

MÉMOIRE DE M2 POUR L'AGRÉGATION
Agrégation 2025

226

Suites vectorielles et réelles définies par une relation de récurrence
 $u_{n+1} = f(u_n)$.
Exemples. Applications à la résolution approchée d'équations.

Kylian Prigent

INTRODUCTION

Dans cette leçon, $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .

On adopte dans ce mémoire une approche inspirées des objectifs de l'agrégation. On définira donc rapidement la notion et ses propriétés les plus fondamentales afin de résérer une place prépondérante aux lieux d'utilisation des suites définies par des relations de récurrence. La première partie pourra à ce titre apparaître plutôt formelle quand les suivantes sont principalement des parties de mises en pratique de la notion suivant deux axes propices à l'illustration des intérêts des suites récurrentes : l'utilisation des suites récurrentes dans le cadre de points particuliers (points fixes, zéro d'une fonction) d'une part et des suites récurrentes convergentes dans le cadre de l'analyse numérique d'autre part.

Les suites récurrentes apparaissent naturellement en mathématiques dès que l'on traite de systèmes dynamiques discrets puisqu'ils s'agit de systèmes qui évoluent explicitement en fonction de leurs états précédents : que ce soit des chaînes de Markov (marche aléatoire sur \mathbb{Z} par exemple), des méthodes itératives, des systèmes différentiels,... Ceci peut mener à des études d'évolutions de populations comme le processus de Galton-Watson de disparition des patronymes dans l'Angleterre moderne ; et qui sont tant celles qui ont motivé la théorie historiquement que celles qui recèlent les idées importantes pour une première mise en théorie.

Les suites définies par une relation de récurrence peuvent encore servir à la mise en place de méthodes itératives en analyse numérique et en optimisation.

Elles fournissent encore des exemples très classiques de borne : la suite de Fibonacci est par exemple la pire ennemie de l'algorithme d'Euclide.

Dans une première partie on introduit la notion de suites récurrentes et on dégage les premières propriétés des suites définies par des relations de récurrence : leur lien avec la continuité, leur rapport aux fonctions qui les définissent, leurs interprétations en terme d'algèbre linéaire (matrices, polynômes), et un exemple géométrique d'utilisation de suites récurrentes.

Après avoir introduit la notion, on s'intéresse à des points particuliers dont les suites récurrentes permettent de démontrer l'existence et l'unicité. Les suites d'itérées (suites récurrentes par excellence) permettent de démontrer le théorème du point fixe de Banach, résultat central pour établir pléthore de théorème d'existence et d'unicité. La version constructive (algorithmique) du théorème de Jordan-Chevalley découle d'un autre résultat de suites récurrentes : la méthode de Newton. On conclut cette partie avec des mesures invariantes de chaînes de Markov homogènes qui constituent une émanation dans le domaine des probabilités de notre notion.

En dernier lieu on discute la convergences de méthodes numériques. Les méthodes itératives définissent des suites par une relation de récurrence et on peut exhiber une condition nécessaire et suffisante de convergente de la suite ainsi construite. La convergence de ces suites constitue un des enjeux de l'analyse numérique, particulièrement indispensable lorsque la taille des matrices explosent et que les méthodes exactes d'algèbre linéaire ne permettent plus de fournir la/une solution en un temps raisonnable. Enfin les méthodes de gradient permettent par des suites récurrentes d'approcher le minimum d'une fonction.

Développements choisis

Méthode du gradient à pas optimal (lemme 54 + théorème 55)
Perron-Frobenius et mesures invariantes (théorème 41 + application 42)

Autre développement possible

Convergence d'une suite de polygones (lemme 16 + théorème 17)

TABLE DES MATIÈRES

I Généralités sur les suites récurrentes	4
I.A Suites récurrentes réelles	4
I.B Suites récurrentes vectorielles	5
I.C Une suite de polygone	8
II Suites récurrentes et points particuliers	10
II.A Point fixe de Banach-Picard	10
II.B Point d'annulation, méthode de Newton	12
II.C Mesures invariantes de chaînes de Markov sur un espace d'état fini	15
III Méthodes numériques, convergence	20
III.A Méthodes itératives	20
III.B Une méthode de gradient	23

I GÉNÉRALITÉS SUR LES SUITES RÉCURRENTES

I.A Suites récurrentes réelles

[Amr11]

Dans toute cette leçon on s'intéressera aux suites définies par une relation de récurrence. Il est donc naturelle des commencer par se donner la définition de tels objets.

Définition 1

Soient I un intervalle de \mathbb{R} . La suite $(u_n)_{n \geq 0}$ est dite suite récurrente si elle est définie par la donnée de u_0 et la relation de récurrence $u_{n+1} = f(u_n)$, où $f : I \rightarrow \mathbb{R}$ est une application continue.

Les suites donnent un critère extrêmement utile pour vérifier la continuité d'une application en un point. On l'énonce ci-après

Théorème 2: critère séquentiel de continuité

Soient $I =]a, b[$ un intervalle de \mathbb{R} et $l \in \mathbb{R}$. Une fonction $f : I \rightarrow \mathbb{R}$ est continue au point l si et seulement si, quelle que soit la suite (u_n) d'éléments de I convergeant vers l , la suite $(f(u_n))$ converge vers $f(l)$.

On en déduit aussitôt le résultat important suivant.

Corollaire 3:

Si la suite (u_n) d'éléments de I converge vers $l \in I$, alors nécessairement $f(l) = l$. Autrement dit l est un point fixe de f .

On donne un premier exemple immédiat d'utilisation de ce résultat.

Exemple 4

Si la suite (u_n) définie par $u_0 = 1$ et $u_{n+1} = u_n^2 - u_n - 3$, sa limite vérifie nécessairement $l = l^2 - l - 3$. Or les racines du polynôme $X^2 - 2X - 3$ sont -1 et 3 . Il en résulte par le corollaire 3 que l est nécessairement égale à -1 ou 3 .

Proposition 5

Soit (u_n) une suite réelle définie par la donnée de u_0 et la relation de récurrence $u_{n+1} = f(u_n)$. On suppose de plus que $f(I) \subset I$.

1. Si f est croissante sur I , alors la suite (u_n) est monotone.
2. Si f est décroissante sur I , alors les suites extraits (u_{2n}) et (u_{2n+1}) sont monotones de sens de variation opposés.

Démonstration. 1. Puisque la fonction f est croissante sur I , la différence $u_{n+2} - u_{n+1} = f(u_{n+1}) - f(u_n)$ est de même signe que $u_{n+1} - u_n$. Par récurrence immédiate, on déduit que $u_{n+1} - u_n$ est de même signe que $u_{N+1} - u_N$ et donc la suite (u_n) est monotone.

2. Pour tout $n \in \mathbb{N}$, on a :

$$u_{2n+2} = f \circ f(u_{2n}) \quad \text{et} \quad u_{2n+3} = f \circ f(u_{2n+1})$$

Or $f \circ f$ est croissante comme composée de deux fonctions décroissantes. Le résultat précédent assure alors que les deux suites extraites sont monotones. Comparons leur sens de variation. On a $u_{2n+3} - u_{2n+1} = f(u_{2n+2}) - f(u_{2n+1})$, donc $u_{2n+3} - u_{2n+1}$ est de signe contraire à $u_{2n+2} - u_{2n+1}$, donc les suites extraites (u_{2n}) et (u_{2n+1}) ont des sens de variation opposés. ■

On peut donner l'exemple ultra-classique de suite définie par relation de récurrence suivant.

Exemple 6:

Étude de la suite définie par

$$u_{n+1} = \sin(u_n) \quad , \quad u_0 \in [-\pi/2, \pi/2].$$

Pour déterminer les points fixes de la fonction f , on étudie φ donnée sur \mathbb{R} par $\varphi(x) = \sin(x) - x$. Cette fonction est dérivable sur \mathbb{R} et on a $\varphi'(x) = \cos(x) - 1 \leq 0$, donc φ est décroissante. Comme $\varphi(0) = 0$, on conclut que si la suite (u_n) converge, alors nécessairement sa limite est égale à 0.

La fonction sinus est croissante sur $[-\pi/2, \pi/2]$, et pour tout $n \in \mathbb{N}$, on a :

$$u_n = \sin(u_{n-1}) \in [-\pi/2, \pi/2]$$

(car $\sin(u_{n-1}) \in [-1, 1]$), on peut donc appliquer la proposition précédente. On en déduit que (u_n) définit une suite décroissante si $\varphi(u_0) = u_1 - u_0 \leq 0$, i.e. si $u_0 \in [0, \pi/2]$, et une suite croissante si $\varphi(u_0) = u_1 - u_0 \geq 0$, i.e. si $u_0 \in [-\pi/2, 0]$. Dans les deux cas, comme la suite (u_n) est bornée, elle converge d'après le théorème de Bolzano-Weierstrass, et sa limite est égale à 0.

Mais les suites définies par récurrence ne sauraient être restreintes au seul domaine de l'analyse. Elles apparaissent naturellement partout en mathématiques.

Les deux items suivants proviennent du cours de Calcul Formel de Delphine Boucher, mais je suis sûr que tout.e étudiant.e lisant ceci trouvera des exemples dans sa propre option, bien que celui-ci amène un peu de diversité dans la leçon.

Exemple 7: Algorithme d'Euclide étendu

On définit l'algorithme d'Euclide étendu de a par b par le processus suivant :

- on pose $r_0 = a \quad r_1 = b \quad u_0 = 1 \quad u_1 = 0 \quad v_0 = 0 \quad v_1 = 1$.
- pour $i \geq 0$ tant que $r_{i+1} \neq 0$, on pose :
 - r_{i+2} est le reste dans la division euclidienne de r_1 par r_{i+1} et q_{i+2} le quotient.
 - On a alors $r_i = q_{i+2}r_{i+1} + r_{i+2}$
 - $u_{i+2} = u_i - q_{i+2}u_{i+1}$
 - $v_{i+2} = v_i - q_{i+2}v_{i+1}$

Et bien dans l'algorithme d'Euclide étendu on peut exprimer chaque étape récursivement à partir de l'étape précédente :

$$\begin{pmatrix} r_{i+1} & u_{i+1} & v_{i+1} \\ r_{i+2} & u_{i+2} & v_{i+2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -q_{i+2} \end{pmatrix} \begin{pmatrix} r_i & u_i & v_i \\ r_{i+1} & u_{i+1} & v_{i+1} \end{pmatrix}.$$

Ainsi la suite des coefficients et restes d'Euclide vérifient :

$$\begin{pmatrix} r_{i+1} & u_{i+1} & v_{i+1} \\ r_{i+2} & u_{i+2} & v_{i+2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -q_{i+2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -q_{i+1} \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & -q_2 \end{pmatrix} \begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \end{pmatrix}.$$

Proposition 8

Le calcul du PGCD du $(n+1)^{\text{e}}$ et du n^{e} terme de la suite de Fibonacci nécessite n passages dans la boucle dans l'algorithme précédent d'Euclide.

C'est le pire cas !

Jusqu'ici nous avons vu les suites récurrentes définies par le seul prédécesseur du terme courant. Mais il est assez peu imaginable que toutes les suites définies par des relations de récurrences ne dépendent que du terme précédent. On s'intéresse donc dans un deuxième temps à définir proprement, étendez rigoureusement, les suites récurrentes dont le terme courant dépend des h prédécesseurs. On appelle ces suites des "suites récurrentes vectorielles" parce qu'elles peuvent être vues comme des suites récurrentes d'ordre 1, c'est-à-dire ce que l'on a vu jusqu'à présent, en les représentant sous forme de vecteurs. On expliquera cette manipulation et le lecteur est invité à remarquer qu'on l'a en réalité déjà fait dans le dernier exemple. Si on extrait les termes d'indice $i+2$ de la matrice de l'algorithme d'Euclide alors les suites dépendent de leurs deux prédécesseurs immédiats.

I.B Suites récurrentes vectorielles

[Gou08]

Définition 9

Soient (E, d) un espace métrique et $h \in \mathbb{N}^*$. Une suite (u_n) à valeurs dans E est dite récurrente d'ordre h si on peut écrire

$$\forall n \geq h, \quad u_n = f(u_{n-1}, u_{n-2}, \dots, u_{n-h}) \quad (1)$$

où f est une application de E^h dans E .

