19 avril 1999

Table des matières

Introduction Générale

La motivation originelle de ce travail était l'étude mathématique des réseaux de neurones. Nous voulions tout d'abord nous intéresser à un modèle d'intelligence artificielle ayant refait surface à la fin des années 80 : le perceptron. Celui-ci est utilisé en automatique, en approximation fonctionnelle et, par association, comme outil pour les séries temporelles non linéaires de prévision.

Dans la première partie, nous proposons une démonstration élémentaire des propriétés d'approximation uniforme des fonctions continues de $\mathbb{R}^d$ dans $\mathbb{R}$ (ou dans $\mathbb{R}^q$) par des perceptrons à une couche cachée. Nous récupérons également, le cas échéant, les propriétés d'approximation uniforme des dérivées. Contrairement aux approches antérieures, notre méthode étant constructive, elle nous permet de donner également des vitesses de convergence du perceptron (et de toutes ses dérivées) vers la fonction (et toutes ses dérivées).

Historiquement, le perceptron était implémenté par la fonction de Heaviside. Celle-ci n'étant pas continue, les utilisateurs du perceptron comme fonction de régression non linéaire pour des modèles de séries chronologiques se trouvaient donc confrontés à des problèmes d'existence de probabilités invariantes pour des chaînes de Markov non fellériennes. Ils contournaient souvent le problème en remplaçant la fonction de Heaviside par des fonctions sigmoïdes, puisque celles-ci sont continues et approchent la fonction de Heaviside avec un degré de précision arbitraire au sens de la convergence des intégrales.

Notre ambition était d'attaquer ce problème de front. Nous avons donc cherché à obtenir des critères permettant d'assurer l'existence d'une loi stationnaire, mais aussi d'obtenir des propriétés, plus fines, de stabilité et de récurrence positive, pour certains modèles markoviens non fellériens que nous précisons. Pour ce faire, nous étendons les méthodes topologiques, employées habituellement pour traiter les modèles fellériens, à nos modèles non fellériens. C'est l'objet de la deuxième partie.

La troisième partie est consacrée à l'étude des systèmes dynamiques à perturbation non évanescente. Ce sujet trouve son origine dans les problèmes de quantification vectorielle où entrent en compétition les algorithmes récursifs, stochastiques ou non, comme l'algorithme de Kohonen issu des réseaux de neurones, et les systèmes dynamiques, perturbés ou pas, comme l'algorithme de Lloyd également neuronal.

Première partie

Approximations de fonctions par des perceptrons à une couche cachée

Chapitre 1

Introduction et notations

Cette partie est la traduction française de (Attali et Pagès, 1997) publié dans *Neural Networks* sous le titre : "Approximations of Functions by a Multilayer Perceptron : a New Approach".

1.1 Présentation du perceptron

1.1.1 Le neurone formel

Le *neurone biologique* est une cellule vivante qui, recevant de multiples informations sous forme d'impulsions électriques par les *dendrites*, restitue, ou non, un signal transmis par l'*axone*. C'est ce fonctionnement schématique qu'essaye de reproduire le *neurone formel*, à savoir ; agrégation pondérée des influx ou *inputs* et restitution, ou non, d'un signal unique ou *output* au-delà d'un *seuil d'activation* donné.

On formule ce principe comme suit. On représente les d entrées (ou inputs) ainsi que leurs pondérations sous la forme de deux vecteurs de $\mathbb{R}^d$, respectivement $x := (x_1, \cdots, x_d)$ et $\alpha := (\alpha_1, \cdots, \alpha_d)$. Ce que le neurone formel "comprend" c'est :

$$\alpha.x := \sum_{i=1}^{d} \alpha_i x_i.$$

Si le seuil d'activation du neurone est θ, la sortie (ou output) est alors :

$$y := \mathbf{1}_{[\theta,+\infty[}(\alpha.x) = \mathbf{1}_{[0,+\infty[}(\alpha.x - \theta),$$

$y = 1$ signifiant la présence d'un signal de sortie et $y = 0$ son absence. On modélise souvent le seuil en ajoutant un neurone $i = 0$ d'entrée $x_0 = -1$ et de poids θ.

Par suite, la fonction de Heaviside $\mathbf{1}_{[0,+\infty[}$ a fait place à de nombreuses fonctions. Les plus classiques d'entre elles sont les fonctions sigmoïdes de température T :

$$\Phi(u) := \frac{1}{1 + e^{-\frac{u}{T}}},$$

puisque elles sont très régulières et que leur limite, quand la température T tend vers 0, est précisément la fonction de Heaviside.

1.1.2 Le perceptron simple

Le *perceptron simple*, notion introduite par (Rosenblatt, 1962), est juste un assemblage de neurones formels pour fournir plusieurs sorties, ce que l'on peut représenter sous la forme d'un vecteur de $\mathbb{R}^p$, $y := (y_1, \cdots, y_p)$. Le neurone formel est donc un perceptron simple à une *unité*.

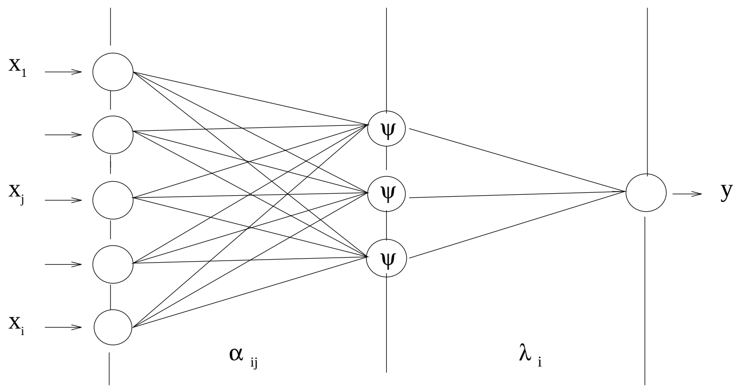
Une propriété remarquable du perceptron simple est qu'il peut être un séparateur linéaire. En effet, si deux ensembles A et B de $\mathbb{R}^d$ sont linéairement séparables, il existe une forme linéaire strictement positive sur A et strictement négative sur B. En d'autres termes, il existe un vecteur $\alpha := (\alpha_1, \cdots, \alpha_d)$ de $\mathbb{R}^d$ tel que, pour tout x de $\mathbb{R}^d$, la fonction $x \mapsto \alpha.x$ soit strictement positive si $x \in A$ et strictement négative si $x \in B$. Ainsi $y := \mathbf{1}_{[0,+\infty[}(\alpha.x)$ vaut 1 sur A et 0 sur B.

L'engouement pour le perceptron simple prend fin quand (Minsky et Papert, 1969) montrent qu'il est incapable de séparer les deux ensembles non linéairement séparables, mais pourtant très simples $A := \{(0,1); (1,0)\}$ et $B := \{(0,0); (1,1)\}$, qui sont la représentation géométrique du XOR (ou exclusif).

1.1.3 Le perceptron multi-couches

Le perceptron renaît presque simultanément en France, avec (Le Cun, 1985), et aux États-Unis, avec (Mc Clelland et Rumelhart, 1986), qui introduirent le *perceptron multi-couches*.

Ils résolurent alors le problème du XOR, en rajoutant une seule couche de neurones, dite couche cachée, entre les entrées et la sortie selon un schéma du type :



La figure précédente représente un perceptron associé à une fonction d'activation ψ et possédant $(d+1)$ unités sur la couche d'entrée, $(n+1)$ unités sur la couche cachée, et 1 unité sur la couche de sortie.

Ici $\alpha_i := (\alpha_{i1}, \cdots, \alpha_{id})$ et $x := (x_1, \cdots, x_d)$. On choisit $\alpha_{i0} := \theta_i$ et $x_0 := 1$ pour modéliser les biais. Un tel perceptron est alors une fonction de $\mathbb{R}^d$ dans $\mathbb{R}$ de la forme :

$$y(x) := \sum_{i=0}^{n} \lambda_i \psi\left(\alpha_i.x + \theta_i\right) = \sum_{i=0}^{n} \lambda_i \psi\left(\sum_{j=1}^{d} \alpha_{ij} x_j + \theta_i\right).$$

Dès lors, on commence à pressentir les propriétés d'approximateur universel de fonctions des perceptrons à une couche cachée. C'est l'objet de cette partie.

1.2 Le perceptron à une couche cachée comme approximateur universel de fonctions

L'appellation "perceptron" désignera toujours dans la suite des perceptrons à une couche cachée, puisque notre propos ne porte que sur ceux-ci.

1.2.1 Résultats antérieurs

La densité des perceptrons dans l'ensemble des fonctions continues pour la topologie de la convergence uniforme sur tout compact a été établie, en tout premier lieu, par Hornik, Stinchcombe et White (1989). Ils ont montré que, si la fonction d'activation ψ est une fonction de "seuil" – à savoir croissante et pourvue de limites finies en $\pm\infty$ – la famille des perceptrons associée est dense dans $C(\mathbb{R}^d, \mathbb{R})$ muni de la topologie de la convergence uniforme sur les compacts. Depuis, bon nombre d'améliorations ont été apportées à ce résultat. Leshno, Lin, Pinkus et Schocken (1993) étendirent ce résultat aux fonctions ψ non polynomiales, bornées sur les compacts et dont l'adhérence des points de discontinuités est de mesure Lebesgue nulle, les fonctions de seuil entrant évidemment dans ce cadre. Plus récemment, Hornik (1993) a établi que, si ψ est analytique et non polynomiale, le résultat reste vrai avec un biais universel $\theta \in \mathbb{R}$ (ce biais étant en fait un point où ψ et toutes ses dérivées sont non nulles). Il laissa comme conjecture que le résultat reste vrai pour les fonctions ψ simplement C^∞ et non polynomiales. Il étendit également le résultat aux fonctions Riemann intégrables non polynomiales ψ, le bais n'étant plus, cette fois, universel. C'est, jusqu'à maintenant, le résultat le plus général concernant l'approximation de fonctions par des perceptrons. Quand ψ est C^∞, la démonstration s'appuie sur le théorème de représentation de Riesz (Rudin, 1973), qui assure que toute forme linéaire continue de l'espace $C_c(\mathbb{R}^d, \mathbb{R})$ des fonctions continues à support compact de $\mathbb{R}^d$ dans $\mathbb{R}$ (muni de la topologie de la convergence uniforme sur tout compact) admet une représentation intégrale par rapport à une mesure μ. L'extension aux fonctions Riemann intégrables repose ensuite sur certaines propriétés régularisantes de la convolution par un "mollifieur" (i.e. approximation uniforme de l'unité sur un compact).

L'approximation uniforme d'une fonction de classe C^p et de toutes ses dérivées a d'abord été établie par Hornik, Stinchcombe et White (1990) dans le cas où la fonction d'activation ψ est non polynomiale et analytique. Ils ne donnent ni borne d'erreur (en fonction du nombre d'unités $n+1$ de la couche cachée), ni biais universel. La question des bornes d'erreur, qui est fortement liée au dimensionnement de la couche cachée, est cruciale pour l'implémentation pratique. Sur ce sujet précis, on peut citer les travaux de Barron (1993) qui donne des bornes dans le cas de l'approximation L^q ($q \geq 1$) lorsque la fonction d'activation ψ est sigmoïdale. Hornik, Stinchcombe, White et Auer (1994) étendent le résultat de Barron aux fonctions non sigmoïdales et donnent des bornes d'erreur, toujours dans le cadre de l'approximation L^q, pour la fonction approximée ainsi que pour ses dérivées. Les bornes données par Barron et Hornik et al. semblent ne pas dépendre de la dimension de l'espace des entrées. Pourtant elles le sont à travers l'espace des fonctions approchables qui se réduit lorsque la dimension de l'espace des entrées augmente. En effet, ils réalisent leurs approximations sur un ensemble de fonctions définies par des séries construites à partir des coefficients de Fourier, et de telles séries demandent de plus en plus de régularité à mesure que la dimension de l'espace des entrées augmentent.

Enfin des résultats complémentaires peuvent être trouvés dans Light (1992) ou Ellacott (1994).

1.2.2 Résumé des résultats

Nous donnons ici une démonstration élémentaire de la propriété d'approximation des fonctions continues par des perceptrons au sens de la convergence uniforme sur les compacts avec un biais universel θ, lorsque la fonction d'activation ψ est de classe C^∞ et a toutes ses dérivées non nulles en θ. Quoique plus longue que l'élégante démonstration de Hornik (1993), notre démonstration est constructive et s'appuie uniquement sur des notions élémentaires comme les développements de Taylor, les déterminants de Vandermonde et les propriétés d'approximation des polynômes. Une réponse (partielle) à la conjecture de Hornik est proposée (voir les sections 2.1 et 2.2). L'intérêt principal de notre approche est de fournir des bornes d'approximation uniforme de la fonction et de toutes ses dérivées en fonction du nombre d'unités de la couche cachée. En retour, il est alors possible de dimensionner la couche cachée pour obtenir une précision donnée.

Nous retrouvons également les résultats d'approximation L^q pour $q \geq 1$ qui découlent classiquement des résultats d'approximation uniforme. Nos bornes d'erreur dépendent de la dimension de l'espace des entrées (mais pas de l'ensemble des fonctions approximables).

L'idée consiste à approximer d'abord les polynômes. Plus précisément, nous montrons qu'un polynôme à d variables de degré global n peut être approximé sur tout compact et avec toutes ses dérivées partielles par un perceptron ayant autant d'unités sur sa couche cachée que la dimension N_n^d des polynômes à d variables de degré global inférieur ou égal à n. Ce résultat montre que les perceptrons à une couche cachée sont au moins aussi performants pour approximer une fonction continue que les polynômes. Cela est vrai également pour l'approximation des dérivées (voir les sections 2.1 et 2.2). Les bornes d'erreur sont alors données par la théorie de l'approximation polynomiale. Ainsi, la borne d'erreur pour l'approximation uniforme sur un compact K d'une fonction de classe C^p sur $\mathbb{R}^d$ par un perceptron à n unités sur la couche cachée est celle donnée par le polynôme de meilleure approximation uniforme et s'exprime en $O\left(\frac{1}{n^{\frac{p+1}{d}}}\right)$.

En ce qui concerne l'approximation d'une fonction avec ses (premières) dérivées, le cas unidimensionnel est plus satisfaisant que le cas multidimensionnel en termes de bornes d'erreur, ce qui n'est guère surprenant.

Dans le cas unidimensionnel, nous obtenons une borne en $O\left(\frac{1}{n^{p+1-\ell}}\right)$ lorsque nous approximons une fonction de classe C^p avec ses ℓ premières dérivées. L'idée est d'intégrer le polynôme de meilleure approximation de la ℓ-ième dérivée (Lorentz, 1966).

Concernant le cas multidimensionnel $(d \geq 2)$, les vitesses de convergence que nous proposons sont celles obtenues avec les polynômes multinomiaux de Bernstein qui peuvent approcher une fonction et toutes ses dérivées avec une borne d'erreur en $O\left(\frac{1}{n^{\frac{1}{2d}}}\right)$ (voir l'annexe pour la démonstration de ce résultat).

L'organisation de cette partie est la suivante : le chapitre 2 traite de l'approximation uniforme, sur un compact de $\mathbb{R}$ (section 2.1), sur un compact de $\mathbb{R}^d$ (section 2.2) et pour les perceptrons à sorties dans un espace multidimensionnel (section 2.3). Nous donnons les bornes d'erreur dans chacun des cas au chapitre 3. Enfin, l'annexe est consacrée à la démonstration de la propriété d'approximation d'une fonction et de toutes ses dérivées partielles par des polynômes multinomiaux de Bernstein.

1.2.3 Notations

• Nous appellerons“$(n+1,\psi)$-perceptrons” les perceptrons sans biais (ou plutôt avec le biais universel 0) à $n+1$ unités sur la couche cachée et dont la fonction d'activation est ψ. Ce sont donc les fonctions de la forme :

$$x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i.x), \quad x \in \mathbb{R}^d$$

• K désigne un compact de $\mathbb{R}^d$, $d \geq 1$.
• $|x|$ désignera indifféremment la valeur absolue si $x \in \mathbb{R}$ ou une norme arbitraire si $x \in \mathbb{R}^d$.
• $C(\mathbb{R}^d, \mathbb{R})$ désigne l'espace des fonctions continues de $\mathbb{R}^d$ dans $\mathbb{R}$.
• Pour toute $f \in C(\mathbb{R}^d, \mathbb{R})$, nous définissons $\|f\|_K := \sup_{x\in K} |f(x)|$.
• Pour $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ lipschitzienne, nous définissons $[f] := \sup_{x\neq y} \dfrac{|f(x)-f(y)|}{|x-y|}$.
• Pour f_k et f dans $C(\mathbb{R}^d, \mathbb{R})$, nous dirons que f_k U_K-converge vers f et nous écrirons $f_k \overset{U_K}{\longrightarrow} f$ si f_k converge uniformément vers f sur le compact K.
• $C^n(\mathbb{R}^d, \mathbb{R})$ désignera l'espace des fonctions de $\mathbb{R}^d$ dans $\mathbb{R}^d$ n fois continuement différentiables.
• Pour f_k et f dans $C^n(\mathbb{R}^d, \mathbb{R})$, où $n \geq 1$, nous dirons que f_k $U_K^{(n)}$-converge vers f et nous écrirons $f_k \overset{U_K^{(n)}}{\longrightarrow} f$ si pour tout $(\ell_1, \cdots, \ell_d) \in \{0, \cdots, n\}^d$ avec $\ell = \ell_1 + \cdots + \ell_d \leq n$:

$$\frac{\partial^\ell f_k}{\partial x_1^{\ell_1} \cdots \partial x_d^{\ell_d}} \overset{U_K}{\longrightarrow} \frac{\partial^\ell f}{\partial x_1^{\ell_1} \cdots \partial x_d^{\ell_d}},$$

avec la convention naturelle $\dfrac{\partial^0 f}{\partial x_1^0 \cdots \partial x_d^0} = f$.
• $\mathbb{R}_n[x_1, \cdots, x_d]$ désigne l'espace des fonctions polynômes à d variables de degré global inférieur ou égal à n, et $\mathbb{R}[x_1, \cdots, x_d]$ désigne l'espace des fonctions polynômes à d variables. Pour un polynôme P, d^0P désignera le degré global de P.

Chapitre 2

Les résultats d'approximation

Nous allons traiter séparément l'approximation sur un compact de $\mathbb{R}$ et celle sur un compact de $\mathbb{R}^d$, essentiellement parce que cette dernière est beaucoup plus technique, mais aussi parce que les résultats concernant les vitesses d'approximation de la fonction et de ses dérivées seront plus performants dans le cadre unidimensionnel.

Nous verrons également que notre démarche permet de montrer, à peu de frais, la propriété d'approximation universelle des perceptrons à sorties multi-dimensionnelles.

2.1 Approximation sur les compacts de $\mathbb{R}$

Nous allons tout d'abord démontrer le théorème d'approximation d'une fonction et de ses dérivées par des perceptrons. Notre approche étant constructive, elle permettra, dans la section 2.3, de fournir des vitesses de convergence. Nous répondrons également à la conjecture de Hornik sur les fonctions d'activation C^∞ non analytiques, puis, à l'aide du théorème de Muntz, nous établirons le résultat le plus fin concernant l'approximation d'une fonction continue (sans ses dérivées). Celui-ci ne permet cependant pas d'obtenir de vitesses de convergence. Nous dirons juste un mot de l'approximation L^q qui n'est qu'un simple corollaire des résultats d'approximation uniforme, via les théorèmes de densité classiques.

2.1.1 Résultat principal

Pour toute fonction ψ de classe C^p et pour tout polynôme $P \in \mathbb{R}[X]$ de degré $n \leq p$, nous exhibons une suite de perceptrons à $n+1$ unités sur la couche cachée et de biais universel 0 ($(n+1,\psi)$-perceptrons) qui $U_K^{(p)}$-converge vers P sur tout compact K de $\mathbb{R}$.

Proposition 2.1 *Soit $\psi \in C^p(\mathbb{R},\mathbb{R})$ telle que :*

$$\forall k, \quad 0 \leq k \leq p, \;\; \psi^{(k)}(0) \neq 0.$$

Soit $n \in \{0, \cdots, p\}$ et $(c_i)_{0 \leq i \leq p}$ $p+1$ réels non nuls et 2 à 2 distincts.

Alors, pour tout polynôme P de degré n, il existe $n+1$ fractions rationnelles $\lambda_i(h) := Q_i(\frac{1}{h})$ où Q_i sont des polynômes de degré n, telles que :

$$\forall K \text{ compact de } \mathbb{R}, \quad \sum_{i=0}^{n} \lambda_i(h)\psi(c_i h x) \xrightarrow{U_K^{(p)}} P(x) \text{ quand } h \to 0.$$

Démonstration : 1) Convergence du perceptron :

Soit $P(x) = \sum_{i=0}^{n} a_i x^i$, $n \leq p$, le polynôme que nous voulons approximer. Soit $(\alpha_i, \lambda_i)_{i\in\{0,\cdots,n\}}$ $2(n+1)$ nombres réels arbitraires. La formule de Taylor-Young appliquée à ψ à l'ordre n donne, pour $0 \leq i \leq n$:

$$\psi(\alpha_i x) - \psi(0) - \alpha_i x \psi'(0) - \cdots - \frac{\alpha_i^n x^n}{n!}\psi^{(n)}(0) = (\alpha_i x)^n \varepsilon(\alpha_i x), \tag{2.1}$$

avec $\lim_{x\to 0} \varepsilon(x) = 0$. Posant $A_K := \sup_{x\in K} |x^n|$ et sommant sur i les équations (2.1), on obtient :

$$\left|\sum_{i=0}^{n}\lambda_i\psi(\alpha_i x) - \psi(0)\sum_{i=0}^{n}\lambda_i - \cdots - x^n\frac{\psi^{(n)}(0)}{n!}\sum_{i=0}^{n}\lambda_i\alpha_i^n\right| \leq A_K \sum_{i=0}^{n}|\lambda_i\alpha_i^n\varepsilon(\alpha_i x)|. \tag{2.2}$$

On résoud alors le système linéaire de Vandermonde en $\lambda_0, \cdots, \lambda_n$:

$$(S_n) \equiv \begin{cases} \lambda_0 + \quad \cdots \quad +\lambda_n = \dfrac{a_0}{\psi(0)} \\ \vdots \qquad\qquad\qquad \vdots \\ \lambda_0\alpha_0^n + \quad \cdots \quad +\lambda_n\alpha_n^n = \dfrac{a_n.n!}{\psi^{(n)}(0)} \end{cases}$$

La solution du système (S_n) est donnée par :

$$\lambda_i(\alpha_0, \cdots, \alpha_n) = \begin{vmatrix} 1 & \cdots & 1 & \dfrac{a_0}{\psi(0)} & 1 & \cdots & 1 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_0^n & \cdots & \alpha_{i-1}^n & \dfrac{a_n.n!}{\psi^{(n)}(0)} & \alpha_{i+1}^n & \cdots & \alpha_n^n \end{vmatrix}_{(n+1)\times(n+1)} \frac{1}{\prod_{i>j}(\alpha_i - \alpha_j)}. \tag{2.3}$$

La clé de la démonstration est de montrer que $\lambda_i(\alpha_0, \cdots, \alpha_n)\alpha_i^n$ a une limite finie quand $\alpha := (\alpha_0, \cdots, \alpha_n) \to 0$, pour au moins une certaine classe de $(n+1)$-uplets α. L'inégalité (2.2) donnera alors le résultat annoncé. Soit $\alpha_i := c_i h$, où $c_i > 0$, $c_i \neq c_j$ si $i \neq j$ et $h > 0$.

En développant le déterminant du numérateur dans (2.3) par rapport à la colonne i, on trouve, en posant $a'_j := \dfrac{a_j.j!}{\psi^{(j)}(0)}$ et $\lambda_i(h) := \lambda_i(c_0 h, \cdots, c_n h)$:

$$\lambda_i(h) = \frac{1}{h^{\frac{n(n+1)}{2}}\prod_{i>j}(c_i - c_j)}\left((-1)^{i+1}a'_0 h^{\frac{n(n+1)}{2}}\begin{vmatrix} c_0 & \cdots & c_{i-1} & c_{i+1} & \cdots & c_n \\ \vdots & & \vdots & \vdots & & \vdots \\ c_0^n & \cdots & c_{i-1}^n & c_{i+1}^n & \cdots & c_n^n \end{vmatrix}_{n\times n} + \cdots + \right.$$

$$\left. +(-1)^{i+n+1}a'_n h^{\frac{n(n-1)}{2}}\begin{vmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 \\ \vdots & & \vdots & \vdots & & \vdots \\ c_0^{n-1} & \cdots & c_{i-1}^{n-1} & c_{i+1}^{n-1} & \cdots & c_n^{n-1} \end{vmatrix}_{n\times n}\right),$$

On remarque alors que $\lambda_i(h) = Q_i(\frac{1}{h})$ où les Q_i sont des polynômes de degré n. Donc $\lim_{h\to 0} \lambda_i(h)(c_i h)^n$ est finie car $\lambda_i(h)(c_i h)^n$ est un polynôme en h.

D'autre part, comme $\lim_{h\to 0}\sup_{x\in K}|\varepsilon(c_ihx)|=0$, on a :

$$A_K\sum_{i=0}^{n}|\lambda_i(h)(c_ih)^n||\varepsilon(c_ihx)| \xrightarrow{U_K} 0.$$

Mais (2.2) implique :

$$\left|\sum_{i=0}^{n}\lambda_i(h)\psi(c_ihx)-P(x)\right| \le A_K\sum_{i=0}^{n}|\lambda_i(h)(c_ih)^n||\varepsilon(c_ihx)|.$$

Finalement, $\forall K$ compact, $\sum_{i=0}^{n}\lambda_i(h)\psi(c_ihx) \xrightarrow{U_K} P(x)$.

2) U_K-convergence des dérivées d'ordre $k\in\{0,\cdots,p\}$:

Pour $k\le n$, la formule de Taylor-Young à l'ordre $n-k$ appliquée à $\psi^{(k)}$ donne :

$$\left|\psi^{(k)}(c_ihx)-\psi^{(k)}(0)-\cdots-\frac{(c_ih)^{n-k}\psi^{(n)}(0)}{(n-k)!}x^{n-k}\right|\le A_K^k|(c_ih)^{n-k}\varepsilon_k(c_ihx)|, \tag{2.4}$$

avec $\lim_{y\to 0}\varepsilon_k(y)=0$, et $A_K^k:=\sup_{x\in K}|x^{n-k}|$.

On vérifie facilement, à partir de la définition des $\lambda_i(h)$ par le système (S_n), que :

$$\sum_{\ell=k}^{n}\left(\sum_{i=0}^{n}\lambda_i(h)(c_ih)^\ell\right)\psi^{(\ell)}(0)\frac{x^{\ell-k}}{(\ell-k)!}=\frac{n!}{(n-k)!}a_nx^{n-k}+\cdots+k!a_k=P^{(k)}(x).$$

Multipliant alors chaque équation (2.4) par $\lambda_i(h)(c_ih)^k$ et sommant sur i, on obtient :

$$\left|\sum_{i=0}^{n}\lambda_i(h)(c_ih)^k\psi^{(k)}(c_ihx)-P^{(k)}(x)\right|\le A_K^k\sum_{i=0}^{n}|\lambda_i(h)(c_ih)^n||\varepsilon_k(c_ihx)|,$$

ce qui donne le résultat pour $k\le n$ car le membre de droite tend vers 0 quand $h\to 0$.
Si $n=p$ c'est terminé, le résultat a été démontré pour $0\le k\le n$.

Prenons maintenant $k\in\{n+1,\cdots,p\}$, et h tel que $h\le\min_i\frac{1}{c_i}$. Il vient :

$$P^{(k)}(x)=0 \text{ et } \sup_{x\in K}\left|\sum_{i=0}^{n}\lambda_i(h)(c_ih)^k\psi^{(k)}(c_ihx)\right|\le\left\|\psi^{(k)}\right\|_K\sum_{i=0}^{n}\left|\lambda_i(h)(c_ih)^k\right|. \tag{2.5}$$

$\lambda_i(h)(c_ih)^k$ est un polynôme de valuation $k-n>1$, donc $\lim_{h\to 0}\lambda_i(h)(c_ih)^k=0$.
On en déduit finalement :

$$\sum_{i=0}^{n}\lambda_i(h)(c_ih)^k\psi^{(k)}(c_ihx)\xrightarrow{U_K}0=P^{(k)}(x) \text{ quand } h\to 0. \quad \square$$

On peut maintenant énoncer le théorème d'approximation sur $\mathbb{R}$:

Théorème 2.1 *Soit $\psi\in C^\infty(\mathbb{R},\mathbb{R})$ telle que $\forall k\in\mathbb{N}$, $\psi^{(k)}(0)\ne 0$. Alors pour tout $\eta>0$, pour tout compact K de $\mathbb{R}$ et pour tout $p\in\mathbb{N}$, l'espace*

$\left\{x\mapsto\sum_{i=0}^{n}\lambda_i\psi(\alpha_ix),\ \alpha_i\in]0,\eta[,\quad n\in\mathbb{N},\ \lambda_i\in\mathbb{R}\right\}$ *est dense dans* $\left(C^p(\mathbb{R},\mathbb{R}),U_K^{(p)}\right)$.

Démonstration : D'après la proposition 2.1, il suffit de démontrer que toute fonction $f \in C^p[0,1]$ est une $U_{[0,1]}^{(p)}$-limite de polynômes. Ce résultat s'obtient, par exemple, à l'aide des polynômes de Bernstein :

$$B_n(f) := \sum_{k=0}^{n} C_n^k f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \quad \text{(Lorentz, 1966 et annexe A)} . \quad \square$$

Remarques : • L'approximation des dérivées peut être obtenue de façon plus directe (mais non constructive) en étendant la méthode de Hornik (1993). L'idée est simplement de substituer au théorème de représentation de Riesz le fait que le dual topologique de $\left(C^p(\mathbb{R},\mathbb{R}), U_K^{(p)}\right)$ est l'espace des distributions d'ordre p (dérivées p-ièmes des mesures au sens des distributions, Schwartz (1973)).
• Les polynômes de Bernstein ne sont pas le meilleur choix en terme de vitesse de convergence (voir le chapitre 3).

2.1.2 Conjecture de Hornik : une réponse

(Hornik, 1993) conjecture que le théorème 2.1 reste vrai pour toute fonction ψ C^∞ non polynomiale sous réserve de bien "choisir" le bais θ. Le corollaire suivant donne une réponse positive à cette conjecture lorsque ψ est non polynomiale *sur tout intervalle ouvert* ce qui semble un peu plus contraignant.

Corollaire 2.1 *Si $\psi \in C^\infty(\mathbb{R},\mathbb{R})$ est non polynomiale sur tout intervalle ouvert, alors il existe $\theta \in \mathbb{R}$ tel que $\forall k \in \mathbb{N},\ \psi^{(k)}(\theta) \neq 0$. Donc, avec les notations du théorème 2.1, l'espace*

$$\left\{ x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i x + \theta),\ \alpha_i \in]0,\eta[, \quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R} \right\} \text{ est dense dans } \left(C^p(\mathbb{R},\mathbb{R}), U_K^{(p)}\right).$$

Démonstration : Soit $O_k := \left\{\theta / \ \psi^{(k)}(\theta) \neq 0\right\}$. Comme $\psi^{(k)}$ est continue et que $\mathbb{R}^*$ est ouvert, $O_k = \left(\psi^{(k)}\right)^{-1}(\mathbb{R}^*)$ est ouvert.

Les O_k sont partout denses. En effet, si O_{k_0} ne l'était pas, il existerait un sous-intervalle J sur lequel $\psi^{(k_0)} = 0$ et ψ serait polynomiale sur J.

Le théorème de Baire entraîne alors que, comme toute intersection dénombrable d'ouverts denses est dense, $A := \bigcap_{k=0}^{+\infty} O_k$ est dense (et donc non vide). $\square$.

2.1.3 Extension de la propriété d'approximation

Nous venons de voir que si ψ est C^∞ et n'est polynomiale sur aucun intervalle ouvert, il existe toujours un biais θ tel que toutes les dérivées de ψ soient non nulles en θ. Les perceptrons implémentés par $\psi(.+\theta)$ sont alors denses dans l'espace des fonctions continues muni de la topologie de la convergence uniforme sur les compacts.

Cependant, un biais étant donné, par exemple $\theta = 0$, l'hypothèse de non nullité de toutes les dérivées en 0 n'est pas nécessaire pour assurer la densité uniforme des fonctions perceptrons implémentées par $\psi(.)$ dans l'espace des fonctions continues, comme le montre la proposition suivante qui s'appuie sur le théorème de Muntz.

Proposition 2.2 *Soit $\psi \in C^\infty(\mathbb{R},\mathbb{R})$. On suppose que :*

i) $\psi(0) \neq 0$,
ii) il existe une suite croissante d'entiers pairs non nuls $(j_k)_{k\in\mathbb{N}}$ telle que :

$$\forall k \in \mathbb{N},\ \psi^{(j_k)}(0) \neq 0 \ et \ \sum_{k\geq 0} \frac{1}{j_k} = +\infty,$$

iii) il existe une suite croissante d'entiers impairs $(\ell_k)_{n\in\mathbb{N}}$ telle que :

$$\forall k \in \mathbb{N}, \psi^{(\ell_k)}(0) \neq 0 \ et \ \sum_{k\geq 0} \frac{1}{\ell_k} = +\infty.$$

Alors pour tout $\eta > 0$ et pour tout compact K de $\mathbb{R}$, l'espace
$\left\{x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i x),\ \alpha_i \in]0,\eta[,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ *est dense dans* $(C(\mathbb{R},\mathbb{R}), U_K)$.

Démonstration : D'après la proposition 2.1, nous pouvons trouver, pour tout k, une suite de (j_k+1,ψ)-perceptrons satisfaisant :

$$\sum_{i=0}^{j_k} \lambda_i(h)\psi(c_i h x) \xrightarrow{U_K} x^{j_k}, \quad \text{quand } h \to 0.$$

D'après le théorème de Muntz (Lorentz, 1966), on sait que l'espace engendré par $\mathbf{1}$ et les $(x^{j_k})_{k\in\mathbb{N}}$ est $U_{[0,1]}$-dense dans $C(\mathbb{R},\mathbb{R})$. Cette famille étant paire, elle peut donc approcher uniformément toutes les fonctions paires sur $[-1,1]$.

Les hypothèses $i)$ et $ii)$ permettent donc d'assurer la densité uniforme des perceptrons dans l'espace des fonctions continues et paires sur $[-1,1]$.

De même, nous pouvons trouver, une suite de (ℓ_k+1,ψ)-perceptrons telle que :

$$\sum_{i=0}^{\ell_k} \lambda_i(h)\psi(c_i h x) \xrightarrow{U_K} x^{\ell_k},$$

et, toujours d'après le théorème de Muntz, l'espace engendré par $\mathbf{1}$ et les $(x^{\ell_k})_{k\in\mathbb{N}}$ est $U_{[0,1]}$-dense dans $C(\mathbb{R},\mathbb{R})$. L'espace engendré par les $(x^{\ell_k})_{k\in\mathbb{N}}$ est donc $U_{[0,1]}$-dense dans l'espace des fonctions de $C(\mathbb{R},\mathbb{R})$ nulles en 0. Cette famille étant impaire, elle permet donc approcher uniformément toutes les fonctions impaires sur $[-1,1]$.

On peut donc approcher uniformément sur $[-1,1]$ toutes les fonctions continues en approchant simultanément leurs parties paires et leurs parties impaires.

Une homothétie de $[-1,1]$ permet enfin d'assurer le résultat sur tout compact K. $\square$

Remarques : • Pour obtenir la $U_K^{(p)}$-densité des perceptrons dans $C^p(\mathbb{R},\mathbb{R})$ par cette approche, il est nécessaire que les p premières dérivées de ψ soient non nulles en 0. Par suite, pour obtenir la $U_K^{(p)}$-densité pour tout $p \in \mathbb{N}$, les hypothèses du théorème 2.1 nous semblent nécessaires.
• La démarche de la proposition 2.2 étant non constructive, elle ne permet pas d'obtenir de vitesses de convergence.

La densité des perceptrons dans $L^q(\mu)$ est un simple corollaire de la proposition 2.2 :

Corollaire 2.2 *Soit μ une mesure de Borel sur $\mathbb{R}$. Alors, sous les hypothèses de la proposition 2.2, $\forall q \in [1, +\infty[$, l'ensemble*

$$\left\{ x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i x),\ \alpha_i \in]0, \eta[, \quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R} \right\} \text{ est dense dans } (L^q(\mu), \|.\|_q).$$

Démonstration : L'espace $C(\mathbb{R}, \mathbb{R})$ est dense dans $(L^q(\mu), \|.\|_q)$. Le résultat est alors donné par la proposition 2.2. □

2.2 Approximation sur les compacts de $\mathbb{R}^d$

Nous suivons exactement le même plan que pour la section précédente.

2.2.1 Résultat principal

Le schéma de la démonstration est analogue à celui du cadre unidimensionnel, quoique plus technique, l'idée étant toujours d'approximer les polynômes.

Proposition 2.3 *Soit $\psi \in C^p(\mathbb{R}, \mathbb{R})$ telle que $\forall k, \quad 0 \leq k \leq p,\ \psi^{(k)}(0) \neq 0$. Soit $n \in \{0, \cdots, p\}$ et $P \in \mathbb{R}_n[X_1, \cdots, X_d]$, l'ensemble des polynômes de degré global inférieur ou égal à n. Soit $N_n^d := dim_{\mathbb{R}}\ \mathbb{R}_n[X_1, \cdots, X_d] = card \left\{ (k_1, \cdots, k_d) \in \mathbb{N}^d /\ 0 \leq k_1 + \cdots + k_d \leq n \right\}$.*

Alors il existe N_n^d vecteurs $c_i \in \mathbb{R}^d$, $1 \leq i \leq N_n^d$, et N_n^d fractions rationnelles $\lambda_i(h) := Q_i(\frac{1}{h})$, où $Q_1, \cdots, Q_{N_n^d} \in \mathbb{R}_n[X]$ tels que:

$$\forall K \text{ compact de } \mathbb{R}^d, \sum_{1 \leq i \leq N_n^d} \lambda_i(h) \psi(hc_i.x) \xrightarrow{U_K^{(p)}} P(x_1, \cdots, x_d) \text{ quand } h \to 0.$$

Démonstration : Soit $P(x_1, \cdots, x_d) := \sum_{0 \leq k_1 + \cdots + k_d \leq n} a_{k_1, \cdots, k_d} x_1^{k_1} \cdots x_d^{k_d}$ le polynôme à approximer. La formule de Taylor-Young appliquée à ψ à l'ordre n s'écrit :

$$\psi(\alpha_i^1 x_1 + \cdots + \alpha_i^d x_d) = \psi(0) + \cdots + \frac{(\alpha_i^1 x_1 + \cdots + \alpha_i^d x_d)^n}{n!} \psi^{(n)}(0) +$$
$$+ (\alpha_i^1 x_1 + \cdots + \alpha_i^d x_d)^n \varepsilon(\alpha_i^1 x_1 + \cdots + \alpha_i^d x_d).$$

Multipliant chaque équation par λ_i et sommant sur i on trouve :

$$\left| \sum_{1 \leq i \leq N_n^d} \lambda_i \psi(\alpha_i^1 x_1 + \cdots + \alpha_i^d x_d) - \psi(0) \sum_{1 \leq i \leq N_n^d} \lambda_i - \cdots - \frac{\psi^{(n)}(0)}{n!} \sum_{1 \leq i \leq N_n^d} \lambda_i (\alpha_i^1 x_1 + \cdots + \alpha_i^d x_d)^n \right|$$
$$\leq \sum_{1 \leq i \leq N_n^d} \left| \lambda_i (\alpha_i^1 x_1 + \cdots + \alpha_i^d x_d)^n \varepsilon(\alpha_i^1 x_1 + \cdots + \alpha_i^d x_d) \right|.$$

Le coefficient de $x_1^{k_1} \cdots x_d^{k_d}$ dans l'inégalité précédente est :

$$\sum_{1 \leq i \leq N_n^d} \frac{\lambda_i (\alpha_i^1)^{k_1} \cdots (\alpha_i^d)^{k_d}}{k_1! \cdots k_d!} \psi^{(k_1 + \cdots + k_d)}(0).$$

Nous avons alors à résoudre le système de dimension N_n^d en λ_i suivant :

$$\sum_{1\leq i\leq N_n^d} \lambda_i(\alpha_i^1)^{k_1}\cdots(\alpha_i^d)^{k_d} = \frac{k_1!\cdots k_d! a_{k_1,\cdots,k_d}}{\psi^{(k_1+\cdots+k_d)}(0)}, \quad 0\leq k_1+\cdots+k_d\leq n.$$

La partie la plus technique de la démonstration est de montrer que nous pouvons exhiber une suite de N_n^d vecteurs $\alpha_i = (\alpha_i^1,\cdots,\alpha_i^d)_{0\leq i\leq N_n^d}$ telle que le déterminant (de dimension N_n^d), $det\left[(\alpha_i^1)^{k_1}\cdots(\alpha_i^d)^{k_d}\right]$ soit non nul. Le lemme suivant donne une condition suffisante pour que ce résultat soit assuré :

Lemme 2.1 *Soient* $(\alpha_{j_1}^1,\cdots,\alpha_{j_d}^d)$, $0\leq j_1+\cdots+j_d\leq n$, N_n^d *vecteurs de* $\mathbb{R}^d$ *vérifiant :* $\forall \ell,\ 1\leq \ell\leq d,\ \alpha_{j_\ell}^\ell \neq \alpha_{j'_\ell}^\ell$ *si* $j_\ell\neq j'_\ell$, *alors :*

$$det\left[(\alpha_{j_1}^1)^{k_1}\cdots(\alpha_{j_d}^d)^{k_d}\right]_{\substack{0\leq k_1+\cdots+k_d\leq n\\ 0\leq j_1+\cdots+j_d\leq n}} \neq 0.$$

Démonstration du lemme 2.1 : on va démontrer ce résultat par récurrence sur la dimension de l'espace.
Pour $d=1$, on retrouve un déterminant de Vandermonde non nul.
Supposons le résultat vrai pour $d-1$.
Il faut alors montrer que la matrice $A := \left[(\alpha_{j_1}^1)^{k_1}\cdots(\alpha_{j_d}^d)^{k_d}\right]_{\substack{0\leq k_1+\cdots+k_d\leq n\\ 0\leq j_1+\cdots+j_d\leq n}}$, (ou sa transposée ${}^tA$) a un noyau réduit à 0 quand $\alpha_{j_i}^i \neq \alpha_{j'_i}^i,\ j_i\neq j'_i,\ 1\leq i\leq d$.

Le sytème ${}^tAx=0$, $x\in\mathbb{R}^{N_n^d}$ s'écrit :

$$\forall\ 0\leq j_1+\cdots+j_d\leq n, \quad \sum_{0\leq k_1+\cdots+k_d\leq n} x_{k_1,\cdots,k_d}(\alpha_{j_1}^1)^{k_1}\cdots(\alpha_{j_d}^d)^{k_d} = 0. \tag{2.6}$$

Notre objectif est de prouver que $\forall\ 0\leq k_1+\cdots+k_d\leq n,\ x_{k_1,\cdots,k_d}=0$.

En factorisant $(\alpha_{j_1}^1)^{k_1}\cdots(\alpha_{j_{d-1}}^{d-1})^{k_{d-1}}$ dans l'équation (2.6), on trouve :

$$\forall\ 0\leq j_1+\cdots+j_d\leq n, \quad \sum_{0\leq k_1+\cdots+k_{d-1}\leq n} (\alpha_{j_1}^1)^{k_1}\cdots(\alpha_{j_{d-1}}^{d-1})^{k_{d-1}} P_{k_1,\cdots,k_{d-1}}(\alpha_{j_d}^d) = 0, \tag{2.7}$$

avec $P_{k_1,\cdots,k_{d-1}}\in\mathbb{R}_{n-(k_1+\cdots+k_{d-1})}[X]$.

Pour $j_d=0$, (2.7) est un système linéaire carré en les variables $P_{k_1,\cdots,k_{d-1}}(\alpha_0^d)$:

$$\forall\ 0\leq j_1+\cdots+j_{d-1}\leq n, \quad \sum_{0\leq k_1+\cdots+k_{d-1}\leq n} (\alpha_{j_1}^1)^{k_1}\cdots(\alpha_{j_{d-1}}^{d-1})^{k_{d-1}} P_{k_1,\cdots,k_{d-1}}(\alpha_0^d) = 0. \tag{2.8}$$

L'hypothèse de récurrence assure que le système (2.8) est de Cramer, d'où :

$$\forall\ 0\leq k_1+\cdots+k_{d-1}\leq n,\ P_{k_1,\cdots,k_{d-1}}(\alpha_0^d) = 0.$$

Pour $k_1+\cdots+k_{d-1}=n$, le degré de $P_{k_1,\cdots,k_{d-1}}$ est 0, donc $P_{k_1,\cdots,k_{d-1}}=0$. L'expression (2.7) devient alors :

$$\forall\ 0\leq j_1+\cdots+j_d\leq n, \quad \sum_{0\leq k_1+\cdots+k_{d-1}\leq n-1} (\alpha_{j_1}^1)^{k_1}\cdots(\alpha_{j_{d-1}}^{d-1})^{k_{d-1}} P_{k_1,\cdots,k_{d-1}}(\alpha_{j_d}^d) = 0. \tag{2.9}$$

Prenant $j_d = 1$, (2.9) devient un système linéaire carré en les variables $P_{k_1,\cdots,k_{d-1}}(\alpha_1^d)$:

$$\forall\ 0 \leq j_1 + \cdots + j_{d-1} \leq n-1,\ \sum_{0\leq k_1+\cdots+k_{d-1}\leq n-1} (\alpha_{j_1}^1)^{k_1} \cdots (\alpha_{j_{d-1}}^{d-1})^{k_{d-1}} P_{k_1,\cdots,k_{d-1}}(\alpha_1^d) = 0. \qquad (2.10)$$

De nouveau, l'hypothèse de récurrence assure que (2.10) est de Cramer, et donc :

$$\forall\ 0 \leq k_1 + \cdots + k_{d-1} \leq n-1,\ P_{k_1,\cdots,k_{d-1}}(\alpha_1^d) = 0.$$

Quand $k_1+\cdots+k_{d-1} = n-1$, le degré de $P_{k_1,\cdots,k_{d-1}}$ est 1 et nous avons à la fois que $P_{k_1,\cdots,k_{d-1}}(\alpha_0^d)$ et $P_{k_1,\cdots,k_{d-1}}(\alpha_1^d)$ sont nuls. Mais alors, $P_{k_1,\cdots,k_{d-1}}$ est nul, car $\alpha_0^d \neq \alpha_1^d$.

Finalement, une récurrence descendante jusqu'à $k_1 + \cdots + k_{d-1} = 0$ donne :

$$\forall\ 0 \leq k_1 + \cdots + k_{d-1} \leq n,\ P_{k_1,\cdots,k_{d-1}} = 0, \text{ i.e. tous les } x_{k_1,\cdots,k_d} \text{ sont nuls .} \quad \square$$

La fin de la démonstration est alors la même que pour la dimension 1. Nous prenons $(j_1, \cdots, j_d)$ tels que $0 \leq j_1 + \cdots + j_d \leq n$:

$$(\alpha_{j_1}^1(h), \cdots, \alpha_{j_d}^d(h)) := (c_{j_1}^1 h, \cdots, c_{j_d}^d h) = h c_{j_1,\cdots,j_d},$$

où les N_n^d vecteurs $c_{j_1,\cdots,j_d} := (c_{j_1}^1, \cdots, c_{j_d}^d)$ satisfont la condition du lemme.

Il est alors facile de voir que la solution du système en $\lambda_{j_1,\cdots,j_d}(h)$ est un polynôme de degré n en $\dfrac{1}{h}$. Donc $\lambda_{j_1,\cdots,j_d} h^n$ tend vers une limite finie $n_{j_1,\cdots,j_d}$ quand h tend vers zéro. Nous avons alors pour tout compact K :

$$\begin{cases} \lambda_{j_1,\cdots,j_d}(h)(c_{j_1}^1 h x_1 + \cdots + c_{j_d}^d h x_d)^n \xrightarrow{U_K} n_{j_1,\cdots,j_d}(c_{j_1}^1 x_1 + \cdots + c_{j_d}^d x_d)^n, \\ \varepsilon(c_{j_1}^1 h x_1 + \cdots + c_{j_d}^d h x_d) \xrightarrow{U_K} 0. \end{cases}$$

Donc, quand h tend vers zéro :

$$\sum_{0\leq j_1+\cdots+j_d\leq n} \left|\lambda_{j_1,\cdots,j_d}(h) h^n (c_{j_1}^1 x_1 + \cdots + c_{j_d}^d x_d)^n \varepsilon(c_{j_1}^1 h x_1 + \cdots + c_{j_d}^d h x_d)\right| \xrightarrow{U_K} 0.$$

Cela donne : $\displaystyle\sum_{0\leq j_1+\cdots+j_d} \lambda_{j_1,\cdots,j_d}(h)\psi(h c_{j_1,\cdots,j_d}.(x_1, \cdots, x_d)) \xrightarrow{U_K} P(x_1, \cdots, x_d)$, quand $h \to 0$.

La convergence uniforme de toutes les dérivées partielles jusqu'à l'ordre p fonctionne alors de la même manière que pour la proposition 2.1. $\square$

Théorème 2.2 *Soit $\psi \in C^\infty(\mathbb{R}, \mathbb{R})$ telle que $\forall k \in \mathbb{N}$, $\psi^{(k)}(0) \neq 0$. Alors pour tout $\eta > 0$ et pour tout compact K de $\mathbb{R}^d$, l'espace*

$\left\{x \mapsto \displaystyle\sum_{i=0}^n \lambda_i \psi(\alpha_i.x),\quad x \in \mathbb{R}^d,\ \alpha_i := (\alpha_i^1, \cdots, \alpha_i^d) \in (]0, \eta[)^d,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ *est dense dans* $\left(C^p(\mathbb{R}^d, \mathbb{R}), U_K^{(p)}\right)$ *pour tout $p \in \mathbb{N}$.*

Démonstration : Les polynômes de Bernstein en dimension d sur le simplexe $S_d := \left\{(x_1, \cdots, x_d)/\ \forall i,\quad x_i \geq 0 \text{ et } \sum_{i=1}^{d} x_i \leq 1\right\}$ sont définis par :

$$B_n(f) := \sum_{|k|=n} C_n^k x_1^{k_1} \cdots x_d^{k_d} \left(1 - (x_1 + \cdots + x_d)\right)^{k_{d+1}} f\left(\frac{k_1}{n}, \cdots, \frac{k_d}{n}\right),$$

avec $k := (k_1, \cdots, k_d, k_{d+1})$, $|k| = k_1 + \cdots + k_d + k_{d+1}$ et $C_n^k = \dfrac{n!}{k_1! \cdots k_d! k_{d+1}!}$. Pour tout $p \in \mathbb{N}$, ces polynômes sont denses dans $\left(C^p(\mathbb{R}^d, \mathbb{R}), U_{S_d}^{(p)}\right)$ (Lorentz, 1966).

Pour approximer une fonction sur un compact K quelconque, il suffit alors de faire une homothétie-translation de l'ensemble S_d pour qu'il contienne le compact K. $\square$

2.2.2 Conjecture de Hornik dans le cas multi-dimensionnel

Le corollaire suivant apporte une réponse à la conjecture de Hornik dans le cas multidimensionnel.

Corollaire 2.3 *Comme pour le théorème 2.1, si $\psi \in C^\infty(\mathbb{R}, \mathbb{R})$ est non polynomiale sur tout intervalle ouvert, il existe $\theta \in \mathbb{R}$ tel que $\forall k\ \psi^{(k)}(\theta) \neq 0$. On a alors, avec les notations du théorème 2.2, l'espace*
$\left\{x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i.x + \theta),\ \alpha_i = (\alpha_i^1, \cdots, \alpha_i^d) \in (]0, \eta[)^d,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ *est dense dans* $\left(C^p(\mathbb{R}^d, \mathbb{R}), U_K^{(p)}\right)$ *pour tout $p \in \mathbb{N}$.*

Démonstration : Voir le corollaire 2.2. $\square$

2.2.3 Extension du résultat d'approximation dans le cas multi-dimensionnel

De même que dans le cas unidimensionnel, on peut, à l'aide du théorème de Muntz, généraliser le résultat d'approximation comme suit :

Proposition 2.4 *Soit $\psi \in C^\infty(\mathbb{R}, \mathbb{R})$. On suppose que :*

i) $\psi(0) \neq 0$,

ii) il existe une suite croissante d'entiers pairs non nuls $(j_k)_{k \in \mathbb{N}}$ telle que :

$$\forall k \in \mathbb{N},\ \psi^{(j_k)}(0) \neq 0 \ \text{et} \ \sum_{k \geq 0} \frac{1}{j_k} = +\infty,$$

iii) il existe une suite croissante d'entiers impairs $(\ell_k)_{n \in \mathbb{N}}$ telle que :

$$\forall k \in \mathbb{N}, \psi^{(\ell_k)}(0) \neq 0 \ \text{et} \ \sum_{k \geq 0} \frac{1}{\ell_k} = +\infty.$$

Alors, pour tout $\eta > 0$, pour tout compact K de $\mathbb{R}$, l'espace
$\left\{x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i.x),\ \alpha_i = (\alpha_i^1, \cdots, \alpha_i^d) \in (]0, \eta[)^d,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ *est dense dans* $\left(C(\mathbb{R}^d, \mathbb{R}), U_K\right)$.

Démonstration : Par souci de commodité, nous allons démontrer ce résultat sur $\mathbb{R}^2$, l'extension à $\mathbb{R}^d$ ne posant qu'un problème d'écriture.

D'après la proposition 2.2, pour tout $\eta > 0$ et pour tout compact K de $\mathbb{R}$, l'espace

$\left\{x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i x),\ \alpha_i \in]0, \eta[,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ est dense dans $(C(\mathbb{R}, \mathbb{R}), U_K)$.

En particulier, pour tout $\eta > 0$ et pour tout compact K de $\mathbb{R}$, l'espace

$\left\{x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i x),\ \alpha_i \in]0, \eta[,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ est donc dense dans $(\mathbb{R}[x], U_K)$.

Donc, pour tout $k \in \mathbb{N}$, pour tout $\eta > 0$ et pour tout compact de K_2 de $\mathbb{R}^2$, l'espace

$\left\{(x, y) \mapsto \sum_{i=0}^{n} \lambda_i \psi\left(\alpha_i(x + ky)\right),\ \alpha_i \in \left]0, \frac{\eta}{k}\right[,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ est dense dans $(\mathbb{R}[x + ky], U_{K_2})$,

puisque pour k fixé dans $\mathbb{N}$, $K_k := \{x + ky, \quad (x, y) \in K_2\}$ est un compact de $\mathbb{R}$.

