \end{array}$$

$$\longrightarrow \begin{pmatrix} 3 & -1 & 2 & | & r \\ & & & | & \\ -0 & \frac{5}{3} & -\frac{7}{3} & | & s + \frac{r}{3} \\ & & & | & \\ 0 & \frac{7}{3} & \frac{1}{3} & | & t - \frac{r}{3} \end{pmatrix} \begin{array}{l} \to L_1 \\ \\ \to L_2 = L_2 - \frac{(-1)}{3} L_1 \\ \\ \to L_3 = L_3 - \frac{1}{3} L_1 \end{array}$$

$$\longrightarrow \begin{pmatrix} 3 & -1 & 2 & | & r \\ & & & | & \\ 0 & \frac{5}{3} & -\frac{7}{3} & | & s + \frac{r}{3} \\ & & & | & \\ 0 & 0 & \frac{54}{15} & | & \frac{15t - 12r - 21s}{15} \end{pmatrix} \begin{array}{l} \to L_1 \\ \\ \to L_2 \\ \\ \to L_3 = L_3 - \frac{7}{5} L_2 \end{array}$$

On obtient le système d'équations suivant

$$\begin{cases} 3x - y + 2z & = & r \\ \\ \frac{5}{3}y - \frac{7}{3}z & = & s + \frac{9}{3} \\ \\ \frac{54}{15}z & = & \frac{15t - 12r - 21s}{15} \end{cases}$$

En résolvant ce système on obtient la solution suivante :

$$x = \begin{pmatrix} x \\ \\ y \\ \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{18}(8r + 5s - t) \\ \\ \frac{1}{18}(-2r + s + 7t) \\ \\ \frac{1}{18}(-4r + -7s + 5t) \end{pmatrix}$$

Exercice 27 *Soit le système linéaire $Ax = b$*

$$\begin{pmatrix} 6 & 1 & 1 \\ 2 & 4 & 0 \\ 1 & 2 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 12 \\ 0 \\ 6 \end{pmatrix}$$

1. *Résoudre le système linéaire par la méthode d'élimination de Gauss.*
2. *Factoriser la matrice A et résoudre le système linéaire.*

Solution

1. Commençons d'abord par rendre la matrice $(A|b)$ triangulaire supérieure

$$(A|b) = \begin{pmatrix} 6 & 1 & 1 & | & 12 \\ -2 & 4 & -0 & | & 0 \\ 1 & 2 & 6 & | & 6 \end{pmatrix} \quad \begin{matrix} \to L_1 \\ \to L_2 \\ \to L_3 \end{matrix}$$

$$\longrightarrow \begin{pmatrix} 6 & 1 & 1 & | & 12 \\ & & & | & \\ 0 & \frac{11}{3} & -\frac{1}{3} & | & -4 \\ & & & | & \\ 0 & \frac{11}{6} & \frac{35}{6} & | & 4 \end{pmatrix} \quad \begin{matrix} \to L_1 \\ \\ \to L_2 = L_2 - \frac{-1}{3}L_1 \\ \\ \to L_3 = L_3 - \frac{1}{6}L_1 \end{matrix}$$

$$\longrightarrow \left(\begin{array}{ccc|c} 6 & 1 & 1 & 12 \\ & & & \\ 0 & \frac{11}{3} & -\frac{1}{3} & -4 \\ & & & \\ 0 & 0 & 6 & 6 \end{array}\right) \quad \begin{array}{l} \to L_1 \\ \\ \to L_2 \\ \\ \to L_3 = L_3 - \frac{1}{2}L_2 \end{array}$$

On obtient le système d'équations suivant

$$\left\{\begin{array}{lcl} 6x_1 + x_2 + x_3 & = & 12 \\ \frac{11}{3}x_2 - \frac{1}{3}x_3 & = & -4 \\ 6x_3 & = & 6 \end{array}\right.$$

En résolvant ce système on obtient la solution suivante :

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$$

2. On pose $A = LU$, avec

$$U = \begin{pmatrix} 6 & 1 & 1 \\ 0 & \frac{11}{3} & -\frac{1}{3} \\ 0 & 0 & 6 \end{pmatrix} \text{ et } L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{1}{6} & \frac{1}{2} & 1 \end{pmatrix}$$

on a alors

$$LUx = b \Longleftrightarrow \left\{\begin{array}{l} y = Ux \\ Ly = b \end{array}\right.$$

$$Ly = b \Longleftrightarrow \left\{\begin{array}{l} y_1 = 12 \\ \frac{1}{3}y_1 + y_2 = 0 \\ \frac{1}{6}y_1 + \frac{1}{2}y_2 + y_3 = 6 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} y_1 = 12 \\ y_2 = -4 \\ y_3 = 6 \end{array}\right.$$

$$Ux = y \Longleftrightarrow \left\{\begin{array}{l} 6x_1 + x_2 + x_3 = 12 \\ \frac{11}{3}x_2 - \frac{1}{3}x_3 = -4 \\ 6x_3 = 6 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} x_1 = 2 \\ x_2 = -1 \\ x_3 = 1 \end{array}\right.$$

Alors la solution de ce système est

$$x = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$$

Exercice 28 *Résoudre lorsqu'il est possible les deux systèmes $Ax = b$ et $Bx = b$, par la factorisation LU avec :*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 7 & 8 & 9 \end{pmatrix}, B = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 2 & 4 & 5 \end{pmatrix} \quad et \quad b = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Solution

1. Soit la matrice

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 7 & 8 & 9 \end{pmatrix}$$

Vérifions si on peut faire la factorisation LU de A.

$$\Delta_3 = \det(A) = \det\begin{pmatrix} 4 & 5 \\ 8 & 9 \end{pmatrix} - 2\det\begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix} + 7\det\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}$$
$$= -2 \neq 0$$

Par suite A est inversible.

$$\Delta_1 = \det(A_1) = 1 \neq 0$$

$$\Delta_2 = \det(A_2) = \det\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = 0$$

Donc on ne peut pas factoriser A sans utiliser la technique du pivot. En effet,

$$A = \begin{pmatrix} 1 & 2 & 3 \\ \\ 2 & 4 & 5 \\ \\ 7 & 8 & 9 \end{pmatrix} \quad \begin{matrix} \to L_1 \\ \\ \to L_2 = L_2 - \dfrac{-2}{1} L_1 \\ \\ \to L_3 = L_3 - \dfrac{7}{1} L_1 \end{matrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & -1 \\ 0 & -6 & -12 \end{pmatrix}$$

Soit la matrice

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 2 & 4 & 5 \end{pmatrix}$$

Vérifions si on peut faire la factorisation LU de B.

$$\Delta_3 = \det(B) = \det\begin{pmatrix} 8 & 9 \\ 4 & 5 \end{pmatrix} - 7\det\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} + 2\det\begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix}$$
$$= 6 \neq 0$$

Donc B est inversible.

$$\Delta_1 = \det(B_1) = 1 \neq 0$$
$$\Delta_2 = \det(B_2) = \det\begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix} = -6 \neq 0$$

Donc on peut factoriser B. La factorisation LU de la matrice B est donc

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 2 & 4 & 5 \end{pmatrix} \qquad \begin{matrix} \to L_1 \\ \to L_2 \\ \to L_3 \end{matrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -12 \\ 0 & 0 & -1 \end{pmatrix} \qquad \begin{matrix} \to L_1 \\ \to L_2 = L_2 - 7L_1 \\ \to L_3 = L_3 - 2L_2 \end{matrix}$$

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 7 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}, U = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -12 \\ 0 & 0 & -1 \end{pmatrix}$$

2. Résoudre le système $Bx = b$.

$$LUx = b \Longleftrightarrow \begin{cases} y = Ux \\ Ly = b \end{cases}$$

$$Ly = b \Longleftrightarrow \begin{cases} y_1 = 1 \\ \\ 7y_1 + y_2 = 0 \\ \\ 2y_1 + y_3 = 1 \end{cases} \Longrightarrow \begin{cases} y_1 = 1 \\ y_2 = -7 \\ y_3 = -1 \end{cases}$$

$$Ux = y \Longleftrightarrow \begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ \\ -6x_2 - 12x_3 = -7 \\ \\ -x_3 = -1 \end{cases} \Longrightarrow \begin{cases} x_1 = -\dfrac{1}{3} \\ x_2 = -\dfrac{5}{6} \\ x_3 = 1 \end{cases}$$

Alors la solution de ce système est

$$x = \begin{pmatrix} -\dfrac{1}{3} \\ \\ -\dfrac{5}{6} \\ \\ 1 \end{pmatrix}$$

Exercice 29 *On considère le système $Ax = b$, où*

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 2 & 1 & 1 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 8 \\ 4 \end{pmatrix}$$

1. *Résoudre le système par la méthode de LU.*
2. *Résoudre le système $A^2x = b$ sans calculer A^2.*

Solution

1. Calculons d'abord les matrices L et U, pour cela on va faire la triangularisation de la matrice A.

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 2 & 3 \\ 2 & 1 & 1 \end{pmatrix} \quad \begin{array}{l} \to L_1 \\ \to L_2 \\ \to L_3 \end{array}$$

$$\longrightarrow \begin{pmatrix} 3 & 2 & 1 \\ 0 & \frac{4}{3} & \frac{8}{3} \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{pmatrix} \quad \begin{array}{l} \to L_1 \\ \to L_2 = L_2 - \frac{1}{3}L_1 \\ \to L_3 = L_3 - \frac{2}{3}L_1 \end{array}$$

$$\longrightarrow \begin{pmatrix} 3 & 2 & 1 \\ 0 & \frac{4}{3} & \frac{8}{3} \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{array}{l} \to L_1 \\ \to L_2 = L_2 \\ \to L_3 = L_3 + \frac{1}{4}L_2 \end{array}$$

La factorisation LU de la matrice A est donc

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 7 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}, U = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -12 \\ 0 & 0 & -1 \end{pmatrix}$$

Résoudre le système $Ax = b$.

$$LUx = b \Longleftrightarrow \begin{cases} y = Ux \\ Ly = b \end{cases}$$

$$Ly = b \Longleftrightarrow \begin{cases} y_1 = 4 \\ \frac{1}{3}y_1 + y_2 = 8 \\ \frac{2}{3}y_1 - \frac{1}{4}y_2 + y_3 = 4 \end{cases} \Longrightarrow \begin{cases} y_1 = 4 \\ y_2 = \frac{20}{3} \\ y_3 = 3 \end{cases}$$

$$Ux = y \Longleftrightarrow \left\{\begin{array}{l} 3x_1 + 2x_2 + x_3 = 4 \\ \\ \dfrac{4}{3}x_2 + \dfrac{8}{3}x_3 = \dfrac{20}{3} \\ \\ x_3 = 3 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} x_1 = 1 \\ x_2 = -1 \\ x_3 = 3 \end{array}\right.$$

Alors la solution de ce système est

$$x = \begin{pmatrix} 1 \\ \\ -1 \\ \\ 3 \end{pmatrix}$$

2.

$$A^2x = b \Longleftrightarrow AAx = b \Longleftrightarrow \left\{\begin{array}{l} z = Ax \\ Az = b \end{array}\right.$$