Les premières valeurs u_0, \dots, u_{h-1} étant données, la relation de récurrence d'ordre h permet de calculer de manière itérative tous les autres termes de la suite.

La proposition suivante permet souvent de calculer la limite d'une suite récurrente si on sait par ailleurs qu'elle converge.

Proposition 10

Soient (E, d) un espace métrique et (u_n) une suite récurrente d'ordre $h \in \mathbb{N}^*$ vérifiant

$$\forall n \geq h, \quad u_n = f(u_{n-1}, u_{n-2}, \dots, u_{n-h})$$

où f est une application de E^h dans E .

Si la suite (u_n) converge vers une limite l et si l'application f est continue au point (l, \dots, l) , alors on a $f(l, \dots, l) = l$.

Démonstration. Il suffit de faire tendre n vers $+\infty$ dans la relation (1) et d'utiliser la continuité de f au point (l, \dots, l) . ■

Définition 11

On dit qu'une suite (u_n) à valeurs dans E vérifie une récurrence linéaire (homogène) d'ordre h à coefficients constants si

$$\forall n \geq h, \quad u_n = a_1 u_{n-1} + a_2 u_{n-2} + \dots + a_h u_{n-h} \quad (2)$$

avec $(a_1, \dots, a_h) \in \mathbb{C}^h$

Proposition 12

L'équation

$$X^h - a_1 X^{h-1} - \dots - a_h = 0$$

s'appelle équation caractéristique de la récurrence (2). Si on note r_1, \dots, r_q ses racines et $\alpha_1, \dots, \alpha_q$ ses multiplicités respectives, alors l'ensemble des suites (u_n) vérifiant (2) est l'ensemble des suites de la forme

$$u_n = P_1(n)r_1^n + \dots + P_q(n)r_q^n$$

où pour tout $1 \leq i \leq q$, P_i est un polynôme vérifiant $\deg(P_i) < \alpha_i$.

Démonstration. Si la série entière $\sum u_n z^n$ a un rayon de convergence non nul alors la récurrence (1) entraîne, après produit par z^n et sommation sur n pour z dans le disque de convergence que sa somme f vérifie :

$$f(z) - P_h(z) = a_1 z [f(z) - P_{h-1}(z)] + \dots + a_{h-1} z^{h-1} [f(z) - P_1(z)] + a_h z^h f(z)$$

où pour tout k , $1 \leq k \leq h$, $P_k(z) = u_0 + u_1 z + \dots + u_{k-1} z^{k-1}$. Ceci s'écrit aussi $f(z)Q(z) = P(z)$ où

$$Q(z) = 1 - a_1 z - \dots - a_h z^h \quad \text{et} \quad P(z) = P_h(z) - a_1 z P_{h-1}(z) - \dots - a_{h-1} z^{h-1} P_1(z).$$

Ainsi, $f(z) = P(z)/Q(z)$ est une fraction rationnelle. Par ailleurs, 0 n'est pas un pôle de $f(z)$ donc $P(z)/Q(z)$ est développable en une série entière $\sum v_n z^n$ dont le rayon de convergence est non nul. Sa somme g vérifie $Q(z)g(z) = P(z)$ sur son disque de convergence, autrement dit g vérifie la même égalité que f dans la première équation, donc la suite

(v_n) vérifie la récurrence (2). Par ailleurs, la première égalité vérifiée par g montre que pour tout k , $0 \leq k < h$, le coefficient de z^k dans $g(z) - P_h(z)$ est nul. Autrement dit, $v_k = u_k$ lorsque $0 \leq k < h$. Finalement, la suite (v_n) vérifie la même récurrence que (u_n) avec les mêmes conditions initiales. Ces suites sont donc égales, ce qui montre finalement que $\sum u_n z^n$ a bien un rayon de convergence non nul et que sa somme f vérifie $f(z) = P(z)/Q(z)$ sur son disque de convergence.

On veut maintenant calculer explicitement u_n . Soient x_1, \dots, x_{n+p} les racines de l'équation caractéristique de (2), d'ordre de multiplicité $\alpha_1, \dots, \alpha_p$. de sorte que

$$R(X) = X^h - a_1 X^{h-1} - \dots - a_h = \prod_{i=1}^p (X - x_i)^{\alpha_i} \quad \text{donc} \quad Q(X) = X^h R\left(\frac{1}{X}\right) = \prod_{i=1}^p (1 - x_i X)^{\alpha_i}.$$

Le polynôme P vérifie $\deg(P) < h = \deg(Q)$ par construction, on peut donc écrire la décomposition en éléments simples de P/Q sous la forme

$$\frac{P(X)}{Q(X)} = \sum_{i=1}^p \left(\sum_{j=1}^{\alpha_i} \frac{c_{i,j}}{(X - 1/x_i)^j} \right) \quad (c_{i,j} \in \mathbb{C}).$$

On en déduit que le développement en série entière de $f(z) = P(z)/Q(z)$ s'écrit

$$f(z) = \sum_{i=1}^p \left[\sum_{j=1}^{\alpha_i} \frac{(-1)^j c_{i,j} x_i^j}{(j-1)!} \left(\sum_{n=0}^{+\infty} \frac{(n+j-1)!}{n!} x_i^n z^n \right) \right].$$

On tire, en prenant le coefficient de z^n dans cette expression :

$$u_n = \sum_{i=1}^p \left[\sum_{j=1}^{\alpha_i} \frac{(-1)^j c_{i,j} x_i^j}{(j-1)!} (n+j-1) \dots (n+1) \right] x_i^n,$$

donc $u_n = \sum_{i=1}^p T_i(n) x_i^n$ où pour tout i , T_i est un polynôme de degré $< \alpha_i$. En désignant par Γ l'espace vectoriel des suites vérifiant cette dernière condition, nous avons prouvé que l'espace vectoriel U des suites complexes vérifiant la récurrence linéaire (2) est tel que $U \subset \Gamma$. Nous voulons prouver la condition suffisante, c'est-à-dire $\Gamma \subset U$. Pour cela, on remarque que l'application

$$\begin{array}{ccc} U & \rightarrow & \mathbb{C}^h \\ (u_n)_{n \in \mathbb{N}} & \mapsto & (u_0, u_1, \dots, u_{h-1}) \end{array}$$

est linéaire et bijective car uniquement déterminée par ses premiers termes. Ainsi, U est de dimension h , et comme Γ est de dimension $\alpha_1 + \dots + \alpha_p = h$, l'inclusion $U \subset \Gamma$ entraîne $U = \Gamma$. ■

Exemple 13

Réurrences linéaires à coefficients constants d'ordre 2 :

$$u_0, u_1, \quad \forall n \leq 2, \quad u_n = a u_{n-1} + b u_{n-2} \quad (\flat)$$

L'équation caractéristique correspondante est

$$x^2 - ax - b = 0 \quad (E)$$

- Si (E) possède deux racines (réelles) distinctes x_1, x_2 , alors les suites vérifiant (\flat) sont de la forme

$$u_n = \lambda x_1^n + \mu x_2^n$$

où λ et μ sont déterminés par u_0 et u_1 .

- Si (E) possède une racine double x , alors les suites vérifiant (\flat) sont de la forme

$$u_n = (n\lambda + \mu)x^n$$

où λ et μ sont déterminés par u_0 et u_1 .

- Si (E) possède deux racines complexes $\rho e^{i\theta}, \rho e^{-i\theta}$, alors les suites vérifiant (\flat) sont de la forme

$$u_n = \rho^n (\gamma \cos(n\theta) + \delta \sin(n\theta))$$

où γ et δ sont déterminés par u_0 et u_1 .

Mais au fait, on est dans la partie suites récurrentes vectoriel mais on n'a pas vu le lien avec les vecteurs pour l'instant ... On illustre ce lien dans la remarque suivante.

Remarque 14

On appelle suites vectorielles les suites de cette partie car on peut "vectorialiser" ces suites comme on le fait usuellement avec les équations différentielles pour les linéariser et pouvoir ainsi utiliser des méthodes comme les schémas d'Euler.

Si on se donne (u_n) un suite récurrente linéaire d'ordre h , alors elle vérifie (2) et en posant

$$v_n = \begin{pmatrix} u_{n-h+1} \\ u_{n-h+2} \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix},$$

On a que la suite (v_n) vérifie

$$v_{n+1} = Av_n$$

où A est la matrice compagnon

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \\ a_h & a_{h-1} & \dots & a_2 & a_1 \end{pmatrix} \in \mathcal{M}_h(\mathbb{K})$$

I.C Une suite de polygone

On rappelle la définition d'isobarycentre d'une famille de points dans un espace affine :

Définition 15

Soit A_1, \dots, A_n des points d'un espace affine \mathcal{E} . Alors l'isobarycentre de cette famille de points et le point $G \in \mathcal{E}$ tel que

$$\overrightarrow{A_1G} + \overrightarrow{A_2G} + \dots + \overrightarrow{A_nG} = \overrightarrow{0}.$$

On introduit ensuite un lemme technique qui nous permettra de démontrer le théorème important cette partie.

Lemme 16: déterminant circulant [Gou08]

Soient $A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \\ a_n & a_1 & \dots & a_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_2 & a_3 & \dots & a_1 \end{bmatrix}$, et $P_A = \sum_{k=1}^n a_k X^{k-1}$.

Alors $\det A = \prod_{k=1}^n P_A(\omega^k)$, avec $\omega = e^{\frac{2i\pi}{n}}$.

Démonstration.

Notons $\Omega = [\omega^{(i-1)(j-1)}] = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega & \dots & \omega^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \dots & (\omega^{n-1})^{n-1} \end{bmatrix}$.

Les ω^i étant distincts deux à deux, $\det(\Omega)$ est un déterminant de VANDERMONDE non-nul. L'idée est de calculer le déterminant de $A\Omega$.

$$\det(A\Omega) = \begin{vmatrix} P(1) & P(\omega) & \dots & P(\omega^{n-1}) \\ P(1) & P(\omega)\omega & \dots & P(\omega^{n-1})\omega^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ P(1) & P(\omega)\omega^{n-1} & \dots & P(\omega^{n-1})\omega^{(n-1)(n-1)} \end{vmatrix} = \prod_{i=0}^{n-1} P(\omega^i) \begin{vmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega & \dots & \omega^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-1} & \dots & (\omega^{n-1})^{n-1} \end{vmatrix}$$

Donc

$$\det(A) = \frac{\det(A\Omega)}{\det(A)} = \prod_{i=0}^{n-1} P(\omega^i)$$

■

Théorème 17: [PI24]

Soit $x^{(0)} = (x_1, \dots, x_n) \in \mathbb{C}^n$. On définit pour tout $k \in \mathbb{N}$

$$x^{(k+1)} = \left(\frac{x_1^{(k)} + x_2^{(k)}}{2}, \dots, \frac{x_{n-1}^{(k)} + x_n^{(k)}}{2}, \frac{x_n^{(k)} + x_1^{(k)}}{2} \right)$$

Alors $(x^{(k)})_{k \in \mathbb{N}}$ converge vers $(g, \dots, g) \in \mathbb{C}^d$, où g est l'isobarycentre de x_0, \dots, x_d .

Démonstration. Pour expliquer ce qu'il se passe géométrique, on identifie le plan complexe \mathbb{C} au plan \mathbb{R}^2 ; et si l'on prend un polygone et que l'on crée une suite de polygones en reliant les milieux des côtés, alors la suite de polygones "converge" vers l'isobarycentre.

