

SUR QUELQUES ALGORITHMES RÉCURSIFS POUR LES PROBABILITÉS NUMÉRIQUES *

GILLES PAGÈS¹

Abstract. The aim of this paper is to take an in-depth look at the long time behaviour of some continuous time Markovian dynamical systems and at its numerical analysis. We first propose a short overview of the main ergodicity properties of time continuous homogeneous Markov processes (stability, positive recurrence). The basic tool is a Lyapunov function. Then, we investigate if these properties still hold for the time discretization of these processes, either with constant or decreasing step (ODE method in stochastic approximation, Euler scheme for diffusions). We point out several advantages of the weighted empirical random measures associated to these procedures, especially with decreasing step, in terms of convergence and of rate of convergence. Several simulations illustrate these results.

AMS Subject Classification. 65C30, 62L20.

Reçu le 26 février 2001. Révisé le 17 septembre 2001.

1. INTRODUCTION

Le but de ce mémoire est de faire le point sur les méthodes permettant *l'analyse numérique du comportement en temps long* des algorithmes stochastiques et des diffusions. L'essentiel de cette introduction est consacré à des rappels et à des commentaires sur ces processus dans le cadre des probabilités numériques. Il nous est paru ensuite utile de brosser une rapide synthèse des méthodes permettant l'étude des propriétés d'ergodicité (réurrence positive, stabilité) des processus de Markov à temps continu en présence d'une fonction de Lyapunov. C'est l'objet des sections 2 et 3. Nous avons volontairement privilégié les diffusions et les équations différentielles ordinaires (ces dernières étant l'objet à temps continu sous-jacent aux algorithmes stochastiques). Has'minskii a joué dans ce domaine un rôle majeur et son ouvrage [25] reste une référence essentielle (voir aussi, plus récemment [10]). Les différentes approches numériques sont abordées à la section 4, accompagnées de quelques premiers éléments de comparaison.

1.1. Algorithmes stochastiques (markoviens) sur \mathbb{R}^d

On considère $(X_n)_{n \geq 0}$ une suite de vecteurs aléatoires de \mathbb{R}^d définie récursivement par

$$\forall n \in \mathbb{N}, \quad X_{n+1} = X_n + \gamma_{n+1} H(X_n, \varepsilon_{n+1}). \quad (1)$$

Mots-clés et phrases : Ergodicity, stability, Markov process, diffusion, stochastic algorithm, ODE method, Euler scheme, empirical measure.

* *Journées MAS 2000, 8–11 septembre 2000, Rennes.*

¹ Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, Université Pierre et Marie Curie, Université Paris 6, Case 188, 4 place Jussieu, 75252 Paris Cedex 05, France ; e-mail: gpa@ccr.jussieu.fr

- La suite $(\gamma_n)_{n \geq 1}$, appelée *pas* de l'algorithme stochastique, vérifie $\gamma_n > 0$, $\sum_{n \geq 1} \gamma_n = +\infty$;
- la suite $(\varepsilon_n)_{n \geq 1}$ est une suite de variables aléatoires i.i.d., de loi μ , indépendante de la variable initiale X_0 ;
- la fonction H est l'*observation* de l'algorithme stochastique. Elle est définie par

$$H : \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}^d \\ (x, u) \mapsto H(x, u)$$

et vérifie, pour tout $x \in \mathbb{R}^d$, $H(x, \cdot) \in \mathcal{L}^1(\mu)$ et $\mu(\mathrm{d}u)$ -p.s., $H(\cdot, u)$ est continue en x .

Le terme observation indique ici que l'utilisateur est capable, disposant de X_n de "fabriquer" $H(X_n, \varepsilon_{n+1})$, soit expérimentalement *via* un dispositif physique, soit par simulation. Dans ce dernier cas, cela signifie que la fonction H est numériquement calculable et que la loi μ des ε_k est *simulable* (à un coût raisonnable s'entend). On appelle *fonction moyenne* de l'algorithme la fonction h définie par

$$h(x) := \mathbb{E}(H(x, \varepsilon_1)) = \int_{\mathbb{R}^q} H(x, u) \mu(\mathrm{d}u).$$

La fonction h , elle, n'est généralement pas calculable numériquement (toujours à un coût raisonnable). En effet, lorsqu'elle l'est, il n'y a généralement pas de raison de mettre en œuvre l'algorithme (1) puisqu'il peut alors être remplacé – souvent avantageusement – par son pendant déterministe

$$x_{n+1} = x_n + \gamma_{n+1} h(x_n). \quad (2)$$

Ainsi on a typiquement recours à un algorithme stochastique de type (1) lorsque le calcul numérique de H et la simulation d'échantillons de la loi μ ont un coût faible comparé à l'intégration numérique d'une fonction par rapport à μ , sachant que, précisément, l'intégration numérique a un coût prohibitif sur \mathbb{R}^d lorsque d augmente. D'autres considérations peuvent motiver le choix d'une procédure récursive stochastique, notamment l'usage du "bruit" pour éviter des convergences vers des points "indésirables". Mais, à nouveau, si h est numériquement évaluable, on privilégie alors une perturbation linéaire de (2) par un bruit convenablement choisi. De telles procédures ne se ramènent pas toujours à (1) (recuit simulé).

Revenons à l'algorithme (1) dans sa forme originelle. On peut formellement le réécrire sous la forme

$$\forall n \in \mathbb{N}, \quad X_{n+1} = X_n + \gamma_{n+1} h(X_n) + \gamma_{n+1} (H(X_n, \varepsilon_{n+1}) - h(X_n)).$$

Les termes $H(X_n, \varepsilon_{n+1}) - h(X_n)$, $n \geq 0$, ont une espérance conditionnelle nulle par rapport à $\mathcal{F}_n^\varepsilon := \sigma(X_0, \varepsilon_1, \dots, \varepsilon_n)$, filtration propre de ε (et X_0), notée \mathcal{F}_n dans la suite. Ils sont par ailleurs \mathcal{F}_{n+1} -mesurables. Ce sont donc les accroissements d'une \mathcal{F}_n -martingale $(M_n)_{n \geq 1}$. Finalement, l'algorithme stochastique (1) se réécrit

$$X_{n+1} = X_n + \gamma_{n+1} h(X_n) + \gamma_{n+1} \Delta M_{n+1}. \quad (3)$$

Les outils sur lesquels repose l'étude de cet algorithme stochastique diffèrent selon que le pas $(\gamma_n)_{n \geq 1}$ est constant ou tend vers 0.

- $\boxed{\gamma_n = \gamma > 0}$:
– L'algorithme stochastique (1) est alors une \mathcal{F}_n -chaîne de Markov homogène de semi-groupe de transition fellérien donné par

$$Q_\gamma(f)(x) := \int_{\mathbb{R}^q} f(x + \gamma H(x, u)) \mu(\mathrm{d}u).$$

- On peut également le voir, *via* (3), comme le schéma d’Euler de pas γ , *perturbé* par $\gamma\Delta M_{n+1}$, de l’Équation Différentielle Ordinaire associée à h (en abrégé EDO_h), à savoir

$$EDO_h \equiv \dot{x} = h(x).$$

- $\boxed{\gamma_n \rightarrow 0}$:

- On a alors à faire à une \mathcal{F}_n -chaîne de Markov inhomogène, dont la transition à la date n est donnée par

$$\mathbb{E}(f(X_{n+1})/\mathcal{F}_n) = Q_{\gamma_{n+1}}f(X_n).$$

- L’équation (3) est alors celle du schéma d’Euler à *pas “décroissant”* (et perturbé) de l’ EDO_h .

Le fait d’assimiler l’équation (3) à un schéma d’Euler perturbé d’ EDO_h est connu sous le nom de *méthode(s) de l’EDO*. On peut alors interpréter une procédure stochastique récursive comme un moyen d’analyser le comportement en temps long de l’ $EDO_h \equiv \dot{x} = h(x)$ lorsque la fonction h – supposée continue – n’est pas numériquement connue. Sur un plan mathématique, ces différentes méthodes de l’ EDO consistent à déterminer dans quelle mesure ceci est possible et jusqu’à quel point.

Précisons à ce stade qu’il n’est pas dans nos objectifs de proposer un exposé général sur les méthodes d’approximation stochastique, notamment concernant le comportement trajectorien *p.s.* de ces algorithmes (convergence, récurrence par chaîne, vitesse en loi, etc.) qui ne sera évoqué ici qu’à titre de comparaison, de façon tout à fait anecdotique. Pour se familiariser (en français) avec les principaux aspects mathématiques de l’approximation stochastique, on se reportera par exemple à des traités généraux comme [9] et [17] ou à des synthèses plus spécialisées comme [7] (notamment pour les liens avec les systèmes dynamiques).

1.2. Diffusions (browniennes) sur \mathbb{R}^d

On considère le processus $(X_t)_{t \geq 0}$ défini par l’Équation Différentielle Stochastique ($EDS_{b,\sigma}$) suivante

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

où

- $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est une fonction continue ;
- $\sigma : \mathbb{R}^d \rightarrow \mathcal{M}(d \times q)$ est également continue, b et σ vérifiant $|b(x)| + \|\sigma(x)\| = O(|x|)$;
- $(W_t)_{t \geq 0}$ est un mouvement brownien standard sur \mathbb{R}^q .

On s’intéresse ici au problème de *probabilités numériques* suivant : les fonctions b et σ étant connues, comment déterminer numériquement la mesure invariante de cette diffusion, autrement dit déterminer son comportement en temps long.

Pour ce faire on discrétise le problème, en utilisant par exemple le schéma d’Euler de pas γ constant :

$$X_{n+1}^\gamma = X_n^\gamma + \gamma b(X_n^\gamma) + \sqrt{\gamma} \sigma(X_n^\gamma)U_{n+1}, \tag{4}$$

où $(U_n)_{n \geq 1}$ est une suite de variables aléatoires i.i.d., centrées, normalisées, de loi μ vérifiant $\int_{\mathbb{R}^q} |u|^2 \mu(du) < +\infty$. Il s’ensuit que $(X_n^\gamma)_{n \geq 0}$ est une chaîne de Markov homogène, de transition fellérienne donnée par

$$Q_\gamma(f)(x) = \int_{\mathbb{R}^q} f(x + \gamma b(x) + \sqrt{\gamma} \sigma(x)u) \mu(du).$$

Constatant que les algorithmes stochastiques comme les diffusions font intervenir des processus de Markov homogènes, nous allons commencer par quelques rappels, notamment concernant l’existence d’un éventuel régime stationnaire et les liens de ce régime avec le générateur infinitésimal du processus.

2. PROCESSUS DE MARKOV HOMOGENÈS

Un processus $(X_t)_{t \geq 0}$ est un processus de Markov homogène si

$$\forall s, t \geq 0, \quad \mathbb{E}(f(X_{t+s})/\mathcal{F}_s) = \mathbb{E}(f(X_{t+s})/X_s) = P_t f(X_s)$$

où $P_t f(x) := \mathbb{E}_x(f(X_t))$ vérifie

$$P_t \mathbf{1} = \mathbf{1} \quad \text{et} \quad P_t \circ P_s = P_{t+s}$$

($(P_t)_{t \geq 0}$ est le semi-groupe de transition du processus de Markov). Nous allons donner quelques exemples de processus de Markov homogènes.

1. La solution de l'EDO $\dot{x} = h(x)$, avec h fonction localement lipschitzienne vérifiant $|h(x)| = O(|x|)$, est un processus de Markov homogène (bien évidemment déterministe !), dont le semi-groupe de transition est donné par

$$P_t f(x) = f(\Phi_t(x)) \quad \text{où} \quad (\Phi_t(x))_{t \in \mathbb{R}_+} \quad \text{désigne le flot de l'EDO.}$$

2. On considère l'EDO $\dot{x} = 2\sqrt{x}$ sur \mathbb{R}_+ . Dans ce cas le flot n'existe pas, mais on peut trouver des solutions qui forment un processus de Markov homogène, à savoir si X^{x_0} désigne une solution issue de x_0 :

$$X_t^{x_0} = \begin{cases} (\sqrt{x_0} + t)^2 & \text{si } x_0 \neq 0 \\ ((t - T_0)_+)^2 & \text{si } x_0 = 0, \text{ où } T_0 \text{ suit une loi exponentielle de paramètre 1.} \end{cases}$$

Le semi-groupe de transition de ce processus de Markov s'écrit

$$P_t f(x) = f(x + 2x\sqrt{t} + t^2) \mathbf{1}_{\{x > 0\}} + \left(f(0)e^{-t} + \int_0^t f((t - \theta)^2) e^{-\theta} d\theta \right) \mathbf{1}_{\{x = 0\}}.$$

3. Toute solution de l'EDS

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t,$$

où b et σ sont toutes deux localement lipschitziennes et vérifient de plus $(|b| + \|\sigma\|)(x) = O(|x|)$, est un processus de Markov homogène (cf. [19, 40]).

Les semi-groupes apparaissant dans les trois exemples ci-dessus sont tous trois *fellériens*, à savoir

- (a) $\forall t > 0, \quad P_t(\mathcal{C}_0) \subset \mathcal{C}_0$, où $\mathcal{C}_0 := \mathcal{C}_0(\mathbb{R}^d, \mathbb{R})$ (fonctions continues de \mathbb{R}^d dans \mathbb{R} , qui tendent vers 0 à l'infini) ;
- (b) $\forall f \in \mathcal{C}_0, \quad P_t f \rightarrow f$ quand $t \rightarrow 0$.

Par extension un processus de Markov dont le semi-groupe de transition est fellérien sera lui-même appelé fellérien. Dans la suite, les processus de Markov considérés seront, sauf mention explicite, tous supposés fellériens. On montre classiquement qu'un tel processus admet *une version continue à droite, limitée à gauche* ("càdlàg", cf. [40], Th. 2.7, p. 86).

2.1. Régime stationnaire

Lorsque l'on simule un processus de Markov homogène (ou plutôt une chaîne de Markov), on constate généralement que, soit le processus atteint systématiquement des valeurs numériques de plus en plus grandes en valeur absolue au fil du temps, soit il tend à se stabiliser, sans pour autant converger en général. Dans le second cas de figure, si l'on simule un grand nombre de trajectoires du processus, on constate en traçant les histogrammes de fréquence à plusieurs instants successifs que la loi de X_t ne dépend plus de t lorsque t devient grand : le processus tend vers un régime stationnaire.