Alors

$$z = x = \begin{pmatrix} 1 \\ \\ -1 \\ \\ 3 \end{pmatrix}$$

car le système $Ax = b$ admet une unique solution.
Il suffit donc de résoudre le système $Ax = z$ avec le même L et U.

$$\left\{\begin{array}{l} y = Lz \\ Ux = y \end{array}\right.$$

$$Ly = z \Longleftrightarrow \left\{\begin{array}{l} y_1 = 1 \\ \\ \dfrac{1}{3}y_1 + y_2 = -1 \\ \\ \dfrac{2}{3}y_1 - \dfrac{1}{4}y_2 + y_3 = 3 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} y_1 = 1 \\ y_2 = -\dfrac{4}{3} \\ y_3 = 2 \end{array}\right.$$

$$Ux = y \Longleftrightarrow \left\{\begin{array}{l} 3x_1 + 2x_2 + x_3 = 1 \\ \\ \dfrac{4}{3}x_2 + \dfrac{8}{3}x_3 = -\dfrac{4}{3} \\ \\ x_3 = 2 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} x_1 = 3 \\ x_2 = -5 \\ x_3 = 2 \end{array}\right.$$

Alors la solution de ce système est

$$x = \begin{pmatrix} 3 \\ \\ -5 \\ \\ 2 \end{pmatrix}$$

Exercice 30 *Soit α un paramètre réel et soient les matrices A_α, P et le vecteur b définis par*

$$A_\alpha = \begin{pmatrix} 2 & 4 & 1 \\ \alpha & -2 & -1 \\ 2 & 3 & 2 \end{pmatrix}, P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \text{ et } b = \begin{pmatrix} 0 \\ -\frac{3}{2} \\ -1 \end{pmatrix}$$

1. *À quelle condition sur α, la matrice A_α est inversible ?*
2. *À quelle condition sur α, la matrice A_α admet-elle une décomposition LU (sans pivot) ?*
3. *Soit $\alpha = -1$. Calculer, si elle existe, la décomposition LU de la matrice $M = PA_\alpha$.*
4. *Soit $\alpha = -1$. Résoudre le système linéaire $Ax = b$ en résolvant le système linéaire $Mx = Pb$.*

Solution

1. La matrice A_α est inversible si et seulement si $\det(A_\alpha) \neq 0$.

$$\det(A_\alpha) = 2\det\begin{pmatrix} -2 & -1 \\ 3 & 2 \end{pmatrix} - \alpha\det\begin{pmatrix} 4 & 1 \\ 3 & 2 \end{pmatrix} + 2\det\begin{pmatrix} 4 & 1 \\ -2 & -1 \end{pmatrix}$$
$$= -6 - 5\alpha$$

Donc la matrice A_α est inversible si et seulement si $\alpha \neq -\frac{6}{5}$.

2. La matrice A_α admet une décomposition LU si et seulement si les mineurs principaux de A_α sont non nuls.

$$A_1 = 2 \neq 0$$
$$A_2 = \det\begin{pmatrix} 2 & 4 \\ \alpha & -2 \end{pmatrix}$$
$$= -4(1+\alpha)$$

Par conséquent, la matrice A_α admet une décomposition LU (sans pivot) si et seulement si $\alpha \neq -1$.

3. Si $\alpha = -1$, la matrice A_α n'admet pas de décomposition LU sans pivot. La matrice P échange les lignes 2 et 3 de la matrice A et on obtient la matrice

$$M = PA_{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 4 & 1 \\ -1 & -2 & -1 \\ 2 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 1 \\ 2 & 3 & 2 \\ -1 & -2 & -1 \end{pmatrix}$$

La matrice M admet une décomposition LU (sans pivot) et l'on a
par conséquent, on obtient la décomposition LU suivante de la matrice M

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix}, U = \begin{pmatrix} 2 & 4 & 1 \\ 0 & -1 & -1 \\ 0 & 0 & -\frac{1}{2} \end{pmatrix}$$

$$M = \begin{pmatrix} 2 & 4 & 1 \\ 2 & 3 & 2 \\ -1 & -2 & -1 \end{pmatrix} \begin{array}{l} \to L_1 \\ \to L_2 \\ \to L_3 \end{array}$$

$$\longrightarrow \begin{pmatrix} 2 & 4 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & -\frac{1}{2} \end{pmatrix} \begin{array}{l} \to L_1 \\ \to L_2 = L_2 - \frac{2}{2}L_1 \\ \to L_3 = L_3 - \frac{(-1)}{2}L_1 \end{array}$$

4. Pour résoudre le système linéaire $Mx = Pb$, il suffit de résoudre les deux systèmes triangulaires suivants :

$$Ly = b \Longleftrightarrow \begin{cases} y_1 = 0 \\ y_1 + y_2 = -1 \\ -\frac{1}{2}y_1 + y_3 = -\frac{3}{2} \end{cases} \Longrightarrow \begin{cases} y_1 = 0 \\ y_2 = -1 \\ y_3 = -\frac{3}{2} \end{cases}$$

$$Ux = y \Longleftrightarrow \begin{cases} 2x_1 + 4x_2 + x_3 = 0 \\ -x_2 + x_3 = -1 \\ -\frac{1}{2}x_3 = -\frac{3}{2} \end{cases} \Longrightarrow \begin{cases} x_1 = -\frac{19}{2} \\ x_2 = 4 \\ x_3 = 3 \end{cases}$$

Alors la solution de ce système est

$$x = \begin{pmatrix} -\frac{19}{2} \\ 4 \\ 3 \end{pmatrix}$$

Exercice 31 *On considère la matrice*

$$A = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}, \; b = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

1. *Montrer que A est symétrique définie positive.*
2. *Trouver une matrice triangulaire inférieure $L = (l_{ij}) \in \mathcal{M}_n(\mathbb{R})$ telle que $l_{ii} > 0$ et $LL^T = A$ (factorisation de Cholesky de A).*
3. *Résoudre le système linéaire $Ax = b$. En déduire le déterminant de A.*

Solution

1. On a

$$A^T = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix} = A$$

Alors A est symétrique.

$$\Delta_1 = a_{11} = 1 > 0$$

$$\Delta_2 = \det \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} = \frac{3}{4} > 0$$

$$\Delta_3 = \det \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} - \frac{1}{2} \det \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} + \frac{1}{2} \det \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix} = \frac{1}{2} > 0$$

Alors A est définie positive.

2.

$$l_{11} = \sqrt{a_{11}} = 1$$

$$l_{21} = \frac{a_{21}}{l_{11}} = \frac{1}{2}$$

$$l_{31} = \frac{a_{31}}{l_{11}} = \frac{1}{2}$$

$$l_{22} = \sqrt{a_{22} - l_{21}^2} = \frac{\sqrt{3}}{2}$$

$$l_{32} = \frac{a_{32} - l_{31}l_{21}}{l_{22}} = \frac{1}{2\sqrt{3}}$$

$$l_{33} = \sqrt{a_{33} - l_{31}^2 - l_{32}^2} = \frac{\sqrt{2}}{\sqrt{3}}$$

Alors

$$L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ \frac{1}{2} & \frac{1}{2\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}$$

3. Résoudre le système $Ax = b$. On a

$$L^T = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{1}{2\sqrt{3}} \\ 0 & 0 & \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}$$

$$LL^T x = b \Longleftrightarrow \begin{cases} y = L^T x \\ Ly = b \end{cases}$$

$$Ly = b \Longleftrightarrow \begin{cases} y_1 = 1 \\ \frac{1}{2}y_1 + \frac{\sqrt{3}}{2}y_2 = 0 \\ \frac{1}{2}y_1 + \frac{1}{2\sqrt{3}}y_2 + \frac{\sqrt{2}}{\sqrt{3}}y_3 = 1 \end{cases} \Longrightarrow \begin{cases} y_1 = 1 \\ y_2 = -\frac{1}{\sqrt{3}} \\ y_3 = \frac{2\sqrt{3}}{3\sqrt{2}} \end{cases}$$

$$L^T x = y \Longleftrightarrow \begin{cases} x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 = 1 \\ \frac{\sqrt{3}}{2}x_2 + \frac{1}{2\sqrt{3}}x_3 = -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}}x_3 = \frac{2\sqrt{3}}{3\sqrt{2}} \end{cases} \Longrightarrow \begin{cases} x_1 = 1 \\ x_2 = -1 \\ x_3 = 1 \end{cases}$$

Alors la solution de ce système est

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

$$\det(LL^T) = \det(L) \times \det(L^T) = (\det(L))^2 = \frac{1}{2}$$

2.1.6 Exercices supplémentaires

Exercice 32 *Résoudre le système suivant par la méthode de Gauss, avec $a, b \in \mathbb{R}$.*

$$\begin{cases} ax + by + z &= 1 \\ x + aby + z &= b \\ x + by + az &= 1 \end{cases}$$

Exercice 33 *On considère la matrice*

$$A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & -12 & \\ -3 & 2 & 0 \end{pmatrix}, \; b = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

1. *Montrer que A est symétrique définie positive.*
2. *Résoudre le système linéaire par la méthode de Cholesky.*
3. *En déduire le déterminant de A.*

2.2 Méthodes itératives

2.2.1 Le principe

Étant donné le système $Ax = b$ à résoudre, les méthodes itératives de résolution des systèmes linéaires consistent à calculer les valeurs successives d'une suite de vecteurs $x^{(k)}$ convergeant vers la solution x quand $k \longrightarrow \infty$.

$$\lim_{k \to +\infty} x^{(k)} = x \qquad (1)$$

On décompose A en $A = M - N$ où M est une matrice inversible alors

$$Ax = b \Longleftrightarrow Mx = Nx + b$$

conduit à l'itération

$$Mx^{(k+1)} = Nx^{(k)} + b, \qquad k \geq 0$$
$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b$$

Alors

$$M^{-1}N = B \text{ et } M^{-1}b = c$$

La suite $(x^{(k)})_{k\in\mathbb{N}}$ est obtenue à partir d'un vecteur initial arbitraire $x^{(0)}$, par une relation de récurrence de la forme

$$x^{(k+1)} = Bx^{(k)} + c, \quad \forall k \in \mathbb{N} \qquad (2)$$

où la matrice B est carrée, appelée matrice d'itération de la méthode, et le vecteur c dépendant de la matrice A et du second membre de b du système à résoudre.

2.2.2 Étude de la convergence

Définition 3 *On dit que la méthode itérative est convergente si l'on a* (1) *pour tout initialisation $x^{(0)}$ dans $\mathbb{R}^n$.*

Définition 4 *Une méthode itérative de la forme (2) est dite consistante avec le système linéaire $Ax = b$ si B et c vérifient*

$$x = Bx + c$$

le vecteur x étant la solution du système $Ax = b$, ou de manière équivalente

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

Définition 5 *On appelle erreur (resp. résidu) à l'itération k ($k \in \mathbb{N}$) de la méthode itérative le vecteur*

$$e^{(k)} = x^{(k)} - x \text{ (resp. } r^{(k)} = b - Ax^{(k)})$$

où $x = A^{-1}b$ est la solution du système $Ax = b$.