Il suffit de remarquer que

$$\forall k \in \mathbb{N} \quad x^{(k+1)} = Ax^{(k)} \quad \text{avec } A = \begin{bmatrix} 1/2 & 1/2 & 0 & \dots & 0 \\ 0 & 1/2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & 1/2 \\ 1/2 & 0 & \dots & 0 & 1/2 \end{bmatrix}$$

Donc la suite est définie par

$$x^{(k)} = A^k x^{(0)}$$

Etudions donc la matrice A , ce qui va nous permettre d'utiliser (enfin) l'astucieux lemme énoncé plus haut. Le polynôme caractéristique de A évalué en $\lambda \in \mathbb{C}$ donne :

$$\chi_A(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 1/2 - \lambda & 1/2 & & & \\ & 1/2 - \lambda & \ddots & & \\ & & \ddots & \ddots & \\ & & & 1/2 & \\ 1/2 & & & & 1/2 - \lambda \end{vmatrix}$$

Quoi qu'il en soit, on reconnaît un déterminant circulant, et le lemme nous donne :

$$\chi_A(\lambda) = (-1)^n \prod_{i=0}^{n-1} \left(\lambda - \frac{1 + \omega^i}{2} \right) \quad \text{avec } \omega = e^{\frac{2i\pi}{n}}$$

Donc A est diagonalisable (car χ_A est scindé simple) : $A = PDP^{-1}$ avec $P \in Gl_n(\mathbb{R})$. Donc $A^k \rightarrow A^\infty$, avec $A^\infty = PD^\infty P^{-1}$, avec $D^\infty = \text{diag}(1, 0, \dots, 0)$. Donc

$$x^{(k)} \rightarrow x^\infty = A^\infty x^{(0)}$$

x^∞ est évidemment un point fixe de A , c'est-à-dire (en termes algébriques) associé à la valeur propre 1. Le sous-espace propre de A associé à 1 est de dimension 1, et le vecteur $(1, 1, \dots, 1)$ en fait évidemment partie. Donc

$$\exists z \in \mathbb{C} \text{ tel que } x^\infty = (z, z, \dots, z)$$

Pour conclure, on remarquera que l'isobarycentre g de $x^{(0)}$ est aussi (par associativité) l'isobarycentre de $x^{(1)}$ et donc de $x^{(k)}$ pour tout k . En passant à la limite, g est l'isobarycentre de x^∞ , donc $x^\infty = (g, g, \dots, g)$.

■

II SUITES RÉCURRENTES ET POINTS PARTICULIERS

II.A Point fixe de Banach-Picard

[Rou14]

Définition 18

Soient (E, d_E) , (F, d_F) deux espaces métriques. On dit qu'une application $f : E \rightarrow F$ est k-contractante s'il existe $k \in]0, 1[$, $\forall (x, y) \in E^2$, $d_F(f(x), f(y)) \leq kd_E(x, y)$.

Théorème 19: Point fixe de Banach-Picard

Soient (X, d) un espace métrique complet et $f : X \rightarrow X$ une application contractante. Alors

1. f admet un unique point fixe,
2. pour tout $x_0 \in X$, la suite des itérées de x_0 par f , définie par $x_n := f^n(x_0)$ pour tout $n \in \mathbb{N}$, converge vers a
3. la convergence est géométrique,

$$\forall n \in \mathbb{N} \quad d(x_n, a) \leq \frac{k^n}{1-k} d(x_1, x_0)$$

Démonstration. Soit $x_0 \in X$. On considère la suite $(x_n)_{n \in \mathbb{N}}$ des itérées de x_0 par f , alors pour tous n et p entiers naturels,

$$d(x_{n+p}, x_n) \leq \sum_{i=0}^{p-1} d(x_{n+i}, x_{n+i+1}) = \sum_{i=0}^{p-1} d(x_{n+i}, f(x_{n+i})) \leq \sum_{i=0}^{p-1} d(x_0, x_1) \times k^{n+i} = k^n \frac{1-k^p}{1-k} d(x_0, x_1) \leq \frac{k^n}{1-k} d(x_0, x_1)$$

Comme $0 \leq k < 1$ la suite est de Cauchy dans X complet donc converge vers $x \in X$ qui est point fixe par le corollaire 3.

Pour l'unicité, on suppose que f admet deux points fixes x et y . Alors par contractance et caractérisation du point fixe, on a :

$$d(x, y) = d(f(x), f(y)) \leq kd(x, y) < d(x, y)$$

Alors $x = y$ et le point fixe est unique.

Ne reste plus à montrer que la vitesse de convergence.

En fait on l'a déjà trouvée avec notre calcul au début de la preuve. Il suffit de faire tendre p vers $+\infty$ pour obtenir la vitesse de convergence énoncée. ■

Corollaire 20

Le théorème précédent reste valable même si on a seulement f^p est contractante pour un entier naturel p .

Démonstration.

Analyse : Supposons que f ait un point fixe notée a , alors on a $f(a) = a$. Mais alors

$$a = f(a) = f^{\circ}f(a) = \dots = f^p(a)$$

Donc a est un point fixe de f^p qui est contractante. Par le théorème précédent, ce point fixe (de f^p) est unique. Donc a est le point fixe de f^p .

Synthèse : Soit a l'unique point fixe de f^p , vérifions que a est point fixe de f .

$$f(a) = f(f^p(a)) = f^p(f(a))$$

On en déduit que $f(a)$ est point fixe de f^p . Par unicité, il s'agit donc de a . ■

On peut se poser la question de la nécessité de chacune des hypothèses du théorème 19 de point fixe de Banach-Picard. On va voir au travers des exemples qui vont suivre que les hypothèses sont optimales.

On commence par enlever la complétude de X .

Exemple 21

On prend $E =]0, 1[$ qui n'est clairement pas complet car la suite $(\frac{1}{n})_{n \geq 1}$ est de Cauchy mais ne converge pas vers un élément de X . On considère $f : x \mapsto \frac{x}{2}$. Elle est 1/2-contractante, mais elle n'a pas de point fixe dans X .

On notera que le domaine de définition et le domaine d'arrivée de la fonction que l'on considère doivent également être les mêmes.

Exemple 22

On prend $X = [0, 1]$ et $f : x \mapsto 1 + \frac{x}{2}$. Alors $f(X) = [1, 3/2]$. L'espace X n'est pas stable par f et f n'admet pas de point fixe bien qu'elle soit contractante et que X soit un espace métrique complet.

On peut ensuite s'intéresser à l'hypothèse de contractance sur f .

Exemple 23

On prend $X = [0, 1]$ qui est complet. Et on considère $f : x \mapsto x$. Alors f n'est pas contractante. Et elle n'a pas un unique point fixe puisque tous les points de X sont des points fixes.

On notera que dans les exemples suivants, on a pris des espaces X dont on savait immédiatement s'ils étaient ou non complets car on a pris $X \subset \mathbb{R}$. Or \mathbb{R} est complet donc X est complet si et seulement si X est fermé.

Enfin le contre-exemple suivant montre que la contractance ne pas être réduite à une inégalité stricte.

Exemple 24

Ce théorème est faux si l'on suppose seulement $d_F(f(x), f(y)) < d_E(x, y)$. Par exemple, la fonction définie sur \mathbb{R} par $f(x) = \sqrt{1 + x^2}$ vérifie cette inégalité mais n'admet pas de point fixe.

Ce théorème du point fixe s'applique souvent sous les hypothèses suivantes :

- X est une partie convexe fermée d'un Banach ;
- f est différentiable sur un ouvert contenant X ;
- il existe $k < 1$ tel que $\|df_x\| \leq k$ pour tout $x \in X$

En effet, on dispose alors du théorème des accroissements finis qui assure la k -contractance de f . Et il reste alors simplement à s'assurer que $f(X) \subset X$.

On peut utiliser le point fixe dans des domaines comme celui des équations différentielles. Il permet alors de démontrer le théorème de Cauchy-Lipschitz.

On s'intéresse au problème de Cauchy

$$\begin{cases} X' = f(t, X) \\ X(t_0) = x_0 \end{cases} \quad (3)$$

où $f : U \rightarrow \mathbb{R}^n$ avec U un ouvert de $\mathbb{R} \times \mathbb{R}^n$ et $(t_0, x_0) \in U$.

Définition 25

On dit que f est localement lipschitzienne en X si pour tout $(t_1, x_1) \in U$, il existe un voisinage V de x_1 , un voisinage W de t_1 et $k > 0$ tels que pour tous $x, y \in V$ et tout $t \in W$, $\|f(t, x) - f(t, y)\| \leq k\|x - y\|$.

Théorème 26: Théorème de Cauchy-Lipschitz local

Si f est continue sur U et localement lipschitzienne en X , alors (3) admet une unique solution maximale.

Démonstration. f est continue sur U donc (3) est équivalent à

$$X(t) = x_0 + \int_{t_0}^t f(u, X(u)) \, du. \quad (4)$$

Soit $V \in \mathcal{V}(x_0), W \in \mathcal{V}(t_0)$ et $k > 0$ comme dans la définition du caractère localement lipschitzien de f , on peut supposer $W \times V$ borné. On note $M := \sup_{W \times V} \|f\|$.

On se place sur un cylindre de sécurité : soit $r > 0$ tel que $\overline{B}(x_0, r) \subset V$ et soit $T > 0$ tel que $[t_0 - T, t_0 + T] \subset W$. On note \mathcal{F} l'espace des fonctions continues de $[t_0 - T, t_0 + T]$ dans $\overline{B}(x_0, r)$ muni de la norme infinie, il s'agit alors d'un espace complet.

On définit l'application Φ de \mathcal{F} dans \mathcal{F} par

$$\Phi(Y)(t) := x_0 + \int_{t_0}^t f(u, Y(u)) \, du.$$

Il faut d'abord que \mathcal{F} soit stable par Φ .

$$\|\Phi(Y)(t) - x_0\| \leq |t - t_0|M \leq TM$$

donc en choisissant $T \leq \frac{r}{M}$, $\Phi(Y)$ est bien à valeurs dans $\overline{B}(x_0, r)$ (ce choix garantit aussi que le cylindre considéré est bien un cylindre de sécurité).

Le but est maintenant de montrer que Φ admet un point fixe en utilisant le théorème de Picard. En effet, l'équation (4) implique qu'une fonction X de classe C^1 est solution de (3) si et seulement si elle est point fixe de Φ .

On va montrer que Φ admet une itérée contractante. Soit $Y, Z \in \mathcal{F}$, montrons par récurrence sur $p \in \mathbb{N}$ que

$$\forall t \in [t_0 - T, t_0 + T], \quad \|\Phi^p(Y)(t) - \Phi^p(Z)(t)\| \leq \frac{k^p |t - t_0|^p}{p!} \|Y - Z\|_\infty.$$

Cette inégalité est vraie pour $p = 0$ et

$$\begin{aligned} \|\Phi^{p+1}(Y)(t) - \Phi^{p+1}(Z)(t)\| &\leq \left| \int_{t_0}^t \|f(u, \Phi^p(Y)(u)) - f(u, \Phi^p(Z)(u))\| \, du \right| \\ &\leq \left| \int_{t_0}^t k \|\Phi^p(Y)(u) - \Phi^p(Z)(u)\| \, du \right| \\ &\leq \left| \int_{t_0}^t k \frac{k^p |u - t_0|^p}{p!} \|Y - Z\|_\infty \, du \right| \\ &= \frac{k^{p+1} |t - t_0|^{p+1}}{(p+1)!} \|Y - Z\|_\infty, \end{aligned}$$

d'où le résultat. On a donc, pour tout $p \in \mathbb{N}$ et tous $Y, Z \in \mathcal{F}$,

$$\|\Phi(Y) - \Phi(Z)\|_\infty \leq \frac{k^p T^p}{p!} \|Y - Z\|_\infty.$$

Or $\frac{k^p T^p}{p!} \xrightarrow[p \rightarrow +\infty]{} 0$ donc il existe $p \in \mathbb{N}$ tel que $\frac{k^p T^p}{p!} < 1$. D'après le théorème de point fixe de Picard, Φ admet un unique point fixe X sur \mathcal{F} , qui est donc l'unique solution de (3) sur $[t_0 - T, t_0 + T]$.