Si $\alpha_i \in \left]0, \frac{\eta}{k}\right[$, alors $(\alpha_i^1, \alpha_i^2) := (\alpha_i, k\alpha_i) \in (]0, \eta[)^2$. Comme $\bigcup_{k \in \mathbb{N}^*} \mathbb{R}[x + ky] = \mathbb{R}[x, y]$, on a en fait prouvé que pour tout $\eta > 0$ et pour tout compact de K_2 de $\mathbb{R}^2$, l'espace

$\left\{(x, y) \mapsto \sum_{i=0}^{n} \lambda_i \psi\left((\alpha_i^1, \alpha_i^2).(x, y)\right),\ (\alpha_i^1, \alpha_i^2) \in (]0, \eta[)^2,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ est dense dans $(\mathbb{R}[x, y], U_{K_2})$ donc dans $(C(\mathbb{R}^2, \mathbb{R}), U_{K_2})$. □

Corollaire 2.4 *Soit μ une mesure de Borel sur $\mathbb{R}^d$. Alors, sous les hypothèses de la proposition 2.4, $\forall q \in \mathbb{R},\ 1 \leq q < +\infty$, l'espace*

$\left\{x \mapsto \sum_{i=0}^{n} \lambda_i \psi(\alpha_i.x),\quad x \in \mathbb{R}^d,\ \alpha_i := (\alpha_i^1, \cdots, \alpha_i^d) \in (]0, \eta[)^d,\quad n \in \mathbb{N},\ \lambda_i \in \mathbb{R}\right\}$ *est dense dans* $(L^q(\mu), \|.\|_q)$.

Démonstration : Comme simple corollaire de la proposition 2.4, puisque l'espace $C(\mathbb{R}^d, \mathbb{R})$ est dense dans $(L^q(\mu), \|.\|_q)$. □

2.3 Perceptron à sorties multi-dimensionnelles

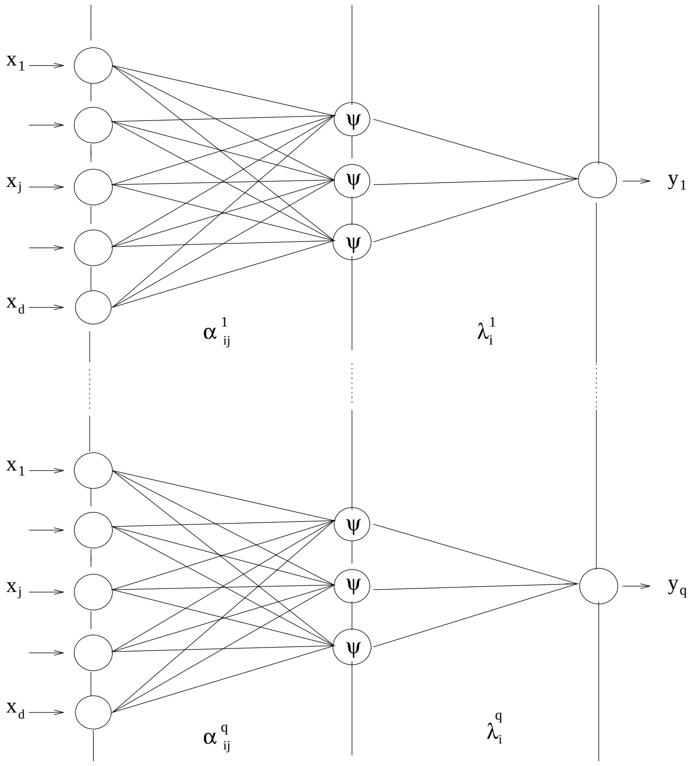
L'objectif de cette section est de montrer que tous les résultats précédents s'étendent à l'approximation de fonctions à valeurs vectorielles.

Définition 2.1 *Nous appellerons (n, q, ψ)-perceptron, $q \geq 2$, toute fonction de la forme :*

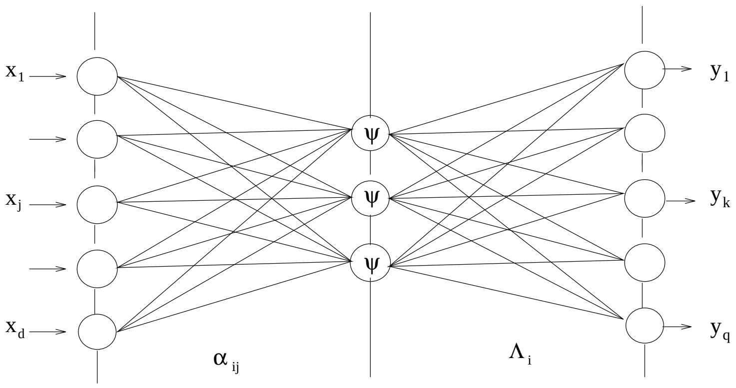
$$x \mapsto \sum_{i=0}^{n} \Lambda_i \psi(\alpha_i.x),$$

où $x \in \mathbb{R}^d$, $(\Lambda_i)_{1 \leq i \leq n} \in \mathbb{R}^q$, $(\alpha_i)_{1 \leq i \leq n} \in \mathbb{R}^d$.

Une telle fonction est en fait la sortie d'un perceptron à d unités sur la couche d'entrée, $n+1$ unités sur la couche cachée et q unités sur la couche de sortie. Une préoccupation naturelle est de savoir si les résultats obtenus pour les (n, ψ)-perceptrons s'étendent aux (n, q, ψ)-perceptrons avec $q \geq 2$. Cette question est importante en pratique. En effet, une approche naïve serait d'approximer les fonctions continues de $\mathbb{R}^d$ dans $\mathbb{R}^q$ en faisant un montage en parallèle de q perceptrons, chacun d'entre eux ayant d unités sur la couche d'entrée et une unité sur la couche de sortie, comme sur la figure suivante :



Cependant, cette démarche n'est pas satisfaisante car il n'y a aucune connection entre les q perceptrons. Notre idée est alors de montrer que nous pouvons approximer les q coordonnées de la fonction vectorielle en prenant le même vecteur (α_i) pour chaque perceptron du montage en parallèle de la figure précédente afin d'obtenir un vrai perceptron ayant d unités sur la couche d'entrée, $n+1$ unités sur la couche cachée et q unités sur la couche de sortie, avec $\Lambda_i = (\lambda_i^1, \cdots, \lambda_i^q)$, comme indiqué par la figure suivante :



Ce résultat est, en fait, une conséquence presque immédiate de la proposition 2.3.

Théorème 2.3 *Soit $\psi \in C^\infty(\mathbb{R},\mathbb{R})$ telle que $\forall k \in \mathbb{N},\ \psi^{(k)}(0) \neq 0$. Alors pour tout $\eta > 0$ et pour tout compact K de $\mathbb{R}^d$ et pour tout $n \in \mathbb{N}$, l'espace*

$$\left\{ x \mapsto \sum_{i=1}^{m} \Lambda_i \psi(\alpha_i.x), \quad x \in \mathbb{R}^d,\ \alpha_i := (\alpha_i^1, \cdots, \alpha_i^d) \in (]0,\eta[)^d, \quad m \in \mathbb{N},\ \Lambda_i \in \mathbb{R}^q \right\}$$ *est dense dans* $(C^n(\mathbb{R}^d,\mathbb{R}^q), U_{K,q}^{(n)})$.

Démonstration : On se convainc aisément à la lecture de la démonstration de la proposition 2.3 que le choix du vecteur $(\alpha_i)_{1\leq i\leq n}$ est libre sous réserve que les composantes de α_i vérifient les conditions du lemme 2.1. Il suffit alors d'utiliser les mêmes vecteurs $(\alpha_i)_{1\leq i\leq n}$ pour chacune des composantes f_j, $1 \leq j \leq q$. $\square$

Chapitre 3

Dimensionnement de la couche cachée

Ce chapitre est consacré à l'obtention de bornes pour l'erreur commise lors de l'approximation des fonctions par des $(n+1)$-perceptrons et quand n tend vers $+\infty$. La proposition 2.3 permet d'affirmer qu'un perceptron à N_n^d unités sur la couche cachée (ou n unités dans le cadre unidimensionnel) approxime (au sens de la convergence uniforme sur les compacts) les fonctions continues sur $\mathbb{R}^d$ au moins aussi bien que les polynômes de degré global n. Pour une fonction f continue donnée, l'erreur d'approximation commise par le perceptron (quand $n \to +\infty$) est donc bornée par celle commise par le polynôme de meilleure approximation. Inversement, ces résultats peuvent être compris comme des bornes du nombre d'unités nécessaires sur la couche cachée pour obtenir une erreur donnée. Nous allons donner ici deux types de résultats, selon que l'on ne veuille approximer que la fonction f elle-même ou bien la fonction f et toutes ses dérivées.

Si K est un compact de $\mathbb{R}$ (ou de $\mathbb{R}^d$), nous noterons $M_K := \sup\limits_{x \in K} |x|$ et $\delta_K := \sup\limits_{(x,y) \in K^2} |x-y|$.

3.1 Le cadre unidimensionnel

Dans cette partie, K désigne toujours un compact $\mathbb{R}$.

Théorème 3.1 *Soit $\psi \in C^\infty(\mathbb{R},\mathbb{R})$ telle que $\forall k \in \mathbb{N}$ $\psi^{(k)}(0) \neq 0$. Soit $f \in C^p(\mathbb{R},\mathbb{R})$ telle que $f^{(p)}$ est lipschitzienne. Soit ε_n une suite de nombres réels strictement positifs tels que $\lim\limits_{n\to\infty} \varepsilon_n = 0$. Alors :*

(a) il existe une suite $(\phi_n)_{n\geq 0}$ de $(n+1,\psi)$-perceptrons telle que :

$$\|f-\phi_n\|_K \leq [f^{(p)}]A_p M_K^{p+1}\frac{(1+\varepsilon_n)}{n^{p+1}} = O\left(\frac{1}{n^{p+1}}\right),$$

où A_p est une constante ne dépendant que de p et de la norme choisie sur $\mathbb{R}$.

(b) il existe une suite $(\Phi_n)_{n\geq 0}$ de $(n+1,\psi)$-perceptrons telle que :

$$\forall k \in \{0,\cdots,p\}, \quad \left\|f^{(k)} - \Phi_n^{(k)}\right\|_K \leq [f^{(p)}]A_0 M_K \max\left(1,(\delta_K)^p\right)\frac{(1+\varepsilon_n)}{n} = O\left(\frac{1}{n}\right).$$

Démonstration : (a) Pour tout $n \in \mathbb{N}$, le polynôme de meilleure approximation uniforme de degré n de f sur K, soit $P_n(f)$ vérifie :

$$\|f - P_n(f)\|_K \leq [f^{(p)}] A_p M_K^{p+1} \frac{1}{n^{p+1}},$$

où A_p ne dèpend que de p et de la norme choisie sur $\mathbb{R}$ (Lorentz, 1986). Le résultat est alors assuré par la proposition 2.1 car nous pouvons choisir ϕ_n telle que :

$$\forall\, n \in \mathbb{N}, \quad \|\phi_n - P_n(f)\|_K \leq [f^{(p)}] A_p M_K^{p+1} \frac{\varepsilon_n}{n^{p+1}}.$$

(b) Il existe un polynôme $Q_n(f)$ tel que :

$$\forall k \in \{0, \cdots, p\}, \quad \left\| f^{(k)} - Q_n^{(k)}(f) \right\|_K \leq [f^{(p)}] A_0 M_K \max\left(1, (\delta_K)^p\right) \frac{1}{n}.$$

La proposition 2.1 fournit encore le résultat. □

Remarque importante : En fait, tous les cas intermédiaires sont possibles. C'est-à-dire que si nous ne voulons approximer que les ℓ ($\ell \leq p$) premières dérivées de $f \in C^p(\mathbb{R}, \mathbb{R})$ alors que $f^{(p)}$ est lipschitzienne, nous obtenons un ordre d'approximation en $O\left(\frac{1}{n^{p+1-\ell}}\right)$ pour les ℓ premières dérivées. Il ne sera pas possible de donner une vitesse d'approximation que pour un certain nombre de dérivées dans le cadre multidimensionnel. En effet, notre résultat s'appuie sur les polynômes de Bernstein, et ceux-ci approximent toutes les dérivées de f.

Remarques : • Les polynômes de meilleure approximation $P_n(f)$ ne sont généralement pas explicitables. En revanche, les polynômes de Tchebichev $T_n(f)$ le sont et vérifient :

$$\|T_n(f) - f\|_K \leq (3 + \ln(n))\, \|P_n(f) - f\|_K \quad \text{(Lorentz, 1986).}$$

Nous pouvons donc construire explicitement une suite ϕ_n de $(n+1, \psi)$-perceptrons vérifiant :

$$\|f - \phi_n\|_K \leq [f^{(p)}] A_p M_K^{p+1} \frac{(3 + \ln(n))(1 + \varepsilon_n)}{n^{p+1}},$$

où $(\varepsilon_n)_{n \geq 0}$ est une suite de réels strictement positifs telle que $\lim\limits_{n \to +\infty} \varepsilon_n = 0$.

• De même, les $Q_n(f)$ ne sont pas explicites, mais on peut expliciter une suite de polynômes $R_n(f)$ de degré n qui le soient telle que :

$$\forall k \in \{0, \cdots, p\}, \quad \left\| f^{(k)} - R_n^{(k)}(f) \right\|_K \leq [f^{(p)}] A_0 M_K \max\left(1, (\delta_K)^p\right) \frac{(3 + \ln(n))}{n},$$

Il est alors possible de construire une suite Φ_n telle que :

$$\forall k \in \{0, \cdots, p\}, \quad \left\| f^{(k)} - \Phi_n^{(k)} \right\|_K \leq [f^{(p)}] A_0 M_K \max\left(1, (\delta_K)^p\right) \frac{(3 + \ln(n))(1 + \varepsilon_n)}{n},$$

où $\lim\limits_{n \to +\infty} \varepsilon_n = 0$.

3.2 Le cadre multidimensionnel

Dans cette partie, K désigne un compact de $\mathbb{R}^d$, $d \geq 2$.

Théorème 3.2 *Soit $\psi \in C^\infty(\mathbb{R}, \mathbb{R})$ telle que $\forall k \in \mathbb{N}$ $\psi^{(k)}(0) \neq 0$. Soit $f \in C^p(\mathbb{R}^d, \mathbb{R})$ telle que pour tout $k_1, \cdots, k_d$, $0 \leq k := k_1 + \cdots + k_d \leq p$, $\dfrac{\partial^k f}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}}$ est lipschitzienne avec :*

$$\left[\frac{\partial^k f}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}}\right] \leq \rho.$$

Soit ε_n une suite de nombres réels strictement positifs telle que $\lim_{n\to\infty} \varepsilon_n = 0$. Alors:

(a) il existe une constante $A_{d,p}$ ne dépendant que de d, p et de la norme choisie sur $\mathbb{R}^d$, et une suite $(\phi_n)_{n\in\mathbb{N}}$ de $(n+1, \psi)$-perceptrons telles que :

$$\|f - \phi_n\|_K \leq \rho A_{d,p} M_K^{p+1} \frac{(1+\varepsilon_n)}{n^{\frac{p+1}{d}}} = O\left(\frac{1}{n^{\frac{p+1}{d}}}\right).$$

(b) il existe une suite $(\Phi_n)_{n\in\mathbb{N}}$ de $(n+1, \psi)$-perceptrons, une constante $a_{d,p}$ ne dépendant que de d, de p et de la norme choisie sur $\mathbb{R}^d$ et une constante C_K^d ne dépendant que de d, du compact K et de la norme choisie sur $\mathbb{R}^d$ telles que pour tout $0 \leq k := k_1 + \cdots + k_d \leq p$:

$$\left\|\frac{\partial^k f}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} - \frac{\partial^k \Phi_n}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}}\right\|_K \leq \rho a_{d,p} C_K^d \frac{(1+\varepsilon_n)}{n^{\frac{1}{2d}}} = O\left(\frac{1}{n^{\frac{1}{2d}}}\right).$$

Démonstration : (a) Pour tout $j \in \mathbb{N}$, le polynôme de meilleure approximation uniforme de f sur K de degré global j, $P_j(f)$ vérifie :

$$\|f - P_j(f)\|_K \leq \rho B_{d,p} M_K^{p+1} \frac{1}{j^{p+1}},$$

où $B_{d,p}$ ne dépend que de d, de p et de la norme choisie sur $\mathbb{R}^d$ (Lorentz, 1986, page 89). Soit un entier n, il existe j_n tel que :

$$N_{j_n}^d \leq n < N_{j_n+1}^d.$$

Nous avons donc :

$$\|f - P_{j_n}(f)\|_K \leq \rho B_{d,p} M_K^{p+1} \frac{1}{{j_n}^{p+1}}.$$

D'après la proposition 2.3, nous savons qu'il est possible de construire une suite $\left(\phi_{N_{j_n}^d}\right)_{n\in\mathbb{N}}$ de $\left(N_{j_n}^d, \psi\right)$-perceptrons telle que :

$$\left\|\phi_{N_{j_n}^d} - P_{j_n}(f)\right\|_K \leq \rho B_{d,p} M_K^{p+1} \frac{\varepsilon_n}{{j_n}^{p+1}}.$$

En sommant ces deux inégalités, nous obtenons :

$$\left\|\phi_{N_{j_n}^d} - f\right\|_K \leq \rho B_{d,p} M_K^{p+1} \frac{1+\varepsilon_n}{{j_n}^{p+1}}. \tag{3.1}$$

Il est facile de montrer par récurrence sur la dimension d de l'espace de départ que :

$$\forall d \geq 1,\ N_j^d \underset{j \to +\infty}{\sim} \frac{j^d}{d}.$$

D'où, pour tout $d \geq 1$, $N_{j_n}^d \underset{n \to +\infty}{\sim} \dfrac{{j_n}^d}{d!}$. Mais $N_{j_n}^d \leq n < N_{j_n+1}^d$, nous avons donc également $n \underset{n \to +\infty}{\sim} \dfrac{{j_n}^d}{d!}$. Par suite, $\dfrac{1}{j_n} \underset{n \to +\infty}{\sim} \dfrac{1}{(d!)^{\frac{1}{d}} n^{\frac{1}{d}}}$. Nous pouvons donc trouver une constante c_d ne dépendant que de d telle que :

$$\frac{1}{j_n} \leq \frac{c_d}{(d!)^{\frac{1}{d}} n^{\frac{1}{d}}}. \tag{3.2}$$

Reportant cette dernière inégalité dans (3.1), nous trouvons finalement :

$$\left\| \phi_{N_{j_n}^d} - f \right\|_K \leq \rho M_K^{p+1} \underbrace{\frac{c_d^{p+1} B_{d,p}}{(d!)^{\frac{p+1}{d}}}}_{A_{d,p}} \frac{1+\varepsilon_n}{n^{\frac{p+1}{d}}},$$

qui est le résultat désiré car $N_{j_n}^d \leq n$.

(b) De nouveau, pour tout entier n, nous pouvons trouver un entier j_n tel que :

$$N_{j_n}^d \leq n < N_{j_n+1}^d.$$

Par le théorème A.1 de l' Annexe, nous savons que sur S_d, les polynômes de Bernstein $B_{j_n}(f)$vérifient :

$$\forall k = k_1 + \cdots + k_d \leq p,\ \left\| \frac{\partial^k f}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} - \frac{\partial^k B_{j_n}(f)}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} \right\|_{S_d} \leq \rho\ \mu\ \alpha_{d,p} \frac{1}{\sqrt{j_n}},$$

où $\mu = 1$ si l'on choisit $|x| := \max\limits_{1 \leq i \leq d} |x_i|$ comme norme sur $\mathbb{R}^d$ (voir l' Annexe).

D'aprés la proposition 2.3, nous pouvons alors construire une suite $\left(\Phi_{N_{j_n}^d}\right)_{n \in \mathbb{N}}$ de $\left(N_{j_n}^d, \psi\right)$-perceptrons telle que :

$$\forall k = k_1 + \cdots + k_d \leq p,\ \left\| \frac{\partial^k \Phi_{N_{j_n}^d}}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} - \frac{\partial^k B_{j_n}(f)}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} \right\|_{S_d} \leq \rho\ \mu\ \alpha_{d,p} \frac{\varepsilon_n}{\sqrt{j_n}}.$$

Nous avons donc :

$$\forall k = k_1 + \cdots + k_d \leq p,\ \left\| \frac{\partial^k \Phi_{N_{j_n}^d}}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} - \frac{\partial^k f}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} \right\|_{S_d} \leq \rho\ \mu\ \alpha_{d,p} \frac{1+\varepsilon_n}{\sqrt{j_n}}.$$

Lors du point (a), nous avons vu (inegalité (3.2)):

$$\frac{1}{j_n} \leq \frac{c_d}{(d!)^{\frac{1}{d}} n^{\frac{1}{d}}},$$

ce qui reporté plus haut donne :

$$\forall k = k_1 + \cdots + k_d \leq p, \quad \left\| \frac{\partial^k \Phi_{N^d_{j_n}}}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} - \frac{\partial^k f}{\partial {x_1}^{k_1} \cdots \partial {x_d}^{k_d}} \right\|_{S_d} \leq \rho \underbrace{\frac{\mu \sqrt{c_d} \alpha_{d,p}}{(d!)^{\frac{1}{2d}}}}_{a_{d,p}} \frac{1 + \varepsilon_n}{n^{\frac{1}{2d}}},$$

qui est le résultat voulu sur S_d car $N^d_{j_n} \leq n$.

Pour obtenir le résultat sur tout compact K, il suffit alors de faire une homothétie-translation de S_d afin que ce nouveau compact contienne K. Ce faisant, on introduit une nouvelle constante C^d_K ne dépendant que du rapport (diamètre de K/diamètre de S_d). $\square$

Remarque : Comme nous l'avions vu lors de la partie précédente, on n'obtient pas ici de borne d'approximation que pour un certain nombre de dérivées. De plus, l'erreur d'approximation commise est pire que celle en $O\left(\frac{1}{n^{\frac{1}{d}}}\right)$ que nous pouvions attendre au vu des résultats en dimension 1.

Remarque sur les perceptrons à sorties multidimensionnelles : En ce qui concerne le dimensionnement de la couche cachée, tous les résultats précédents s'étendent au cas de l'approximation des fonctions vectorielles, l'erreur d'approximation ne dépendant que de la dimension de l'espace des entrées.

Annexe A

Convergence des polynômes multinomiaux de Bernstein

Nous montrons ici que les polynômes multinomiaux de Bernstein sur le simplexe S_d de $\mathbb{R}^d$ peuvent approximer une fonction de classe C^p avec toutes ses dérivées et avec une erreur en $\dfrac{1}{\sqrt{j}}$, où j est le degré global du polynôme, mais le nombre de coefficients est N_j^d qui dépend de la dimension d de l'espace des entrées.

Nous allons montrer ce résultat dans $\mathbb{R}^2$ mais la démarche est exactement la même en dimension supérieure.

Ce résultat est probablement déjà connu des spécialistes de l'approximation polynomiale mais nous n'en avons trouvé aucune référence. Nous donnons donc notre propre démonstration.

Par souci de simplification, nous choisissons $|(x,y)| := \max(|x|,|y|)$ comme norme sur $\mathbb{R}^2$. Si f est lipschitzienne, son coefficient de Lipschitz est donc défini par :

$$[f] := \sup_{(x_1,y_1)\neq(x_2,y_2)} \frac{|f(x_1,y_1)-f(x_2,y_2)|}{\max(|x_1-x_2|,|y_1-y_2|)}.$$

Définition A.1 *Les polynômes de Bernstein associés à une fonction ϕ bornée de 2 variables, sont définis sur le simplexe $S_2 := \{(x,y)/\ x\geq 0,\ y\geq 0$ et $x+y\leq 1\}$ par:*

$$\forall (x,y)\in S_2,\quad B_j(\phi)(x,y) := \sum_{|k|=j} C_j^k x^{k_1} y^{k_2} \left(1-(x+y)\right)^{k_3} \phi\left(\frac{k_1}{j},\frac{k_2}{j}\right),$$

avec $k := (k_1,k_2,k_3)$, $|k| = k_1+k_2+k_3$ et $C_j^k = \dfrac{j!}{k_1!k_2!k_3!}$.

Démontrons tout d'abord un premier lemme :

Lemme A.1 *(a) Pour toute fonction ϕ bornée sur S_2 :*

$$\forall j\in\mathbb{N}^*,\ \|B_j(\phi)\|_{S_2} \leq \|\phi\|_{S_2}. \tag{A.1}$$

(b) Pour toute fonction lipschitzienne ϕ sur S_2 :

$$\forall j\in\mathbb{N}^*,\ \|B_j(\phi)-\phi\|_{S_2} \leq \frac{[\phi]}{\sqrt{j}}. \tag{A.2}$$

Démonstration: (a) Comme $\displaystyle\sum_{|k|=j} C_j^k x^{k_1} y^{k_2}\left(1-(x+y)\right)^{k_3} = 1$, l'inégalité (A.1) est clairement vraie.

Maintenant si ϕ est lipschitzienne , on a :

$$\begin{aligned}
\|B_j(\phi)-\phi\|_{S_2} &= \left\| \sum_{|k|=j} C_j^k x^{k_1} y^{k_2} \left(1-(x+y)\right)^{k_3} \left(\phi\left(\frac{k_1}{j},\frac{k_2}{j}\right)-\phi(x,y)\right)\right\|_{S_2} \\
&\leq [\phi] \left\| \sum_{|k|=j} C_j^k x^{k_1} y^{k_2} \left(1-(x+y)\right)^{k_3} \max\left(\left|x-\frac{k_1}{j}\right|,\left|y-\frac{k_2}{j}\right|\right)\right\|_{S_2} \\
&\leq [\phi] \left\| \sum_{|k|=j} C_j^k x^{k_1} y^{k_2} \left(1-(x+y)\right)^{k_3} \left(\left|x-\frac{k_1}{j}\right|+\left|y-\frac{k_2}{j}\right|\right)\right\|_{S_2}.
\end{aligned}$$

On remarque alors que :

$$\sum_{|k|=j} C_j^k x^{k_1} y^{k_2} \left(1-(x+y)\right)^{k_3} \left|x-\frac{k_1}{j}\right| = \mathbb{E}\left|\frac{X_1}{j}-x\right|, \tag{A.3}$$

où (X_1,X_2) suit une loi multinomiale $\mathcal{M}(j,x,y)$. X_1 suit alors une binomiale $\mathcal{B}(j,x)$, d'où :

$$\mathbb{E}\left|\frac{X_1}{j}-x\right| \leq \frac{1}{\sqrt{j}}\sigma(X_1) = \sqrt{\frac{x(1-x)}{j}} \leq \frac{1}{2\sqrt{j}}, \quad \text{car } 0 \leq x \leq 1.$$

De la même manière :

$$\sum_{|k|=j} C_j^k x^{k_1} y^{k_2} \left(1-(x+y)\right)^{k_3} \left|y-\frac{k_2}{j}\right| \leq \frac{1}{2\sqrt{j}}.$$

Finalement on trouve : $\|B_j(\phi)-\phi\|_{S_2} \leq [\phi]\times\frac{1}{\sqrt{j}}$. □

Remarque : Il est facile de voir qu'en dimension d, le résultat reste vrai avec une borne $[\phi]\times\frac{d}{2\sqrt{j}}$.

L'idée de la démonstration est de voir les dérivées des polynômes de Bernstein comme de nouveaux polynômes de Bernstein de fonctions issues de f. Pour fixer les idées, nous allons d'abord démontrer le résultat pour les deux dérivées partielles premières.

Proposition A.1 (premier ordre) *Soit $f \in C^1(\mathbb{R}^2,\mathbb{R})$ ayant des dérivées partielles lipschitziennes.*

Alors : $$\begin{cases} \left\|\frac{\partial B_j(f)}{\partial x}-\frac{\partial f}{\partial x}\right\|_{S_2} = O\left(\frac{1}{\sqrt{j}}\right) \\ \left\|\frac{\partial B_j(f)}{\partial y}-\frac{\partial f}{\partial y}\right\|_{S_2} = O\left(\frac{1}{\sqrt{j}}\right) \end{cases}$$

Démonstration : Un petit calcul algébrique montre que :

$$\frac{\partial B_j(f)}{\partial x} = B_{j-1}\left(G_j(f)\right), \tag{A.4}$$

où $G_j(f) := \dfrac{j}{j-1}.\dfrac{f\left(\frac{j-1}{j}\left(x+\frac{1}{j-1}\right),\frac{j-1}{j}y\right)-f\left(\frac{j-1}{j}x,\frac{j-1}{j}y\right)}{\frac{1}{j-1}}$.

On écrit alors :

$$\begin{aligned}
\left\|\frac{\partial B_j(f)}{\partial x}-\frac{\partial f}{\partial x}\right\|_{S_2} &= \left\|B_{j-1}\left(G_j(f)\right)-\frac{\partial f}{\partial x}\right\|_{S_2} \\
&\leq \left\|B_{j-1}\left(G_j(f)-\frac{\partial f}{\partial x}\right)\right\|_{S_2} + \left\|B_{j-1}\left(\frac{\partial f}{\partial x}\right)-\frac{\partial f}{\partial x}\right\|_{S_2}.
\end{aligned} \tag{A.5}$$

La démonstration consiste à démontrer que les termes de droite de l'inégalité (A.5) sont en $O\left(\frac{1}{\sqrt{j}}\right)$ quand $j \to +\infty$.

Pour le premier terme, l'inégalité (A.1) donne :

$$\left\| B_{j-1}\left(G_j(f) - \frac{\partial f}{\partial x}\right)\right\|_{S_2} \leq \left\| G_j(f) - \frac{\partial f}{\partial x}\right\|_{S_2}.$$

La formule des accroissements finis permet d'écrire :

$$G_j(f) = \frac{\partial f}{\partial x}\left(\frac{j-1}{j}\varepsilon_j(x), y\right) \text{ où } \varepsilon_j(x) \in \left[x, x + \frac{1}{j-1}\right].$$

Par suite :

$$\begin{aligned} G_j(f) - \frac{\partial f}{\partial x} &= \frac{\partial f}{\partial x}\left(\frac{j-1}{j}\varepsilon_j(x), y\right) - \frac{\partial f}{\partial x}(x,y) \\ &\leq \left[\frac{\partial f}{\partial x}\right]\left|\frac{j-1}{j}\varepsilon_j(x) - x\right| \\ &\leq \left[\frac{\partial f}{\partial x}\right] \times \frac{1}{j}. \end{aligned}$$

Pour le second terme majorant de (A.5), l'inegalité (A.2) donne :

$$\left\| B_{j-1}\left(\frac{\partial f}{\partial x}\right) - \frac{\partial f}{\partial x}\right\|_{S_2} \leq \left[\frac{\partial f}{\partial x}\right] \times \frac{1}{\sqrt{j-1}}.$$

Finalement :

$$\left\| \frac{\partial B_j}{\partial x} - \frac{\partial f}{\partial x}\right\|_{S_2} = O\left(\frac{1}{\sqrt{j-1}}\right) = O\left(\frac{1}{\sqrt{j}}\right).$$

On a la même borne en y en utilisant un opérateur H_j de définition analogue à celle de G_j. □

Avant de démontrer le résultat général, donnons quelques inégalités qui nous seront utiles par la suite :

$\forall \phi \in C^1(\mathbb{R}^2, \mathbb{R})$, et $\forall n \in \mathbb{N}^*$:

$$\left\{ \begin{array}{l} \|G_n(\phi)\|_{S_2} \leq \left\|\frac{\partial \phi}{\partial x}\right\|_{S_2} \\ \|H_n(\phi)\|_{S_2} \leq \left\|\frac{\partial \phi}{\partial y}\right\|_{S_2} \end{array} \right. . \tag{A.6}$$

$\forall \phi \in C^1(\mathbb{R}^2, \mathbb{R})$, et $\forall n \in \mathbb{N}^*$:

$$\left\{ \begin{array}{l} \frac{\partial G_n(\phi)}{\partial x} = \frac{n-1}{n} G_n\left(\frac{\partial \phi}{\partial x}\right), \ \frac{\partial H_n(\phi)}{\partial x} = \frac{n-1}{n} H_n\left(\frac{\partial \phi}{\partial x}\right) \\ \frac{\partial G_n(\phi)}{\partial y} = \frac{n-1}{n} G_n\left(\frac{\partial \phi}{\partial y}\right), \ \frac{\partial H_n(\phi)}{\partial y} = \frac{n-1}{n} H_n\left(\frac{\partial \phi}{\partial y}\right) \end{array} \right. . \tag{A.7}$$

$\forall \phi \in C^1(\mathbb{R}^2, \mathbb{R})$ avec des dérivées partielles lipschitziennes, et $\forall n \in \mathbb{N}^*$:

$$\left\{ \begin{array}{l} \left\|G_n(\phi) - \frac{\partial \phi}{\partial x}\right\|_{S_2} \leq \left[\frac{\partial \phi}{\partial x}\right] \times \frac{1}{n} \\ \left\|H_n(\phi) - \frac{\partial \phi}{\partial y}\right\|_{S_2} \leq \left[\frac{\partial \phi}{\partial y}\right] \times \frac{1}{n} \end{array} \right. . \tag{A.8}$$

Proposition A.2 *Soit $p \geq 2$ et soit $f \in C^p(\mathbb{R}^2, \mathbb{R})$. Supposons que f et toutes ses dérivées partielles $\dfrac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}}$ sont lipschitziennes, $1 \leq k_1 + k_2 \leq p$. Alors :*

$$\forall\, k_1, k_2,\ 0 \leq k_1 + k_2 \leq p,\ \left\| \frac{\partial^{k_1+k_2} B_j(f)}{\partial x^{k_1} \partial y^{k_2}} - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2} = O\left(\frac{1}{\sqrt{j}}\right).$$

Démonstration: Une récurrence évidente fondée sur (A.4) permet d'établir que :

$$\frac{\partial^{k_1+k_2} B_j(f)}{\partial x^{k_1} \partial y^{k_2}} = B_{j-(k_1+k_2-1)}\left(M^j_{k_1,k_2}(f)\right),$$

où l'opérateur M est défini comme suit :

$$M^j_{k_1,k_2}(f) := G_{j-(k_1+k_2-1)} \circ \cdots \circ G_{j-k_2} \circ H_{j-(k_2-1)} \circ \cdots \circ H_j(f).$$

Par inégalité triangulaire, on obtient ainsi :

$$\begin{aligned}\left\| \frac{\partial^{k_1+k_2} B_j(f)}{\partial x^{k_1} \partial y_{k_2}} - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2} \quad \leq \quad & \underbrace{\left\| B_{j-(k_1+k_2-1)}\left(M^j_{k_1,k_2}(f) - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right) \right\|_{S_2}}_{(a)} \\ + \quad & \underbrace{\left\| B_{j-(k_1+k_2-1)}\left(\frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right) - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2}}_{(b)}.\end{aligned}$$

1) Majoration de (a)

L'inégalité (A.1) donne :

$$(a) \leq \left\| M^j_{k_1,k_2}(f) - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2}.$$

La démarche consiste alors à démontrer que :

$$\left\| M^j_{k_1,k_2}(f) - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2} = O\left(\frac{1}{j}\right).$$

On va procéder par récurrence.

$$(H_k) \text{ est vrai} \equiv \forall\ k_1, k_2, \text{ avec } k_1 + k_2 = k \leq p,\ \left\| M^j_{k_1,k_2}(f) - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2} = O\left(\frac{1}{j}\right).$$

Pour $k_1 + k_2 = 1$ le résultat découle des inégalités (A.8). Supposons que (H_{k-1}) est vraie $k \geq 2$.

Soient k_1, k_2 tels que $k_1 + k_2 = k \geq 2$. On peut supposer sans perte de généralité que $k_1 \geq 1$. Comme $M^j_{k_1,k_2}(f) = G_{j-(k_1+k_2-1)} \circ M^j_{k_1-1,k_2}(f)$, on peut écrire :

$$\begin{aligned}\left\| M^j_{k_1,k_2}(f) - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2} \quad \leq \quad & \underbrace{\left\| G_{j-(k_1+k_2-1)}\left(M^j_{k_1-1,k_2}(f) - \frac{\partial^{k_1+k_2-1} f}{\partial x^{k_1-1} \partial y^{k_2}} \right) \right\|_{S_2}}_{(\alpha)} \\ + \quad & \underbrace{\left\| G_{j-(k_1+k_2-1)}\left(\frac{\partial^{k_1+k_2-1} f}{\partial x^{k_1-1} \partial y^{k_2}} \right) - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2}}_{(\beta)}.\end{aligned} \tag{A.9}$$

L'inégalité (A.6) implique :

$$(\alpha) \leq \left\| \frac{\partial}{\partial x}\left(M^j_{k_1-1,k_2}(f) \right) - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2},$$

et les équations (A.7) donnent :

$$\frac{\partial}{\partial x}\left(M^j_{k_1-1,k_2}(f)\right) = \frac{(j-(k_1-1+k_2))}{j} M^j_{k_1-1,k_2}\left(\frac{\partial f}{\partial x}\right).$$

Nous avons donc :

$$(\alpha) \leq \frac{k_1+k_2-1}{j}\left\|M^j_{k_1-1,k_2}\left(\frac{\partial f}{\partial x}\right)\right\|_{S_2} + \left\|M^j_{k_1-1,k_2}\left(\frac{\partial f}{\partial x}\right) - \frac{\partial^{k_1+k_2-1}}{\partial x^{k_1-1}\partial y^{k_2}}\left(\frac{\partial f}{\partial x}\right)\right\|_{S_2}.$$

Une récurrence basée sur l'equation (A.6) montre que :

$$\left\|M^j_{k_1-1,k_2}(f)\right\|_{S_2} \leq \left\|\frac{\partial^{k_1+k_2-1} f}{\partial x^{k_1-1}\partial y^{k_2}}\right\|_{S_2},$$

ce qui, appliqué à $\dfrac{\partial f}{\partial x}$, donne :

$$\left\|M^j_{k_1-1,k_2}\left(\frac{\partial f}{\partial x}\right)\right\|_{S_2} \leq \left\|\frac{\partial^{k_1+k_2} f}{\partial x^{k_1}\partial y^{k_2}}\right\|_{S_2}.$$

D' autre part, l'hypothèse de récurrence (H_{k-1}) appliquée à $\dfrac{\partial f}{\partial x}$ permet d'affirmer que :

$$\left\|M^j_{k_1-1,k_2}\left(\frac{\partial f}{\partial x}\right) - \frac{\partial^{k_1+k_2-1}}{\partial x^{k_1-1}\partial y^{k_2}}\left(\frac{\partial f}{\partial x}\right)\right\|_{S_2} = O\left(\frac{1}{j}\right),$$

ce qui donne pour le premier terme de l'inégalité (A.9) :

$$(\alpha) \leq \frac{k_1+k_2-1}{j}\left\|\frac{\partial^{k_1+k_2-1} f}{\partial x^{k_1-1}\partial y^{k_2}}\right\|_{S_2} + O\left(\frac{1}{j}\right) = O\left(\frac{1}{j}\right).$$

D'autre part, (A.8) implique aisément que :

$$(\beta) \leq \left[\frac{\partial^{k_1+k_2} f}{\partial x^{k_1}\partial y^{k_2}}\right] \times \frac{1}{j} = O\left(\frac{1}{j}\right).$$

Finalement $(a) = (\alpha) + (\beta) = O\left(\dfrac{1}{j}\right)$.

2) Majoration de (b)

Elle se fait aisément en remarquant que (A.2) implique:

$$(b) \leq \left[\frac{\partial^{k_1+k_2} f}{\partial x^{k_1}\partial y^{k_2}}\right] \times \frac{1}{\sqrt{j-(k_1+k_2-1)}} = O\left(\frac{1}{\sqrt{j}}\right).$$

En combinant tout ce qui précède, on arrive à :

$$\left\|\frac{\partial^{k_1+k_2} B_j(f)}{\partial x^{k_1}\partial y^{k_2}} - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1}\partial y^{k_2}}\right\|_{S_2} \leq (a) + (b) = O\left(\frac{1}{j}\right) + O\left(\frac{1}{\sqrt{j}}\right) = O\left(\frac{1}{\sqrt{j}}\right),$$

ce qui achève la preuve. □

Nous sommes maintenant en mesure de démontrer le résultat final en utilisant la proposition A.2 et le théorème de Banach-Steinhaus.

Théorème A.1 *Soient $p \in \mathbb{N}^*$ et $f \in C^p(\mathbb{R}^2, \mathbb{R})$. Supposons que f et toutes ses dérivées partielles sont ρ-lipschitziennes, i.e. :*

$$\forall\ k_1, k_2 \in \mathbb{N},\ 0 \leq k_1 + k_2 \leq p \Longrightarrow \left[\frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}}\right] \leq \rho \quad \left(\textit{avec par convention } \frac{\partial^0 f}{\partial x^0 \partial y^0} = f\right).$$

Alors il existe une constante $\alpha_{2,p}$ ne dépendant que de p telle que :

$$\forall\ k_1, k_2 \in \mathbb{N},\ 0 \leq k_1 + k_2 \leq p,\ \left\| \frac{\partial^{k_1+k_2} B_j(f)}{\partial x^{k_1} \partial y^{k_2}} - \frac{\partial^{k_1+k_2} f}{\partial x^{k_1} \partial y^{k_2}} \right\|_{S_2} \leq \rho \alpha_{2,p} \frac{1}{\sqrt{j}},$$

Démonstration: Soient k_1 et k_2 deux entiers vérifiant $0 \leq k_1 + k_2 \leq p$.

Soit $C^p_{Lip,0}$ l'espace des fonctions f de classe C^p dont toutes les dérivées partielles sont lipschitziennes et telles que $f(0,0) = 0$. On munit $C^p_{Lip,0}$ de la topologie induite par la norme :

$$\forall g \in C^p_{Lip,0},\ ||g||_{C^p_{Lip,0}} := max\left(\sup_{0 \leq i_1+i_2 \leq p} \left[\frac{\partial^{i_1+i_2} g}{\partial x^{i_1} \partial y^{i_2}}\right], \sup_{1 \leq i_1+i_2 \leq p} \left| \frac{\partial^{i_1+i_2} g}{\partial x^{i_1} \partial y^{i_2}}(0,0) \right| \right).$$

$\left(C^p_{Lip,0}, ||.||_{C^p_{Lip,0}}\right)$ est clairement un espace de Banach.

Définissons pour tout j, les applications linéaires de $C^p_{Lip,0}$ dans l'espace des fonctions continues :

$$N_j^{k_1,k_2}(.) := \sqrt{j} \left(\frac{\partial^{k_1+k_2} B_j(.)}{\partial x^{k_1} \partial y^{k_2}} - \frac{\partial^{k_1+k_2} .}{\partial x^{k_1} \partial y^{k_2}} \right).$$

Il est clair que, pour tout j, $N_j^{k_1,k_2}(.)$ est bornée sur la sphère unité de $C^p_{Lip,0}$. Par conséquent, pour tout j, $N_j^{k_1,k_2}(.)$ est continue.

En outre, d'après la proposition A.2, $\sup_j ||N_j^{k_1,k_2}(f)||_{S_2} < +\infty$ pour tout $f \in C^p_{Lip,0}$.

Le théorème de Banach-Steinhaus (Rudin, 1973) implique alors que pour tous k_1, k_2 tels que $0 \leq k_1 + k_2 \leq p$:

$$\exists\ \alpha_2^{k_1,k_2} /\ \forall j,\ \forall f \in C^p_{Lip,0},\ ||N_j^{k_1,k_2}(f)||_{S_2} \leq \alpha_2^{k_1,k_2} ||f||_{C^p_{Lip,0}}.$$

Posant $\alpha_{2,p} := \sup_{0 \leq k_1+k_2 \leq p} \alpha_2^{k_1,k_2}$, on obtient :

$$\forall f \in C^p_{Lip,0},\ \forall j \text{ et } \forall\ k_1, k_2,\ 0 \leq k_1 + k_2 \leq p,\ ||N_j^{k_1,k_2}(f)||_{S_2} \leq \alpha_{2,p} ||f||_{C^p_{Lip,0}}. \tag{A.10}$$

Pour finir, supposons que f est de classse C^p avec toutes ses dérivées partielles lipschitziennes, c'est-à-dire que l'on peut trouver ρ vérifiant :

$$\forall\ i_1, i_2 \text{ avec } 0 \leq i_1 + i_2 \leq p,\ \left[\frac{\partial^{i_1+i_2} f}{\partial x^{i_1} \partial y^{i_2}}\right] \leq \rho.$$

Pour $1 \leq i_1 + i_2 \leq p$, on a nécessairement soit $i_1 \geq 1$ soit $i_2 \geq 1$. On peut supposer sans perte de généralité que $i_1 \geq 1$.

$\dfrac{\partial^{i_1+i_2-1} f}{\partial x^{i_1-1} \partial y^{i_2}}$ étant ρ-lipschitzienne, il vient :

$$\frac{\left| \frac{\partial^{i_1+i_2-1} f}{\partial x^{i_1-1} \partial y^{i_2}}(x,0) - \frac{\partial^{i_1+i_2-1} f}{\partial x^{i_1-1} \partial y^{i_2}}(0,0) \right|}{|x|} \leq \rho.$$

En prenant la limite de cette inégalité quand x tend vers 0 on obtient :

$$\forall\ i_1, i_2 \text{ avec } 1 \leq i_1 + i_2 \leq p,\ \left| \frac{\partial^{i_1+i_2} f}{\partial x^{i_1} \partial y^{i_2}}(0,0) \right| \leq \rho.$$

Cela prouve que $f - f(0,0) \in C^p_{Lip,0}$, avec $\|f - f(0,0)\|_{C^p_{Lip,0}} \leq \rho$.

L'inégalité (A.10) donne ainsi :

$$\forall j \text{ et } \forall\ k_1, k_2 \text{ avec } 0 \leq k_1 + k_2 \leq p,\ \|N_j^{k_1,k_2}(f - f(0,0))\|_{S_2} \leq \rho\alpha_{2,p}.$$

Remarquant que $N_j^{k_1,k_2}(f(0,0)) = 0$, nous obtenons le résultat final. □

Remarque : Il est vraisemblable que l'on peut établir l'existence de la constante $a_{2,p}$ par un calcul direct, sans recourir au théorème de Banach-Steinhaus. Il reste que la valeur de la constante est de peu d'intérêt.

Bibliographie de la première partie

Attali, J.G., Pagès, G. (1997) Approximations of Functions by a Multilayer Perceptron: a New Approach. *Neural Networks*, 10, pp. 1069-1081.

Barron, A.R. (1993). Universal Approximation Bounds for Superposition of a Sigmoidal Function. *IEEE Transactions on Information Theory*, 39, pp.930-945.

Chui, C.K., Li, X., Mhaskar, H.N. (1995). Limitations of the approximation capabilities of neural networks with one hidden layer. Preprint available from the third author at Dept. of Math., California State University, Los Angeles, California, 90032, USA.

DeVore, R.A., Howard, R. Micchelli, C.A. (1994). Optimal nonlinear approximation. *Manuscripta math.*, 63, pp.469-478.

Ellacott, S.W. (1994). In A. Iserles (Ed.), *Acta Numerica*, Cambride University Press, pp.145-202.

S.N. Ethier, T.G. Kurtz, *Markov Processes, Characterization and Convergence*, Wiley series in Probability and Mathematical Statistics, New-York, 1986, 534p.

Hornik, K. (1993). Some New Results on Neural Network Approximation. *Neural Networks*, 6, pp.1069-1072.

Hornik, K., Stinchcombe, M., White, H. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2, pp.359-366.

Hornik, K., Stinchcombe, M., White, H. (1990). Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks. *Neural Networks*, 3, pp.551-560.

Hornik, K., Stinchcombe, M., White, H., Auer, P. (1994). Degree of Approximation Results for Feedforward Networks Approximating Unknown Mappings and Their Derivatives. *Neural Computation*, 6, pp.1262-1275.

Le Cun, Y. (1985) Une procédure d'apprentissage pour réseaux de neurones à seuil asymétrique. *Proceedings of Cognitiva 85.* pp.599-604.

Leshno, M., Lin, V.Ya., Pinkus, A., Schocken, S. (1993). Multilayer Feedforward Networks with Non-Polynôme Activation Function Can Approximate Any Function. *Neural Networks*, 6, pp.861-867.

Light, W. (1992). In E.W. Cheney, C.K. Chui and L.L. Schumaker

Lorentz, G.G. (1966). *Approximation of Functions*, New-York: Chelsea Publishing Company.

Lorentz, G.G. (1986). *Bernstein Polynomials*, New-York: Chelsea Publishing Company.

Mc Clelland, J.L., Rumelhart, D.E. (1986) Parallel distributed processing, explorations in the microstructure of cognition. *Foundations*, 1, *Psychological and Biological models*, 2.

Mhaskar, H.N. (1993). Neural networks for localised approximation of smooth and analytic functions. *Neural networks for signal processing*, III, pp.190-196.

Mhaskar, H.N., Micchelli, C.A. (1994). Dimension-independent bounds of the degree of approximation by neural networks. *IBM J. Research and Development*, 38, pp.277-283.

Mhaskar, H.N. (1996). On smooth activation functions. To appear in a special issue of *Annals of mathematics and artificial intelligence* published by Baltzer Science Publishers, entitled *Mathematics of neural networks*, (eds. J.C. Mason and S.W. Ellacott), and comprising the proceedings of the 1995 conference on Mathematics of Neural Networks and Applications at Lady Margaret Hall, Oxford, UK.

Minsky, M., Papert, S. (1969) *Perceptrons*, Cambridge (USA): MIT Press.

Rosenblatt, F. (1962) *Principles of Neurodynamics*, New-York: Spartan.

Rudin, W. (1973). *Functional Analysis*, New-York: McGraw-Hill.

Scarselli, F., Tsoi, A.C. (1995). Universal approximation using feedforward neural networks: a survey of some existing methods, and some new results. Report available from the authors: F. Scarselli, Dipartimento di Ingegneria dei Sistemi e Informatica, Universit di Firenze, v.S. Marta, n.3, 50139 Firenze, Italy. A.C. Tsoi, Dept. Elec. and Comp. Eng., University of Queensland, St. Lucia, Queensland, 4072 Australia.

Schwartz, L. (1973) *Théorie des distributions*, Paris: Hermann.

Timan, A.F. (1963). *Theory of Approximation of Functions of a Real Variable*, New-York: Macmillan Co.

Deuxième partie

Méthodes topologiques pour des modèles markoviens homogènes non fellériens

Chapitre 1

Introduction

Historiquement, les premières études de la stationnarité chaînes de Markov sont dues à Doob et à Chung et portaient sur des chaînes à valeurs dans un espace dénombrable (voir (Doob, 1953) et (Chung, 1967). Cependant, comme beaucoup d'auteurs le soulignaient, de nombreux modèles étaient amenés à évoluer dans un espace d'états général, ou du moins continu comme $\mathbb{R}^d$. Il était donc naturel de chercher à étendre les notions développées pour les chaînes à espace d'états dénombrable à un cadre plus général. La première extension de la notion d'irréductibilité est due à Doeblin (voir (Doeblin, 1937) et (Doeblin, 1938)), permettant ainsi à Doob, puis à Harris (Harris, 1957) et à Chung, d'étendre l'approche des chaînes à espace d'états dénombrable à certaines chaînes évoluant dans des espaces quelconques, non forcément topologiques. Cette approche a par la suite été améliorée dans (Orey, 1971) puis revisitée dans (Nummelin, 1984) et (Revuz, 1984). Enfin, on peut trouver les exposés les plus récents dans (Duflo, 1997) et (Meyn et Tweedie, 1993).

Cette démarche s'appuie principalement sur la recherche, pour les chaînes irréductibles, d'un petit ensemble récurrent pour lequel l'espérance du temps de retour de la chaîne dans ce dernier est bornée. Elle conduit, le plus souvent, à des résultats puissants assurant non seulement l'existence et l'unicité de la loi stationnaire, mais aussi des vitesses de convergence en variation des itérées de la transition $P(x, dy)$ vers la loi stationnaire.

Il existe cependant une deuxième approche, plus récente mais désormais classique, dont les précurseurs sont (Jamison, 1964), (Foguel, 1969), (Lin, 1970), (Rosenblatt, 1974) ou encore (Sine, 1975), portant sur les modèles markoviens fellériens définis sur un espace métrique polonais S muni de sa tribu borélienne $\mathcal{B}(S)$, c'est-à-dire dont l'opérateur de transition vu comme application fonctionnelle de l'espace des fonctions boréliennes bornées de S dans $\mathbb{R}$ conserve l'espace des fonctions continues bornées de S dans $\mathbb{R}$. Cette approche, particulièrement bien adaptée à l'étude des chaînes à valeurs dans $\mathbb{R}^d$, s'appuie sur des méthodes de topologie faible sur l'espace des mesures. Ainsi, pour obtenir l'existence d'une loi stationnaire pour un modèle à valeurs dans un espace métrique polonais S de transition fellérienne $P(x, dy)$, il suffit en fait d'établir la tension pour un certain $x_0 \in S$ de la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ (voir e.g. Duflo, 1990). Par exemple, l'existence de lois stationnaires pour les modèles de régression Non Linéaire $NLR(1)$ de la forme $X_{t+1} = F(X_t) + \varepsilon_{t+1}$, où X_t est à valeurs dans $\mathbb{R}^d$ et les ε_t sont des variables i.i.d., peut se montrer aisément pourvu que F soit d'une part continue, pour assurer que la transition est fellérienne et qu'elle ait, d'autre part, une croissance suffisamment lente à l'infini, pour assurer la tension de la suite $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$. Plus généralement pour les modèles $ARCH(1)$ de la forme $X_{t+1} = F(X_t) + G(X_t)\varepsilon_{t+1}$, il suffit que F et G soient continues et possèdent un bon comportement à l'infini. Il est important de noter que, pour ces deux types de modèles, si le bruit ε_t a une densité par rapport à la mesure de Lebesgue, les transitions deviennent fortement fellériennes, i.e. qu'elle tranforment les fonctions boréliennes bornées de $\mathbb{R}^d$ dans $\mathbb{R}$ en fonctions continues bornées de $\mathbb{R}^d$ dans $\mathbb{R}$.

Malheureusement, cette démarche topologique élémentaire est mise en défaut dès que les modèles ne sont plus fellériens. Ainsi, dans le cas des modèles $NLR(1)$ ou $ARCH(1)$, dès qu'il existe un point de discontinuité à F ou G, on ne peut plus appliquer cette méthode.

Il est cependant connu, via les techniques de petits ensembles récurrents, que certains modèles $NLR(1)$, où F est une fonction de seuil, possèdent des lois stationnaires alors qu'ils ne sont pas fellériens. Ces modèles de régression apparaîssent naturellement en réseaux de neurones (Neural Auto-Regressive Models ou NAR) et occupent une place importante en statistique.

Notre objectif est donc d'étendre les méthodes de topologie faible sur l'espace des probabilités, afin de traiter un certain nombre de modèles non fellériens.

Pour ce faire, nous allons introduire la notion de chaîne de Markov quasi-(fortement) fellérienne qui est une généralisation du caractère (fortement) fellérien. (cf chapitre 2).