On déduit de ces définitions qu'une méthode itérative consistante de la forme (2) converge si et seulement si

$$\lim_{k \to +\infty} e^{(k)} = 0$$

Soit encore si

$$\lim_{k \to +\infty} r^{(k)} = \lim_{k \to +\infty} Ae^{(k)} = 0$$

Théorème 12 *Si une méthode de la forme (2) est consistante, celle ci est convergente si et seulement si $\rho(B) < 1$.*

Démonstration . Voir [5]

Lemme 1 *Pour toute matrice carrée B, et toute norme matricielle*

$$\lim_{k \to +\infty} \left\|B^k\right\|^{\frac{1}{k}} = \rho(B) \qquad (3)$$

Démonstration . Voir [4]

2.2.3 Critère d'arrêt

En pratique, il conviendrait de mettre fin aux calculs à la première itération pour laquelle l'erreur est "suffisamment petite", c'est-à-dire le premier entier naturel k tel que

$$\left\|e^{(k)}\right\| = \left\|x^{(k)} - x\right\| \leq \epsilon$$

où ϵ est une tolérance fixée et $\|.\|$ est une norme vectorielle donnée. Cependant on ne sait généralement pas évaluer l'erreur, puisque la solution x n'est pas connue, et il faut donc avoir recours à un autre critère d'arrêt. Deux possibilités s'imposent alors.

$1^{\text{ière}}$ **possibilité :** Tout d'abord, les résidus $r^{(k)} = b - Ax^{(k)}$ étant très facile à calculer, on peut tester si

$$\left\|r^{(k)}\right\| \leq \delta$$

avec δ une tolérance fixée. Puisque l'on a

$$\left\|e^{(k)}\right\| = \left\|x^{(k)} - x\right\| = \left\|x^{(k)} - A^{-1}b\right\| = \left\|A^{-1}r^{(k)}\right\| \leq \left\|A^{-1}\right\| \left\|r^{(k)}\right\|$$

on doit choisir δ tel que

$$\delta \leq \frac{\epsilon}{\|A^{-1}\|}$$

Ce critère peut par conséquent être trompeur si la norme de A^{-1} est grande et qu'on ne dispose pas d'une bonne estimation de cette dernière. Il est en général plus judicieux de considérer dans le test d'arrêt un résidu normalisé

$$\frac{\left\|r^{(k)}\right\|}{\|r^0\|} \leq \delta$$

ou encore

$$\frac{\|r^{(k)}\|}{\|b\|} \leq \delta$$

$2^{\text{ième}}$ **possibilité :** Correspond au choix de l'initialisation $x^{(0)}$. Dans ce cas on obtient le contrôle suivant de l'erreur relative

$$\frac{\|e^{(k)}\|}{\|x\|} \leq \|A^{-1}\| \frac{\|r^{(k)}\|}{\|x\|} \leq cond(A)\delta$$

où $cond(A)$ désigne le conditionnement de la matrice A relativement à la norme subordonnée $\|.\|$ considérée (voir Annexe).

Un autre critère parfois utilisé dans la pratique est basé sur l'incrément $x^{(k+1)} - x^{(k)}$,$k \in \mathbb{N}$. L'erreur d'une méthode itérative de la forme (2) vérifiant la récurrence $x^{(k+1)} = Bx^{(k)}$, $\forall k \in \mathbb{N}$, on obtient par l'utilisation de l'inégalité triangulaire

$$\|e^{(k+1)}\| \leq \|B\| \|e^{(k)}\| \leq \|B\| \left(\|e^{(k+1)}\| + \|x^{(k+1)} - x^{(k)}\|\right), \qquad \forall k \in \mathbb{N}$$

d'où

$$\|e^{(k+1)}\| \leq \frac{\|B\|}{1 - \|B\|} \left(\|x^{(k+1)} - x^{(k)}\|\right) \qquad \forall k \in \mathbb{N}$$

Théorème 13 ***(Convergence et équivalence)***
Soit $B \in \mathcal{M}_n(\mathbb{C})$ les assertions suivantes sont équivalentes :

1. $\lim_{k\to+\infty} B^k = 0$ *(autrement dit* $\lim_{k\to+\infty} \|B^k\| = 0$ *pour* $\|.\|$ *quelconque sur* $\mathcal{M}_n(\mathbb{C})$*).*
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* $\|.\|$ *quelconque sur* $\mathcal{M}_n(\mathbb{C})$*).*
3. $\rho(B) < 1$
4. *Il existe une norme matricielle telle que* $\|B\| < 1$.

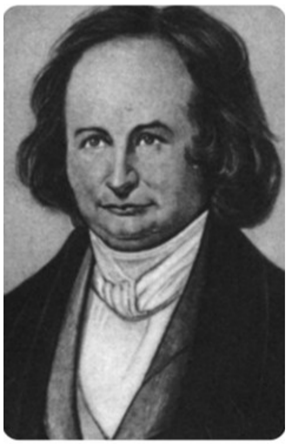
Démonstration . Voir [6]

Théorème 14 *Soit A une matrice hermitienne définie positive, si la matrice hermitienne $M^* + N$ est définie positive, alors $\rho(B) < 1$.*

Démonstration . Voir [6]

2.2.4 Méthode de Jacobi

Charles Gustave Jacob Jacobi, ou **Carl Gustav Jakob Jacobi** (10 décembre 1804 - 18 février 1851), est un mathématicien allemand surtout connu pour ses travaux sur les intégrales elliptiques, les équations aux dérivées partielles et leur application à la mécanique analytique. Il était le frère du physicien Moritz von Jacobi, découvreur de la galvanoplastie.



Principe de la méthode

Soit encore $Ax = b$ le système linéaire à résoudre. La méthode de Jacobi correspond à la décomposition de la matrice A sous la forme $A = D - E - F$ avec D matrice diagonale constituée des éléments diagonaux a_{ii} de A, E et F sont des matrices triangulaires définies par

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

On suppose que A ne possède pas de zéros sur sa diagonale. Dans cette méthode, on prend

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

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

$$\begin{cases} x^{(0)} & \text{donné} \\ x^{(k+1)} = D^{-1}(E+F)x^{(k)} + D^{-1}b & \text{pour tout entier } k \geq 0 \end{cases}$$

La matrice d'itération de Jacobi est donnée par

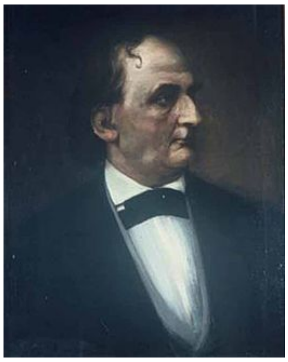
$$J = D^{-1}(E+F)$$

On observe que la méthode de Jacobi correspond à l'écriture ligne par ligne suivante

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j\neq i}^{n} a_{ij} x_j^{(k)} \right), i = 1, \cdots, n, k \in \mathbb{N}$$

2.2.5 Méthode de Gauss-Seidel

Philipp Ludwig von Seidel (24 octobre 1821 à Deux-Ponts et mort le 13 août 1896 à Munich), était un mathématicien, physicien de l'optique et astronome allemand. Seidel est connu pour la méthode de Gauss-Seidel sur la résolution des équations numériques. Il étudia les aberrations optiques du premier ordre qu'il décompose en cinq polynômes dits polynômes de Seidel.



Principe de la méthode

Dans cette méthode, on prend

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

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

$$\begin{cases} x^{(0)} & \text{donné} \\ x^{(k+1)} = (D-E)^{-1}Fx^{(k)} + (D-E)^{-1}b & \text{pour tout entier } k \geq 0 \end{cases}$$

La matrice d'itération de Gauss-Seidel est donnée par

$$G = (D-E)^{-1}F$$

On observe que la méthode de Gauss-Seidel correspond à l'écriture ligne par ligne suivante

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right), i = 1, \cdots, n, k \in \mathbb{N}$$

2.2.6 Convergence de la méthode de Jacobi et de Gauss-Seidel

Soit $e^{(k)}$ le vecteur erreur.

$$e^{(k+1)} = Je^{(k)} = J^{(k+1)}e^{(0)} \ \left(\text{ resp. } e^{(k+1)} = Ge^{(k)} = G^{(k+1)}e^{(0)}\right)$$

L'algorithme de Jacobi (resp. Gauss-Seidel) converge si

$$\lim_{k\to+\infty} \left\|e^{(k)}\right\| = 0 \Longleftrightarrow \lim_{k\to+\infty} \left\|J^{(k)}\right\| = 0 \ \left(\text{resp. } \lim_{k\to+\infty} \left\|G^{(k)}\right\| = 0\right)$$

Théorème 15 *Une condition nécessaire et suffisante pour que*

$$\lim_{k\to+\infty} \left\|J^{(k)}\right\| = 0 \ \left(\textit{resp. } \lim_{k\to+\infty} \left\|G^{(k)}\right\| = 0\right)$$

est que

$$\rho(J) < 1 \ \left(\textit{ resp.}\rho(G) < 1\right)$$

Théorème 16 *Les méthodes de Gauss-Seidel et Jacobi convergent* $\forall x^{(0)}$ *pour les systèmes linéaires dont la matrice est à diagonale strictement dominante.*

Démonstration . Voir [5]

Théorème 17 *Soit* A *une matrice tridiagonale, alors les rayons spectraux des méthodes de Gauss-Seidel et de Jacobi sont reliés par*

$$\rho(G) = [\rho(J)]^2$$

Ainsi pour ce type de matrice, les deux méthodes convergent ou divergent simultanément et si elles convergent, Gauss-Seidel est plus rapide.

Démonstration . Voir [6]

2.2.7 Méthode de Relaxation

Principe de la méthode

C'est une méthode intermédiaire entre les deux méthodes précédentes dépendant d'un paramètre w. Ce paramètre est déterminé pour obtenir une convergence plus rapide. Dans cette méthode, on prend

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

où ω est un paramètre de relaxation.

La méthode de relaxation s'exprime alors par la suite

$$\begin{cases} x^{(0)} & \text{donné} \\ x^{(k+1)} = \left(\frac{1}{\omega}D - E\right)^{-1}\left(\frac{1-\omega}{\omega}D + F\right)x^{(k)} + \left(\frac{1}{\omega}D - E\right)^{-1} b & \text{pour tout entier } k \geq 0 \end{cases}$$

La matrice de relaxation est donnée par

$$\mathcal{L}_\omega = \left(\frac{1}{\omega}D - E\right)^{-1}\left(\frac{1-\omega}{\omega}D + F\right)$$

La méthode de Gauss-Seidel correspond au cas $\omega = 1$.

La méthode de relaxation s'écrit ligne par ligne comme une extrapolation de la composante obtenue par Gauss-Seidel. On a

$$x_i^{(k+1)}(\text{relax}) = \omega x_i^{(k+1)}(\text{Gauss-Seidel}) + (1-\omega)x_i^{(k)}(\text{relax})$$

Proposition 3 *Lorsque* $\rho(\mathcal{L}_\omega) \geq |1-\omega|$ *la méthode de relaxation ne peut converger que si* $0 < \omega < 2$.