Nous allons commencer par préciser dans le cadre markovien homogène cette notion de régime stationnaire. Dans un second temps nous allons explorer les situations dans lesquelles il est envisageable d'identifier ce régime stationnaire par l'observation suffisamment prolongée d'une seule trajectoire (propriété d'ergodicité). Les méthodes numériques pour y parvenir seront développées dans le cas des diffusions et des algorithmes stochastiques à la section 4.

On considère un processus de Markov homogène fellérien $(X_t)_{t \geq 0}$, et l'on appelle μ la loi de X_0 , *i.e.* $X_0 \stackrel{\text{Loi}}{\sim} \mu$. On a alors $X_t \stackrel{\text{Loi}}{\sim} \mu P_t$, $\forall t \geq 0$. Plus généralement la loi initiale μ et le semigroupe de transition $(P_t(x, dy))_{t \geq 0}$ caractérisent entièrement la loi d'un processus de Markov homogène. (*cf.*, *e.g.* [40], Th. 1.5, p. 77). On note \mathbb{P}_μ la loi du processus de loi initiale μ (et de transition $P_t(x, dy)$). Lorsque $\mu = \delta_x$, on pose $\mathbb{P}_{\delta_x} = \mathbb{P}_x$. On vérifie ainsi que $P^t(f)(x) := \mathbb{E}_x(f(X_t))$ (dès que ces expressions ont un sens).

Définition 1. On appelle mesure invariante pour le processus X toute probabilité ν sur \mathbb{R}^d vérifiant

$$\forall t \geq 0, \quad \nu P_t = \nu.$$

Si ν est invariante, dès que $X_0 \stackrel{\text{Loi}}{\sim} \nu$, pour tout $t \geq 0$, $X_t \stackrel{\text{Loi}}{\sim} \nu$. On vérifie dans ce cas que, sous \mathbb{P}_ν , le processus $(X_t)_{t \geq 0}$ est un processus *stationnaire*.

On définit alors l'ensemble

$$\mathcal{I}_P := \{ \nu \text{ probabilité sur } \mathbb{R}^d, \quad \nu P_t = \nu \quad \forall t \geq 0 \}.$$

L'ensemble \mathcal{I}_P est un *convexe, fermé pour la topologie de la convergence étroite* via la propriété de Feller, éventuellement vide. Si \mathcal{I}_P est non vide et compact (toujours pour la topologie de la convergence étroite), il existe d'après le théorème de Krein–Millman au moins une probabilité ν^* *extrémale* dans \mathcal{I}_P (*cf.* [18]). Par extrémale, on entend qu'il est impossible d'écrire ν^* comme combinaison convexe de deux probabilités de \mathcal{I}_P distinctes de ν^* . On montre alors que la probabilité ν^* est nécessairement *ergodique* (abus de langage pour signifier l'ergodicité du processus de Markov lui-même). Soyons plus précis sur ce que l'on entend par là. Tout d'abord, on se place sur l'espace canonique des trajectoires (càdlàg) du processus, soit $\Omega := \mathbb{D}(\mathbb{R}_+, \mathbb{R}^d) := \{ \omega : \mathbb{R}_+ \rightarrow \mathbb{R}^d, \text{ càdlàg} \}$ et l'on note X le processus canonique défini sur Ω par $X_t(\omega) := \omega(t)$, $\omega \in \Omega$. Soit $(\mathcal{F}_t^X)_{t \geq 0}$ la filtration naturelle de X (rendue continue à droite) et $\mathcal{F}_\infty^X := \sigma(X_t, t \in \mathbb{R}_+)$. L'espace Ω est muni canoniquement de l'opérateur de décalage défini par $(\theta_t(X))_s = X_{t+s}$. Sous la probabilité \mathbb{P}_{ν^*} définie sur $(\Omega, \mathcal{F}_\infty^X)$, on vérifie aisément que le processus canonique X est stationnaire : en d'autres termes $\theta_t^{-1}(\mathbb{P}_{\nu^*}) = \mathbb{P}_{\nu^*}$ pour tout $t \in \mathbb{R}_+$. Enfin, le caractère extrémal de ν^* entraîne l'ergodicité de X sur $(\Omega, \mathcal{F}_\infty^X, \mathbb{P}_{\nu^*})$: pour tout élément $A \in \mathcal{F}_\infty^X$ invariant par l'opérateur de décalage $(\theta_t)_{t \in \mathbb{R}_+}$, $\mathbb{P}_{\nu^*}(A) = 0$ ou 1 (voir l'annexe A pour une démonstration, ou encore [43], Th.7.4.8, Chap. 7, p. 420).

Le *théorème ergodique ponctuel de Birkhoff* (*cf.*, *e.g.*, Ref. [29]) permet alors de relier le comportement en temps long du processus de Markov à cette mesure invariante extrémale : en effet, $(X_t)_{t \geq 0}$ étant ergodique sous \mathbb{P}_{ν^*} , pour toute fonctionnelle $F \in L^1(\mathbb{P}_{\nu^*})$,

$$\mathbb{P}_{\nu^*}\text{-p.s.} \quad \frac{1}{t} \int_0^t F(X_{s+}) ds \rightarrow \mathbb{E}_{\mathbb{P}_{\nu^*}}(F).$$

Si l'on se restreint à des fonctionnelles élémentaires $F(X) := f(X_0)$, la convergence ci-dessus s'écrit

$$\forall f \in L^1(\nu^*) \quad \nu^*(dx)\text{-p.s.} \quad \mathbb{P}_x\text{-p.s.} \quad \frac{1}{t} \int_0^t f(X_s^x) ds \rightarrow \int f d\nu^*.$$

Exemple. On considère une $EDO_h \dot{x} = h(x)$, où h est une fonction vérifiant $\{h = 0\} = \{0\}$. On fait l'hypothèse que ODE_h admet un flot $(\Phi_t(x))_{(t,x) \in \mathbb{R} \times \mathbb{R}^d}$ vérifiant $\lim_{t \rightarrow +\infty} \Phi_t(x) = 0$ (le flot converge vers son unique équilibre). En d'autres termes 0 est un attracteur d' ODE_h . Il est alors immédiat de vérifier que $\mathcal{I}_P = \{\delta_0\}$, et donc, au

vu de ce qui précède, la solution de l'EDO_h est ergodique sous \mathbb{P}_0 . Le théorème de Birkhoff conduit dans ce contexte au résultat suivant : pour toute fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$,

$$\frac{1}{t} \int_0^t f(0) ds \rightarrow f(0) \quad \text{quand } t \rightarrow +\infty \dots$$

L'information fournie par une telle convergence est évidemment nulle.

Le théorème ergodique ponctuel seul est donc insuffisant pour rendre compte du fait que le processus tend vers son régime stationnaire.

Dans le but de caractériser de façon pratique les mesures invariantes d'un processus de Markov, on va introduire la notion de générateur infinitésimal d'un processus de Markov.

2.2. Générateur infinitésimal et stationnarité

Soit $(X_t)_{t \geq 0}$ un processus de Markov homogène, de semi-groupe de transition $(P_t)_{t \geq 0}$ fellérien. On appelle *générateur infinitésimal* de $(P_t)_{t \geq 0}$ un couple $(A, \mathcal{D}(A))$ où $\mathcal{D}(A) \subset \mathcal{C}_0$ et A désigne un opérateur $A : \mathcal{D}(A) \rightarrow \mathcal{C}_0$ vérifiant

$$\forall f \in \mathcal{D}(A), \quad \forall t \geq 0 \quad P_t f(x) = f(x) + \int_0^t P_s(Af)(x) ds. \tag{5}$$

En particulier $\frac{d}{dt} P_t(f)(x)|_{t=0} := \lim_{t \rightarrow 0} \frac{P_t(f)(x) - f(x)}{t} = Af(x)$. Par abus de langage on appelle couramment A seul le générateur infinitésimal.

Exemples.

- Dans le cas d'un processus de Markov défini par une $EDS_{b,\sigma}$, on pose $\mathcal{D}(A) := \mathcal{C}_K^2$. La formule d'Itô entraîne immédiatement que le générateur est donné par la formule suivante

$$Af(x) = (\nabla f|b)(x) + \frac{1}{2} \text{Tr}(\sigma^* D^2 f(x) \sigma).$$

- Dans le cas d'une EDO_h , on pose naturellement $\mathcal{D}(A) := \mathcal{C}_K^1$ et $Af(x) = (\nabla f|h)(x)$.

Le lien entre mesure invariante et générateur infinitésimal est le suivant

$$(\nu \in \mathcal{I}_P) \iff (\forall f \in \mathcal{D}(A) \quad \nu(Af) = 0). \tag{6}$$

Esquissons une preuve (formelle mais incomplète) de cette caractérisation :

- \implies : On réécrit la définition (5) du générateur infinitésimal à laquelle on applique le théorème de Fubini.
- \impliedby : Partant de (5), il vient que $t \mapsto P_t(f)(x)$ est dérivable et

$$\begin{aligned} \frac{dP_s}{ds} \Big|_{s=t} f(x) &= P_t(Af)(x) \\ \text{donc, formellement, } A(P_t f)(x) &= P_t(Af)(x) \\ \text{d'où } P_t f(x) &= f(x) + \int_0^t A(P_s f)(x) ds. \end{aligned}$$

Le théorème de Fubini achève la preuve.

Le terme *formellement* cache deux difficultés. La première est d'ordre technique : le raisonnement ci-dessus tient pour acquis que $P_t(f) \in \mathcal{D}(A)$ ce qui n'est généralement pas le cas avec notre définition élémentaire du domaine du générateur (par exemple si b et σ ne sont pas \mathcal{C}^2 pour les EDS ou lorsque h n'est pas \mathcal{C}^1). La seconde est plus profonde et apparaît lorsque $(A, \mathcal{D}(A))$ ne caractérisent pas le semi-groupe $(P_t)_{t \geq 0}$. Un exemple élémentaire d'une telle situation "mal posée" est fournie par les EDO_h admettant plusieurs solutions issues d'un même point (d'équilibre). Ceci peut arriver lorsque la fonction h est simplement continue sans être localement lipschitzienne (cf. Ex. 2, Sect. 2).

Pour une démonstration rigoureuse et complète de la caractérisation (6), on pourra consulter [19] (Chap. 4). En cas de problème “mal posé”, on fait appel au théorème de Echeverria–Weiss (1982) qui construit “à la main” un processus de Markov de semi-groupe P_t , stationnaire de loi initiale ν (cf. [19], Th. 9.17, p. 248).

3. VERS LE RÉGIME STATIONNAIRE

On travaille avec $(X_t)_{t \geq 0}$ processus de Markov homogène, de semi-groupe de transition fellerien $(P_t)_{t \geq 0}$. Pour tout $t \geq 0$, on définit la mesure aléatoire $\nu_t(\omega, y)$ par

$$\nu_t(\omega, dy) := \frac{1}{t} \int_0^t \delta_{X_s(\omega)}(dy) ds,$$

ce qui correspond, en termes d’intégrale par rapport à cette mesure, à

$$\nu_t(\omega, f) = \frac{1}{t} \int_0^t f(X_s) ds.$$

Ces moyennes empiriques convergent-elles quand t tend vers $+\infty$ vers une probabilité invariante pour X ? Autrement dit

$$\forall x \in \mathbb{R}^d, \quad \mathbb{P}_x\text{-}p.s. \quad \nu_t(\omega, dy) \longrightarrow \mathcal{I}_P \neq \emptyset ?$$

On va énoncer deux séries d’hypothèses sous lesquelles on sait répondre à cette question.

3.1. Positive récurrence et théorème ergodique

C’est le premier cadre dans lequel on peut établir, sous des hypothèses raisonnables, la convergence *p.s.* de la famille des ν_t vers une unique mesure invariante. La méthode s’appuie sur le théorème ergodique et permet d’obtenir une propriété de positive récurrence.

La première hypothèse est l’*irréductibilité*. Il n’est pas dans notre propos d’explorer ici la notion d’irréductibilité (cf. par exemple [35] dans un cadre à temps discret). Pour les processus de Markov à temps continu, on s’appuie généralement sur le critère suivant : un processus de Markov est irréductible *dès que*, à un instant $t_0 > 0$ donné, le semi-groupe de transition admet une densité strictement positive, *i.e.*

$$\exists t_0 > 0 \quad \text{tel que} \quad \begin{cases} P_{t_0}(x, dy) &= p_{t_0}(x, y) dy \\ dy\text{-}p.p., & \forall x \in \mathbb{R}^d, \quad p_{t_0}(x, y) > 0. \end{cases} \quad (7)$$

Cette condition sert à établir l’unicité de l’(éventuelle) probabilité invariante.

Une famille de processus de Markov classiquement connue pour vérifier cette condition est la classe des *diffusions uniformément elliptiques à coefficients dans C_b^∞* . Par uniformément elliptique, on entend $\sigma\sigma^* \geq \varepsilon_0 I_d$, $\varepsilon_0 > 0$. Une fonction a la régularité C_b^∞ , si elle est bornée, indéfiniment différentiable, à dérivées partielles successives toutes bornées. On trouvera dans [2] les inégalités de minoration sur $p_t(x, y)$ permettant d’établir ce résultat. Ces conditions peuvent cependant être affaiblies en troquant la bornitude pour la à croissance linéaire et l’uniforme ellipticité pour l’hypo-ellipticité sous condition de Hörmander forte. Pour des résultats dans cette direction, consulter notamment [1] et [8].

En revanche, les EDO_h ne vérifient *jamais* d’hypothèse de cette nature.