Cette solution se prolonge en une solution maximale. Supposons qu'il existe deux tels prolongements X_1 et X_2 sur deux intervalles I_1 et I_2 . L'intervalle $I_1 \cap I_2$ est non vide car il contient $[t_0 - T, t_0 + T]$. Soit J le plus grand intervalle inclus dans $I_1 \cap I_2$ et contenant $[t_0 - T, t_0 + T]$ tel que $X_1 = X_2$ sur J . Alors J est fermé dans $I_1 \cap I_2$ car $X_1 - X_2$ est continue. Si $J \neq I_1 \cap I_2$, alors on peut appliquer l'unicité locale précédemment démontrée en l'une des bornes de J et contredire la maximalité de J . Donc $J = I_1 \cap I_2$, d'où on déduit $X_1 = X_2$ sur $I_1 \cap I_2$ et, par définition de solution maximale, $I_1 = I_2$. Finalement, X se prolonge en une unique solution maximale. ■

II.B Point d'annulation, méthode de Newton

Théorème 27: Méthode de Newton [Rou14]

- Soit $f : [c, d] \rightarrow \mathbb{R}$, une fonction de classe C^2 , telle que $f(c) < 0 < f(d)$, et $f'(x) > 0$, pour tout $x \in [c, d]$. Soit a l'unique solution de $f(x) = 0$ et $F(x) = x - f(x)/f'(x)$. Alors pour $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$, il existe un réel strictement positif α tel que $I =]a - \alpha, a + \alpha[$ soit stable par F et pour tout $x_0 \in I$, (x_n) converge à l'ordre 2 vers a .
- De plus, si f est convexe, pour tout $x_0 \in [a, d]$, la méthode converge et on a :

$$0 \leq x_{n+1} - a \leq C(x_n - a)^2 \text{ et}$$

$$x_{n+1} - a \sim \frac{1}{2} \frac{f''(a)}{f'(a)} (x_n - a)^2 \text{ si } n \rightarrow +\infty.$$

Démonstration.

- Comme $f(a) = 0$, $\forall x \in [c, d]$, on a :

$$F(x) - a = x - a - \frac{f(x) - f(a)}{f'(x)} = \frac{f(a) - f(x) + (x - a)f'(x)}{f'(x)}$$

Mais alors par la formule de Taylor-Lagrange à l'ordre 2,

$$\exists z \in]a, x[\text{ tel que } F(x) - a = \frac{f''(z)(x - a)^2}{2f'(x)}$$

On pose $C = \frac{\max_{x \in [c, d]} f''(x)}{2 \min_{x \in [c, d]} f'(x)}$, il en résulte :

$$\forall x \in [c, d] \quad |F(x) - a| \leq C|x - a|^2$$

Soit $\alpha > 0$ tel que $C\alpha < 1$, et suffisamment petit pour avoir

$$I =]a - \alpha, a + \alpha[\subset [c, d].$$

Alors pour tout $x \in I$, on a :

$$|F(x) - a| \leq C\alpha^2 < \alpha$$

d'où

$$F(I) \subset I$$

Si $x_0 \in I$, pour tout $n \in \mathbb{N}$ on a donc $x_n \in I$ et :

$$|x_{n+1} - a| = |F(x_n) - a| \leq C|x_n - a|^2$$

d'où

$$C|x_n - a| \leq (C|x_0 - a|)^{2^n} \leq (C\alpha)^{2^n}$$

et la convergence d'ordre 2 de x_n vers a puisque $C\alpha < 1$.

- Pour $a \leq x \leq d$, on a $f(x) \geq 0$ et $f'(x) > 0$, d'où

$$F(x) = x - \frac{f(x)}{f'(x)} \leq x \quad (\text{inégalité stricte si } x > a)$$

et d'autre part, on a

$$F(x) - a = \frac{1}{2} \frac{f''(z)}{f'(x)} (x - a)^2 \geq 0 \quad (\text{stricte si } x > a)$$

Par conséquent, $F([a, d]) \subset [a, d]$ et pour $x_0 \in]a, d[$, on a donc $x_n \in]a, d[\forall n \in \mathbb{N}$ (si $x_0 = a$, la suite est constante). Comme la suite $(x_n)_{n \in \mathbb{N}}$ est décroissante, elle admet une limite l qui vérifie $F(l) = l$, donc $f(l) = l$ et l ne peut être que a .

La convergence de $(x_n)_{n \in \mathbb{N}}$ vers a est quadratique car on a :

$$0 \leq x_{n+1} - a \leq C(x_n - a)^2$$

Or si $a < x_0 \leq d$, pour tout $n \in \mathbb{N}$ on a $x_n > a$ et

$$\frac{x_{n+1} - a}{(x_n - a)^2} = \frac{1}{2} \frac{f''(z_n)}{f'(x_n)} \quad \text{avec } a < z_n < x_n$$

Par conséquent,

$$\frac{x_{n+1} - a}{(x_n - a)^2} = \frac{1}{2} \frac{f''(z_n)}{f'(x_n)} \xrightarrow{n \rightarrow +\infty} \frac{1}{2} \frac{f''(a)}{f'(a)}$$

■

La méthode de Newton est un résultat d'une extrême utilité en mathématiques. Il s'invite dans de très nombreux domaines, dont certains assez surprenants. C'est ainsi que l'on retrouve cette méthode dans la preuve constructive du théorème de Jordan-Chevalley-Dunford.

Notez que ce théorème est souvent attribué à tort au mathématicien étatsunien Dunford qui n'a pourtant contribuer ni à établir l'énoncé du théorème ni à en démontrer le contenu. C'est le mathématicien français Camille Jordan qui le premier démontra ce théorème puis le Bourbaki Claude Chevalley qui le plongea dans le contexte de la théorie des groupes algébriques. On remerciera au passage Mathieu Romagny pour les anecdotes fort intéressantes et le travail de contextualisation des mathématiques qu'il se donne la peine de faire et qui sont toujours pertinent.

Théorème 28: Jordan-Chevalley [Rom24]

Soit $k = \mathbb{C}$ et $A = \mathcal{M}_n(k)$. (ou k un corps parfait et A une k -algèbre de dimension finie). Alors pour tout $a \in A$ il existe un unique couple (s, n) d'éléments de A tels que :

- s est semi-simple
- n est nilpotent
- $sn = ns$
- $a = s + n$

De plus, s et n sont des polynômes en a et μ_s est la partie sans facteur carré de μ_a .

Remarque 29

Les relations de divisibilité suivantes

$$\mu_s | \mu_a | \mu_s^d \quad \text{où } d = \dim_k(A)$$

disent que μ_s et μ_a ont même liste de facteurs irréductibles. On montre ces relations.

Démonstration.

On a d'une part :

$$0 = \mu_a(a) = \mu_a(s + n) = \mu_a(s) + n(\dots)$$

Donc :

$$\mu_a(s) = -n(\dots)$$

or s est diagonalisable sur une clôture algébrique de k et n est nilpotent. Donc $\mu_a(s) = 0$ et donc $\mu_s | \mu_a$
On a d'autre part :

$$0 = \mu_s(s) = \mu_s(a - n) = \mu_s(a) - n(\dots)$$

Donc :

$$\mu_s(a) = n(\dots)$$

or n est nilpotent, donc en élevant à la puissance d on obtient :

$$\mu_s(a)^d = 0.$$

■

Preuve du théorème.

On montre l'existence (seule partie qui nous intéresse dans cette leçon) par une idée de Chevalley : on va fabriquer s comme racine de $P(x) = 0$ par la méthode de Newton !

où $P = \text{rad}(\mu_a)$

Clairement si on écrit $\mu_a = P_1^{\alpha_1} \dots P_r^{\alpha_r}$ alors on a $P = P_1 \dots P_r = \frac{\mu_a}{\mu_a \wedge \mu_a'}$.

On notera que la dernière égalité est vraie en dimension 0 car la polynôme μ_a est nécessairement non constant. Par construction il existe r (par exemple $r = \max(\alpha_i)$) tel que $\mu_a | P^r$.

En conséquence, $P(a)^r = 0$. On note $\varepsilon = P(a)$. Il s'agit donc d'un élément nilpotent. Et si x est un élément de l'idéal engendré par ε^n alors on note $x = \mathcal{O}(\varepsilon^n)$. Les facteurs P_i sont à racines simples que \bar{k} car k est parfait.

En particulier, P et P' sont premiers entre eux. Donc il en est de même pour μ_a et P' . On écrit alors une relation de Bezout entre ces deux éléments :

$$U\mu_a + VP' = 1$$

et l'on a donc :

$$U(a)\mu_a(a) + V(a)P'(a) = V(a)P'(a) = 1.$$

Donc $P'(a) \in A^\times$.

Il est venu l'heure de comprendre ce que fait ce théorème ici, on introduit l'algorithme de Newton :

$$\begin{cases} a_0 &= a \\ a_{n+1} &= a_n - \frac{P(a_n)}{P'(a_n)} \end{cases}$$

et on montre par récurrence les trois points suivants :

- (i) a_n est bien défini

(ii) $P(a_n) = \mathcal{O}(\varepsilon^{2^n})$

(iii) $a_n - a = \mathcal{O}(\varepsilon)$

Pour le cas $n = 0$, on a $P(a_0) = P(a) = \varepsilon = \mathcal{O}(\varepsilon)$ et $a_0 - a = 0 = \mathcal{O}(\varepsilon)$.

Procérons maintenant à l'hérédité.

1. $P'(a_{n+1}) = P'(a_n - \frac{P(a_n)}{P'(a_n)})$

Mais $P(X, Y) = P(X) + YP'(X) + Y^2Q(X, Y)$. Donc :

$$P'(a_{n+1}) = P'(a_n) - \frac{P(a_n)}{P'(a_n)} \dots$$

est par hypothèse de récurrence la somme d'un inversible ($P'(a_n)$) et d'un nilpotent qui commutent. Donc $P'(a_{n+1})$ est inversible.

2. $P(a_{n+1}) = P(a_n - \frac{P(a_n)}{P'(a_n)}) = P'(a_n) - P'(a_n) + \left(-\frac{P(a_n)}{P'(a_n)}\right)^2 Q(a_n, -\frac{P(a_n)}{P'(a_n)})$.

Mais par hypothèse de récurrence, $\frac{P(a_n)}{P'(a_n)} = \mathcal{O}(\varepsilon^{2^n})$. Donc $P(a_{n+1}) = \mathcal{O}(\varepsilon^{2^{n+1}})$.

3.

$$a_{n+1} - a = a_n - a - \frac{P(a_n)}{P'(a_n)} = \mathcal{O}(\varepsilon) + \mathcal{O}(\varepsilon^{2^n}) = \mathcal{O}(\varepsilon).$$

Si $n \geq \lceil \log_2(r) \rceil$ alors on a $\varepsilon^{2^n} = 0$.

Donc l'algorithme stationne à $s = a_\infty = a_n$ qui est dès lors racine de P par le deuxième point de la récurrence. De plus $n := a - s = a - a_\infty = \mathcal{O}(\varepsilon)$ est nilpotent.

On a ainsi montré l'existence d'une décomposition de Jordan-Chevalley. L'unicité est la partie facile mais hors sujet ici. ■

Remarque 30

Cet algorithme est effectif et il stationne donc converge très vite. De plus il est totalement inutile de connaître les valeurs propres de a pour en effectuer la décomposition, ce qui est une très bonne chose car la détermination algorithmique des valeurs propres est à éviter à tout prix du fait de son coût proéminent.

II.C Mesures invariantes de chaînes de Markov sur un espace d'état fini

[Rom19][App13]

Cadre probabiliste. Dans toute cette partie, on va considérer E un ensemble fini à N éléments que l'on identifiera à $\llbracket 1, N \rrbracket$, et les matrices dont on traitera seront toujours des matrices carrées $P = P(x, y)$ à N lignes et N colonnes avec $(x, y) \in E^2$.

Tous les résultats seront énoncés dans ce cadre qu'on ne rappellera pas.

On va maintenant s'intéresser à des points fixes particuliers, ils apparaissent en probabilités, plus particulièrement lorsque l'on parle de chaînes de Markov : il s'agit des mesures invariantes. On commence par rappeler ce qu'est une matrice stochastique.