Le chapitre 3 est essentiellement consacré à la démonstration du théorème 3.1 assurant l'existence d'une loi stationnaire pour les modèles quasi-fellériens. De même que dans le cadre fellérien, ce théorème s'appuie sur une hypothèse de tension, pour un certain $x_0 \in S$, de la suite des mesures de probabilité $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$, où $P(x, dy)$ est la transition de la chaîne considérée. Nous serons donc amenés à rappeler les critères de Pakes et de Hajek qui sont les deux principaux critères permettant d'assurer cette tension.

Le chapitre 4 sera consacré à l'application du résultat précédent aux modèles $NLR(1)$ et $NLR(p)$ à valeurs dans $\mathbb{R}^d$ et le chapitre 5 à son application aux modèles $ARCH(1)$, $ARCH(p)$ et $GARCH(p, q)$ à valeurs dans $\mathbb{R}^d$. Lors de ces deux chapitres, une attention toute particulière sera portée aux modèles implémentés par des fonctions neuronales $NAR(p)$, $NARCH(p)$ et $NGARCH(p, q)$.

Au chapitre 6, nous étendrons le théorème 3.1 lorsque la chaîne évolue sur un espace métrique localement compact et séparable S et que sa transition vérifie le critère de Has'minskii. Cette approche sera un peu différente puisque le critère de Has'minskii n'est pas réellement un critère de tension, pour tout $x \in S$, des mesures $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n\in\mathbb{N}^*}$. Nous illustrerons rapidement ce critère pour les modèles NLR, $ARCH$ et $GARCH$, toujours à valeurs dans $\mathbb{R}^d$.

Par suite, nous nous intéresserons à l'obtention de propriétés de loi des grands nombres, stabilité et récurrence positive, pour les modèles quasi-fellériens et quasi-fortement fellériens (chapitre 7).

Dans le cadre quasi-fellérien, un raffinement de la démarche mise en évidence pour démontrer le théorème 3.1 nous permettra d'étendre le résultat de stabilité, notion originellement introduite pour les modèles fellériens, à nos modèles. Nous montrerons, à travers d'exemples, qu'il est possible d'obtenir des modèles stables non fellériens et non positivement récurrents validant ainsi cette approche. Nous n'étudierons pas la stabilité des modèles quasi-fortement fellériens puisque leur stabilité entraîne leur récurrence positive.

Pour obtenir de la récurrence positive nous étudierons séparément les cadres quasi-fellérien et quasi-fortement fellérien. Dans le cadre quasi-fellérien (standard ou Has'minskii), nous obtiendrons la récurrence positive, en étendant le résultat de (Meyn et Tweedie, 1993) d'existence d'un petit ensemble ouvert pour les modèles fellériens aux modèles quasi-fellériens. Enfin, dans le cadre quasi-fortement fellérien (standard ou Has'minskii), une approche par l'indécomposabilité (Breiman, 1968) nous permettra d'obtenir la récurrence positive sous des hypothèses plus faibles que dans le cadre quasi-fellérien et sans invoquer l'existence d'un petit ensemble. Nous verrons que ces résultats n'englobent pas ce que l'on peut obtenir par l'approche par "petits ensembles récurrents", mais qu'ils permettent de traiter des modèles que cette approche ne permet pas.

Enfin, au chapitre 8, nous montrerons que, sous le critère de Hajek, la propriété de stabilité d'un modèle peut être étendue afin d'obtenir une loi des grands nombres pour un certain nombre de fonctions non bornées que nous préciserons, améliorant ainsi le résultat de (Duflo, 1990, page 175). Ce dernier point permettra de donner une démonstration élémentaire ainsi qu'un raffinement des propriétés probabilistes du mélange lipschitzien.

Chapitre 2

Présentation des modèles

Ce chapitre a pour objet d'introduire les extensions des modèles fellériens et fortement fellériens que nous nous proposons d'étudier dans la suite.

Les modèles généraux que nous allons introduire dans ce chapitre sont définis sur un espace polonais S (i.e. séparable, métrique et complet) muni de sa tribu borélienne $\mathcal{B}(S)$, mais la plupart des exemples et applications, notamment aux chapitres 4 et 5 (modèles $NLR(1)$, $ARCH(1)$ ...), seront à valeurs dans $\mathbb{R}^d$ supposé muni de sa tribu borélienne $\mathcal{B}(\mathbb{R}^d)$

Notations : • d désignera la distance sur S.
• $|\,.\,|$ désignera indifféremment la valeur absolue de $\mathbb{R}$ ou la norme sur $\mathbb{R}^d$.
• $\mathcal{K}$ désignera l'ensemble des compacts de S.

2.1 Les modèles quasi-fellériens

Définition 2.1 *Un modèle markovien homogène de transition $P(x, dy)$ est quasi-fellérien s'il existe un espace topologique W, une fonction borélienne $H : S \to W$ telle que, pour tout compact K, $H(K)$ soit relativement compact et une famille $(Q(w, dy))_{w\in W}$ de probabilités sur $(S, \mathcal{B}(S))$, vérifiant :*

i) l'opérateur Q défini sur les fonctions boréliennes $g : S \to \mathbb{R}$ par $Qg(w) := \int_S g(y)Q(w, dy)$ est fellérien au sens où :

$$\forall g \in C_b\left(S, \mathbb{R}\right),\ Qg \in C_b\left(W, \mathbb{R}\right),$$

ii) $\forall g$ borélienne bornée de S dans $\mathbb{R}$, $Pg(x) = Qg\left(H(x)\right)$,

iii) $\forall w \in \bigcup_{K\in\mathcal{K}} \overline{H(K)}$, $Q\left(w, D_H\right) = 0$, où D_H désigne l'ensemble des points de discontinuité de H.

Remarques : • On a toujours $\bigcup_{K\in\mathcal{K}} \overline{H(K)} \subset \overline{H(S)}$. Si H est continue, on a toujours égalité, en revanche dès que H est discontinue en au moins un point, l'inclusion peut être stricte. Nous verrons un tel exemple lors du chapitre 4.
• Cette définition a bien un sens car D_H est toujours mesurable (Billingsley, 1968, page 225).
• L'écriture de la transition P en Q et H n'est évidemment pas unique.
• Une transition fellérienne P est quasi-fellérienne avec $W = S$, $H = Id_S$, $D_H = \emptyset$ et $Q = P$.
• Il est à noter, dans cette définition, que Q n'est à proprement parler une transition que si $W = S$.

• Dans nos applications, W sera généralement un espace $\mathbb{R}^p$ ou un de ses sous-ensembles.

On peut tout de suite traduire ces conditions pour les modèles $NLR(1)$ et $ARCH(1)$.

2.1.1 Les modèles $NLR(1)$ quasi-fellériens

Un modèle de régression non linéaire $NLR(1)$ est défini par la donnée d'une fonction borélienne $F : \mathbb{R}^d \to \mathbb{R}^d$, d'un bruit blanc $(\varepsilon_t)_{t\geq 1}$ de loi μ définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et de l'équation récursive :

$$X_{t+1} = F(X_t) + \varepsilon_{t+1},\ X_0 \in \mathbb{R}^d. \tag{2.1}$$

$(X_t)_{t\geq 0}$ est clairement un processus de Markov homogène par rapport à la filtration du bruit blanc $\mathcal{F}_t := \sigma(\varepsilon_1, \cdots, \varepsilon_t)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Sa transition $P(x, dy)$ est donc définie sur les fonctions boréliennes bornées $g : \mathbb{R}^d \to \mathbb{R}$ par :

$$Pg(x) = \int g(F(x) + y)\mu(dy). \tag{2.2}$$

On peut alors établir le résultat suivant :

Proposition 2.1 *Si l'ensemble D_F des points de discontinuité de F vérifie :*

$$\forall w \in \bigcup_{K\in\mathcal{K}} \overline{F(K)}, \quad \mu\left(D_F - w\right) = 0. \tag{2.3}$$

Alors le modèle $NLR(1)$ est quasi-fellérien.

Démonstration : Si l'on définit Q pour les fonctions boréliennes bornées $g : \mathbb{R}^d \to \mathbb{R}$ par :

$$\forall w \in \mathbb{R}^d,\ Qg(w) = \int g(w + y)\mu(dy),$$

il est clair que $Q(w, dy)$ est une probabilité (ici $W = \mathbb{R}^d$ et Q reste une transition) et on a $Pg(x) = Qg\left(F(x)\right)$. D'autre part il évident, par convergence dominée, que Q est fellérienne. L'hypothèse $Q\left(w, D_F\right) = 0$ pour tout $w \in \bigcup_{K\in\mathcal{K}} \overline{F(K)}$ se traduisant pour le modèle $NLR(1)$ par

$$\forall w \in \bigcup_{K\in\mathcal{K}} \overline{F(K)},\ \mu\left(D_F - w\right) = 0,$$

on a donc le résultat annoncé. □

2.1.2 Les modèles $ARCH(1)$ quasi-fellériens

Un modèle $ARCH(1)$ est défini par la donnée d'une fonction borélienne $F : \mathbb{R}^d \to \mathbb{R}^d$ et d'une fonction borélienne $G : \mathbb{R}^d \to S^+(d, \mathbb{R})$ ensemble des matrices définies positives, d'un bruit blanc $(\varepsilon_t)_{t\geq 1}$ de loi μ définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et de l'équation récursive :

$$X_{t+1} = F(X_t) + G(X_t)\varepsilon_{t+1},\ X_0 \in \mathbb{R}^d. \tag{2.4}$$

On désignera par G^* l'application de $\mathbb{R}^d$ dans $S^+(d, \mathbb{R})$ qui à x fait correspondre l'inverse matriciel de $G(x)$, $\left(G(x)\right)^{-1}$, ceci afin de ne pas confondre avec l'éventuelle fonction réciproque de G, G^{-1}.

On définira la norme d'une matrice de dimension d par la norme d'opérateur associée :

$$\|A\| := \sup_{|u|=1} \frac{|Au|}{|u|}.$$

La transition $P(x, dy)$ de ce modèle est définie par :

$$\forall g \text{ borélienne bornée, } Pg(x) = \int g\left(F(x) + G(x)y\right)\mu(dy). \tag{2.5}$$

Proposition 2.2 *Si l'ensemble* $D_{F,G} := D_F \cup D_G$ *vérifie :*

$$\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{F(K)} \times \overline{G(K)},\ \mu\left((w^2)^{-1}\left(D_{F,G} - w^1\right)\right) = 0, \tag{2.6}$$

alors le modèle $ARCH(1)$ *est quasi-fellérien.*

Remarque : Ici $W = \mathbb{R}^d \times S^+(d, \mathbb{R})$, et $\overline{G(K)} \subset S^+(d, \mathbb{R})$.

Démonstration de la proposition 2.2 : Si l'on définit Q pour les fonctions boréliennes bornées $g : \mathbb{R}^d \to \mathbb{R}$ par :

$$\forall (w^1, w^2) \in \mathbb{R}^d \times S^+(d, \mathbb{R}),\ Qg(w^1, w^2) = \int g(w^1 + w^2 y)\mu(dy), \tag{2.7}$$

on a $Pg(x) = Qg\left(F(x), G(x)\right)$. Q est clairement fellérienne au sens de la définition 2.1 (ici Q n'est plus une transition car $E = \mathbb{R}^d \times S^+(d, \mathbb{R})$).
L'hypothèse $\forall w \in \bigcup_{K \in \mathcal{K}} \overline{F(K)} \times \overline{G(K)},\ Q\left(w, D_{F,G}\right) = 0$ se traduit alors pour le modèle $ARCH(1)$ par

$$\forall w := (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{F(K)} \times \overline{G(K)},\ \mu\left((w^2)^{-1}\left(D_{F,G} - w^1\right)\right) = 0. \quad \square$$

2.2 Les modèles quasi-fortement fellériens

Définition 2.2 *Un modèle markovien homogène de transition $P(x, dy)$ est quasi-fortement fellérien s'il existe un espace topologique W, une fonction borélienne $H : S \to W$ telle que, pour tout compact K, $H(K)$ soit relativement compact et une famille de probabilités sur $(S, \mathcal{B}(S))$, $\left(Q(w, dy)\right)_{w \in W}$ vérifiant :*

i) l'opérateur Q défini sur les fonctions boréliennes $g : S \to \mathbb{R}$ par $Qg(w) := \int_S g(y)Q(w, dy)$ est fortement fellérien au sens où :

$$\forall g \text{ borélienne bornée de } S \text{ dans } \mathbb{R},\ Q(g) \in C_b\left(W, \mathbb{R}\right),$$

ii) $\forall g$ borélienne bornée de S dans $\mathbb{R}$, $Pg(x) = Qg\left(H(x)\right)$,

iii) $\forall w \in \bigcup_{K \in \mathcal{K}} \overline{H(K)}$, $Q\left(w, D_H\right) = 0$, où D_H désigne toujours l'ensemble des points de discontinuité de H.

Remarques : • Les modèles quasi-fortement fellériens sont évidemment quasi-fellériens.
• De même que pour le cadre quasi-fellérien, la notion de transition quasi-fortement fellérienne généralise la notion de transition fortement fellérienne.

• Il est à noter que les chaînes quasi-fortement fellériennes ne sont généralement pas fellériennes. Il suffit en effet que D_H soit non vide.

Nous allons illustrer cette notion avec les modèles $NLR(1)$ et $ARCH(1)$, ce qui revient essentiellement à considérer que le bruit a une densité par rapport à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$.

Il est cependant possible de trouver d'autres modèles quasi-fortement félleriens dont les transitions ne soient pas absolument continues par rapport à la mesure de Lebesgue.

2.2.1 Les modèles $NLR(1)$ quasi-fortement fellériens

On considère un modèle $NLR(1)$:

$$X_{t+1} = F(X_t) + \varepsilon_{t+1},\ X_0 \in \mathbb{R}^d,$$

et on suppose que la loi μ du bruit s'écrit $\mu(dy) = u(y)dy$.

Sa transition $P(x, dy)$ est donc définie pour les fonctions boréliennes bornées g de $\mathbb{R}^d$ dans $\mathbb{R}$ par :

$$Pg(x) = \int g(F(x) + y)u(y)dy. \tag{2.8}$$

Proposition 2.3 *Si l'ensemble D_F des points de discontinuité de F vérifie :*

$$\lambda_d\left(D_F\right) = 0. \tag{2.9}$$

Alors le modèle $NLR(1)$ est quasi-fortement fellérien.

Démonstration : On définit toujours Q pour les fonctions boréliennes bornées g de $\mathbb{R}^d$ dans $\mathbb{R}$ par :

$$Qg(w) = \int g(w + y)u(y)dy,$$

et on a encore $Pg(x) = Qg\left(F(x)\right)$.

D'autre part $Qg(w) = \left(\widehat{g} * u\right)(-w)$ où $\widehat{g}(x) = g(-x)$ et $*$ désigne la convolution sur $\mathbb{R}^d$. Qg est donc continue car g est dans L^∞ et u dans L^1. Q est donc fortement fellérienne. La mesure de Lebesgue étant invariante par translation, l'hypothèse $\lambda_d\left(D_F\right) = 0$ permet donc d'affirmer que :

$$\forall w \in \bigcup_{K \in \mathcal{K}} \overline{F(K)},\ Q\left(w, D_F\right) = 0,$$

car toutes les probabilités $Q(w, dy)$ sont absolument continues par rapport à la mesure de Lebesgue. □

2.2.2 Les modèles $ARCH(1)$ quasi-fortement fellériens

Proposition 2.4 *On suppose toujours que la loi du bruit $\left(\varepsilon_t\right)_{t\geq 1}$ s'écrit $\mu(dy) = u(y)dy$. Si l'ensemble $D_{F,G} := D_F \cup D_G$ vérifie :*

$$\lambda_d\left(D_{F,G}\right) = 0, \tag{2.10}$$

Alors le modèle $ARCH(1)$ est quasi-fortement fellérien.

Démonstration : On a pour g borélienne bornée,

$$\forall (w^1, w^2) \in \mathbb{R}^d \times S^+(d,\mathbb{R}),\ Qg(w^1, w^2) = \int g(w^1 + w^2 y) u(y) dy,$$

et il s'agit encore de montrer que Q est fortement fellérienne.

La fonction u est intégrable donc il existe une suite $(u_j)_{j\in\mathbb{N}}$ de fonctions continues convergeant vers u dans L^1. On définit alors :

$$Q_j g(w^1, w^2) := \int g(w^1 + w^2 y) u_j(y) dy = \int g(y) u_j((w^2)^{-1}(y - w^1)) \frac{dy}{\text{dét } w^2}.$$

Pour tout $j \in \mathbb{N}$, $Q_j g$ est continue sur $\mathbb{R}^d \times S^+(d,\mathbb{R})$, et :

$$\left| Q_j g - Qg \right| (w^1, w^2) \leq \| u_j - u \|_{L^1} \| g \|_\infty,$$

ce qui prouve que $Q_j g$ converge vers Qg uniformément sur tout compact de $\mathbb{R}^d \times S^+(d,\mathbb{R})$ et par suite que Qg est continue sur $\mathbb{R}^d \times S^+(d,\mathbb{R})$.

La mesure de Lebesgue étant invariante par homothétie-translation et les probabilités $Q(w, dy)$ étant dominées par la mesure de Lebesgue, l'hypothèse $\lambda_d\left(D_{F,G}\right) = 0$ permet d'assurer que :

$$\forall w := (w^1, w^2) \in \bigcup_{K\in\mathcal{K}} \overline{F(K)} \times \overline{G(K)},\ Q\left(w, D_{F,G}\right) = 0. \quad \square$$

Chapitre 3

Existence de probabilités invariantes pour les modèles quasi-fellériens

L'essentiel de ce chapitre est consacré a établir le théorème d'existence de probabilités invariantes pour les modèles quasi-fellériens proposés au chapitre précédent. Ce théorème, fondé sur des méthodes de convergence étroite de mesures, fournit en outre une localisation du support de la ou des mesures invariantes obtenues.

Une hypothèse essentielle consiste à supposer l'existence d'un x_0 dans S tel que la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ soit tendue. Ce point dépend, pour l'essentiel, du comportement de H à l'infini et somme toute assez peu de la régularité fellérienne de P.

Néanmoins, les principaux critères de tension des modèles à valeurs dans S seront rappelés (Pakes, Hajek) ou adaptés (Has'minskii) à notre situation pour fournir des conditions vérifiables à notre théorème d'existence.

3.1 Existence d'une loi stationnaire

Théorème 3.1 *On considère un modèle markovien de transition $P(x, dy)$ quasi-fellérienne relativement à la fonction H et au noyau Q comme explicité dans la définition 2.1.*
On suppose en outre qu'il existe $x_0 \in S$ tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue. Alors
(a) le modèle admet une probabilité invariante.
(b) Pour toute loi invariante ν et pour tout borélien C vérifiant :

$$\forall w \in \bigcup_{K\in\mathcal{K}} \overline{H(K)},\ Q\,(w, C) = 0, \tag{3.1}$$

on a $\nu(C) = 0$. En particulier on a donc $\nu(D_H) = 0$.

Démonstration du théorème 3.1 : Il existe par hypothèse au moins une valeur d'adhérence ν à la suite de probabilités $\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)$. L'objectif est de démontrer, comme dans le cadre fellérien usuel, que ν est nécessairement une loi stationnaire.

L'essentiel du travail consiste à montrer que $\nu\,(D_H) = 0$. On va, pour ce faire, établir directement le point (b). Soit un borélien C de S vérifiant (3.1). Montrons que $\nu(C) = 0$.

Considérons dans un premier temps un fermé $F \subset C$. F vérifie également (3.1), i.e. :

$$\forall w \in \bigcup_{K \in \mathcal{K}} \overline{H(K)},\ Q\,(w, F) = 0.$$

Soit $(h_p)_{p \in \mathbb{N}}$ la suite de fonctions S dans $\mathbb{R}$ définie par :

$$\forall p \in \mathbb{N}^*,\ \forall x \in S,\ h_p := 1 - \min\left(pd\,(x, F)\,, 1\right).$$

Les fonctions h_p sont continues pour tout $p \in \mathbb{N}$ (et même p-lipschitziennes). Elles vérifient en outre $\|h_p\|_\infty \leq 1$ et $h_p \searrow \mathbf{1}_{\overline{F}} = \mathbf{1}_F$ car F est fermé.

On définit alors la suite de fonctions $(Qh_p)_{p \in \mathbb{N}}$ de W dans $\mathbb{R}$. Les fonctions Qh_p sont continues car Q est fellérienne. D'autre part pour tout $w \in W$, $Q(w, dy)$ est une probabilité, donc d'après le théorème de convergence dominée :

$$\forall\ w \in \bigcup_{K \in \mathcal{K}} \overline{H(K)},\ Qh_p(w) \searrow Q(w, F).$$

Soit K un compact de S. $Qh_p \searrow 0$ sur $\overline{H(K)}$ puisque $Q(w, F) = 0$ si $w \in \overline{H(K)}$. Enfin, $\overline{H(K)}$ est compact par hypothèse, le théorème de Dini implique alors que la convergence précédente est uniforme sur $\overline{H(K)}$ i.e.

$$\sup_{w \in \overline{H(K)}} Qh_p(w) \to 0 \text{ quand } p \to +\infty. \tag{3.2}$$

Posons $\nu_n := \frac{1}{n} \sum_{k=0}^{n-1} P^k(x_0, dy)$ avec la convention $P^0(x_0, dy) = \delta_{x_0}(dy)$. La suite $(\nu_n)_{n \in \mathbb{N}^*}$ est tendue donc :

$$\forall \varepsilon > 0,\ \exists K_\varepsilon \text{ compact de } S \text{ tel que, } \forall n \in \mathbb{N}^*,\ \nu_n({}^cK_\varepsilon) \leq \varepsilon. \tag{3.3}$$

Par suite, pour tous n, $p \geq 1$,

$$\nu_n(Ph_p) = \int_{K_\varepsilon} Ph_p d\nu_n + \int_{{}^cK_\varepsilon} Ph_p d\nu_n \leq \sup_{y \in K_\varepsilon} Ph_p(y) + \nu_n({}^cK_\varepsilon) \leq \sup_{y \in K_\varepsilon} Ph_p(y) + \varepsilon.$$

Or $Ph_p = Qh_p \circ H$ donc, pour tous n, $p \geq 1$,

$$\nu_n(Ph_p) \leq \sup_{w \in \overline{H(K_\varepsilon)}} Qh_p(w) + \varepsilon.$$

Soit une suite extraite $\left(\nu_{\varphi(n)}\right)_{n \in \mathbb{N}^*}$ convergeant étroitement vers une probabilité ν. Comme pour toute fonction f bornée on a :

$$|\nu_n(Pf) - \nu_n(f)| \leq \frac{2\|f\|_\infty}{n},$$

il est immédiat que :

$$\nu(h_p) = \lim_{n \to \infty} \nu_{\varphi(n)}(h_p) = \lim_{n \to \infty} \nu_{\varphi(n)}(Ph_p) \leq \sup_{w \in \overline{H(K_\varepsilon)}} Qh_p(w) + \varepsilon.$$

Il vient donc, d'après (3.2) :

$$\forall\ \varepsilon > 0,\ \limsup_{p \to \infty} \nu(h_p) \leq \varepsilon,$$

et partant, $\nu(F) = \lim_{p\to\infty} \nu(h_p) = 0$

On a donc prouvé que pour tout fermé $F \subset C$, $\nu(F) = 0$. ν étant une probabilité, elle est intérieurement régulière sur les fermés i.e. :

$$\nu(C) := \sup\{\nu(F),\ F \subset C,\ F \text{ fermé}\} \quad \text{(Billingsley, 1968, page 7).}$$

Finalement $\nu(C) = 0$ et cela prouve, en particulier, que $\nu(D_H) = 0$.

Soit maintenant $h \in C_b$. Il est immédiat que $Ph = Qh \circ H$ est continue en tout point où H est continue. Ph est donc ν-*p.s.* continue. Par suite :

$$\nu_{\varphi(n)}(Ph) \overset{n\to+\infty}{\longrightarrow} \nu(Ph) \text{ car } \nu_{\varphi(n)} \Rightarrow \nu.$$

Comme :

$$|\nu_n(Ph) - \nu_n(h)| \overset{n\to+\infty}{\longrightarrow} 0 \text{ et } \nu_{\varphi(n)}(h) \overset{n\to+\infty}{\longrightarrow} \nu(h),$$

on en déduit $\nu(Ph) = \nu(h)$ et enfin $\nu P = \nu$ puisque h est quelconque dans C_b. □

3.2 Retour sur les critères de tension

3.2.1 Rappels des critères de Pakes et Hajek

Les critères usuels d'existence d'une loi stationnaire pour un modèle fellérien à valeurs dans S s'appuient sur des critères de tension, pour tout x dans S, de la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy)\right)_{n\in\mathbb{N}^*}$. Ils sont donc directement adaptables à nos problèmes.

Ces critères sont principalement basés sur le comportement de $PV - V$ à l'infini pour une fonction V "de rappel" vérifiant, pour un $a \in S$ (donc pour tous) :

$$\lim_{d(x,a)\to\infty} V(x) = +\infty.$$

Une telle fonction est souvent appelée fonction de Lyapounov pour la transition P.

Théorème 3.2 *(Ethier et Kurtz, 1986, Duflo, 1990)*
On considère une chaîne Markov homogène de transition $P(x,dy)$ sur S et une fonction de Lyapounov V.

Critère de Pakes (Pakes, 1969) *On suppose que :*

i) $\forall x \in S, \quad PV(x) \le V(x) + \beta,$
ii) pour un $a \in S,\ \lim_{d(x,a)\to\infty} (PV - V)(x) = -\infty.$

Alors,
pour tout x dans S la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy)\right)_{n\in\mathbb{N}^*}$ *est tendue.*

Critère de Hajek (Hajek, 1982) *On suppose que :*

$$\forall x \in S, \quad PV(x) \le \alpha V(x) + \beta \text{ avec } 0 < \alpha < 1.$$

Alors,
(a) pour tout x dans S la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy)\right)_{n\in\mathbb{N}^*}$ *est tendue,*

(b) de plus, pour toute valeur d'adhérence ν, on a $\nu(V) \leq \dfrac{\beta}{1-\alpha}$.

Démonstration : La démonstration du critère de Pakes est bien connue et le critère de Hajek en est un corollaire. Démontrons toutefois rapidement le point (b) du critère de Hajek.

Il est clair que l'on a pour tout n de $\mathbb{N}$:

$$P^n V \leq \alpha^n V + \frac{\beta}{1-\alpha}.$$

On a alors pour tout réel $M > 0$:

$$P^n(V \wedge M) \leq P^n V \wedge M \leq \left(\alpha^n V + \frac{\beta}{1-\alpha}\right) \wedge M.$$

Si ν est une loi stationnaire, il vient :

$$\nu(V \wedge M) \leq \alpha^n \nu(V \wedge M) + \frac{\beta}{1-\alpha} \wedge M.$$

Passant alors à la limite sur n puis sur M, on obtient $\nu(V) \leq \dfrac{\beta}{1-\alpha}$. $\square$

Commentaire sur les critères : Il est clair que le critère de Pakes suppose la sous-V-linéarité de PV, en d'autres termes, il suppose d'avoir $\limsup\limits_{d(x,a)\to+\infty} \dfrac{PV}{V}(x) \leq 1$. D'autre part, pour que le critère de Hajek de tension soit vérifié il faut que $\limsup\limits_{d(x,a)\to+\infty} \dfrac{PV}{V}(x) < 1$ (et c'est même suffisant si PV est bornée sur les compacts). Ce critère permettant par ailleurs d'obtenir des résultats plus forts (intégrabilité de V sous une loi stationnaire, stabilité voire récurrence géométrique), il est alors clair que le critère de Pakes ne trouve son utilité que dans le cas où l'on ne peut obtenir que $\limsup\limits_{d(x,a)\to+\infty} \dfrac{PV}{V}(x) = 1$.

Pour exemple, si $PV(x) \leq V(x) - \ln(1 + V(x))$, le critère de Pakes est vérifié par V alors que le critère de Hajek ne l'est pas forcément. Nous verrons un exemple plus précis lors de l'étude des modèles $NLR(1)$.

Précisons qu'un troisième critère, le critère de Has'minskii, sera explicité plus loin.

3.2.2 Application aux modèles p-dépendants

Il est possible d'introduire, en toute généralité, les modèles p-dépendants sur S par la donnée d'une famille de probabilités $R(x^1, \cdots, x^p, dy)$ sur S où pour tout $i \in \{1, \cdots, p\}$, $x^i \in S$. On peut alors définir, pour une fonction borélienne bornée $h : S \to \mathbb{R}$, la fonction $Rh : (S)^p \to \mathbb{R}$ par

$$Rh(x^1, \cdots, x^p) = \int h(y) R(x^1, \cdots, x^p, dy). \tag{3.4}$$

Le processus $(X_t)_{t\geq p}$ est alors défini de manière analogue à la 1-dépendance, c'est-à-dire que si $(\mathcal{F}_t)_{t\geq 0}$ est la filtration associée aux X_t, la loi de X_{t+p} conditionnelle à $\mathcal{F}_{t+p-1}$ est $R(X_{t+p-1}, \cdots, X_t, dy)$.

Ce modèle n'est évidemment pas une chaîne de Markov, mais on peut étudier le modèle surdimensionné $(Y_t)_{t\geq p}$ défini par :

$$\forall t \geq p-1,\ Y_t := (X_t, \cdots, X_{t-p+1}) \in (S)^p. \tag{3.5}$$

$(Y_t)_{t\geq p}$ obéit alors à la dynamique markovienne de transition P sur $(S)^p$ donnée par :

$$P(x^1, \cdots, x^p, dz) = R(x^1, \cdots, x^p, dz^1) \otimes \delta_{x^1}(z^2) \cdots \otimes \delta_{x^{p-1}}(z^p). \tag{3.6}$$

Pour une fonction borélienne bornée $g : (S)^p \to \mathbb{R}$, on a donc :

$$Pg(x) = \int g(y, x^1, \cdots, x^{p-1}) R(x^1, \cdots, x^p, dy). \tag{3.7}$$

L'illustration des critères de Pakes et Hajek pour les modèles p-dépendants est très classique (Duflo, 1990). Nous en donnons ici une démonstration élémentaire ne nécessitant pas de raisonner sur une transition $\widehat{P}$ itérée de P. L'idée est simplement de trouver la "bonne" fonction de Lyapounov pour le modèle surdimensionné.

Proposition 3.1 *On considère un modèle p-dépendant de transition $R(x^1, \cdots, x^p, dy)$ sur S et une fonction de Lyapounov V.*

Critère de Pakes *On suppose qu'il existe p réels strictement positifs $(\alpha_i)_{1\leq i\leq p}$ vérifiant $\sum_{i=1}^{p} \alpha_i \leq 1$ tels que :*

i) $\forall (x^1, \cdots, x^p) \in (S)^p\,,\ RV(x^1, \cdots, x^p) \leq \sum_{i=1}^{p} \alpha_i V(x^i) + \beta,$

ii) $\lim\limits_{\sup_{1\leq i\leq p} d(x^i, a) \to \infty} \left(RV(x^1, \cdots, x^p) - \sum_{i=1}^{p} \alpha_i V(x^i) \right) = -\infty.$

Alors, pour tout x dans $(S)^p$, la suite de probabilités $\left(\frac{1}{n} \sum_{k=0}^{n-1} P^k(x, dy) \right)_{n\in\mathbb{N}^*}$ est tendue où P est la transition du modèle surdimensionné.

Critère de Hajek *S'il existe p réels positifs $(\alpha_i)_{1\leq i\leq p}$ vérifiant $\sum_{i=1}^{p} \alpha_i < 1$ tels que :*

$$\forall (x^1, \cdots, x^p) \in (S)^p\,,\ RV(x^1, \cdots, x^p) \leq \sum_{i=1}^{p} \alpha_i V(x^i) + \beta,$$

alors, pour tout x dans $(S)^p$, la suite de probabilités $\left(\frac{1}{n} \sum_{k=0}^{n-1} P^k(x, dy) \right)_{n\in\mathbb{N}^*}$ est tendue où P désigne la transition du modèle surdimensionné.

Remarques : • Dans le critère de Pakes, on est obligé de supposer que tous les α_i sont *strictement* positifs pour que l'hypothèse *ii*) ait un sens.

• Pour le critère de Hajek, ce n'est plus nécessaire. Cependant, puisque $\sum_{i=1}^{p} \alpha_i < 1$, on peut toujours trouver p réels strictement positifs α'_i tels que :

$$\forall i,\ 1 \leq i \leq p,\ \alpha'_i > \alpha_i \text{ et } \sum_{i=1}^{p} \alpha'_i < 1.$$

Démonstration de la proposition 3.1: Hypothèses à la Pakes. Il ne s'agit en fait que de trouver une bonne fonction de Lyapounov sur $(S)^p$ construite à partir de V.

Soit $x := (x^1, \cdots, x^p) \in (S)^p$. On définit V_p sur $(S)^p$ par:

$$V_p(x^1, \cdots, x^p) := (\alpha_1 + \cdots + \alpha_p)V(x^1) + (\alpha_2 + \cdots + \alpha_p)V(x^2) + \cdots + \alpha_p V(x^p).$$

On a alors:

$$PV_p(x^1, \cdots, x^p) = (\alpha_1 + \cdots + \alpha_p)RV(x^1, \cdots, x^p) + (\alpha_2 + \cdots + \alpha_p)V(x^1) + \cdots + \alpha_p V(x^{p-1}).$$

Comme $\sum_{i=1}^{p} \alpha_i \leq 1$, on tire de ce qui précède:

$$PV_p(x^1, \cdots, x^p) \leq RV(x^1, \cdots, x^p) + (\alpha_2 + \cdots + \alpha_p)V(x^1) + \cdots + \alpha_p V(x^{p-1}),$$

Il est clair que:

$$PV_p(x) - V_p(x) \leq RV(x^1, \cdots, x^p) - \sum_{i=1}^{p} \alpha_i V(x^i).$$

Les hypothèses $i)$ et $ii)$ entraîne donc de manière évidente que:

$$PV_p(x) - V_p(x) \leq \beta \text{ et } \lim_{d_p(x,a_p)\to\infty} PV_p(x) - V_p(x) = -\infty,$$

où $a_p := (a, \cdots, a)$ et $d_p(x, a_p) := \sup_{1\leq i\leq p} d(x^i, a)$ qui est évidemment une distance sur $(S)^p$.

V_p étant clairement une fonction de Lyapounov, le critère de Pakes est vérifié pour le modèle surdimensionné.

Hypothèses à la Hajek. D'après la remarque suivant l'énoncé du théorème, on peut supposer, sans perte de généralité, que tous les α_i sont strictement positifs.

Reprenons V_p définie dans le cadre précédent. On a toujours:

$$PV_p(x^1, \cdots, x^p) = (\alpha_1 + \cdots + \alpha_p)RV(x^1, \cdots, x^p) + (\alpha_2 + \cdots + \alpha_p)V(x^1) + \cdots + \alpha_p V(x^{p-1}),$$

Mais par hypothèse:

$$RV(x^1, \cdots, x^p) \leq \sum_{i=1}^{p} \alpha_i V(x^i) + \beta,$$

d'où l'on tire avec ce qui précède:

$$\begin{aligned} PV_p(x^1, \cdots, x^p) \ &\leq (\alpha_1(\alpha_1 + \cdots + \alpha_p) + \alpha_2 + \cdots + \alpha_p)\, V(x^1) + \\ &(\alpha_2(\alpha_1 + \cdots + \alpha_p) + \alpha_3 + \cdots + \alpha_p)\, V(x^2) + \cdots + \\ &\alpha_p\, (\alpha_1 + \cdots + \alpha_p)\, V(x^p) + \beta. \end{aligned} \tag{3.8}$$

Comme $\alpha_1 + \cdots + \alpha_p < 1$ et que pour tout i, $1 \leq i \leq p$, $\alpha_i > 0$, il est clair que:

$$\begin{cases} \alpha_1' := \dfrac{\alpha_1(\alpha_1 + \cdots + \alpha_p) + \alpha_2 + \cdots + \alpha_p}{\alpha_1 + \cdots + \alpha_p} < 1 \\ \alpha_2' := \dfrac{\alpha_2(\alpha_1 + \cdots + \alpha_p) + \alpha_3 + \cdots + \alpha_p}{\alpha_2 + \cdots + \alpha_p} < 1 \\ \qquad \vdots \\ \alpha_p' := \dfrac{\alpha_p(\alpha_1 + \cdots + \alpha_p)}{\alpha_p} < 1. \end{cases}$$

Posant alors $\alpha' := \sup_{1 \leq i \leq p} \alpha'_i$, on a $\alpha' < 1$ et l'inégalité (3.8) permet alors de conclure à :

$$PV_p(x^1, \cdots, x^p) \leq \alpha' V_p(x^1, \cdots, x^p) + \beta. \quad \square$$

Chapitre 4

Loi stationnaire pour les modèles NLR

Rappelons que les modèles de régression non linéaires sont définis par la donnée d'une fonction borélienne $F : \mathbb{R}^d \to \mathbb{R}^d$, d'un bruit blanc $(\varepsilon_t)_{t\geq 1}$ de loi μ et de l'équation récursive :

$$X_{t+1} = F(X_t) + \varepsilon_{t+1}. \tag{4.1}$$

Pour les fonctions boréliennes bornées $g : \mathbb{R}^d \to \mathbb{R}$ on a :

$$Pg(x) = \int g(F(x) + y)\mu(dy).$$

Dans ce chapitre D_F désignera l'ensemble des points de discontinuités de F. Pour ne pas alourdir inutilement la notation, des hypothèses simples relatives aux critères de tension sur les mesures empiriques associées à P seront spécifiées en fin de chapitre (il ne s'agira que de rappels).

Nous allons traiter d'abord le cas général où μ est une loi quelconque éventuellement singulière, puis nous traduirons les hypothèses dans le cas particulier important où μ est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Nous donnerons de multiples exemples et illustrations.

Une attention toute particulière sera portée aux modèles de régression d'ordre p, $NLR(p)$.

4.1 Cadre général (bruit éventuellement singulier)

4.1.1 Résultat principal

Théorème 4.1 *On considère un modèle $NLR(1)$ et on suppose que :*

i) l'ensemble D_F vérifie :

$$\forall w \in \bigcup_{K\in\mathcal{K}} \overline{F(K)}, \quad \mu\left(D_F - w\right) = 0.$$

ii) $F : \mathbb{R}^d \to \mathbb{R}^d$ est borélienne, bornée sur les compacts,

iii) il existe x_0 tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue.

Alors

(a) le modèle admet une probabilité invariante.

(b) En outre, pour toute loi invariante ν et pour tout borélien C vérifiant

$$\forall w \in \bigcup_{K \in \mathcal{K}} \overline{F(K)}, \quad \mu\,(C - w) = 0, \tag{4.2}$$

$\nu(C) = 0$. En particulier $\nu(D_F)$.

Démonstration : L'hypothèse $i)$ assure selon la proposition 2.1 que le modèle $NLR(1)$ est quasi-fellérien. Ici $W = \mathbb{R}^d$. F étant bornée sur les compacts, $\overline{F(K)}$ est un compact de $\mathbb{R}^d$ si K est compact de $\mathbb{R}^d$. On est alors avec $ii)$ et $iii)$ sous les conditions d'application du théorème 3.1. □

Commentaires sur les hypothèses : • *Retour sur la condition $iii)$:* L'hypothèse de tension de la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ dépend essentiellement du comportement de F à l'infini.

• *A propos de la condition $i)$:* Pour toute fonction $F : \mathbb{R}^d \to \mathbb{R}^d$, on a évidemment :

$$F(\mathbb{R}^d) \subset \bigcup_{K \in \mathcal{K}} \overline{F(K)} \subset \overline{F(\mathbb{R}^d)}.$$

Si F est continue, comme $\overline{F(K)} = F(K)$, il y a toujours égalité. En revanche si F est discontinue en au moins un point, l'inclusion peut être stricte. Ainsi (sur $\mathbb{R}$) la fonction

$$F(x) = \begin{cases} e^{-|x|}, \text{ si } x \neq 0 \\ \frac{1}{2}, \text{ si } x = 0 \end{cases} \tag{4.3}$$

vérifie $\bigcup_{K \in \mathcal{K}} \overline{F(K)} =]0,1]$, $F(\mathbb{R}^d) =]0,1[$ et $\overline{F(\mathbb{R}^d)} = [0,1]$.

• Plus précisément, il peut arriver que l'hypothèse $i)$ soit vérifiée et qu'il existe $w \in \overline{F(\mathbb{R}^d)} \setminus \bigcup_{K \in \mathcal{K}} \overline{F(K)}$ tel que $\mu\,(D_F - w) \neq 0$. Ainsi considérons $X_{t+1} = F(X_t) + \varepsilon_{t+1}$ où F est définie par (4.3) et $\mu = \delta_0$ (il s'agit donc en fait du système dynamique déterministe $x_{t+1} = F(x_t)$!).

D'après le théorème 4.1, comme

$$D_F = \{0\} \quad \text{et} \quad \bigcup_{K \in \mathcal{K}} \overline{F(K)} =]0,1],$$

le système admet au moins une loi stationnaire (par exemple $\nu = \delta_{x^*}$, où $x^* \neq 0$ est l'unique point fixe du sytème déterministe). Comme μ ne charge que 0, on est bien dans la situation recherchée.

• *Conditions nécessaires induites sur μ par la condition $i)$:* Il est important de noter que l'hypothèse $i)$ induit de fortes contraintes sur μ, notamment, la restriction de μ à l'ensemble :

$$D_F - \bigcup_{K \in \mathcal{K}} \overline{F(K)} = \left\{ y - w,\ y \in D_F,\ w \in \bigcup_{K \in \mathcal{K}} \overline{F(K)} \right\}$$

est diffuse. En effet, soit $x \in D_F - \bigcup_{K \in \mathcal{K}} \overline{F(K)}$; $x = y - w$, $\mu(\{x\}) \leq \mu(D_F - w) = 0$.

Exemple : Si l'on pose :

$$\begin{cases} F(x) := -\dfrac{n}{n+1}(x-n)+1, \ x \in]n, n+1], \ n \in \mathbb{N}^*, \\ F(x) = \dfrac{1}{2}, \ x \in]0,1], F \text{ paire et } F(0) = 0, \end{cases}$$

il vient $\bigcup_{K \in \mathcal{K}} \overline{F(K)} =]0,1]$, et $D_F = \mathbb{Z}$.

Par suite μ doit être diffuse sur $D_F - \bigcup_{K \in \mathcal{K}} \overline{F(K)} = \mathbb{Z} -]0,1] = \mathbb{R}$, donc, dans cet exemple, aucune partie discrète n'est autorisée à μ.

4.1.2 Application à des modèles à valeurs dans un sous-ensemble de $\mathbb{R}^d$

Il est important de bien spécifier l'espace des états de la chaîne avant de traiter l'existence éventuelle d'une loi stationnaire. En effet, il est possible que la chaîne évolue sur un sous-ensemble S fermé (donc complet) de $\mathbb{R}^d$ sur laquelle elle est quasi-fellérienne alors qu'elle ne l'est pas sur $\mathbb{R}^d$, le cas le plus flagrant étant, bien sûr quand le modèle est à espace d'états fini comme pour l'application suivante :

Application 4.1 *Le modèle :*

$$\begin{cases} X_{t+1} = sign(X_t) + \varepsilon_{t+1}, \\ X_0 = 0 \end{cases},$$

où les ε_t sont i.i.d. et suivent une loi $\mathcal{B}er\left(\pm 1, \frac{1}{2}\right)$, admet une loi stationnaire.

Pour tout $t \in \mathbb{N}$, X_t est à valeurs dans $\{-2,-1,0,1,2\}$. La chaîne $(X_t)_{t\geq 1}$ est donc à valeurs dans $S := \{-2,-1,0,1,2\}$ qui est bien sûr polonais. Elle est alors évidemment quasi-fellérienne, puisque fellérienne et même fortement fellérienne.

Cependant, il n'est pas utile de faire appel au caractère fellérien d'une telle chaîne, puisque l'on peut expliciter la loi stationnaire. Les probabilités conditionnelles s'écrivent :

$$\begin{cases} \mathbb{P}\left(X_{t+1} = 2/X_t = 1 \text{ ou } 2\right) = 1/2, \\ \mathbb{P}\left(X_{t+1} = 0/X_t = 1 \text{ ou } 2\right) = 1/2, \\ \mathbb{P}\left(X_{t+1} = \pm 1/X_t = 0\right) = 1/2, \\ \mathbb{P}\left(X_{t+1} = 0 \text{ ou } -2/X_t = -1 \text{ ou } -2\right) = 1/2. \end{cases}$$

La matrice de transition P est donc :

$$P = \frac{1}{2}\begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

La loi $\nu := \left(\frac{1}{9}, \frac{2}{9}, \frac{1}{3}, \frac{2}{9}, \frac{1}{9}\right)$ est alors l'unique loi invariante car la matrice P est irréductible (ce qui assure en outre que P est géométriquement récurrente). □

Pourtant, si l'on considère ce modèle comme étant à valeurs dans $\mathbb{R}$, on trouve :

$$D_{sign(.)} = \{0\} \text{ et } \bigcup_{K \in \mathcal{K}} \overline{sign(K)} = \{-1, 0, 1\},$$

les conditions d'application du théorème 4.1 ne sont donc pas remplies puisque :

$$\nu\left(D_{sign(.)} - 1\right) = \nu\left(D_{sign(.)} + 1\right) = \frac{2}{9} > 0 \text{ et } \nu\left(D_{sign(.)}\right) = \frac{1}{3} > 0. \quad \square$$

Le modèle précédent étant à espace d'états fini, il n'est pas nécessaire, comme nous venons de le voir, de faire appel au théorème 4.1. Donnons donc un exemple où nous utilisons réellement le caractère quasi-fellérien de la chaîne.

Application 4.2 *Soit la fonction $F : \mathbb{R}^2 \to \mathbb{R}^2$ définie par :*

$$F(x^1, x^2) := \begin{pmatrix} f(x^1) \\ \frac{1}{4}\left([x^1] + [x^2]\right) \end{pmatrix},$$

où $[\,.\,]$ désigne la partie entière et $f : \mathbb{R} \to \mathbb{R}$.

Soit $(\varepsilon_t^1)_{t\in\mathbb{N}}$ une suite de variables i.i.d. de loi μ vérifiant :

$$\forall w^1 \in \mathbb{R}, \quad \mu\left((D_f \cup \mathbb{Z}) - w^1\right) = 0,$$

où D_f est l'ensemble des points de discontinuité de f.

Soit également $(\varepsilon_t^2)_{t\in\mathbb{N}}$ une suite de variables i.i.d. indépendante de $(\varepsilon_t^1)_{t\in\mathbb{N}}$ de loi $\sum_{n\in\mathbb{Z}} a_n\delta_n$, avec, pour tout $n \in \mathbb{Z}$, $a_n > 0$ et $\sum_{n\in\mathbb{Z}} a_n = 1$.

Si, en outre, il existe x_0 tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue, alors le modèle $NLR(1)$ défini par :

$$X_{t+1} := F(X_t) + \begin{pmatrix} \varepsilon_{t+1}^1 \\ \varepsilon_{t+1}^2 \end{pmatrix},$$

admet une loi stationnaire.

Remarque : La tension des mesures empiriques est vérifiée, par exemple, si ε_1^1 et ε_1^2 ont un moment d'ordre 1 et s'il existe $0 < \alpha < 1$ et β tels que :

$$\forall x \in \mathbb{R}, \quad |f(x)| \leq \alpha|x| + \beta,$$

puisque le modèle vérifie alors le critère de Hajek (voir la section 4.3).

Démonstration de l'application 4.2 : On se place sur l'espace $S := \mathbb{R} \times \frac{\mathbb{Z}}{4}$ puisque dès la date $t = 1$, la chaîne est à valeurs dans S. S est clairement séparable complet.

Montrons que le modèle est quasi-fellérien sur S. L'ensemble D_F des points de discontinuité de F s'écrit :

$$D_F = (D_f \cup \mathbb{Z}) \times \frac{\mathbb{Z}}{4}.$$

On a alors $\forall w := (w^1, w^2) \in S$,

$$\mu \otimes \left(\sum_{n \in \mathbb{Z}} a_n \delta_n \right) (D_F - w) = \mu\left((D_f \cup \mathbb{Z}) - w^1 \right) \times \sum_{n \in \mathbb{Z}} a_n \delta_n \left(\frac{\mathbb{Z}}{4} \right) = 0.$$

Soit maintenant une fonction $h : S \to \mathbb{R}$ continue bornée et $x := (x^1, x^2) \notin D_F$. Soit également $(x_p)_{p \in \mathbb{N}} := \left((x_p^1, x_p^2) \right)_{p \in \mathbb{N}}$ une suite de points convergeant vers (x^1, x^2). Comme x^1 n'appartient pas à $\mathbb{Z}$ et que, pour tout $p \in \mathbb{N}$, x_p^2 appartient à $\dfrac{\mathbb{Z}}{4}$, il existe un entier $p_0 \geq 0$ tel que :

$$\forall p \geq p_0, \quad [x_p^1] = [x^1] \quad \text{et} \quad [x_p^2] = [x^2].$$

On a alors :

$$\forall p \geq p_0, \quad Ph(x_p) := \sum_{n \in \mathbb{Z}} a_n \int_{\mathbb{R}} h \left(f(x_p^1) + u, \frac{1}{4} \left([x^1] + [x^2] \right) + n \right) \mu(du). \tag{4.4}$$

Il est clair que pour tout $n \in \mathbb{Z}$, les fonctions de $\mathbb{R}$ dans $\mathbb{R}$ définies par :

$$g_n(y) := \int_{\mathbb{R}} h \left(f(y) + u, \frac{1}{4} \left([x^1] + [x^2] \right) + n \right) \mu(du),$$

sont continues sur $\mathbb{R} \setminus D_f$ (puisque la deuxième coordonnée ne dépend pas de y).

Comme $\sum_{n \in \mathbb{Z}} a_n = 1 < +\infty$ et que h est bornée, (4.4) implique alors que la fonction g de $\mathbb{R}$ dans $\mathbb{R}$ définie par :

$$g(y) := \sum_{n \in \mathbb{Z}} a_n \int_{\mathbb{R}} h \left(f(y) + u, \frac{1}{4} \left([x^1] + [x^2] \right) + n \right) \mu(du),$$

est continue sur $\mathbb{R} \setminus D_f$ comme limite uniforme de fonctions continues sur $\mathbb{R} \setminus D_f$.

Comme $x^1 \notin D_f$, on a enfin :

$$\lim_{p \to +\infty} Ph(x_p) = Ph(x).$$

Ph est donc continue en tout point où F est continue, ce qui signifie que le modèle est quasi-fellérien.

On est alors sous les conditions d'application du théorème 2.1. □

Si l'on traitait ce modèle comme étant à valeurs dans $\mathbb{R}^2$, alors :

$$\mathbb{R} \times \mathbb{Z} \subset D_F.$$

On aurait donc, pour tout $w^1 \in \bigcup_{K \in \mathcal{K}} \overline{f(K)}$ et pour tout $w^2 \in \mathbb{Z}$:

$$\mu \otimes \left(\sum_{n \in \mathbb{Z}} a_n \delta_n \right) (D_F - w) = \mu(\mathbb{R}) \times \sum_{n \in \mathbb{Z}} a_n \delta_n (\mathbb{Z}) = 1 > 0,$$

et on ne serait pas sous les conditions d'application du théorème 4.1 permettant d'assurer l'existence d'une loi stationnaire.

4.1.3 Application aux fonctions F telles que D_F est dénombrable

Proposition 4.1 *On considère un modèle $NLR(1)$ à valeurs dans $\mathbb{R}^d$. On suppose que :*

i) l'ensemble D_F est dénombrable et la restriction de μ à $D_F - \bigcup_{K \in \mathcal{K}} \overline{F(K)}$ est diffuse,

ii) $F : \mathbb{R}^d \to \mathbb{R}^d$ est borélienne, bornée sur les compacts,

iii) il existe x_0 tel que $\left(\frac{1}{n} \sum_{k=0}^{n-1} P^k(x_0, dy) \right)_{n \in \mathbb{N}^*}$ est tendue.

Alors,

(a) il existe une loi stationnaire au modèle.

(b) En outre, pour toute loi stationnaire ν, $\nu(D_F) = 0$.

Démonstration : Il est clair que pour tout w de $\bigcup_{K \in \mathcal{K}} \overline{F(K)}$, $D_F - w$ est dénombrable, donc $\forall w \in \bigcup_{K \in \mathcal{K}} \overline{F(K)}, \quad \mu(D_F - w) = 0.$ □

Application 4.3 *La proposition 4.1 s'applique à toute forme de loi diffuse. Ainsi, si μ est de la forme :*

$$\mu := f.\mu_0, \quad f \geq 0 \text{ et } \int f d\mu_0 = 1,$$

où μ_0 est la mesure uniforme sur l'ensemble de Cantor (étendu à $\mathbb{R}$ tout entier) définie par : $\mu_0(]a,b]) = L(b) - L(a)$ où L est la fonction de Lebesgue, le modèle $NLR(1)$ admet une loi stationnaire dès que les hypothèses ii) et iii) sont remplies, ce qui est le cas, par exemple, pour les fonctions :

$$F_1(x) := \sum_{n=0}^{+\infty} \frac{1}{2^n} \sin \frac{n}{nx-1} \quad \text{et} \quad F_2(x) := \sum_{n=0}^{+\infty} \frac{1}{2^n} \sin \frac{1}{x - r_n},$$

où $(r_n)_{n \geq 0}$ désigne une numérotation des rationnels.

Cependant, il n'est pas nécessaire pour autant que μ soit diffuse pour entrer dans le champ d'application du théorème 4.1, comme le montre l' application suivante :

4.1.4 Application aux modèles $NLR(p)$

Le théorème 4.1 permet de traiter en outre le cas des modèles p-dépendants. Un modèle de régression non linéaire p-dépendant, que nous noterons $NLR(p)$, est défini comme suit :

$$\forall t \geq p-1, \ X_{t+1} = f(X_t, \cdots, X_{t-p+1}) + \varepsilon_{t+1}, \tag{4.5}$$

où X_t est à valeurs dans $\mathbb{R}^d$ et les ε_t sont i.i.d.

Un tel modèle n'est évidemment pas une chaîne de Markov, mais on entend par loi stationnaire d'un tel modèle, la mesure invariante du modèle surdimensionné $(Y_t)_{t \geq p}$ défini par :

$$\forall t \geq p-1, \ Y_t := (X_t, \cdots, X_{t-p+1}) \in (\mathbb{R}^d)^p. \tag{4.6}$$

$(Y_t)_{t \geq p}$ obéit en effet à la dynamique markovienne donnée par :

$$\forall t \geq p-1, \ Y_{t+1} = F(Y_t) + E_{t+1}, \tag{4.7}$$

où $F : (\mathbb{R}^d)^p \to (\mathbb{R}^d)^p$ est définie par

$$F(x^1, \cdots, x^p) = {}^t\left(f(x^1, \cdots, x^p), x^1, \cdots, x^{p-1}\right) \tag{4.8}$$

et $(E_t)_{t \geq p}$ est un bruit blanc (dégénéré dans $(\mathbb{R}^d)^p$) de la forme :

$$E_t = (\varepsilon_t, 0, \cdots, 0). \tag{4.9}$$

On notera encore sa loi sur $(\mathbb{R}^d)^p$ par μ.