L’autre hypothèse, classique dans le domaine des algorithmes stochastiques, a pour fonction d’assurer l’existence d’une probabilité invariante. Elle consiste à supposer l’existence d’une *fonction de Lyapunov* pour le processus de Markov :

CRITÈRE DE HAS' MINSKII

On suppose qu'il existe une fonction $V : \mathbb{R}^d \longrightarrow \mathbb{R}_+^*$ telle que

- (i) $\lim_{|x| \rightarrow +\infty} V(x) = +\infty$;
- (ii) $AV \leq C^{te}$ et $\limsup_{|x| \rightarrow +\infty} AV(x) < 0$ (cette condition traduit une force de rappel) ;
- (iii) $P_t V(x) = V(x) + \int_0^t P_s AV(x) ds$.

EXEMPLE DE FONCTION DE LYAPUNOV

La fonction $V(x) := \frac{1}{2}|x|^2 + 1$ vérifie la condition (ii) de Lyapunov pour l'EDS $_{b,\sigma}$ si et seulement si

$$\forall x \in \mathbb{R}^d \quad (x|b(x)) + \frac{1}{2}Tr(\sigma\sigma^*)(x) \leq C^{te} \text{ et } \limsup_{|x| \rightarrow +\infty} \left[(x|b(x)) + \frac{1}{2}Tr(\sigma\sigma^*)(x) \right] < 0$$

(on rappelle que, par ailleurs $\|b(x)\| + \|\sigma(x)\| = O(|x|)$.) La condition (iii) s'obtient *via* un procédé d'approximation de la fonction V : on établit d'abord (iii) pour les fonctions $x \mapsto \varphi_K(|x|^2)$ où $\varphi_K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est \mathcal{C}^2 , croissante et vérifie $\varphi_K(v) = v$ si $v \in [0, K]$, $\varphi_K(v) = \varphi_K(K+1)$ si $v \geq K+1$ (pour plus détails on peut se reporter à l'annexe B).

Ainsi, la fonction V ci-dessus est une fonction de Lyapunov pour le processus d'Ornstein–Uhlenbeck

$$dX_t = -\frac{1}{2}X_t dt + \sigma dW_t, \quad \text{avec } \sigma \in \mathcal{S}_+(d, \mathbb{R}).$$

Nous sommes maintenant en mesure d'énoncer un résultat d'existence et d'unicité de la mesure invariante pour $(X_t)_{t \geq 0}$. La notation $\mathcal{L}_{\text{bloc}}^1(\nu)$ dans le théorème ci-après désigne *l'ensemble des fonctions ν -intégrables, bornées sur les compact de \mathbb{R}^d* .

Théorème 1. *Soit $(X_t)_{t \geq 0}$ un processus de Markov fellerien. S'il vérifie le critère de Has'minskii ci-avant (fonction de Lyapunov) et l'hypothèse d'irréductibilité (7), alors*

- (a) $\mathcal{I}_P = \{\nu\}$;
- (b) $\forall f \in \mathcal{L}_{\text{bloc}}^1(\nu), \forall x \in \mathbb{R}^d, \mathbb{P}_x$ -p.s. $\frac{1}{t} \int_0^t f(X_s) ds \longrightarrow \int f d\nu$ (propriété dite de récurrence positive).

Démonstration. Pour simplifier on se place dans le cas où $AV(x) \rightarrow -\infty$; pour le cas général, on pourra s'appuyer sur [19] (Th. 9.9, p. 243 et Ex. 47, p. 272).

(a) On commence par montrer que $\mathcal{I}_P \neq \emptyset$: soit $x_0 \in \mathbb{R}^d$ fixé. On raisonne sur la mesure moyenne du semi-groupe de transition.

$$\frac{1}{t} \int_0^t P_s(C^{te} - AV)(x_0) ds = \frac{V(x_0) - P_t V(x_0)}{t} + C^{te} \leq \frac{V(x_0)}{t} + C^{te}.$$

Par suite

$$\sup_{t \geq 1} \frac{1}{t} \int_0^t P_s(C^{te} - AV)(x_0) ds < +\infty, \quad \text{d'où} \quad \Longrightarrow \quad \left(\frac{1}{t} \int_0^t P_s(x_0, dy) \right)_{t \geq 1} \text{ est tendue.}$$

Soit alors ν une valeur d'adhérence faible lorsque $t \rightarrow +\infty$, autrement dit

$$\frac{1}{t_p} \int_0^{t_p} P_s(x_0, dy) \xrightarrow{(\mathbb{R}^d)} \nu \text{ quand } t_p \rightarrow +\infty.$$

Soit $f \in \mathcal{C}_0$. Alors

$$\frac{1}{t_p} \int_0^{t_p} P_s(f)(x_0) ds - \frac{1}{t_p} \int_0^{t_p} P_s(P_{t_p}f)(x_0) ds = O\left(\frac{2\|f\|_\infty}{t_p}\right) \xrightarrow{p \rightarrow +\infty} 0.$$

Mais $f \in \mathcal{C}_0$ entraîne que $P_{t_p}f \in \mathcal{C}_0$, donc chacun des deux termes du membre de gauche converge – respectivement vers $\nu(f)$ et $\nu(P_{t_p}f)$ – ce qui conduit à l'égalité

$$\nu(f) = \nu(P_{t_p}f),$$

et donc l'ensemble \mathcal{I}_P des mesures invariantes n'est pas vide.

Il s'agit maintenant de montrer que \mathcal{I}_P est réduit à un singleton. Soit $\nu \in \mathcal{I}_P$. D'après l'hypothèse d'irréductibilité (7) sur X , on a en t_0 ,

$$\nu(f) = \nu(P_{t_0}(f)) = \int \left(\int p_{t_0}(x, y) f(y) dy \right) \nu(dx) = \int f(y) \underbrace{\left(\int p_{t_0}(x, y) \nu(dx) \right)}_{>0} dy.$$

Par suite, $\nu(dy) \sim \lambda_d(dy)$, où λ_d désigne la mesure de Lebesgue sur \mathbb{R}^d . Or, on montre classiquement que si un semi-groupe de transition vérifie que ses probabilités invariantes sont équivalentes deux à deux, alors celui-ci admet en fait *au plus une* probabilité invariante (cf. par exemple [17], Prop. 6.1.9, p. 188). L'unicité de ν est donc établie.

(b) On se place sur l'espace canonique (*i.e.* $X_t(\omega) := \omega(t)$, $\omega \in \mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$). Soit $f \in \mathcal{L}_{bloc}^1(\nu)$; la fonction $t \mapsto f(X_t(\omega))$ est localement bornée comme composée de fonctions localement bornées, donc localement intégrable sur \mathbb{R}_+ . On pose

$$B_f := \left\{ \omega \in \Omega / \frac{1}{t} \int_0^t f(X_s(\omega)) ds \longrightarrow \int f d\nu \right\}.$$

La probabilité ν étant ergodique, le théorème ergodique ponctuel entraîne que $\mathbb{P}_\nu(B_f) = 1$, autrement dit

$$\nu(dx)\text{-p.s.} \quad \mathbb{P}_x(B_f) = 1.$$

D'autre part, on vérifie que l'événement B_f est invariant par l'opérateur de décalage car la convergence au sens de Césaro ne dépend pas du comportement d'une fonction à horizon fini. En d'autres termes $B_f = \theta_t^{-1}(B_f)$ d'où il découle *via* la propriété de Markov que

$$\mathbb{P}_{x_0}(B_f) = \mathbb{E}_{x_0}(\mathbf{1}_{B_f} \circ \theta_{t_0}) = \mathbb{E}_{x_0}(\mathbb{P}_{X_{t_0}}(B_f)) = \int_{\mathbb{R}^d} \mathbb{P}_x(B_f) P_{t_0}(x_0, dx).$$

Or $P_{t_0}(x_0, dy) \sim \nu$ puisque toutes deux sont équivalentes à la mesure de Lebesgue λ_d . En conséquence $\int_{\mathbb{R}^d} \mathbb{P}_x(B_f) P_{t_0}(x_0, dx) = 1$.

Cette démarche – ainsi que celle qui suit dans laquelle des théorèmes de convergence de martingales se substitueront au théorème ergodique – peut également être mise en œuvre dans des cadres “moins” fellériens. Signalons, dans cette direction, les travaux récents d'Attali (cf. [3]) sur des modèles markoviens quasi-fellériens. *Grosso modo*, il est possible, lorsque le bruit n'est pas dégénéré, de considérer des modèles mettant en jeu des fonctions localement Riemann intégrables en lieu et place de fonctions continues.

3.2. Stabilité et méthodes de martingales

Ce paragraphe va permettre de traiter le cas des *EDO*, inaccessible à la méthode précédente pour cause de défaut systématique d'irréductibilité, ainsi que celui des diffusions éventuellement dégénérées. On va travailler dans le cadre d'une *hypothèse de Lyapunov "renforcée"* : on suppose qu'il existe une fonction deux fois continûment différentiables $V : \mathbb{R}^d \rightarrow \mathbb{R}_+^*$ vérifiant

$$\begin{cases} (i) & \lim_{|x| \rightarrow +\infty} V(x) = +\infty \\ (ii) & AV \leq \beta - \alpha V^a, \quad \alpha > 0, \quad \beta \in \mathbb{R}, \quad a \in]\frac{1}{2}, 1] \\ (iii) & |b|^2 + \|\sigma\|^2 + |\nabla V|^{2a} = O(V^a). \end{cases} \quad (8)$$

(À noter que $\underline{v} := \min V > 0$ sous ces hypothèses.) Pour illustrer cette condition de Lyapunov renforcée, signalons que si $V(x) = |x|^2 + 1$, la condition (iii) se traduit sur le coefficient de dérive b par $b(x) \asymp -|x|^r \frac{x}{|x|}$, avec $2a - 1 < r \leq a$.

Théorème 2. *Soit $(X_t)_{t \geq 0}$ un processus de Markov fellerien. Sous l'hypothèse de Lyapunov renforcée (8), on a*

$$(a) \quad \forall x \in \mathbb{R}^d, \quad \mathbb{P}_x\text{-p.s.}, \quad \begin{cases} \sup_{t \geq 1} \frac{1}{t} \int_0^t \sqrt{V}(X_s) ds < +\infty \text{ et } \sup_{t \in \mathbb{R}} \mathbb{E}(V(X_t)) < +\infty & \text{si } a = 1 \\ \sup_{t \geq 1} \frac{1}{t} \int_0^t V^{a-\frac{1}{2}-\varepsilon}(X_s) ds < +\infty \text{ pour tout } \varepsilon \in]0, a - 1/2] & \text{si } 1/2 < a < 1. \end{cases}$$

En particulier, $\forall x \in \mathbb{R}^d, \quad \mathbb{P}_x(d\omega)\text{-p.s.}, (\nu_t(\omega, dy))_{t \geq 1}$ est tendue.

$$(b) \quad \forall x \in \mathbb{R}^d, \quad \mathbb{P}_x\text{-p.s.}, \quad \forall f \in \mathcal{C}_K^2 \quad \frac{1}{t} \int_0^t Af(X_s) ds \xrightarrow{t \rightarrow +\infty} 0.$$

La caractérisation (6) du paragraphe 2.2 des probabilités invariantes entraîne alors (pour les *EDO* et les *EDS*) que,

$$\forall x \in \mathbb{R}^d, \quad \mathbb{P}_x\text{-p.s.} \quad \text{dist}_{\text{étroite}}(\nu_t(\omega, dx), \mathcal{I}_{EDO/EDS}) \xrightarrow{t \rightarrow +\infty} 0. \quad (9)$$

En outre l'ensemble $\mathcal{I}_{EDO/EDS}$ est un convexe faiblement compact.

La propriété (9) est appelée *stabilité*.

Démonstration. Pour simplifier, on se contente ici de traiter le cas (plus simple) $a = 1$ (pour le cas $1/2 < a < 1$ voir annexe C).

Étape 1. On applique la formule d'Itô à la fonction V et à la diffusion issue de $x \in \mathbb{R}^d$

$$V(X_t) = V(x) + \int_0^t AV(X_s) ds + \underbrace{\int_0^t (\nabla V^* \sigma)(X_s) dW_s}_{\text{martingale locale}}$$

(où ∇V^* désigne le transposé de ∇V). On vérifie en localisant la martingale locale et en s'appuyant sur (ii) que $V(X_t) \in L^1(\mathbb{P}_x)$ i.e. $P_t V(x)$ est fini. Par un procédé d'approximation de V (cf. annexe B), on obtient ensuite la propriété de représentation de $P_t V$ attendue :

$$P_t V(x) = V(x) + \int_0^t P_s AV(x) ds$$

d'où $\frac{d}{dt} P_t V = P_t AV \leq \beta - \alpha P_t V.$

Le lemme de Gronwall fournit finalement

$$\forall t \geq 0, P_t V(x) \leq e^{-\alpha t} \left[V(x) + \frac{\beta}{\alpha} (e^{\alpha t} - 1) \right] \leq \max \left(V(x), \frac{\beta}{\alpha} \right) \quad (10)$$

si bien que, d'une part, pour tout $t \geq 0$, $\sup_t \mathbb{E}_x(V(X_t)) < +\infty$.

Étape 2. D'autre part, l'équation (10) et l'inégalité de convexité $\sqrt{1+u} \leq 1+u/2$, $u \geq -1$ entraînent

$$\begin{aligned} P_t \sqrt{V}(x) &\leq \sqrt{P_t V}(x) \leq \left(V(x) - (1 - e^{-\alpha t}) \left(V - \frac{\beta}{\alpha} \right) \right)^{\frac{1}{2}} \\ &\leq \sqrt{V}(x) \left(1 - \frac{1 - e^{-\alpha t}}{2} \left(1 - \frac{\beta}{\alpha V(x)} \right) \right) \leq \sqrt{V}(x) \frac{1 + e^{-\alpha t}}{2} + \frac{\beta}{\alpha} \frac{1 - e^{-\alpha t}}{2\sqrt{V}} \end{aligned}$$

d'où l'on déduit immédiatement que $A\sqrt{V} = \lim_{t \rightarrow 0} \frac{P_t \sqrt{V} - \sqrt{V}}{t} \leq \tilde{\beta} - \tilde{\alpha} \sqrt{V}$ avec $\tilde{\alpha} := \alpha/2$ et $\tilde{\beta} := \beta/(2\sqrt{V})$.