Démonstration . Voir [4]

Théorème 18 *On suppose que* A *est symétrique définie positive. Alors la méthode de relaxation converge pour tout* $\omega \in]0, 2[$ *(en particulier la méthode de Gauss-Seidel converge).*

Démonstration . Voir [4]

Remarque 12 *On peut montrer que si $D^{-1}E$ et $D^{-1}F$ sont à éléments positives ou nuls, alors, dès que la méthode de Jacobi converge, il en est de même de celle de Gauss-Seidel. De plus, la convergence de celle ci est au moins aussi rapide. Ce phénomène n'est pas systématique comme on peut le voir dans l'exercice 36.*

Cependant, pour de nombreux systèmes, surtout lorsqu'ils proviennent de la discrétisation d'équations aux dérivées partielles, la méthode de Gauss-Seidel converge plus vite que celle de Jacobi. De plus pour une large famille de tels systèmes, on peut montrer qu'il existe un paramètre ω^* optimal pour lequel la méthode de relaxation est considérablement plus efficace. Le nombre d'itérations nécessaires pour atteindre une précision donnée chute considérablement lorsque ω est voisin de ω^*.

On va énoncer sans démonstration un résultat dans ce sens. Pour cela, on a besoin de la définition suivante :
A étant décomposée comme précédemment, on note

$$L = D^{-1}E, \quad U = D^{-1}F$$

d'où

$$A = D(I - L - U), \quad J = L + U$$

$$\mathcal{L}_\omega = (I - \omega L)^{-1}\left((1-\omega)I + \omega U\right)$$

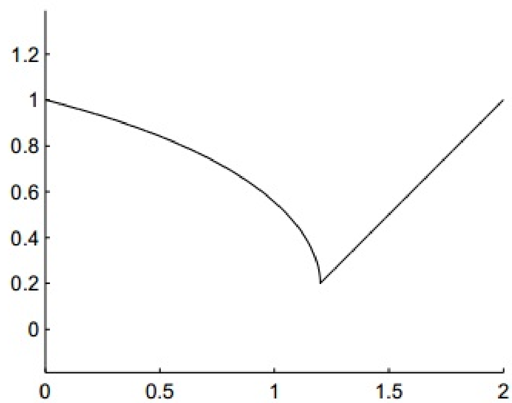
Définition 6 *On dira que A est de type (V), si $\alpha \neq 0$, les valeurs propres de $J(\alpha) = \alpha L + \dfrac{1}{\alpha} U$ sont indépendantes de α.*

Remarque 13 *On montre que les matrices tridiagonales sont de type (V). De nombreuses matrices provenant de la discrétisation d'équations aux dérivées partielles sont également de type (V). Cette notion a été introduite par Varga qui utilise la terminologie* ***"consistently ordered matrices"****.*

On a le résultat général suivant :

Théorème 19 *Soit A de type (V). Alors*

1. $\rho(G) = \rho(J)^2$
   Donc la méthode de Gauss-Seidel converge si et seulement si celle de Jacobi converge et elle converge alors deux fois plus vite.
2. *Si, de plus, les valeurs propres de J sont réelles et $\rho(J) < 1$, alors le graphe de $\rho(\mathcal{L}_\omega)$ a l'allure suivante :*



Plus précisément, on a

$$\omega^* = \frac{2}{1+\sqrt{1-\rho(J)^2}}, \qquad \rho(\ \mathcal{L}^*_\omega) = \left(\frac{\rho(J)}{1+\sqrt{1-\rho(J)^2}}\right)^2$$

et

$$\begin{cases} \omega - 1 & si\ \omega^* \leq \omega \leq 2 \\ 1-\omega+\frac{1}{2}\omega^2\rho(J)^2+\omega\rho(J)\sqrt{1-\omega+\frac{1}{4}\omega^2\rho(J)^2} & si\ 0 \leq \omega \leq \omega^* \end{cases}$$

Remarque 14 *Le plus souvent, ω^* est déterminé expérimentalement. Noter à ce propos que la dérivée à gauche de $\rho(\ \mathcal{L}_\omega)$ en $\omega = \omega^*$ est ∞. On aura donc plutôt intérêt à surévaluer ω^* car une erreur à droite sur ω^* sera moins sensible qu'une erreur à gauche.*

Nombre d'opérations

Il faut noter que chaque itération nécessite (Dans Gauss-Seidel par exemple) : $N(N-1)$ multiplications$+(N-1)$ additions$+1$ division $\simeq 2N^2$ opérations, si la matrice est pleine (matrice contenant moins de zéros).

Pour être comparable à la méthode de Gauss, il faut donc que la précision voulue soit obtenue en moins de $\frac{N}{3}$ itérations. Dans la pratique, on constate que les méthodes itératives sont surtout avantageuses lorsque la matrice est creuse (une matrice contenant beaucoup de zéros) : à chaque itération, le nombre d'opérations est d'ordre kN où k est une constante fixe directement proportionnelle au nombre d'éléments non nuls. De plus, il suffit de stocker les éléments non nuls de A. C'est le dernier point qui est essentiel dans le choix des méthodes itératives pour les grands systèmes creux : Une factorisation de type Gauss ou Cholesky, même si elle préserve la structure-bande, remplit "les trous" à l'intérieur de la bande et nécessite donc plus d'espace mémoire.

Pour N grand, on a le tableau suivant :

Méthodes	Nombre d'itérations	Nombre d'opérations
Jacobi	$\frac{4}{\pi^2}N^2\log N$	$k\frac{4}{\pi^2}N^3\log N$
Gauss Seidel	$\frac{2}{\pi^2}N^2\log N$	$k\frac{2}{\pi^2}N^3\log N$
Relaxation	$\frac{1}{\pi}N\log N$	$k\frac{1}{\pi}N^2\log N$

2.2.8 Exercices corrigés

Exercice 34 *On considère la matrice suivante*

$$\begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix}$$

1. *Etudier la convergence des méthodes de Jacobi et Gauss-Seidel pour cette matrice.*
2. *Vérifier que $\rho(G) = \rho(J)^2$, où G et J désignent respectivement les matrices d'itération des méthodes de Gauss-Seidel et Jacobi.*

Solution

1. Soit $A = D - E - F$ avec

$$D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}, E = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \text{ et } F = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Pour la méthode de Jacobi :
Soit $J = D^{-1}(E + F)$ la matrice de Jacobi

$$D^{-1} = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix} \text{ et } (E + F) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Alors

$$J = \begin{pmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{pmatrix}$$

Pour étudier la convergence de la méthode de Jacobi, on va calculer le rayon spectral de J.

$$det(J - \lambda I) = \det \begin{pmatrix} -\lambda & \frac{1}{4} & 0 \\ \frac{1}{4} & -\lambda & \frac{1}{4} \\ 0 & \frac{1}{4} & -\lambda \end{pmatrix}$$

$$= \lambda \left(-\lambda^2 + \frac{1}{8} \right)$$

Les valeurs propres sont : $\lambda_1 = 0$, $\lambda_2 = \dfrac{1}{\sqrt{8}}$ et $\lambda_3 = -\dfrac{1}{\sqrt{8}}$. Alors

$$\rho(J) = \max \left\{ 0, \frac{1}{\sqrt{8}} \right\} = \frac{1}{\sqrt{8}} < 1$$

Donc la méthode de Jacobi converge.

Pour la méthode de Gauss-Seidel :
Soit $G = (D - E)^{-1}$

$$(D - E) = \begin{pmatrix} 4 & 0 & 0 \\ -1 & 4 & 0 \\ 0 & -1 & 4 \end{pmatrix}$$

$$\det (D - E) = 64 \neq 0$$

$$(D - E)^{-1} = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ \frac{1}{16} & \frac{1}{4} & 0 \\ \frac{1}{64} & \frac{1}{16} & \frac{1}{4} \end{pmatrix}$$

Alors

$$G = \begin{pmatrix} 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{16} & \frac{1}{4} \\ 0 & \frac{1}{64} & \frac{1}{16} \end{pmatrix}$$

Pour étudier la convergence de la méthode de Gauss-Seidel, on va calculer le rayon spectral de G.

$$\det(G - \lambda I) = \begin{pmatrix} -\lambda & \frac{1}{4} & 0 \\ 0 & \frac{1}{16} - \lambda & \frac{1}{4} \\ 0 & \frac{1}{64} & \frac{1}{16} - \lambda \end{pmatrix}$$

$$= \lambda^2 \left(-\lambda + \frac{1}{8} \right)$$

Les valeurs propres sont : $\lambda_1 = \lambda_2 = 0$ et $\lambda_3 = \dfrac{1}{8}$. Alors

$$\rho(G) = \max \left\{ 0, \frac{1}{8} \right\} = \frac{1}{8} < 1$$

Donc la méthode de Gauss-Seidel converge.

2. On a
$$\rho(G) = \frac{1}{8} = \left(\frac{1}{\sqrt{8}}\right)^2 = \rho(J)^2$$
Alors la méthode de Gauss-Seidel converge plus rapidement que celle de Jacobi.

Exercice 35 *Soit α et β deux réels. On considère les matrices*
$$A_\alpha = \begin{pmatrix} 2 & \alpha & 0 \\ \alpha & 2 & \alpha \\ 0 & \alpha & 2 \end{pmatrix}, C_\beta = \begin{pmatrix} 1 & \beta & \beta \\ \beta & 1 & \beta \\ \beta & \beta & 1 \end{pmatrix}$$

1. *Pour quelles valeurs de α (resp. β) la matrice A_α (resp. C_β) est elle définie positive ?*
2. *Écrire la matrice d'itération de la méthode de Jacobi associée à A_α (resp. C_β). Pour quelles valeurs de α (resp. β) cette méthode converge- t-elle ?*
3. *Écrire la matrice d'itération de la méthode de Gauss-Seidel associée à A_α et calculer son rayon spectral. Pour quelles valeurs de α a-t-on la convergence de cette méthode ?*

Solution

1. Pour A_α :
$$\Delta_1 = 2 > 0$$
$$\Delta_2 = \det\begin{pmatrix} 2 & \alpha \\ \alpha & 2 \end{pmatrix} = 4 - \alpha^2 > 0 \text{ ssi } -2 < \alpha < 2$$
$$\Delta_3 = \det\begin{pmatrix} 2 & \alpha & 0 \\ \alpha & 2 & \alpha \\ 0 & \alpha & 2 \end{pmatrix} = 8 - 4\alpha^2 > 0 \text{ ssi } -\sqrt{2} < \alpha < \sqrt{2}$$
Donc A_α est définie positive si et seulement si $\alpha \in \,]-\sqrt{2}, \sqrt{2}\,[$.
Pour C_β :
$$\Delta_1 = 1 > 0$$
$$\Delta_2 = \det\begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix} = 1 - \beta^2 > 0 \text{ ssi } -1 < \beta < 1$$
$$\Delta_3 = \det\begin{pmatrix} 1 & \beta & 0 \\ \beta & 1 & \beta \\ 0 & \beta & 1 \end{pmatrix} = \left(\beta + \frac{1}{2}\right)(\beta - 1)^2 > 0 \text{ ssi } \beta > -\frac{1}{2}$$
Donc la matrice C_β est définie positive si et seulement si $\beta \in \,]-\frac{1}{2}, 1\,[$.
2. Soit
$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, E = \begin{pmatrix} 0 & 0 & 0 \\ -\alpha & 0 & 0 \\ 0 & -\alpha & 0 \end{pmatrix} \text{ et } F = \begin{pmatrix} 0 & -\alpha & 0 \\ 0 & 0 & -\alpha \\ 0 & 0 & 0 \end{pmatrix}$$

Pour A_α :

$$J = D^{-1}(E+F) = \begin{pmatrix} 0 & -\frac{\alpha}{2} & 0 \\ -\frac{\alpha}{2} & 0 & -\frac{\alpha}{2} \\ 0 & -\frac{\alpha}{2} & 0 \end{pmatrix}$$

Pour étudier la convergence de la méthode de Jacobi, on va calculer le rayon spectral de J.

$$\det(J-\lambda I) = \begin{pmatrix} -\lambda & -\frac{\alpha}{2} & 0 \\ -\frac{\alpha}{2} & -\lambda & -\frac{\alpha}{2} \\ 0 & -\frac{\alpha}{2} & -\lambda \end{pmatrix} = -\lambda\left(\lambda^2 - \frac{\alpha^2}{2}\right)$$

Les valeurs propres sont : $\lambda_1 = 0$, $\lambda_2 = \frac{\alpha}{\sqrt{2}}$ et $\lambda_3 = -\frac{\alpha}{\sqrt{2}}$.