Cette représentation, très classique, nous permet donc de ramener l'étude de $(X_t)_{t\in\mathbb{N}}$ à celle d'un nouveau modèle autorégressif de chaîne de Markov dont la transition est donnée par :

$$Pg\left(x^1, \cdots, x^p\right) = \int g\left(F(x^1, \cdots, x^p) + (u, 0, \cdots, 0)\right)\mu(du). \tag{4.10}$$

Avec les notations précédentes :

Théorème 4.2 *Soit un modèle $NLR(p)$ d'équation récursive (4.5). On suppose que :*

i) l'ensemble D_f vérifie :

$$\forall w \in \bigcup_{K\in\mathcal{K}} \overline{F(K)}, \quad \mu\left((D_f - w) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1}\right)\right) = 0,$$

ii) $f : \mathbb{R}^d p \to \mathbb{R}^d$ est borélienne, bornée sur les compacts,

iii) il existe x_0 tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue, où $P(x, dy)$ désigne la transition associée définie par (4.10).

Alors

(a) le modèle surdimensionné (4.7) admet une probabilité invariante.

(b) En outre, pour toute loi invariante ν et pour tout borélien C vérifiant

$$\forall w \in \bigcup_{K\in\mathcal{K}} \overline{F(K)}, \quad \mu\left((C - w) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1}\right)\right) = 0, \tag{4.11}$$

$\nu(C) = 0$. En particulier $\nu(D_F) = 0$.

Démonstration : f étant borélienne bornée sur les compacts, il en est de même pour F. Le résultat découle alors du théorème 4.1 puisque la loi μ de E_1 est concentrée sur $\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1}$. □

Application aux modèles $NAR(p)$

Les modèles $NAR(p)$ (Neural Auto Regressive) sont des processus autorégressifs p-dépendants dont la fonction de régression est un perceptron à une couche cachée (voir la partie I pour leur définition). Il est normal d'espérer qu'ils soient très efficaces puisque les perceptrons sont, comme nous l'avons vu lors de la partie I, des approximateurs universels. En outre, étant presque aussi simples que des $AR(p)$ linéaires, les $NAR(p)$ sont donc très attractifs pour effectuer des calculs de prévisions. On pourra, par exemple, se réferer à (Cottrell, Girard B., Girard Y., Mangeas et Muller, 1995) qui utilisent des modèles $NAR(p)$ pour modéliser la consommation journalière d'électricité en vue de prévisions et qui montrent leur compétitivité par rapport aux modèles linéaires.

Enfin, leur caractère paramétrique rend possible le calcul explicite de certains estimateurs, comme ceux des moindres carrés par exemple (Mangeas et Yao, 1996).

Cependant, les praticiens emploient souvent des fonctions d'activation ψ non continues, la plus courante d'entre elles étant la fonction de Heaviside valant 0 sur $\mathbb{R}_-$ et 1 sur $\mathbb{R}_+^*$. Il est donc important de pouvoir assurer l'existence d'une loi stationnaire stricte pour ce genre de modèles.

Application 4.4 Les modèles $NAR(p)$ *On considère un modèle $NAR(p)$ à valeurs dans $\mathbb{R}^d$ défini par :*

$$X_{t+1} = \sum_{i=0}^{n} \lambda_i \psi(A_{i,1}X_t + \cdots + A_{i,p}X_{t-p+1}) + \varepsilon_{t+1},$$

où les $(A_{i,j})_{\substack{1\le j\le p\\ 0\le i\le n}}$ sont des matrices de $\mathcal{M}_d(\mathbb{R})$ vérifiant, pour tout $i \in \{0, \cdots, n\}$, $A_{i,1}$ est inversible.
On suppose que :

i) le bruit suit une loi μ telle que l'ensemble D_ψ des points de discontinuités de ψ vérifie:

$$\forall i,\ 0 \le i \le n,\ \forall w \in \mathbb{R}^d, \quad \mu\left(A_{i,1}^{-1} D_\psi - w\right) = 0,$$

ii) ψ est bornée sur les compacts,

iii) il existe $x_0 = (x_0^1, \cdots, x_0^p) \in \left(\mathbb{R}^d\right)^p$ tel que la suite des probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue, où $P(x, dy)$ désigne la transition du modèle surdimensionné.

Alors,

(a) il existe une loi stationnaire au modèle surdimensionné.

(b) En outre toute loi invariante ν vérifie :

$$\forall i \in \{0, \cdots, n\}, \quad \nu\left(\left\{(x^1, \cdots, x^p) \in \left(\mathbb{R}^d\right)^p \ / \ \sum_{j=1}^{p} A_{i,j}x^j \in D_\psi\right\}\right) = 0.$$

Démonstration : Soit pour tout $i \in \{0, \cdots, n\}$, les ensembles D_i définis par :

$$D_i := \left\{(x^1, \cdots, x^p) \in \left(\mathbb{R}^d\right)^p \ / \ \sum_{j=1}^{p} A_{i,j}x^j \in D_\psi\right\}.$$

Soit Π_i la projection sur $\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1}$ parallèlement à $A_{i,1}x^1 + \cdots + A_{i,p}x^p = 0$. On a alors,

$$\Pi_i\left(x^1, \cdots, x^p\right) = \left(A_{i,1}^{-1}\sum_{i=1}^{p} A_{i,j}x^j, 0, \cdots, 0\right),$$

$$\text{et } (D_i - w) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1}\right) \subset \Pi_i\left(D_i - w\right) \begin{aligned}[t] &= \Pi_i\left(D_i\right) - \Pi\left(w\right) \\ &= A_{i,1}^{-1}D_\psi - A_{i,1}^{-1}\sum_{j=1}^{p} A_{i,j}w^j. \end{aligned}$$

Donc $\mu\left((D_i - w) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1}\right)\right) \le \mu\left(A_{i,1}^{-1}D_\psi - A_{i,1}^{-1}\sum_{j=1}^{p} A_{i,j}w^j\right) = 0$ par hypothèse.

Comme $D_F = \bigcup_{i=0}^{n} D_i$, on a donc :

$$\forall w \in \mathbb{R}^d, \quad \mu\left((D_F - w) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1}\right)\right) = 0.$$

Il est par ailleurs clair que si ψ est bornée sur les compacts, il en est de même pour F, on se trouve donc sous les conditions d'application du théorème 4.2. □

Commentaire sur le cas $A_{i,1}$ non inversible : Prenons $n = 0$ et $A_{0,1} = A_1 = 0$ par souci de simplification. L'ensemble D_F des points de discontinuité de F est alors donné par :

$$D_F = \left\{(x^1, \cdots, x^p)/\ A_2 x^2 + \cdots + A_p x^p \in D_\psi\right\},$$

où D_ψ est l'ensemble des points de discontinuité de ψ.

On remarque alors que, dès que D_ψ est non vide, on a :

$$D_F \cap \mathbb{R}^d \times \{0_{\mathbb{R}^d}\} \times \cdots \times \{0_{\mathbb{R}^d}\} = \mathbb{R}^d \times \{0_{\mathbb{R}^d}\} \times \cdots \times \{0_{\mathbb{R}^d}\}.$$

Comme μ est concentrée sur $\mathbb{R}^d \times \{0_{\mathbb{R}^d}\} \times \cdots \times \{0_{\mathbb{R}^d}\}$, on a donc :

$$\mu(D_F) = 0 \Leftrightarrow D_\psi = \emptyset.$$

On ne peut donc appliquer le théorème 4.1 que si ψ est continue. On retrouve alors les critères habituels donnant l'existence d'une loi stationnaire pour les modèles NAR(p) fellériens.

4.2 Les modèles NLR dans le cas Lebesgue

Il s'agit d'un cas particulier important du résultat général puisque l'invariance de la mesure de Lebesgue par homothétie-translation permet d'avoir un énoncé très simple des hypothèses.

Théorème 4.3 *Soit un modèle $NLR(1)$. On suppose que :*

i) la loi μ des ε_t est absolument continue par rapport à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$,
ii) $\lambda_d(D_F) = 0$,
iii) F est bornée sur les compacts,
iv) il existe x_0 tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue.

Alors

(a) le modèle admet au moins une loi stationnaire ν,

(b) toute loi stationnaire α vérifie en outre $\alpha \ll \lambda_d$.

Démonstration : Il s'agit d'un simple corollaire du théorème 3.1.

Les hypothèses $i)$ et $ii)$ permettent d'assurer par la proposition 2.3 que le modèle est quasi-fortement fellérien. On est alors avec les hypothèses $iii)$ et $iv)$ sous les conditions d'application du théorème 3.1.

Il est par ailleurs classique, via le théorème de Fubini, que si le bruit a une loi absolument continue par rapport à la mesure de Lebesgue alors toute loi stationnaire est absolument continue par rapport à la mesure de Lebesgue. □

Ce résultat généralise et simplifie celui de (Broniatowski et Diebolt, 1987) :

Théorème 4.4 (Broniatowski et Diebolt, 1987)
On considère un modèle $NLR(1)$ à valeurs dans $\mathbb{R}$ tel que le bruit ε_1 possède une densité m par rapport à la mesure de Lebesgue. On suppose que :

i) F admet un nombre fini de discontinuités de première espèce et est absolument continue ailleurs,

ii) $\limsup_{|x|\to+\infty} \dfrac{|F(x)|}{|x|} < 1$,

iii) il existe $A > 0$ tel que $m(x)$ soit dérivable pour $|x| \geq A$ et sa dérivée $m'(x)$ est bornée pour $|x| \geq A$,

iv) la restriction de la fonction $\dfrac{|m'(x)|}{m(x)}$ à l'intervalle $[A, +\infty[$ est croissante et à variation régulière d'ordre $\gamma > 0$,

v) la fonction $|F'(x)|$ est à croissance lente à l'infini,

vi) $\liminf_{|x|\to+\infty} \dfrac{m(x)}{1 - M(x)} > 0$, *et* $\liminf_{|x|\to-\infty} \dfrac{m(x)}{M(x)} > 0$, *où M est la fonction de répartition de ε_1,*

vii) la densité de probabilité m est une fonction paire bornée continue et sa restriction $m_{|\mathbb{R}^+}$ est décroissante.

Alors la chaîne est ergodique.

L'hypothèse $ii)$ est le critère de Hajek. Sous cette hypothèse, nous montrons que le modèle $NLR(1)$ a une loi stationnaire dès que ε_1 possède une densité m positive et que F est d'ensemble de discontinuité de mesure de Lebesgue nulle. Nous verrons en outre, lors de l'application 7.1 du paragraphe 7.6, qu'il suffit que la densité m soit strictement positive pour obtenir l'ergodicité de la chaîne.

On a donc une hypothèse sur F beaucoup plus fine que $i)$ et $v)$ et on ne tient jamais compte du comportement de m, qui est le sujet des hypothèses $iii)$, $iv)$, $vi)$ et $vii)$.

Enfin, nos résultats sont vrais pour les modèles à valeurs dans $\mathbb{R}^d$, alors que le précédent théorème ne s'applique que sur $\mathbb{R}$.

On peut toutefois donner des résultats dans le même esprit que le théorème 4.4. L'application ci-dessous montre que,

Application 4.5 *Dans le cadre réel, si $\mu \ll \lambda_1$ et que F est croissante ou plus généralement que F n'a que des discontinuités de première espèce (et sous réserve que $ii)$ soit vérifiée), le modèle $NLR(1)$ admet au moins une loi stationnaire puisque D_F est dénombrable donc Lebesgue négligeable.*

Application 4.6 *En dimension supérieure la notion de fonction croissante s'étend généralement en :*

$$\forall x := (x^1, \cdots, x^d),\ \forall y := (y^1, \cdots, y^d), \forall i\ x^i \leq y^i \Rightarrow \forall i\ F_i(x) \leq F_i(y).$$

On peut montrer (Vitali et Sansone, 1951) que de telles fonctions ont un ensemble D_F λ_d-négligeable car porté par une réunion dénombrable d'hyperplans parallèles aux axes naturels.

4.2.1 Application aux modèles $NAR(p)$ dans le cas Lebesgue

La forme particulière des modèles $NAR(p)$ permet de simplifier l'énoncé de l'application 4.4 dans le cas Lebesgue.

Application 4.7 *Soit le modèle $NAR(p)$ à valeurs dans $\mathbb{R}^d$ défini par :*

$$X_{t+1} = \sum_{i=0}^{n} \lambda_i \psi(A_{i,1}X_t + \cdots + A_{i,p}X_{t-p+1}) + \varepsilon_{t+1},$$

avec, pour tout i, $0 \leq i \leq n$, $A_{i,1}$ inversible.
On suppose que :

i) $\varepsilon_t \ll \lambda_d$,
ii) $\lambda_d(D_\psi) = 0$,
iii) ψ est bornée sur les compacts,
iv) il existe $x_0 = (x_0^1, \cdots, x_0^p) \in \left(\mathbb{R}^d\right)^p$ tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue, où $P(x, dy)$ est la probabilité de transition du modèle surdimensionné.

Alors,

(a) Le modèle surdimensionné admet une loi stationnaire ,

(b) toute loi stationnaire ν est absolument continue par rapport à la mesure de Lebesgue sur $\left(\mathbb{R}^d\right)^p$.

Démonstration : Corollaire immédiat de l'application 4.4. □

On peut également étudier des modèles $NLR(p)$ un peu plus compliqués que les $NAR(p)$:

Application 4.8 *Soit le modèle à valeurs dans $\mathbb{R}$:*

$$X_{t+1} = f(X_t^2 - X_{t-1}) + \varepsilon_{t+1}.$$

on suppose que :

i) $\mu := \mathcal{L}(\varepsilon_t) \ll \lambda_1$,
ii) $\lambda_1(D_f) = 0$,
iii) f est bornée sur les compacts,
iv) il existe $x_0 = (x_0^1, x_0^2) \in \left(\mathbb{R}^d\right)^2$ tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue, où $P(x, dy)$ est la probabilité de transition du modèle surdimensionné.

Alors le modèle surdimensionné admet une loi stationnaire.

Démonstration : Le modèle surdimensionné $Y_t = (X_t, X_{t-1})$ s'écrit :

$$Y_{t+1} = F(Y_t) + E_{t+1},$$

où $F(x^1, x^2) = \left(f\left((x^1)^2 - x^2\right), x^1\right)$ et $E_t = (\varepsilon_t, 0)$. On note toujours μ la loi de E_t sur $\mathbb{R}^2$ (μ est portée par l'axe des abscisses).

F est bornée sur les compacts. D'autre part, il est facile de voir que l'on a :

$$D_F = \left\{(x^1, x^2) \in \mathbb{R}^2 / \ (x^1)^2 - x^2 \in D_f\right\}.$$

Soit λ_1 la mesure de Lebesgue sur l'axe des abscisses de $\mathbb{R}^2$ et $w := (w^1, w^2) \in \mathbb{R}^2$.

$$\lambda_1 (D_F - w) = \lambda_1 ((D_F - w) \cap (\mathbb{R} \times \{0\})) = \lambda_1 \left(\left\{x^1 / \ (x^1 - w^1)^2 - w^2 \in D_f\right\}\right) = 0. \quad \square$$

On peut évidemment généraliser à des modèles de la forme :

$$X_{t+1} = f \left(P \left(X_t, \cdots, X_{t-p+1}\right)\right) + \varepsilon_{t+1},$$

où $P \in \mathbb{R}[X^1, \cdots, X^p]$ est tel que, pour tout $(x^2, \cdots, x^p) \in \mathbb{R}^{p-1}$, $P(x, x^2, \cdots, x^p)$ est un polynôme en x de degré supérieur à 1.

4.3 Traduction des critères de tension

Nous allons traduire, pour les modèles NLR, les critères de Pakes et Hajek explicités à la section 3.2. Ce sont ces critères qui permettent de vérifier les hypothèses de tension sur les probabilités empiriques des énoncés des paragraphes 3.2.1 et 3.2.2.

4.3.1 Une formulation des critères de Pakes et Hajek pour les modèles $NLR(1)$

On emploie généralement une norme convenablement choisise comme fonction de Lyapounov pour les modèles spécifiés, mais il est évidemment possible de trouver d'autres fonctions de Lyapounov.

Proposition 4.2 *On considère un modèle $NLR(1)$ sur $\mathbb{R}^d$. On suppose que le bruit ε_1 a un moment d'ordre 1.*

Critère de Pakes *On suppose :*

i) $\forall x \in \mathbb{R}^d, \quad |F(x)| \leq |x| + \beta,$
ii) $\lim\limits_{|x| \to \infty} (|F(x)| - |x|) = -\infty.$

Alors pour tout x dans $\mathbb{R}^d$ la suite de probabilités $\left(\dfrac{1}{n} \sum\limits_{k=0}^{n-1} P^k(x, dy)\right)_{n \in \mathbb{N}^*}$ *est tendue.*

Critère de Hajek *On suppose :*

$$\forall x \in \mathbb{R}^d, \quad |F(x)| \leq \alpha |x| + \beta \textit{ avec } 0 \leq \alpha < 1.$$

Alors pour tout x dans $\mathbb{R}^d$ la suite de probabilités $\left(\dfrac{1}{n} \sum\limits_{k=0}^{n-1} P^k(x, dy)\right)_{n \in \mathbb{N}^*}$ *est tendue.*

Remarque : Soit le modèle $NLR(1)$ sur $\mathbb{R}$ avec $F(x) := |x| - \ln(1 + |x|)$. il est clair que $V(.) := |\,.\,|$ vérifie le critère de Pakes de tension sans vérifier le critère de Hajek de tension.

4.3.2 Application aux modèles $NLR(p)$

Proposition 4.3 *On considère un modèle $NLR(p)$ sur $\mathbb{R}^d$. On suppose que ε_1 a un moment d'ordre 1.*

Critère de Pakes *On suppose qu'il existe p réels strictement positifs $(\alpha_i)_{1\leq i\leq p}$ vérifiant $\sum_{i=1}^{p}\alpha_i \leq 1$ tels que :*

i) $\forall (x^1,\cdots,x^p)\in \left(\mathbb{R}^d\right)^p, \quad |f(x^1,\cdots,x^p)| \leq \sum_{i=1}^{p}\alpha_i|x^i| + \beta,$

ii) $\lim_{\sup_{1\leq i\leq p}|x^i| \to +\infty} \left(|f(x^1,\cdots,x^p)| - \sum_{i=1}^{p}\alpha_i|x^i|\right) = -\infty,$

Alors pour tout x dans $\left(\mathbb{R}^d\right)^p$ la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1}P^k(x,dy)\right)_{n\in\mathbb{N}^*}$ est tendue, où $P(x,dy)$ est la transition du modèle surdimensionné.

Critère de Hajek *On suppose qu'il existe p réels positifs $(\alpha_i)_{1\leq i\leq p}$ vérifiant $\sum_{i=1}^{p}\alpha_i < 1$ tels que :*

$$\forall (x^1,\cdots,x^p)\in \left(\mathbb{R}^d\right)^p, \quad |f(x^1,\cdots,x^p)| \leq \sum_{i=1}^{p}\alpha_i|x^i| + \beta.$$

Alors pour tout x dans $\left(\mathbb{R}^d\right)^p$ la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1}P^k(x,dy)\right)_{n\in\mathbb{N}^*}$ est tendue, où $P(x,dy)$ est la transition du modèle surdimensionné.

Chapitre 5

Loi stationnaire pour les modèles $ARCH$ et $GARCH$

Les modèles $ARCH$ et $GARCH$ ont été introduits à partir de motivations économétriques et financières, notamment pour modéliser la volatilité d'actifs financiers à des fins de prévision.

Ces modèles sont en fait une généralisation des modèles $NLR(1)$ mais il était tout de même utile de préciser les critères d'existence d'une loi stationnaire pour les modèles $NLR(1)$.

Au chapitre 2, nous avions défini les modèles $ARCH(1)$ sous leur forme la plus générale (voir par exemple Duflo, 1990, Guégan, 1994) :

Définition 5.1 *Un modèle $ARCH(1)$ est défini par :*

$$X_{t+1} = F(X_t) + G(X_t)\varepsilon_{t+1}, \tag{5.1}$$

$X_t \in \mathbb{R}^d$, les ε_t sont des vecteurs aléatoires i.i.d. et de loi μ sur $\mathbb{R}^d$.
$F : \mathbb{R}^d \to \mathbb{R}^d$ et $G : \mathbb{R}^d \to S^+(d, \mathbb{R})$ ensemble des matrices définies positives.

Un modèle $ARCH(1)$ est un modèle markovien homogène de transition $P(x, dy)$ donnée par :

$$\forall g \text{ borélienne bornée, } Pg(x) = \int g\left(F(x) + G(x)y\right)\mu(dy).$$

On peut définir également les modèles $GARCH(1,1)$ de manière générale :

Définition 5.2 *Un modèle $GARCH(1,1)$ est défini par :*

$$X_{t+1} = F(X_t, \varepsilon_t) + G(X_t, \varepsilon_t)\varepsilon_{t+1}, \tag{5.2}$$

$X_t \in \mathbb{R}^d$, les ε_t sont des vecteurs aléatoires i.i.d. et de loi μ sur $\mathbb{R}^d$.
$F : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ et $G : \mathbb{R}^d \times \mathbb{R}^d \to S^+(d, \mathbb{R})$.

Sous cette forme, le modèle $GARCH(1,1)$ semble être une généralisation du modèle $ARCH(1)$. Toutefois, si l'on considère $(X_t)_{t\in\mathbb{N}}$ un processus $GARCH(1,1)$ et que l'on pose $Y_t := (X_t, \varepsilon_t)$, on a alors :

$$Y_{t+1} = \widehat{F}(Y_t) + \widehat{G}(Y_t)E_{t+1},$$

avec :

$$\widehat{F}(x^1, x^2) := \begin{pmatrix} F(x^1, x^2) \\ 0_{\mathbb{R}^d} \end{pmatrix}, \quad \widehat{G}(x^1, x^2) := \begin{pmatrix} G(x^1, x^2) & Id_{\mathcal{M}_d(\mathbb{R})} \\ Id_{\mathcal{M}_d(\mathbb{R})} & 0_{\mathcal{M}_d(\mathbb{R})} \end{pmatrix} \quad \text{et} \quad E_{t+1} := \begin{pmatrix} \varepsilon_{t+1} \\ 0 \end{pmatrix}.$$

Le modèle $(Y_t)_{t\in\mathbb{N}}$ est donc un modèle $ARCH(1)$. Nous énoncerons donc le résultat général pour les modèles $ARCH(1)$ puis nous donnerons sa traduction pour les $ARCH(p)$ et les $GARCH(p,q)$.

Rappelons que G^* désigne l'application de $\mathbb{R}^d$ dans $\mathcal{S}^+(d,\mathbb{R})$ qui à x fait correspondre l'inverse matriciel de $G(x)$.

La norme d'opérateur sur $\mathcal{M}_d(\mathbb{R})$ est définie par :

$$\forall A \in \mathcal{M}_d(\mathbb{R}), \quad \|A\| := \sup_{|u|=1} \frac{|Au|}{|u|}.$$

Le plan de ce chapitre est le même que celui de la partie précédente. Nous étudierons tout d'abord le cas où le bruit suit une loi quelconque μ, puis nous aborderons le cas particulier important où le bruit a une densité par rapport à la mesure de Lebesgue.

Dans toute la suite, $D_{F,G} := D_F \cup D_G$ désignera l'ensemble des discontinuités de l'application (F,G).

Comme pour les modèles NLR, des hypothèses simples sur F et G assurant la tension des probabilités empiriques associées à la transition P seront rapidement énoncées en fin de chapitre.

5.1 Le cadre général

5.1.1 Résultat principal

Théorème 5.1 *On considère un modèle $ARCH(1)$ et on suppose que :*

i) l'ensemble $D_{F,G}$ vérifie :

$$\forall (w^1,w^2) \in \bigcup_{K\in\mathcal{K}} \overline{F(K)} \times \overline{G(K)}, \quad \mu\left((w^2)^{-1}\left(D_{F,G} - w^1\right)\right) = 0,$$

ii) les fonctions F, G et G^* sont bornées sur les compacts,

iii) il existe x_0 tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0,dy)\right)_{n\in\mathbb{N}^*}$ est tendue.

Alors

(a) le modèle admet au moins une loi stationnaire.

(b) En outre, pour toute loi stationnaire ν et pour tout borélien C vérifiant :

$$\forall (w^1,w^2) \in \bigcup_{K\in\mathcal{K}} \overline{F(K)} \times \overline{G(K)}, \ \mu\left((w^2)^{-1}\left(C - w^1\right)\right) = 0, \tag{5.3}$$

on a $\nu(C) = 0$. En particulier $\nu(D_{F,G}) = 0$.

Démonstration : L'hypothèse $i)$ entraîne via la proposition 2.2 que le modèle est quasi-fellérien.

Il s'agit maintenant de démontrer que si K est un compact de $\mathbb{R}^d$, $\overline{F(K)} \times \overline{G(K)}$ est un compact de $\mathbb{R}^d \times S^+(d,\mathbb{R})$. F est bornée sur les compacts donc $\overline{F(K)}$ est un fermé borné donc un compact de $\mathbb{R}^d$.

G et G^* étant bornées sur K :

$$\exists a_K > 0 / \quad \forall w \in G(K), \quad a_K|u|^2 \le |w.u| \le \frac{|u|^2}{a_K}, \quad \forall u \in \mathbb{R}^d.$$

D'où :

$$\forall w \in \overline{G(K)}, \quad a_K |u|^2 \leq |w.u| \leq \frac{|u|^2}{a_K}, \quad \forall u \in \mathbb{R}^d,$$

ce qui prouve que :

$$\forall w \in \overline{G(K)}, w \in S^+(d, \mathbb{R}).$$

On est alors avec *iii*) sous les conditions d'application du théorème 3.1. □

5.1.2 Application aux fonctions F et G telles que $D_{F,G}$ est dénombrable

Lorsque $D_{F,G}$ est dénombrable, le théorème 5.4 permet de traiter les modèles $ARCH$ à bruit diffus. On peut, de même que pour les modèles NLR, donner l'exemple :

Application 5.1 *Si μ est de la forme :*

$$\mu := f.\mu_0,$$

où μ_0 est la mesure uniforme sur l'ensemble de Cantor étendu à $\mathbb{R}$ tout entier, si F, G et G^* sont bornées sur les compacts et que l'hypothèse iii) du théorème 5.1 est remplie, alors il existe une loi stationnaire.

5.1.3 Application aux modèles $ARCH(p)$

Un modèle $ARCH(p)$ est un modèle issu de p vecteurs aléatoires de $\mathbb{R}^d$, $X_0, \cdots, X_{p-1}$ et d'équation récursive :

$$\forall t \geq p-1, \quad X_{t+1} = f\left(X_t \cdots, X_{t-p+1}\right) + g\left(X_t \cdots, X_{t-p+1}\right) \varepsilon_{t+1} \tag{5.4}$$

où X_t est à valeurs dans $\mathbb{R}^d$, pour tout $(x^1, \cdots, x^p) \in \left(\mathbb{R}^d\right)^p$, $g\left(x^1 \cdots, x^p\right)$ est une matrice définie positive et les ε_t sont i.i.d. de loi μ.

On entend par loi stationnaire d'un tel modèle, la mesure invariante du modèle surdimensionné $(Y_t)_{t \geq p}$ défini par :

$$\forall t \geq p-1, \; Y_t := (X_t, \cdots, X_{t-p+1}) \in (\mathbb{R}^d)^p. \tag{5.5}$$

$(Y_t)_{t \geq p}$ obéit en effet à la dynamique markovienne donnée par :

$$Y_{t+1} = F\left(Y_t\right) + G\left(Y_t\right) E_{t+1}, \tag{5.6}$$

où $F(x^1, \cdots, x^p) = {}^t\left(f(x^1, \cdots, x^p), x^1, \cdots, x^{p-1}\right)$,

$G(x^1, \cdots, x^p)$ est la matrice formée de blocs diagonaux $d \times d$ tous égaux à $g(x^1, \cdots, x^p)$ (il n'y a pas unicité dans le choix de G mais si l'on veut conserver l'inversiblilité, ce choix est le plus naturel)

et $E_t = (\varepsilon_t, 0, \cdots, 0)$. Par abus, on notera toujours par μ sa loi sur $(\mathbb{R}^d)^p$.

Ce nouveau modèle est alors un modèle $ARCH(1)$ dont la transition est donnée par :

$$Pg\left(x^1, \cdots, x^p\right) = \int g\left(F(x^1, \cdots, x^p) + G(x^1, \cdots, x^p)(u, 0, \cdots, 0)\right) \mu(du). \tag{5.7}$$

Théorème 5.2 *Soit un modèle $ARCH(p)$ d'équation récursive (5.4). On suppose que :*

i) l'ensemble $D_{f,g}$ vérifie :

$$\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{F(K)} \times \overline{G(K)}, \quad \mu \left((w^2)^{-1} \left(\left(D_{f,g} - w^1 \right) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1} \right) \right) \right) = 0,$$

ii) f, g et g^* sont boréliennes, bornées sur les compacts,

iii) il existe x_0 tel que $\left(\frac{1}{n} \sum_{k=0}^{n-1} P^k(x_0, dy) \right)_{n \in \mathbb{N}^*}$ est tendue, où $P(x, dy)$ désigne la transition associée définie par (5.7).

Alors

(a) le modèle surdimensionné (5.6) admet une probabilité invariante.

(b) En outre, pour toute loi invariante ν et pour tout borélien C vérifiant

$$\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{F(K)} \times \overline{G(K)}, \quad \mu \left((w^2)^{-1} \left(\left(C - w^1 \right) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p-1} \right) \right) \right) = 0, \tag{5.8}$$

$\nu(C) = 0$. En particulier $\nu(D_{F,G})$.

Démonstration : Comme corollaire du théorème 5.1. □

5.1.4 Application aux modèles $GARCH(p, q)$

Un modèle $GARCH(p, q)$ est défini par :

$$\forall t \geq \sup(p, q) - 1, \quad X_{t+1} = f\left(X_t \cdots, X_{t-p+1}, \varepsilon_t, \cdots, \varepsilon_{t-q+1}\right) + g\left(X_t \cdots, X_{t-p+1}, \varepsilon_t, \cdots, \varepsilon_{t-q+1}\right) \varepsilon_{t+1}, \tag{5.9}$$

où X_t est à valeurs dans $\mathbb{R}^d$, pour tout $(x^1, \cdots, x^{p+q}) \in \left(\mathbb{R}^d\right)^{p+q}$, $g\left(x^1 \cdots, x^{p+q}\right)$ est une matrice définie positive et les ε_t sont i.i.d.

On s'intéresse alors à la probabilité invariante du modèle surdimensionné $(Y_t)_{t \geq \sup(p,q)}$ défini par :

$$\forall t \geq \sup(p, q) - 1, \; Y_t := (X_t, \cdots, X_{t-p+1}, \varepsilon_t, \cdots, \varepsilon_{t-q+1}) \in (\mathbb{R}^d)^{p+q}. \tag{5.10}$$

$(Y_t)_{t \geq \sup(p,q)}$ est un processus $ARCH(1)$ d'équation :

$$Y_{t+1} = F\left(Y_t\right) + G\left(Y_t\right) E_{t+1}. \tag{5.11}$$

F est définie par $F(x^1, \cdots, x^{p+q}) = {}^t\left(f(x^1, \cdots, x^{p+q}), x^1, \cdots, x^{p-1}, 0, x^{p+1}, \cdots, x^{p+q-1}\right)$.

La fonction G est une matrice de taille $(p+q)d \times (p+q)d$ définie par :

$$G := \begin{pmatrix} g & 0 & \dots & 0 & Id & 0 & \\ 0 & \ddots & \ddots & & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & Id \\ 0 & & \ddots & g & \ddots & & \\ Id & \ddots & & \ddots & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & & \ddots & \ddots & 0 \\ & 0 & Id & 0 & \dots & 0 & 0 \end{pmatrix},$$

0 et Id désignent respectivement la matrice nulle et la matrice identité de $\mathcal{M}_d(\mathbb{R})$. Ainsi, la diagonale de G est constituée de p matrices égales à g et de q matrices égales à la matrice nulle de $\mathcal{M}_d(\mathbb{R})$.

Enfin, $E_t = (\varepsilon_t, \underbrace{0, \cdots, 0}_{p+q-1 \text{ fois}})$. On notera encore μ sa loi.

la transition de ce modèle $ARCH(1)$ est alors donnée par :

$$Pg\left(x^1, \cdots, x^{p+q}\right) = \int g\left(F(x^1, \cdots, x^{p+q}) + G\left(x^1, \cdots, x^{p+q}\right)(u, 0, \cdots, 0)\right)\mu(du). \tag{5.12}$$

Théorème 5.3 *On considère un modèle $GARCH(p,q)$ d'équation récursive (5.9). On suppose que :*

i) l'ensemble $D_{f,g}$ vérifie :

$$\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{F(K)} \times \overline{G(K)}, \quad \mu\left((w^2)^{-1}\left(\left(D_{f,g} - w^1\right) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p+q-1}\right)\right)\right) = 0,$$

ii) f, g et g^* sont boréliennes, bornées sur les compacts,

iii) il existe x_0 tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n \in \mathbb{N}^*}$ est tendue, où $P(x, dy)$ est la transition du modèle surdimensionné définie par (5.12).

Alors

(a) le modèle surdimensionné (5.11) admet une probabilité invariante.

(b) De plus, pour toute loi invariante ν et pour tout borélien C vérifiant :

$$\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{F(K)} \times \overline{G(K)}, \ \mu\left((w^2)^{-1}\left(\left(C - w^1\right) \cap \left(\mathbb{R}^d \times \{0_{\mathbb{R}^d}\}^{p+q-1}\right)\right)\right) = 0, \tag{5.13}$$

$\nu(C) = 0$. En particulier $\nu(D_{F,G}) = 0$.

Démonstration : Simple corollaire du théorème 5.1. $\square$

5.2 Les modèles $ARCH$ dans le cas Lebesgue

5.2.1 Le cas général

Théorème 5.4 *On considère un modèle $ARCH(1)$ et on suppose que :*

i) $\mu \ll \lambda_d$,
ii) $\lambda_d(D_{F,G}) = 0$,
iii) F, G et G^* sont bornées sur les compacts,
iv) il existe $x_0 \in \mathbb{R}^d$ tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n \in \mathbb{N}^*}$ est tendue.

Alors

(a) le modèle admet au moins une loi stationnaire ν,

(b) toute loi stationnaire est absolument continue par rapport à λ_d.

Démonstration : Appliquer la proposition 2.4 et le théorème 3.1. $\square$

Evidemment, ce théorème ne permet pas de traiter tous les modèles $ARCH(1)$ non fellériens puisque nous avons besoin d'hypothèses concernant l'ensemble des points de discontinuités de F et G.

Cependant, comme nous le précisions au chapitre 4, l'étude de l'espace des états est primordiale. Ainsi, les résultats de Gouriéroux et Montfort (1990) concernant les modèles $QTARCH$ semblent ne pas rentrer dans le cadre du théorème 5.4. Ces modèles sont définis comme suit :

$$X_{t+1} = \sum_{j=1}^{J} \alpha_j \mathbf{1}_{A_j}(X_t) + \sum_{j=1}^{J} \beta_j \mathbf{1}_{A_j}(X_t)\varepsilon_{t+1}, \tag{5.14}$$

où $X_t \in \mathbb{R}^d$, les ε_t sont i.i.d. et distribués selon une loi équivalente à la mesure de Lebesgue sur $\mathbb{R}^d$ et les A_j sont mesurables et forment une partition de $\mathbb{R}^d$.

L'existence d'une loi stationnaire ne demandant aucune contrainte de plus, le résultat de Gouriéroux et Montfort s'énonce donc comme suit :

Théorème 5.5 (Gouriéroux et Montfort, 1990) *Tout modèle $QTARCH$ admet une loi stationnaire.*

Démonstration : On considère les variables :

$$Z_t = (Z_{1t}, \cdots, Z_{Jt}) \text{ où } Z_{jt} = \mathbf{1}_{A_j}(X_t).$$

Z_t est alors une chaîne de Markov à espace d'états fini possédant une loi stationnaire.

Il suffit alors de remarquer que Z_t est stationnaire si et seulement si X_t l'est pour avoir le résultat. $\square$

Ces modèles sont en fait des $ARCH(1)$ avec des fonctions F et G définies par :

$$F(x) := \sum_{j=1}^{J} \alpha_j \mathbf{1}_{A_j}(x) \text{ et } G(x) := \sum_{j=1}^{J} \beta_j \mathbf{1}_{A_j}(x).$$

On remarque qu'aucune contrainte sur la continuité des fonctions F et G n'est requise, ce résultat semble donc plus fort que ce que nous pourrions obtenir pour les modèles $QTARCH$. Cependant, ces modèles ne sont pas réellement des modèles $ARCH(1)$ dont l'espace d'états est $\mathbb{R}^d$ puisque la chaîne sous-jacente $(Z_t)_{t\in\mathbb{N}}$ est à espace d'états fini et nous savons que ce type de chaînes rentrent dans le cadre des chaînes fellériennes.

Bien sûr, la méthode exposée ci-dessus ne peut s'appliquer que si F et G ne prennent qu'un nombre fini de modalités (voire dénombrable) sinon on ne peut plus définir la nouvelle chaîne Z_t.

De plus, si l'on considère un modèle $QTARCH$ p-dépendant :

$$X_{t+1} = \sum_{j=1}^{J} \alpha_j \mathbf{1}_{A_j}(a_1X_t + \cdots + a_pX_{t-p+1}) + \sum_{j=1}^{J} \beta_j \mathbf{1}_{A_j}(a_1X_t + \cdots + a_pX_{t-p+1})\varepsilon_{t+1},$$

le modèle surdimensionné $Y_t := (X_t, \cdots, X_{t-p+1})$ n'est pas un $QTARCH$ et on ne peut plus appliquer la même démarche.

5.2.2 Application aux modèles $NARCH(p)$ dans le cas Lebesgue

Comme lors du paragraphe 4.2.1, il est possible de simplifier les énoncés des théorèmes 5.2 et 5.3 pour les modèles $ARCH(p)$ et $GARCH(p,q)$ implémentés par des perceptrons et avec un bruit de densité dominée par la mesure de Lebesgue.

Nous allons traiter un exemple simple de modèle $NARCH(p)$ implémenté par deux perceptrons à une unité sur la couche cachée, l'extension au cas de plusieurs unités ne posant aucun problème particulier :

Application 5.2 *Soit le modèle $NARCH(p)$ à valeurs dans $\mathbb{R}^d$ défini par :*

$$X_{t+1} = \psi_1(A_1 X_t + \cdots + A_p X_{t-p+1}) + \psi_2(B_1 X_t + \cdots + B_p X_{t-p+1})\varepsilon_{t+1},$$

où $\psi_1 : \mathbb{R}^d \to \mathbb{R}^d$, $\psi_1 : \mathbb{R}^d \to \mathcal{S}^+(d,\mathbb{R})$ et les $(A_i)_{1\le i\le p}$ et les $(B_i)_{1\le i\le p}$ sont des matrices de $\mathcal{M}_d(\mathbb{R})$ avec A_1 et B_1 inversibles.
On suppose que :

i) $\mu \ll \lambda_d$,
ii) $\lambda_d(D_{\psi_1,\psi_2}) = 0$,
iii) ψ_1, ψ_2 et ψ_2^* sont bornées sur les compacts,
iv) il existe $x_0 = (x_0^1, \cdots, x_0^p) \in \left(\mathbb{R}^d\right)^p$ tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue, où $P(x,dy)$ est la probabilité de transition du modèle surdimensionné.

Alors,

(a) le modèle surdimensionné admet une loi stationnaire,

(b) toute loi stationnaire est absolument continue par rapport à la mesure de Lebesgue sur $\left(\mathbb{R}^d\right)^p$

Démonstration : On reprend la même démarche que pour les modèles $NAR(p)$. □

Remarque : On pourrait évidemment traiter de même un modèle du type :

$$X_{t+1} = \psi_1(A_1 X_t + \cdots + A_p X_{t-p+1}) + \psi_2(B_1 X_t + \cdots + B_p X_{t-p+1})\varepsilon_{t+1}.$$

5.2.3 Application aux modèles $NGARCH(p,q)$ dans le cas Lebesgue

Application 5.3 *Soit le modèle $NGARCH(p,q)$ à valeurs dans $\mathbb{R}^d$ défini par :*

$$\begin{aligned} X_{t+1} &= \psi(A_1 X_t + \cdots + A_p X_{t-p+1} + A_{p+1}\varepsilon_t + \cdots + A_{p+q}\varepsilon_{t-q+1}) + \\ &\qquad \psi(B_1 X_t + \cdots + B_p X_{t-p+1} + B_{p+1}\varepsilon_t + \cdots + B_{p+q}\varepsilon_{t-q+1})\varepsilon_{t+1}, \end{aligned}$$

où $\psi_1 : \mathbb{R}^d \to \mathbb{R}^d$, $\psi_1 : \mathbb{R}^d \to \mathcal{S}^+(d,\mathbb{R})$ et les $(A_i)_{1\le i\le p}$ et les $(B_i)_{1\le i\le p}$ sont des matrices de $\mathcal{M}_d(\mathbb{R})$ avec A_1 et B_1 inversibles.
On suppose que :

i) $\mu \ll \lambda_d$,
ii) $\lambda_d(D_{\psi_1,\psi_2}) = 0$,
iii) ψ_1, ψ_2 et ψ_2^* sont bornées sur les compacts,
iv) il existe $x_0 = (x_0^1, \cdots, x_0^p) \in \left(\mathbb{R}^d\right)^p$ tel que $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$ est tendue, où $P(x,dy)$ est la probabilité de transition du modèle surdimensionné.

Alors,

(a) le modèle surdimensionné admet une loi stationnaire,

(b) toute loi stationnaire est absolument continue par rapport à la mesure de Lebesgue sur $\left(\mathbb{R}^d\right)^{p+q}$

5.3 Traduction des critères de tension

5.3.1 Une formulation des critères de Pakes et Hajek pour les modèles $ARCH(1)$

On va formuler les critères pour $V(x) := |x|$ mais ce n'est pas la seule fonction de Lyapounov possible.

Proposition 5.1 *On considère un modèle $ARCH(1)$ sur $\mathbb{R}^d$. On suppose que le bruit ε_1 a un moment d'ordre 1.*

Critère de Pakes *Sous les hypothèses :*

i) $\forall x \in \mathbb{R}^d, \quad |F(x)| + \|G(x)\| \, \mathbb{E}|\varepsilon_1| \leq |x| + \beta,$
ii) $\lim_{|x|\to\infty} \left(|F(x)| - |x| + \|G(x)\| \, \mathbb{E}|\varepsilon_1|\right) = -\infty.$

Alors pour tout x dans $\mathbb{R}^d$ la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy)\right)_{n\in\mathbb{N}^*}$ *est tendue.*

Critère de Hajek *Sous l'hypothèse :*

$$\forall x \in \mathbb{R}^d, \quad |F(x)| + \|G(x)\| \, \mathbb{E}|\varepsilon_1| \leq \alpha|x| + \beta \text{ avec } \alpha < 1.$$

Alors pour tout x dans $\mathbb{R}^d$ la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy)\right)_{n\in\mathbb{N}^*}$ *est tendue.*

5.3.2 Application aux modèles $ARCH(p)$ et $GARCH(p,q)$

Un modèle $ARCH(p)$ étant un $GARCH(p,0)$, nous énonçons directement le résultat pour les $GARCH(p,q)$.

Proposition 5.2 *On considère un modèle $GARCH(p,q)$ à valeurs dans $\mathbb{R}^d$. On suppose que ε_1 a un moment d'ordre 1.*

Critère de Pakes *S'il existe $p+q$ réels strictement positifs $(\alpha_i)_{1\leq i\leq p+q}$ vérifiant $\sum_{i=1}^{p+q} \alpha_i \leq 1$ tels que :*

i) $\forall (x^1,\cdots,x^{p+q}) \in \left(\mathbb{R}^d\right)^{p+q}, \quad |f(x^1,\cdots,x^{p+q})| + \|g(x^1,\cdots,x^{p+q})\| \, \mathbb{E}|\varepsilon_1| \leq \sum_{i=1}^{p+q} \alpha_i |x^i| + \beta,$

ii) $\lim_{\sup_{1\leq i\leq p+q} |x^i| \to \infty} \left(|f(x^1,\cdots,x^{p+q})| - \sum_{i=1}^{p+q} \alpha_i |x^i| + \|g(x^1,\cdots,x^{p+q})\| \, \mathbb{E}|\varepsilon_1| \right) = -\infty,$

Alors pour tout x dans $\left(\mathbb{R}^d\right)^{p+q}$ *la suite de probabilités* $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy)\right)_{n\in\mathbb{N}^*}$ *est tendue, où* $P(x,dy)$ *est la transition du modèle surdimensionné.*

Critère de Hajek *On suppose qu'il existe $p+q$ réels positifs $(\alpha_i)_{1\leq i\leq p+q}$ vérifiant* $\sum_{i=1}^{p+q}\alpha_i < 1$ *tels que :*

$$\forall (x^1,\cdots,x^{p+q})\in\left(\mathbb{R}^d\right)^{p+q},\quad |f(x^1,\cdots,x^{p+q})| + \|g(x^1,\cdots,x^{p+q})\|\ \mathbb{E}|\varepsilon_1| \leq \sum_{i=1}^{p+q}\alpha_i|x^i| + \beta.$$

Alors pour tout x dans $\left(\mathbb{R}^d\right)^{p+q}$ *la suite de probabilités* $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy)\right)_{n\in\mathbb{N}^*}$ *est tendue, où* $P(x,dy)$ *est la transition du modèle surdimensionné.*

Chapitre 6

Existence d'une loi stationnaire dans le cadre Has'minskii

Dans ce chapitre, nous allons travailler sans l'hypothèse de tension de $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0, dy)\right)_{n\in\mathbb{N}^*}$, d'où la spécificité des méthodes que nous allons développer maintenant. Cette séparation se répètera lors de l'étude de la stabilité et de la récurrence positive des modèles au chapitre 7.

Le critère de Has'minskii est un critère d'existence d'une loi stationnaire pour des chaînes de Markov fellériennes dont le noyau de transition, en tant qu'application fonctionnelle, possède la propriété de conserver les fonctions continues de limite nulle à l'infini.

Ici, nous avons besoin de supposer que l'espace polonais S métrisé par la distance d est, en outre, localement compact. On le suppose toujours muni de sa tribu borélienne $\mathcal{B}(S)$.

Définition 6.1 *On notera $C_0(S,\mathbb{R})$ l'ensemble des fonctions continues de S dans $\mathbb{R}$ telles que pour un $a \in S$ (donc pour tous) :*

$$\forall f \in C_0(S,\mathbb{R}), \qquad \lim_{d(x,a)\to+\infty} f(x) = 0.$$

Théorème 6.1 Critère de Has'minskii *(Ethier et Kurtz, 1986)*
Soit une chaîne Markov homogène de transition $P(x,dy)$ sur S. On suppose que :

i) $\forall f \in C_0(S,\mathbb{R})$, $Pf \in C_0(S,\mathbb{R})$,
ii) pour $a \in S$, il existe une fonction de Lyapounov V telle que :

$$\forall x \in S, \quad PV(x) \leq V(x) + c \quad et \quad \limsup_{d(x,a)\to+\infty} (PV - V)(x) < 0.$$

Alors il existe une loi stationnaire.

Remarques : • $C_0(S,\mathbb{R})$ n'existe que si S n'est pas compact. Cependant, on supposera toujours implicitement que c'est le cas puisque, sinon, toute suite de probabilités est automatiquement tendue sur S et le critère de Has'minskii devient inutile.
• Ce critère ressemble beaucoup au critère de Pakes, mais il permet de traiter des cas que le critère de Pakes ne permet pas de traiter et réciproquement. Ainsi, considérons le modèle $NLR(1)$ fellérien à valeurs dans $\mathbb{R}$:

$$X_{t+1} = F(X_t) + \varepsilon_{t+1},$$

où $F(x) := |x| - 2$ et $\varepsilon_1 \in [-1, 1]$.

On a $P|\,.\,|(x) = \mathbb{E}|F(x) + \varepsilon_1| \leq |x| + 3$, et $\limsup\limits_{|x| \to +\infty} P|\,.\,|(x) - |x| \leq -1$.

En outre, comme $\lim\limits_{|x| \to +\infty} F(x) = +\infty$, il est facile de voir, comme nous le verrons au paragraphe 6.3.1, que :

$$\forall g \in C_0(\mathbb{R}, \mathbb{R}), \; Pg \in C_0(\mathbb{R}, \mathbb{R}).$$

Le critère de Has'minskii est donc vérifié avec $V(x) := |x|$ pour ce modèle alors que le critère de Pakes ne l'est pas puisque $\limsup\limits_{|x| \to +\infty} P|\,.\,|(x) - |x| \geq -3$.

Si l'on considère maintenant le modèle $NLR(1)$ sur $\mathbb{R}$ avec $F := 0$ et ε_1 intégrable, alors le critère de Pakes est vérifié avec $|\,.\,|$ car F est évidemment bornée. D'autre part on a pour toute fonction $f \in C_0(\mathbb{R}, \mathbb{R})$, $Pf(x) = \mathbb{E}f(\varepsilon_1)$. La condition $i)$ n'est donc pas remplie si f est strictement positive et de limite nulle à l'infini.

L'inconvénient du critère de Has'minskii est qu'il ne s'appuie pas sur la tension sur S des suites de probabilités $\left(\frac{1}{n} \sum_{k=0}^{n-1} P^k(x, dy)\right)_{n \in \mathbb{N}^*}$. Il n'est donc pas utilisable en l'état. On peut cependant l'adapter à chacun de nos modèles en modifiant légèrement les hypothèses.

6.1 Existence d'une loi stationnaire

Nous allons donner, dans l'esprit du critère de Has'minskii, des résultats d'existence d'une loi stationnaire pour les modèles quasi-fellériens.

Théorème 6.2 *On considère une transition quasi-fellérienne $P(x, dy) = Q(H(x), dy)$. On se fixe $a \in S$ et on suppose que :*

i) pour toute fonction $f : S \to \mathbb{R}$ borélienne bornée telle que $\lim\limits_{d(x,a) \to +\infty} f(x) = 0$, on a $\lim\limits_{d(x,a) \to +\infty} Pf(x) = 0$,

ii) il existe une fonction de Lyapounov V telle que :

$$\forall x \in S, \quad PV(x) \leq V(x) + \beta \text{ et } \limsup_{d(x,a) \to +\infty} (PV - V)(x) < 0.$$

Alors le modèle admet une probabilité invariante.

Remarque : Le caractère quasi-fellérien du modèle est renforcé par la deuxième hypothèse pour avoir une généralisation de l'hypothèse $i)$ du critère de Has'minskii. Il s'agit en fait d'une hypothèse de continuité à l'infini.

Démonstration du théorème 6.2 : Nous allons intégrer les ingrédients propres à la démonstration du critère de Has'minskii (Ethier et Kurtz, 1986) à la démonstration initiale sur les modèles quasi-fellériens (théorème 3.1).

Étape 1 : On se place sur le compactifié d'Alexandrov $\widehat{S} := S \cup \{\infty\}$. On définit alors sur $\widehat{S}$ un nouveau noyau de tansition $\widehat{P}(x, dy)$ par :

$$\begin{cases} \widehat{P}(x, dy) = P(x, dy), \; x \neq \infty \\ \widehat{P}(\infty, dy) = \delta_\infty \end{cases}$$

Soit $x_0 \in S$. On considère la suite de probabilités $(\widehat{\nu}_n)_{n\in\mathbb{N}^*}$ sur $\widehat{S}$ définie pour tout n dans $\mathbb{N}^*$ par :

$$\widehat{\nu}_n(dy) := \frac{1}{n}\sum_{k=0}^{n-1}\widehat{P}^k(x_0, dy).$$

$(\widehat{\nu}_n)_{n\in\mathbb{N}^*}$ est nécessairement tendue car $\widehat{S}$ est compact (on notera en outre que $\widehat{\nu}_n(S) = 1$ pour tout n). Il existe donc au moins une valeur d'adhérence $\widehat{\nu}$.

Montrons que $\widehat{\nu}$ est une loi stationnaire pour la transition $\widehat{P}(x, dy)$.

Soit C un compact de S inclus dans D_H et soit $(h_p)_{p\in\mathbb{N}}$ la suite de fonctions définie par :

$$\forall p \in \mathbb{N}^*, \quad \forall x \in S, \quad h_p := 1 - \min\left(pd(x, {}^cC), 1\right).$$

Comme S est localement compact, h_p est donc dans $C_0(S,\mathbb{R})$, pour tout $p \in \mathbb{N}^*$. On a de plus $h_p \searrow \mathbf{1}_C$ et, pour tout p, $\|h_p\|_\infty \le 1$.

Soit alors la suite de fonctions de W dans $\mathbb{R}$, $(Qh_p)_{p\in\mathbb{N}}$. Les fonctions Qh_p sont continues car Q est fellérienne au sens de la définition 2.1.

D'autre part pour tout $w \in W$, $Q(w, dy)$ étant une probabilité, on a :

$$\forall\ w \in \bigcup_{K\in\mathcal{K}} \overline{H(K)},\ Qh_p(w) \searrow Q(w, C) = 0.$$

Soit K un compact de S. Comme $\overline{H(K)}$ est un compact de W, le théorème de Dini implique que la convergence est uniforme sur $\overline{H(K)}$:

$$\sup_{w\in\overline{H(K)}} Qh_p(w) \to 0, \text{ quand } p \to +\infty. \tag{6.1}$$

On a $\lim\limits_{d(x,a)\to+\infty} Ph_1(x) = 0$, car $h_1 \in C_0(S,\mathbb{R})$. D'où :

$$\forall\varepsilon > 0,\ \exists K_\varepsilon \text{ compact } /\ \forall x \in\ {}^cK_\varepsilon,\ Ph_1(x) \le \varepsilon.$$

Mais $(Ph_p)_{p\ge1}$ est une suite décroissante de fonctions, d'où :

$$\forall\varepsilon > 0,\ \exists K_\varepsilon \text{ compact } /\ \forall p \in \mathbb{N},\ \forall x \in\ {}^cK_\varepsilon,\ Ph_p(x) \le \varepsilon. \tag{6.2}$$

Or $Ph_p = Qh_p \circ H$, on déduit alors de (6.1) et (6.2) que :

$$\sup_{x\in S} Ph_p(x) \to 0, \text{ quand } p \to +\infty. \tag{6.3}$$

Soit $\left(\widehat{\nu}_{\varphi(n)}\right)_{n\in\mathbb{N}^*}$ une suite extraite de $(\widehat{\nu}_n)_{n\in\mathbb{N}^*}$ convergeant étroitement vers $\widehat{\nu}$ sur $\widehat{S}$. Comme $C_0(S,\mathbb{R}) \subset C(\widehat{S},\mathbb{R})$, on a :

$$\widehat{\nu}(h_p) = \lim_{n\to+\infty} \widehat{\nu}_{\varphi(n)}(h_p) = \lim_{n\to+\infty} \widehat{\nu}_{\varphi(n)}(\widehat{P}h_p) \le \sup_{x\in S} \widehat{P}h_p(x) = \sup_{x\in S} Ph_p(x).$$

Cette dernière propriété combinée avec (6.3) donne :

$$\widehat{\nu}(C) = \lim_{p\to\infty} \widehat{\nu}(h_p) = 0.$$

Comme cela est vrai pour tout compact C inclus dans D_H, la régularité intérieure de la probabilité $\widehat{\nu}$ (Billingsley, 1968, page 7) permet encore de conclure que $\widehat{\nu}(D_H) = 0$.