Étape 3 (Tension *p.s.*). Il vient

$$0 \leq \sqrt{V}(X_t) = \sqrt{V}(x) + \int_0^t A\sqrt{V}(X_s) ds + \underbrace{\int_0^t (\nabla \sqrt{V}^* \sigma)(X_s) dW_s}_{=: O(\sqrt{V}(X_s))}$$

d'où

$$\frac{1}{t} \int_0^t -A\sqrt{V}(X_s) ds \leq \frac{\sqrt{V}(x)}{t} + \frac{1}{t} \int_0^t O(\sqrt{V}(X_s)) dW_s.$$

Si l'on pose $M_t := \int_1^t \frac{O(\sqrt{V}(X_s))}{s} dW_s$, alors $\langle M \rangle_\infty = \int_1^\infty \frac{O(V(X_s))}{s^2} ds \in L^1$. Ceci entraîne que M_t converge presque sûrement vers M_∞ élément de L^2 . Le lemme de Kronecker (*cf.* [24]) permet alors d'écrire que $\frac{1}{t} \int_0^t O(\sqrt{V}(X_s)) dW_s \xrightarrow{p.s.} 0$. Il en découle immédiatement que

$$\limsup_t \frac{1}{t} \int_0^t -A\sqrt{V}(X_s) ds \stackrel{p.s.}{\leq} 0 \quad i.e. \quad \limsup_t \frac{1}{t} \int_0^t \sqrt{V}(X_s) ds \leq \frac{\tilde{\beta}}{\tilde{\alpha}}.$$

Étape 4 (Identification). Soit $f \in \mathcal{C}_K^2$; on vérifie que

$$\frac{1}{t} \int_0^t Af(X_s) ds = \frac{f(X_t) - f(x)}{t} - \frac{1}{t} \int_0^t f'(X_s) \cdot \sigma(X_s) dW_s \stackrel{p.s.}{\rightarrow} 0.$$

En effet, le premier terme à droite de l'égalité est borné par $2\|f\|_\infty/t$ et le second terme est une (vraie) martingale de carré intégrable qui se traite comme son analogue de l'étape 3 (en plus simple).

Étape 5 (Compacité de \mathcal{I}_P). La convexité de \mathcal{I}_P est évidente, sa fermeture découle du caractère fellérien du semi-groupe de transition. Soit maintenant $\nu \in \mathcal{I}_P$. D'après (10), pour tout $K > 0$,

$$\nu(V \wedge K) = \nu(P_t(V \wedge K)) \leq \int P_t(V)(x) \wedge K \nu(dx) \leq \nu \left(\left(e^{-\alpha t} V + \frac{\beta}{\alpha} \right) \wedge K \right).$$

On en déduit en faisant tendre successivement t puis K vers $+\infty$ que $\sup_{\nu \in \mathcal{I}_P} \nu(V) \leq \beta/\alpha$.

Pour un exposé détaillé sur les méthodes de stabilité dans le cadre des chaînes de Markov homogènes à valeurs \mathbb{R}^d , on peut se reporter à [17] (voir aussi [3] dans un cadre quasi-fellérien).

La question naturelle non résolue par le théorème précédent est celle de l'unicité (hors hypothèse d'irréductibilité évidemment).

- Dans le cas des *EDO*, si $\text{Card}(\{x/h(x) = 0\}) \geq 2$, il n'y a jamais unicité. La non-unicité est donc la situation générique. Nous verrons dans la section 4.1 comment utiliser un résultat de type (9) dans un tel cadre.
- Dans le cas des *EDS*, si $\forall x, y \in \mathbb{R}^d, X_t^x - X_t^y \xrightarrow{\mathbb{P}} 0$, alors il y a unicité (y compris donc dans des situations non uniformément elliptiques). On parle parfois pour de tels processus de *diffusion asymptotiquement plate*. Un exemple de telle *EDS* est

$$dX_t = b(X_t)dt + \sigma(X_t) dW_t,$$

où b et σ sont "fortement décroissante" au sens suivant : il existe une constante réelle $c > 0$ telle que

$$\frac{1}{2} \text{Tr}((\sigma(x) - \sigma(y))(\sigma(x) - \sigma(y))^* + (b(x) - b(y)|x - y) \leq -c|x - y|^2.$$

Le processus d'Ornstein-Uhlenbeck vérifie une telle propriété indépendamment de la structure de son coefficient de diffusion. Voir par exemple [4] pour de plus amples développements sur cette propriété.

4. MÉTHODES NUMÉRIQUES

Dans la section précédente, nous avons établi l'existence d'une mesure invariante pour ce processus de Markov, et, dans certains cas, son unicité. Nous allons revenir maintenant à notre objectif initial : les méthodes numériques permettant de simuler cette mesure invariante. Pour ce faire, commençons par déterminer ceux des objets que nous avons rencontrés qui sont simulables et ceux qui ne le sont pas.

On se donne dans tout ce paragraphe un processus de Markov homogène fellérien X .

- Le semi-groupe de transition du processus de Markov, $P_t(x, dy)$, $t \geq 0$, est le plus souvent non "directement" simulable (ceci est équivalent à la simulation exacte de trajectoires du processus de Markov X à des instants arbitraires, voire à tout instant).
- Pour le générateur infinitésimal A , la question n'a pas de sens en tant que telle.
- On a associé à X dans la première partie, une famille paramétrée $(Q_\gamma(x, dy))_{\gamma > 0}$ de transitions markoviennes fellériennes, dans les deux cadres qui sont les nôtres ; plus précisément :

$$Q_\gamma(f)(x) := \begin{cases} \int_{\mathbb{R}^q} f(x + \gamma H(x, u)) \mu(du) & \text{(Algorithme stochastique et EDO)} \\ \int_{\mathbb{R}^q} f(x + \gamma b(x) + \sqrt{\gamma} \sigma(x) u) \mu(du) & \text{(Schéma d'Euler } (\gamma) \text{ d'une EDS).} \end{cases}$$

Pour tout $\gamma > 0$, la chaîne de Markov homogène $(X_n^\gamma)_{n \geq 0}$ associée à Q_γ est simulable par construction même.

Parallèlement, les transitions $Q_\gamma(x, dy)$ jouissent d'une propriété d'approximation à travers leur pseudo-générateur infinitésimal A_γ . Plus précisément,

$$\forall f \in \mathcal{D}(A), \quad A_\gamma(f) := \frac{Q_\gamma(f) - f}{\gamma} \xrightarrow{\gamma \rightarrow 0} A(f).$$

En revanche, et ce un point est essentiel, les Q_γ n'ont aucune propriété de semi-groupe en γ i.e.

$$\boxed{Q_{\gamma_1} \circ Q_{\gamma_2} \neq Q_{\gamma_1 + \gamma_2}.$$

Ce qui se transfère généralement du semi-groupe P_t à temps continu aux transitions Q_γ , au moins pour γ assez petit, ce sont les propriétés d'irréductibilité et l'existence de fonctions de Lyapunov.

4.1. Approximation par des chaînes de Markov homogènes

Pour tout $\gamma > 0$, $(X_n^\gamma)_{n \geq 0}$ est une chaîne de Markov homogène fellerienne (simulable). Il s'agit d'établir le lien entre les mesures invariantes associées à la chaîne $(X_n^\gamma)_{n \geq 0}$, et celles du processus de Markov $(X_t)_{t \geq 0}$.

Théorème 3. *S'il existe $V : \mathbb{R}^d \rightarrow \mathbb{R}_+^*$, $\gamma_0 > 0$ et $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$ tels que*

$$\forall \gamma \in]0, \gamma_0], \quad A_\gamma(V) \leq \beta - \alpha V,$$

et si

$$\forall f \in \mathcal{D}(A), \quad A_\gamma(f) \xrightarrow{\gamma \rightarrow 0} A(f)$$

alors, pour γ assez petit, il vient :

(a) $(X_n^\gamma)_{n \geq 0}$ est stable, c'est-à-dire,

$$\forall x \in \mathbb{R}^d, \quad \mathbb{P}_x(dw)\text{-p.s.} \quad \text{dist}_{\text{étroite}} \left(\frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_k^\gamma(\omega)}, \mathcal{I}_\gamma \right) \xrightarrow{n \rightarrow +\infty} 0,$$

où $\mathcal{I}_\gamma := \{\nu_\gamma / \nu_\gamma Q_\gamma = \nu_\gamma\} \neq \emptyset$, convexe compact.

(b) $\lim_{\gamma \rightarrow 0^+} \text{dist}_{\text{étroite}}(\mathcal{I}_\gamma, \mathcal{I}_A) = 0$, autrement dit si $\gamma_n \rightarrow 0$, alors

$$\nu_{\gamma_n} \xrightarrow{(\mathbb{R}^d)} \mathcal{I}_A$$

où $\mathcal{I}_A = \{\nu / \nu(A) = 0\}$ désigne l'ensemble des probabilités invariantes du processus X .

Applications.

- **EDS UNIFORMÉMENT ELLIPTIQUES** : alors $\mathcal{I}_{EDS} = \{\nu\}$, $\mathcal{I}_\gamma = \{\nu_\gamma\}$ (au moins pour γ assez petit) et

$$\nu_\gamma \xrightarrow{(\mathbb{R}^d)} \nu \quad \text{quand } \gamma \rightarrow 0 \quad (\text{Talay 90 [44]}).$$

Dans cet article originel [44], $V(x) = x^2$, b et σ sont infiniment dérivables à dérivées toutes bornées et σ est bornée et uniformément elliptique. On peut alors préciser l'ordre des vitesses de convergence le long de fonctions-test $f \in \mathcal{C}^\infty$ à dérivées sous-polynomiales :

$$\text{dist}_{\text{étroite}}(\nu_\gamma, \nu) = \begin{cases} O(\gamma) & \text{schéma d'Euler} \\ O(\gamma^2) & \text{schéma du second ordre (non développé ici).} \end{cases}$$

- **ALGORITHMES STOCHASTIQUES** : (Benaïm 96 [6], Fort–Pagès 96 [20], mais aussi, dans le cadre proche des systèmes dynamiques perturbés, Kifer [28], etc.). La difficulté propre à ce cadre est que généralement \mathcal{I}_{EDO} est infini. Ouvrons une brève parenthèse pour expliquer comment utiliser sur un plan pratique ce type de résultat dans le cadre de l'approximation stochastique.

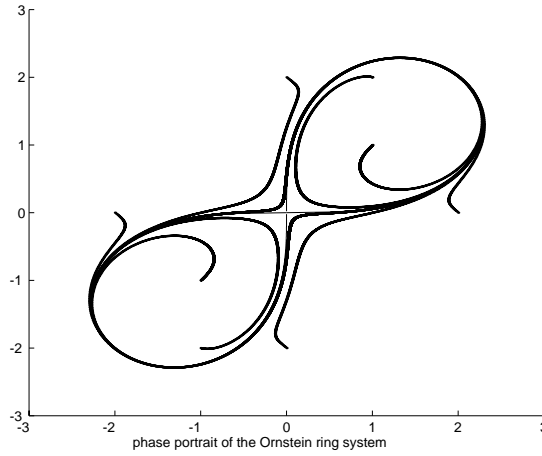


FIGURE 1. Flot de l’anneau d’Ornstein.

- THÉORÈME DE RÉCURRENCE DE POINCARÉ (cf., e.g. [29]) :
 - Si $\nu \in \mathcal{I}_{EDO}$, alors $\text{supp}\nu \subset \mathbf{B}_{EDO_h}$ où \mathbf{B}_{EDO_h} désigne le *noyau de Birkhoff* d’ EDO_h , i.e.

$$\mathbf{B}_{EDO_h} := \overline{\{x_0 \in \mathbb{R}^d / x_0 \text{ est valeur d'adhérence d'une trajectoire d}'EDO_h \text{ issue de } x_0\}}. \quad (11)$$

- En particulier, pour tout voisinage (ici compact) $K_\varepsilon := \{x \in \mathbb{R}^d / d(x, \mathbf{B}_{EDO_h}) \leq \varepsilon\}$ de \mathbf{B}_{EDO_h} ,

$$\nu_\gamma(K_\varepsilon) \xrightarrow{\gamma \rightarrow 0} 1.$$

Ceci illustre le fait que, l’algorithme à pas constant “petit” passe asymptotiquement l’essentiel de son temps dans un voisinage arbitrairement petit du noyau de Birkhoff. Cet ensemble apparaît donc comme l’ensemble (ou comme contenant l’ensemble) des *cibles véritables* de l’algorithme.

- ILLUSTRATION (L’ANNEAU D’ORNSTEIN) :
On considère un algorithme linéairement perturbé

$$\begin{aligned} X_{n+1} &= X_n + \gamma_{n+1} (h(X_n) + \varepsilon_{n+1}), \quad \varepsilon_n \text{ i.i.d.}, \quad \mathbb{E}(\varepsilon_1) = 0, \quad \text{Var}(\varepsilon_1) = I_2, \\ \gamma_n &> 0, \quad \sum_n \gamma_n = +\infty \end{aligned} \quad (12)$$

construit à partir du système différentiel de l’anneau d’Ornstein (dont le flot est représenté dans la Fig. 1 ; une expression analytique de h est disponible dans [20]). Nous allons comparer sur cet exemple le comportement *p.s.* de l’algorithme (item (a) de la proposition ci-après) et celui des limites faibles de ses probabilités invariantes ν_γ à pas $\gamma_n = \gamma > 0$ constant (item (b)). Le point (a) s’appuie sur des techniques non développées ici.

Proposition 1 (Fort–Pagès 99 [20]).

- (a) Pas décroissant ($\gamma_n \rightarrow 0$, $\sum_n \gamma_n^2 < +\infty$) : l’ensemble \mathcal{X}^∞ des valeurs d’adhérence d’une trajectoire de l’algorithme à pas décroissant est (*p.s.*) donné par

$$\mathcal{X}^\infty = L \quad \text{ou} \quad \mathcal{X}^\infty = L^+ \quad \text{ou} \quad \mathcal{X}^\infty = L^-$$

où L (resp. L^\pm) désigne la lemniscate (resp. ses boucles gauches et droites).