On a

$$\rho(J) = \max\left\{0, \frac{|\alpha|}{\sqrt{2}}\right\} = \frac{|\alpha|}{\sqrt{2}}$$

Alors

$$\rho(J) < 1 \Longleftrightarrow \frac{|\alpha|}{\sqrt{2}} < 1 \Longleftrightarrow -\sqrt{2} < \alpha < \sqrt{2}$$

Alors la méthode de Jacobi converge pour $\alpha \in \left]-\sqrt{2}, \sqrt{2}\right[$.

Pour C_β :

$$J = D^{-1}(E+F) = \begin{pmatrix} 0 & -\beta & -\beta \\ -\beta & 0 & -\beta \\ -\beta & -\beta & 0 \end{pmatrix}$$

Pour étudier la convergence de la méthode de Jacobi, on va calculer le rayon spectral de J.

$$\det(J-\lambda I) = \begin{pmatrix} -\lambda & -\beta & -\beta \\ -\beta & -\lambda & -\beta \\ -\beta & -\beta & -\lambda \end{pmatrix} = -\lambda\left(\lambda^2 - \beta^2\right)$$

Les valeurs propres sont : $\lambda_1 = 0$, $\lambda_2 = \beta$ et $\lambda_3 = -\beta$.

3. Soit $G = (D-E)^{-1} F$

$$D - E = \begin{pmatrix} 2 & 0 & 0 \\ \alpha & 2 & 0 \\ 0 & \alpha & 2 \end{pmatrix}$$

$$\det(D-E) = 8 \neq 0$$

$$(D-E)^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{\alpha}{4} & \frac{1}{2} & 0 \\ \frac{\alpha^2}{8} & -\frac{\alpha}{4} & \frac{1}{2} \end{pmatrix}$$

Alors

$$G = \begin{pmatrix} 0 & -\frac{\alpha}{2} & 0 \\ 0 & \frac{\alpha^2}{4} & -\frac{\alpha}{2} \\ 0 & \frac{\alpha^3}{8} & \frac{\alpha^2}{4} \end{pmatrix}$$

Calculons le rayon spectral de G pour étudier la convergence de la méthode de Gauss-Seidel.

$$\det(G - \lambda I) = \begin{vmatrix} -\lambda & -\frac{\alpha}{2} & 0 \\ 0 & \frac{\alpha^2}{4} - \lambda & -\frac{\alpha}{2} \\ 0 & \frac{\alpha^3}{8} & \frac{\alpha^2}{4} - \lambda \end{vmatrix} = \lambda^2 \left(\frac{\alpha^2}{2} - \lambda \right)$$

Les valeurs propres sont : $\lambda_1 = \lambda_2 = 0$ et $\lambda_3 = \dfrac{\alpha^2}{2}$.

Alors

$$\rho(G) = \frac{\alpha^2}{2} < 1 \Longleftrightarrow -\sqrt{2} < \alpha < \sqrt{2}$$

Donc la méthode de Gauss-Seidel converge pour $\alpha \in \left] -\sqrt{2}, \sqrt{2} \right[$

Exercice 36 *Soit*

$$A = \begin{pmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix}$$

1. *Montrer que $\rho(J) < 1 < \rho(G)$, où G et J désignent respectivement les matrices d'itérations des méthodes de Gauss-Seidel et Jacobi.*
2. *Soit maintenant*

$$A = \begin{pmatrix} 2 & -1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & 2 \end{pmatrix}$$

Montrer que $\rho(G) < 1 < \rho(J)$

Solution

1. Soit

$$A = \begin{pmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix}$$

Calculons le rayon spectral de J ensuite de G. Soient

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, E = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -2 & -2 & 0 \end{pmatrix} \text{ et } F = \begin{pmatrix} 0 & -2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

On a

$$J = D^{-1}(E+F) = \begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix}$$

$$\det(J-\lambda I) = \begin{pmatrix} -\lambda & -2 & 2 \\ -1 & -\lambda & -1 \\ -2 & -2 & -\lambda \end{pmatrix} = -\lambda^3$$

Les valeurs propres sont :$\lambda_1 = \lambda_2 = \lambda_3 = 0$.
Alors

$$\rho(J) = 0 < 1$$

Soit $G = (D-E)^{-1} F$

$$D - E = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & 2 & 1 \end{pmatrix}$$

$$\det(D-E) = 1 \neq 0$$

$$(D-E)^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}$$

Alors

$$G = \begin{pmatrix} 0 & -2 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\det(G-\lambda I) = \begin{pmatrix} -\lambda & -2 & 2 \\ 0 & 2-\lambda & -3 \\ 0 & 0 & 2-\lambda \end{pmatrix} = -\lambda\,(2-\lambda)^2$$

Les valeurs propres sont :$\lambda_1 = 0$ et $\lambda_2 = \lambda_3 = 2$.
Alors

$$\rho(G) = 2 > 1$$

Donc

$$\rho(J) < 1 < \rho(G)$$

2. Soit

$$A = \begin{pmatrix} 2 & -1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & 2 \end{pmatrix}$$

Calculons le rayon spectral de J ensuite de G. Soient

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, E = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \text{ et } F = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & -2 \\ 0 & 0 & 0 \end{pmatrix}$$

On a

$$J = D^{-1}(E+F) = \begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & -1 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$$

$$\det(J-\lambda I)=\begin{pmatrix} -\lambda & \frac{1}{2} & -\frac{1}{2} \\ -1 & -\lambda & -1 \\ 0 & \frac{1}{2} & -\lambda \end{pmatrix}=-\lambda^3-\lambda+\frac{1}{4}$$

Les valeurs propres sont : $\lambda_1 = 0.236733$, $\lambda_2 = -0.118366 + \mathbf{i}\ 1.0208$ et $\lambda_3 = \overline{\lambda_2}$.
Alors

$$\rho(J) = \max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} = |\lambda_2| = |\lambda_3| = 1.02764 > 1$$

Soit $G = (D-E)^{-1} F$

$$D-E=\begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ -1 & -1 & 2 \end{pmatrix}$$

$$\det(D-E) = 8 \neq 0$$

$$(D-E)^{-1}=\begin{pmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

Alors

$$G=\begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} \end{pmatrix}$$

$$\det(G-\lambda I)=\begin{pmatrix} -\lambda & \frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{1}{2}-\lambda & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2}-\lambda \end{pmatrix}=-\lambda\left(\lambda+\frac{1}{2}\right)^2$$

Les valeurs propres sont : $\lambda_1 = 0$ et $\lambda_2 = \lambda_3 = -\frac{1}{2}$.

Alors

$$\rho(G) = \frac{1}{2} < 1$$

Donc

$$\rho(G) < 1 < \rho(J)$$

Exercice 37 *Considérons le système linéaire $Ax = b$ avec*

$$A = \begin{pmatrix} \alpha & 0 & \gamma \\ 0 & \alpha & \beta \\ 0 & \delta & \alpha \end{pmatrix}$$

avec α, β, γ et δ des paramètres réels.
Donner des conditions suffisantes sur ces coefficients pour avoir la convergence de la méthode de Jacobi (resp. Gauss-Seidel).

Solution

1. La méthode de Jacobi :
   Une condition suffisante pour que la méthode de Jacobi converge est que la matrice A soit à diagonale dominante stricte ce qui équivaut à imposer

$$\begin{cases} |\alpha| > |\gamma| \\ |\alpha| > |\beta| \\ |\alpha| > |\delta| \end{cases}$$

   Cela veut dire que

$$|\alpha| > \max\{|\beta|, |\gamma|, |\delta|\}$$

2. La méthode de Gauss-Seidel :
   La condition précédente est aussi suffisante pour la convergence de Gauss-Seidel.
   Une autre condition pour la convergence de cette méthode est que la matrice A soit symétrique définie positive.
   La symétrie :

$$\begin{cases} \gamma = 0 \\ \beta = \delta \end{cases}$$

   et on obtient

$$A = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & \beta & \alpha \end{pmatrix}$$

   Elle est définie positive si ses valeurs propres sont positives.
   On a

$$\lambda_1 = \alpha, \lambda_2 = \alpha - \beta \text{ et } \lambda_3 = \alpha + \beta$$

   Donc il faut que $\alpha > 0$ et $\alpha > |\beta|$.
   On note dans ce cas que lorsque A est symétrique définie positive elle est aussi à diagonale dominante stricte.

Exercice 38 *Considérons le système linéaire suivant*

$$\begin{pmatrix} 6 & 1 & 1 \\ 2 & 4 & 0 \\ 1 & 2 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 12 \\ 0 \\ 6 \end{pmatrix}$$

1. *Étudier la convergence de la méthode de Jacobi puis approcher la solution avec 3 itérations à partir de $x^{(0)} = (2, 2, 2)$.*
2. *Faire la même chose avec la méthode de gauss-Seidel.*

Solution

1.

$$Ax = b \Longleftrightarrow \begin{cases} 6x_1 + x_2 + x_3 = 12 \\ 2x_1 + 4x_2 = 0 \\ x_1 + 2x_2 + 6x_3 = 6 \end{cases}$$

$$\Longleftrightarrow \begin{cases} x_1 = -\dfrac{1}{6}x_2 - \dfrac{1}{6}x_3 + 2 \\ \\ x_2 = -\dfrac{1}{2}x_1 \\ \\ x_3 = -\dfrac{1}{6}x_1 - \dfrac{1}{3}x_2 + 1 \end{cases}$$

d'où le procédé itératif de Jacobi

$$\begin{cases} x_1^{(k+1)} = -\dfrac{1}{6}x_2^{(k)} - \dfrac{1}{6}x_3^{(k)} + 2 \\ \\ x_2^{(k+1)} = -\dfrac{1}{2}x_1^{(k)} \\ \\ x_3^{(k+1)} = -\dfrac{1}{6}x_1^{(k)} - \dfrac{1}{3}x_2^{(k)} + 1 \end{cases}$$

On en déduit alors la matrice de Jacobi

$$J = \begin{pmatrix} 0 & -\dfrac{1}{6} & -\dfrac{1}{6} \\ \\ -\dfrac{1}{2} & 0 & 0 \\ \\ -\dfrac{1}{6} & -\dfrac{1}{3} & 0 \end{pmatrix}$$

$$\det(J - \lambda I) = \det \begin{pmatrix} -\lambda & -\dfrac{1}{6} & -\dfrac{1}{6} \\ \\ -\dfrac{1}{2} & -\lambda & 0 \\ \\ -\dfrac{1}{6} & -\dfrac{1}{3} & -\lambda \end{pmatrix} = -\lambda^3 + \frac{\lambda}{9} - \frac{1}{36}$$

Les valeurs propres sont : $\lambda_1 = 0.210425 - \mathbf{i}\ 0.147395$, $\lambda_2 = \overline{\lambda_1}$ et $\lambda_3 = -0.42085$. Alors

$$\rho(J) = \max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} = |\lambda_3| = 0.42085 < 1$$

Donc la méthode de Jacobi converge.