Soit h une fonction de $C_0(S,\mathbb{R})$. D'après ce qui précède, $\widehat{P}h$ est $\widehat{\nu}$-*p.s.* continue (car $\widehat{P}h$ est continue en ∞). On en déduit, comme pour le théorème 3.1, que $\widehat{\nu}\left(\widehat{P}h\right) = \widehat{\nu}\,(h)$ prouvant ainsi que $\widehat{\nu}$ est une loi stationnaire pour la transition $\widehat{P}(x,dy)$.

Étape 2 : Montrons maintenant que la trace $\widehat{\nu}$ sur S est non nulle.

Soit $g(x) := PV(x) - V(x)$. Comme par hypothèse $\limsup\limits_{d(x,a)\to+\infty} g(x) < 0$, il existe un réel $\alpha > 0$ et un compact K de S tels que :

$$\forall x \in^c K,\ \ g(x) < -\alpha.$$

On écrit alors pour tout $x \in S$ et pour tout $k \in \mathbb{N}$:

$$P^k g(x) \leq \beta P^k(\mathbf{1}_K)(x) - \alpha P^k(\mathbf{1}_{^cK})(x) \leq (\alpha+\beta)P^k(\mathbf{1}_K)(x) - \beta.$$

On somme alors ces équations sur k pour $x = x_0$ et on arrive à :

$$\widehat{\nu}_{n,x_0}(K) \geq \frac{1}{n(\alpha+\beta)}\sum_{k=0}^{n-1} P^k g(x_0) + \frac{\beta}{\alpha+\beta} \geq \frac{P^nV(x_0) - V(x_0)}{n(\alpha+\beta)} + \frac{\beta}{\alpha+\beta}.$$

V et P^nV étant positives, en passant à la limite sur n on obtient :

$$\widehat{\nu}_{x_0}(S) \geq \widehat{\nu}_{x_0}(K) \geq \frac{\beta}{\alpha+\beta}. \tag{6.4}$$

On vérifie ensuite aisément que $\nu := \dfrac{\widehat{\nu}_{x_0}(.\cap S)}{\widehat{\nu}_{x_0}(S)}$ est une probabilité invariante pour la transition initiale P (Ethier et Kurtz, 1986, page 272). □

6.2 Redéfinition du critère de Has'minskii et synthèse des résultats précédents

On peut remarquer qu'on a utilisé une condition ressemblant beaucoup au critère de Pakes et que nous appellerons désormais critère de Has'minskii :

Définition 6.2 Critère de Has'minskii
On considère un modèle markovien de transition $P(x,dy)$ sur S. On fixe $a \in S$ et on suppose que :

i) pour toute fonction $f : S \to \mathbb{R}$ borélienne bornée telle que $\lim\limits_{d(x,a)\to+\infty} f(x) = 0$, on a $\lim\limits_{d(x,a)\to+\infty} Pf(x) = 0$,

ii) il existe une fonction de Lyapounov V telle que :

$$\forall x \in S, \quad PV(x) \leq V(x) + \beta \ \textit{et} \ \limsup_{d(x,a)\to+\infty} (PV - V)(x) < 0.$$

On peut alors synthétiser tous les résultats obtenus concernant l'existence d'une loi stationnaire sous la forme :

Résumé des résultats d'existence d'une loi stationnaire :

Soit un modèle markovien de transition $P(x,dy)$ quasi-fellérienne à valeurs dans un espace S polonais. On suppose en outre que :

- (α) *Soit S est localement compact et le critère de Has'minskii est vérifié,*
- (β) *Soit il existe $x_0 \in S$ tel que la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x_0,dy)\right)_{n\in\mathbb{N}^*}$ est tendue.*

Alors il existe une loi stationnaire.

Il est donc important de traduire le critère de Has'minskii pour les modèles spécifiés, ce que nous allons faire maintenant.

6.3 Le critère de Has'minskii pour les modèles NLR

6.3.1 Une formulation du critère de Has'minskii pour les modèles $NLR(1)$

Proposition 6.1 *On considère un modèle $NLR(1)$ où ε_1 est de loi μ intégrable. On suppose que :*

i) $\lim\limits_{|x|\to+\infty} |F(x)| = +\infty$,
ii) $\forall x \in \mathbb{R}^d, \quad |F(x)| \leq |x| + \beta$,
iii) $\limsup\limits_{|x|\to+\infty} (|F(x)| - |x|) < -\mathbb{E}|\varepsilon_1|$.

Alors le modèle vérifie le critère de Has'minskii avec $V(x) := |x|$.

Démonstration : Il est clair que l'hypothèse *ii)* du critère de Has'minskii est vérifiée avec $V(x) := |x|$.

Par ailleurs, μ étant tendue, on peut écrire :

$$\forall \varepsilon > 0,\ \exists K_\varepsilon \text{ compact de } \mathbb{R}^d / \ \mu\left({}^cK_\varepsilon\right) < \varepsilon.$$

Soit f borélienne bornée telle que $\lim\limits_{|x|\to+\infty} f(x) = 0$. On a :

$$\begin{aligned} Pf(x) &= \int f\left(F(x)+y\right)\mu(dy) \\ &= \int_{{}^cK_\varepsilon} f\left(F(x)+y\right)\mu(dy) + \int_{K_\varepsilon} f\left(F(x)+y\right)\mu(dy) \\ &\leq \|f\|_\infty \mu\left({}^cK_\varepsilon\right) + \int_{K_\varepsilon} f\left(F(x)+y)\right)\mu(dy) \\ &\leq \varepsilon\|f\|_\infty + \int_{K_\varepsilon} f\left(F(x)+y)\right)\mu(dy). \end{aligned}$$

Comme $\lim\limits_{|x|\to\infty} F(x) = +\infty$, il est clair que $f\left(F(x)+y\right)$ converge uniformément vers 0 par rapport à y dans K_ε quand $|x|$ tend vers l'infini. On a donc :

$$\forall \varepsilon > 0,\ \limsup_{|x|\to+\infty} Pf(x) \leq \varepsilon\|f\|_\infty,$$

et, par suite, $\lim\limits_{|x|\to+\infty} Pf(x) = 0$. $\square$

6.3.2 Application aux modèles $NLR(p)$

Proposition 6.2 *On considère un modèle $NLR(p)$ où ε_1 est de loi μ intégrable. On suppose qu'il existe p réels strictement positifs $(\alpha_i)_{1\leq i\leq p}$ vérifiant $\sum_{i=1}^{p}\alpha_i \leq 1$ tels que :*

i) $\displaystyle\lim_{\sup_{1\leq i\leq p}|x^i|\to+\infty} |f(x^1,\cdots,x^p)| = +\infty,$

ii) $\forall(x^1,\cdots,x^p)\in\left(\mathbb{R}^d\right)^p,\quad |f(x^1,\cdots,x^p)| \leq \sum_{i=1}^{p}\alpha_i|x^i| + \beta,$

iii) $\displaystyle\limsup_{\sup_{1\leq i\leq p}|x^i|\to+\infty}\left(|f(x^1,\cdots,x^p)| - \sum_{i=1}^{p}\alpha_i|x^i|\right) < -\mathbb{E}|\varepsilon_1|.$

Alors le modèle surdimensionné vérifie le critère de Has'minskii avec $V(x) := |x|$..

6.4 Le critère de Has'minskii pour les modèles $ARCH$ et $GARCH$

6.4.1 Une formulation du critère de Has'minskii pour les modèles $ARCH(1)$

Proposition 6.3 *On considère un modèle $ARCH(1)$ où ε_1 est de loi μ intégrable. On suppose que :*

i) G est borélienne bornée sur $\mathbb{R}^d$ et $\displaystyle\lim_{|x|\to+\infty}|F(x)| = +\infty,$

ii) $\forall x\in\mathbb{R}^d,\quad |F(x)| + \|G(x)\|\ \mathbb{E}|\varepsilon_1| \leq |x| + \beta,$

iii) $\displaystyle\limsup_{|x|\to+\infty}\left(|F(x)| - |x| + \|G(x)\|\ \mathbb{E}|\varepsilon_1|\right) < 0.$

Alors le critère de Has'minskii est vérifié.

Démonstration : Comme $\lim_{|x|\to+\infty} F(x) = +\infty$, et que G est bornée sur $\mathbb{R}^d$, il est clair que pour tout compact K de $\mathbb{R}^d$, on a :

$$\lim_{|x|\to+\infty}\inf_{y\in K}|F(x)+G(x)y)| = +\infty.$$

A ce stade, on peut reproduit la démonstration du théorème 6.1. □

6.4.2 Application aux modèles $ARCH(p)$ et $GARCH(p,q)$

On traduit directement ce critère pour les modèles $GARCH(p,q)$, puisque les $ARCH(p)$ sont des $GARCH(p,0)$.

Proposition 6.4 *On considère un modèle $GARCH(p,q)$ où ε_1 est de loi μ intégrable. On suppose qu'il existe $p+q$ réels strictement positifs $(\alpha_i)_{1\leq i\leq p+q}$ vérifiant $\sum_{i=1}^{p+q}\alpha_i \leq 1$ tels que:*

i) g est bornée et $\displaystyle\lim_{\sup_{1\leq i\leq p+q}|x^i|\to+\infty}|f(x^1,\cdots,x^{p+q})| = +\infty,$

ii) $\forall (x^1, \cdots, x^{p+q}) \in \left(\mathbb{R}^d\right)^{p+q}, \quad |f(x^1, \cdots, x^{p+q})| + \|g(x^1, \cdots, x^{p+q})\| \ \mathbb{E}|\varepsilon_1| \leq \sum_{i=1}^{p+q} \alpha_i |x^i| + \beta,$

iii) $\limsup_{\sup_{1\leq i\leq p+q} |x^i| \to +\infty} \left(|f(x^1, \cdots, x^{p+q})| - \sum_{i=1}^{p+q} \alpha_i |x^i| + \|g(x^1, \cdots, x^{p+q})\| \ \mathbb{E}|\varepsilon_1| \right) < 0.$

Alors le critère de Has'minskii est vérifié par le modèle surdimensionné.

Chapitre 7

Stabilité et récurrence positive

Ce chapitre est consacré à l'obtention de lois fortes des grands nombres partant de n'importe quelle valeur initiale pour les chaînes précédemment étudiées, S désignant toujours une espace polonais. Deux types de propriétés peuvent être obtenus, la stabilité et la récurrence positive ou récurrence au sens de Harris.

Définition 7.1 (Duflo, 1997) *On dira qu'une chaîne de Markov homogène $(X_t)_{t\in\mathbb{N}}$ est stable s'il existe une probabilité ν telle que :*

$$\forall g : S \to \mathbb{R} \text{ continue bornée }, \ \forall x \in S, \quad \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Définition 7.2 *Une chaîne de Markov homogène $(X_t)_{t\in\mathbb{N}}$ est récurrente positive s'il existe une probabilité ν telle que :*

$$\forall g : S \to \mathbb{R} \text{ borélienne bornée}, \ \forall x \in S, \quad \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Remarques : • La ν-récurrence d'une chaîne entraîne sa ν-stabilité.
• Si une chaîne est stable ou récurrente positive et si ν est une loi stationnaire, ν est alors l'unique loi stationnaire.
• Si une chaîne est récurrente positive, alors ν est forcément une loi stationnaire et c'est donc l'unique loi stationnaire.
• De même, si une chaîne est stable et fellérienne alors ν est l'unique loi stationnaire. Cependant, si la chaîne n'est pas fellérienne, il est possible d'avoir la ν-stabilité sans que ν soit une loi stationnaire, comme l'illustre l'exemple ci-après :

Exemple 7.1 *Soit le modèle markovien homogène défini sur $\mathbb{R}$ par :*

$$X_{t+1} := F(X_t)\,\varepsilon_{t+1},$$

où les $(\varepsilon_t)_{t\in\mathbb{N}^*}$ sont i.i.d. de loi normale $\mathcal{N}(0,1)$ et F est la fonction de $\mathbb{R}$ dans $\mathbb{R}$ vérifiant :

$$\begin{cases} F(x) := x, \ x \neq 0, \\ F(0) := 1. \end{cases}$$

Ce modèle est stable pour la mesure de Dirac en 0, δ_0, i.e. :

$$\forall g : \mathbb{R} \to \mathbb{R} \text{ continue bornée }, \ \forall x \in \mathbb{R}, \quad \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} g(0),$$

alors que δ_0 n'est pas une loi stationnaire du modèle.

Démonstration : En calculant l'espérance sous $\mathbb{P}_x$ de $|X_n|$, on trouve :

$$\begin{cases} \mathbb{E}_x|X_n| := x\,(\mathbb{E}|\varepsilon_1|)^n = x\left(\sqrt{\dfrac{2}{\pi}}\right)^n, \ x \neq 0, \\ \mathbb{E}_0|X_n| := (\mathbb{E}|\varepsilon_1|)^n = \left(\sqrt{\dfrac{2}{\pi}}\right)^n. \end{cases}$$

On a donc : $\forall x \in \mathbb{R},\ \sum_{k=0}^{+\infty} \mathbb{E}_x|X_k| < +\infty$, ce qui prouve que : $\forall x \in \mathbb{R},\ X_n \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} 0.$

Le théorème de Césaro donne alors :

$$\forall g : \mathbb{R} \to \mathbb{R} \text{ continue bornée },\ \forall x \in \mathbb{R},\ \ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} g(0).$$

Or ce modèle n'admet aucune loi stationnaire. En effet, comme toutes les transitions sont équivalentes à la mesure de Lebesgue, une telle loi stationnaire ν serait équivalente à la mesure de Lebesgue (ce qui n'est, à l'évidence, pas le cas de la Dirac en 0) et donc unique. En outre (voir par exemple le théorème 7.4 ci-après) la chaîne serait alors récurrente positive de loi stationnaire $\nu \neq \delta_0$. Ceci est contradictoire avec le résultat obtenu. □

Remarque : On a vu que la ν-stabilité pour une loi ν non loi stationnaire n'est possible que si la chaîne n'est pas fellérienne. En fait, pour nos modèles, qui sont une extension des modèles fellériens, la ν-stabilité entraînera toujours que ν est une loi stationnaire. On pourrait effectivement vérifier sur l'exemple précédent, que pour une fonction continue g, Pg est généralement discontinue en 0, prouvant par là même que ce modèle ne rentre pas dans notre cadre d'étude.

Rappelons également le résultat élémentaire énoncé par exemple dans (Duflo, 1990, page 176)

Proposition 7.1 *Si une chaîne de Markov est stable et fortement fellérienne alors elle est récurrente positive.*

Enfin, on peut montrer à l'aide du théorème ergodique que la récurrence positive nous permet d'avoir la loi forte des grands nombres pour les fonctions intégrables sous ν, i.e. :

Proposition 7.2 *Si une chaîne de Markov $(X_t)_{t\in\mathbb{N}}$ à valeurs dans S est récurrente positive, alors :*

$$\forall g\ \in L^1(\nu),\ \forall x \in S,\ \ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Le point faible de l'approche que nous allons développer pour établir la stabilité et la récurrence positive est de ne fournir aucune indication de vitesse de convergence vers la loi stationnaire. Elle mérite néanmoins d'être développée car, d'une part fondée sur des méthodes d'analyse fonctionnelle, elle est particulièrement bien adaptée aux chaînes à valeurs dans $\mathbb{R}^d$, mais surtout, elle permet de traiter avec succès des modèles que l'approche classique basée sur l'existence d'un petit ensemble récurrent ne permet pas de traiter.

7.1 Stabilité des modèles quasi-fellériens

La stabilité est une notion purement topologique puisque l'on s'intéresse à des propriétés de loi des grands nombres pour des fonctions continues. On parle également souvent de récurrence topologique ou de récurrence dans les ouverts. A l'instar des méthodes de tension, l'approche par la stabilité est très facilement maniable pour les chaînes fellériennes, il s'agit en fait de la même approche mais en considérant une autre suite de mesures empiriques. Cette approche topologique est donc couramment utilisé dans le cas fellérien, on pourra ainsi se référer, par exemple, aux travaux de (Skorokhod, 1986), (Borovkov, 1990 et 1991) ou encore (Loskot et Rudnicki, 1995).

Notre onbjectif est ici d'étendre l'approche par la stabilité aux chaînes quasi-fellériennes. Rappelons, pour fixer les idées, le critère de stabilité dérivé du critère de Hajek énoncé notamment dans (Duflo, 1990, page 175).

Théorème 7.1 *Soit un modèle markovien homogène de transition $P(x, dy)$ sur $\mathbb{R}^d$. On suppose que :*

i) P est fellérienne,
ii) il existe une fonction de Lyapounov V telle que :

$$\forall x \in S,\ \ PV(x) \leq \alpha V(x) + \beta,\ \ \textit{avec } \alpha \in [0, 1[,$$

iii) il existe au plus une loi stationnaire.

Alors le modèle est stable.

Remarques : • $i)$ et $ii)$ entraînent l'existence d'une loi stationnaire.
• On peut assurer la condition $iii)$ par exemple par la notion d'indécomposabilité que nous aborderons plus loin, ou par l'hypothèse de mélange lipschitzien.
• Nous montrerons au chapitre 8 que, sous les hypothèses du théorème 7.1, on a en fait, pour tout $b < 1$ et pour toute fonction ν-*p.s.* continue g telle que $|g| \leq a(V^b + 1)$:

$$\forall x \in S,\ \ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Avertissement : Comme aux chapitres 3, 4 et 5, les questions relatives à la tension seront traitées à part, essentiellement parce qu'elles n'ont rien à voir avec le caractère fellérien ou non des modèles étudiés.

Théorème 7.2 *Soit un transition quasi-fellérienne $P(x, dy) = Q(H(x), dy)$. On suppose que :*

i) pour tout x de S, $(\Lambda_n(\omega, du))_{n\in\mathbb{N}} := \left(\frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_k(\omega)}(du)\right)_{n\in\mathbb{N}}$ est $\mathbb{P}_x(d\omega)$-p.s. tendue,
ii) il existe au plus une loi stationnaire.

Alors le modèle est stable.

Démonstration : On va montrer que $\mathbb{P}_x(d\omega)$-*p.s.* les valeurs d'adhérence de $(\Lambda_n(\omega, du))_{n\in\mathbb{N}}$ sont des lois stationnaires. L'existence d'au plus une loi stationnaire assurera alors la stabilité.

Soit un fermé C inclus dans D_H et soit $(h_p)_{p\in\mathbb{N}}$ la suite de fonctions continues de S dans $\mathbb{R}$ définie par :

$$\forall p \in \mathbb{N}^*,\ h_p := 1 - \min\left(pd\left(x,\ {}^cC\right), 1\right).$$

$h_p \searrow \mathbf{1}_C$ et vérifie en outre pour tout p, $\|h_p\|_\infty \le 1$. D'après la démonstration du théorème 3.1, on sait que les fonctions Qh_p convergent uniformément vers 0 sur tout compact. En considérant les compacts de la forme $\overline{H(K)}$, on en déduit que Ph_p converge uniformément vers 0 sur tout compact K de S.

On sait qu'il existe un ensemble dénombrable $\mathcal{G} := \{g_i,\ i \in \mathbb{N}\}$ de fonctions continues bornées de S dans $\mathbb{R}$ tel que deux probabilités égales sur $\mathcal{G}$ coïncident.

Définissons alors l'ensemble :

$$\widehat{\Omega} := \left\{ \omega \in \Omega / \left(\Lambda_n(\omega, du)\right)_{n\in\mathbb{N}} \text{ tendue, } \Lambda_n(\omega, h_p - Ph_p) \underset{n\to+\infty}{\longrightarrow} 0,\ \Lambda_n(\omega, g_i - Pg_i) \underset{n\to+\infty}{\longrightarrow} 0 \right\}.$$

Il est classique, d'après la théorie des martingales que, pour toute fonction g borélienne bornée :

$$\text{pour tout } x \text{ de } S, \quad \mathbb{P}_x(d\omega)\text{-}p.s., \quad \Lambda_n(\omega, g - Pg) \underset{n\to+\infty}{\longrightarrow} 0 .$$

L'hypothèse $i)$ permet alors d'affirmer que, pour tout x de S, $\mathbb{P}_x\left(\widehat{\Omega}\right) = 1$.

Soit maintenant $\omega \in \widehat{\Omega}$. Il existe une loi $\nu(\omega, du)$ et une sous-suite $\left(\Lambda_{\varphi(n)}(\omega, du)\right)$ de $(\Lambda_n(\omega, du))$ telles que :

$$\Lambda_{\varphi(n)}(\omega, du) \Longrightarrow \nu(\omega, du).$$

$\Lambda_{\varphi(n)}(\omega, du)$ converge étroitement, elle est donc en particulier tendue, d'où :

$$\forall \varepsilon > 0,\ \exists K_\varepsilon \text{ compact de } S \text{ tel que, } \forall n \in \mathbb{N}^*,\ \Lambda_{\varphi(n)}(\omega,\ {}^cK_\varepsilon) \le \varepsilon.$$

ε et K_ε étant donnés, on a pour tout $n \in \mathbb{N}^*$:

$$\Lambda_{\varphi(n)}(\omega, Ph_p) \le \|Ph_p\|_{K_\varepsilon} + \varepsilon\|h_p\|_\infty = \|Ph_p\|_{K_\varepsilon} + \varepsilon.$$

Choisissant maintenant p_ε tel que pour tout $p \ge p_\varepsilon$, $\|Ph_p\|_{K_\varepsilon} \le \varepsilon$, on trouve :

$$\forall p \ge p_\varepsilon,\ \forall n \in \mathbb{N}^*,\ \Lambda_{\varphi(n)}(\omega, Ph_p) \le 2\varepsilon. \tag{7.1}$$

On sait d'une part que $\Lambda_{\varphi(n)}(\omega, h_p - Ph_p) \longrightarrow 0$. D'autre part, comme h_p est continue, $\Lambda_{\varphi(n)}(\omega, h_p) \longrightarrow \nu(\omega, h_p)$. On tire alors de (7.1) que :

$$\forall p \ge p_\varepsilon,\ \nu(\omega, h_p) \le 2\varepsilon.$$

Passant à la limite sur p on obtient :

$$\nu(\omega, C) \le 2\varepsilon, \tag{7.2}$$

mais, ε étant arbitraire, on a en fait prouvé que $\nu(\omega, C) = 0$. La régularité intérieure de $\nu(\omega, du)$ permet encore de conclure que $\nu(\omega, D_H) = 0$.

Soit alors $g_i \in \mathcal{G}$. Comme Pg_i est $\nu(\omega, du)$-$p.s.$ continue, $\Lambda_{\varphi(n)}(\omega, Pg_i) \longrightarrow \nu(\omega, Pg_i)$. Donc $\nu(\omega, g_i - Pg_i) = 0$. Comme g_i est quelconque dans $\mathcal{G}$, on a en fait prouvé que $\nu(\omega, du)$ est une loi stationnaire.

Enfin, il existe au plus une loi stationnaire ν, et pour tout x de S, $\mathbb{P}_x\left(\widehat{\Omega}\right) = 1$. On déduit alors la stabilité :

$$\text{pour tout } x \text{ de } S, \quad \mathbb{P}_x(d\omega)\text{-}p.s.,\ \Lambda_n(\omega, du) \underset{n\to+\infty}{\Longrightarrow} \nu. \quad \square$$

Remarques : • Lors de l'étude de la récurrence positive, nous verrons qu'il existe bel et bien des modèles non fellériens stables mais non récurrents positifs, alors que la propriété de stabilité était originellement restreinte à des modèles fellériens (voir l'application 7.5).
• De manière analogue au cadre fellérien classique, on montre également qu'un modèle markovien quasi-fortement fellérien stable est récurrent positif. Nous verrons cependant, au paragraphe 7.6, que les modèles quasi-fortement fellériens sont récurrents positifs sous des hypothèses moins contraignantes que celle du théorème 7.2.

7.2 Unicité et indécomposabilité

Nous allons employer ici la notion d'indécomposabilité initialement introduite dans (Breiman, 1968). Cette notion est un critère d'unicité de la loi stationnaire, mais permet aussi de récupérer très rapidement la récurrence positive de certaines chaînes, notamment dans le cas où toutes les transitions sont équivalentes entre elles. Bien que proche de la notion d'irréductibilité, elle se révèle un peu plus faible et plus commode à manier.

Définition 7.3 (Breiman, 1968)
Une chaîne de Markov homogène de transition $P(x,dy)$ sur un espace mesurable $(E,\mathcal{E})$ est dite décomposable *s'il existe deux ensembles A_1, $A_2 \in \mathcal{E}$ disjoints non vides tels que :*

$$\forall x \in A_i,\ \forall i \in \{1,2\},\ P(x,A_i) = 1.$$

Dans le cas contraire, la chaîne est dite indécomposable.

Théorème 7.3 (Breiman, 1968)
Soit $(X_t)_{t\in\mathbb{N}}$ une chaîne de Markov homogène indécomposable de transition $P(x,dy)$. Alors,

(a) il existe au plus une loi stationnaire,
(b) s'il existe une loi stationnaire ν (elle est donc unique), la chaîne est ergodique sous $\mathbb{P}_\nu$. En particulier :

$$\forall g\ \in L^1(\nu),\ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_\nu\text{-}p.s.}{\longrightarrow} \nu(g).$$

On remarque que la notion d'indécomposabilité est l'extension naturelle de l'irréductibilité pour les chaînes de Markov à espace d'états dénombrable.

On note également que l'indécomposabilité apparaît comme un affaiblissement de la notion plus usuelle de ϕ-irréductibilité (Harris, 1956) et dont nous rappelons la définition ci-dessous.

Définition 7.4 *Soit une chaîne de Markov homogène de transition $P(x,dy)$ à valeurs dans S et ϕ une mesure positive non identiquement nulle. On dit que la chaîne est ϕ-irréductible si,*

$$\forall x \in S,\ \phi(dy) \ll \sum_{n=1}^{+\infty} a_n P^n(x,dy),$$

où les a_n sont strictement positifs et vérifient $\displaystyle\sum_{n=1}^{+\infty} a_n = 1.$

Proposition 7.3 *Si une chaîne de Markov homogène est ϕ-irréductible, alors elle est indécomposable.*

Démonstration : On considère deux ensembles A_1 et $A_2 \in \mathcal{E}$ tels que :

$$\forall x \in A_i,\ i \in \{1,2\},\ P(x, A_i) = 1.$$

Il est clair que :

$$\forall n \in \mathbb{N}^*,\ \forall x \in A_i,\ i \in \{1,2\},\ P^n(x, A_i) = 1,$$

ce qui est équivalent à :

$$\forall n \in \mathbb{N}^*,\ \forall x \in A_i,\ i \in \{1,2\},\ P^n(x, {}^c A_i) = 0.$$

La chaîne étant ϕ-irréductible, on a donc $\phi({}^c A_i) = 0,\ i \in \{1,2\}$. La mesure ϕ n'étant pas nulle, ${}^c A_1 \cup {}^c A_2 \neq E$, i.e. $A_1 \cap A_2 \neq \emptyset$. La chaîne est donc indécomposable. □

Enfin, il est clair qu'une chaîne récurrente positive de loi stationnaire ν est toujours ν-irréductible et donc indécomposable (Meyn et Tweedie, 1993). Nous en rappelons la preuve pour mémoire.

Proposition 7.4 *Soit $(X_t)_{t\in\mathbb{N}}$ une chaîne de Markov homogène de transition $P(x, dy)$ à valeurs dans S. Alors si la chaîne est ν-récurrente positive, elle est ν-irréductible.*

Démonstration : La chaîne étant récurrente positive, on a :

$$\forall A \in \mathcal{B}(S), \forall x \in S,\ \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_A(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(A).$$

La convergence ayant aussi lieu dans L^1 on a :

$$\forall A \in \mathcal{B}(S),\ \forall x \in S,\ \frac{1}{n} \sum_{k=0}^{n-1} P^k(x, A) \longrightarrow \nu(A). \tag{7.3}$$

Soit maintenant $A \in \mathcal{B}(S)$ tel que $\nu(A) > 0$. D'après (7.3), il existe forcément pour tout $x \in S$ un entier $k_0(x) > 1$ (et même une infinité) tel que $P^{k_0(x)}(x, A) > 0$.

Par suite, pour toute suite d'entiers strictement positifs $(a_n)_{n\in\mathbb{N}^*}$ vérifiant $\sum_{n=1}^{+\infty} a_n = 1$, on a $\sum_{n=1}^{+\infty} a_n P^n(x, A) > 0$, ce qui prouve la ν-irréductibilité puisque x est quelconque. □

En revanche, ce résultat est en général faux pour les chaînes ν-stables. Soit en effet la chaîne de Markov à valeurs dans $\mathbb{R}$ définie par :

$$X_{t+1} := \frac{1}{2}(X_t + \varepsilon_{t+1}),$$

où X_0 est une variable aléatoire réelle et les ε_t sont i.i.d. de loi $\mathcal{B}(\pm 1, \frac{1}{2})$.

On remarque alors que l'on peut décomposer la chaîne sur $\mathbb{Q}$ et $\mathbb{R} \setminus \mathbb{Q}$. Par ailleurs, ce modèle étant un mélange lipschitzien (voir le chapitre 8), il est donc ν-stable, de loi stationnaire $\nu := U\left[-\frac{1}{2}, \frac{1}{2}\right]$.

7.3 Récurrence positive dans le cadre abstrait

L'approche par l'indécomposabilité permet de démontrer aisément le théorème classique donnant la récurrence positive des chaînes dès qu'il existe une loi stationnaire et que les transitions sont toutes équivalentes entre elles. On a même le résultat plus général suivant :

Théorème 7.4 *Soit une chaîne de Markov homogène de transition $P(x, dy)$ et à valeurs dans S. On suppose que :*

i) il existe une loi ϕ et un entier $n_0 > 0$ tels que :

$$\forall x \in S,\ P^{n_0}(x, dy) = \theta(x, y)\ \phi(dy),$$

où $\theta : S \times S \to \mathbb{R}_+$ est borélienne et $\forall x \in S,\ \theta(x, y) > 0,\ \phi(dy) - p.p.$,

ii) il existe au moins une loi stationnaire ν.

Alors la chaîne est récurrente positive, i.e. :

$$\forall g\ \in L^1(\nu),\ \forall x \in S,\ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Démonstration : Il est clair que l'hypothèse $i)$ entraîne la ϕ-irréductibilité et partant l'indécomposabilité. En outre, toujours sous $i)$, si une loi stationnaire existe, elle est alors unique et équivalente à ϕ. ν est donc nécessairement l'unique loi stationnaire.

La chaîne est $\mathbb{P}_\nu$-ergodique d'après le théorème 7.3. Soit maintenant $g \in L^1(\nu)$ et l'ensemble A_g défini par :

$$A_g := \left\{\omega :\ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k)(\omega) \longrightarrow \nu(g)\right\}.$$

On a évidemment $\mathbb{P}_\nu(A_g) = 1$ car la chaîne est $\mathbb{P}_\nu$-ergodique, ce qui signifie en particulier que :

$$\nu(dy)\text{-}p.s., \quad \mathbb{P}_y(A_g) = 1. \tag{7.4}$$

Soit maintenant $x \in S$. Il est clair que A_g est invariant pour l'opérateur de décalage sur l'espace canonique car la convergence ne dépend pas des premiers termes, ce qui implique que :

$$\mathbb{P}_x(A_g) = \int P^{n_0}(x, dy)\mathbb{P}_y(A_g). \tag{7.5}$$

Mais $P^{n_0}(x, dy) \sim \phi(dy) \sim \nu(dy)$, ce qui donne, combiné avec (7.4) :

$$P^{n_0}(x, dy)\text{-}p.s., \quad \mathbb{P}_y(A_g) = 1.$$

Cette dernière propriété remplaçée dans (7.5) fournit :

$$\mathbb{P}_x(A_g) = 1.$$

x et g étant quelconques on a donc établi la récurrence positive. □

7.4 Récurrence positive dans le cadre quasi-fellérien standard

Nous définissons le cadre standard comme celui où pour tout x de S, la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n\in\mathbb{N}}$ est tendue par opposition au cadre Has'minskii que nous traitons ci-après.

7.4.1 Résultat principal

Dans ce paragraphe, nous allons généraliser les critères de récurrence positive des modèles fellériens que l'on peut trouver dans (Meyn et Tweedie, 1993).

Si une chaîne de Markov homogène est ϕ-irréductible, on peut démontrer (Meyn et Tweedie, 1993) qu'il existe toujours une mesure maximale d'irréductibilité ψ, c'est-à-dire, telle que toute mesure d'irréductibilité est absolument continue par rapport à ψ.

Dans la suite, on ne travaillera qu'avec une mesure maximale d'irréductibilité ψ à partir de laquelle on définit les "petits ensembles" :

Définition 7.5 *Soit une chaîne de Markov homogène de transition $P(x, dy)$ ψ-irréductible. On dira que $A \in \mathcal{B}(S)$ est un petit ensemble s'il est non vide et vérifie :*

$$\inf_{x\in A} K_a(x, dy) \geq \delta\psi(dy),$$

où $K_a(x, dy) := \sum_{k=1}^{+\infty} a_k P^k(x, dy)$ *et* $\sum_{k=1}^{+\infty} a_k = 1,\ \forall k \geq 1,\ a_k > 0,\ \delta > 0.$

Remarques : • $K_aK_b(x, dy) := \int K_a(x, dz)K_a(z, dy) = K_{a*b}(x, dy)$.

• La notion de petit ensemble ne dépend que de $P(x, dy)$ et pas de la suite $(a_n)_{n\in\mathbb{N}^*}$ choisie.

• Tout sous-ensemble non vide d'un petit ensemble est lui-même un petit ensemble.

On peut établir pour les chaînes fellériennes ψ-irréductibles que dès que le support de ψ est d'intérieur non vide, il existe un petit ensemble ouvert chargé par ψ (Meyn et Tweedie, 1993, page 136).

A partir de ce petit ensemble ouvert et sous l'hypothèse supplémentaire de tension, pour tout x dans S, de la suite de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n\in\mathbb{N}}$, il est facile de montrer la récurrence positive (Meyn et Tweedie, 1993, page 208).

Notre but est d'étendre ce résultat aux chaînes quasi-fellériennes. Pour ce faire, nous procédons en deux étapes :

_ Nous montrons d'abord dans la proposition 7.5 que si, une chaîne quasi-fellérienne de transition $P(x, dy) = Q(H(x), dy)$ étant donnée, il existe un ouvert O tel $O \cap (S\backslash D_H)$ soit un petit ensemble chargé par ψ et sous l'hypothèse de tension, pour tout x dans S, de $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n\in\mathbb{N}}$ alors la chaîne est récurrente positive.

Nous montrons ensuite, en adaptant les méthodes de Meyn et Tweedie, l'existence d'un tel ouvert O dès que la chaîne quasi-fellérienne est ψ-irréductible avec $\operatorname{supp}\{\psi\}$ d'intérieur non vide.

Ces deux résultats conduisent au théorème suivant donnant des conditions pour assurer la récurrence positive d'un modèle quasi-fellérien.

Théorème 7.5 *On considère une chaîne de Markov homogène quasi-fellérienne de transition $P(x, dy) = Q(H(x), dy)$. On suppose que :*

i) $\forall x \in S,\ \left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n\in\mathbb{N}}$ *est tendue,*

ii) la chaîne est ψ-irréductible et $\operatorname{supp}\{\psi\}$ *est d'intérieur non vide.*

Alors le modèle est récurrent positif.

Démonstration : Cette démonstration repose sur les deux propositions suivantes :

Proposition 7.5 *Soit une chaîne de Markov homogène de transition $P(x,dy) = Q(H(x),dy)$ quasi-fellérienne. On suppose que :*

i) il existe une loi stationnaire ν et :

$$\forall x \in S,\ \frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy) \Longrightarrow \nu,$$

ii) P est ψ-irréductible

iii) il existe un ouvert O tel que $O \cap (S \backslash D_H)$ soit un petit ensemble chargé par ψ.

Alors la chaîne est récurrente positive.

Proposition 7.6 *Soit une chaîne de Markov homogène de transition $P(x,dy) = Q(H(x),dy)$ quasi-fellérienne. On suppose que :*

i) il existe une unique loi stationnaire ν,

ii) P est ψ-irréductible avec $\mathrm{supp}\{\psi\}$ *d'intérieur non vide.*

Alors il existe un ouvert O tel que $O \cap (S \backslash D_H)$ soit un petit ensemble chargé par ψ.

Démonstration de la proposition 7.5 : Soit $g \in L^1(\nu)$ et l'ensemble :

$$A_g := \left\{\omega :\ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k)(\omega) \longrightarrow \nu(g)\right\}.$$

La chaîne étant $\mathbb{P}_\nu$-ergodique car irréductible, on a $\mathbb{P}_\nu(A_g) = 1$. Par suite, si l'on note $R_g := \{x \in S / \ \mathbb{P}_x(A_g) = 1\}$, on a alors $\nu(R_g) = 1$. Comme $\psi \ll \nu$, on a $\psi(R_g) > 0$.

Soit $x \in O \cap (S \backslash D_H)$. Comme A_g est invariant pour l'opérateur de décalage sur l'espace canonique, on a :

$$\forall k \geq 0, \quad \mathbb{P}_x(A_g) = \int P^k(x,dy)\mathbb{P}_y(A_g),$$

et donc :

$$\begin{aligned}
\mathbb{P}_x(A_g) = \int K_a(x,dy)\mathbb{P}_y(A_g) &\geq \int_{R_g} K_a(x,dy)\mathbb{P}_y(A_g) \\
&\geq \delta \int_{R_g} \psi(dy)\mathbb{P}_y(A_g) \\
&= \delta\psi(R_g) > 0. \qquad (7.6)
\end{aligned}$$

Soit maintenant $x \in S$. Il vient d'autre part :

$$\begin{aligned}
\mathbb{P}_x(A_g) &= \frac{1}{n}\sum_{k=1}^{n-1} \int P^k(x,dy)\mathbb{P}_y(A_g) \\
&\geq \frac{1}{n}\sum_{k=1}^{n-1} \int_{O\cap(S\backslash D_H)} P^k(x,dy)\mathbb{P}_y(A_g) \\
&\geq \left(\frac{1}{n}\sum_{k=1}^{n-1} P^k\left(x, O \cap (S\backslash D_H)\right)\right) \times \delta\psi(R_g). \qquad (7.7)
\end{aligned}$$

Par construction des modèles quasi-fellériens, on a pour tout $k \geq 1$, $P^k(x, D_H) = 0$ et donc $P^k(x, O \cap (S \backslash D_H)) = P^k(x, O)$.

Par ailleurs, $\liminf_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n-1} P^k(x, O) \geq \nu(O) > 0$, car $\psi(O \cap (S \backslash D_H)) > 0$.

On tire alors de (7.7) :

$$\forall x \in S, \ \mathbb{P}_x(A_g) \geq \delta\psi(R_g) \times \nu(O) > 0. \tag{7.8}$$

D'après la propriété de Markov, le processus $M_k := \mathbb{P}_{X_k}(A_g) = \mathbb{E}_{X_k}(\mathbf{1}_{A_g})$ est une martingale car A_g est invariant pour l'opérateur de décalage sur l'espace canonique. Par suite :

$$\forall x \in S, \ \mathbb{P}_{X_k}(A_g) \stackrel{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \mathbf{1}_{A_g}.$$

Il est alors clair via (7.8) que :

$$\forall x \in S, \ \mathbf{1}_{A_g} > 0 \quad \mathbb{P}_x\text{-}p.s. \quad \text{i.e.} \quad \mathbf{1}_{A_g} = 1 \quad \mathbb{P}_x\text{-}p.s.$$

On a alors :

$$\forall x \in \mathbb{R}^d, \quad \mathbb{P}_x(A_g) = 1,$$

ce qui est exactement la propriété de récurrence positive puisque toute fonction borélienne bornée est dans $L^1(\nu)$. □

Démonstration de la proposition 7.6 : A son tour, celle-ci repose sur plusieurs lemmes.

Lemme 7.1 *Soit une transition quasi-fellérienne $P(x, dy) = Q(H(x), dy)$. Alors :*

(a) pour toute fonction $f : S \to \mathbb{R}$ bornée, continue sur $S \backslash D_H$, pour tout $k \in \mathbb{N}^*$, $P^k f$ est continue sur $S \backslash D_H$.
(b) $\forall F$ fermé, $\forall k \in \mathbb{N}^*$, $x \mapsto P^k(x, F)$ est semi-continue supérieurement sur $S \backslash D_H$,
(c) $\forall O$ ouvert, $\forall k \in \mathbb{N}^*$, $x \mapsto P^k(x, O)$ est semi-continue inférieurement sur $S \backslash D_H$.

Démonstration : Montrons que, si f est s.c.i sur $S \backslash D_H$, Pf est s.c.i sur $S \backslash D_H$.

Soit une fonction f s.c.i. sur $S \backslash D_H$. La fonction f est donc limite croissante sur $S \backslash D_H$ d'une suite de fonctions $(g_p)_{p \in \mathbb{N}}$ lipschitziennes bornées sur $S \backslash D_H$.

Lors de la démonstration du théorème 3.1, on montre que pour toute fonction $g : S \to \mathbb{R}$ continue bornée, Pg est continue sur $S \backslash D_H$.

On a donc :

$$Pg(x_n) \to Pg(x) \quad \text{dès que} \quad x_n \to x.$$

On peut alors écrire :

$$\forall x \in S \backslash D_H, P(x_n, dy) \Rightarrow P(x, dy) \text{ si } x_n \to x.$$

La convergence étroite précédente nous permet d'écrire que si g est une fonction $P(x, dy)$-$p.s.$ continue, on a encore :

$$Pg(x_n) \to Pg(x) \text{ si } x_n \to x. \tag{7.9}$$

Or, par hypothèse $P(x, D_H) = 0$ donc si f est une fonction continue sur $S \backslash D_H$ et bornée, Pf est, d'après ce qui précède, $P(x, dy)$-$p.s.$ continue.

Donc pour tout $p \in \mathbb{N}$, Pg_p est continue sur $S \backslash D_H$. La fonction $x \mapsto Pf(x)$ est donc s.c.i sur $S \backslash D_H$ puisque limite croissante de fonctions continues sur $S \backslash D_H$. Le point (a) est donc vérifié.

Les deux autres points en découlent alors facilement. $\square$

Le lemme suivant et la fin de la démonstration de la proposition 7.6 sont des extensions aux chaînes quasi-fellériennes du lemme **6.2.7** et de la proposition **6.2.8** de (Meyn et Tweedie, 1993, pages 135 et 136).

Lemme 7.2 *Soit une transition quasi-fellérienne $P(x, dy) = Q(H(x), dy)$ et soit A un borélien tel que $A \cap (S \backslash D_H)$ est un petit ensemble pour P. Alors $\overline{A} \cap (S \backslash D_H)$ est un petit ensemble pour P.*

Démonstration : Par hypothèse $\inf_{x \in A \cap (S \backslash D_H)} K_a(x, dy) \geq \delta \psi(dy)$, avec $\delta > 0$.

Soit $B \in \mathcal{B}(S)$ et soit $F \subset B$ un fermé. Le lemme 7.1 nous permet d'affirmer que la fonction $x \mapsto K_a(x, F)$ est une limite uniforme de fonctions s.c.s. en tout point de $S \backslash D_H$. Elle est donc elle-même s.c.s. en tout point de $S \backslash D_H$, d'où :

$$\inf_{x \in \overline{A} \cap (S \backslash D_H)} K_a(x, F) = \inf_{x \in A \cap (S \backslash D_H)} K_a(x, F).$$

F étant inclus dans B, on en déduit que :

$$\inf_{x \in \overline{A} \cap (S \backslash D_H)} K_a(x, B) \geq \delta \psi(F),$$

et comme la mesure ψ est intérieurement régulière,

$$\inf_{x \in \overline{A} \cap (S \backslash D_H)} K_a(x, B) \geq \delta \psi(B),$$

ce qui prouve que $\overline{A} \cap (S \backslash D_H)$ est un petit ensemble. $\square$

Fin de la démonstration de la proposition 7.6 : La chaîne étant ψ-irréductible, il existe un petit ensemble A chargé par ψ (Meyn et Tweedie, 1993, page 133).

Définissons pour tout $k \in \mathbb{N}^*$ les ensembles :

$$A_k := \left\{ x \in S /\ K_a(x, A) > \frac{1}{k} \right\}.$$

On a clairement :

$$\bigcup_{k \in \mathbb{N}^*} A_k = \{x \in S /\ K_a(x, A) > 0\} = S \quad \text{car} \quad \psi(A) > 0.$$

Il existe donc $k_0 \in \mathbb{N}^*$ tel que pour tout $k \geq k_0$, $A_k \neq \emptyset$. Soit $k \geq k_0$ et $x \in A_k$, on a

$$K_{a*a}(x, dy) = \int K_a(x, dz) K_a(z, dy) \geq \int_A K_a(x, dz) K_a(z, dy) \geq \delta \psi(dy) \times K_a(x, A) \geq \frac{\delta}{k} \psi(dy),$$

donc pour tout $k \geq k_0$, A_k est un petit ensemble chargé par ψ.

Soit $O := \widehat{\mathrm{supp}\{\psi\}}^{\circ}$. O est non vide par hypothèse et donc chargé par ψ d'où, pour tout $k \geq k_0$, $A_k \cap O$ est chargé par ψ. $A_k \cap O$ étant par ailleurs un petit ensemble car inclus dans le petit ensemble A_k, $A_k \cap O$ est donc un petit ensemble chargé par ψ. En outre, d'après la proposition 3.1, $\nu(D_H) = 0$, d'où $\psi(D_H) = 0$ car $\psi \ll \nu$. Donc pour tout $k \geq k_0$, $A_k \cap O \cap (S \backslash D_H)$ est un petit ensemble chargé par ψ.

Le lemme 7.2 permet alors d'affirmer que, pour tout $k \geq k_0$, les ensembles :

$$B_k := \overline{A_k \cap O} \cap (S \backslash D_H),$$

sont des petits ensembles chargés par ψ.

Posons, pour tout $k \geq k_0$, $F_k := \overline{A_k \cap O}$. Les F_k sont fermés et vérifient $\bigcup_{k \geq k_0} F_k = \overline{O}$. L'intérieur de la réunion des F_k est donc non vide. Le théorème de Baire implique alors que l'intérieur d'au moins l'un des F_k est non vide.

Soit F_{k_1} $(k_1 \geq k_0)$ l'un des fermés d'intérieur non vide et O_{k_1} son intérieur. $O_{k_1} \cap (S \backslash D_H)$ est chargé par ψ puisque $O_{k_1} \subset \mathrm{supp}\{\psi\}$ et c'est un petit ensemble puisque inclus dans B_{k_1}. □

7.4.2 Application aux modèles $ARCH(1)$

Application 7.1 *On considère un modèle $ARCH(1)$ et on suppose que :*

i) $\mu \sim \lambda_d$,
ii) $\lambda_d(D_{F,G}) = 0$,
iii) F, G et G^* sont bornées sur les compacts,
iv) $\forall x \in \mathbb{R}^d$, $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n \in \mathbb{N}}$ *est tendue.*

Alors, le modèle est récurrent positif.

Démonstration : Il s'agit d'un corollaire immédiat du théorème 7.5 puisque l'hypothèse *ii)* induit que la transition est λ_d-irréductible. □

Il n'est pas nécessaire que le modèle $ARCH(1)$ ait un bruit de densité équivalente à la mesure de Lebesgue pour assurer sa récurrence positive. Cependant, l'hypothèse d'irréductibilité pour une loi quelconque est très difficile à traduire.

Il n'est cependant pas obligatoire que la loi du bruit soit équivalente à la mesure de Lebesgue pour que la chaîne soit ψ-irréductible avec une loi ψ possédant une densité par rapport à la mesure de Lebesgue, comme le montre l'exemple suivant :

Exemple 7.2 *On considère le modèle $NLR(1)$ à valeurs dans $\mathbb{R}$ où :*

i) F est définie par :

$$\begin{cases} F(x) := \frac{x}{2}, \ x > 1, \\ F(x) := f(x), \ -1 \leq x \leq 1, \\ F(x) := \frac{x}{2}, \ x < -1, \end{cases}$$

où f est une fonction Riemann-intégrable de $[-1,1]$ dans lui-même.

ii) la loi μ des ε_t est porté par l'intervalle $[-2,2]$ sur lequel elle a une densité strictement positive par rapport à la mesure de Lebesgue.

Alors le modèle est récurrent positif.

Démonstration : Il est facile de vérifier le critère de Hajek pour ce modèle. D'autre part F est telle que $\lambda(D_F) = 0$. Enfin, il est clair que la transition est irréductible par rapport à la mesure de Lebesgue sur $[-1,1]$. □

7.4.3 Application à des modèles à valeurs dans un sous-ensemble de $\mathbb{R}^d$

Comme nous l'avons déjà noté, la bonne spécification de l'espace des états est importante. Ainsi, l'application 4.2 fournit l'exemple d'une chaîne à valeurs dans un sous-ensemble S de $\mathbb{R}^d$ qui est quasi-fellérienne sur S sans l'être sur $\mathbb{R}^d$.

Dans le même esprit, il peut arriver que la chaîne soit ψ-irréductible et que le support de ψ soit d'intérieur vide sur $\mathbb{R}^d$ et non vide sur S.

A cet effet, reprenons l'application 4.2 en la modifiant légèrement.

Application 7.2 *Soit la fonction $F : \mathbb{R}^2 \to \mathbb{R}^2$ définie par :*

$$F(x^1, x^2) := \left(\begin{array}{c} f(x^1) \\ \frac{1}{4}([x^1] + [x^2]) \end{array} \right),$$

où $[\,.\,]$ désigne la partie entière et f est une fonction Riemann-intégrable vérifiant, pour tout $x^1 \in \mathbb{R}$, $|f(x^1| \leq \alpha |x^1| + \beta$, avec $\alpha < 1$.

Soit $(\varepsilon_t^1)_{t\in\mathbb{N}}$ une suite de variables i.i.d. de loi μ équivalente à la mesure de Lebesgue λ_1 sur $\mathbb{R}$ et possédant un moment d'ordre 1.

Soit également $(\varepsilon_t^2)_{t\in\mathbb{N}}$ une suite de variables i.i.d. indépendante de $(\varepsilon_t^1)_{t\in\mathbb{N}}$ de loi $\sum_{n\in\mathbb{Z}} a_n \delta_n$, avec, pour tout $n \in \mathbb{Z}$, $a_n > 0$, $\sum_{n\in\mathbb{Z}} a_n = 1$ et $\sum_{n\in\mathbb{Z}} |n| a_n < +\infty$ (e.g. ε_1^2 a un moment d'ordre 1).

Alors, le modèle $NLR(1)$ à valeurs dans $\mathbb{R}^2$ défini par :

$$X_{t+1} := F(X_t) + \left(\begin{array}{c} \varepsilon_{t+1}^1 \\ \varepsilon_{t+1}^2 \end{array} \right),$$

est récurrent positif.

Démonstration : On se place toujours sur l'espace (métrique séparable complet) $S := \mathbb{R} \times \frac{\mathbb{Z}}{4}$, puisque dès la date $t = 1$, la chaîne est à valeurs dans S.

L'application 4.2 permet d'affirmer que ce modèle est quasi-fellérien sur S et la tension des lois empiriques est assurée par le critère de Hajek.

Enfin, il est facile de voir que ce modèle est $\psi := \lambda_1 \otimes \left(\sum_{n\in\mathbb{Z}} a_n \delta_{\frac{n}{4}} \right)$-irréductible et que son support est d'intérieur non vide dans S.

On est alors sous les conditions d'application du théorème 7.5. □

Il est clair que ψ est de support d'intérieur vide sur $\mathbb{R}^2$ ce qui illustre, ici encore, la nécessité de bien spécifier l'espace des états.

7.4.4 Exemple d'un modèle quasi-fellérien récurrent positif et non quasi-fortement fellérien

En fait, les deux modèles précédents sont quasi-fortement fellériens. Il est donc utile de donner un exemple de chaîne quasi-fellérienne qui ne soit pas quasi-fortement fellérienne et pour laquelle la condition de ψ-irréductibilité avec $\mathrm{supp}\{\psi\}$ d'intérieur non vide soit absolument indispensable. En effet, comme nous le verrons plus loin, la ψ-irréductibilité sans se soucier du support de ψ est suffisante pour les modèles quasi-fortement fellériens.

Application 7.3 *Soit la transition sur* $[0,1]$ *définie par :*

$$\forall x \in [0,1], \quad P(x,dy) := (1-H(x)) \sum_{n=0}^{+\infty} a_n \delta_{r_n}(dy) + H(x)\delta_{\frac{H(x)}{2}}(dy),$$

où H *est borélienne de* $[0,1]$ *dans* $[0,1[$, δ_x *désigne la mesure de Dirac en* x, $(r_n)_{n\geq 0}$ *désigne une numérotation des rationnels de l'intervalle* $[0,1]$ *et, pour tout* $n \in \mathbb{N}$, *les* a_n *sont positifs ou nuls et vérifient* $\sum_{n=0}^{+\infty} a_n = 1$.
Si en outre :

i) $\overline{\{r_n / \ a_n > 0\}}$ *est d'intérieur non vide,*
ii) l'ensemble D_H *des points de discontinuité de* H *vérifie :*

$$D_H \subset \left(\mathbb{R} \setminus \mathbb{Q} \cap \left]\frac{1}{2}, 1\right[\right), \tag{7.10}$$

alors, le modèle est récurrent positif.

Démonstration : Si f est une fonction borélienne bornée sur $[0,1]$, on a :

$$Pf(x) := (1-H(x)) \sum_{n=0}^{+\infty} a_n f(r_n) + H(x) f\left(\frac{H(x)}{2}\right).$$

En posant naturellement

$$\forall w \in [0,1], \quad Q(w,dy) := (1-w) \sum_{n=0}^{+\infty} a_n \delta_{r_n}(dy) + w\delta_{\frac{w}{2}}(dy),$$

on a $P(x,dy) = Q\left(H(x),dy\right)$, où $Q(w,dy)$ est fellérienne.
Comme l'hypothèse (7.10) implique :

$$\forall w \in [0,1], \quad Q(w, D_H) = 0,$$

le modèle est donc quasi-fellérien.