Définition 31

On dit qu'une matrice P est stochastique si tous ses coefficients $P(x, y)$ sont positifs ou nuls et

$$\forall x \in E \quad \sum_{y \in E} P(x, y) = 1.$$

Les fonctions $f : E \mapsto \mathbb{R}$ seront désignées par des vecteurs colonnes et le dual de l'espace des fonctions réelles sur E est l'espace des mesures finies sur E . Les éléments de ce dernier seront notés en lignes. La dualité est donnée par la formule :

$$\langle \mu, \mathbf{1} \rangle = \mathbf{1} = \sum_{x \in E} \mu(x) f(x)$$

On note **1** la fonction constante égale à 1 et on remarque que les mesures de probabilités correspondent aux mesures positives telles que $\langle \mu, \mathbf{1} \rangle = 1$.

Enfin, la loi d'une variable aléatoire X est la mesure de probabilité telle que :

$$\forall x \in E \quad \mu(x) = \mathbb{P}(X = x).$$

Remarque 32

Si μ est une mesure de probabilité sur E et si P est une matrice stochastique, alors μP désigne la mesure de probabilité :

$$\mu P(y) = \sum_x \mu(x)P(x, y). \quad (5)$$

Une fois ce cadre probabiliste posé, on peut s'intéresser aux objets de notre propos : les chaînes de Markov.

Chaînes de Markov.

Définition 33

On dit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires à valeurs dans E est une chaîne de Markov si pour tout $k \in \mathbb{N}$ la loi conditionnelle de X_k sachant (X_0, \dots, X_{k-1}) est la loi conditionnelle de X_k sachant X_{k-1}

En d'autres termes, la suite de variables aléatoires est sans mémoire.

Exemple 34

Les suites de variables aléatoires indépendantes sont les exemples les plus simples de chaînes de Markov.

On s'intéressera dans la suite aux chaînes de Markov homogènes (sous-entendu homogène en temps, c'est-à-dire que la loi conditionnelle $\mathcal{L}(X_k | X_{k-1})$ ne dépend pas de k . C'est par exemple le cas des suites de variables aléatoires indépendantes identiquement distribuées.).

On commence ainsi à voir où les suites récurrentes vont pouvoir entrer en jeu.

Définition 35

Pour une chaîne de Markov homogène, on note P la matrice stochastique

$$P_{x,y} = P(x, y) = \mathbb{P}(X_1 = y | X_0 = x).$$

Une chaîne de Markov homogène de matrice de transition P est donc une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ telle que la loi μ_{n+1} de X_{n+1} est définie par :

$$\mu_{n+1} = \mu_n P.$$

Remarque 36

La connaissance de la matrice P et de la loi de X_0 permet de déterminer la loi de (X_0, \dots, X_n) .

Avec les notations introduites au début de cette partie, on a la proposition suivante :

Proposition 37

Soit (X_n) une chaîne de Markov homogène de matrice de transition P . Si μ_0 désigne la loi de la variable aléatoire X_0 , alors la loi de X_n est $\mu_0 P^n$

Démonstration. On désigne par μ_n la loi de X_n . Alors la loi de X_{n+1} est $\mu_{n+1} = \mu_n P$. En effet :

$$\mu_{n+1}(y) = \mathbb{P}(X_{n+1} = y) = \sum_x \mathbb{P}(X_{n+1} = y | X_n = x) \mathbb{P}(X_n = x) = \sum_x \mu_n(x)P(x, y) = \mu_n P(y).$$

Il suffit alors d'itérer pour trouver le résultat avancé. ■

Les chaînes de Markov homogènes sont donc des suites définies par récurrence.

Matrices strictement positives. On va maintenant s'intéresser à la convergence des mesures de probabilités des chaînes de Markov homogènes. Pour cela on fait un détour par un résultat d'analyse matricielle pour les matrices strictement positives : le théorème de Perron-Frobenius. Ce théorème établit des résultats sur les valeurs propres d'une telle matrice ainsi que sur la dimension de l'espace propre associé à l'unique (il le démontre) valeur propre de module maximale et enfin, assurant l'existence d'une mesure de probabilité (c'est ce dernier mot qui est important puisqu'alors toutes les composantes du vecteur doivent être dans $[0, 1]$ et de somme 1), que cet espace propre est engendré par un vecteur positif.

Définition 38

Une matrice A est dite strictement positive, et on note $A > 0$, si tous ses coefficients sont strictement positifs.

Lemme 39

On a les encadrements suivants pour le rayon spectral de A :

$$\min_{1 \leq i \leq n} \left(\sum_{j=1}^n a_{i,j} \right) \leq \rho(A) \leq \max_{1 \leq i \leq n} \left(\sum_{j=1}^n a_{i,j} \right)$$

$$\min_{1 \leq j \leq n} \left(\sum_{i=1}^n a_{i,j} \right) \leq \rho(A) \leq \max_{1 \leq j \leq n} \left(\sum_{i=1}^n a_{i,j} \right)$$

De plus, S'il existe $x \in \mathbb{R}^n$ strictement positif et α, β dans \mathbb{R}_+ tels que $\alpha x \leq Ax \leq \beta x$ [resp. $\alpha x < Ax < \beta x$], on a alors $\alpha \leq \rho(A) \leq \beta$ [resp. $\alpha < \rho(A) < \beta$].

Démonstration. Pour le premier point, on peut écrire :

On sait déjà que $\rho(A) \leq \|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n a_{i,j}$. On note $\alpha_i = \sum_{j=1}^n a_{i,j}$ pour $1 \leq i \leq n$ et $\alpha = \min_{1 \leq i \leq n} (\alpha_i)$. Pour $\alpha = 0$, le résultat est évident. Pour $\alpha > 0$, on a $\alpha_i > 0$ pour tout i et la matrice $B = \left(\left(\frac{\alpha}{\alpha_i} a_{i,j} \right) \right)_{1 \leq i,j \leq n}$ est telle que $0 \leq B \leq A$, $\sum_{j=1}^n b_{i,j} = \alpha$ pour tout i compris entre 1 et n , ce qui nous donne $\alpha = \rho(B) \leq \rho(A)$. En raisonnant avec ${}^t A$, considérant que $\rho({}^t A) = \rho(A)$ et $\|{}^t A\|_\infty = \|A\|_1$, on obtient le deuxième encadrement.

Pour le second point, l'encadrement $\alpha x \leq Ax \leq \beta x$ équivaut à $\alpha x_i \leq (Ax)_i \leq \beta x_i$ pour tout i compris entre 1 et n , ce qui entraîne :

$$\alpha \leq \min_{1 \leq i \leq n} \frac{(Ax)_i}{x_i} \leq \rho(A) \leq \max_{1 \leq i \leq n} \frac{(Ax)_i}{x_i} \leq \beta.$$

On procède de même pour les inégalités strictes. ■

Lemme 40

Soient $A \in \mathcal{M}_n(\mathbb{R})$ strictement positive et $x \in \mathbb{C}^n$ un vecteur propre non nul associé à une valeur propre λ telle que $|\lambda| = \rho(A)$. Dans ce cas, $\rho(A)$ est valeur propre de A avec $|x|$ comme vecteur propre associé, ce vecteur étant strictement positif et il existe un réel θ tel que $x = e^{i\theta}|x|$.

Démonstration. On a $\rho(A) > 0$ du fait que $A > 0$ par le premier point du lemme précédent. De $Ax = \lambda x$ avec $|\lambda| = \rho(A)$, on déduit que $\rho(A)|x| = |Ax| \leq \|A\||x| = A|x|$, donc $y = A|x| - \rho(A)|x|$ est positif. Si ce vecteur est non nul, on a alors $Ay > 0$ car A est strictement positive. En effet, pour un certain k on a alors $y_k > 0$ et il en résulte alors que pour tout i , comme A est strictement positive et y est positif, on a :

$$(Ax)_i = \sum_{j=1}^n a_{i,j} y_j \geq a_{i,k} y_k > 0.$$

Ceci signifie qu'en notant $x' = A|x|$, on a $\rho(A)x' < Ax'$ avec $x' > 0$ (le vecteur x est non nul en temps que vecteur propre) qui entraîne que $\rho(A) < \rho(A)$ par le deuxième point du lemme précédent. C'est impossible. On a donc $y = 0$, ou $A|x| = \rho(A)|x|$, ce qui signifie que $\rho(A)$ est valeur propre de A avec $|x|$ comme vecteur propre associé. En écrivant que $|x| = \frac{1}{\rho(A)} A|x|$, on déduit que $|x| > 0$. De plus :

$$A|x| = \rho(A)|x| = |\lambda x| = |Ax|,$$

c'est-à-dire :

$$\forall i \in \llbracket 1, n \rrbracket, \quad \sum_{j=1}^n |a_{i,j}x_j| = \left| \sum_{j=1}^n a_{i,j}x_j \right|.$$

Il s'agit d'un cas d'égalité dans l'inégalité triangulaire, donc tous les $a_{i,j}x_j$ ont le même argument. Ainsi il existe $\theta \in]-\pi, \pi]$ tel que :

$$\forall k \in \llbracket 1, n \rrbracket \quad \forall j \in \llbracket 1, n \rrbracket \quad a_{k,j}x_j = e^{i\theta_k} a_{k,j}|x_j|.$$

En simplifiant par $a_{k,j}$ des deux côtés de l'égalité (on peut le faire car $A > 0$), on a :

$$x = e^{i\theta}|x|.$$

■

Théorème 41: Perron-Frobenius

Si $A \in \mathcal{M}_n(\mathbb{R})$ est strictement positive. Alors :

- (i) $\rho(A)$ est l'unique valeur propre de A de module maximum ;
- (ii) l'espace propre associé à $\rho(A)$ est une droite vectorielle engendrée par un vecteur strictement positif ;
- (iii) $\rho(A)$ est valeur propre simple de A .

Démonstration. (i) Si λ est une valeur propre de la matrice A telle que $|\lambda| = \rho(A)$ et si x est un vecteur propre non nul associé, on a alors $x = e^{i\theta}|x|$ avec $A|x| = \rho(A)|x|$. Le rayon spectral $\rho(A)$ est donc valeur propre de A . De plus, avec :

$$\lambda x = Ax = A(e^{i\theta}|x|) = e^{i\theta}A|x| = e^{i\theta}\rho(A)|x| = \rho(A)x$$

on déduit que $\lambda x = \rho(A)x$ avec $x \neq 0$, et $\lambda = \rho(A)$.

Donc $\rho(A)$ est l'unique valeur propre de A de module maximal.

- (ii) En notant $E_{\rho(A)}$ l'espace propre associé à la valeur propre $\rho(A)$, tout vecteur non nul x dans $E_{\rho(A)}$ est tel que $|x| > 0$ par le lemme précédent, et aucune des composantes de x n'est nulle. S'il existe deux vecteurs x, y linéairement indépendants dans $E_{\rho(A)}$, le vecteur $z = x_1y - y_1x$ est alors non nul (linéaire indépendance) et dans $E_{\rho(A)}$ avec $z_1 = 0$, ce qui est exclus. On a donc que x et y sont linéairement dépendants. Comme ils sont quelconques dans $E_{\rho(A)}$, on en conclut $\dim(E_{\rho(A)}) = 1$; et donc $E_{\rho(A)} = \text{Vect}(|x|)$
- (iii) Pour $n = 1$, il est clair que $\rho(A)$ est valeur propre simple de A . On suppose donc que $n \geq 2$. Si la multiplicité (algébrique) de $\rho(A)$ comme valeur propre de A est $m \geq 2$, alors en se donnant un générateur $x > 0$ de l'espace propre $E_{\rho(A)}$ (de multiplicité géométrique 1), il existe $y \in \mathbb{C}^n$ linéairement indépendant de x tel que $Ay = x + \rho(A)y$ (la matrice A est semblable à une matrice de la forme

$$\begin{pmatrix} \rho(A) & 1 & 0 \\ 0 & \rho(A) & 0 \\ 0 & 0 & B \end{pmatrix}$$

En notant \bar{y} le vecteur conjugué de y dans \mathbb{C}^n , on a

$$A\bar{y} = \overline{Ay} = \overline{x + \rho(A)y} = x + \rho(A)\bar{y}$$

puisque A et x sont réels.

Le vecteur $z = \frac{1}{2}(y + \bar{y}) = \Re(y)$ est alors réel et $Az = x + \rho(A)z$.