- (b) Pas constant $\gamma_n = \gamma$: l'origine O est le seul support possible d'une probabilité invariante obtenue comme valeur d'adhérence faible des ν_γ , en d'autres termes

$$\nu_\gamma \xrightarrow{(\mathbb{R}^d)} \delta_0.$$

(On notera que O est un point selle du système différentiel sous-jacent.)

On constate donc sur cet exemple que les mesures invariantes de l'algorithme à pas constant ν^γ voient beaucoup mieux la cible (le zéro de h) lorsque γ est petit que l'algorithme à pas constant. Nous verrons plus loin qu'il est cependant possible d'exploiter l'algorithme à pas non constant pour mettre en évidence la cible *p.s.*

Démonstration (esquisse).

- (a) On montre *via* la méthode de l'équation différentielle – dite aussi méthode de Kushner et Clark (cf. [30]) – que l'ensemble \mathcal{X}^∞ des valeurs d'adhérence d'une trajectoire de l'algorithme à pas (suffisamment) décroissant est compact, connexe. Il est en outre intérieurement transitif par chaîne (cf. Benaïm [7]). Il est alors possible de caractériser \mathcal{X}^∞ comme l'un des ensembles suivants :

$$\mathcal{X}^\infty = L \text{ ou } \mathcal{X}^\infty = L^+ \text{ ou } \mathcal{X}^\infty = L^- \text{ ou } \mathcal{X}^\infty = \{A\} \text{ ou } \mathcal{X}^\infty = \{B\} \text{ ou } \mathcal{X}^\infty = \{O\}$$

où A et B sont les centres des boucles et O leur point de contact.

Les points A et B étant des répulseurs d' EDO_h en lesquels le bruit est non évanescent et O étant un point selle de même nature, ils constituent donc (cf., e.g., Brandière–Duflo 96 [12]) des pièges répulsifs et excités, nécessairement évités *p.s.* par l'algorithme.

- (b) Le noyau de Birkhoff est donné par $\mathbf{B}_{EDO} = \{A, O, B\}$ (première réduction) mais A et B sont maxima locaux en lesquels le bruit est non évanescent donc $\{O\}$ est le seul support possible d'une probabilité invariante obtenue comme valeur d'adhérence faible des ν_γ (cf. [20] ou [6]). En d'autres termes

$$\lim_{\gamma \rightarrow 0} \nu_\gamma = \delta_O.$$

4.2. Approximation par des chaînes de Markov inhomogènes

On considère une chaîne de Markov inhomogène simulable $(X_n)_{n \geq 0}$ fabriquée à partir de la famille de transitions Q_γ de la façon suivante :

$$\begin{cases} \mathbb{E}(f(X_{n+1})/\mathcal{F}_n^X) = Q_{\gamma_{n+1}} f(X_n) \\ \gamma_n \rightarrow 0, \quad \sum_{n \geq 1} \gamma_n = +\infty. \end{cases}$$

Dans le cas des *diffusions*, ceci peut correspondre par exemple à une discrétisation d'Euler à pas décroissant sous la forme

$$X_{n+1} = X_n + \gamma_{n+1} b(X_n) + \sqrt{\gamma_{n+1}} \sigma(X_n) U_{n+1}, \quad X_0 \in \mathbb{R}^d \tag{13}$$

où $(U_n)_{n \geq 1}$ est un bruit blanc, indépendant de X_0 , de carré intégrable.

Dans le cas des *algorithmes stochastiques*, la chaîne de Markov inhomogène est donc donnée par la formule récursive (1) *i.e.*

$$X_{n+1} = X_n + \gamma_{n+1} H(X_n, \varepsilon_{n+1}), \quad X_0 \in \mathbb{R}^d,$$

où $(\varepsilon_n)_{n \geq 1}$ est une suite de variables aléatoires (intégrables) indépendantes et de même loi, indépendantes de X_0 .

4.2.1. Convergence faible de la chaîne inhomogène

Les deux principaux résultats – adaptés à leurs cadres d'utilisation respectifs – sont les suivants :

Théorème 4. (*Algorithmes stochastiques : “méthode de l'EDO pour la convergence faible”, Kushner 77, cf. [30]*) Soit $L \subset \mathbb{R}^d$ un attracteur compact d'EDO $_h \equiv \dot{x} = h(x)$ au sens où

$$\text{dist}(\Phi_t(x), L) \xrightarrow{t \rightarrow +\infty} 0 \quad \text{uniformément en } x \text{ sur les compacts de } \mathbb{R}^d.$$

Si $(X_n)_{n \geq 0}$ est tendue et $\sup_{n \geq 0} \mathbb{E}|H(X_n, \varepsilon_{n+1})|^2 < +\infty$ (ce qui est vrai sous une hypothèse de type Lyapunov quadratique), alors

$$\text{dist}(X_n, L) \xrightarrow{\mathbb{P}} 0.$$

Théorème 5. (*“Méthode de l'EDS”, Bouton 87, cf. [11], Kushner 87, cf. [31, 32], et plus récemment, Pelletier 98, cf. [36], Basak–Hu–Wei 97, cf. [5]*). Sous les hypothèses

- $\mathcal{I}_{\text{diffusion}} = \{\nu\}$ et $P_t(x, dy) \implies \nu(dy)$ uniformément sur les compacts ;
- Sous une hypothèse de Lyapunov, à savoir qu'il existe $V : \mathbb{R}^d \rightarrow \mathbb{R}_+^*$, $\alpha > 0$, $\beta \in \mathbb{R}$, tels que

$$\lim_{|x| \rightarrow +\infty} V(x) = +\infty \quad \text{et} \quad A_\gamma(V) = \frac{Q_\gamma V - V}{\gamma} \leq \beta - \alpha V,$$

on a

$$X_n \xrightarrow{\mathcal{L}} \nu.$$

4.2.2. Approximation par mesure empirique dans le cadre inhomogène

Après l'exposé de ces deux premières approches numériques, l'une par des chaînes de Markov homogènes, l'autre par des chaînes de Markov inhomogènes, on est tenté d'imaginer une synthèse entre ces deux méthodes, qui en auraient simultanément les qualités : viser juste trajectoire par trajectoire (propriétés de type ergodique, éventuellement sous une forme affaiblie comme la stabilité). L'objectif doit être atteint sans (trop) perdre en vitesse de convergence évidemment.

C'est l'objet de ce paragraphe, où, pour ne pas trop nous disperser, nous allons nous concentrer plutôt sur les diffusions. Cependant, le type de résultat que nous obtiendrons admet un analogue pour les algorithmes stochastiques (établi antérieurement d'ailleurs).

Nous allons considérer des chaînes de Markov inhomogènes *stables*, soit donc $(X_n)_{n \geq 0}$ une chaîne inhomogène simulable (et fellérienne), telle que

$$\mathbb{E}(f(X_{n+1})/\mathcal{F}_n^X) = Q_{\gamma_{n+1}}(f)(X_n),$$

avec $\gamma_n \downarrow 0$ et $\sum_n \gamma_n = +\infty$.

Dans le cas des *diffusions*, ceci correspond à la discrétisation suivante :

$$X_{n+1} = X_n + \gamma_{n+1}b(X_n) + \sqrt{\gamma_{n+1}}\sigma(X_n)U_{n+1},$$

avec $(U_n)_{n \geq 1}$ bruit blanc de carré intégrable. Dans celui des *algorithmes stochastiques*, elles sont définies par la formule récursive

$$X_{n+1} = X_n + \gamma_{n+1}H(X_n, \varepsilon_{n+1}),$$

avec $(\varepsilon_n)_{n \geq 1}$ suite de variables aléatoires i.i.d.

On introduit les mesures empiriques pondérées

$$\nu_n^\eta(\omega, dx) := \frac{\eta_1 \delta_{X_0(\omega)} + \eta_2 \delta_{X_1(\omega)} + \cdots + \eta_n \delta_{X_{n-1}(\omega)}}{\eta_1 + \eta_2 + \cdots + \eta_n},$$

où $\eta_n \geq 0$ et $H_n := \eta_1 + \cdots + \eta_n \rightarrow +\infty$. Les choix qui sont intuitivement naturels pour les poids η_n sont $\eta_n := \gamma_n$ ou bien $\eta_n := 1$.

L'intégrale d'une fonction f contre la mesure aléatoire ν_n^η est alors donnée par

$$\nu_n^\eta(\omega, f) = \frac{1}{H_n} \sum_{k=0}^{n-1} \eta_{k+1} f(X_k(\omega)),$$

ce qui conduit immédiatement à la formule récursive suivante

$$\nu_{n+1}^\eta(f) = \frac{\eta_{n+1}}{H_{n+1}} f(X_n(\omega)) + \left(1 - \frac{\eta_{n+1}}{H_{n+1}}\right) \nu_n^\eta.$$

4.2.2.1 Le cadre diffusion

Les résultats que l'on va énoncer dans ce cadre-là concernant la mesure invariante de la diffusion sont établis sous les hypothèses suivantes :

HYPOTHÈSE DE LYAPUNOV (CADRE DIFFUSION)

Il existe une fonction $V : \mathbb{R}^d \rightarrow \mathbb{R}_+^*$, de classe \mathcal{C}^2 , telle que

$$(\mathcal{L}) \equiv \begin{cases} (i) & \lim_{|x| \rightarrow +\infty} V(x) = +\infty \text{ et } \|D^2V\|_\infty < +\infty, \\ (ii) & |b|^2 + \|\sigma\sigma^*\| + |\nabla V|^2 = O(V), \\ (iii) & (\nabla V|b) + \|D^2V\|_\infty \text{Tr}(\sigma\sigma^*) \leq \beta - \alpha V, \text{ avec } \alpha > 0. \end{cases}$$

Il existe alors $\gamma_0 > 0$ tel que, pour tout $\gamma \in]0, \gamma_0]$,

$$A_\gamma := \frac{Q_\gamma V - V}{\gamma} \leq \tilde{\beta} - \tilde{\alpha}V \quad \text{et} \quad AV \leq \tilde{\beta} - \tilde{\alpha}V.$$

Théorème 6. (Cadre diffusion, Lamberton–Pagès 99, cf. [33]) *Sous les hypothèses suivantes*

- Hypothèse de Lyapunov (\mathcal{L}) ;
- $(U_1 \in L^2 \text{ et } \text{Tr}(\sigma\sigma^*) = o(\sqrt{V}))$ ou $(U_1 \in L^{4+\varepsilon} \text{ pour un } \varepsilon > 0)$;
- $\gamma_n := \frac{c_\lambda}{n^\lambda}$, $c_\lambda > 0$ et $\eta_n := \frac{1}{n^\mu}$ avec $(0 < \lambda < 1 \text{ et } -\infty < \mu \leq 1)$ ou $(\lambda = \mu = 1)$;

la discrétisation d'Euler à pas décroissant est stable : pour toute valeur initiale $X_0 = x_0 \in \mathbb{R}^d$,

$$\mathbb{P}(d\omega)\text{-p.s.}, \quad \text{dist}_{\text{étroite}}(\nu_n^\eta(\omega, dx), \mathcal{I}_{\text{diffusion}}) \xrightarrow{n \rightarrow +\infty} 0. \quad (14)$$

Démonstration (squelette). Les étapes importantes peuvent être schématisées ainsi

- pour $\gamma > 0$ assez petit, V est une fonction de Lyapunov pour $A_\gamma = \frac{Q_\gamma - Id}{\gamma}$:

$$A_\gamma V \leq \beta - \alpha V ;$$

- on estime ensuite la vitesse de convergence de A_γ vers A : pour $f \in \mathcal{C}_K^2$,

$$A_\gamma(f) = A(f) + \sqrt{\gamma}O(\sqrt{V}) + \gamma O(V) ;$$

- le dernier élément-clé est un contrôle dynamique des erreurs de prédiction :

$$\begin{aligned} \|f(X_{n+1}) - \mathbb{E}(f(X_{n+1})/\mathcal{F}_n^X)\|_2 &\leq \|f(X_{n+1}) - f(X_n)\|_2 \\ &\leq \|f\|_{\text{Lip}} \|X_{n+1} - X_n\|_2 \\ &\leq \sqrt{\gamma_{n+1}} \mathbb{E}(V(X_n)). \end{aligned}$$

Remarque. L'énoncé ci-dessus correspond essentiellement au cas $a = 1$ du théorème 2. Une extension au cas $1/2 < a < 1$ est possible.

4.2.2.2 Le cadre algorithme stochastique

Un résultat de stabilité trajectorielle tout à fait analogue – que nous n'énoncerons pas en détail ici – a été obtenu dans [21] (Fort–Pagès 98) pour les algorithmes stochastiques. Évidemment, l'ensemble des probabilités stationnaires de l'EDO étant rarement réduit à un singleton dans les situations d'intérêt, l'usage pratique d'un tel résultat se justifie par la mise en évidence *p.s. sur chaque trajectoire* des “cibles” de l'algorithme telles qu'elles ont été définies par (11), en l'espèce du noyau de Birkhoff \mathbf{B}_{EDO_h} . Ceci s'applique y compris dans des situations où l'algorithme lui-même n'est pas borné. L'intérêt, comme on a pu le constater précédemment dans le cas de l'anneau d'Ornstein, c'est que ce noyau est habituellement beaucoup plus petit que l'ensemble des valeurs d'adhérence d'une trajectoire (bornée) de l'algorithme.

En résumé, cette version du théorème 6 indique que, *p.s.*, on peut *observer sur une seule trajectoire* le phénomène asymptotique mis en évidence sur les mesures invariantes ν_γ des algorithmes à pas constant lorsque $\gamma \rightarrow 0$. C'est un avantage indéniable puisque dans le cadre homogène, la mesure invariante ν_γ d'un algorithme à pas constant γ était déjà un objet asymptotique.

La comparaison avec le résultat de convergence en probabilité du théorème 4 renvoie, en pire, à celle du comportement *p.s.* de l'algorithme lui-même : l'objet limite qui apparaît dans ce théorème est un attracteur L d'EDO_h – *a priori* beaucoup plus gros que le noyau de Birkhoff – qui, en outre, est moins bien caractérisé que dans les résultats relatifs à la convergence *p.s.* trajectorielle de l'algorithme (récurrence par chaîne, *cf.* [7]).