Par ailleurs, pour tout x de $[0,1]$, $H(x) \neq 1$. La transition est donc ψ-irréductible avec $\psi := \sum_{n=0}^{+\infty} a_n \delta_{r_n}$. En outre, l'hypothèse $i)$ implique que le support de ψ est d'intérieur non vide.
La tension des lois empiriques est évidente puisque la chaîne est à valeurs dans le compact $[0,1]$. Le théorème 7.5 permet donc d'affirmer que ce modèle est récurrent positif.

Enfin, il est clair que P n'est pas quasi-fortement fellérienne. $\square$

Remarque : Il est très difficile d'identifier la loi stationnaire puisque que celle-ci dépend du choix de la fonction H.

7.4.5 Sur la nécessité de l'hypothèse concernant l'intérieur du support de la mesure d'irréductibilité

Étudions en détail l'exemple cité dans (Meyn et Tweedie, 1993, page 136) à ce propos.

Application 7.4 *La transition sur* $[0,1]$ *définie par :*

$$\forall x \in [0,1], \quad P(x, dy) := x\delta_0(dy) + (1-x)\delta_{\frac{x}{2}}(dy),$$

où δ_x *désigne encore la mesure de Dirac au point* x *est fellérienne non récurrente positive (mais elle est stable).*

Démonstration : Pour toute fonction continue bornée $f : [0,1] \to [0,1]$, on a :

$$\forall x \in [0,1], \quad Pf(x) = xf(0) + (1-x)f\left(\frac{x}{2}\right).$$

Cette chaîne est fellérienne et à valeurs dans le compact $[0,1]$. Enfin $\psi := \delta_0$ est une mesure maximale d'irréductibilité, mais évidemment le support de δ_0 est d'intérieur vide.

Calculons maintenant les itérées $P^k f_0$ de la fonction borélienne $f_0 := \mathbf{1}_{\mathbb{Q}\cap]0,1[}(x)$.

$$\text{pour } k=1, \quad \begin{cases} Pf_0 = (1-x)f_0\left(\dfrac{x}{2}\right), \ \forall x \in]0,1], \\ Pf_0(0) = Pf_0(1) = 0. \end{cases}$$

et il est facile de voir que, plus généralement, pour tout $k \in \mathbb{N}^*$:

$$\begin{cases} P^k f_0 = \displaystyle\prod_{i=0}^{k-1}\left(1-\frac{x}{2^i}\right) f_0\left(\frac{x}{2^k}\right), \ \forall x \in]0,1], \\ P^k f_0(0) = P^k f_0(1) = 0. \end{cases}$$

Pour tout $x \in]0,1[$, on pose $a(x) := \displaystyle\prod_{i=0}^{+\infty}\left(1-\frac{x}{2^i}\right) > 0$. On trouve par conséquent :

$$\begin{cases} \displaystyle\lim_{n\to+\infty}\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k f_0(x)\right) = a(x) > 0, \ \forall x \in \mathbb{Q}\cap]0,1[, \\ \displaystyle\lim_{n\to+\infty}\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k f_0(0)\right) = \lim_{n\to+\infty}\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k f_0(1)\right) = 0. \end{cases}$$

La chaîne n'est donc pas récurrente. En revanche, étant fellérienne, irréductible et de loi stationnaire δ_0, elle est δ_0-stable. □

Remarque : Cet exemple est en fait un cas particulier du modèle cité dans l'application 7.3 avec $H(x) = x$ et tous les a_n nuls sauf pour le point 0 où $a := 1$. Cela prouve donc que l'hypothèse *i)* de l'application 7.3, qui permet d'assurer que le support de ψ est d'intérieur non vide, est bien nécessaire.

7.4.6 Stabilité sans récurrence positive pour un modèle non fellérien

Application 7.5 *A partir du modèle précédent, la transition sur* $[0,1]$ *définie par :*

$$\forall x \in [0,1], \quad P(x, dy) := H(x)\delta_0(dy) + (1 - H(x))\, \delta_{\frac{H(x)}{2}}(dy),$$

où $H\left(\frac{\pi}{4}\right) = 0$ *et, pour tout* $x \neq \frac{\pi}{4}$, $H(x) = \frac{x}{2}$, *est quasi-fellérienne, stable et non récurrrente positive.*

Démonstration : L'hypothèse $ii)$ de l'application 7.3 est vérifiée puisque $D_H := \left\{\frac{\pi}{4}\right\}$. Le modèle est donc quasi-fellérien et δ_0-irréductible. On montre alors, de manière analogue au point précédent, qu'il est stable mais pas récurrent positif. □

7.4.7 Sur la nécessité du caractère quasi-fellérien pour obtenir la récurrence positive dans le théorème 7.5

On peut également construire une transition qui n'est pas quasi-fellérienne mais qui vérifie par ailleurs les autres hypothèses du théorème 7.5 sans être récurrente positive.

Soit la transition $P(x, dy)$ sur $[0,1]$ définie par :

$$\begin{cases} P(x, dy) := xU[0,1](dy) + (1-x)\delta_{\frac{x}{2}}(dy), \ \forall x \in]0,1], \\ P(0, dy) := U[0,1](dy), \end{cases}$$

où $U[0,1]$ désigne la loi uniforme sur $[0,1]$ et δ_x désigne la mesure de Dirac au point x,

Pour une fonction f continue bornée de $[0,1]$ dans lui-même, on a :

$$\begin{cases} Pf(x) = x \displaystyle\int_0^1 f(y)\ dy + (1-x) f\left(\frac{x}{2}\right), \ \forall x \in]0,1], \\ Pf(0) = \displaystyle\int_0^1 f(y)\ dy. \end{cases}$$

On remarque que Pf est continue sauf éventuellement en 0, que $\forall x \in [0,1]$, $P(x, \{0\}) = 0$ et que la chaîne est $U[0,1]$-irréductible. De plus, $\forall x \in [0,1]$, les suites de probabilités $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n\in\mathbb{N}}$ sont tendues.

Si l'on calcule les itérées par P de la fonction $f_0 = \mathbf{1}_{\mathbb{Q}\cap]0,1[}(x)$, on a pour tout $k \in \mathbb{N}^*$:

$$\begin{cases} P^k f_0 = \displaystyle\prod_{i=0}^{k-1}\left(1 - \frac{x}{2^i}\right) f_0\left(\frac{x}{2^k}\right), \ \forall x \in]0,1], \\ P^k f_0(0) = 0. \end{cases}$$

On prouve donc, comme pour le point précédent, que la chaîne n'est pas récurrente positive.

Il reste que, pour toute fonction f continue bornée, Pf ne peut être discontinue qu'en 0 et que $P(x, \{0\}) = 0$ pour tout $x \in [0,1]$.

Si ce modèle était quasi-fellérien, le théorème 7.5 s'appliquerait pour assurer la récurrence positive.

Par ailleurs, on se convainc plus directement qu'il ne l'est pas, sur la représentation en Q et H suivante :

$$\forall w \in [0,1], \quad Q(w, dy) := wU[0,1](dy) + (1-w)\delta_{\frac{w}{2}}(dy),$$

et

$$H : [0,1] \to [0,1] \quad \text{définie par} \quad \begin{cases} H(x) := x, \ \forall x \in]0,1], \\ H(0) := 1. \end{cases}$$

En effet, pour $f := Id_{[0,1]}$, $PId_{[0,1]}$ est discontinue en 0, donc $\{0\} \in D_H$. Par ailleurs Q étant fellérienne, on a $Q\left(\lim_{x\to 0} H(x), \{0\}\right) = \delta_0$. Donc $Q\left(\lim_{x\to 0} H(x), D_H\right) = 1 \neq 0$, et les conditions du théorème 3.1 ne sont pas respectées.

De surcroît, comme ce modèle n'admet même pas de loi stationnaire, il illustre également l'optimalité du théorème 3.1.

7.5 Récurrence positive dans le cadre quasi-fellérien Has'minskii

Comme lors du chapitre 6, nous considérons ici que S est en outre localement compact.

7.5.1 Résultat Principal

Théorème 7.6 *On considère un modèle quasi-fellérien de transition $P(x,dy) = Q(H(x),dy)$. On fixe $a \in S$ et on suppose que :*

i) pour toute fonction $f : S \to \mathbb{R}$ borélienne bornée telle que $\lim_{d(x,a)\to+\infty} f(x) = 0$, on a $\lim_{d(x,a)\to+\infty} Pf(x) = 0$,

ii) il existe une fonction de Lyapounov V telle que :

$$\forall x \in S, \quad PV(x) \leq V(x) + \beta \text{ et } \limsup_{d(x,a)\to+\infty} (PV - V)(x) < 0.$$

iii) la chaîne est ψ-irréductible où $\text{supp}\{\psi\}$ est d'intérieur non vide.

Alors le modèle est récurrent positif.

Démonstration : D'après le théorème 6.2, il existe une probabilité invariante ν, nécessairement unique par irréductibilité. Soit une fonction $g \in L^1(\nu)$. On considère alors les ensembles :

$$A_g := \left\{\omega : \ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k)(\omega) \longrightarrow \nu(g)\right\} \text{ et } R_g := \{x \in S / \ \mathbb{P}_x(A_g) = 1\}.$$

La proposition 7.6 entraîne l'existence d'un ouvert tel que $O \cap (S \backslash D_H)$ soit un petit ensemble chargé par ψ.

Il est immédiat que, comme dans le cadre classique (cf proposition 7.5), A_g et R_g vérifient :

$$\forall x \in S, \ \mathbb{P}_x(A_g) \geq \frac{1}{n}\sum_{k=0}^{n-1} P^k(x,O) \times \delta\phi(R_g), \tag{7.11}$$

le problème étant que, cette fois, on ne peut plus affirmer que l'on a :

$$\liminf_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} P^k(x,O) \geq \nu(O) > 0.$$

On reprend alors le noyau de transition $\widehat{P}(x,dy)$, défini sur $\widehat{S}$ par :

$$\begin{cases} \widehat{P}(x,dy) = P(x,dy), \ x \neq \infty \\ \widehat{P}(\infty,dy) = \delta_\infty \end{cases},$$

et les suites de probabilités $(\widehat{\nu}_{n,x})_{n\in\mathbb{N}^*}$ sur $\left(\widehat{S},\mathcal{B}\left(\widehat{S}\right)\right)$ définies par :

$$\forall x \in S, \ \forall n \in \mathbb{N}, \ \widehat{\nu}_{n,x} := \frac{1}{n}\sum_{k=0}^{n-1}\widehat{P}^k(x,dy).$$

On sait que $(\widehat{\nu}_{n,x})_{n\in\mathbb{N}^*}$ est nécessairement tendue car $\widehat{S}$ est compact. Il existe donc au moins une valeur d'adhérence $\widehat{\nu}_x$ (éventuellement dépendante de x).

Soit $\left(\widehat{\nu}_{\phi(n),x}\right)_{n\in\mathbb{N}^*}$ une suite extraite de $(\widehat{\nu}_{n,x})_{n\in\mathbb{N}^*}$ convergeant étroitement vers $\widehat{\nu}_x$. On a alors pour tout ouvert O de S :

$$\liminf_{n\to+\infty}\widehat{\nu}_{\phi(n),x}\left(O\right) \geq \widehat{\nu}_x\left(O\right).$$

Lors de la démonstration du théorème 6.2, on montre (inégalité (6.4)), que :

$$\forall x \in S, \ \widehat{\nu}_x(S) \geq \frac{\beta}{\alpha+\beta},$$

et que la trace de $\widehat{\nu}_x$ sur S est une mesure stationnaire finie.

La chaîne étant irréductible, l'unicité de la loi stationnaire ν entraîne que :

$$\widehat{\nu}_x = \widehat{\nu}_x(S)\ \nu \geq \frac{\beta}{\alpha+\beta}\ \nu.$$

Par suite :

$$\liminf_{n\to+\infty}\widehat{P}^k(x,O) = \liminf_{n\to+\infty}P^k(x,O) \geq \widehat{\nu}_x(O) \geq \frac{\beta}{\alpha+\beta}\times\nu(O) > 0,$$

ce qui, combinée avec (7.11) conduit à :

$$\forall x \in S, \ \mathbb{P}_x(A_g) \geq \frac{\beta}{\alpha+\beta}\times\nu(O)\times\delta\phi(R_g) > 0.$$

On peut alors reproduire à l'identique la fin de la démonstration de la proposition 7.5. □

7.5.2 Application aux modèles $ARCH(1)$

De même que lors de la section précédente, on peut donner l'application et l'exemple suivant :

Application 7.6 *On considère un modèle $ARCH(1)$ où F et G sont d'ensemble de discontinuité de mesure de Lebesgue nulle et F et G^* sont bornées sur les compacts. On suppose que :*

i) G *est borélienne bornée sur* $\mathbb{R}^d$ *et* $\lim_{|x|\to+\infty}|F(x)| = +\infty$,

ii) $\forall x \in \mathbb{R}^d, \quad |F(x)| + \|G(x)\|\ \mathbb{E}|\varepsilon_1| \leq |x| + \beta$,

iii) $\limsup_{|x|\to+\infty}\left(|F(x)| - |x| + \|G(x)\|\ \mathbb{E}|\varepsilon_1|\right) < 0$,

iv) $\mu \sim \lambda_d$.

Alors le modèle est récurrent positif.

Exemple 7.3 *On considère le modèle $NLR(1)$ à valeurs dans $\mathbb{R}$ où :*

i) F est définie par :

$$\begin{cases} F(x) := x - 3, \ x > 2, \\ F(x) := f(x), \ -1 \leq x \leq 1, \\ F(x) := x + 3, \ x < -2, \end{cases},$$

où f est une fonction Riemann-intégrable à valeurs dans $[-1, 1]$

ii) la loi μ des ε_t est portée par l'intervalle $[-2, 2]$ sur lequel elle a une densité strictement positive par rapport à la mesure de Lebesgue.

Alors le modèle est récurrent positif.

Démonstration : Le critère de Has'minskii est bien vérifié puisque $\mathbb{E}|\varepsilon_1| < 2$. □

De même que lors de la section précédente, on pourrait trouver un modèle quasi-fellérien récurrent positif sans être quasi-fortement fellérien et qui vérifie le critère de Has'minskii.

7.6 Récurrence positive dans le cadre quasi-fortement fellérien standard

7.6.1 Résultat principal

Ici, la récurrence positive s'obtient sous des hypothèses plus faibles que dans le cadre quasi-fellérien. En effet, la condition de ψ-irréductibilité avec supp$\{\psi\}$ d'intérieur non vide n'est plus nécessaire. En fait, la ψ-irréductibilité suffit et même l'indécomposabilité. Cependant, même si cette notion semble être un affaiblissement de l'irréductibilité, en montrant la récurrence positive, on montre a posteriori l'irréductibilité (voir la proposition 7.4).

Théorème 7.7 *On considère un modèle markovien homogène quasi-fortement fellérien de transition $P(x, dy) = Q(H(x), dy)$ sur S. On suppose que :*

i) $\forall x \in S$, $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n\in\mathbb{N}}$ est tendue,

ii) la transition $P(x, dy)$ est indécomposable.

Alors, le modèle est récurrent positif.

Démonstration : On est évidemment sous les conditions d'application du théorème 3.1 donc il existe une loi stationnaire ν, unique car $P(x, dy)$ est indécomposable (théorème 7.3), et vérifiant $\nu(D_H) = 0$.

En outre, $\forall x \in S$, $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy)\right)_{n\in\mathbb{N}}$ est tendue, donc :

$$\forall x \in S, \ \frac{1}{n}\sum_{k=0}^{n-1} P^k(x, dy) \Rightarrow \nu. \tag{7.12}$$

Soit g une fonction borélienne bornée, Pg est ν-*p.s.* continue. Appliquée à Pg, la relation (7.12) entraîne :

$$\forall x \in S, \ \frac{1}{n}\sum_{k=0}^{n-1} P^k(Pg)(x) \longrightarrow \nu(Pg) = \nu(g). \tag{7.13}$$

Par suite :

$$\forall x \in S,\ \frac{1}{n}\sum_{k=0}^{n-1} P^k g(x) = \frac{1}{n}\sum_{k=1}^{n} P^k g(x) + O\left(\frac{1}{n}\right) \longrightarrow \nu(g), \tag{7.14}$$

on a donc convergence de $\left(\frac{1}{n}\sum_{k=0}^{n-1} P^k(x,dy)\right)_{n\in\mathbb{N}}$ vers ν le long de toutes les fonctions g boréliennes bornées.

Soit g borélienne bornée et $A_g := \left\{\omega :\ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k)(\omega) \longrightarrow \nu(g)\right\}$. $\mathbb{P}_\nu(A_g) = 1$ car la chaîne est $\mathbb{P}_\nu$-ergodique. L'invariance de A_g par l'opérateur de décalage entraîne :

$$\forall x \in S,\ \forall k > 0,\ \mathbb{P}_x(A_g) = \int P^k(x,dy)\mathbb{P}_y(A_g).$$

En remplaçant $y \mapsto \mathbb{P}_y(A_g)$ dans (7.14) on trouve :

$$\mathbb{P}_x(A_g) = \frac{1}{n}\sum_{k=0}^{n-1} P^k\left(\mathbb{P}_.(A_g)\right)(x) \longrightarrow \nu\left(\mathbb{P}_.(A_g)\right) = \mathbb{P}_\nu(A_g) = 1.$$

Cela prouve que pour tout $x \in S$, $\mathbb{P}_x(A_g) = 1$ et donc la récurrence positive du modèle. □

Evidemment les modèles $ARCH(1)$ de transition indécomposable et dont le bruit a une densité absolument continue par rapport à la mesure de Lebesgue rentrent dans le cadre du théorème 7.7. Cependant, comme nous l'indiquions en tête de cette section, ces modèles sont, a posteriori, ψ-irréductibles où ψ est absolument continue par rapport à la mesure de Lebesgue, donc de support d'intérieur non vide. Ces modèles peuvent donc être traités par le théorème 7.5, le seul problème étant de prouver leur irréductibilité a priori.

7.6.2 Exemple d'un modèle quasi-fortement fellérien de loi stationnaire de support d'intérieur vide

Pour illustrer l'amélioration apportée par rapport au cadre quasi-fellérien, nous allons donc donner un exemple de chaîne quasi-fortement fellérienne récurrente positive de mesure maximale d'irréductibilité (donc de loi stationnaire) d'intérieur vide.

Application 7.7 *Soit la transition sur* $[0,1]$ *définie par :*

$$P(x,dy) := \left(1 - \mu\left([0,H(x)]\right)\right)\delta_0(dy) + \mathbf{1}_{[0,H(x)]}(y)\mu(dy),$$

où μ *est une loi diffuse quelconque sur* $[0,1]$*. On suppose en outre :*

i) l'ensemble D_H *des points de discontinuité de* H *vérifie* $0 \notin D_H$ *et* $\mu(D_H) = 0$*,*

ii) $\forall x \in [0,1], \quad H(x) \neq 1$.

Alors, le modèle est récurrent positif.

Démonstration : Pour $f : [0,1] \to \mathbb{R}$ borélienne bornée, on a :

$$Pf(x) := \left(1 - \mu\left([0,H(x)]\right)\right)f(0) + \int_0^{H(x)} f(y)\mu(dy).$$

μ étant diffuse, Pf est donc continue en tout point où H est continue. L'hypothèse $i)$ permet alors d'affirmer que le modèle est quasi-fortement fellérien. Par ailleurs, l'hypothèse $ii)$ prouve qu'il est δ_0-irréductible. Il est donc récurrent positif.

Enfin, dans le cas où $H(0) := 0$, δ_0 est l'unique loi stationnaire (de support d'intérieur vide dans $[0,1]$). □

Remarque : Si $H(0) \neq 0$, la loi stationnaire ν a une partie diffuse. Le modèle étant alors ν-irréductible (puisque récurrent positif comme nous venons de le voir), on est sous les conditions d'application du théorème 7.5. Cependant, la loi ν dépendant très fortement de la forme de H, il est difficile de la déterminer a priori. Il est donc également difficile de prouver que l'on est bien sous les conditions d'application du théorème 7.5. Le cadre quasi-fortement fellérien est donc une manière de simplifier l'approche d'un modèle.

7.7 Récurrence positive dans le cadre quasi-fortement fellérien Has'minskii

On a toujours besoin de supposer que S est polonais et localement compact.

Théorème 7.8 *On considère un modèle quasi-fortement fellérien. On fixe $a \in S$ et on suppose que :*

i) pour toute fonction $f : S \to \mathbb{R}$ borélienne bornée telle que $\lim_{d(x,a)\to+\infty} f(x) = 0$, on a $\lim_{d(x,a)\to+\infty} Pf(x) = 0$,

ii) il existe une fonction de Lyapounov V telle que :

$$\forall x \in S, \quad PV(x) \le V(x) + \beta \text{ et } \limsup_{d(x,a)\to+\infty} (PV - V)(x) < 0.$$

iii) la transition est indécomposable.

Alors, le modèle est récurrent positif.

Démonstration : On sait qu'il existe une unique loi stationnaire ν. Pour g borélienne bornée, on définit toujours les ensembles A_g et R_g par :

$$A_g := \left\{\omega :\ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k)(\omega) \longrightarrow \nu(g)\right\}, \quad R_g := \{y \in S/\ \mathbb{P}_y(A_g) = 1\}.$$

$\mathbb{P}_\nu(A_g) = 1$ et $\nu(R_g) = 1$ car la chaîne est $\mathbb{P}_\nu$-ergodique.

Soit $x \in S$, nous voulons toujours montrer que $\mathbb{P}_x(A_g) = 1$.

Nous savons (démonstration du théorème 7.5) qu'il suffit en fait de montrer que la fonction $x \mapsto \mathbb{P}_x(A_g)$ de S dans $\mathbb{R}$ est minorée par une constante strictement positive.

On reconsidère le noyau de transition $\widehat{P}(x,dy)$, et la suite de probabilités $(\widehat{\nu}_{n,x})_{n\in\mathbb{N}^*}$ sur $\left(\widehat{S}, \mathcal{B}\left(\widehat{S}\right)\right)$ définis dans la démonstration du théorème 7.6.

$(\widehat{\nu}_{n,x})_{n\in\mathbb{N}^*}$ admet au moins une valeur d'adhérence étroite $\widehat{\nu}_x$ sur $\widehat{S}$.

Considérons $\left(\widehat{\nu}_{\phi(n),x}\right)_{n\in\mathbb{N}^*}$ une suite extraite de $(\widehat{\nu}_{n,x})_{n\in\mathbb{N}^*}$ qui converge étroitement vers $\widehat{\nu}_x$ et K un compact. L'ensemble $R_g \cap K$ étant bornée, $\widehat{P}\mathbf{1}_{R_g\cap K}(.)$ est une fonction $\widehat{\nu}_x$-*p.s.* continue sur $\widehat{S}$ (voir la démonstration du théorème 6.2). On peut donc écrire :

$$\lim_{n\to+\infty} \widehat{\nu}_{\phi(n),x}\left(\widehat{P}\mathbf{1}_{R_g\cap K}\right) = \widehat{\nu}_x\left(\widehat{P}\mathbf{1}_{R_g\cap K}\right) = \widehat{\nu}_x(R_g \cap K). \tag{7.15}$$

On a vu lors de la démonstration du théorème 7.6 :

$$\exists\alpha > 0,\ \widehat{\nu}_x(R_g \cap K) = \nu(R_g \cap K) \times \widehat{\nu}_x(S) \geq \nu(R_g \cap K) \times \frac{\beta}{\alpha+\beta}. \tag{7.16}$$

De (7.15) et (7.16) on tire :

$$\lim_{n\to+\infty} \widehat{\nu}_{\phi(n),x}\left(\widehat{P}\mathbf{1}_{R_g\cap K}\right) \geq \nu(R_g \cap K) \times \frac{\beta}{\alpha+\beta}. \tag{7.17}$$

Mais $\nu(R_g) = 1$ car le modèle est ergodique sous $\mathbb{P}_\nu$. Faisant tendre K vers S, on déduit de (7.17) que :

$$\limsup_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \widehat{P}^k(x, R_g) \geq \frac{\beta}{\alpha+\beta},$$

et par suite (comme R_g est dans S) :

$$\limsup_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} P^k(x, R_g) \geq \frac{\beta}{\alpha+\beta}. \tag{7.18}$$

D'autre part :

$$\mathbb{P}_x(A_g) = \frac{1}{n}\sum_{k=0}^{n-1}\int P^k(x,dy)\mathbb{P}_y(A_g) \geq \frac{1}{n}\sum_{k=0}^{n-1} P^k(x, R_g),$$

d'où l'on tire avec (7.18) et comme x est quelconque :

$$\forall x \in S,\ \mathbb{P}_x(A_g) \geq \frac{\beta}{\alpha+\beta} > 0. \tag{7.19}$$

La fin de la démonstration de la proposition 7.5 permet ici encore de conclure. □

7.8 Récurrence positive des modèles p-dépendants.

Nous donnons juste ici une illustration pour les modèles $NARCH(p)$, tous les résultats précédents étant évidemment transposables pour les modèles p-dépendants.

Application 7.8 *Soit le modèle $NARCH(p)$ à valeurs dans $\mathbb{R}^d$ défini par :*

$$X_{t+1} = f(A_1X_t + \cdots + A_pX_{t-p+1}) + g(B_1X_t + \cdots + B_pX_{t-p+1})\varepsilon_{t+1},$$

où les $(A_i)_{1\leq i\leq p}$ et les $(B_i)_{1\leq i\leq p}$ sont des matrices de $\mathcal{M}_d(\mathbb{R})$ avec A_1 et B_1 inversibles. On suppose en outre que :

i) ε_t a une loi équivalente à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$,
ii) $\lambda_d(D_{f,g}) = 0$,
iii) f, g et g^* sont bornées sur les compacts,
iv) le critère de Pakes ou le critère de Has'minskii est vérifié.

Alors le modèle surdimensionné est récurrent positif.

Démonstration : Sous ces hypothèses, on sait (application 5.2) que la chaîne admet au moins une probabilité invariante. On sait aussi que pour tout x dans $\mathbb{R}^d$, $P^p(x,dy)$ est équivalente à la mesure de Lebesgue sur $\mathbb{R}^d$, où $P(x,dy)$ est la transition du modèle surdimensionné. On est alors sous les hypothèses de l'application 7.6.

Le modèle surdimensionné $Y_t = (X_t, \cdots, X_{t-p+1})$ est donc récurrent positif, i.e. :

$\forall g$ borélienne bornée de $\left(\mathbb{R}^d\right)^p$ dans $\mathbb{R}$, $\forall x := (x^1, \cdots, x^p) \in \left(\mathbb{R}^d\right)^p$,

$$\frac{1}{n}\sum_{k=0}^{n-1} g(Y_k) \overset{\mathbb{P}_x\text{-p.s.}}{\longrightarrow} \nu(g). \quad \square$$

7.9 Résultats antérieurs

On peut énoncer le résultat connu démontré par exemple dans (Duflo, 1997) ou (Meyn et Tweedie, 1993) concernant la récurrence des chaînes à valeurs dans un espace E, non forcément topologique.

Théorème 7.9 *Soit un chaîne de Markov $(X_t)_{t\in\mathbb{N}}$ de transition $P(x,dy)$ sur un espace E. On suppose :*

i) il existe un petit ensemble récurrent C vers lequel tout point $x \in E$ conduit presque sûrement,

ii) $\sup_{x\in C} \mathbb{E}_x(T_C) < +\infty$, où $T_C := \inf\{t/\ t > 0,\ X_t \in C\}$ est le temps de retour dans C.

Alors la chaîne est récurrente positive.

Si l'on cherche à appliquer ce résultat à des chaînes à valeurs dans $\mathbb{R}^d$ possédant une fonction de Lyapounov, comme les modèles NLR et $ARCH$ étudiés précédemment, une formulation naturelle est donnée par l'énoncé ci-après :

Théorème 7.10 *Soit une chaîne de Markov homogène $(X_t)_{t\in\mathbb{N}}$ de transition $P(x,dy)$ sur S polonais. On suppose que :*

i) il existe une fonction de Lyapounov V et deux réels $a,\ M > 0$ tels que :

$$\left\{\begin{array}{l} PV(x) \leq V(x) - a,\ si\ V(x) > d, \\ PV(x) \leq M,\ si\ V(x) \leq d. \end{array}\right. ,$$

ii) il existe un réel δ strictement positif et une probabilité ϕ tels que :

$$\forall x \in \{V \leq d\},\ P(x,dy) \geq \delta\phi(dy).$$

Alors le modèle est récurrent.

Ce résultat découle du théorème 7.9 en prenant comme petit ensemble $C := \{V \leq d\}$.

L'hypothèse $i)$ ressemble beaucoup à ce que nous avons appelé le critère de Has'minskii sauf qu'aucune hypothèse sur le comportement à l'infini de Pf n'est demandé si f est une fonction de limite nulle à l'infini. De plus, aucune hypothèse de type fellérien n'est demandée.

Cependant, l'hypothèse $ii)$ induit de fortes contraintes sur la chaîne puisque l'on impose que l'ensemble $\{V \leq d\}$ soit un petit ensemble. Or, si nous faisons appel à la notion de petit ensemble (via l'irréductibilité) dans le cadre quasi-fellérien, nous n'avons jamais besoin de le spécifier, et moins encore de chercher à borner le temps de retour de la chaîne dans celui-ci.

De plus, dans le cadre quasi-fortement fellérien, la notion de petit ensemble et même d'irréductibilité disparaît pour faire place à l'indécomposabilité, plus facile à manier en pratique.

En fait, nous allons illustrer, au travers des exemples, que ces deux types de résultats ne se recouvrent pas.

Ainsi, l'exemple 7.4 ci-dessous est un corollaire du théorème 7.10 et ne peut pas être traité dans un cadre quasi-fellérien puisqu'aucune hypothèse de continuité sur F n'est requise.

Exemple 7.4 *Soit un modèle $NLR(1)$ sur $\mathbb{R}^d$ où ε_1 a un moment d'ordre 1. On suppose que :*

i) il existe un réel $a > 0$ tel que :

$$\forall x \in \mathbb{R}^d,\ |F(x)| \leq |x| - a - \mathbb{E}|\varepsilon_1|,$$

ii) la loi de ε_1 a une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ minorée sur tout compact K de $\mathbb{R}^d$ par une constante $\delta_K > 0$.

Alors le modèle est récurrent positif.

En revanche, comme corollaire de l'application 7.6 du paragraphe 7.6, on peut donner le résultat suivant, ne rentrant pas dans le cadre du théorème 7.10 :

Exemple 7.5 *Soit un modèle $NLR(1)$ sur $\mathbb{R}^d$ où ε_1 a un moment d'ordre 1. On suppose que :*

i) $\lambda_d(D_F) = 0$ et $\lim\limits_{x\to+\infty} |F(x)| = +\infty$,

ii) il existe un réel $a > 0$ tel que :

$$\forall x \in \mathbb{R}^d,\ |F(x)| \leq |x| - a - \mathbb{E}|\varepsilon_1|,$$

iii) la loi de ε_1 a une densité par rapport à la mesure de Lebesgue λ_d de $\mathbb{R}^d$ et la transition est indécomposable (e.g. le bruit a une densité strictement positive par rapport à le mesure de Lebesgue voir aussi l'exemple 7.3),

Alors le modèle est récurrent positif.

De même, l'application 7.3 est très difficilement abordable de manière générale par les méthodes de petits ensembles récurrents. En effet, la donnée de H est nécessaire pour identifier un éventuel petit ensemble récurrent pour lequel on peut borner l'espérance du temps de retour, d'autant plus que H peut se rapprocher une infinité de fois non dénombrable de 1.

Notre approche des modèles quasi-fellériens s'avère donc performante dès que le bruit est relativement pauvre ou que la forme de la transition n'est pas assez explicite. En efffet, dans ces cas-là, il devient très difficile d'identifier les petits ensembles récurrents, ce qui est nécessaire si l'on cherche à étudier le temps de retour de la chaîne dans l'un d'entre eux.

En outre, si l'on s'intéresse simplement à l'existence d'une probabilité invariante, toutes les hypothèses du type irréductibilité ou indécomposabilité sont inutiles.

Cependant, comme nous le précisions précédemment, l'approche quasi-fellérienne ne permet pas d'avoir de vitesse de convergence vers la loi stationnaire. Il faut alors revenir vers les méthodes de petits ensembles développées dans cette optique.

Chapitre 8

Stabilité étendue. Application au mélange lipschitzien

Lors du chapitre 7, nous avions vu que la propriété de loi des grands nombres pour une chaîne ν-récurrente positive était valable pour un certain nombre de fonctions non bornées. Plus précisément :

Proposition 8.1 *Si une chaîne de Markov $(X_t)_{t\in\mathbb{N}}$ est récurrente positive, alors :*

$$\forall g \ \in L^1(\nu),\ \forall x \in S,\quad \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Evidemment, il est souvent difficile d'identifier précisément $L^1(\nu)$, mais on peut en exhiber des sous-ensembles.

Ainsi, nous avions vu en section 3.2 (critère de Hajek) que s'il existe une fonction de Lyapounov V telle que :

$$\forall x \in S, \quad PV(x) \leq \alpha V(x) + \beta \text{ avec } \alpha < 1,$$

alors pour pour toute loi stationnaire ν, on a $\nu(V) \leq \dfrac{\beta}{1-\alpha} < +\infty$.

Donc, si une chaîne ν-récurrente positive vérifie le critère de Hajek pour une fonction V, l'ensemble des fonctions g boréliennes qui vérifient $|g(x)| \leq cte(V(x)+1)$ est un sous-ensemble de $L^1(\nu)$.

La propriété de loi des grands nombres est alors vraie pour de telles fonctions, i.e. :

Proposition 8.2 *Soit une chaîne de Markov $(X_t)_{t\in\mathbb{N}}$. On suppose que :*

i) la chaîne est ν-récurrente positive,
ii) il existe une fonction de Lyapounov V telle que :

$$\forall x \in S, \quad PV(x) \leq \alpha V(x) + \beta \textit{ avec } \alpha < 1.$$

Alors, pour toute fonction borélienne g telle que $|g(x)| \leq cte(V(x)+1)$:

$$\forall x \in S, \quad \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Cette propriété de loi des grands nombres pour des fonctions non bornées est très importante dès lors que l'on veut faire des statistiques sur les modèles, par exemple, si l'on s'intéresse à l'estimation des moments.

A cette fin, il est donc intéressant, pour une chaîne de Markov ν-stable, où ν est une loi stationnaire, puisque nous avons vu qu'il est possible qu'une chaîne soit stable sans posséder de loi stationnaire, d'étendre les conclusions du théorème 7.1 (Duflo, 1990).

Ce résultat a déjà été utilisé dans (Mangeas et Yao, 1996) et (Attali et Yao, 1998).

8.1 Extension de la propriété de stabilité

Théorème 8.1 *Soit un modèle markovien homogène $(X_t)_{t\in\mathbb{N}}$ de transition quasi-fellérienne $P(x,dy) = Q(H(x),dy)$ sur S. On suppose que :*

i) il existe une fonction de Lyapounov V continue telle que :

$$\forall x \in S,\ PV(x) \le \alpha V(x) + \beta,\ \textit{avec } \alpha \in]0,1[,$$

ii) il existe au plus une loi stationnaire.

Alors,

(a) le modèle est stable de loi stationnaire ν.
(b) En outre, pour toute fonction g ν-p.s. continue telle que $|g(x)| \le cte(V^b(x)+1)$ avec $b < 1$:

$$\forall x \in S,\quad \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Démonstration : Le point (a) a déjà été montré au chapitre 7. Démontrons maintenant le point (b).

Notre démarche repose sur la démonstration de (Duflo, 1996, page 15) (Étape 1) ainsi que sur une extension de la propriété de quasi-fellériennité du modèle aux fonctions continues non bornées de la forme V^b avec $b < 1$ (Étape 2).

Étape 1 : Soit $x \in S$ et $b < 1$. On va tout d'abord montrer que :

$$\forall x \in S,\ \frac{1}{n}\sum_{k=0}^{n-1}\Big(V^b(X_k) - PV^b(X_k)\Big) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} 0.$$

Il est clair que : $\displaystyle\frac{1}{n}\sum_{k=0}^{n-1}\Big(V^b(X_k) - PV^b(X_k)\Big) =$

$$\underbrace{\frac{1}{n}\sum_{k=1}^{n-1}\Big(V^b(X_k) - PV^b(X_{k-1})\Big)}_{A_n} + \underbrace{\frac{V^b(X_0)}{n} - \frac{PV^b(X_{n-1})}{n}}_{B_n}.$$

Convergence de A_n : Soit M_n la $\mathcal{F}_n$-martingale définie par :

$$M_n := \sum_{k=1}^{n}\frac{1}{k}\Big(V^b(X_k) - PV^b(X_{k-1})\Big)$$

On va montrer, en utilisant le théorème de Chow, que M_n converge $\mathbb{P}_x$-*p.s.*.

$$\text{On calcule } \mathbb{E}_x \left|M_n - M_{n-1}\right|^{\frac{1}{b}} = \mathbb{E}_x \left|\frac{V^b(X_n) - PV^b(X_{n-1})}{n}\right|^{\frac{1}{b}}. \tag{8.1}$$

L'inégalité triangulaire pour la norme $\|\,.\|_{L^b}$ $\left(\frac{1}{b} > 1\right)$ donne :

$$\left(\mathbb{E}_x \left|V^b(X_n) - PV^b(X_{n-1})\right|^{\frac{1}{b}}\right)^b \leq \left(\mathbb{E}_x \left(V\left(X_n\right)\right)\right)^b + \left(\mathbb{E}_x \left(PV^b(X_{n-1})\right)^{\frac{1}{b}}\right)^b.$$

En appliquant l'inégalité de Jensen à PV^b, on trouve :

$$\left(PV^b(X_{n-1})\right)^{\frac{1}{b}} \leq PV(X_{n-1}), \tag{8.2}$$

ce qui, remplacé dans l'inégalité ci-dessus, donne :

$$\left(\mathbb{E}_x \left|V^b(X_n) - PV^b(X_{n-1})\right|^{\frac{1}{b}}\right)^b \leq 2\left(P^n V(x)\right)^b. \tag{8.3}$$

Comme $PV(x) \leq \alpha V(x) + \beta$ avec $\alpha < 1$, on peut écrire, pour tout $n \in \mathbb{N}$:

$$P^n V(x) \leq \alpha^n V(x) + \frac{\beta}{1-\alpha} \leq V(x) + \frac{\beta}{1-\alpha} := C(x). \tag{8.4}$$

On obtient alors avec (8.3) et (8.1) :

$$\mathbb{E}_x \left|M_n - M_{n-1}\right|^{\frac{1}{b}} \leq \frac{2^{\frac{1}{b}} C(x)}{n^{\frac{1}{b}}},$$

et partant,

$$\sum_{k=1}^{+\infty} \mathbb{E}_x \left|M_{n+1} - M_n\right|^{\frac{1}{b}} < +\infty.$$

Par suite :

$$\sum_{k=1}^{+\infty} \mathbb{E}_x \left(\left|M_{n+1} - M_n\right|^{\frac{1}{b}} / \mathcal{F}_{n-1}\right) < +\infty,\ \mathbb{P}_x\text{-}p.s.$$

A ce stade, on applique le théorème de Chow (Hall et Heyde, 1980) pour obtenir la convergence $\mathbb{P}_x$-*p.s.* de M_n. Le lemme de Kronecker permet alors de conclure que :

$$\frac{1}{n} \sum_{k=1}^{n-1} \left(V^b(X_k) - PV^b(X_{k-1})\right) \xrightarrow{\mathbb{P}_x\text{-}p.s.} 0.$$

Convergence de B_n : On a de manière évidente $\frac{V^b(X_0)}{n} \xrightarrow{\mathbb{P}_x\text{-}p.s.} 0$.
Par ailleurs, on a montré précédemment, que pour tout $n \in \mathbb{N}^*$:

$$\mathbb{E}_x \left(PV^b(X_{n-1})\right)^{\frac{1}{b}} \leq C(x).$$

On voit alors que $\sum_{k=1}^{n-1}\left(\frac{PV^b(X_k)}{k}\right)^{\frac{1}{b}}$ converge en norme $L^1(\mathbb{P}_x)$ donc converge $\mathbb{P}_x$-*p.s.* On en déduit que $\frac{PV^b(X_{n-1})}{n} \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} 0$, et finalement :

$$B_n = \frac{V^b(X_0)}{n} - \frac{PV^b(X_{n-1})}{n} \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} 0.$$

Étape 2 : On sait déjà que la chaîne est stable de loi stationnaire ν. Par définition de nos modèles, pour toute fonction f continue bornée de S dans $\mathbb{R}$, Pf est continue sauf éventuellement sur D_H. On sait également que $\nu(D_H) = 0$. On veut maintenant montrer que PV^b est continue sauf sur D_H ce qui prouvera que PV^b est ν-*p.s.* continue.

Soit alors $x \in S\backslash D_H$ et $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de S telle que $\lim_{n\to+\infty} x_n = x$. Il est clair que $P(x_n, dy) \Rightarrow P(x, dy)$.

D'autre part, le critère de Hajek et la continuité de V entraînent :

$$\sup_{n\in\mathbb{N}} P\left((V^b)^{\frac{1}{b}}(x_n)\right) = \sup_{n\in\mathbb{N}} PV(x_n) \leq \sup_{n\in\mathbb{N}} (\alpha V(x_n) + \beta) < +\infty,$$

ce qui nous assure l'equiintégrabilité de V^b par rapport aux probabilités $P(x_n, dy)$.

On en déduit alors que $\lim_{n\to+\infty} PV^b(x_n) = PV^b(x)$, ce qui prouve que PV^b est continue sauf éventuellement sur D_H.

On a $PV \leq \alpha V + \beta$. En appliquant l'inégalité de Jensen à PV^b on trouve :

$$PV^b \leq (PV)^b \leq (\alpha V + \beta)^b \leq \alpha^b V^b + \beta^b.$$

Par suite, si l'on pose : $h := \frac{V^b - PV^b + \beta^b}{1-\alpha^b}$, il vient

$$V^b \leq h \quad \text{et} \quad \forall x \in S,\ \frac{1}{n}\sum_{k=0}^{n-1} h(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \frac{\beta^b}{1-\alpha^b}, \tag{8.5}$$

d'après l'étape 1. Or la chaîne est stable de loi invariante ν. De plus, d'après le critère de Hajek $\nu(V) < +\infty$, d'où $\nu(V^b) < +\infty$. h est donc intégrable sous ν et on a :

$$\nu(h) = \nu\left(\frac{V^b - PV^b + \beta^b}{1-\alpha^b}\right) = \frac{\beta^b}{1-\alpha^b} \text{ car } \nu(PV^b) = \nu(V^b),$$

ce qui donne, reporté dans l'équation (8.5) :

$$\forall x \in S,\ \frac{1}{n}\sum_{k=0}^{n-1} h(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(h).$$

Comme h est ν-*p.s.* continue et $h \geq V^b \geq 0$, que par ailleurs la chaîne est stable, on a classiquement par equiintégrabilité :

$$\forall x \in S,\ \frac{1}{n}\sum_{k=0}^{n-1} V^b(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(V^b),$$

et, plus généralement, pour toute fonction ν-*p.s.* continue g telle que $|g(x)| \leq cte(V^b(x)+1)$:

$$\forall x \in S,\ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g). \quad \square$$

Remarque : La chaîne étant stable de loi stationnaire ν, elle est ergodique sous $\mathbb{P}_\nu$. En particulier, on a donc :

$$\forall g \in L^1(\nu),\ \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_\nu\text{-}p.s.}{\longrightarrow} \nu(g).$$

La loi des grands nombres reste donc vraie pour ν-presque tout x pour toute fonction g intégrable sous ν, donc pour au moins un x_0 puisque ν n'est pas nulle. Cette remarque nous sera utile pour traiter le mélange lipschitzien.

8.2 Le mélange lipschitzien

8.2.1 Résultat général

La propriété de loi des grands nombres pour les modèles de mélange lipschitzien citée dans (Duflo, 1990) se voit alors comme un simple corollaire du théorème 8.1 du paragraphe 8.1.

Nous évitons l'introduction des fonctions $Li(r,s)$ dont le maniement est très technique, la contre-partie de cette méthode étant que l'on ne récupère pas la vitesse de convergence pour les fonctions $Li(r,s)$, propriété qui s'obtient en utilisant l'inégalité de Hölder (Duflo, 1997, page 201).

Cependant, notre objectif n'était pas de s'intéresser aux vitesses de convergence mais plutôt d'étendre la classe de fonctions vérifiant la loi des grands nombres en utilisant une approche basée sur l'utilisation du théorème ergodique, métode employée notamment par L. Elie dans son cours de DEA. Cette amélioration, qui est l'objet du théorème 8.2 ci-dessous, est citée dans (Duflo, 1996).

Théorème 8.2 *On se place sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel on peut construire pour tout x de $\mathbb{R}^d$ une version $(X_t^x)_{t\in\mathbb{N}}$ de la chaîne de Markov de transition $P(x,dy)$ issue de x. Soit $a \geq 1$, on suppose :*

i) $\sup_{n\in\mathbb{N}} \mathbb{E}\left|X_n^0\right|^a < +\infty$,
ii) $\exists\ \alpha \in]0,1[$ *tel que* $\forall x,y \in \mathbb{R}^d, \quad \mathbb{E}\left|X_n^x - X_n^y\right|^a \leq \alpha^n \left|x-y\right|^a.$

Alors,

(a) la chaîne est stable de loi invariante ν.
(b) De plus, $\forall g \in C(\mathbb{R}^d, \mathbb{R})$ telle que $|g(x)| \leq cte\,(|x|^a+1)$ on a :

$$\forall x \in \mathbb{R}^d,\quad \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g).$$

Démonstration : Vérifions tout d'abord les hypothèses du théorème 8.1.

L'hypothèse *ii)* appliquée entre 0 et x donne :

$$\mathbb{E}\left|X_n^x - X_n^0\right|^a \leq \alpha^n \left|x\right|^a.$$

Par ailleurs, l'inégalité triangulaire pour la norme L^a permet d'écrire :

$$(\mathbb{E}|X_n^x|^a)^{\frac{1}{a}} - \left(\mathbb{E}\left|X_n^0\right|^a\right)^{\frac{1}{a}} \leq \left(\mathbb{E}\left|X_n^x - X_n^0\right|^a\right)^{\frac{1}{a}}.$$

En posant $c := \sup_{n\in\mathbb{N}} \mathbb{E}\left|X_n^0\right|^a$, on tire de ces deux dernières inégalités :

$$P|\,.\,|^a(x) = \mathbb{E}|X_1^x|^a \leq \left(\alpha^{\frac{1}{a}}|x|^{\frac{1}{a}} + c^{\frac{1}{a}}\right)^a,$$

ce qui prouve qu'il existe $\beta \in \mathbb{R}$ tel que :

$$P|\,.\,|^a(x) \leq \alpha|x|^a + \beta,$$

le critère de Hajek est donc vérifié avec $V(x) := |x|^a$.

D'autre part, il est classique que si f est une fonction lipschitzienne bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, Pf est également lipschitzienne, le modèle est alors fellérien. Il existe donc au moins une loi stationnaire ν.

Enfin, celle-ci est forcément unique. On peut le vérifier par exemple par une méthode de point fixe (Sunyach, 1975), ou en remarquant que l'hypothèse de mélange lipschitzien implique :

$$\forall x, y \in \mathbb{R}^d,\ X_n^x - X_n^y \xrightarrow{\mathcal{L}} 0,$$

qui est un critère d'unicité de la loi stationnaire lorsque l'existence est déjà assurée (Duflo, 1990).

On est alors sous les conditions d'application du théorème 8.1, i.e. pour toute fonction g continue telle que $|g(x)| \leq cte(|x|^b + 1)$ avec $b < a$:

$$\forall x \in \mathbb{R}^d, \quad \frac{1}{n}\sum_{k=0}^{n-1} g(X_k) \overset{\mathbb{P}_x\text{-}p.s.}{\longrightarrow} \nu(g). \tag{8.6}$$

Reste à démontrer la propriété de loi des grands nombres pour $|\,.\,|^a$. De l'inégalité :

$$\forall x, y \in \mathbb{R}^d,\ \big|\,|x|^a - |y|^a\big| \leq a|x-y|\left(|x|^{a-1} + |y|^{a-1}\right),$$

on déduit, pour tout k, $0 \leq k \leq n-1$:

$$\forall x, y \in \mathbb{R}^d,\ \big|\,|X_k^x|^a - |X_k^y|^a\big| \leq a\,|X_k^x - X_k^y|\left(|X_k^x|^{a-1} + |X_k^y|^{a-1}\right). \tag{8.7}$$

D'une part la série de terme général $\mathbb{E}|X_n^x - X_n^y|$ converge ce qui implique $|X_n^x - X_n^y| \xrightarrow{p.s.} 0$, d'autre part, comme $a-1 < a$, l'équation (8.6) permet d'écrire :

$$\sum_{k=0}^{n-1} |X_k^x|^{a-1} \sim n\nu(|\,.\,|^{a-1})\ \ p.s., \text{ et } \sum_{k=0}^{n-1} |X_k^y|^{a-1} \sim n\nu(|\,.\,|^{a-1})\ \ p.s..$$

En sommant sur k, $0 \leq k \leq n-1$ les équations (8.7) on obtient alors, compte tenu de ce qui précède :

$$\frac{1}{n}\sum_{k=0}^{n-1} \big|\,|X_k^x|^a - |X_k^y|^a\big| \xrightarrow{p.s.} 0 \tag{8.8}$$

D'après la remarque suivant la démonstration du théorème 8.1, on sait que :

$$\exists x_0 \in \mathbb{R}^d / \ \frac{1}{n}\sum_{k=0}^{n-1} |X_k|^a \overset{\mathbb{P}_{x_0}\text{-}p.s.}{\longrightarrow} \nu(|\,.\,|^a).$$

Prenant alors $y = x_0$ dans (8.8), on trouve finalement :

$$\forall x \in \mathbb{R}^d, \ \frac{1}{n}\sum_{k=0}^{n-1} |X_k|^a \overset{\mathbb{P}_{x}\text{-}p.s.}{\longrightarrow} \nu(|\,.\,|^a). \quad \square$$

8.2.2 Application aux modèles itératifs p-dépendants

Un modèle itératif p-dépendant est défini par l'équation récursive :

$$\forall t \geq p-1, \ X_{t+1} = F\left(X_t, \cdots, X_{t-p+1}, \varepsilon_{t+1}\right), \tag{8.9}$$

où X_t est à valeurs dans $\mathbb{R}^d$, les ε_t sont i.i.d. à valeurs dans $\mathbb{R}^q$ et F est borélienne de $\left(\mathbb{R}^d\right)^p \times \mathbb{R}^q$ dans $\mathbb{R}^d$.

On entend par loi stationnaire de ce modèle, la mesure invariante du modèle surdimensionné $(Y_t)_{t\geq p}$ défini par :

$$\forall t \geq p-1, \ Y_t := (X_t, \cdots, X_{t-p+1}) \in (\mathbb{R}^d)^p. \tag{8.10}$$

$(Y_t)_{t\geq p}$ suit alors la dynamique markovienne :

$$\forall t \geq p-1, \ Y_{t+1} = F_p\left(Y_t, \varepsilon_{t+1}\right), \tag{8.11}$$

où $F_p : \left(\mathbb{R}^d\right)^p \times \mathbb{R}^q \to (\mathbb{R}^d)^p$ est définie par :

$$F_p(x^1, \cdots, x^p, u) = \ {}^t\left(F(x^1, \cdots, x^p, u), x^1, \cdots, x^{p-1}\right). \tag{8.12}$$

Théorème 8.3 *Soit un modèle itératif p-dépendant défini comme ci-dessus et soit $a \geq 1$. On suppose :*

i) $\mathbb{E}\,|F_p(0, \cdots, 0, \varepsilon_1)|^a < +\infty$,

ii) il existe p réels positifs $(\alpha_i)_{1\leq i\leq p}$ vérifiant $\sum_{i=1}^{p} \alpha_i < 1$ tels que, pour tout $x := (x^1, \cdots, x^p)$, $y := (y^1, \cdots, y^p)$ dans $(\mathbb{R}^d)^p$:

$$\mathbb{E}\left|F_p(x^1, \cdots, x^p, \varepsilon_1) - F_p(y^1, \cdots, y^p, \varepsilon_1)\right|^a \leq \sum_{i=1}^{p} \alpha_i \left|x^i - y^i\right|^a.$$

Alors le modèle surdimensionné vérifie l'hypothèse de mélange lipschitzien.

Démonstration : Il ne s'agit en fait que de vérifier les hypothèses du théorème 8.2. Il est clair que l'on peut supposer tous les α_i strictement positifs. On définit alors la norme $|.|_p$ sur $(\mathbb{R}^d)^p$ par :

$$|x|_p := |(x^1, \cdots, x^p)|_p := \left((\alpha_1 + \cdots + \alpha_p)|x^1|^a + (\alpha_2 + \cdots + \alpha_p)|x^2|^a + \cdots + \alpha_p|x^p|^a\right)^{\frac{1}{a}}.$$

Reprenant la démarche employée pour le critère de Hajek lors de la démonstration de la proposition 3.1, on aboutit, pour $n \geq p-1$, à :

$$\exists\ \alpha \in]0,1[\text{ tel que } \forall x, y \in \left(\mathbb{R}^d\right)^p, \quad \mathbb{E}\left|Y_n^x - Y_n^y\right|_p^a \leq \alpha^{n-p+1}\left|x-y\right|_p^a,$$

ce qui prouve que le modèle surdimensionné vérifie l'hypothèse ii) du théorème 8.2. L'hypothèse i) se vérifiant aisément, on a alors le résultat demandé. □

Bibliographie de la deuxième partie

Attali, J.G., Yao, J.F. (1998) Stabilité des modèles AR non-linéaires à régime markovien, note à paraître dans les *Comptes Rendus de l'Académie des Sciences, Paris.*

Billingsley, P. (1968) *Convergence of probability measures*, John Wiley and sons, New-York, 253p.

Borovkov, A.A. (1990) Ergodicity and stability of multidimensional Markov chains, *Theory Proba. Appl.*, 35, pp.542-546.

Borovkov, A.A. (1991) Lyapounov functions and multidimensional Markov chains, *Theory Proba. Appl.*, 36, pp.1-18.

Breiman, L. (1968) *Probability*, Addison-Wesley, Reading, 421p.

Broniatowski, M., Diebolt, J. (1987) Loi stationnaire et loi des fluctuations pour le processus autorégressif général d'ordre un, *Comptes rendus de l'Académie des Sciences*, t. 305, Série I, pp.203-206.

Çinlar, E. (1975) *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ.

Chung, K.L. (1967) *Markov Chains with Stationnary Transition Probabilities*, Springer-Verlag, Berlin, 2nd edition.

Doeblin, W. (1937) Sur le proprétés asymptotiques de mouvement régis par certains types de chaînes simples constantes, *Bull. Math. Soc. Roum. Sci.*, 39(1), pp.57-115, (2), pp.3-61.