Voici maintenant le tableau qui représente la solution approchée après 3 itérations :

$x^{(0)}$	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$
2	1.333	2.1666	1.925
2	-1	-0.666	-1.083
2	0	1.111	0.8611

2. Pour la méthode de Gauss-Seidel je laisse le calcul au lecteur.

Exercice 39 *Soit le système linéaire*

$$\begin{cases} 4x_1 + 3x_2 + 3x_3 = 10 \\ 3x_1 + 4x_2 + 3x_3 = 10 \\ 3x_1 + 3x_2 + 4x_3 = 10 \end{cases} \tag{2.2.1}$$

1. *Écrire les méthodes de Jacobi et de Gauss-Seidel pour ce système.*
2. *Étudier la convergence des deux méthodes.*
3. *Calculer les* 3 *premiers itérés.*

Solution

1. Méthode de Jacobi :

$$(2.2.1) \Longleftrightarrow \begin{cases} x_1 = -\frac{3}{4}x_2 - \frac{3}{4}x_3 + \frac{10}{4} \\ \\ x_2 = -\frac{3}{4}x_1 - \frac{3}{4}x_3 + \frac{10}{4} \\ \\ x_3 = -\frac{3}{4}x_1 - \frac{3}{4}x_2 + \frac{10}{4} \end{cases}$$

D'où le système itéré

$$\begin{cases} x_1^{(k+1)} = -\frac{3}{4}x_2^{(k)} - \frac{3}{4}x_3^{(k)} + \frac{5}{2} \\ \\ x_2^{(k+1)} = -\frac{3}{4}x_1^{(k)} - \frac{3}{4}x_3^{(k)} + \frac{5}{2} \\ \\ x_3^{(k+1)} = -\frac{3}{4}x_1^{(k)} - \frac{3}{4}x_2^{(k)} + \frac{5}{2} \end{cases}$$

d'où

$$J = \begin{pmatrix} 0 & -\frac{3}{4} & -\frac{3}{4} \\ \\ -\frac{3}{4} & 0 & -\frac{3}{4} \\ \\ -\frac{3}{4} & -\frac{3}{4} & 0 \end{pmatrix}, b^* = \begin{pmatrix} \frac{5}{2} \\ \\ \frac{5}{2} \\ \\ \frac{5}{2} \end{pmatrix}$$

Méthode de Gauss-Seidel :

$$(2.2.1) \Longleftrightarrow \begin{cases} x_1 = -\dfrac{3}{4}x_2 - \dfrac{3}{4}x_3 + \dfrac{5}{2} \\[2ex] x_2 = -\dfrac{3}{4}x_1 - \dfrac{3}{4}x_3 + \dfrac{5}{2} \\[2ex] x_3 = -\dfrac{3}{4}x_1 - \dfrac{3}{4}x_2 + \dfrac{5}{2} \end{cases}$$

D'où le système itéré

$$\begin{cases} x_1^{(k+1)} = -\dfrac{3}{4}x_2^{(k)} - \dfrac{3}{4}x_3^{(k)} + \dfrac{5}{2} \\[2ex] x_2^{(k+1)} = -\dfrac{3}{4}x_1^{(k+1)} - \dfrac{3}{4}x_3^{(k)} + \dfrac{5}{2} \\[2ex] x_3^{(k+1)} = -\dfrac{3}{4}x_1^{(k+1)} - \dfrac{3}{4}x_2^{(k+1)} + \dfrac{5}{2} \end{cases} \Longleftrightarrow \begin{cases} x_1^{(k+1)} = -\dfrac{3}{4}x_2^{(k)} - \dfrac{3}{4}x_3^{(k)} + \dfrac{5}{2} \\[2ex] x_2^{(k+1)} = \dfrac{9}{16}x_1^{(k)} - \dfrac{3}{16}x_3^{(k)} + \dfrac{5}{8} \\[2ex] x_3^{(k+1)} = \dfrac{9}{64}x_1^{(k)} + \dfrac{45}{64}x_2^{(k)} + \dfrac{65}{32} \end{cases}$$

d'où

$$G = \begin{pmatrix} 0 & -\dfrac{3}{4} & -\dfrac{3}{4} \\[2ex] 0 & \dfrac{9}{16} & -\dfrac{3}{16} \\[2ex] 0 & \dfrac{9}{64} & \dfrac{45}{64} \end{pmatrix}, b^* = \begin{pmatrix} \dfrac{5}{2} \\[2ex] \dfrac{5}{8} \\[2ex] \dfrac{65}{32} \end{pmatrix}$$

2. Calculons le rayon spectral des deux méthodes

$$\det(J - \lambda I) = \det \begin{pmatrix} -\lambda & -\dfrac{3}{4} & -\dfrac{3}{4} \\[2ex] -\dfrac{3}{4} & -\lambda & -\dfrac{3}{4} \\[2ex] -\dfrac{3}{4} & -\dfrac{3}{4} & -\lambda \end{pmatrix}$$

$$= \left(-\lambda - \frac{3}{4}\right)\left(\lambda^2 - \frac{3}{4}\lambda + \frac{18}{16}\right)$$

Les valeurs propres sont : $\lambda_1 = \lambda_2 = \dfrac{3}{4}$ et $\lambda_3 = -\dfrac{3}{2}$.

Donc

$$\rho(J) = \max\left\{\frac{3}{4}, \frac{3}{2}\right\} = \frac{3}{2} > 1$$

Alors la méthode de Jacobi ne converge pas.

$$\det(G - \lambda I) = \det \begin{pmatrix} -\lambda & -\dfrac{3}{4} & -\dfrac{3}{4} \\[2ex] 0 & \dfrac{9}{16} - \lambda & -\dfrac{3}{16} \\[2ex] 0 & \dfrac{9}{64} & \dfrac{45}{64} - \lambda \end{pmatrix}$$

$$= -\lambda\left(\lambda^2 - \frac{81}{64}\lambda + \frac{27}{64}\right)$$

Les valeurs propres sont : $\lambda_1 = 0$, $\lambda_2 = 0.6328125 - \mathbf{i}\ 0.14637$ et $\lambda_3 = \overline{\lambda}_2$.

$$\rho(G) = \max\{0, |\lambda_2|, |\lambda_3|\} = |\lambda_2| = 0.649519696 < 1$$

alors la méthode de Gauss-Seidel converge.

3. Pour calculer les 3 premiers itérés on utilise la méthode de Gauss-Seidel car elle converge

$x^{(0)}$	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$
5/2	0.5078125	0.79895	0.95313
5/8	0.595703	0.64651	0.7231675
65/32	1.67236	1.41598	1.242776

Exercice 40 *Soit A une matrice carrée d'ordre 3 telle que $A = I_3 - E - F$ avec*

$$E = \begin{pmatrix} 0 & -2 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ et } F = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & 0 \end{pmatrix}$$

1. *Montrer que A est inversible.*
2. *Soit $0 < \omega < 2$. Montrer que la matrice $\frac{1}{\omega}I_3 - E$ est inversible si et seulement si $\omega \neq \frac{\sqrt{2}}{2}$.*
3. *Pour $0 < \omega < 2$, $\omega \neq \frac{\sqrt{2}}{2}$, on considère pour la résolution de $Ax = b$, la méthode itérative définie par*

$$x^{(0)} \in \mathbb{R}^\star \text{ et } \forall k \in \mathbb{N}, \left(\frac{1}{\omega}I_3 - E\right)x^{(k+1)} = \left(F + \frac{1-\omega}{\omega}I_3\right)x^{(k)} + b'$$

et on pose

$$\mathcal{L}_\omega = \left(\frac{1}{\omega}I_3 - E\right)^{-1}\left(F + \frac{1-\omega}{\omega}I_3\right)$$

Calculer en fonction de ω les valeurs propre de $\mathcal{L}_\omega$ et son rayon spectral.

4. *Pour quelles valeurs de ω cette méthode converge -t-elle ?*
5. *Déterminer $\omega^* \in]0, 2[$ vérifiant*

$$\rho(\ \mathcal{L}_{\omega^*}\ ) = \min\{\ \rho(\ \mathcal{L}_\omega\ )\ /\omega \in\ ]0, 2[\ , \omega \neq \tfrac{\sqrt{2}}{2}\}$$

Solution

Soit

$$A = I_3 - E - F = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

1.
$$\det(A) = 1 \neq 0$$
Alors A est inversible.

2. on a $0 < \omega < 2$, calculons le déterminant de $\frac{1}{\omega}I_3 - E$. Soit
$$\left(\frac{1}{\omega}I_3 - E\right) = \begin{pmatrix} \frac{1}{\omega} & 2 & 0 \\ 0 & \frac{1}{\omega} & 0 \\ 0 & 0 & \frac{1}{\omega} \end{pmatrix}$$
$$\det\left(\frac{1}{\omega}I_3 - E\right) = \frac{1}{\omega^3}\left(1 - 2\omega^2\right)$$
$$\det\left(\frac{1}{\omega}I_3 - E\right) = 0 \Longleftrightarrow \left(\frac{1}{\omega^3} = 0\right) \vee \left(1 - 2\omega^2 = 0\right)$$
Puisque $0 < \omega < 2$, on prend la solution $1 - 2\omega^2 = 0$, ce qui donne $\omega = \frac{\sqrt{2}}{2}$.
Donc pour que $\left(\frac{1}{\omega}I_3 - E\right)$ soit inversible il faut que $\omega \neq \frac{\sqrt{2}}{2}$.