Comme $x > 0$, il existe un réel $\alpha > 0$ tel que $v = z + \alpha x > 0$. Alors :

$$Av = Az + \alpha Ax = x + \rho(A)z + \alpha\rho(A)x = x + \rho(A)v > \rho(A)v.$$

Ceci nous donne $\rho(A) > \rho(A)$ par le second point du premier lemme technique précédent. C'est impossible. Donc $\rho(A)$ est valeur propre simple de A .

■

On peut dès cet instant constater que le théorème de Perron-Frobenius nous sera utile dans le cadre des chaînes de Markov car une matrice stochastique est évidemment positive (pas nécessairement strictement mais il ne manque pas grand chose). De plus, le rayon spectral est majoré par 1 (on peut penser aux disques de Gershgorin pour localiser les valeurs propres) et on connaît un vecteur propre pour la valeur propre 1.

Application 42

Si (X_n) est une chaîne de Markov de matrice de transition ergodique (i.e. $\exists n \in \mathbb{N} \quad P^n > 0$), alors il existe une unique loi invariante :

$$\pi^* = (\pi_1^*, \dots, \pi_N^*)$$

(i.e. $\pi^*P = \pi^*$) et

$$P^n \xrightarrow{n \rightarrow +\infty} P^\infty = \begin{pmatrix} \pi^* \\ \vdots \\ \pi^* \end{pmatrix}$$

et pour toute mesure de probabilité $\pi^{(0)}$ la suite des itérées $\pi^{(n)} = \pi^{(0)}P^n$ converge vers π^* .

Démonstration. Comme $P^n > 0$ et que $\rho(P) = 1$, il existe une probabilité invariante π pour P^n . Mais π est aussi une probabilité invariante pour tout $k \geq n$ puisque

$$P^{n+1} = PP^n = \begin{pmatrix} L_1 \\ \vdots \\ L_N \end{pmatrix} \begin{pmatrix} C_1 & \dots & C_N \end{pmatrix} = (L_i C_j)_{(i,j) \in \llbracket 1, N \rrbracket^2}$$

et $C_j > 0$ pour tout $j \in \llbracket 1, N \rrbracket$ et $\sum_j L_{i,j} = 1$ pour tout $i \in \llbracket 1, N \rrbracket$. Par Perron-Frobenius. L'espace propre de P^n associé à la valeur propre 1 est de dimension 1. Mais $Px = x$ entraîne $P^n x = x$. Donc $E_1(P) \subseteq E_1(P^n)$. De plus, $P\mathbf{1} = \mathbf{1}$. Donc $E_1(P) = E_1(P^n)$.

On écrit alors la décomposition de Jordan de P :

$$P = Q \begin{pmatrix} 1 & & & \\ & J_{\lambda_1} & & \\ & & \ddots & \\ & & & J_{\lambda_r} \end{pmatrix} Q^{-1}$$

avec $|\lambda_i| < 1$ car $\rho(P) = 1$.

Donc on a

$$P^k \longrightarrow P^\infty = Q \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} Q^{-1}.$$

Donc P^∞ est stochastique et de rang 1. Il en résulte que l'on peut écrire :

$$P^\infty = \begin{pmatrix} \pi^* \\ \vdots \\ \pi^* \end{pmatrix}$$

où π^* est une mesure de probabilité.

Ceci équivaut à $P^\infty = \mathbf{1} \cdot \pi^*$. Il s'en suit :

$$\pi^* P^\infty = \pi^* \mathbf{1} \pi^* = \pi^*$$

car $\pi^* \mathbf{1} = \sum_i \pi_i^* = 1$ puisque P^∞ est stochastique.

Mais $P^\infty = PP^\infty$. Donc :

$$\pi^* = \pi^* P^\infty = \pi^* P^\infty P = \pi^* P.$$

D'où, comme la dimension de l'espace propre associé à la valeur propre 1 est 1, π^* est l'unique mesure invariante qui soit une mesure de probabilité.

Enfin, soit $\pi^{(0)}$ une mesure de probabilité quelconque, $\pi^{(n)} = \pi^{(0)}P^n \longrightarrow \pi_0 P^\infty = \pi_0 \mathbf{1} \pi^* = \sum_{j=1}^N \pi_j^{(0)} \pi^*$.

Or $\pi^{(0)}$ est une mesure de probabilité. Donc $\sum_{j=1}^N \pi_j^{(0)} = 1$ et $\pi^{(n)} \longrightarrow \pi^*$. ■

On obtient ainsi un théorème de point fixe sur les mesures de probabilité dans le contexte des chaînes de Markov.

III MÉTHODES NUMÉRIQUES, CONVERGENCE

III.A Méthodes itératives

[All02]

Dans toute cette sous-partie $A \in \mathcal{M}_n(\mathbb{K})$ est une matrice inversible, et $b \in \mathbb{R}^n$ est un vecteur.

Définition 43

Si $(M, N) \in \mathrm{GL}_n(\mathbb{R}) \times \mathcal{M}_n(\mathbb{R})$ est tel que l'on a la décomposition $A = M - N$, une méthode itérative associée à M et N est définie par :

$$\begin{cases} x_0 \in \mathbb{K}^n & \text{donné} \\ \forall k \in \mathbb{N}^* & Mx_{k+1} = Nx_k + b \end{cases}$$

Les méthodes itératives vont consister en un découpage de la matrice inversible A sous la forme précédente (en pratique, on choisira toujours une telle décomposition de sorte que la matrice M soit facilement inversible). L'objectif est de résoudre le système $Ax = b$. En effet, si la suite (x_k) converge vers x (l'unique solution du système (A est inversible), alors par passage à la limite on a $Mx = Nx + b$ et donc $Ax = b$.

Définition 44

On dit que la méthode itérative associée à (M, N) converge si pour tout $x_0 \in \mathbb{K}^n$, la suite (x_k) de solution approchée de $Ax = b$ converge vers la solution exacte.

Remarque 45

On utilise souvent la notation suivante :

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

Elle est en effet très pratique pour définir les méthodes de Jacobi (la plus élémentaire), de Gauss-Seidel (un petit peu) plus perfectionnée ou les méthodes de relaxation.

On donne trois exemple de méthodes itératives de la plus élémentaire à la plus "sophistiquée".

Exemple 46: Méthodes de Jacobi

La méthode de Jacobi est la méthode itérative où l'on prend :

$$M = D = (a_{i,j}\delta_{i,j})_{1 \leq i, j \leq n} \quad \text{et} \quad N = D - A = E + F$$

Il s'agit de la méthode la plus naïve. Elle exige (et ce n'est pas la seule) que la diagonale de A soit composée d'éléments tous non nuls.

Exemple 47: Méthode de Gauss-Seidel

La méthode de Gauss-Seidel est la méthode itérative où l'on prend :

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

On résume dans le tableau suivant les principales caractéristiques des méthodes itératives les plus naïves. On a choisi comme exemple de méthode de relaxation de relaxer la méthode de Gauss-Seidel, mais on aurait bien évidemment pu

choisir de relaxer la méthode de Jacobi. On notera également que la méthode de relaxation n'est définie que pour $\omega \neq 0$

Nom de la méthode	Décomposition $A = M - N$	Matrice $M^{-1}N$ de la méthode itérative	Description d'une itération
Jacobi	$A = D - (E + F)$	$J = D^{-1}(E + F) = I - D^{-1}A$	$Du_{k+1} = (E + F)u_k + b$
Gauss-Seidel	$A = (D - E) - F$	$\mathcal{L}_1 = (D - E)^{-1}F$	$(D - E)u_{k+1} = Fu_k + b$
Relaxation	$A = \left(\frac{1}{\omega}D - E\right) - \left(\frac{1-\omega}{\omega}D + F\right)$	$\mathcal{L}_\omega = \left(\frac{1}{\omega}D - E\right)^{-1}\left(\frac{1-\omega}{\omega}D + F\right)$	$\left(\frac{1}{\omega}D - E\right)u_{k+1} = \left(\frac{1-\omega}{\omega}D + F\right)u_k + b$

On constate que toutes les méthodes évoquées ici nécessite que la diagonale de A ne contienne aucun 0. Cette hypothèse n'est pas affolante en ce sens que comme A est inversible, quitte à faire des combinaisons linéaires sur les lignes de la matrices A , elle répondra à cette prérogative. Si on souhaite résoudre le système linéaire $Ax = b$, alors il faut effectuer sur b les mêmes transformations que sur A et on peut alors utiliser les méthodes itératives pour calculer des solutions approchées du système linéaire. Ceci a un vrai intérêt lorsque la taille des matrices explose. En effet, on dispose de méthodes de calcul exacte de solution du système linéaire, mais si la taille de la matrice explose, alors les algorithmes de calcul de la solution tourne très longtemps avant de renvoyer la solution. Alors, souhaitant avoir un résultat en un temps raisonnable, on est souhaiterait interrompre l'exécution après un certain temps. Hélas, si on utilise une méthode exacte : Gauss, LU, Cholesky, QR (on ne fera pas l'affront de parler de Cramer pour des grandes matrices) et qu'on l'arrête avant qu'elle n'ait finie, alors on juste réussi à perdre du temps. Ceci justifie de s'intéresser aux méthodes itératives que l'on vient d'introduire.

Une fois dit ceci, on serait bien avisé de savoir qu'une méthode converge avant de l'utiliser pour trouver une solution approchée d'un système linéaire. On souhaite donc établir un critère de convergence des méthodes itératives. Par chance, il en existe. On s'intéresse donc maintenant à déterminer un tel critère.

On commence par établir quelque liens entre le rayon spectrale d'une matrices et les normes de matrices (en particulier subordonnées à des normes vectorielles).

Lemme 48

Soient $A \in \mathcal{M}_n(\mathbb{K})$ et $||| \cdot |||$ une norme matricielle subordonnée ou non. Alors

$$\rho(A) \leqslant |||A|||$$

De plus, étant donné $A \in \mathcal{M}_n(\mathbb{K})$ et $\epsilon > 0$, il existe au moins une norme matricielle subordonnée telle que

$$|||A||| \leqslant \rho(A) + \epsilon$$

Démonstration. Par le théorème de Schur (procédé de triangularisation d'une matrice complexe), il existe $U \in \mathrm{GL}_n(\mathbb{C})$ tel que

$$T = U^{-1}AU = \begin{pmatrix} \lambda_1 & t_{1,2} & \dots & t_{1,n} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{n-1,n} \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

Soit $\delta > 0$ choisi tel que :

$$\forall i \in \{1, \dots, n-1\}, \quad \sum_{j=i+1}^n \delta^{j-i} |t_{i,j}| \leqslant \epsilon$$

On note alors

$$D_\delta = \begin{pmatrix} 1 & & & \\ & \delta & (0) & \\ & (0) & \ddots & \\ & & & \delta^{n-1} \end{pmatrix}$$

$$T_\delta = D_\delta^{-1}TD_\delta = (UD_\delta)^{-1}A(UD_\delta) = \begin{pmatrix} \lambda_1 & \delta t_{1,2} & \dots & \delta^{n-1} t_{1,n} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \delta t_{n-1,n} \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

On définit la norme suivante : pour tout $B \in \mathcal{M}_n(\mathbb{K})$

$$\|B\| = \max\{\|(UD_\delta)^{-1}B(UD_\delta)x\|_\infty ; x \in \mathbb{K}^n, \|x\|_\infty = 1\}$$

Par construction, on a donc

$$\|A\| \leq \rho(A) + \epsilon$$

et cette norme est subordonnée pour la norme sur \mathbb{K}^n suivante

$$\|x\| = \|(UD_\delta)^{-1}x\|_\infty$$

En effet, en posant $x = (UD_\delta)^{-1}y$

$$\begin{aligned}\|B\| &= \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \frac{\|(UD_\delta)^{-1}B(UD_\delta)x\|_\infty}{\|x\|_\infty} \\ &= \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \frac{\|(UD_\delta)^{-1}By\|_\infty}{\|(UD_\delta)^{-1}y\|_\infty} \\ &= \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \frac{\|By\|}{\|y\|}\end{aligned}$$

■

Théorème 49: Critère de convergence des méthodes itératives

La méthode itérative associée à (M, N) converge si et seulement si $\rho(M^{-1}N) < 1$.