Doeblin, W. (1938) Éléments d'une théorie générale des chaînes simples constantes de Markov, *Annales Scientifiques de l'École Normale Supérieure*, 57(III), pp.61-111.

Doob, J.L. (1953) *Stochastic Processes*, John Wiley and Sons, New York.

Duflo, M. (1990) *Méthodes Récursives Aléatoires*, Collection *Techniques Stochastiques*, Masson, 359p.

Duflo, M. (1996) *Algorithmes Stochastiques*, Coll. *Mathématiques & Applications 23*, Springer-Verlag, Paris, 319p.

Duflo, M. (1997) *Random Iterative Models*, Coll. *Applications of Mathematics, Stochastic Modelling and Applied Probability*, Springer-Verlag, 385p.

Elie, L. Modèles Autorégressifs non Linéaires, cours du DEA de *Statistique et Modèles Aléatoires en Économie et en Finance* des Universités Paris 1-Paris 7.

Ethier, S.N., Kurtz, T.G. (1986) *Markov Processes, Characterization and Convergence*, Wiley series in *Probability and Mathematical Statistics*, New-York, 534p.

Foguel, S.R. (1969) *The Ergodic Theory of Markov Processes*, Van Nostrand Reinhold, New

York.

Gouriéroux, C. Montfort, A. (1990) Qualitative Threshold Arch Models, Institut National de la Statistique et des Etudes Economiques, *Série des documents de travail du CREST et du Département de la Recherche*, n. 9009.

Guégan, D. (1994) *Séries chronologiques non linéaires à temps discret*, Coll. *Statistique mathématique et Probabilité*, Edition Economica, Paris, 301p.

Hajek, B. (1982) Hitting-time and occupation time bounds implied by drift analysis with applications, *Adv. Appl. Proba.*, 29, pp.504-518.

Hall, P., Heyde, C.C. (1980) *Martingale Limit Theory and its Application*, Coll. *Probability and Mathematical Statistics*, Academic Press edition, San Diego, 308p.

Harris, T.E. (1956) The existence of stationnary measures for certain Markov processes, *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, pp. 113-124.

Jamison, B. (1967) Irreducible Markov Operators on $C(S)$, *Proc. Amer. Math. Soc.*, 24, pp.366-370.

Lin, M. (1970) Conservative Markov processes on a topological space, Israel J. Math., 8, pp.165-186.

Meyn, S.P., Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability*, Springer-Verlag, 550p.

Loskot, K., Rudnicki R. (1995) Limit theorems for stochastically perturbed dynamic systems, *J. Appl. Proba.*, 32, pp.459-469.

Mangeas, M., Yao, J.F. (1996) Sur l'estimateur des moindres carrés d'un modèle auto-régressif fonctionnel, *Comptes Rendus de l'Académie des Sciences, Paris*, I-324, pp.471-474.

Neveu, J. (1964) *Bases mathématiques du calcul des probabilités*, Masson, Paris, 203p.

Neveu, J. (1972) *Martingales à temps discret*, Masson, Paris.

Nummelin, E. (1978) *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press, Cambridge.

Orey, S. (1971) *Limit Theorems for Markov Chain Transition Probabilities*, Van Nostrand Reinhold, London.

Revuz, D. (1984) *Markov Chains*, North-Holland, Amsterdam, 2nd edition.

Pakes, A.G. (1969) Some conditions for ergodicity and recurrence of Markov chains, *Oper. Res.*, 17, pp.1048-1061.

Rosenblatt, M. (1974) *Recurrent points and transition function acting on continuous functions*, Z. Wahrscheinlichkeitstheorie und Verw. Geb., 30, pp.173-183.

Rosenthal, J.S. (1995) Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo, *Journal of the American Statistical Association 90*, pp.558-566.

Sine R. (1975) *Sample path convergence of stable Markov processes II*, Indiana University Math., 60, pp.247-252.

Skorokhod, A.V. (1987) Topologically recurrent Markov chains : ergodic properties, *Theory Proba. Appl.*, 31, pp.563-571.

Sunyach, C. (1975) Une classe de chaînes de Markov récurrentes sur un espace métrique complet, *Annales de l'Institut Henri Poincaré*, Section B, Vol. XI, n. 4, pp.325-343.

Vitali, G., Sansone, G. (1951) *Moderna Teoria delle Funzioni di Variabile Reale*, Coll. *Consiglio Nazionale delle Ricerche, Monografie di Matematica Applicata*, Nicola Zanichelli Editore, Bologna, 3nd edition.

Troisième partie

Systèmes dynamiques perturbés par un bruit non évanescent

Chapitre 1

Introduction

De nombreux problèmes issus de la physique, de l'automatique ou encore de l'analyse numérique ont poussé les probabilistes et les statisticiens à s'intéresser à des dynamiques stochastiques de la forme :

$$X_{t+1}^{\gamma} = F^{\gamma}\left(X_t^{\gamma}, \varepsilon_{t+1}\right),$$

où γ est une paramètre réel destiné à tendre vers 0, X_0^{γ} est une variable aléatoire et les $(\varepsilon_t)_{t\in\mathbb{N}^*}$ sont des variables i.i.d de loi μ intégrable centrée et indépendantes de X_0^{γ}. Dans cette modélisation rentrent notamment deux types de problèmes importants, les algorithmes stochastiques et les systèmes dynamiques perturbés. Les algorithmes stochastiques correspondent à des fonctions F^{γ} de la forme :

$$F^{\gamma}(x, \varepsilon) = x + \gamma H(x, \varepsilon).$$

Dans la littérature, il y a principalement deux approches des algorithmes stochastiques, l'idée étant, le plus souvent, de comparer l'algorithme à l'équation différentielle ordinaire associée :

$$\frac{dx}{dt} = h(x) \quad \text{avec} \quad h(x) := \mathbb{E}\left(H(x, \varepsilon_1)\right).$$

Plus précisément, on cherche à montrer que, sous certaines hypothèses, l'algorithme stochastique ne charge que les équilibres stables.

Une première approche (dite à pas décroissant) consiste à choisir un paramètre γ_n tendant vers 0 et à montrer que, presque sûrement, les trajectoires de $X_n^{\gamma_n}$ ne convergent pas vers les équilibres instables de l'équation différentielle ordinaire associée. Parmi ces travaux, on peut citer ceux de (Nevels'on et Has'minskii, 1972), (Benveniste, Métivier et Priouret, 1987), (Pemantle, 1990), (Lazarev, 1992) ou encore (Brandière, 1996).

Une deuxième approche (dite à pas constant) a récemment été développée dans (Fort et Pagès, 1998) et (Benaïm, 1998). L'idée est de s'intéresser aux valeurs d'adhérence faibles ν^0 des probabilités stationnaires ν^{γ} quand le pas γ de l'algorithme tend vers 0, en particulier sous l'existence d'une fonction de Lyapounov globale V pour l'équation différentielle associée. On montre, par exemple, que si l'algorithme est de type (pseudo)-gradient, ν^0 ne peut charger que les minima locaux de la fonction de Lyapounov V.

Un deuxième type de problème est l'étude des systèmes dynamiques perturbés, c'est-à-dire quand les fonctions F^{γ} s'écrivent :

$$F^{\gamma}(x, \varepsilon) = f(x) + \gamma e(x, \varepsilon).$$

De même que pour les algorithmes stochastiques, il existe deux types d'approche, l'idée, cette fois, étant de comparer le comportement du système dynamique perturbé a celui du système dynamique déterministe :

$$x_{t+1} = f(x_t).$$

A propos de l'étude des systèmes perturbés par un bruit décroissant vers 0 (i.e. quand on choisit γ_n tendant vers 0 et que l'on étudie les trajectoires), on peut citer les travaux de (Kesten, 1972), (Pierre Loti Viaud, 1995) ou (Brandière, 1998). On y montre qu'au voisinage d'un point fixe instable de f supposée continue, l'ensemble des trajectoires du sytème convergeant vers ce point fixe est de probabilité nulle.

L'objet de cette partie est l'étude des systèmes dynamiques perturbés par un bruit non évanescent, c'est-à-dire que l'on va s'intéresser à l'asymptotique, quand γ tend vers 0, des probabilités invariantes ν^γ.

On considère donc une famille de processus :

$$(S_\gamma) \equiv \qquad X_{t+1}^\gamma = f(X_t^\gamma) + \gamma e\left(X_t^\gamma, \varepsilon_{t+1}\right),$$

où X_0^γ est une variable aléatoire, les $(\varepsilon_t)_{t\in\mathbb{N}^*}$ sont des variables i.i.d indépendantes de X_0^γ et de loi μ et e et f sont des fonctions mesurables. On s'intéresse, après avoir assuré leur existence, à l'asymptotique d'une famille ν^γ de lois stationnaires pour (S_γ) quand γ tend vers 0.

On montre alors, sous certaines hypothèses, la tension des ν^γ pour $\gamma \in]0, \gamma_0]$. Puis, on vérifie que toute valeur d'adhérence ν^0 est une loi stationnaire du sytème déterministe :

$$(S_0) \equiv \qquad x_{t+1} = f(x_t).$$

Par suite, tout l'intérêt consiste à localiser plus précisément le support des probabilités candidates ν^0, puisque celles-ci forment a priori un sous-ensemble particulier de l'ensemble des lois stationnaires de (S_0). En effet, sous l'effet de la perturbation, un certain nombre de points ou d'ensembles ne peuvent appartenir au support d'une telle valeur d'adhérence. Ce type de problématique a déjà été abordé par des méthodes de grandes déviations dans la littérature, notamment par Y. Kifer dans divers articles et dans son livre (Kifer, 1988) essentiellement consacré à ce type de systèmes dynamiques perturbés (par un bruit décroissant vers 0 ou par un bruit non évanescent). Des résultats analogues ont été obtenus dans (Benaïm, 1998), cette fois dans le cadre des algorithmes stochastiques. Tous ces résultats s'appuient, par nature, sur l'existence de moments exponentiels pour la perturbation. Un question naturelle est alors de savoir quel type de résultat il est possible d'obtenir sous des hypothèses de moments moins restrictives, de type quadratique par exemple. Notre propos est ici d'aborder le même type de problématique dans le cadre des systèmes dynamiques perturbés par un bruit non évanescent. Nous nous appuierons sur l'existence, locale ou globale selon les cas, d'une fonction de Lyapounov V du système dynamique (S_0), supposée de classe C^2, vérifiant $V \circ f \leq V$ et $\{V \circ f = V\} = \{f = Id\}$.

Nous donnons bien sûr des exemples non triviaux de sytèmes dynamiques pour lesquels il est possible de trouver une telle fonction V, comme l'algorithme de Lloyd issu des réseaux de neurones ou n'importe quel système dynamique perturbé construit à partir du flot d'une équation différentielle possédant une fonction de Lyapounov.

Il est alors facile, sous ces hypothèses, de montrer que le support de ν^0 est inclus dans l'ensemble des points fixes de f.

Sous des hypothèses sur f et e, nous montrons que les points fixes également maxima locaux stricts de V ne peuvent pas être chargés par ν^0, et, lorsque f et e sont plus régulières, nous montrons le même résultat pour les points fixes de f points selles de V.

Même si nous transposons aux systèmes dynamiques perturbés par un bruit non évanescent l'approche de (Fort et Pagès, 1998), nous obtenons des différences notables dans la démarche liées au caractère essentiellement local des algorithmes stochastiques.

De plus, contrairement à la majorité des études citées ci-dessus qui ne portent que sur un cadre fellérien, les résultats de la partie II concernant les modèles quasi-fellériens vont nous permettre de traiter certains cas où f et e ne sont pas continues.

Notations : • La valeur absolue sur $\mathbb{R}$ ainsi que toutes les normes euclidiennes sur les espaces $\mathbb{R}^p$ (ou leurs duaux) seront notées $|\,.\,|$, et les normes matricielles associées seront notées $\|\,.\,\|$.

• Si E et F sont deux sous-ensembles d'un espace vectoriel réel de dimension finie, $C(E, F)$, $C^k(E, F)$ et $C_b(E, F)$) désigneront respectivement l'ensemble des fonctions continues, de classe C^k et continues bornées de E dans F.

Chapitre 2

Existence et tension d'une famille de lois invariantes ν^γ et asymptotique quand $\gamma \to 0$

2.1 Existence et tension d'une famille de lois invariantes

2.1.1 Résultat général

Soit $(P^\gamma(x,dy))_{\gamma\in]0,\gamma_0]}$ une famille de transitions markoviennes quasi-fellériennes sur $\left(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\right)$. On s'intéresse à l'existence et à la tension d'une suite de probabilités invariantes (i.e. vérifiant $\nu^\gamma = \nu^\gamma P^\gamma$).

A l'aide du théorème 3.1 de la partie II, nous pouvons étendre au cadre quasi-fellérien l'adaptation due à (Fort et Pagès, 1998) d'un critère d'existence de loi stationnaire (cf. Ethier et Kurtz, 1986) sous l'existence d'une fonction de Lyapounov. Cette démarche est, comme nous allons le voir, très facilement adaptable aux systèmes dynamiques perturbés à bruit non évanescent.

Théorème 2.1 *Soit une famille de transitions* $(P^\gamma(x,dy))_{\gamma\in]0,\gamma_0]}$ *et soit* $V : \mathbb{R}^d \to \mathbb{R}_+$ *une fonction de Lyapounov (i.e.* $V \in C(\mathbb{R}^d, \mathbb{R}_+)$ *et* $\lim\limits_{|x|\to+\infty} V(x) = +\infty$.*)*

Critère de Pakes *On suppose :*

i) $\forall \gamma \in]0,\gamma_0]$, $P^\gamma(x,dy)$ *est quasi-fellérienne,*

ii) il existe une fonction $\psi : \mathbb{R}^d \to \mathbb{R}$ *telle que :*

$$\forall \gamma \in]0,\gamma_0], \quad P^\gamma V - V \leq \lambda(\gamma)\psi + \mu(\gamma),$$

avec $\psi \leq C$ *et* $\lim\limits_{|x|\to+\infty} \psi(x) = -\infty$, $\lambda(\gamma) > 0$ *et* $\mu(\gamma) \in \mathbb{R}$.

Alors,

(a) il existe une famille $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ *de lois* P^γ*-invariantes,*

(b) si en outre $\sup\limits_{\gamma\in]0,\gamma_0]} \dfrac{\mu(\gamma)}{\lambda(\gamma)} < +\infty$, *alors toute famille* $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ *de lois* P^γ*-invariantes est tendue.*

Critère de Hajek *On suppose :*

i) $\forall \gamma \in]0,\gamma_0]$, $P^\gamma(x,dy)$ *est quasi-fellérienne,*

ii) il existe $\alpha(\gamma) \in]0,1[$ *et* $\beta(\gamma) \in \mathbb{R}$ *tels que :*

$$\forall \gamma \in]0,\gamma_0], \quad P^\gamma V \leq \alpha(\gamma) V + \beta(\gamma).$$

Alors,

(a) il existe une famille $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ *de lois* P^γ*-invariantes,*

(b) si en outre $\sup_{\gamma\in]0,\gamma_0]} \dfrac{\beta(\gamma)}{1-\alpha(\gamma)} < +\infty$, *alors toute famille* $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ *de lois* P^γ*-invariantes est tendue et vérifie* $\sup_{\gamma\in]0,\gamma_0]} \nu^\gamma(V) \leq \sup_{\gamma\in]0,\gamma_0]} \dfrac{\beta(\gamma)}{1-\alpha(\gamma)} < +\infty.$

2.1.2 Applications aux SDP

Un système dynamique perturbé par un bruit non évanescent (SDP) est une famille de modèles définie par la donnée d'une fonction borélienne $f : \mathbb{R}^d \to \mathbb{R}^d$, d'un bruit blanc $(\varepsilon_t)_{t\geq 1}$ à valeurs dans $\mathbb{R}^p$ de loi μ définie sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, d'une fonction borélienne $e : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^d$ et des équations récursives :

$$(S_\gamma) \equiv \qquad X^\gamma_{t+1} = f(X^\gamma_t) + \gamma e\left(X^\gamma_t, \varepsilon_{t+1}\right), \quad X_0 \in \mathbb{R}^d.$$

On note $P^\gamma(x,dy)$ la transition du modèle de régression (S_γ). Pour toute fonction $g : \mathbb{R}^d \to \mathbb{R}$ borélienne bornée, on a donc :

$$P^\gamma g(x) := \int_{\mathbb{R}^p} g\left(f(x) + \gamma e(x,y)\right) \mu(dy).$$

Le théorème 2.1 appliqué aux systèmes dynamiques perturbés conduit à des conditions permettant d'assurer l'existence d'une suite ν^γ de lois P^γ-invariantes, ainsi qu'à des conditions assurant la tension de toutes les suites ν^γ.

Pour bien comprendre le cadre quasi-fellérien, nous allons tout d'abord donner les conditions assurant l'existence et la tension des loi invariantes dans le cadre fellérien.

Le cadre fellérien

Proposition 2.1 *Soit un système dynamique perturbé* (S_γ) *et soit une fonction de Lyapounov* $V \in C^1(\mathbb{R}^d, \mathbb{R}_+)$ *telle que* ∇V *est bornée. On suppose que :*

i) f *est continue, et pour tout* $x_0 \in \mathbb{R}^d$, *les fonctions* $x \mapsto e(x,y)$ *sont* $\mu(dy)$*-p.s. continues en* x_0*,*

ii) $\forall x \in \mathbb{R}^d, \quad \displaystyle\int_{\mathbb{R}^p} |e(x,y)|\,\mu(dy) < +\infty.$

Critère de Pakes *Si, en outre, il existe* $a_0 > 0$ *tel que :*

$$\lim_{|x|\to+\infty} V \circ f(x) - V(x) + a_0 \int_{\mathbb{R}^p} |e(x,y)|\,\mu(dy) = -\infty,$$

alors,

(a) il existe $\gamma_0 > 0$ *et une famille* $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ *de lois* P^γ*-invariantes,*

(b) toute famille $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ *de lois* P^γ*-invariantes est tendue.*

Critère de Hajek *Si, en outre, il existe $a_0 > 0$, $\alpha \in]0,1[$ et $\beta \in \mathbb{R}$ tels que :*

$$V \circ f(x) + a_0 \int_{\mathbb{R}^p} |e(x,y)|\, \mu(dy) \leq \alpha V(x) + \beta,$$

alors,

(a) il existe $\gamma_0 > 0$ et une famille $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes,
(b) toute famille $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes est tendue et vérifie :

$$\sup_{\gamma \in]0,\gamma_0]} \nu^\gamma(V) \leq \sup_{\gamma \in]0,\gamma_0]} \frac{\beta(\gamma)}{1-\alpha(\gamma)} < +\infty.$$

Démonstration : (a) et (b) L'hypothèse $i)$ permet classiquement d'assurer que toutes les transitions $P^\gamma(x,dy)$ sont fellériennes.

Par hypothèse, il existe un réel positif M tel que $|\nabla V| \leq M$. L'inégalité des accroissements finis donne alors :

$$P^\gamma V(x) = \int_{\mathbb{R}^p} V\left(f(x) + \gamma e(x,y)\right) \mu(dy) \leq V \circ f(x) + \gamma M \int_{\mathbb{R}^p} |e(x,y)|\, \mu(dy).$$

On est alors clairement, avec $\gamma_0 := \dfrac{a_0}{M}$, sous les conditions d'application du théorème 2.1. □

Le cadre quasi-fellérien

On peut maintenant affaiblir les hypothèses en utilisant les résultats de la partie II.

Proposition 2.2 *Soit un système dynamique perturbé (S_γ) et soit une fonction de Lyapounov $V \in C^1(\mathbb{R}^d, \mathbb{R}_+)$ telle que ∇V est bornée. On suppose que :*

i) $e(x,y) := \widehat{e}\,(F(x), y)$ où F est une fonction de $\mathbb{R}^d$ dans lui-même et pour tout $x_0 \in \mathbb{R}^d$, les fonctions $x \mapsto \widehat{e}(x,y)$ sont $\mu(dy)$-p.s. continues en x_0,

ii) $\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{f(K)} \times \overline{F(K)}, \quad \int_{\mathbb{R}^p} \mathbf{1}_{D_{f,F}} \left(w^1 + \gamma \widehat{e}(w^2, y)\right) \mu(dy) = 0,$
où $D_{f,F}$ désigne la réunion de l'ensemble des points de discontinuité de f et de F,

iii) $\forall x \in \mathbb{R}^d, \quad \int_{\mathbb{R}^p} |e(x,y)|\, \mu(dy) < +\infty.$

Critère de Pakes *Si, en outre, il existe $a_0 > 0$ tel que :*

$$\lim_{|x| \to +\infty} V \circ f(x) - V(x) + a_0 \int_{\mathbb{R}^p} |e(x,y)|\, \mu(dy) = -\infty,$$

alors,

(a) il existe $\gamma_0 > 0$ et une famille $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes,
(b) toute famille $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes est tendue.

Critère de Hajek *Si, en outre, il existe $a_0 > 0$, $\alpha \in]0,1[$ et $\beta \in \mathbb{R}$ tels que :*

$$V \circ f(x) + a_0 \int_{\mathbb{R}^p} |e(x,y)|\, \mu(dy) \leq \alpha V(x) + \beta,$$

alors,

(a) il existe $\gamma_0 > 0$ et une famille $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ de lois P^γ-invariantes,
(b) toute famille $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ de lois P^γ-invariantes est tendue et vérifie :

$$\sup_{\gamma\in]0,\gamma_0]} \nu^\gamma(V) \leq \sup_{\gamma\in]0,\gamma_0]} \frac{\beta(\gamma)}{1-\alpha(\gamma)} < +\infty.$$

Démonstration : D'après la proposition 2.1, les probabilités $\left(Q^\gamma\left((w^1,w^2),dy\right)\right)_{(w^1,w^2)\in(\mathbb{R}^d)^2}$ définies sur $\mathbb{R}^p$ par :

$$Q^\gamma\left((w^1,w^2),dy\right) = w^1 + \gamma b\left(w^2,dy\right)\mu(dy),$$

sont fellériennes en tant qu'applications fonctionnelles sur l'ensemble des fonctions continues de $\mathbb{R}^d$ dans $\mathbb{R}$. De plus :

$$P^\gamma(x,dy) := Q^\gamma\left(H(x),dy\right),$$

avec $H := (f,F)$. L'hypothèse *ii)* permet alors d'affirmer que les transitions P^γ sont quasi-fellériennes. $\square$

2.2 Asymptotique de ν^γ quand $\gamma \to 0$

Dans cette section, nous restons dans le cadre quasi-fellérien. Cependant, comme nous allons le voir, il est nécessaire de faire la distinction entre le cas où f est continue et celui où elle ne l'est pas.

2.2.1 Lois limites quand f est continue

Dans ce paragraphe, nous allons voir que, conformément à l'intuition, les valeurs d'adhérence des lois stationnaires du système dynamique perturbé quand γ tend vers zéro sont des lois stationnaires du sytème déterministe associé :

$$(S_0) \equiv \qquad x_{t+1} = f(x_t).$$

Théorème 2.2 *On considère un système dynamique perturbé par un bruit constant (S_γ). On suppose :*

i) f est continue,

ii) $e(x,y) := \widehat{e}\left(F(x),y\right)$ où F est une fonction de $\mathbb{R}^d$ dans lui-même et pour tout $x_0 \in \mathbb{R}^d$, les fonctions $x \mapsto \widehat{e}(x,y)$ sont $\mu(dy)$-p.s. continues en x_0,

iii) $\forall (w^1,w^2) \in \bigcup_{K\in\mathcal{K}} \overline{f(K)} \times \overline{F(K)}, \quad \int_{\mathbb{R}^p} \mathbf{1}_{D_F}\left(w^1 + \gamma\widehat{e}(w^2,y)\right)\mu(dy) = 0,$
où D_F désigne l'ensemble des points de discontinuité de F,

iv) il existe γ_0 et une famille tendue $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ de lois P^γ-invariantes.

Alors, toute valeur d'adhérence étroite ν^0 de $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ vérifie :

$$\forall g \text{ borélienne bornée de } \mathbb{R}^d \text{ dans } \mathbb{R},\ \nu^0(g) = \nu^0(g\circ f),$$

i.e. $\nu^0 = f(\nu^0)$.

Démonstration : Soit une fonction $g : \mathbb{R}^d \to \mathbb{R}$ lipschitzienne bornée et soit $[g]_1$ son coefficient de Lipschitz. Pour $\gamma \in]0, \gamma_0]$, on a :

$$P^\gamma g(x) = \int_{\mathbb{R}^p} g\left(f(x) + \gamma e(x,y)\right) \mu(dy).$$

Soit $\varepsilon > 0$ quelconque. Il existe un compact K_ε de $\mathbb{R}^d$ tel que $\mu(K_\varepsilon) < \varepsilon$. On a alors, pour tout x dans $\mathbb{R}^d$:

$$\begin{aligned}
|P^\gamma g(x) - g\circ f(x)| &= \left|\int_{\mathbb{R}^p} \left(g\left(f(x) + \gamma e(x,y)\right) - g\circ f(x)\right) \mu(dy)\right| \\
&\leq \int_{\mathbb{R}^p} \left|g\left(f(x) + \gamma e(x,y)\right) - g\circ f(x)\right| \mu(dy) \\
&= \int_{K_\varepsilon} \left|g\left(f(x) + \gamma e(x,y)\right) - g\circ f(x)\right| \mu(dy) \\
&\quad + \int_{{}^cK_\varepsilon} \left|g\left(f(x) + \gamma e(x,y)\right) - g\circ f(x)\right| \mu(dy) \\
&\leq \gamma[g]_1 \int_{K_\varepsilon} |e(x,y)|\mu(dy) + 2\mu(K_\varepsilon)\|g\|_\infty \\
&\leq \gamma[g]_1 \int_{K_\varepsilon} |e(x,y)|\mu(dy) + 2\varepsilon\|g\|_\infty.
\end{aligned}$$

Si ν^γ est une loi invariante pour P^γ, alors :

$$\begin{aligned}
|\nu^\gamma(g) - \nu^\gamma(g\circ f)| &= |\nu^\gamma(P^\gamma g) - \nu^\gamma(g\circ f)| \\
&\leq \nu^\gamma\, |P^\gamma g - g\circ f| \\
&\leq \gamma[g]_1 \int_{K_\varepsilon} |e(x,y)|\mu(dy) + 2\varepsilon\|g\|_\infty.
\end{aligned}$$

g et $g\circ f$ étant continues, en passant à la limite sur γ, on déduit de ce qui précède que, pour toute valeur d'adhérence étroite ν^0 de $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$, on a :

$$\left|\nu^0(g) - \nu^0(g\circ f)\right| \leq 2\varepsilon\|g\|_\infty. \tag{2.1}$$

D'où, puisque ε est quelconque,

$$\nu^0(g) = \nu^0(g\circ f). \tag{2.2}$$

La propriété (2.2) étant vraie pour toute fonction g lipschitzienne bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on en déduit classiquement :

$$\forall g \in C_b(\mathbb{R}^d, \mathbb{R}),\ \nu^0(g) = \nu^0(g\circ f),$$

et, comme ν^0 est une probabilité :

$$\forall g \text{ borélienne bornée de } \mathbb{R}^d \text{ dans } \mathbb{R},\ \nu^0(g) = \nu^0(g\circ f). \quad \square$$

Remarques : • Il est clair que le théorème 2.2 implique que :

$$\forall A \in \mathcal{B}(\mathbb{R}^d),\ \nu^0(A) = \nu^0\left(f^{-1}(A)\right).$$

Pour $A := \overline{f(\mathbb{R}^d)}$, on remarque alors que $\mathrm{supp}\{\nu^0\} \subset \overline{f(\mathbb{R}^d)}$.

• Il n'est pas nécessaire ici de supposer $\displaystyle\int_{\mathbb{R}^p} |e(x,y)|\mu(dy) < +\infty$.

Quand f n'est pas continue, on ne peut rien dire du support des lois limites sans l'hypothèse supplémentaire d'existence d'une fonction de Lyapounov globale pour le système dynamique, puisqu'il est impossible d'assurer que $g\circ f$ est ν^0-*p.s.* continue. C'est ce que nous allons étudier dans la section suivante.

2.2.2 Localisation du support de ν^0

Cas où f est continue

En corollaire de ce qui précède, on peut donner, à l'aide du théorème de récurrence de Poincaré (Krengel, 1985, page 17), un première localisation du support des lois limites sans hypothèses supplémentaires.

Corollaire 2.1 *On considère un système dynamique perturbé (S_γ). On suppose que :*

i) f est continue,

ii) $e(x,y) := \widehat{e}\,(F(x),y)$ où F est une fonction de $\mathbb{R}^d$ dans lui-même et pour tout $x_0 \in \mathbb{R}^d$, les fonctions $x \mapsto \widehat{e}(x,y)$ sont $\mu(dy)$-p.s. continues en x_0,

iii) $\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{f(K)} \times \overline{F(K)}, \quad \int_{\mathbb{R}^p} \mathbf{1}_{D_F}\left(w^1 + \gamma \widehat{e}(w^2,y)\right)\mu(dy) = 0,$

iv) il existe γ_0 et une famille tendue $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes.

Alors, toute valeur d'adhérence étroite ν^0 de $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ vérifie :

$$\operatorname{supp}\left\{\nu^0\right\} \subset B(f) := \left\{x \in \mathbb{R}^d / \ \exists \phi(n) \nearrow +\infty \quad f^{\phi(n)}(x) \longrightarrow x\right\},$$

où $B(f)$ est appelé noyau de Birkoff du système dynamique.

Démonstration : Simple application du théorème de récurrence de Poincaré puisque, d'après le théorème 2.2, ν^0 est une loi stationnaire du système dynamique déterministe. $\square$

On peut préciser le support de ν^0 sous l'existence d'une fonction de Lyapounov globale pour le système dynamique.

Proposition 2.3 *On considère un SDP à (S_γ), et on suppose :*

i) f est continue,

ii) $e(x,y) := \widehat{e}\,(F(x),y)$ où F est une fonction de $\mathbb{R}^d$ dans lui-même et pour tout $x_0 \in \mathbb{R}^d$, les fonctions $x \mapsto \widehat{e}(x,y)$ sont $\mu(dy)$-p.s. continues en x_0,

iii) $\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{f(K)} \times \overline{F(K)}, \quad \int_{\mathbb{R}^p} \mathbf{1}_{D_F}\left(w^1 + \gamma \widehat{e}(w^2,y)\right)\mu(dy) = 0,$

iv) il existe $\gamma_0 > 0$ et une famille tendue $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes,

v) il existe une fonction de Lyapounov globale $V \in C(\mathbb{R}^d, \mathbb{R}_+)$ vérifiant $V \circ f \leq V$.

Alors, pour toute valeur d'adhérence étroite ν^0 de $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$;

$$\operatorname{supp}\{\nu^0\} \subset \{V \circ f = V\} \cap \overline{f(\mathbb{R}^d)}.$$

Démonstration : Soit φ une fonction continue strictement croissante de $\mathbb{R}_+$ dans $\mathbb{R}$, telle que $\varphi(0) = 0$ et $\lim_{x \to +\infty} \varphi(x) = 1$.

$\varphi(V)$ étant continue bornée, le théorème 2.2 permet d'affirmer que, pour toute valeur d'adhérence ν^0 de $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$:

$$\nu^0\left(\varphi(V)\right) = \nu^0\left(\varphi(V \circ f)\right). \tag{2.3}$$

Or φ est strictement croissante et $V \circ f \leq V$, donc $\varphi(V \circ f) \leq \varphi(V)$. L'égalité (2.3) permet alors d'affirmer que :

$$\mathrm{supp}\{\nu^0\} \subset \{\varphi(V \circ f) = \varphi(V)\}.$$

Enfin φ étant bijective, il est clair que :

$$\{\varphi(V \circ f) = \varphi(V)\} = \{V \circ f = V\}. \quad \square$$

Cas où f est discontinue

Proposition 2.4 *On considère un SDP à (S_γ), et on suppose :*

i) $e(x,y) := \widehat{e}\,(F(x),y)$ où F est une fonction de $\mathbb{R}^d$ dans lui-même et pour tout $x_0 \in \mathbb{R}^d$, les fonctions $x \mapsto \widehat{e}(x,y)$ sont $\mu(dy)$-p.s. continues en x_0,

ii) $\forall (w^1, w^2) \in \bigcup_{K \in \mathcal{K}} \overline{f(K)} \times \overline{F(K)}, \quad \int_{\mathbb{R}^p} \mathbf{1}_{D_{f,F}} \left(w^1 + \gamma \widehat{e}(w^2, y)\right) \mu(dy) = 0,$

iii) il existe $\gamma_0 > 0$ et une famille tendue $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes,

iv) il existe une fonction de Lyapounov globale $V \in C(\mathbb{R}^d, \mathbb{R}_+)$ vérifiant $V \circ f \leq V$ et :

$$(x_n \to x \quad \text{et} \quad V(x_n) - V \circ f(x_n) \to 0) \Rightarrow (V(x) - V \circ f(x) = 0).$$

Alors, pour toute valeur d'adhérence étroite ν^0 de $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$;

$$\mathrm{supp}\{\nu^0\} \subset \{V \circ f = V\}.$$

Démonstration : En reprenant la démonstration du théorème 2.2, on remarque que, pour toute fonction g lipschitzienne bornée :

$$\lim_{\gamma \to 0} |\nu^\gamma(g \circ f) - \nu^\gamma(g)| = 0, \tag{2.4}$$

la difficulté de la démonstration provenant du fait que l'on ne peut pas affirmer que $g \circ f$ est ν^0-*p.s.* continue si g est continue.

Considérons toujours une fonction φ lipschitzienne et strictement croissante de $\mathbb{R}_+$ dans $\mathbb{R}$, telle que $\varphi(0) = 0$ et $\lim_{x \to +\infty} \varphi(x) = 1$ et définissons la fonction :

$$\widehat{\varphi}(x) := \liminf_{y \to x} \left(\varphi\,(V)\,(y) - \varphi\,(V \circ f)\,(y)\right).$$

On a évidemment :

$$0 \leq \widehat{\varphi} \leq \varphi\,(V) - \varphi\,(V \circ f).$$

$\widehat{\varphi}$ étant semi-continue inférieurement, on peut donc écrire, compte tenu de l'équation (2.4),

$$0 = \lim_{\gamma \to 0} \nu^\gamma\left(\varphi(V) - \varphi(V \circ f)\right) \geq \liminf_{\gamma \to 0} \nu^\gamma\,(\widehat{\varphi}) \geq \nu^0\,(\widehat{\varphi}) \geq 0, \tag{2.5}$$

ce qui signifie que :

$$\mathrm{supp}\{\nu^0\} \subset \{\widehat{\varphi} = 0\}. \tag{2.6}$$

Pour finir, si $x \in \{\widehat{\varphi} = 0\}$, il existe une suite $(x_n)_{n \in \mathbb{N}}$ tendant vers x telle que :

$$\lim_{n \to +\infty} \left(\varphi\,(V)\,(x_n) - \varphi\,(V \circ f)\,(x_n)\right) = 0.$$

Comme φ est bijective et continue, on a donc :

$$\lim_{n \to +\infty} \left(V(x_n) - V \circ f(x_n)\right) = 0,$$

ce qui prouve, par hypothèse, que $V(x) = V \circ f(x)$ soit encore $x \in \{V = V \circ f\}$. $\square$

2.2.3 Exemples et applications

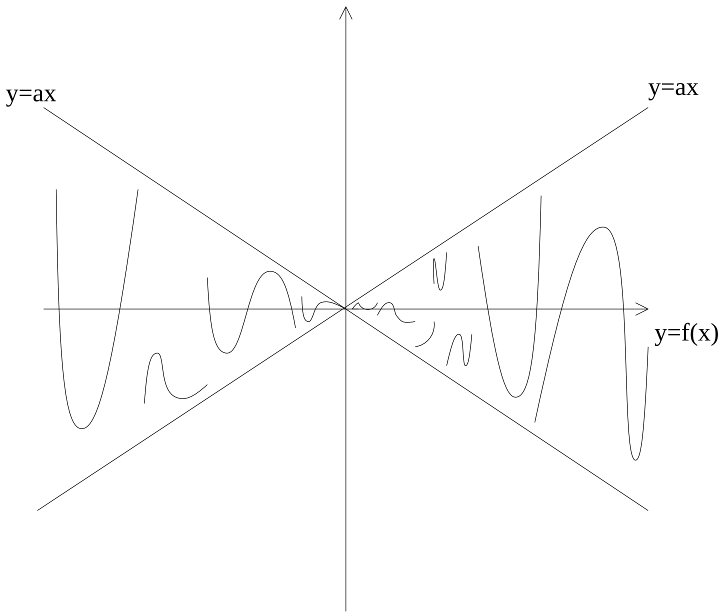
Sous l'existence d'une fonction de Lyapounov globale V, le support des lois limites est inclus dans $\{V \circ f = V\} \cap \overline{f(\mathbb{R}^d)}$. Si l'on cherche à identifier les points succeptibles d'être chargés par les lois limites candidates, il est important de trouver une fonction de Lyapounov V pour le SDP qui soit la plus discriminante possible. En d'autres termes, on cherche V telle que, f étant donnée, $\{V \circ f = V\}$ soit le plus petit possible.

Un exemple dans le cadre quasi-fellérien avec f discontinue

Nous allons illustrer l'existence d'une fonction de Lyapounov dans le cadre de la proposition 2.4. Soit une fonction f borélienne de $\mathbb{R}^d$ dans $\mathbb{R}^d$ telle que :

$$|f(x)| \leq a|x|, \quad 0 < a < 1,$$

ce qui est le cas, par exemple, pour une fonction à valeurs réelles dont le graphe est de la forme :



Nous n'avons besoin d'aucune hypothèse supplémentaire puisque, pour la fonction $V(x) := |x|^2$, il est clair que les hypothèses de la proposition 2.4 sont remplies.

Les lois limites d'un tel système dynamique vont donc toutes converger étroitement vers la mesure de dirac en 0, δ_0. On remarque a posteriori pour ces systèmes que, dès que g est continue bornée, $g \circ f$ est δ_0-$p.s.$ continue, puisque f est continue en 0 (qui peut d'ailleurs être le seul point de continuité de f).

L' algorithme de Lloyd en dimension 1 (Kieffer, 1982)

Il s'agit d'un algorithme issu des réseaux de neurones servant à déterminer numériquement la meilleure approximation (au sens de l'erreur quadratique) d'une variable aléatoire réelle X de loi μ diffuse de carré intégrable avec Conv $(\text{supp}\{\mu\}) = [m, M]$ (où Conv désigne l'enveloppe convexe), par une variable en escalier $Q(X)$ à $(n+1)$ paliers. Cet algorithme est défini comme suit :

Définition de f : On se donne $(m, M) \in \mathbb{R} \cup (-\infty, +\infty)$ et on définit l'ensemble :

$$\mathcal{O}_n := \left\{ (x^1, \cdots, x^n) / \quad m < x^1 < \cdots < x^n < M \right\}.$$

Pour $x := (x^1, \cdots, x^n) \in \mathcal{O}_n$, f est une fonction différentiable de $\mathcal{O}_n$ dans $\mathcal{O}_n$ définie par :

$$f(x^1, \cdots, x^n) = (v^1, \cdots, v^n),$$

où :

$$\forall i \in \{1, \cdots, n\}, \quad v_i := \frac{1}{2} \left(g(x^{i-1}, x^i) + g(x^i, x^{i+1}) \right),$$

avec par convention $x^0 := m$ et $x^{n+1} := M$, et g est définie pour tout $m \le \alpha \le \beta \le M$ par :

$$g(\alpha, \beta) := \frac{\displaystyle\int_\alpha^\beta u\mu(du)}{\mu\left(]\alpha, \beta]\right)}, \alpha < \beta, \quad g(\alpha, \alpha) := \alpha.$$

Définition de V : X est toujours la variable aléatoire définie plus haut.

Soit, pour tout $x := (x^1, \cdots, x^n) \in \mathcal{O}_n$, la fonction Q_x définie par :

$$Q_x(u) := g(x^{i-1}, x^i) \text{ si } u \in [x^{i-1}, x^i[.$$

On définit alors, pour tout $x \in \mathcal{O}_n$, la fonction V par :

$$V(x) := \mathbb{E}\left(Q_x(X) - X\right)^2 = \sum_{i=1}^n \int_{x^{i-1}}^{x^i} \left(g(x^{i-1}, x^i) - u \right)^2 \mu(du).$$

Le résultat s'établit à l'aide des deux lemmes suivants, dont les démonstrations reposent sur une simple différentiation de fonction :

Lemme 2.1 *A $a^1, \cdots, a^n$ fixés dans $\mathbb{R}$, la fonction*

$$(b^1, \cdots, b^{n+1}) \mapsto \sum_{i=1}^{n+1} \int_{a^{i-1}}^{a^i} (b^i - u)^2 \mu(du)$$

est minimum en $b^i := g(a^{i-1}, a^i) = \dfrac{\displaystyle\int_{a^{i-1}}^{a^i} u\mu(du)}{\mu\left(]a^{i-1}, a^i]\right)}$, $1 \le i \le n+1$, *avec, par convention,* $a^0 := m$ *et* $a^{n+1} := M$.

Lemme 2.2 *A $b^1, \cdots, b^{n+1}$ fixés dans $\mathbb{R}$, la fonction*

$$(a^1, \cdots, a^n) \mapsto \sum_{i=1}^{n+1} \int_{a^{i-1}}^{a^i} (b^i - u)^2 \mu(du)$$

est minimum en $a^i := \dfrac{b^i + b^{i+1}}{2}$, $1 \le i \le n$.

D'après la définition des v_i, on montre alors :

$$\begin{aligned} V \circ f(x) = \mathbb{E}\left(Q_{f(x)}(X) - X\right)^2 &= \sum_{i=1}^{n+1} \int_{v^{i-1}}^{v^i} \left(g(v^{i-1}, v^i) - u\right)^2 \mu(du) \\ &\leq \sum_{i=1}^{n+1} \int_{v^{i-1}}^{v^i} \left(g(x^{i-1}, x^i) - u\right)^2 \mu(du) \\ &\leq \sum_{i=1}^{n+1} \int_{x^{i-1}}^{x^i} \left(g(x^{i-1}, x^i) - u\right)^2 \mu(du) = V(x), \end{aligned}$$

en appliquant successivement le lemme 2.1 avec, pour tout $1 \leq i \leq n$, $a^i := v^i$, puis le lemme 2.2 avec, pour tout $1 \leq i \leq n+1$, $b^i := g(x^{i-1}, x^i)$.

Il est facile, par symétrie, d'étendre la fonction f au pavé $]m, M[^n$. En choisissant alors $m = -\infty$ et $M = +\infty$, on a :

$$\forall x \in \mathbb{R}^n, \quad V \circ f(x) \leq V(x) \text{ et } V \circ f(x) = V(x) \Rightarrow f(x) = x.$$

Remarques : • Dans le cas où μ a une densité p *log*-concave, on montre que f a alors un unique point fixe sur $\mathcal{O}_n$ (Cohort, 1998).
• Cet algorithme converge généralement sur $\mathcal{O}_n$ exponentiellement vite vers $Q^*(X)$ solution unique au problème posé. C'est notamment le cas lorsque μ a une densité p strictement *log*-concave ou bien *log*-concave et vérifiant $p(m^+) + p(M^-) > 0$ (Lamberton et Pagès, 1996).
• La fonction V est optimale sur $\mathbb{R}^n$ puisque $\{V \circ f = V\} = \{f = Id\}$.

L' algorithme de Lloyd en dimension supérieure

Avant de définir l'algorithme de Lloyd, nous avons besoin de définir la notion de mosaïque de Voronoï.

Définition 2.1 *Soit C un ensemble compact convexe non vide de $\mathbb{R}^d$. On appelle mosaïque de Voronoï associée au n-uplets $x := (x^1, \cdots, x^n) \in \left(\mathbb{R}^d\right)^n$ (cf e.g. Pagès, 1993), la famille d'ouverts convexes de $\mathbb{R}^d$, $(C_i(x))_{1 \leq i \leq d}$ définie par :*

$$\begin{cases} C_i(x) := \{u \in C /\ x^j \neq x^i \Rightarrow |x^i - u| < |x^j - u|\} & \text{si } i := \min\{k / x^k = x^i\}, \\ C_i(x) := \emptyset & \text{s'il existe } j < i \text{ avec } x^j = x^i. \end{cases}$$

Il est alors possible d'étendre l'algorithme de Lloyd à l'approximation quadratique d'une variable aléatoire X à valeurs dans un compact convexe C de $\mathbb{R}^d$ ($d \geq 1$) et de loi μ diffuse de carré intégrable, par une variable $Q(X)$ constante par paliers sur une mosaïque de Voronoï associée a un n-uplets. L'idée est de généraliser les lemmmes 2.1 et 2.2 de la façon suivante :

Lemme 2.3 *A $a := (a^1 \cdots a^n)$ fixés, la fonction*

$$(b^1, \cdots, b^n) \mapsto \sum_{i=1}^{n} \int_{C_i(a)} |b^i - u|^2 \mu(du)$$

est minimum en $b^i := \dfrac{\int_{C_i(a)} u\mu(du)}{\mu\left(C_i(a)\right)}$, $1 \leq i \leq n$.

Démonstration : On a classiquement, pour toute variable aléatoire Y à valeurs dans $\mathbb{R}^d$:

$$Arg \min_{m\in\mathbb{R}^d} \mathbb{E}|m-Y|^2 := \mathbb{E}Y.$$

On applique alors cette propriété, pour tout $i \in \{1, \cdots, n\}$, aux variables aléatoires Y_i de loi $\dfrac{\mathbf{1}_{C_i(a)}(u)\mu(du)}{\mu\left(C_i(a)\right)}$ pour obtenir le résultat recherché. □

Lemme 2.4 *A $(b^1, \cdots, b^n)$ fixés, la fonction*

$$\left(a^1, \cdots, a^n\right) \mapsto \sum_{i=1}^n \int_{C_i(a)} |b^i - u|^2 \mu(du)$$

est minimum en $a^i := b^i$, $1 \leq i \leq n$.

Démonstration : La mosaïque de Voronoï associée au n-uplet $(b^1, \cdots, b^n)$ formant une partition de C, on peut écrire :

$$\sum_{i=1}^n \int_{C_i(a)} |b^i - u|^2 \mu(du) = \sum_{i=1}^n \sum_{j=1}^n \int_{C_i(a)\cap C_j(b)} |b^i - u|^2 \mu(du). \tag{2.7}$$

Par définition de la mosaïque de Voronoï associée au n-uplet $(b^1, \cdots, b^n)$:

$$\forall u \in C_j(b), \quad |b^i - u| \geq |b^j - u|,$$

ce qui, remplacé dans (2.7), donne :

$$\begin{aligned}\sum_{i=1}^n \int_{C_i(a)} |b^i - u|^2 \mu(du) \geq \sum_{i=1}^n \sum_{j=1}^n \int_{C_i(a)\cap C_j(b)} |b^j - u|^2 \mu(du) &= \sum_{j=1}^n \sum_{i=1}^n \int_{C_i(a)\cap C_j(b)} |b^j - u|^2 \mu(du) \\ &= \sum_{j=1}^n \int_{C_j(b)} |b^j - u|^2 \mu(du). \quad \square\end{aligned}$$

Pour $(x^1, \cdots, x^n) \in C$, on définit, de manière analogue à la dimension 1 :

$$Q_x(u) := \frac{\displaystyle\int_{C_i(x)} u\mu(du)}{\mu\left(C_i(x)\right)} \quad \text{si } u \in C_i(x).$$

On définit également :

$$V(x) := \mathbb{E}\left|Q_x(X) - X\right|^2 \quad \text{et} \quad f(x) = f(x^1, \cdots, x^n) := (v^1, \cdots, v^n),$$

où, pour tout $1 \leq i \leq n$, $v^i := \dfrac{\displaystyle\int_{C_i(x)} u\mu(du)}{\mu\left(C_i(x)\right)}$.

En appliquant alors le lemme 2.3 avec, pour tout $1 \leq i \leq n$, $a^i := v^i$, puis le lemme 2.4 avec, pour tout $1 \leq i \leq n$, $b^i := \dfrac{\int_{C_i(x)} u\mu(du)}{\mu\left(C_i(x)\right)}$, on montre :

$$\begin{aligned}
V \circ f(x) = \mathbb{E}\left|Q_{f(x)}(X) - X\right|^2 &= \sum_{i=1}^n \int_{C_i(f(x))} \left|Q_{f(x)}(u) - u\right|^2 \mu(du) \\
&\leq \sum_{i=1}^n \int_{C_i(f(x))} \left|Q_x(u) - u\right|^2 \mu(du) \\
&\leq \sum_{i=1}^n \int_{C_i(x)} \left|Q_x(u) - u\right|^2 \mu(du) = V(x).
\end{aligned}$$

Flot d'une équation différentielle

On considère une équation différentielle $\dot{x} = -h(x)$, où h est une fonction continue de $\mathbb{R}^d$ dans $\mathbb{R}^d$. On fait l'hypothèse qu'un unique semi-flot Φ existe sur $\mathbb{R}^d \times \mathbb{R}_+$, i.e. pour tout $x \in \mathbb{R}^d$, il existe une unique solution $\Phi(x,.)$ dans $C^1(\mathbb{R}_+, \mathbb{R}^d)$ valant x à l'origine et vérifiant :

$$(ODE_h) \equiv \quad \Phi(x,u) = x - \int_0^u h\left((\Phi(x,v)\right)dv.$$

On supposoe également que (ODE_h) admet une fonction de Lyapounov (globale) i.e. il existe $V \in C^1(\mathbb{R}^d, \mathbb{R}_+)$ telle que $\nabla V.h \geq 0$.

Pour $u_0 > 0$ donné, on considère alors le SDP :

$$(S_\gamma) \equiv \qquad X_{t+1}^\gamma = \Phi(X_t^\gamma, u_0) + \gamma e\left(X_t^\gamma, \varepsilon_{t+1}\right), \quad X_0 \in \mathbb{R}^d.$$

Le théorème des accroissements finis nous permet d'écrire :

$$\forall x \in \mathbb{R}^d,\ \exists \widehat{u}_x \in]0, u_0[\ / \quad \frac{V\left(\Phi(x,u_0)\right) - V(x)}{u_0} = -\nabla V\left(\Phi(x,\widehat{u}_x)\right).h\left(\Phi(x,\widehat{u}_x)\right). \tag{2.8}$$

Comme $\nabla V.h \geq 0$, la relation (2.8) entraîne que la fonction $\Phi(\,.\,, u_0)$ vérifie :

$$V \circ \Phi(\,.\,, u_0) \leq V.$$

On peut tout de suite donner un exemple :

La Lemniscate : Soit la fonction L de $\mathbb{R}^2$ dans $\mathbb{R}^2$ définie par :

$$\forall x := (x_1, x_2) \in \mathbb{R}^2, \quad L(x) := \frac{\left(x_1^2 + x_2^2\right)^2}{16} - x_1 x_2.$$

On pose :

$$V(x) := \frac{L^2}{2\left(1+L^2\right)^{\frac{3}{4}}} \quad \text{et} \quad \theta(x) := \frac{4+L^2}{4\left(1+L^2\right)^{\frac{11}{8}}} \begin{pmatrix} x_2 \frac{x_1^2+x_2^2}{4} - x_1 \\ -x_1 \frac{x_1^2+x_2^2}{4} + x_2 \end{pmatrix}.$$

On trouve alors :

$$\nabla V(x) = \frac{4+L^2}{4\left(1+L^2\right)^{\frac{7}{4}}} \begin{pmatrix} x_1 \dfrac{x_1^2+x_2^2}{4} - x_2 \\ x_2 \dfrac{x_1^2+x_2^2}{4} - x_1 \end{pmatrix}.$$

Pour $h := \nabla V + \theta$, (Fort et Pagès, 1998) montrent que :

$$\{\nabla V.h = 0\} = \left\{(0,0)\,;\left(\sqrt{2},\sqrt{2}\right);\left(-\sqrt{2},-\sqrt{2}\right)\right\}.$$

Par conséquent, toute loi limite ν^0_{alg} de l'algorithme stochastique implémenté par h vérifie :

$$\mathrm{supp}\{\nu^0_{\mathrm{alg}}\} \subset \left\{(0,0)\,;\left(\sqrt{2},\sqrt{2}\right);\left(-\sqrt{2},-\sqrt{2}\right)\right\}.$$

Si l'on se donne $u_0 > 0$ et que l'on reprend l'équation (2.8) dans ce cas, on remarque que si x^* vérifie $V\left(\Phi(x^*,u_0)\right) = V(x^*)$, il existe $\widehat{u}_{x^*} \in]0,u_0[$ tel que $\Phi(x^*,\widehat{u}_{x^*}) \in \{\nabla V.h = 0\}$. Par suite,

$$\Phi(x^*,\widehat{u}_{x^*}) \in \left\{(0,0)\,;\left(\sqrt{2},\sqrt{2}\right);\left(-\sqrt{2},-\sqrt{2}\right)\right\}.$$

Supposons que $\Phi(x^*,\widehat{u}_{x^*}) = (0,0)$. Comme, pout tout $u > 0$, $\Phi\left((0,0),u\right) = (0,0)$, on a en particulier :

$$\Phi\left((0,0),\widehat{u}_{x^*}\right) = (0,0) = \Phi\left(x^*,\widehat{u}_{x^*}\right). \tag{2.9}$$

Or, le théorème de Cauchy-Lipschitz implique qu'il existe une unique solution passant par $(0,0)$ en $\widehat{u}_{x^*}$. On tire donc de l'équation (2.9) que $x^* = (0,0)$. En reproduisant la même démarche pour les deux autres points, on arrive finalement à :

$$\forall u_0 > 0, \quad \{V \circ \Phi(\,.\,,u_0) = V\} = \{\Phi(\,.\,,u_0) = Id\} = \left\{(0,0)\,;\left(\sqrt{2},\sqrt{2}\right);\left(-\sqrt{2},-\sqrt{2}\right)\right\}.$$

Donc, pour $u_0 > 0$ fixé, toute loi limite $\nu^{u_0,0}_{sdp}$ du système dynamique perturbé implémenté par $\Phi(\,.\,,u_0)$ vérifie :

$$\mathrm{supp}\{\nu^{\mathrm{u_0},0}_{\mathrm{sdp}}\} \subset \left\{(0,0)\,;\left(\sqrt{2},\sqrt{2}\right);\left(-\sqrt{2},-\sqrt{2}\right)\right\}.$$

Nous reviendrons sur ce système au chapitre 3 pour montrer, comme pour l'algorithme stochastique, que $\nu^{u_0,0}_{sdp} = \delta_{(0,0)}$.

La fonction "tente"

Cet exemple est important pour souligner que certains sytèmes dynamiques ne possèdent pas de fonction de Lyapounov autre que les fonctions constantes.