3. $\mathcal{L}_\omega = \left(\frac{1}{\omega}I_3 - E\right)^{-1}\left(F + \frac{1-\omega}{\omega}I_3\right)$
$$\left(F + \frac{1-\omega}{\omega}I_3\right) = \begin{pmatrix} \frac{1-\omega}{\omega} & 0 & 0 \\ 0 & \frac{1-\omega}{\omega} & 0 \\ -1 & -1 & \frac{1-\omega}{\omega} \end{pmatrix}$$
et
$$\left(\frac{1}{\omega}I_3 - E\right)^{-1} = \begin{pmatrix} \frac{\omega}{1-2\omega^2} & -\frac{2\omega^2}{1-2\omega^2} & 0 \\ -\frac{\omega^2}{1-2\omega^2} & \frac{\omega}{1-2\omega^2} & 0 \\ 0 & 0 & \omega \end{pmatrix}$$
d'où
$$\mathcal{L}_\omega = \begin{pmatrix} \frac{1-\omega}{1-2\omega^2} & -\frac{2\omega(1-\omega)}{1-2\omega^2} & 0 \\ -\frac{\omega(1-\omega)}{1-2\omega^2} & \frac{1-\omega}{1-2\omega^2} & 0 \\ -\omega & -\omega & 1-\omega \end{pmatrix}$$
$$\det(\ \mathcal{L}_\omega\ - I\lambda) = \begin{pmatrix} \frac{1-\omega}{1-2\omega^2} - \lambda & -\frac{2\omega(1-\omega)}{1-2\omega^2} & 0 \\ -\frac{\omega(1-\omega)}{1-2\omega^2} & \frac{1-\omega}{1-2\omega^2} - \lambda & 0 \\ -\omega & -\omega & 1-\omega-\lambda \end{pmatrix}$$

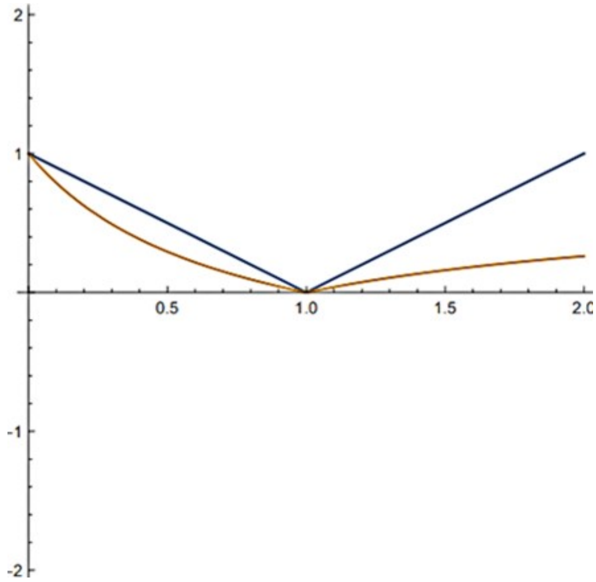
$$= (1-\omega-\lambda)\left(-\lambda+\frac{1-\omega}{1+\sqrt{2}\omega}\right)\left(-\lambda-\frac{1-\omega}{1+\sqrt{2}\omega}\right)$$

Les valeurs propres sont : $\lambda_1 = 1-\omega$, $\lambda_2 = -\frac{1-\omega}{1+\sqrt{2}\omega}$ et $\lambda_3 = \frac{1-\omega}{1+\sqrt{2}\omega}$.

Donc

$$\rho(\ \mathcal{L}_\omega\ ) = \max\left\{|1-\omega|\,, \frac{|1-\omega|}{|1+\sqrt{2}\omega|}\right\}$$

Soit le graphe de $|1-\omega|$ et $\frac{|1-\omega|}{|1+\sqrt{2}\omega|}$



D'après ce graphe on remarque que

$$\rho(\ \mathcal{L}_\omega\ ) = |1-\omega|\,,\ \text{pour } 0<\omega<2$$

4. On a $\rho(\ \mathcal{L}_\omega\ ) < 1$, alors la méthode converge pour $0<\omega<2$.
5. D'après le graphe on remarque aussi que

$$\omega^* = 1 \text{ tel que } \rho(\ \mathcal{L}_{\omega^*}\ ) = \min\left\{\ \rho(\ \mathcal{L}_\omega\ )\ /\omega \in\ ]0,2[\ ,\omega \neq \tfrac{\sqrt{2}}{2}\right\}$$

2.2.9 Exercices supplémentaires

Exercice 41 *Soit $a \in \mathbb{R}$ et $A = \begin{pmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{pmatrix}$*

1. *Pour quelles valeurs de a, A est-elle définie positive ?*
2. *Pour quelles valeurs de a, La méthode de Gauss-Seidel converge-t-elle ?*
3. *Écrire la matrice J de l'itération de Jacobi.*
4. *Pour quelles valeurs de a, la méthode de Jacobi converge-t-elle ?*
5. *Écrire la matrice G de l'itération de Gauss-Seidel.*
6. *Pour quelles valeurs de a, La méthode de Gauss-Seidel converge-t-elle plus vite que celle de Jacobi ?*

Exercice 42 *On considère le système linéaire $Ax = b$, où*

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \text{ et } b = \begin{pmatrix} 3 \\ -5 \\ 5 \end{pmatrix}$$

1. *Écrire la matrice de Jacobi et celle de Gauss-Seidel.*
2. *Étudier la convergence de la méthode de Jacobi et celle de Gauss-Seidel.*
3. *On note par $err^{(k)} = \left\| x^{(k)} - x_* \right\|$ l'erreur commise à l'étape k. Estimer le nombre minimal d'itérations nécessaires pour que l'erreur à l'étape k soit telle que*
$$err^{(k)} \leq 10^{-4} err^{(0)}$$
4. *En utilisant la méthode adéquate, estimer la solution à 10^{-4} près.*

Chapitre 3

Annexe

3.1 Théorème des valeurs intermédiaires

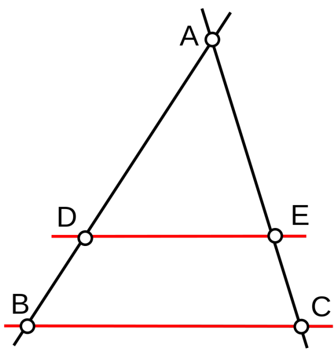
Théorème 20 *Si la fonction f est continue dans $[a, b]$ et $f(a) \times f(b) < 0$ alors $\exists \alpha \in \left]a, b\right[$ tel que $f(\alpha) = 0$. Si de plus f est strictement monotone dans $[a, b]$ alors α est unique dans $[a, b]$.*

3.2 Théorème de Pythagore

Théorème 21 *Si un triangle ABC est rectangle en C, alors $AB^2 = AC^2 + BC^2$.*

3.3 Théorème de Thalès

Théorème 22 *Soit un triangle ABC, et deux points D et E des droites (AB) et (AC) de sorte que la droite (DE) soit parallèle à la droite (BC) comme le montre la figure suivante :*



Alors on a :

$$\frac{AD}{AB} = \frac{AE}{AC} = \frac{DE}{BC}$$

3.4 Éléments matricielles

3.4.1 Inverse des matrices

Définition 7 *Notons $\mathcal{M}_n(\mathbb{R})$ l'ensemble des matrices carrées d'ordre n à coefficients réels. La matrice $A \in \mathcal{M}_n(\mathbb{R})$ est dite inversible s'il existe une matrice $\mathcal{M}_n(\mathbb{R})$ notée A^{-1} telle que*

$$AA^{-1} = A^{-1}A = I_n$$

3.4.2 Matrices triangulaires

Définition 8 *Une matrice $A = (a_{ij})_{1\leq i,j\leq n} \in \mathcal{M}_n(\mathbb{R})$ est dite triangulaire inférieure (resp.triangulaire supérieure), si $\forall i < j$ (resp. $i > j$), $i, j \in \{1, 2, \cdots, n\}$: $a_{ij} = 0$.*

Remarque 15 *Une matrice qui est à la fois triangulaire inférieure et triangulaire supérieure est dite diagonale.*

Proposition 4 *Soit $A = (a_{ij})_{1\leq i,j\leq n} \in \mathcal{M}_n(\mathbb{R})$.*

1. *Si A est triangulaire inférieure (resp. supérieure) et inversible, alors A^{-1} est triangulaire inférieure (resp. supérieure).*
2. *Si A est triangulaire inférieure (resp. supérieure), alors*

$$A \text{ inversible } \iff a_{ii} \neq 0, \forall i, j \in \{1, 2, \cdots, n\}$$

3.4.3 Transposée d'une matrice

Définition 9 *Soit $A \in \mathcal{M}_{n,m(\mathbb{R})}$. On définit la matrice transposée de A et on note A^T la matrice de $\mathcal{M}_{n,m}(\mathbb{R})$ par*

$$A^T = (a_{ij}) \in \mathcal{M}_{m,n}(\mathbb{R})$$

Proposition 5 *Soit $A \in \mathcal{M}_{n,m}(\mathbb{R})$ et $B \in \mathcal{M}_{m,p}(\mathbb{R})$, on a*

$$(AB)^T = B^T A^T$$

Proposition 6 *Soit $A \in \mathcal{M}_n(\mathbb{R})$ inversible, alors A^T est inversible et on a :*

$$(A^T)^{-1} = (A^{-1})^T$$

qui sera noté A^{-T}.

3.4.4 Matrices symétriques

Définition 10 *Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est symétrique si $A^T = A$.*

Proposition 7 *Notons $S_n(\mathbb{R})$ l'ensemble des matrices carrées symétrique d'ordre n.*

Si $A \in \mathcal{M}_n(\mathbb{R})$ inversible et B une matrice quelconque, alors on a : $A + A^T \in S_n(\mathbb{R})$ et les matrices BB^T et B^TB sont symétriques.

3.4.5 Déterminant d'une matrice

On note Ω_n l'ensemble des permutations d'ordre n.

Définition 11 *Soit $A \in \mathcal{M}_n(\mathbb{R})$. On définit le déterminant de A comme suit :*

$$det(A) = \Sigma\epsilon(\sigma) \prod_{\sigma\in\Omega_n} a_{i,\sigma(i)}$$

où $\epsilon(\sigma)$ est la signature de σ.

Proposition 8 *Si $A, B \in \mathcal{M}_n(\mathbb{R})$, on a*

$$det(AB) = det(A) \times det(B) = det(BA)$$

et

$$det(A^T) = det(A)$$

Le déterminant d'une matrice triangulaire ou diagonale est égal au produit des éléments diagonaux. En particulier $det(I_n) = 1$.

3.4.6 Matrices définies positives

Définition 12 *Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ symétrique est dite définie positive si :*

$$\forall x \in \mathbb{R}^n, x^T Ax \geq 0$$

$$\forall x \in \mathbb{R}^n, x^T Ax = 0 \Longleftrightarrow x = 0$$

Proposition 9 *Si A est une matrice carrée réelle inversible, alors $A^T A$ est une matrice symétrique définie positive.*

Proposition 10 *Si $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice définie positive, alors $det(A) > 0$ et $\forall i \in \{1, 2, \cdots, n\}$, $a_{ii} > 0$.*

Remarque 16 *Une matrice définie positive est inversible, mais la réciproque n'est pas vrai.*

Définition 13 *(Mineurs principaux) Soit $A = (a_{ij})_{1\leq i,j\leq n} \in \mathcal{M}_n(\mathbb{R})$ une matrice. On appelle mineurs principaux de A les éléments des matrices*

$$A_k = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}$$

pour k entier compris entre 1 et n.