Revenons aux propriétés de stabilité. En fait, l'histoire ne s'arrête pas là car il est également possible d'éliminer *a priori* certains sous-ensembles du noyau Birkhoff, en l'espèce les répulseurs excités d'EDO_h (*i.e.* en lesquels le bruit de l'algorithme n'est pas évanescent) : en effet ceux-ci ne peuvent *p.s.* pas faire partie du support d'une probabilité invariante obtenue comme limite faible de mesures empiriques pondérées. Ceci nécessite cependant certaines hypothèses restrictives sur les pas (*cf.* [21]). Ce résultat s'étend sous réserve d'hypothèses techniques raisonnables aux points-selle (dûment excités) d'EDO_h dans le cas particulier d'une descente de gradient (*i.e.* $h = -\nabla V$) ou de pseudo-gradient (*i.e.* $(h|\nabla V) \leq 0$ et $\{(h|\nabla V) = 0\} = \{h = 0\}$). Ces résultats sont valides sous des hypothèses de moments polynomiaux peu contraignantes (essentiellement L^4).

L'algorithme linéairement perturbé à pas décroissant relatif à l'anneau d'Ornstein (voir (12) dans la Sect. 4.1 et la Prop. 1 qui suit) est une illustration parfaite de ce phénomène : on montre en effet dans [21] que, si $\varepsilon_1 \in L^4$, $\gamma_n \rightarrow 0$, $\sum_n \gamma_n = \sum_n \gamma_n^2 = +\infty$, alors

$$\mathbb{P}(d\omega)\text{-}p.s. \quad \nu_n^\eta(\omega, dx) \xrightarrow{(\mathbb{R}^d)} \delta_O.$$

(Noter que l'hypothèse sur les pas n'est pas compatible avec l'hypothèse standard sur les pas lorsque l'on recherche une convergence *p.s.*) Ce résultat – et les vitesses de convergence en loi qui l'accompagnent (*cf.* [21]) – peuvent aussi être mis en perspective avec les résultats antérieurs sur la moyennisation des algorithmes stochastiques *p.s.* convergents (*cf.* [39, 41] ou, plus récemment [38]).

4.2.3. Application du théorème 6 au processus d'Ornstein–Uhlenbeck : le TCL-*p.s.*

Revenons au théorème précédent sur les diffusions. Pour en illustrer la portée, nous allons montrer qu'il constitue le cadre naturel du théorème Central-Limite “presque-sûr”.

- Nous rappelons que le processus d’Ornstein–Uhlenbeck est solution de l’EDS

$$dX_t = -\frac{1}{2}X_t dt + dW_t.$$

Il rentre donc dans le cadre étudié ci-dessus, avec

$$b(x) := -\frac{1}{2}x, \quad \sigma(x) := Id \quad \text{et} \quad V(x) := |x|^2 + 1.$$

On a alors $(\nabla V|b) + \|D^2V\|_\infty \text{Tr}(\sigma\sigma^*) = -|x|^2 + 2d = (2d + 1) - V$. D’autre part, on vérifie classiquement que $\mathcal{I}_{O.U.} = \{\mathcal{N}(0, Id)\}$.

Enfin, on considère des pas et des poids identiques donnés par : $\eta_n = \gamma_n = \frac{1}{n}$.

Le théorème précédent permet alors d’écrire que si $(X_n)_{n \geq 0}$ est la chaîne de Markov définie par

$$X_n = X_{n-1} - \frac{1}{2n}X_n + \frac{1}{\sqrt{n}}U_n,$$

avec $(U_n)_{n \geq 1}$ bruit blanc de carré intégrable, alors

$$\mathbb{P}(d\omega)\text{-}p.s., \quad \frac{1}{\ln n} \sum_{k=1}^n \frac{1}{k} \delta_{X_{k-1}(\omega)} \xrightarrow{(\mathbb{R}^d)} \mathcal{N}(0, Id).$$

- Pour démontrer le théorème Central-Limite *p.s.*, on introduit les sommes partielles du TCL standard, et on les réécrit de façon récursive :

$$\begin{aligned} Y_n &:= \frac{U_1 + \dots + U_n}{\sqrt{n}} = \sqrt{\frac{n-1}{n}} Y_{n-1} + \frac{U_n}{\sqrt{n}} \\ &= Y_{n-1} - \frac{1}{2n} Y_{n-1} + \frac{1}{\sqrt{n}} U_n + \underbrace{\left(\sqrt{1 - \frac{1}{n}} - \left(1 - \frac{1}{2n}\right) \right)}_{=: O(1/n^2)} Y_{n-1} \\ &= Y_{n-1} - \frac{1}{2n} Y_{n-1} + \frac{1}{\sqrt{n}} U_n + r_n, \end{aligned}$$

avec $r_n := O(\frac{1}{n^2}) \times Y_n$. Par ailleurs, $\sup_{n \geq 1} \mathbb{E}|Y_n| \leq \sup_{n \geq 1} \sqrt{\mathbb{E}|Y_n|^2} \leq \sqrt{d}$, ce qui entraîne immédiatement

$$\sum_{n \geq 1} |r_n| < +\infty \quad p.s.$$

- L’étape suivante de la démonstration consiste à dire que la suite $(Y_n)_{n \geq 0}$ que l’on vient d’introduire se comporte comme la suite $(X_n)_{n \geq 0}$ associée au processus d’Ornstein–Uhlenbeck : en effet

$$|X_n - Y_n| \leq \left(1 - \frac{1}{2n}\right) |X_{n-1} - Y_{n-1}| + |r_n| \xrightarrow{p.s.} 0.$$

Tout ceci mis bout à bout permet alors d’énoncer le théorème suivant :

Théorème 7 (Théorème Central-Limite “presque sûr”). *Soit $(U_n)_{n \geq 1}$ un bruit blanc de carré intégrable sur \mathbb{R}^d , centré réduit, alors*

$$\mathbb{P}\text{-}p.s. \quad \frac{1}{\ln n} \sum_{k=1}^n \frac{1}{k} \delta_{\frac{U_1 + \dots + U_k}{\sqrt{k}}} \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, Id).$$

Commentaires. Il existe une abondante littérature concernant ce théorème et ses nombreuses généralisations. Citons notamment, sans chercher l'exhaustivité :

- Schatte (88, cf. [42]) sous l'hypothèse $\mathbb{E}|U_1|^3 < +\infty$ (voir aussi Fisher [22]) ;
- Brosamler (88, cf. [13]) sous l'hypothèse $\mathbb{E}|U_1|^{2+\delta} < +\infty$ (version fonctionnelle) ;
- Lacey et Philips (90, cf. [34]) sous l'hypothèse $\mathbb{E}|U_1|^2 < +\infty$;
- Chaâbane *et al.* (96, cf. [15]) ou Touati (98, cf. [46]), pour des accroissements de martingale $\sup_n \mathbb{E}(|\Delta M_{n+1}|^2/\mathcal{F}_n) < +\infty$;
- Pelletier (99, cf. [37]), TCL-*p.s.* des algorithmes stochastiques convergents ;
- Chen et Peng (98 [16]) obtiennent des résultats analogues en théorie des valeurs extrêmes ;
- Berkes et Csáki (01 [14]) pour une mise en perspective de ce type de théorème dans un cadre abstrait.

4.2.4. Vitesse de convergence

À ce stade, pour pouvoir contrôler l'utilisation pratique en simulation de la méthode, il est indispensable de proposer des vitesses de convergence (en loi) : en d'autres termes, à quelle vitesse la variable aléatoire $\nu_n^\eta(f)$ converge-t-elle vers le nombre $\nu(f)$ (la mesure invariante ν étant supposée unique).

On se place dans le cadre diffusion, sous les hypothèses suivantes :

- $\mathcal{I}_{\text{diffusion}} = \{\nu\}$;
- $(\nabla V|b) \leq \beta - \alpha V$ et $\text{Tr}(\sigma\sigma^*) = O(V)$;
- $U_1 \in L^8$, $\mathbb{E}(U_1) = 0$ et $\mathbb{E}(U_1^{\otimes 3}) = 0$;
- $\gamma_n = n^{-\lambda}$, $c_\lambda > 0$, et $\eta_n = n^{-\mu}$; avec ($(0 < \lambda < 1, -\infty < \mu \leq 1)$ ou $(\lambda = \mu = 1)$) et $(\lambda + \mu \leq 1)$.

Soit f une "fonction-test" de la forme $f := \nu(f) + A\varphi$, avec $\varphi \in \mathcal{C}^4$, $\nabla\varphi = O(\|\sigma\|)$ et $D^4\varphi$ lipschitzienne et bornée, on cherche à évaluer – dans l'asymptotique significative – la limite en loi suivante :

$$n^\theta (\nu_n^\eta(f) - \nu(f)) \xrightarrow{(\text{Loi})} ?$$

Remarques.

- Le fait de se restreindre à des "fonctions-test" de la forme $f - \nu(f) = A\varphi$ apparaît naturellement dans la démonstration, fondée sur un développement limité de $\varphi(X_{n+1})$ au point X_n . Sans entrer dans les détails techniques, une fonction f étant donnée, l'existence éventuelle de φ dépend de la vitesse à laquelle le semi-groupe $P_t(x, dy)$ converge vers la mesure invariante ν le long des fonctions de ν -intégrale nulle. En effet, supposons pour simplifier que $\nu(f) = 0$. Partant de l'identité

$$P_t(f)(x) = f(x) + \int_0^t A(P_s(f))(x) ds,$$

on voit que, si $P_t(f)(x)$ tend vers 0 lorsque $t \rightarrow +\infty$, alors, l'identité formelle

$$f(x) = A \left(- \int_0^{+\infty} P_s(f)(x) ds \right) \quad (15)$$

fournit naturellement un candidat pour φ , en l'espèce $\varphi(x) := - \int_0^{+\infty} P_s(f)(x) ds$. L'existence effective de φ et sa régularité (d'où découlera la validité de l'interversion de A et de l'intégrale) dépendent donc de la vitesse à laquelle la convergence de $P_t(f)(x)$ vers 0 a lieu. Cette question est abordée de façon très exhaustive pour les diffusions dans [23] ; voir aussi [44] dans un cadre différent.

- La condition $\lambda + \mu \leq 1$ provient du fait que l'on a besoin de montrer que \mathbb{P} -*p.s.* $\nu_n^{\gamma\eta}(\omega, dx) \xrightarrow{(\mathbb{R}^d)} \nu$, où $(\gamma\eta)_n := \gamma_n\eta_n$.

Théorème 8. Pour tout $n \in \mathbb{N}^*$, on pose $\Gamma_n := \gamma_1 + \dots + \gamma_n$. Selon les valeurs des paramètres λ et μ des pas et des poids, on a les résultats suivants :

- (a) pas rapidement décroissant : si $\gamma_n = c_\lambda n^{-\lambda}$, $1/3 < \lambda \leq 1$ (et $c_\lambda > 0$), la vitesse (en loi) optimale est obtenue avec le poids $\eta := \gamma$ et l'on a

$$\sqrt{\Gamma_n} \nu_n^\gamma(f) \xrightarrow{\text{Loi}} \mathcal{N} \left(0; \int_{\mathbb{R}^d} |\sigma^* \varphi|^2 d\nu \right) ;$$

- (b) pas polynomial optimal : si $\gamma_n = \eta_n = c n^{-1/3}$ ($c > 0$),

$$\sqrt{\Gamma_n} \nu_n^\gamma(f) \xrightarrow{\text{Loi}} \mathcal{N} \left(m_\varphi; \int_{\mathbb{R}^d} |\sigma^* \varphi|^2 d\nu \right) ;$$

- (c) pas lentement décroissant : si $\gamma_n = c_\lambda n^{-\lambda}$, $0 < \lambda < 1/3$ ($c_\lambda > 0$) et $\eta_n = n^{-\mu}$, $\mu \leq -1$,

$$\gamma_n^{-1} \nu_n^\eta(f) \xrightarrow{\mathbb{P}} \frac{1 - \mu}{1 - (\lambda + \mu)} m_\varphi$$

où $m_\varphi := \int_{\mathbb{R}^d} \left(\frac{1}{2} D^2 \varphi(x) \cdot b(x)^{\otimes 2} + \Phi_4(x) \right) \nu(dx)$ avec $\Phi_4(x) = \int_{\mathbb{R}^d} \left(\frac{1}{2} < D^3 \varphi(x); b(x), (\sigma(x) \cdot u)^{\otimes 2} > + \frac{1}{24} D^4 \varphi(x) \times (\sigma(x) \cdot u)^{\otimes 4} \right) \mu(du)$.

Commentaires et aspects numériques

- Dans la convergence en loi (a) (pas rapidement décroissants), il apparaît que plus c_λ est grand, plus rapide est la convergence. Ce constat doit être modulé par le fait que cette vitesse se manifeste également d'autant plus tard sur les simulations. Le paramètre c_λ doit donc être réglé expérimentalement. En l'état actuel des choses nous ne disposons pas d'heuristique pour ajuster c_λ .
- Dans la convergence en probabilité (c) (pas lentement décroissants), un phénomène analogue est observé : la meilleure vitesse est atteinte avec des poids très lourds *i.e.* $n^{-\mu}$ avec $\mu \rightarrow -\infty$. En effet

$$\inf_{\mu \leq -1} \left| \frac{1 - \mu}{1 - (\lambda + \mu)} m_\varphi \right| = \lim_{\mu \rightarrow -\infty} \left| \frac{1 - \mu}{1 - (\lambda + \mu)} m_\varphi \right|.$$

Ceci n'est pas contraignant en termes de stabilité numérique puisque, le calcul récursif de $\nu_n^\eta(f)$ ne requiert que la quantité $\tilde{\eta}_n := \eta_n / H_n \sim \frac{1-\mu}{n}$ ($\mu \neq 1$).

Cependant, on observe sur les simulations que, comme pour la convergence en loi, plus $|\mu|$ est grand, plus l'apparition effective de la vitesse de convergence attendue est longue. Le paramètre μ doit donc être réglé également.