La fonction "tente" (appelée comme cela à cause de l'allure de son tracé) est bien connue des spécialistes des systèmes dynamiques déterministes (Collet et Eckeman, 1980). Elle est définie de la manière suivante :

$$\begin{cases} f(x) = 2x, \quad \forall x \in \left[0,\dfrac{1}{2}\right], \\ f(x) = 2(1-x), \quad \forall x \in \left]\dfrac{1}{2},1\right], \\ f(x) = 0, \quad \forall x \in \mathbb{R}\backslash[0,1]. \end{cases}$$

Très simple dans son expression, elle est dite génératrice de chaos car, pour tout ensemble $U \subset [0,1]$ d'intérieur non vide, on a :

$$\lim_{n\to+\infty} f^n(U) = [0,1],$$

où f^n est la n-ième itérée de f (Devaney, 1989). Elle possède également la propriété remarquable de laisser la mesure de Lebesgue sur $[0, 1]$ invariante. En effet, considérons V une fonction continue vérifiant $V \circ f \leq V$ ainsi que l'ensemble des diadiques :

$$D_2 := \left\{ \sum_{i=0}^{+\infty} \frac{\varepsilon_i}{2^{i+1}}, \quad (\varepsilon_i)_{i \geq 0} \in \{0,1\}^{\mathbb{N}} \right\}.$$

Soit $x \in D_2$. D'une part, il est facile de montrer qu'il existe un entier n_x strictement positif tel que $f^{n_x}(x) = 0$. Comme $V \circ f^{n_x} \leq V$, on a alors $V(0) \leq V(x)$.

D'autre part, pour tout entier $n \in \mathbb{N}^*$:

$$f^n\left(\frac{x}{2^n}\right) = 2^n \frac{x}{2^n} = x \text{ car } \frac{x}{2} \leq \frac{1}{2}.$$

On a donc :

$$V(x) = V \circ f^n\left(\frac{x}{2^n}\right) \leq V\left(\frac{x}{2^n}\right). \tag{2.10}$$

Passant à la limite sur n dans la relation (2.10), on obtient, puisque V est continue, $V(x) \leq V(0)$.

Partant, pour tout x dans D_2, $V(x) = V(0)$. D_2 étant dense dans $[0, 1]$ et V étant continue :

$$\forall x \in [0, 1], \quad V(x) = V(0).$$

Dans ce cas précis, $\overline{f(\mathbb{R}^d)} = [0, 1] \subset \{V \circ f = V\}$. V n'apporte donc aucune information supplémentaire sur le support de ν^0 dont on sait qu'il est toujours inclus dans $\overline{f(\mathbb{R}^d)}$.

On peut toutefois préciser que toute loi stationnaire du système dynamique déterministe peut s'écrire sous la forme :

$$\alpha \lambda_{[0,1]} + \sum_{x \in D_2} \beta_x \delta_x,$$

avec $\alpha \geq 0$, $\forall x \in D_2$, $\beta_x \geq 0$ et $\alpha + \sum_{x \in D_2} \beta_x = 1$.

Chapitre 3

Comportement de ν^γ au voisinage d'un point fixe instable

(Brandière, 1998) montre, sous des hypothèses raisonnables, que, pour les systèmes dynamiques perturbés par un bruit décroissant vers 0, l'ensemble des trajectoires convergeant vers x^*, lorsque x^* est un point fixe instable de f vérifiant $\rho\left(Df(x^*)\right) \geq 1$, où $\rho(Df(x^*))$ désigne le rayon spectral de $Df(x^*)$, est de probabilité nulle.

En ce qui nous concerne, nous avons vu que, lorsqu'il existe une fonction de Lyapounov vérifiant :

$$V \circ f \leq V \text{ et } \{V \circ f = V\} = \{f = Id\},$$

le support de toute valeur d'adhérence étroite ν^0 de ν^γ quand γ tend vers 0 est inclus dans $\{f = Id\}$.

On peut faire les remarques suivantes :

P1 Si V et f sont différentiables, on a :
soit $\nabla V(x^*) = 0$, soit $\nabla V(x^*)$ est un vecteur propre de $Df(x^*)$ associé à la valeur propre 1.

En effet, comme $V \circ f \leq V$, tout point x^* dans $\{f = Id\}$ est un minimum de $W := V - V \circ f$. W étant de classe C^1, on a donc $\nabla W(x^*) = 0$, soit encore :

$$\nabla V(x^*) - {}^t Df(x^*) \nabla V\left(f(x^*)\right) = \nabla V(x^*) - {}^t Df(x^*) \nabla V\left(x^*\right) = 0. \quad \square$$

P2 Si f est C^1 au voisinage d'un point fixe x^* de f vérifiant $\rho\left(Df\left(x^*\right)\right) < 1$, alors x^* est un minimum local strict de V.

En effet, f étant C^1 :

$$\exists\, U \in \mathcal{V}(x^*) / \quad \forall x \in U, \quad \rho\left(Df(x)\right) < 1.$$

f est alors contractante sur U et on a :

$$\forall x \in U, \quad \lim_{n \to +\infty} f^n(x) = x^*,$$

et par suite, comme x^* est le seul point fixe de V sur U :

$$\forall x \in U \setminus \{x^*\}, \quad V(x^*) < V(x). \quad \square$$

Dès lors, s'il existe une fonction de Lyapounov au système vérifiant en outre :

$$\{V \circ f = V\} \subset \{\nabla V = 0\},$$

alors, si x^* est un point fixe de f maximum de V, il vérifie $\rho\left((Df(x^*))^{-1}\right) \leq 1$. Si x^* est un point fixe de f point selle, il vérifie $\rho\left(Df(x^*)\right) \geq 1$.

Notre objectif est donc de montrer que les lois limites du système dynamique perturbé par un bruit non évanescent ne chargent pas les maxima locaux ou les points selles de V dès que la perturbation est suffisante en ces points. Ici, nous revenons au cadre fellérien mais nous indiquerons quand il est possible de traiter un cadre quasi-fellérien.

Notre démarche repose sur la proposition suivante elle-même fondée sur un développement de Taylor-Lagrange à l'ordre 2 :

Proposition 3.1 *Soit $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ une famille tendue de lois P^γ-invariantes d'un système dynamique (S_γ) et soit ν^0 l'une de ses valeurs d'adhérence étroite. On suppose que :*

i) f est continue, et pour tout $x_0 \in \mathbb{R}^d$, les fonctions $x \mapsto e(x,y)$ sont $\mu(dy)$-p.s. continues en x_0,

ii) $\forall x \in \mathbb{R}^d$, $\int_{\mathbb{R}^p} e(x,y)\mu(dy) = 0$,

iii) $\forall K$ compact de $\mathbb{R}^d$, $\left(|e(x,y)|^2\right)_{x\in K}$ est $\mu(dy)$-uniformément intégrable,

iv) $x \mapsto \int_{\mathbb{R}^p} |e(x,y)|^2 \mu(dy)$ est $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$-uniformément intégrable.

Alors, pour toute suite $\gamma_n \searrow 0$ telle que $\nu^{\gamma_n} \Rightarrow \nu^0$ et pour toute fonction $g \in C^2(\mathbb{R}^d,\mathbb{R})$ telle que g et $\nabla^2 g$ sont bornées, on a :

$$\lim_{n\to+\infty} \frac{\nu^{\gamma_n}(g) - \nu^{\gamma_n}(g\circ f)}{\gamma_n^2} = \frac{1}{2}\int_{\mathbb{R}^d} \nu^0(dx) \int_{\mathbb{R}^p} {}^t e(x,y)\nabla^2 g\left(f(x)\right) e(x,y)\mu(dy).$$

Démonstration : Le développement de Taylor-Lagrange de g au point $f(x)$ à l'ordre 2 donne :

$$\begin{aligned}\left| g\left(f(x)+\gamma e(x,y)\right) - g\left(f(x)\right) - \gamma\nabla g\left(f(x)\right) e(x,y) - \frac{1}{2}\gamma^2\, {}^t e(x,y)\nabla^2 g\left(f(x)\right) e(x,y)\right| \\ \leq \frac{1}{2}\gamma^2 |e(x,y)|^2 \left(w\left(\nabla^2 g, f(x), \gamma e(x,y)\right) \wedge \left\|\nabla^2 g\right\|_\infty \right),\end{aligned}$$

où $w(h,x,.)$ désigne le module de continuité de h en x.

En intégrant successivement par rapport à μ et à ν^γ, on obtient, puisque $\int_{\mathbb{R}^p} e(x,y)\mu(dy) = 0$:

$$\left| \frac{\nu^\gamma(g\circ f) - \nu^\gamma(g)}{\gamma^2} - \frac{1}{2}\int_{\mathbb{R}^d}\nu^\gamma(dx)\int_{\mathbb{R}^p} {}^t e(x,y)\nabla^2 g\left(f(x)\right) e(x,y)\mu(dy)\right| \leq \nu^\gamma\left(G_\gamma\right), \tag{3.1}$$

où $G_\gamma(x) := \frac{1}{2}\int_{\mathbb{R}^p} |e(x,y)|^2 \left(w\left(\nabla^2 g, f(x), \gamma e(x,y)\right) \wedge \left\|\nabla^2 g\right\|_\infty\right)\mu(dy)$.

Pour tout x, $G_\gamma(x)$ étant décroissante en γ, on a :

$$\forall \eta \geq \gamma, \quad \nu^\gamma\left(G_\gamma\right) \leq \nu^\gamma\left(G_\eta\right).$$

Les hypothèses $i)$ et $iii)$ permettent d'affirmer que G_η est continue et l'hypothèse $iv)$ qu'elle est $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$-uniformément intégrable. De $\nu^{\gamma_n} \Rightarrow \nu^0$, on tire alors :

$$\lim_{n\to+\infty} \nu^{\gamma_n}\left(G_\eta\right) = \nu^0\left(G_\eta\right).$$

Enfin, le théorème de convergence dominée permet d'affirmer que :

$$\lim_{\eta \to 0} \nu^0 (G_\eta) = 0,$$

ce qui, remplacé dans (3.1), termine la preuve. $\square$

3.1 Étude au voisinage d'un maximum local strict de V

3.1.1 Résultat global dans le cadre fellérien

Sous l'existence d'une fonction de Lyapounov globale pour le système dynamique, on va donner les conditions assurant qu'un maximum local strict n'est chargé par aucune loi limite.

Théorème 3.1 *On considère un SDP (S_γ) tel qu'il existe une famille tendue $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes. On suppose également que les hypothèses de la proposition 3.1 sont remplies. Si en outre :*

i) il existe V de classe C^2 telle que :

$$\begin{array}{ll} (\alpha) & V \circ f \leq V \;\; et \;\; \{V \circ f = V\} = \{f = Id\}, \\ (\beta) & \forall K > 0, \quad \sup_{\{V \leq K\}} \left(|\nabla V| + \left\| \nabla^2 V \right\| \right) < +\infty, \end{array}$$

ii) $\forall v > 0$, $\{f = Id\} \cap \{V \leq v\}$ est fini.

Alors, pour tout point fixe x^* de f, maximum local de V tel que $\nabla^2 V(x^*)$ soit définie négative et pour toute valeur d'adhérence étroite ν^0 de ν^γ, on a :

$$\nu^0 (\{x^*\}) = 0.$$

Démonstration : PREMIÈRE ÉTAPE : Nous allons d'abord transformer V en une fonction V^* de classe C^2 telle que $V^* \circ f \leq V^*$, $\nabla^2 V^*(x^*)$ soit définie négative, et enfin telle que V^*, ∇V^* et $\nabla^2 V^*$ soient bornées.

Soit $K > V(x^*)$ et soit une fonction $\varphi_1 \in C^\infty(\mathbb{R}_+, \mathbb{R}_+)$ telle que :

$$\varphi_1(0) = 0, \; \forall y \in [0, K[, \; \varphi_1'(y) > 0 \text{ et } \forall y \in [K, +\infty[, \; \varphi_1(y) = cte.$$

On définit alors la fonction $V^* := \varphi_1 \circ V$. L'hypothèse $i)$ (β) permet alors d'affirmer que $\nabla V^* := \varphi_1'(V) \nabla V$ et $\nabla^2 V^* := \varphi_1''(V) \nabla V \, {}^t\nabla V + \varphi_1'(V) \nabla^2 V$ sont bornées.

De plus, φ_1 étant croissante, on a encore $V^* \circ f \leq V^*$.

Enfin $\nabla V^*(x^*) = 0$ et $\nabla^2 V^*(x^*) = \varphi_1' (V(x^*)) \nabla^2 V(x^*)$ est définie négative puisque $\nabla^2 V(x^*)$ est définie négative et $\varphi_1' (V(x^*)) > 0$.

DEUXIÈME ÉTAPE : On va maintenant chercher une nouvelle fonction de Lyapounov W telle que x^* soit le seul point fixe au niveau $W = W(x^*)$.

$\nabla^2 V^*(x^*)$ étant définie négative, il existe $\varepsilon_0 > 0$ et une boule $B(x^*, \varepsilon_0)$ tels que :

$$\forall x \in B(x^*, \varepsilon_0) \setminus \{x^*\}, \quad V^*(x) < V^*(x^*).$$

Soit $\alpha_0 := \sup_{x \in B(x^*, \varepsilon_0)} (V^*(x^*) - V^*(x))$. Il est clair que $\alpha_0 > 0$, car $B(x^*, \varepsilon_0)$ est compact et V^* est continue.

Soit maintenant les ensembles définis par :

$$\forall \alpha \in]0, \alpha_0], \quad \Omega_\alpha := \{x \in B(x^*, \varepsilon_0) / \quad V^*(x^*) - \alpha \leq V^*(x)\} .$$

Comme x^* est un point fixe de f et $V^* \circ f$ est continue, il existe un réel $\alpha_1 \in]0, \alpha_0]$ vérifiant :

$$f(\Omega_{\alpha_1}) \subset \Omega_{\alpha_0}.$$

Soit β quelconque dans $]0, \alpha_1[$. Il est possible de trouver une fonction $\varphi_2 \in C^\infty(\mathbb{R}_+, \mathbb{R})$ telle que :

$$\begin{aligned}
&\varphi_2(V^*(x^*) - \alpha_1) := V^*(x^*) - \alpha_1 \quad \text{et} \quad \varphi_2(V^*(x^*)) := V^*(x^*) - \beta,\\
&\varphi_2'(V^*(x^*) - \alpha_1) := 1 \quad \text{et} \quad 0 < \varphi_2'(V^*(x^*)) < 1,\\
&\varphi_2''(V^*(x^*) - \alpha_1) := 0 \quad \text{et} \quad \forall y \in [V^*(x^*) - \alpha_1, V^*(x^*)], \varphi_2''(y) \leq 0.
\end{aligned}$$

On définit alors la fonction W_β par :

$$\begin{aligned}
&\forall x \in \Omega_{\alpha_1}, \quad W_\beta(x) := \varphi_2 \circ V^*(x),\\
&\forall x \in {}^c\Omega_{\alpha_1}, \quad W_\beta(x) := V^*(x).
\end{aligned}$$

W_β est, par construction, de classe C^2 et vérifie $W_\beta \leq V^*$. Montrons que $W_\beta \circ f \leq W_\beta$.

Pour $x \in {}^c\Omega_{\alpha_1}$, on a :

$$W_\beta \circ f(x) \leq V^* \circ f(x) \leq V^*(x) = W_\beta(x).$$

Prenons maintenant $x \in \Omega_{\alpha_1}$. Deux cas de figure peuvent se présenter. Soit $f(x) \in \Omega_{\alpha_1}$, auquel cas, comme φ_2 est croissante sur son domaine de définition :

$$W_\beta \circ f(x) = \varphi_2(V^* \circ f(x)) \leq \varphi_2(V^*(x)) = W_\beta(x).$$

Soit $f(x) \in \Omega_{\alpha_0} \setminus \Omega_{\alpha_1}$, puisque $f(\Omega_{\alpha_1}) \subset \Omega_{\alpha_0}$, et alors :

$$W_\beta \circ f(x) = V^* \circ f(x) \leq V^*(x^*) - \alpha_1 \leq \varphi_2(V^*(x)) = W_\beta(x).$$

D'après l'hypothèse *ii)* il existe β_0 tel que x^* soit le seul point fixe de f au seuil :

$$W_{\beta_0}(x^*) = V^*(x^*) - \beta_0,$$

il existe donc $\eta_0 > 0$ tel que x^* soit le seul point fixe de f vérifiant :

$$W_{\beta_0}(x^*) \in [W_{\beta_0}(x^*) - \eta_0, W_{\beta_0}(x^*) + \eta_0] .$$

Remarquons que, puisque V^*, ∇V^* et $\nabla^2 V^*$ sont bornées, W_{β_0}, ∇W_{β_0} et $\nabla^2 W_{\beta_0}$ le sont également. De plus $\nabla^2 W_{\beta_0}(x^*) = \varphi_2'(V^*(x^*)) \nabla^2 V^*(x^*)$ est définie négative.

Troisième étape : On va encore transformer W_{β_0} pour obtenir une nouvelle fonction de Lyapounov $\widehat{W}$ telle que x^* soit le seul point fixe pour lequel $\nabla^2 \widehat{W}(x^*) \neq 0$.

Pour ce faire, choisissons maintenant une fonction $\varphi_3 \in C^\infty(\mathbb{R}, \mathbb{R})$ telle que :

$$\begin{aligned}
&\forall y \in \left]-\infty, W_{\beta_0}(x^*) - \eta_0\right] \varphi_3(y) := 0, \text{ et } \forall y \in [W_{\beta_0}(x^*) + \eta_0, +\infty[, \varphi_3(y) := 1,\\
&\forall y \in [W_{\beta_0}(x^*) - \eta_0, W_{\beta_0}(x^*) + \eta_0], \varphi_3'(y) > 0.
\end{aligned}$$

La fonction $\widehat{W} := \varphi_3 \circ W_{\beta_0}$ est alors de classe C^2, vérifie $\widehat{W} \circ f \leq \widehat{W}$ puisque φ_3 est croissante et $\nabla^2 \widehat{W}(x^*) = \varphi_3^{'}\left(W_{\beta_0}(x^*)\right) \nabla^2 W_{\beta_0}(x^*)$ est définie négative.

En outre, $\widehat{W}$, $\nabla \widehat{W}$ et $\nabla^2 \widehat{W}$ sont bornées car W_{β_0}, ∇W_{β_0} et $\nabla^2 W_{\beta_0}$ sont bornées.

Enfin, $\widehat{W}$ est constante au voisinage de tout point fixe différent de x^*, donc la Hessienne de $\widehat{W}$ est nulle en ces points.

Quatrième étape : Pour finir, soit ν^0 une valeur d'adhérence de $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ et soit une suite $\gamma_n \searrow 0$ telle que $\nu^{\gamma_n} \Rightarrow \nu^0$.

L'hypothèse $i)$ (α) implique que le support de ν^0 est inclus dans l'ensembles des points fixes de f. En appliquant la proposition 3.1 à $\widehat{W}$, on obtient alors :

$$\lim_{n\to+\infty} \frac{\nu^{\gamma_n}(\widehat{W}) - \nu^{\gamma_n}(\widehat{W}\circ f)}{\gamma_n^2} = \frac{\nu^0\left(\{x^*\}\right)}{2} \int_{\mathbb{R}^p} {}^t e(x^*,y) \nabla^2 \widehat{W}\,(x^*)\, e(x^*,y) \mu(dy). \tag{3.2}$$

Comme $\widehat{W} - \widehat{W} \circ f \geq 0$ et que $\displaystyle\int_{\mathbb{R}^p} {}^t e(x^*,y) \nabla^2 \widehat{W}\,(x^*)\, e(x^*,y)\mu(dy) < 0$, la relation (3.2) permet de conclure que :

$$\nu^0\left(\{x^*\}\right) = 0. \quad \square$$

Remarques : • S'il existe une fonction de Lyapounov V de classe C^2 telle que $V \circ f \leq V$ et qu'en outre $\displaystyle\lim_{|x|\to+\infty} V(x) = +\infty$, il est clair que l'hypothèse $i)$ (β) est vérifiée puisque pour tout $K > 0$, $\{V \leq K\}$ est compact.

• Comme dans le papier (Fort et Pagès, 1998) on peut relaxer en x^* maximum local strict puis imposer que $\displaystyle\int_{\mathbb{R}^p} {}^t e(x^*,y) \nabla^2 \widehat{W}\,(x^*)\, e(x^*,y)\mu(dy) < 0$, ce qui peut arriver sans que $\nabla^2 \widehat{W}\,(x^*) < 0$ (sachant que $\nabla^2 \widehat{W}\,(x^*) \leq 0$). En fait, il suffit de supposer qu'il existe un vecteur propre non nul u_1^* de $\nabla^2 V\,(x^*)$ (donc de $\nabla^2 \widehat{W}\,(x^*)$!) tel que $\mathbb{P}\left(e(x^*,u).u_1^* > 0\right) > 0$.

Application à la Lemniscate : Nous avons vu, au chapitre 2.2, que pour $u > 0$ fixé, le support de toute loi limite $\nu_{sdp}^{u,0}$ du système dynamique perturbé implémenté par $\Phi(\,.\,,u)$ est inclus dans $\left\{(0,0)\,;\left(\sqrt{2},\sqrt{2}\right);\left(-\sqrt{2},-\sqrt{2}\right)\right\}$.

On a :

$$\nabla^2 V\left(\sqrt{2},\sqrt{2}\right) = \nabla^2 V\left(-\sqrt{2},-\sqrt{2}\right) = -\frac{5}{2^{\frac{11}{4}}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Le théorème 3.1 permet alors d'affirmer que :

$$\nu_{sdp}^{u,0}\left(\left\{\left(\sqrt{2},\sqrt{2}\right);\left(-\sqrt{2},-\sqrt{2}\right)\right\}\right) = 0.$$

On retrouve donc le même résultat pour le système dynamique perturbé que pour l'algorithme stochastique, à savoir que pour tout $u > 0$:

$$\nu_{sdp}^{u,\gamma} \Rightarrow \delta_{(0,0)},$$

puisque $\delta_{(0,0)}$ est la seule valeur d'adhérence possible.

3.1.2 Résultat local dans le cadre fellérien

On peut également donner un résultat sous l'existence d'une fonction de Lyapounov locale.

Proposition 3.2 *Soit une famille tendue $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ de lois P^γ-invariantes pour un SDP (S_γ) et soit x^* un point fixe de f. On suppose que les hypothèses de la proposition 3.1 sont remplies. On suppose, en outre, qu'il existe une boule $B(x^*,\varepsilon_0)$ $(\varepsilon_0>0)$ et une fonction V de classe C^2 telles que :*

i) $f^{-1}\left(B(x^*,\varepsilon_0)\right)\subset B(x^*,\varepsilon_0)$,

ii) $\nabla^2 V(x^*)$ est définie négative et :

$$\forall x\in B(x^*,\varepsilon_0)\setminus\{x^*\},\quad V\circ f(x)<V(x)\quad \text{et}\quad V(x)<V(x^*),$$

(x^* est donc un maximum local strict de V).

Alors, pour toute valeur d'adhérence étroite ν^0 de ν^γ, on a :

$$\nu^0\left(\{x^*\}\right)=0.$$

Remarque : L'hypothèse $i)$ est cohérente avec le fait que x^* est un maximum de V. En effet, $\rho\left((Df(x^*))^{-1}\right)\leq 1$ donc f est moralement dilatante au voisinage de x^*. En revanche, lors de l'étude des points selles, nous verrons que cette hypothèse n'est pas raisonnable.

Démonstration de la proposition 3.2 : Soit α_0 défini par :

$$\alpha_0:=\inf_{Fr(B(x^*,\varepsilon_0))}\left(V(x^*)-V(x)\right),$$

où Fr désigne la frontière.

V étant continue, il est clair que $\alpha_0>0$. On définit alors :

$$\forall\alpha\in]0,\alpha_0],\quad \Omega_\alpha:=\left\{x\in B(x^*,\varepsilon_0)/\quad V(x^*)-\alpha\leq V(x)\right\}.$$

Il est clair (voir la deuxième étape de la démonstration du théorème 3.1) qu'il existe un réel $\alpha_1\in]0,\alpha_0]$ tel que :

$$f\left(\Omega_{\alpha_1}\right)\subset\Omega_{\alpha_0}.$$

Soit maintenant une fonction φ de classe C^∞, strictement croissante sur $[V(x^*)-\alpha_1,V(x^*)]$ et telle que :

$$\varphi\left(V(x^*)\right):=1,\quad \text{et}\quad \forall y\in\left]-\infty,V(x^*)-\alpha_1\right],\quad \varphi(y):=0,\quad \text{et}\quad \varphi'\left(V(x^*)\right):=1.$$

On pose enfin :

$$\begin{cases} V^*(x):=\varphi\circ V(x), & \text{pour } x\in\Omega_{\alpha_1}\\ V^*(x):=0, & \text{sinon.}\end{cases}$$

Montrons maintenant que $V^*\circ f\leq V^*$. Il est clair que le résultat est vrai pour $x\in\Omega_{\alpha_1}$. De même, l'hypothèse $i)$ implique clairement le résultat sur ${}^cB(x^*,\varepsilon_0)$ puisque $V^*\circ f(x)$ et $V^*(x)$ sont toutes deux nulles.

Soit alors $x\in B(x^*,\varepsilon_0)\setminus\Omega_{\alpha_1}$, on a donc $V^*(x)=0$. Par définition de Ω_{α_1}, on a également $V(x)<V(x^*)-\alpha_1$. Si $f(x)$ appartenait à Ω_{α_1}, on aurait alors :

$$V(x)<V(x^*)-\alpha_1\leq V\circ f(x),$$

ce qui est contradictoire avec l'hypothèse $ii)$. $f(x)$ n'appartenant pas à Ω_{α_1}, on a donc $V^*\circ f(x)=0$, d'où le résultat.

Pour finir, V^* est évidemment de classe C^2, $\nabla^2 V^*(x^*)$ est définie négative et V^* ainsi que $\nabla^2 V^*$ sont bornées. On conclut, de la même manière que pour le théorème 3.1, par la proposition 3.1. □

Remarque : On note que l'hypothèse *i)* est très contraignante puisqu'elle signifie que partant d'un point de ${}^c B(x^*, \varepsilon_0)$, le système dynamique déterministe ne revient jamais dans $B(x^*, \varepsilon_0)$. C'est là une différence fondamentale entre le système dynamique perturbé et l'algorithme stochastique. En effet, l'algorithme stochastique a un caractère essentiellement local puisque, dès que le pas est suffisamment petit, on reste au voisinage de x^*.

En revanche, suivant la forme de f, il est possible que le système dynamique s'éloigne de x^* et y revienne une infinité de fois, quel que soit le bruit, même si celui-ci est nul. C'est le problème qui se pose, par exemple, pour le système dynamique déterministe implémenté par la fonction tente. En effet, partant d'un irrationnel aussi proche que l'on veut du point fixe 0, le système dynamique déterministe revient une infinité de fois arbitrairement près de 0, puisqu'il "remplit" l'intervalle $[0,1]$. C'est la raison pour laquelle il n'existe aucune fonction de Lyapounov globale, ni même locale pour cette fonction. L'hypothèse *i)* est donc une hypothèse de localisation pour le système dynamique.

3.1.3 Extension des résultats précédents au cadre quasi-fellérien

Le problème ici est que l'on ne peut plus appliquer la proposition 3.1 puisque les fonctions G_γ que nous avions défini par :

$$G_\gamma(x) := \frac{1}{2}\int_{\mathbb{R}^p} |e(x,y)|^2 \left(w\left(\nabla^2 g, f(x), \gamma e(x,y)\right) \wedge \left\|\nabla^2 g\right\|_\infty\right)\mu(dy),$$

ne sont plus continues. On ne peut donc plus affirmer que l'on a :

$$\lim_{n\to+\infty} \nu^{\gamma_n}(G_\eta) = \nu^0(G_\eta).$$

Afin de récupérer cette dernière propriété, il faut assurer que les fonctions G_γ sont ν^0-*p.s.* continues. Cela nous amène donc à énoncer le théorème :

Théorème 3.2 *Soit $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$ une famille tendue de lois P^γ-invariantes d'un système dynamique (S_γ) et soit ν^0 l'une de ses valeurs d'adhérence étroite. On suppose que :*

i) $e(x,y) := \widehat{e}(F(x), y)$ où F est une fonction de $\mathbb{R}^d$ dans lui-même et pour tout $x_0 \in \mathbb{R}^d$, les fonctions $x \mapsto \widehat{e}(x,y)$ sont $\mu(dy)$-p.s. continues en x_0,

ii) $\forall (w^1, w^2) \in \bigcup_{K\in\mathcal{K}} \overline{f(K)} \times \overline{F(K)}, \quad \int_{\mathbb{R}^p} \mathbf{1}_{D_{f,F}}\left(w^1 + \gamma\widehat{e}(w^2,y)\right)\mu(dy) = 0,$

iii) $\forall x \in \mathbb{R}^d, \ \int_{\mathbb{R}^p} e(x,y)\mu(dy) = 0,$

iv) $\forall K$ compact de $\mathbb{R}^d$, $\left(|e(x,y)|^2\right)_{x\in K}$ est $\mu(dy)$-uniformément intégrable,

v) $x \mapsto \int_{\mathbb{R}^p} |e(x,y)|^2 \mu(dy)$ est $(\nu^\gamma)_{\gamma\in]0,\gamma_0]}$-uniformément intégrable,

vi) il existe V de classe C^2 telle que :

$$\begin{aligned}
&(\alpha) \quad V\circ f \le V \ \ et \ \{V\circ f = V\} = \{f = Id\},\\
&(\beta) \quad (x_n \to x \quad et \quad V(x_n) - V\circ f(x_n) \to 0) \Rightarrow (V(x) - V\circ f(x) = 0),\\
&(\delta) \quad \forall K > 0, \quad \sup_{\{V\le K\}} \left(|\nabla V| + \left\|\nabla^2 V\right\|\right) < +\infty,
\end{aligned}$$

vii) $\forall v > 0$, $\{f = Id\} \cap \{V \leq v\}$ *est fini,*

viii) $\{f = Id\} \cap D_{f,F} = \emptyset$

Alors, pour tout point fixe x^* de f, maximum local de V tel que $\nabla^2 V(x^*)$ soit définie négative et pour toute valeur d'adhérence étroite ν^0 de ν^γ, on a :

$$\nu^0\left(\{x^*\}\right) = 0.$$

Démonstration : La proposition 2.4 permet d'affirmer que le support de ν^0 est inclus dans $\{V \circ f = V\} = \{f = Id\}$. Les fonctions G_γ sont donc ν^0-*p.s.* continues par l'hypothèse *viii*). Les conclusions de la proposition 3.1 sont donc valides. Par suite, celle du théorème 3.1 le sont également. □

De la même manière, on pourrait étendre le résultat global de la proposition 3.2 au cadre quasi-fellérien.

3.2 Étude au voisinage d'un point selle de V

Contrairement aux maxima locaux, nous ne pouvons faire ici qu'un traitement par une fonction de Lyapounov globale.

En effet, supposons qu'il existe une boule $B(x^*, \varepsilon_0)$ ($\varepsilon_0 > 0$) et une fonction V de classe C^2 telles que $V \circ f < V$ sur la boule privée de x^*. Alors, si x est un point hors de la boule, que $V(x)$ est strictement plus grand que $V(x^*)$, mais que x est tout de même assez proche de x^*, il est très vraisemblable que les itérés par f de x vont rentrer dans $B(x^*, \varepsilon_0)$, puisque le système déterministe décroît "le long" de V. Cela traduit simplement le fait qu'il peut exister des valeurs propres de $\nabla^2 V(x^*)$ strictement inférieures à 1.

Une hypothèse du type $f^{-1}\left(B(x^*, \varepsilon_0)\right) \subset B(x^*, \varepsilon_0)$, où x^* est un point selle de V, paraît donc incohérente. Cela illustre, une fois encore, la différence fondamentale que nous soulignions précédemment entre l'algorithme stochastique et le système dynamique perturbé.

Nous avons vu, en introduction de ce chapitre que, si x^* est un point fixe de f qui soit un point selle de V, alors $\rho\left(Df(x^*)\right) \geq 1$. Dans ce cas, le polynôme minimal de $Df(x^*)$ peut se décomposer en un produit de deux polynômes à coefficients réels, l'un ayant des racines de modules strictement inférieurs à 1, et l'autre ayant des racines de modules supérieurs ou égal à 1. Nous noterons respectivement $E^*_{<1}$ et $E^*_{\geq 1}$ les espaces correspondant à cette décomposition pour $Df(x^*)$.

Théorème 3.3 *On considère une famille tendue $(\nu^\gamma)_{\gamma \in]0,\gamma_0]}$ de lois P^γ-invariantes d'un SDP (S_γ) quasi-fellérien. On se place sous les hypothèses de la proposition 3.1 et on suppose en outre :*

i) f est de classe C^2,

ii) il existe V de classe C^2 telle que :

$$\begin{array}{ll} (\alpha) & V \circ f \leq V \text{ et } \{V \circ f = V\} = \{f = Id\}, \\ (\beta) & \forall K > 0, \quad \sup\limits_{\{V \leq K\}} \left(|\nabla V| + \left\|\nabla^2 V\right\| + \|Df\| + \left\|D^2 f\right\|\right) < +\infty, \end{array}$$

iii) $\forall v > 0$, $\{f = Id\} \cap \{V \leq v\}$ *est fini.*

Soit x^* un point fixe de f, point selle de V, tel que :

(α) $\nabla^2 V(x^*) - {}^tDf(x^*)\nabla^2 V(x^*)Df(x^*)$ soit non dégénérée,
(β) pour la projection Π^* sur $E^*_{\geq 1}$ parallèlement à $E^*_{<1}$, on a :

$$\int_{\mathbb{R}^p} |\Pi^* e(x^*, y)|\, \mu(dy) > 0.$$

Alors, pour toute valeur d'adhérence étroite ν^0 de ν^γ, on a :

$$\nu^0 \left(\{x^*\}\right) = 0.$$

Remarque : $E^*_{\geq 1}$ est toujours de dimension supérieure ou égale à un. Si $E^*_{<1}$ est réduit à $\{0\}$, Π devient l'idendité de $\mathbb{R}^d$.

Démonstration du théorème 3.3 : PREMIÈRE ÉTAPE : Reprenons la fonction bornée V^* de classe C^2 définie lors de la première étape de la démonstration du théorème 3.1 par $V^* := \varphi_1 \circ V$, où $K > V(x^*)$ et $\varphi_1 \in C^\infty(\mathbb{R}_+, \mathbb{R}_+)$ est telle que :

$$\varphi_1(0) = 0,\ \forall y \in [0, K[,\ \varphi_1'(y) > 0 \text{ et } \forall y \in [K, +\infty[,\ \varphi_1(y) = cte.$$

L'idée est d'appliquer, pour tout entier $p \in \mathbb{N}^*$, la proposition 3.1 aux fonctions $V^* \circ f^p$ où f^p désigne la p-ième itérée de f.

Montrons donc que, pour tout entier $p \in \mathbb{N}^*$, $\nabla^2 \left(V^* \circ f^p\right)$ est bornée.

Il est facile de prouver que :

$$\forall p \in \mathbb{N}^*, \quad \nabla \left(V^* \circ f^p\right)(x) = {}^tDf(x)\, {}^tDf\left(f(x)\right) \cdots {}^tDf\left(f^{p-1}(x)\right) \nabla V^* \left(f^p(x)\right). \tag{3.3}$$

Comme $V \circ f \leq V$, il est clair que $\{V \leq K\}$ est stable par f et donc par f^k pour tout entier $k \in \mathbb{N}^*$. Par hypothèse, Df est bornée sur $\{V \leq K\}$ et, par construction, ∇V^* est bornée sur $\{V \leq K\}$ et nulle sur ${}^c\{V \leq K\}$. Donc, pour tout entier $p \in \mathbb{N}^*$, $\nabla \left(V^* \circ f^p\right)$ est bornée.

En différentiant l'expression (3.3), on obtient :

$$\nabla^2 \left(V^* \circ f^p\right) = A\left(Df, Df \circ f, \cdots, Df \circ f^{p-1}, D^2 f, D^2 f \circ f, \cdots, D^2 f \circ f^{p-1}, \nabla V^* \circ f^p\right) +$$
$$+ {}^tDf(x)\, {}^tDf\left(f(x)\right) \cdots {}^tDf\left(f^{p-1}(x)\right) \nabla^2 V^* \left(f^p(x)\right) Df\left(f^{p-1}(x)\right) Df\left(f^{p-2}(x)\right) \cdots Df(x),$$

où A est une fonction p-multilinéaire.

De même que précédemment, $D^2 f$ et $\nabla^2 V^*$ sont bornées sur $\{V \leq K\}$ et $\nabla^2 V^*$ est nulle sur ${}^c\{V \leq K\}$. Comme $\{V \leq K\}$ est stable par f, pour tout $p \in \mathbb{N}^*$, $\nabla^2 \left(V^* \circ f^p\right)$ est bornée.

Avant d'appliquer la proposition 3.1 à $V^* \circ f^p$, remarquons qu'en x^*, on a $\nabla V^*(x^*) = 0$ et $f(x^*) = x^*$. Par suite, pour tout $p \in \mathbb{N}^*$,

$$\nabla \left(V^* \circ f^p\right)(x^*) = 0$$

et

$$\begin{aligned} \nabla^2 \left(V^* \circ f^p\right)(x^*) &= {}^tDf(x^*)^p \nabla^2 V^*(x^*) Df(x^*)^p \\ &= \varphi_1'\left(V(x^*)\right)\, {}^tDf(x^*)^p \nabla^2 V(x^*) Df(x^*)^p. \end{aligned} \tag{3.4}$$

De plus, si $\widehat{x}$ est un point fixe de f n'appartenant pas à $\{V \leq K\}$, $\nabla^2 \left(V^* \circ f^p\right)(\widehat{x}) = 0$.

Soit alors une loi limite éventuelle ν^0 et une suite $\gamma_n \searrow 0$ telle que $\nu^{\gamma_n} \Rightarrow \nu^0$. Du fait que $\{f = Id\} \cap \{V \leq K\}$ est fini par hypothèse, La proposition 3.1 donne, compte tenu de (3.4), pour tout $p \in \mathbb{N}^*$:

$$\lim_{n\to+\infty} \frac{\nu^{\gamma_n}(V^* \circ f^p) - \nu^{\gamma_n}(V^* \circ f^{p+1})}{\gamma_n^2} =$$

$$\sum_{x_0\in\{f=Id\}\cap\{V\leq K\}} \frac{\nu^0(\{x_0\})}{2} \int_{\mathbb{R}^p} {}^t e(x_0, y)\nabla^2 (V^* \circ f^p)(x_0) e(x_0, y)\mu(dy). \tag{3.5}$$

On a $V^* \geq V^* \circ f$, donc pour tout $p \in \mathbb{N}^*$, $V^* \geq V^* \circ f^p$ ce qui entraîne qu'en tout point fixe x_0, $\nabla^2 V(x_0) - \nabla^2 (V^* \circ f^p)(x_0)$ est une matrice positive (éventuellement non définie). Les relations (3.4) et (3.5) permettent alors d'écrire que, pour tout $p \in \mathbb{N}^*$:

$$\lim_{n\to+\infty} \frac{\nu^{\gamma_n}(V^* \circ f^p) - \nu^{\gamma_n}(V^* \circ f^{p+1})}{\gamma_n^2} \leq \sum_{x_0\in\{f=Id\}\cap\{V\leq K\}\setminus\{x^*\}} \frac{\nu^0(\{x_0\})}{2} \int_{\mathbb{R}^p} {}^t e(x_0, y)\nabla^2 V^*(x_0) e(x_0, y)\mu(dy) +$$

$$+ \frac{\varphi_1'(V(x^*))\,\nu^0(\{x^*\})}{2} \int_{\mathbb{R}^p} {}^t(\nabla f(x^*)^p e(x^*, y))\, \nabla^2 V(x^*)\, (\nabla f(x^*)^p e(x^*, y))\, \mu(dy). \tag{3.6}$$

Deuxième étape : On veut maintenant montrer que :

$$\lim_{p\to+\infty} \int_{\mathbb{R}^p} {}^t(\nabla f(x^*)^p e(x^*, y))\, \nabla^2 V(x^*)\, (\nabla f(x^*)^p e(x^*, y))\, \mu(dy) = -\infty. \tag{3.7}$$

Comme $V \circ f \leq V$, $\nabla^2 V(x^*) - {}^t Df(x^*)\nabla^2 V(x^*) Df(x^*)$ est donc positive. Or, par hypothèse, elle est non dégénérée. $\nabla^2 V(x^*) - {}^t Df(x^*)\nabla^2 V(x^*) Df(x^*)$ est donc définie positive. Par suite, il existe une constante $c > 0$, telle que :

$$\nabla^2 V(x^*) - {}^t Df(x^*)\nabla^2 V(x^*) Df(x^*) \geq c\, I,$$

ce que l'on peut écrire :

$$\forall y \in \mathbb{R}^d, \quad {}^t(Df(x^*)e(x^*, y))\, \nabla^2 V(x^*)\, (Df(x^*)e(x^*, y)) \leq {}^t e(x^*, y)\nabla^2 V(x^*) e(x^*, y) - c|e(x^*, y)|^2.$$

On en déduit que, pour tout $k \in \mathbb{N}^*$:

$$\forall y \in \mathbb{R}^d, \quad {}^t\left(\nabla f(x^*)^k e(x^*, y)\right) \nabla^2 V(x^*) \left(\nabla f(x^*)^k e(x^*, y)\right) \leq$$

$$\leq {}^t\left(\nabla f(x^*)^{k-1} e(x^*, y)\right) \nabla^2 V(x^*) \left(\nabla f(x^*)^{k-1}\right) e(x^*, y) - c\left|\nabla f(x^*)^{k-1} e(x^*, y)\right|^2.$$

En sommant ces inégalités pour k variant de 1 à p, on trouve :

$$\forall y \in \mathbb{R}^d, \quad {}^t(\nabla f(x^*)^p e(x^*, y))\, \nabla^2 V(x^*)\, (\nabla f(x^*)^p e(x^*, y)) \leq {}^t e(x^*, y)\nabla^2 V(x^*) e(x^*, y) - c\sum_{k=0}^{p-1} \left|\nabla f(x^*)^k e(x^*, y)\right|^2. \tag{3.8}$$

Pour tout $k \in \mathbb{N}$, on a :

$$\forall y \in \mathbb{R}^d, \quad \left|\Pi^* \nabla f(x^*)^k e(x^*, y)\right| \leq \|\Pi^*\| \left|\nabla f(x^*)^k e(x^*, y)\right|.$$

Les deux espaces $E^*_{\geq 1}$ et $E^*_{<1}$ sont stables par la projection Π^* sur $E^*_{\geq 1}$ parallèlement à $E^*_{<1}$, donc Π^* et $Df(x^*)$ commutent. D'où, pour tout $k \in \mathbb{N}$:

$$\forall y \in \mathbb{R}^d, \quad \left|\nabla f(x^*)^k e(x^*,y)\right| \geq \frac{1}{\|\Pi^*\|}\left|\Pi^* \nabla f(x^*)^k e(x^*,y)\right| = \frac{1}{\|\Pi^*\|}\left|\nabla f(x^*)^k \Pi^* e(x^*,y)\right|,$$

ce qui réinjecté dans (3.8) donne, pour tout $y \in \mathbb{R}^d$:

$$\begin{aligned} {}^t\left(\nabla f(x^*)^p e(x^*,y)\right) \nabla^2 V(x^*) \left(\nabla f(x^*)^p e(x^*,y)\right) &\leq {}^t e(x^*,y) y \nabla^2 V(x^*) e(x^*,y) - \\ &- \frac{c}{\|\Pi^*\|^2} \sum_{k=0}^{p-1} \left|\nabla f(x^*)^k \Pi^* e(x^*,y)\right|^2. \end{aligned} \tag{3.9}$$

Il reste alors à montrer que :

$$\forall y \in \mathbb{R}^d \text{ tel que } \Pi^* e(x^*,y) \neq 0, \quad \lim_{p \to +\infty} \sum_{k=0}^{p-1} \left|\nabla f(x^*)^k \Pi^* e(x^*,y)\right|^2 = +\infty. \tag{3.10}$$

Ce point est un corollaire simple du lemme suivant :

Lemme 3.1 *Soit F un espace vectoriel euclidien et $|\,.\,|$ sa norme euclidienne. Soit $A \in \mathcal{M}(F)$ telle que $\rho(A^{-1}) \leq 1$ (i.e. toutes les valeurs propres de A sont de modules supérieurs ou égal à 1).*
Alors $G := \left\{ v \in F / \quad \lim_{p \to +\infty} A^p v = 0 \right\} = \{0\}$.

Démonstration du lemme 3.1 : Supposons que $G \neq \{0\}$. Il est facile de voir que G est un sous-espace vectoriel de F. Soit q sa dimension et $(v_1, \cdots, v_q)$ une base de G telle que $\forall\, 1 \leq i \leq q,\ |v_i| = 1$.

Il est clair que G est stable par A. On peut donc définir $B := A_{|G}$ la restriction de A à G. On définit également la norme $|\,.\,|_1$ sur G. pour tout $v := \sum_{i=1}^{q} \lambda_i v_i \in G$ par $|v|_1 := \sum_{i=1}^{q} |\lambda_i|$.

Pour v non nul dans G, on a donc :

$$|B^p v| \leq \sum_{i=1}^{q} |\lambda_i| |B^p v_i| \leq |v|_1 \sup_{1 \leq i \leq q} |B^p v_i|. \tag{3.11}$$

G étant de dimension finie, toutes les normes sur G sont équivalentes. Il existe donc une constante $\alpha > 0$ telle que :

$$\forall v \in G, \quad |v|_1 \leq \alpha |v|,$$

ce qui, reporté dans (3.11) donne :

$$\forall v \in F, \quad |B^p v| \leq \alpha |v| \sup_{1 \leq i \leq q} |B^p v_i|. \tag{3.12}$$

Comme, par définition de G :

$$\lim_{p \to +\infty} \sup_{1 \leq i \leq q} |B^p v_i| = 0,$$

on déduit de (3.12) que :

$$\lim_{p \to +\infty} \|B^p\| = 0, \tag{3.13}$$

ce qui est contradictoire puisque $\|B^p\| \geq \rho(B^p) \geq 1$. □

En considérant $A := \nabla f(x^*)_{|E^*_{\geq 1}}$ la restriction de $\nabla f(x^*)$ au sous-espace $E^*_{\geq 1}$, on est sous les conditions d'application du lemme 3.1.

On prouve ainsi que, pour tout y dans $\mathbb{R}^d$ tel que $\Pi^* e(x^*, y) \neq 0$, $u_k := \left|\nabla f(x^*)^k \Pi^* e(x^*, y)\right|^2$ ne tend pas vers 0 quand k tend vers $+\infty$. La série de terme général u_k est donc divergente et à termes positifs ce qui achève la démonstration de la propriété (3.10).

Partant, pour tout $y \in \mathbb{R}^d$ tel que $\Pi^* e(x^*, y) \neq 0$, on a :

$$\lim_{p \to +\infty} {}^t\left(\nabla f(x^*)^p e(x^*, y)\right) \nabla^2 V(x^*) \left(\nabla f(x^*)^p e(x^*, y)\right) = -\infty. \tag{3.14}$$

Comme, par hypothèse, $\int_{\mathbb{R}^p} |\Pi^* e(x^*, y)| \mu(dy) > 0$, on tire de (3.14) par Beppo-Levi :

$$\lim_{p \to +\infty} \int_{\mathbb{R}^p} {}^t\left(\nabla f(x^*)^p e(x^*, y)\right) \nabla^2 V(x^*) \left(\nabla f(x^*)^p e(x^*, y)\right) \mu(dy) = -\infty,$$

qui est le résultat escompté.

Troisième étape :

Pour finir, on a forcément $\nu^0(\{x^*\}) = 0$, sinon ce qui précède prouverait que le deuxième membre de la relation (3.6) tend vers $-\infty$ quand p tend vers $+\infty$, alors que le premier membre de la relation (3.6) est positif, puisque, pour tout $p \geq 0$, $V^* \circ f^{p+1} \leq V^* \circ f^p$. □

Remarques : • Si x^* est un maximum local strict de V, point fixe de f, on a encore $\nabla V(x^*) = 0$ et $\rho(Df(x^*)) \geq 1$. Le théorème 3.3 s'applique donc aussi aux maxima de V. Cependant, ses hypothèses sont plus contraignantes que celles du théorème 3.1.

D'une part, f est supposée de classe C^2 et des contraintes sur les deux premières différentielles de f sont requises et d'autre part, si x^* est un maximum local strict de V, point fixe de f pour lequel $\nabla^2 V(x^*) - {}^t Df(x^*) \nabla^2 V(x^*) Df(x^*)$ est définie positive, comme $\nabla^2 V(x^*)$ et ${}^t Df(x^*) \nabla^2 V(x^*) Df(x^*)$ sont négatives (car x^* est un maximum de V) ${}^t Df(x^*) \nabla^2 V(x^*) Df(x^*)$ est alors forcément définie négative. Par suite, $V \circ f$ vérifie les hypothèses du théorème 3.1 pour x^*.

A l'inverse, il est possible, a priori, que $\nabla^2 V(x^*) - {}^t Df(x^*) \nabla^2 V(x^*) Df(x^*)$ soit dégénérée sans que $\nabla^2 V(x^*)$ ne le soit. Ainsi, on vérifiera aisément en dimension 2, que c'est le cas si x^* est un maximum de V tel que :

$$Df(x^*) := \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{et} \quad \nabla^2 V(x^*) := -\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix},$$

ou bien si x^* est un point selle de V tel que :

$$Df(x^*) := \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{pmatrix} \quad \text{et} \quad \nabla^2 V(x^*) := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

• Il ne semble pas possible détendre le théorème 3.3 au cadre quasi-fellérien puisque l'on a besoin d'avoir f de classe C^2.

Retour sur la Lemniscate : Nous avons vu, à la section précédente, que, pour tout $u > 0$, $\nu^{u,0} = \delta_{(0,0)}$. Or $(0,0)$ est un point selle de V. Cependant $\nabla^2 V(0,0)$ étant la matrice nulle,

$\nabla^2 V(0,0) -^t Df(0,0)\nabla^2 V(0,0)Df(0,0)$ est dégénérée, et même nulle. Il est donc possible de charger un point selle si celui-ci ne vérifie pas les hypothèses du théorème 3.3.

Dans le cadre des systèmes dynamiques perturbés par un bruit décroissant vers 0, on montrerait (Brandière, 1998) que l'ensemble des trajectoires convergeant presque sûrement vers $(0,0)$ est de probabilité nulle. Cependant, on peut montrer que l'ensemble des trajectoires revenant une infinité de fois au voisinage de $(0,0)$ est de probabilité 1, ce qui explique que, dans le cadre qui nous intéresse, la limite faible des lois des systèmes dynamiques perturbés par un bruit non évanescent soit la probabilité de Dirac en $(0,0)$.

Bibliographie de la troisième partie

Benaïm, M. (1998) Recursive algorithms, urn process and chaining number of chain recurrent sets, Ergodic Theory and Dynamical systems.

Benvéniste, A., Métivier, M., Priouret, P. (1987) *Algorithmes adaptatifs et approximations stochastiques*, Masson.

Brandière, O. (1996) Autour des pièges des algorithmes stochastiques, thèse de l'université Marne-la-Vallée.

Brandière, O. (1998) The dynamic system method and the traps, Advances in Applied Pobability, 30, pp.137-151.

Cohort, P. (1998) Unicité d'un quantificateur localement optimal par le théorème du col, *Prépublication du laboratoire de Probabilités de l'université Paris 6.*

Collet, P., Eckmann, J.P. (1980), *Iterated Maps of the interval as Dynamical systems*, Birkhaüser, Boston.

Devaney, R. (1989) *An introduction to chaotic dynamical systems*, Addison-Wesley Publishing Compagny, Redwood city, 336p.

Fort, J.C., Pagès, G. (1998) Asymptotic behavior of a Markovian stochastic algorithm with constant step, *SIAM Journal of Control and Optimization.*

Kesten , H. (1972) Limit theorems for stochastic growth models I and II, *Adv. Appl. Prob.*, 4, pp.193-232, pp.393-428.

Kieffer, J.C. (1982) Exponential Rate of Convergence for Lloyd's Method I, *IEEE Transactions on Information Theory*, Vol. IT-28, 2.

Kifer, Y. (1988) *Random Perturbations of Dynamical Systems*, *Progress in Probability and Statistics*, Birkhaüser, 294p.

Krengel, U. (1989) *Ergodic Theorems*, De Gruyter Studies in Mathematics, Berlin-New York, 357p.

Lamberton, D., Pagès, G. (1996) On the critical points of the 1-dimensional Competitive Learning Vector Quantization algorithm, Proceedings of ESANN'96, M. Verleysen ed., De Facto, Bruxelles, pp.97-102.

Lazarev, V.A. (1992) Convergence of Stochastic Approximation Procedures in case of Regression Equation with several Roots, *Translated from Problemy Peredachi informatsii*, 28, pp.75-88.

Nevel'son, M.B., Has'minskii, R.Z. (1972) Stochastic approximation and recursive estimations, *Nauka Moscou 1972 translation of math. monographs*, 17, American Mathematical Society.

Pagès, G. (1993) Voronoï tessellation, space quantization algorithms and numerical integration, *Actes de l'ESANN*, M. Verleysen Ed., De Facto, Bruxelles, pp.221-228.

Pagès, G. (1997) A space quantization algorithm for numerical integration, *J. of computational of Appl. Math.*, 88, pp.221-228.

Pemantle, R. (1990) Nonconvergence to unstable points in urn models and stochastic approximations, *Annals of Prob.*, 18, pp.698-712.

Pierre Loti Viaud, D. (1995) Random perturbations of recursive sequences with an application to an epidemic model, J. Appl. Prob., 32.

Remerciements

Je dédie cette thèse à mes parents. A ma mère qui a toujours su me soutenir et m'encourager dans les moments difficiles. Elle m'a communiqué sa joie de vivre et sa volonté de lutter dans l'adversité. A mon père qui m'a épaulé tout au long de ma scolarité. Il m'a transmis son intérêt pour les mathématiques et pour bien d'autres domaines encore. Sans eux, cette thèse n'aurait certainement jamais vu le jour, moi non plus d'ailleurs. Pour ces raisons et pour bien d'autres encore, je leur serai éternellement reconnaissant.

Avant tout, je veux remercier très sincèrement mon directeur de thèse Gilles Pagès qui a su se montrer toujours très disponible. J'ai eu beaucoup de plaisir à travailler sous sa direction et j'espère que les liens que nous avons tissés, tant au niveau professionnel que personnel, perdureront.

Je remercie tout particulièrement Laure Elie de me faire l'honneur de présider ce jury.

Je remercie également Marie Duflo et Philippe Bougerol d'avoir rapporté sur ce travail.

Je tiens à exprimer ma profonde gratitude à Xavier Guyon, que j'ai eu le plaisir et l'honneur de cotoyer durant des années et qui a accepté de participer à ce jury.

Que Michel Benaïm soit également remercié pour sa participation à ce jury.

Enfin, j'adresse tous mes remerciements à Marie Cottrell pour m'avoir si chaleureusement accueilli au sein du SAMOS, ainsi qu'à toute l'équipe avec laquelle j'ai entretenu, et j'entretiens toujours, d'excellents rapports.