Théorème 23 *(Critère de sylvester)*
Soit $A \in \mathcal{M}_n(\mathbb{R})$, il y a équivalence entre :

1. *A définie positive.*
2. *A symétrique et tous les mineurs principaux sont strictement positifs.*

Définition 14 *Une matrice symétrique A est définie positive si toutes ses valeurs propres sont strictement positives.*

3.4.7 Normes matricielles

Définition 15 *Étant donné V un espace vectoriel sur $\mathbb{R}$ ou $\mathbb{C}$, on appelle norme sur V toute application (qu'on notera $\|.\|$) de V dans $[0, +\infty[$ telle que :*

1. $\|x\| = 0 \Longleftrightarrow x = 0$ *pour* $x \in V$.
2. $\forall x \in V$, $\forall \lambda \in \mathbb{R}$ *(ou* $\mathbb{C}$*)*, $\|\lambda x\| = |\lambda| \, \|x\|$ *(homogénéité).*
3. $\forall x, y \in V, \|x + y\| \leq \|x\| + \|y\|$ *(inégalité triangulaire).*

En particulier, on a les normes classiques de $\mathbb{R}^n$ (ou $\mathbb{C}^n$).

$$\|x\|_1 = \sum_{i=1}^{N} |x_i| \, , \ \|x\|_2 = \left(\sum_{i=1}^{N} x_i^2 \right)^{\frac{1}{2}} \, , \ \|x\|_\infty = \max_{1 \leq i \leq N} |x_i|$$

Plus généralement, on peut considérer

$$\|x\|_p = \left(\sum_{i=1}^{N} x_i^p \right)^{\frac{1}{p}} \qquad , 1 \leq p \leq \infty$$

La notation $\|x\|_\infty$ utilisée ci-dessus provient du fait que

$$\forall x \in \mathbb{R}^n, \lim_{p \to \infty} \|x\|_p = \max_{1 \leq i \leq N} |x_i| \, (= \|x\|_\infty)$$

Norme sur l'espace vectoriel des matrices

Une matrice carrée d'ordre N est la donnée de N^2 réels (ou complexes) : elle peut donc être identifiée à un élément de $\mathbb{R}^{N^2}$ fournit une norme sur l'espace des matrices. Ainsi, si $A = [a_{ij}]$, on peut poser

$$\|A\| = \max_{1 \leq i,j \leq N} |a_{ij}| \ \text{ ou } \ \|A\| = \left(\sum_{i,j=1}^{N} a_{ij}^2 \right)^{\frac{1}{2}}$$

Cependant, pour qu'une norme soit pratiquement utiliser, il faut qu'elle soit compatible avec le produit des matrices, ce qui nous conduit à exiger cette propriété :

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

Définition 16 *On appelle norme matricielle toute norme sur l'espace vectoriel des matrices vérifiant ci-dessus.*

Étant donné une norme vectorielle $\|.\|$ définie sur $\mathbb{R}^n$, il existe une norme matricielle naturellement associée à $\|.\|$: on l'appelle norme matricielle induite par $\|.\|$ et elle est définie par :

$$\|A\| = \max_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|} \left(= \max_{\|x\| = 1} \|Ax\| \right)$$

Proposition 11 *1.* $\|A\|_\infty = \max_{1 \leq i \leq N} |a_{ij}|$

2. $\|A\|_1 = \max_{1 \leq j \leq N} |a_{ij}|$

3. $\|A\|_2 = \sqrt{\rho(A^*A)}$ *où* $A^* = \bar{A}^t$ *(adjointe de* A*)*
et $\rho(M)$ *désigne le rayon spectral de la matrice* M *définit par*

$$\rho(M) = \max_{1 \leq i \leq N} \{|\lambda_i| \,;\, \lambda_i \text{ valeurs propres de } M\}$$

Dans le cas particulier où A *est symétrique,* $\|A\|_2 = \rho(A)$.

Remarque 17 — *Toute racine carrée d'une valeur propre de* A^*A *est appelée valeur singulière de* A*. Si* A *est normale (c'est-à-dire* $A^*A = AA^*$*), alors* $\sqrt{\rho(A^*A)} = \rho(A)$*. Dans tout les cas* $\rho(A^*A) = \rho(AA^*)$

— $\|A\|_\infty$ *et* $\|A\|_1$ *se calculent simplement à partir de la matrice* A *contrairement à* $\|A\|_2$
— *On utilise aussi la norme, dite de Fröbenius*

$$\|A\|_F = \left(\sum_{i,j=1}^{N} a_{ij}^2 \right)^{\frac{1}{2}} = (Trace(A^*A))^{\frac{1}{2}}$$

Il s'agit d'une norme matricielle, mais non induite par une norme vectorielle.

Définition 17 *Une matrice* $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{C})$ *est dite " à diagonale dominante " si pour chaque* $i \in [1 : n]$*, on a*

$$\sum_{j=1, j \neq i}^{N} |a_{ij}| \leq |a_{ii}|$$

Elle sera dite " à diagonale dominante stricte " si toutes ces égalités sont strictes.

Corollaire 2 *Une matrice à diagonale dominante stricte est inversible.*

3.4.8 Conditionnement d'une matrice

Quand on veut résoudre numériquement $Ax = b$ avec $A \in \mathcal{M}_n(\mathbb{C})$, $n >> 1$, divers facteurs influent sur la précision du résultat :

— Incertitude sur les données (expérimentales) de A et/ou de b.
— Erreur dans la représentation des coefficients de A et b avec un nombre fini décimales (ordinateur).
— Erreurs d'arrondi lors des calculs.
— Erreurs prévisibles de l'algorithme, notamment lors des calculs approchés par des méthodes itératives.

Considérons le système linéaire

$$Ax = b \tag{3.4.1}$$

Considérons une petite perturbation de b notée δb. La solution correspondante est perturbée soit

$$A(x + \delta x) = b + \delta b \tag{3.4.2}$$

On va mesurer la variation relative de x en fonction de celle de b. Pour cela, on prend $\|.\|$ une norme quelconque sur $\mathbb{R}^n$.
de (3.4.1) et (3.4.2), on tire

$$A(\delta x) = \delta b \Leftrightarrow \delta x = A^{-1}(\delta b)$$

Utilisant la norme matricielle induite, on a

$$\|\delta x\| \leq \left\|A^{-1}\right\| \|\delta b\|$$

par ailleurs, avec (3.4.1)

$$\|b\| \leq \|A\| \|x\|$$

ces deux inégalités donnent

$$\frac{\|\delta x\|}{\|x\|} \leq \left\|A^{-1}\right\| \|A\| \frac{\|\delta b\|}{\|b\|} \tag{3.4.3}$$

cette estimation sur la variation de x est en fait optimale puisqu'on peut trouver x_0 tel que $\|Ax_0\| = \|A\| \|x_0\|$ et δb tel que $\|A^{-1}\| \|\delta b\| = \|A^{-1}(\delta b)\|$.

Ainsi la variation relative en x sera d'autant plus grande que le nombre $\|A^{-1}\| \|A\|$ est plus grand : on l'appelle conditionnement de A relatif à la norme $\|.\|$.

Définition 18 *On appelle conditionnement de la matrice A (dans la norme matricielle $\|.\|$), le nombre*

$$cond(A) = \left\|A^{-1}\right\| \|A\|$$

Si on travail avec $\|.\|_p$, on notera le conditionnement associé $cond_p$.

C'est le même nombre qui intervient lors de la variation des coefficients de A : Supposons, en effet que A soit perturbée par une matrice ΔA, alors

$$Ax = b$$

$$\begin{aligned}
(A + \Delta A)(x + \delta x) &= b \\
\Longrightarrow A(\delta x) + \Delta A(x + \delta x) &= 0 \\
\Longrightarrow \delta x &= -A^{-1}\Delta A(x + \delta x) \\
\Longrightarrow \|\delta x\| &\leq \left\|A^{-1}\right\| \|\Delta A\| \|x + \delta x\| \\
\Longrightarrow \frac{\|\delta x\|}{\|x + \delta x\|} &\leq cond(A)\frac{\|\Delta A\|}{\|A\|}
\end{aligned}$$

Remarque 18 *On ne modifie pas le conditionnement d'une matrice en la multipliant par un scalaire, par contre, on peut diminuer $cond_2 A$ en multipliant certaines lignes ou certaines colonnes par les coefficients non nuls : il s'agit de l'équilibrage de la matrice. C'est une technique de pré-conditionnement très utilisée en pratique.*

Exemple de système linéaire mal conditionné

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 1 \end{pmatrix}, b = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}$$

et la solution

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Le système perturbé

$$\begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 1 \end{pmatrix} \begin{pmatrix} x_1 + \delta x_1 \\ x_2 + \delta x_2 \\ x_3 + \delta x_3 \\ x_4 + \delta x_4 \end{pmatrix} = \begin{pmatrix} 32.1 \\ 22.9 \\ 33.1 \\ 30.9 \end{pmatrix}$$

a pour solution

$$\begin{pmatrix} 9.2 \\ -12.6 \\ 4.5 \\ -1.1 \end{pmatrix}$$

Ainsi une erreur relative de $\frac{1}{200}$ sur les données entraîne une erreur relative de l'ordre de $\frac{10}{1}$ du résultat (erreur amplifiée de 2000).

De même

$$\begin{pmatrix} 10 & 7 & 8.1 & 7.2 \\ 7.08 & 5.04 & 6 & 5 \\ 8 & 5.98 & 9.89 & 9 \\ 6.99 & 4.99 & 9 & 9.98 \end{pmatrix} \begin{pmatrix} x_1 + \Delta x_1 \\ x_2 + \Delta x_2 \\ x_3 + \Delta x_3 \\ x_4 + \Delta x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}$$

a pour solution

$$\begin{pmatrix} -81 \\ 137 \\ -34 \\ 22 \end{pmatrix}$$

pourtant, la matrice est " bonne " (symétrique, de déterminant 1, donc loin de 0). Son inverse est donné par

$$A^{-1} = \begin{pmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{pmatrix}$$

Mais les valeurs propres de A sont

$$\lambda_1 \simeq 0.01015 < \lambda_2 \simeq 0.8431 < \lambda_3 \simeq 3.858 < \lambda_4 \simeq 30.2877$$

si bien que

$$cond_2(A) = \frac{\lambda_4}{\lambda_1} \simeq 2984$$

est grand, d'autre part

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \delta x = \begin{pmatrix} 8.2 \\ -13.6 \\ 3.5 \\ -2.1 \end{pmatrix}, b = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}, \delta b = \begin{pmatrix} 0.1 \\ -0.1 \\ 0.1 \\ -0.1 \end{pmatrix}$$

de sorte que

$$\frac{\|\delta x\|_2}{\|x\|_2} \leq cond_2(A) \frac{\|\delta b\|_2}{\|b\|_2}$$

Bibliographie

[1] Gloria Faccanoni, *Analyse numérique Recueil d'exercices corrigés et aide-mémoire.*, 2013-2014.

[2] Chokri, Bekkey, Zouhaier et Helali, *http ://wims.auto.u-psud.fr/wims/modules/U3/analysis/doczero.fr/doc/files/doczero.pdf.*

[3] Hassen Douzi, *https ://fr.scribd.com/doc/52713874/Poly-AN-Douzi.*

[4] Michel Pierre et Antoine Henrot, *Analyse Numérique, cours de Takéo Takahashi*, cours électif CE33, Ecole des Mines de Nancy, 2013-2014.

[5] Guillaume Legendre, *Méthodes Numériques. Introduction à l'analyse numérique et au calcul scientifique*, cours, DAUPHINE université Paris, 2009-2010.

[6] R. S. Wilson, cf. P. G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*, livre, édition Dunod, 1999.

[7] G. Bontempi, *Calcul Formel et Numérique*, Département d'informatique, Boulevard de Triomphe-CP212, 2014 .