- Le cas $\lambda = 1$ correspond au TCL du TCL *p.s.* On constate donc que la discrétisation d'Euler de l'équation d'Ornstein–Uhlenbeck correspond à la vitesse la plus lente observable dans la gamme de pas “puissances” que nous considérons.
- La meilleure vitesse asymptotique est obtenue en (b) avec un biais et une variance asymptotique non nuls. Il est naturel de la comparer de façon heuristique à celle obtenue en mettant en œuvre l'application du

théorème 3 aux diffusions. Soit f une fonction-test selon [44] de la forme $f = A_\gamma \varphi$.

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \int f \, d\nu &= \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \int f \, d\nu_\gamma + \int f \, d\nu_\gamma - \int f \, d\nu \\ &= \underbrace{\frac{1}{n\gamma} \sum_{k=0}^{n-1} \gamma f(X_k) - \int f \, d\nu_\gamma}_{\text{TCL à la vitesse } \sqrt{n\gamma}} + \underbrace{\int f \, d\nu_\gamma - \int f \, d\nu}_{O(\gamma)} \\ &= O\left(\frac{1}{\sqrt{n\gamma}}\right) + O(\gamma). \end{aligned}$$

En conséquence, on est conduit, n étant fixé, à optimiser en γ une quantité de type $\frac{C_1}{\sqrt{n\gamma}} + C_2\gamma$. L'erreur minimale est donc obtenue pour $\gamma \approx C_3 n^{-\frac{1}{3}}$ ce qui conduit à une vitesse (en loi) en $O(n^{-\frac{1}{3}})$. La vitesse est donc analogue à la vitesse optimale obtenue le long des fonctions-test dans la méthode de stabilité à pas décroissant. L'intérêt des mesures empiriques pondérées du schéma d'Euler à pas décroissant réside donc fondamentalement dans son caractère récursif : il n'y a pas besoin de réajuster le pas du schéma d'Euler à chaque étape, il suffit de le laisser "mouliner" jusqu'à obtenir la précision recherchée.

- L'un des atouts du versant "pas lentement décroissant", c'est que la convergence en probabilité vers une constante réelle se prête très naturellement à la mise en œuvre de la méthode de Romberg puisque ce mode de convergence est compatible avec l'addition (cf. Talay–Tubaro 90 [45], pour la première mise en œuvre répertoriée de la méthode dans le cadre des diffusions). En l'espèce il est possible de gagner un ordre de convergence en considérant simultanément l'algorithme à pas γ et l'algorithme à pas $\gamma/2$ (pour un même poids η et une même trajectoire du bruit blanc U_n). En, effet, il est immédiat au vu de (c) que

$$\tilde{\nu}_n^{\eta,\gamma} = 2\nu_n^{\eta,\gamma/2}(f) - \nu_n^{\eta,\gamma}(f) = o(\gamma_n).$$

Une étude fine des termes d'ordre supérieurs montre, sous des hypothèses techniques omises ici, que, si $\lambda < 1/5$

$$\sqrt{\Gamma_n} \tilde{\nu}_n^{\eta,\gamma} \xrightarrow{\text{Loi}} \mathcal{N}(0; \Sigma_{b,\sigma,\varphi})$$

c'est-à-dire un TCL centré à la vitesse $\sqrt{\Gamma_n} \sim n^{\frac{1-\lambda}{2}}$. Lorsque $\lambda = 1/5$ on obtient un TCL biaisé à la vitesse $n^{\frac{2}{5}}$.

On améliore ainsi la vitesse de convergence, jusqu'à atteindre un optimum de $n^{2/5}$. On peut ainsi réitérer le processus, en appliquant la méthode de Romberg à $\tilde{\nu}_n^{\eta,\gamma}$ pour obtenir, au bout de r itérations, un TCL biaisé à la vitesse $n^{\frac{2r}{2r+1}}$. Lorsque r devient grand on bute de façon prévisible sur la vitesse \sqrt{n} . En pratique, l'apparition tangible de cette vitesse est de plus en plus lointaine et la méthode de Romberg dans sa version multiple semble peu convaincante.

4.3. Quelques simulations illustratives (cadre diffusion)

4.4. TCL à pas rapide *versus* vitesse en probabilité à pas lent

Nous proposons ici une rapide illustration de la structure des vitesses de convergence dans le théorème 8, notamment pour comparer sur des simulations réelles comment se comporte $n \mapsto \nu_n^\gamma(f)$ en fonction de la nature

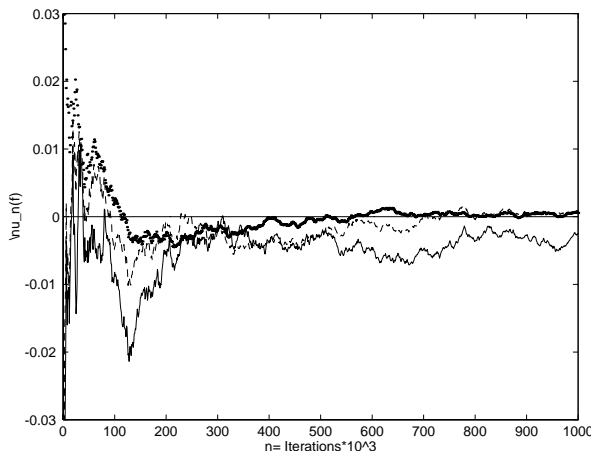


FIGURE 2. $n \mapsto \nu_n^\gamma(f)$ pour différents types de pas polynomiaux.

du pas (rapidement ou lentement décroissant, optimal). À cette fin, on considère le processus d’Ornstein–Uhlenbeck standard $dY_t = -\frac{1}{2}Y_t dt + dW_t$ et la fonction-test

$$f(x) := A \left(\frac{1}{1+x^2} \right) = \frac{1}{1+x^2} + \frac{2}{(1+x^2)^2} - \frac{4}{(1+x^2)^3}.$$

Ensuite pour une vitesse atteignable $\theta \in (0, 1/3[$, on pose

$$\gamma_n^{\text{vite}} := \frac{2\theta}{n^{1-2\theta}} \quad \text{and} \quad \gamma_n^{\text{lent}} := \frac{1-\theta}{1-2\theta} \frac{1}{n^\theta}.$$

Dans les deux cas, les constantes ont été choisies de façon que la vitesse à l’infini soit équivalente à n^θ lorsque $n \rightarrow +\infty$. Lorsque $\theta = 1/3$, on a posé $\gamma_n = \frac{2}{3} \frac{1}{n^{-1/3}}$ pour les mêmes raisons.

Les simulations numériques ont été réalisées pour la vitesse $\theta = 0,30$ proche de la vitesses optimale en régime de pas lent (biais) et rapide (TCL centré) ainsi que pour la vitesse optimale (TCL biaisé), sur $n = 10^6$ itérations. En parallèle, une simulation a été menée avec le pas polynomial conduisant à la vitesse optimale ($\gamma_n = \frac{2}{3}n^{-1/3}$ conduisant à un TCL biaisé à la vitesse $n^{1/3}$).

Dans la figure 2 les traits *épais* et *fin* représentent respectivement les trajectoires de $n \mapsto \nu_n^\gamma(f)$ pour γ^{vite} et γ^{lent} (à la vitesse $n^{-0,30}$) et le trait *pointillé* correspond à la vitesse optimale. On constate que pour une vitesse théorique donnée, la convergence en probabilité semble numériquement plus stable que celle observée avec le TCL centré. Ainsi, dans les simulations ci-dessus, mais ceci est confirmé par ailleurs, la convergence en probabilité supporte même tout à fait la comparaison avec le TCL biaisé de la vitesse optimale.

4.5. Diffusion uni-dimensionnelle possédant plusieurs probabilités invariantes

On considère la fonction $V : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par

$$V(x) = \begin{cases} (x - 3\text{sign}(x))^2 & \text{si } |x| \geq 3 \\ \frac{1}{72}(x^2 - 9)^2 & \text{si } |x| \leq 3. \end{cases}$$

On pose alors $b(x) := -V'(x)$ et $\sigma(x) := cx$ où $0 < c < 2$. On note X_t^x la diffusion brownienne issue de x associée aux fonctions lipschitziennes b et σ . Pour tout $x \in \mathbb{R}$, il existe une unique solution $(X_t^x)_{t \geq 0}$ issue de x .

Comme b et σ s'annulent toutes deux en 0, la fonction nulle est la seule solution issue de 0 et la masse de Dirac δ_0 est donc toujours une probabilité invariante pour la diffusion. En outre, aucune solution issue de $x \neq 0$ ne peut donc traverser 0. Il suffit donc de regarder le comportement de X_t^x pour $x \in \mathbb{R}_+^*$. Signalons d'autre part que pour l'ODE associée au coefficient de diffusion b , l'origine est un équilibre répulsif (± 3 étant eux attractifs pour l'ODE).

On vérifie que $AV(x) \sim -(4 - c^2)x^2$ quand $|x| \rightarrow +\infty$ et donc que $AV \leq \beta - (4 - c^2)V$ pour une constante réelle β si bien que l'hypothèse de Lyapunov (8) du théorème 2 est vérifiée (et *fortiori* son analogue du Th. 1).

Esquissons l'étude qualitative de cette diffusion en fonction du paramètre c . Pour ce faire, nous allons nous appuyer sur la classification de Feller qui permet, en dimension 1, de décrire le comportement d'une diffusion au voisinage de points-barrière et entre ceux-ci à partir de tests simples fondés sur deux fonctions associées à ses coefficients : la *fonction d'échelle* et la *mesure de vitesse* (cf. ci-après). Pour un exposé de cette classification, on pourra consulter [26] (pp. 342-353), et [27] (pp. 220-241).

Cas $1 < c < 2$. La formule d'Itô appliquée à la fonction \ln conduit à

$$\frac{\ln X_t^x}{t} = \frac{\ln x}{t} - \frac{1}{t} \int_0^t \left(\frac{c^2}{2} + g(X_s^x) \right) ds + c \frac{W_t}{t} \tag{16}$$

où $g(x) := V'(x)/x$ est une fonction continue bornée sur \mathbb{R} , atteignant son minimum $-1/2$ en 0. On vérifie alors que $\limsup_{t \rightarrow +\infty} (\ln X_t^x)/t \leq -\frac{c^2 - 1}{2} < 0$, donc X_t^x converge *p.s.* vers 0 (à une vitesse exponentielle, y compris si $c \geq 2$ d'ailleurs).

Cas $c = 1$. On vérifie à la main que la fonction d'échelle de la diffusion

$$p(x) := \int_3^x \exp \left(-2 \int_3^y \frac{b}{\sigma^2}(u) du \right) dy \longrightarrow \pm \infty \quad \text{selon que } x \rightarrow +\infty \text{ ou } 0^+.$$

D'après la classification de Feller, X_t^x vit dans \mathbb{R}_+^* et $\limsup_{t \rightarrow +\infty} X_t^x = +\infty$, $\liminf_{t \rightarrow +\infty} X_t^x = 0$ (en particulier X_t^x ne converge pas vers 0 *p.s.*). D'autre part, la mesure de vitesse de la diffusion, $m(dx) := \frac{2dx}{p'(x)\sigma^2(x)}$, n'est pas finie sur \mathbb{R}_+^* ce qui interdit l'existence d'une probabilité invariante portée par \mathbb{R}_+^* (cf. [27]). La masse de Dirac δ_0 est donc la seule probabilité invariante du système et le théorème 2 entraîne que, pour tout $x \in \mathbb{R}_+^*$,

$$\mathbb{P}_x(d\omega)\text{-}p.s., \quad \frac{1}{t} \int_0^t \delta_{X_s(\omega)} ds \xrightarrow{(\mathbb{R}^d)} \delta_0.$$

Par symétrie ceci s'étend aussitôt à tout $x \in \mathbb{R}$.

Cas $0 < c < 1$. Dans ce cas, $p(0^+) = -\infty, p((+\infty)_-) = +\infty$ et la mesure de vitesse $m(dx)$ est finie sur \mathbb{R}_+^* . La diffusion admet donc une (unique) probabilité invariante $\nu_+(dx)$ portée par \mathbb{R}_+^* obtenue par normalisation de la mesure de vitesse, soit

$$\nu_+(dx) = C \varphi_+(x) dx \quad \text{où} \quad \varphi_+(x) = \exp \left(-\frac{2}{c^2} \int_3^x g(u) du \right) / x^2.$$

On vérifie que φ_+ a une branche infinie en 0 si $c > 1/\sqrt{2}$, une limite finie non nulle si $c = 1/\sqrt{2}$ et une limite nulle si $c < 1/\sqrt{2}$. (La fonction symétrique $\varphi_-(x) = \varphi_+(-x)$ définit de même l'unique probabilité invariante sur \mathbb{R}_-^* .)

D'autre part, en réécrivant l'équation (16) sous la forme

$$\ln X_t^x = \ln x - \int_0^t G(\ln X_s^x) ds + cW_t \quad \text{où } G \text{ est une fonction bornée,}$$

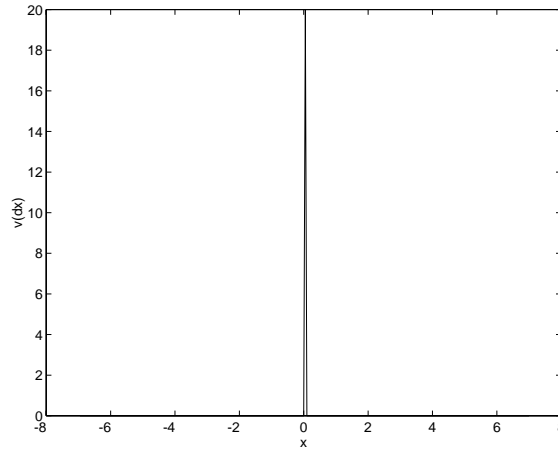


FIGURE 3. $c = 1$.