Démonstration.

Étape 1 - Sens indirect :

On note $e_k = u_k - u$ l'erreur d'approximation au rang $k \in \mathbb{N}^*$. On a donc

$$e_k = (M^{-1}Nu_{k-1} + M^{-1}b) - (M^{-1}Nu + M^{-1}b) = M^{-1}Ne_{k-1}$$

Par récurrence immédiate sur k , on a donc

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

Puisque, par hypothèse, $\rho(M^{-1}N) < 1$, on peut donc, par le lemme, construire une norme subordonnée $\|\cdot\|$ telle que $\|M^{-1}N\| < 1$. En particulier, pour la norme vectorielle associée, on a donc

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

et la suite $(e_k)_{k \in \mathbb{N}}$ converge bien vers 0.

Étape 2 - Sens direct :

Supposons que $\rho(M^{-1}N) \geq 1$.

Soit $\tilde{u} = \tilde{u}_1 + i\tilde{u}_2 \in \mathbb{C}^n$ un vecteur propre complexe de $M^{-1}N$ associée à une valeur propre λ de module supérieur ou égal à 1. Alors la méthode itérative ne converge pas en partant de l'une des valeurs $u_0 = u + \tilde{u}_1$ ou $u_0 = u + \tilde{u}_2$. En effet, on a (on fait la démonstration avec $u_0 = u + \tilde{u}_1$)

$$M^{-1}N(\tilde{u}_1 + i\tilde{u}_2) = \lambda(\tilde{u}_1 + i\tilde{u}_2) \Rightarrow (M^{-1}N)^k(\tilde{u}_1 + i\tilde{u}_2) = \lambda^k(\tilde{u}_1 + i\tilde{u}_2)$$

En particulier, $(M^{-1}N)^k\tilde{u}_1 = \lambda^k\tilde{u}_1$. Par conséquent,

$$\begin{aligned}e_k &= (M^{-1}N)^k e_0 = (M^{-1}N)^k(u_0 - u) \\ &= (M^{-1}N)^k(u + \tilde{u}_1 - u) \\ &= (M^{-1}N)^k\tilde{u}_1 \\ &= \lambda^k\tilde{u}_1\end{aligned}$$

Ce qui implique (puisque $|\lambda| \geq 1$)

$$\|e_k\| = |\lambda|^k \|\tilde{u}_1\| \xrightarrow[k \rightarrow +\infty]{} +\infty$$

■

Théorème 50

Soit A une matrice hermitienne symétrique définie positive, décomposée sous la forme

$$A = M - N, \quad M \in \mathrm{GL}_n(\mathbb{C})$$

Si la matrice $(M^* + N)$ est définie positive, alors

$$\rho(M^{-1}N) < 1$$

Démonstration. On commence par montrer que $M^* + N$ est hermitienne.

$$(M^* + N)^* = M + N^* = A + N + N^* = A^* + N^* + N = (A + N)^* + N = M^* + N$$

On construit la norme vectorielle définie pour tout vecteur x par $\|x\|_A = \sqrt{\langle Ax, x \rangle}$. C'est bien une norme car A est symétrique définie positive. On note également $\|\cdot\|_A$ la norme matricielle subordonnée. On va montrer que $\|M^{-1}N\|_A < 1$ ce qui implique le résultat désiré grâce à la proposition 48. On calcule

$$\|M^{-1}N\|_A^2 = \max_{\|v\|_A} (\|M^{-1}Nv\|_A^2)$$

On prend un vecteur v réalisant ce maximum.

Comme $N = M - A$, on obtient, en posant $w = M^{-1}Av$.

$$\begin{aligned} \|M^{-1}Nv\|_A^2 &= \langle Am^{-1}Nv, M^{-1}Nv \rangle \\ &= \langle Am^{-1}(M - A)v, M^{-1}(M - A)v \rangle \\ &= \langle (A - AM^{-1}A)v, (I - M^{-1}A)v \rangle \\ &= \langle Av, v \rangle - \langle AM^{-1}Av, v \rangle + \langle AM^{-1}Av, M^{-1}Av \rangle - \langle Av, M^{-1}Av \rangle \\ &= 1 - \langle M^{-1}Av, MM^{-1}Av \rangle + \langle AM^{-1}Av, M^{-1}Av \rangle - \langle MM^{-1}Av, M^{-1}Av \rangle \\ &= 1 - \langle w, Mw \rangle + \langle Aw, w \rangle - \langle Mw, w \rangle \\ &= 1 - \langle (M^* + N)w, w \rangle \end{aligned}$$

Or $\langle (M^* + N)w, w \rangle > 0$ car $M^* + N$ est hermitienne définie positive et comme A et M sont inversibles on a aussi $w \neq 0$. Donc $\|M^{-1}N\|_A < 1$ ■

Théorème 51

Soit A une matrice hermitienne symétrique définie positive. La méthode de relaxation converge si $0 < \omega < 2$. Le rayon spectral de la méthode de relaxation vérifie

$$\rho(\mathcal{L}_\omega) \geq |\omega - 1|, \quad \omega \neq 0$$

Par conséquent, la méthode de relaxation ne peut converger que si $0 < \omega < 2$.

Démonstration. Comme A est définie positive, D l'est aussi et donc pour tout $\omega \neq 0$ on a que $\frac{D}{\omega} - E$ est inversible. De plus,

$$M^* + N = \frac{D}{\omega} - E^* + \frac{1-\omega}{\omega}D + F = \frac{2-\omega}{\omega}D$$

puisque $E^* = F$. Donc $M^* + N > 0$ est équivalent à $0 < \omega < 2$.

D'autre part le déterminant de \mathcal{L}_ω vaut

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

On en déduit que $|\rho(\mathcal{L}_\omega)| \geq |\prod_{i=1}^n \lambda_i(\mathcal{L}_\omega)| = |\det(\mathcal{L}_\omega)| = |1 - \omega|^n$ où $\lambda_i(\mathcal{L}_\omega)$ sont les valeurs propres de \mathcal{L}_ω . ■

III.B Une méthode de gradient

[BB18] Les méthodes de gradient sont des méthodes itératives où l'on a remplacé un système linéaire par un problème de minimisation. Le but est de démontrer l'algorithme du gradient à pas optimal qui consiste, comme son nom

l'indique, à déterminer un "pas optimal" minimisant une certaine fonction.

Dans la suite de cette partie on considère $A \in \mathcal{S}_n^{++}(\mathbb{R})$ une matrice symétrique définie positive et $b \in \mathbb{R}^n$ un vecteur non nul. On notera les valeurs propres de A (qui sont réelles car $A \in \mathcal{S}_n^{++}(\mathbb{R})$) donc diagonalisable dans une base orthonormée de vecteurs propres d'après le théorème spectral) :

$$0 < \lambda_1 \leq \dots \leq \lambda_n$$

(et ces relations ont donc bien un sens car les valeurs propres sont réelles.)

On veut trouver une méthode qui converge vers la solution (unique car A est inversible) du système linéaire $Ax = b$. On note \bar{x} cette solution.

Proposition 52

L'application $\langle \cdot, \cdot \rangle_A$ définie par pour tous x et y éléments de \mathbb{R}^n :

$$\langle x, y \rangle_A = \langle Ax, y \rangle = \langle x, Ay \rangle$$

est un produit scalaire.

Démonstration. Comme A est symétrique définie positive, le résultat est immédiat. ■

Dans la suite on va s'intéresser à la fonction $\phi : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{R} \\ x & \mapsto \frac{1}{2}\|x\|_A^2 - {}^t x b \end{cases}$

Proposition 53

Pour tout $x \in \mathbb{R}^n$, on a :

$$\nabla \phi(x) = Ax - b$$

Démonstration. On va calculer la différentielle de ϕ en $x \in \mathbb{R}^n$.

Soit $x \in \mathbb{R}^n$. Pour tout $h \in \mathbb{R}^n$, on a :

$$\begin{aligned} \phi(x + h) &= \frac{1}{2} \langle x + h, x + h \rangle_A - {}^t x b - {}^t h b \\ &= \frac{1}{2} \langle x, x \rangle_A + \frac{1}{2} \langle x, h \rangle_A + \frac{1}{2} \langle h, x \rangle_A + \frac{1}{2} \langle h, h \rangle_A - {}^t x b - {}^t h b \\ &= \frac{1}{2} \|x\|^2 + \langle x, h \rangle_A + \frac{1}{2} \|h\|_A^2 - {}^t x b - {}^t h b \\ &= \phi(x) + \langle x, h \rangle_A - {}^t h b + \underset{\|h\| \rightarrow 0}{o}(\|h\|) \end{aligned}$$

On en déduit la différentielle de ϕ en x appliquée à h :

$$d\phi_x(h) = \langle x, h \rangle_A - {}^t h b = \langle Ax, h \rangle - \langle b, h \rangle = \langle Ax - b, h \rangle$$

Or $d\phi_x$ est une forme linéaire continue sur \mathbb{R}^n qui est un espace de Hilbert donc par théorème de représentation de Riesz, le gradient de ϕ en x est défini de manière unique et donc on peut l'identifier très simplement dans l'expression de $d\phi_x(h)$. ■

Le fait que l'on ait introduit le gradient de l'application ϕ n'est pas un hasard, les méthodes de gradient font naturellement apparaître le gradient de cette fonction.

L'objectif de la méthode du gradient à pas optimal est de déterminer un pas "optimal" en ce sens qui est grand pour que la méthode se rapproche "très vite" de la solution ; et c'est le gradient qui est utilisé pour définir ce pas "optimal". Avant de parler la méthode de gradient à pas optimal, on introduit un lemme technique qui nous servira dans la démonstration de la convergence de la méthode.

Lemme 54: Inégalité de Kantorovitch

$$\forall x \in \mathbb{R}^n \setminus \{0\} \quad \frac{\|x\|^4}{\|x\|_A^2 \|x\|_{A^{-1}}^2} \geq 4 \frac{\lambda_n \lambda_1}{(\lambda_n + \lambda_1)^2}$$

Démonstration. Soit $x \in \mathbb{R}^n \setminus \{0\}$. Par le théorème spectrale, A est diagonalisable en base orthonormée. Notons (e_1, \dots, e_n) une base de vecteurs propres avec e_i vecteur propre pour λ_i . On a alors que e_i est vecteur propre de A^{-1} pour la valeur propre $\frac{1}{\lambda_i}$. On désigne par x_i la i^e coordonnée de x dans la base (e_1, \dots, e_n) . On a donc :

$$\begin{aligned}\|x\|_A\|x\|_{A^{-1}} &= \sqrt{\sum_{i=1}^n \lambda_i x_i^2} \sqrt{\sum_{i=1}^n \frac{1}{\lambda_i} x_i^2} \\ &= \sqrt{\frac{\lambda_n}{\lambda_1}} \sqrt{\sum_{i=1}^n \frac{\lambda_i}{\lambda_n} x_i^2} \sqrt{\sum_{i=1}^n \frac{\lambda_1}{\lambda_i} x_i^2}\end{aligned}$$

Donc en utilisant l'inégalité de Young trivialisée avec $p = q = 2$, on a :

$$\|x\|_A\|x\|_{A^{-1}} \leq \frac{1}{2} \sqrt{\frac{\lambda_n}{\lambda_1}} \left(\sum_{i=1}^n \frac{\lambda_i}{\lambda_n} x_i^2 + \sum_{i=1}^n \frac{\lambda_1}{\lambda_i} x_i^2 \right)$$

C'est-à-dire :

$$\|x\|_A\|x\|_{A^{-1}} \leq \frac{1}{2} \sqrt{\frac{\lambda_n}{\lambda_1}} \sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_n} + \frac{\lambda_1}{\lambda_i} \right) x_i^2 \quad (6)$$

Or l'application $g \begin{cases} [\lambda_1, \lambda_n] & \rightarrow \mathbb{R} \\ t & \mapsto \frac{t}{\lambda_n} + \frac{\lambda_1}{t} \end{cases}$ est convexe comme somme de deux applications convexes. Or on a que :

$$g(\lambda_1) = 1 + \frac{\lambda_1}{\lambda_n} = g(\lambda_n).$$

Donc par convexité de g on obtient :

$$\forall t \in [\lambda_1, \lambda_n] \quad g(t) \leq 1 + \frac{\lambda_1}{\lambda_n}.$$

Et comme toutes les valeurs propres sont dans l'intervalle $[\lambda_1, \lambda_n]$ on a donc :

$$\frac{\lambda_i}{\lambda_n} + \frac{\lambda_1}{\lambda_i} \leq 1 + \frac{\lambda_1}{\lambda_n}. \quad (7)$$

En réinjectant l'inégalité (7) dans l'inégalité (6), on obtient finalement :

$$\begin{aligned}\|x\|_A\|x\|_{A^{-1}} &\leq \frac{1}{2} \sqrt{\frac{\lambda_n}{\lambda_1}} \left(1 + \frac{\lambda_1}{\lambda_n} \right) \sum_{i=1}^n x_i^2 \\ &= \frac{1}{2} \sqrt{\frac{\lambda_n}{\lambda_1}} \left(1 + \frac{\lambda_1}{\lambda_n} \right) \|x\|^2\end{aligned}$$

■

Théorème 55: Méthode du gradient à pas optimal

- (i) L'application ϕ atteint son minimum en \bar{x} et en \bar{x} seulement.
- (ii) Soit $a \in \mathbb{R}^n \setminus \{\bar{x}\}$ et soit $(x_k)_{k \in \mathbb{N}}$ la suite définie par :

$$\begin{cases} x_0 = a \\ \alpha_k = \begin{cases} \frac{\|\nabla \phi(x_k)\|^2}{\|\nabla \phi(x_k)\|_A^2} & \text{si } x_k \neq \bar{x} \\ 0 & \text{sinon} \end{cases} \\ x_{k+1} = x_k - \alpha_k \nabla \phi(x_k) \end{cases}$$

converge vers \bar{x} et

$$\forall k \in \mathbb{N} \quad \|x_{k+1} - \bar{x}\| \leq \sqrt{\frac{\lambda_n}{\lambda_1}} \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^{k+1} \|x_0 - \bar{x}\|.$$

Cette méthode est appelée méthode du gradient à pas optimal.

Commençons par expliquer ce qu'il se passe dans ce théorème. Le premier point ramène en fait le problème de "trouver une solution à un système linéaire" à "trouver le point qui minimise la nouvelle fonction".

Le second point de ce théorème précise la vitesse de convergence de la suite récurrente vers la solution. On peut constater que la vitesse est géométrique, donc la suite va converger rapidement vers la solution. Cette vitesse permet ainsi de concurrencer le très célèbre "pivot de Gauss". Il justifie en outre le commentaire introductif des méthodes numériques sur la recherche de solutions approchées.

Remarque 56

On remarquera que le premier item justifie l'intérêt de considérer le gradient de l'application ϕ . En effet la résolution du système linéaire $Ax = b$ revient à minimiser ϕ (c'est ce premier item qui nous le dit). Or en un point $y \in \mathbb{R}^n$, si $\nabla\phi(y) \neq 0$ alors il indique le sens dans lequel ϕ croît le plus vite.

Étant donné le point x_k , il est donc naturel de chercher le terme x_{k+1} sur la droite affine dirigée par $\nabla\phi(x_k)$ et passant par x_k (on notera que si le gradient est nul alors $x_k = \bar{x}$ et on prend alors $x_{k+1} = x_k$). En fait l'application $f : \begin{array}{ccc} \mathbb{R} & \rightarrow & \mathbb{R} \\ t & \mapsto & \phi(x_k - t\nabla\phi(x_k)) \end{array}$ atteint son minimum en $\alpha_k = \frac{\|\nabla\phi(x_k)\|^2}{\|\nabla\phi(x_k)\|_A^2}$.

Démonstration de la remarque. On démontre l'assertion

"En fait l'application $f : \begin{array}{ccc} \mathbb{R} & \rightarrow & \mathbb{R} \\ t & \mapsto & \phi(x_k - t\nabla\phi(x_k)) \end{array}$ atteint son minimum en $\alpha_k = \frac{\|\nabla\phi(x_k)\|^2}{\|\nabla\phi(x_k)\|_A^2}$."

Soit $t \in \mathbb{R}$ toujours en invoquant la démonstration de la proposition 53, on a :

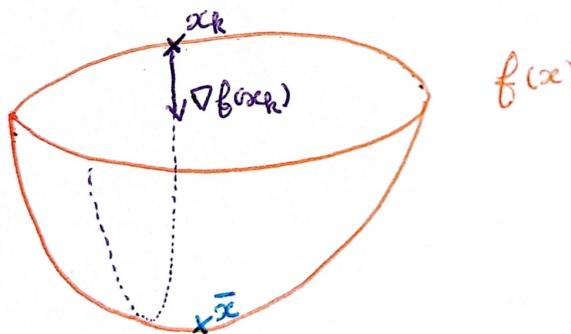
$$\begin{aligned} f(t) &= \phi(x_k) - t\langle Ac_k - b, \nabla\phi(x_k) \rangle + \frac{t^2}{2}\|\nabla\phi(x_k)\|_A^2 \\ &= \phi(x_k) - t\|\nabla\phi(x_k)\|^2 + \frac{t^2}{2}\|\nabla\phi(x_k)\|_A^2 \end{aligned}$$

et le résultat provient de l'étude de ce trinôme du degré deux. Ceci explique la définition de x_{k+1} . Enfin, on notera que la dérivée de f s'annule en α_k par minimalité de f . Donc on obtient la relation d'orthogonalité :

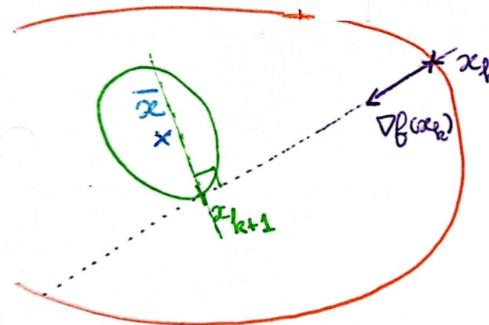
$$\langle \nabla\phi(x_{k+1}), \nabla\phi(x_k) \rangle = 0. \quad (8)$$

Remarque 57

n représente ce qu'il se passe géométriquement dans la méthode du gradient à pas optimal :



Et une "vue du-dessus" donne :



La droite (x_k, x_{k+1}) est tangente à la ligne de niveau de ϕ en x_{k+1} .

NB : les lignes de niveau de ϕ sont des sphères pour la norme $\|\cdot\|_A$.

On justifie le choix d'obtenir x_{k+1} à partir de x_k en déplaçant le point dans le sens opposé à celui donné par le gradient car la première manière d'interpréter le gradient est que géométriquement, il donne le sens de plus grande augmentation.

Démonstration du théorème. 1. Pour démontrer ce premier point on convoque directement la démonstration de la proposition 53. Dans le cas particulier où $x = \bar{x}$, pour $h \neq 0$ on a :

$$\phi(\bar{x} + h) = \phi(\bar{x}) + \frac{1}{2}\|h\|_A^2 > \phi(\bar{x})$$

2. On laisse de côté le cas trivial où pour un entier p on a $x_p = \bar{x}$. Pour alléger les notations, on pose $g_k = \nabla\phi(x_k)$.

Soit $p \in \mathbb{N}$. On a :

$$\|x_{p+1} - \bar{x}\|_A = \langle A(x_{p+1} - \bar{x}), x_{p+1} - x_p \rangle + \langle A(x_{p+1} - \bar{x}), x_p - \bar{x} \rangle.$$

Or $A(x_{p+1} - \bar{x}) = Ax_{p+1} - b = g_{p+1}$ et $x_{p+1} - x_p = -\alpha_p g_p$. Donc en utilisant la relation 8 on obtient :

$$\langle A(x_{p+1} - \bar{x}), x_{p+1} - x_p \rangle = 0.$$

En utilisant la symétrie de A , il s'en suit :

$$\begin{aligned} \|x_{p+1} - \bar{x}\|_A &= \langle A(x_{p+1} - x_p), x_p - \bar{x} \rangle + \langle A(x_p - \bar{x}), x_p - \bar{x} \rangle \\ &= \langle x_{p+1} - x_p, A(x_p - \bar{x}) \rangle + \|x_p - \bar{x}\|_A^2 \\ &= \langle -\alpha_p g_p, g_p \rangle + \|x_p - \bar{x}\|_A^2 \\ &= -\frac{\|g_p\|^4}{\|g_p\|_A^2} + \|x_p - \bar{x}\|_A^2. \end{aligned}$$

Mais par ailleurs

$$\|x_p - \bar{x}\|_A^2 = \langle A(x_p - \bar{x}), x_p - \bar{x} \rangle = \langle A(x_p - \bar{x}), A^{-1}A(x_p - \bar{x}) \rangle = \|g_p\|_{A^{-1}}^2$$

et donc

$$\|x_{p+1} - \bar{x}\|_A^2 = \left(1 - \frac{\|g_p\|^4}{\|g_p\|_A^2 \|g_p\|_{A^{-1}}^2}\right) \|x_p - \bar{x}\|_A^2. \quad (9)$$

L'heure est venue d'utiliser notre lemme technique 54 : l'inégalité de Kantorovitch à $\frac{\|g_p\|^4}{\|g_p\|_A^2 \|g_p\|_{A^{-1}}^2}$. On obtient ainsi :

$$\frac{\|g_p\|^4}{\|g_p\|_A^2 \|g_p\|_{A^{-1}}^2} \geq 4 \frac{\lambda_n \lambda_1}{(\lambda_1 + \lambda_n)^2}$$

Il en résulte (en utilisant les identités remarquables

$$\left(1 - \frac{\|g_p\|^4}{\|g_p\|_A^2 \|g_p\|_{A^{-1}}^2}\right) \leq \left(1 - 4 \frac{\lambda_n \lambda_1}{(\lambda_1 + \lambda_n)^2}\right) = \frac{(\lambda_n - \lambda_1)^2}{(\lambda_n + \lambda_1)^2}$$

En réinjectant dans (9), il vient :

$$\|x_{p+1} - \bar{x}\|_A \leq \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \|x_p - \bar{x}\|_A. \quad (10)$$

et en itérant (10), on a donc :

$$\|x_{p+1} - \bar{x}\|_A \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^{p+1} \|x_0 - \bar{x}\|_A.$$

Enfin en utilisant que l'on a $\sqrt{\lambda_1} \|\cdot\| \leq \|\cdot\|_A \leq \sqrt{\lambda_n} \|\cdot\|$, il vient :

$$\|x_{p+1} - \bar{x}\| \leq \sqrt{\frac{\lambda_n}{\lambda_1}} \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^{p+1} \|x_0 - \bar{x}\|.$$

Et cette dernière inégalité assure la convergence géométrique de la suite $(x_k)_k$ vers \bar{x} et conclut ainsi la preuve. ■

RÉFÉRENCES

- [All02] Grégoire ALLAIRE : *Algèbre linéaire numérique*. ellipses, 2002.
- [Amr11] Mohammed El AMRANI : *Suites et séries numériques, Suites et séries de fonctions*. ellipses, 2011.
- [App13] Walter APPEL : *Probabilités pour les non probabilistes*. HK, 2e éd édition, 2013.
- [BB18] Laurent BERNIS et Julien BERNIS : *Analyse pour l'agrégation de mathématiques*. ellipses, 2018.
- [Gou08] Xavier GOURDON : *Analyse*, volume 1. Calvage et Mounet, 2 édition, 2008.
- [PI24] Thimothée PECATTE et Lucas ISENmann : *L'oral à l'agrégation de mathématiques*. ellipses, 2e éd édition, 2024.
- [Rom19] Jean-Etienne ROMBALDI : *Analyse matricielle*. EDP sciences, 2e éd édition, 2019.
- [Rom24] Matthieu ROMAGNY : Algèbre de dimension finie, théorème de Jordan-Chevalley, exponentielle de matrices. Notes de cours, 2024.
- [Rou14] François ROUVIÈRE : *petit guide de calcul différentiel*. CASSINI, 4 édition, 2014.