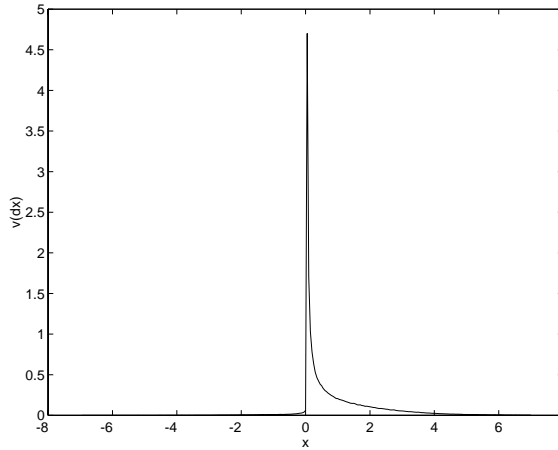


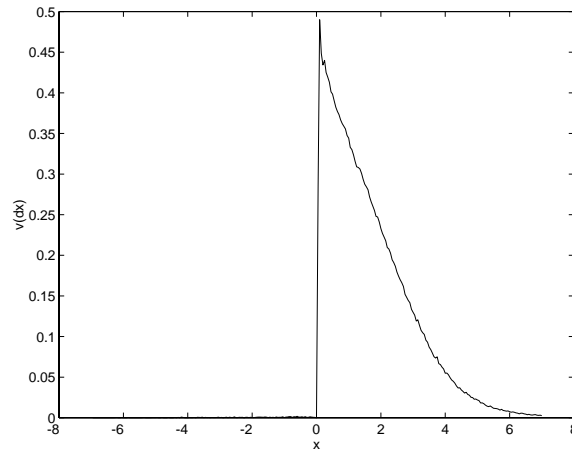
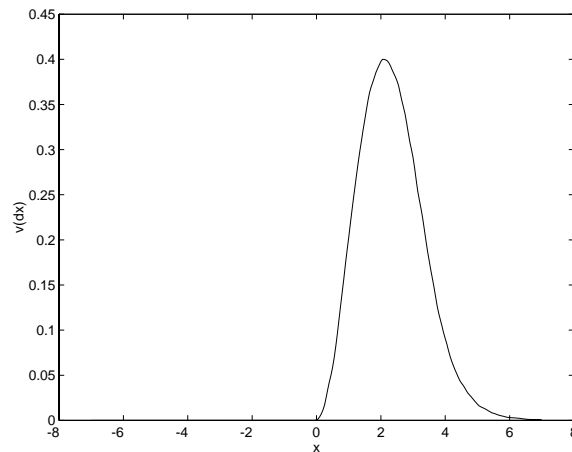
FIGURE 4. $c = 0,85$.

il découle du théorème de Girsanov que $\mathcal{L}(\ln X_t^x)$ est équivalente à la mesure de Lebesgue sur \mathbb{R} . Partant $\mathcal{L}(X_t^x)$ est équivalente à la mesure de Lebesgue sur \mathbb{R}_+^* , donc à ν_+ , pour tout $x \in \mathbb{R}_+^*$. En adaptant à \mathbb{R}_+ la démonstration du point (b) du théorème 1, on montre alors que pour tout $x \in \mathbb{R}_+^*$,

$$\mathbb{P}_x(d\omega)\text{-}p.s., \quad \frac{1}{t} \int_0^t \delta_{X_s^x(\omega)} ds \xrightarrow{(\mathbb{R}^d)} \nu_+.$$

De même, pour tout $x \in \mathbb{R}_-^*$, $\frac{1}{t} \int_0^t \delta_{X_s^x(\omega)} ds \xrightarrow{(\mathbb{R}^d)} \nu_-$.

À la différence de la diffusion elle-même, son schéma d'Euler est susceptible de traverser 0 à horizon fini. Les valeurs d'adhérence de l'algorithme sont donc *a priori* des combinaisons convexes de ν_+ , ν_- et δ_0 . Cependant, on peut espérer que, comme dans le cas des algorithmes stochastiques classiques, les mesures empiriques discrètes ne “visualisent” que les probabilités invariantes “intéressantes”, c'est-à-dire celles apparaissant comme limite des mesures empiriques à temps continu $\frac{1}{t} \int_0^t \delta_{X_s^x(\omega)} ds$, en l'occurrence, une probabilités invariante extrême.

FIGURE 5. $c = 1/\sqrt{2}$.FIGURE 6. $c = 0,5$.

Les simulations ci-dessous représentent les densités φ_+ (ou la masse de Dirac en 0) pour quatre valeurs de c : 1, 0,85, $\frac{1}{\sqrt{2}}$ et 0,5. Elles ont été obtenues avec $n = 10^6$ itérations du schéma d'Euler, avec des pas lentement décroissants à la vitesse $\theta = 0,30$ (*cf.* paragraphe précédent) et des poids égaux aux pas, sans accélération de Romberg. La valeur ponctuelle de la densité a été approchée en 280 points équirépartis entre -7 et 7 , soit 20 points par unité (un noyau de convolution gaussien a été utilisé pour les deux plus petites valeurs de c). La durée d'une simulation fournissant l'une des figures aini paramétrée est de l'ordre de 2 mn (programmation Matlab sur Powerbook G4, cadencé à 500 MHz).

On constate les phénomènes suivants :

- Lorsqu'il n'y a pas convergence vers δ_0 (Figs. 4, 5 et 6), l'algorithme semble "choisir son camp" ce qui conduit, comme pour la mesure empirique de la diffusion initiale, à des mesures invariantes portées par \mathbb{R}_+ ou \mathbb{R}_- . En revanche, contrairement à ce que les figures présentées pourraient suggérer, ce camp n'est pas forcément celui de la valeur initiale ; ceci dépend surtout de la taille du pas en début de simulation et de la valeur initiale.

- Le cas $c = 1$ (Fig. 3) est représentatif de la visualisation d'une masse de Dirac en 0. Pour autant, s'il est clair que, lorsque $c \leq 1/\sqrt{2}$, celle-ci n'intervient pas dans la probabilité limite, il est plus hasardeux de trancher à partir des seules simulations lorsque $c > 1/\sqrt{2}$ puisque la densité φ_+ a une branche infinie en 0. On peut néanmoins penser que l'algorithme converge vers les mêmes cibles que la mesure empirique de la diffusion.
- Globalement, le comportement de l'algorithme en présence de probabilités invariantes multiples semble donc satisfaisant. Un tel jugement devra néanmoins être confirmé par des simulations multi-dimensionnelles, la dimension 1 jouant clairement un rôle particulier dans le problème *via* le théorème des valeurs intermédiaires.

Avertissement. Attention à l'échelle – variable – des ordonnées dans la lecture des figures.

ANNEXE A

De l'identité $\theta_t^{-1}(A) = A$ on déduit que $A \in \sigma(X_s, s \geq t)$ pour tout $t \in \mathbb{R}_+$. A est donc dans la tribu asymptotique du processus. On considère $M_t := \mathbb{E}_{\nu^*}(\mathbf{1}_A/\mathcal{F}_t)$. M_t est une martingale bornée donc *p.s.* convergente vers $\mathbb{E}_{\nu^*}(\mathbf{1}_A/\mathcal{F}_\infty) = \mathbf{1}_A$ puisque \mathcal{F}_∞ contient la tribu asymptotique. D'autre part, l'invariance de A et la propriété de Markov entraînent que

$$M_t = \mathbb{E}_{\nu^*}(\mathbf{1}_A/\mathcal{F}_t) = \mathbb{E}_{\nu^*}(\mathbf{1}_A \circ \theta_t/\mathcal{F}_t) = \mathbb{E}_{X_t}(\mathbf{1}_A) \quad \mathbb{P}_{\nu^*}\text{-}p.s.$$

La martingale $(M_t)_{t \geq 0}$ s'écrit donc $M_t = \Psi(X_t)$ où $\Psi(x) := \mathbb{P}_x(A)$. La fonction Ψ est harmonique pour le semi-groupe P_t . D'autre part $\Psi(X_n)$ converge $\mathbb{P}_{\nu^*}\text{-}p.s.$, donc en loi, vers $\mathbf{1}_A$ alors même que ν^* étant une probabilité invariante, $\Psi(X_0) \in \{0, 1\}$ *p.s.* Soit $B := \{\Psi = 1\}$; ainsi $\Psi = \mathbf{1}_B$ si bien que $P_t(\cdot, B) = \mathbf{1}_B$ par harmonicité de Ψ . Si $\nu^*(B) \neq 0$ et 1, alors les probabilités conditionnelles $\nu^*(\cdot/B)$ et $\nu^*(\cdot/^cB)$ sont également invariantes. En effet, pour tout borélien C de \mathbb{R} ,

$$\begin{aligned} \nu^*(C \cap B) &= \nu^*P_t(C \cap B) = \nu^*(\mathbf{1}_B P_t(C \cap B)) + \underbrace{\nu^*(\mathbf{1}_{cB} P_t(C \cap B))}_{\leq \nu^*(\mathbf{1}_{cB} \mathbf{1}_B) = 0} \\ &= \nu^*(\mathbf{1}_B P_t(C \cap B)) = \nu^*(\mathbf{1}_B P_t(C \cap B)) + \underbrace{\nu^*(\mathbf{1}_B P_t(C \cap^c B))}_{=0} \\ &= \nu^*(\mathbf{1}_B P_t(C)). \end{aligned}$$

Or $\nu^* = \nu^*(B)\nu^*(\cdot/B) + \nu^*(^cB)\nu^*(\cdot/^cB)$ ce qui contredit le caractère extrémal de ν^* .

ANNEXE B

On localise l'intégrale stochastique d'où l'on tire, *via* la condition de rappel et le lemme de Fatou,

$$\sup_{s \in [0, t]} \mathbb{E}(V(X_s)) \leq C t.$$

On considère pour tout $K > 0$, une fonction $\varphi_K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, \mathcal{C}^2 , croissante vérifiant

$$\varphi_{|[0, K]}(v) = v, \quad \varphi_{|[K+1, +\infty[} = \varphi(K + 1), \quad 0 \leq \varphi' \leq 1.$$

On pose $V_K := \varphi_K \circ V$. On vérifie que $\nabla V_K = \varphi'(V)\nabla V$ est à support compact (donc bornée) et que

$$AV_K = \varphi'(V)AV + \varphi_K''(V)\nabla V^{\otimes 2}.$$

En prenant l'espérance dans la formule d'Itô appliquée à V_K et X_t^x il vient

$$P_t V_K(x) = V_K(x) + \int_0^t AV_K(x) ds$$

car le terme martingale locale est une vraie martingale. On vérifie ensuite que $V_K \uparrow V$ quand $K \rightarrow +\infty$ si bien que $P_t V_K \uparrow P_t V(x)$ *via* le théorème de Beppo Levi. Enfin $AV_K \rightarrow AV$ et $AV_K \leq C V$ d'après 8(iii). On conclut par convergence dominée.

ANNEXE C

Lorsque $a \neq 1$ on ne peut plus espérer établir que $\sup_t \mathbb{E}_x(V(X_t))$ soit fini. Soit $\varepsilon \in]0, a-1/2[$ et $\theta := 1-a/2-\varepsilon$. On repart de la formule de représentation

$$P_t V(x) = V(x) + \int_0^t P_s AV(x) ds. \quad (17)$$

L'inégalité de Jensen entraîne

$$\begin{aligned} P_t V^\theta(x) &\leq (P_t V)^\theta(x) = \left(V(x) + \int_0^t P_s AV(x) ds \right)^\theta \\ &\leq V^\theta(x) \left(1 + \int_0^t P_s V(x) ds / V(x) \right)^\theta \\ &\leq V^\theta(x) \left(1 + \theta \int_0^t P_s V(x) ds / V(x) \right) \\ &\leq V^\theta(x) + \theta V^{\theta-1}(x) \int_0^t P_s AV(x) ds \end{aligned}$$

si bien que,

$$AV^\theta = \lim_{t \rightarrow 0} \frac{P_t V^\theta - V^\theta}{t} \leq \theta V^{\theta-1} AV \leq \beta \theta V^{\theta-1} - \alpha \theta V^{a+\theta-1} \leq \tilde{\beta} - \tilde{\alpha} V^{a+\theta-1}$$

avec $\tilde{\alpha} := \theta \alpha$ et $\tilde{\beta} := \beta \theta v^{\theta-1}$.

Il vient alors, sachant que $\theta + \frac{a-1}{2} = \frac{1}{2} - \varepsilon$,

$$0 \leq V^\theta(X_t^x) = V^\theta(x) + \int_0^t AV^\theta(X_s^x) ds + \int_0^t \underbrace{((\nabla V^\theta)^* \sigma)(X_s^x)}_{=: O(V^{\frac{1}{2}-\varepsilon}(X_s^x))} dW_s,$$

d'où

$$\frac{1}{t} \int_0^t -AV^\theta(X_s^x) ds \leq \frac{V^\theta(x)}{t} + \frac{1}{t} \int_0^t O(V^{\frac{1}{2}-\varepsilon}(X_s^x)) dW_s.$$

Si l'on pose $M_t := \int_1^t \frac{O(V^{\frac{1}{2}-\varepsilon}(X_s^x))}{s} dW_s$, alors

$$\langle M \rangle_\infty = \int_1^t \frac{O(V^{1-2\varepsilon}(X_s^x))}{s^2} ds.$$

Or, il découle immédiatement de (17) et de l'hypothèse de Lyapunov que $P_t V(x) \leq V(x) + \beta t$. En conséquence, à nouveau *via* l'inégalité de Jensen, $P_s V^{1-2\varepsilon}(x) \leq (V(x) + \beta s)^{1-2\varepsilon}$. Par suite

$$\mathbb{E}\langle M \rangle_\infty = \int_1^{+\infty} \frac{O(P_s V^{1-2\varepsilon}(x))}{s^2} ds \leq C^{te} \int_1^{+\infty} \frac{ds}{s^{1+2\varepsilon}} < +\infty.$$

Ceci entraîne que M_t converge \mathbb{P}_x -*p.s.* sûrement vers M_∞ élément de L^2 . Le lemme de Kronecker permet alors d'écrire que $\frac{1}{t} \int_0^t O(V^{\frac{1}{2}-\varepsilon}(X_s)) dW_s \xrightarrow{p.s.} 0$. Il en découle immédiatement que

$$\limsup_t \frac{1}{t} \int_0^t -AV^\theta(X_s^x) ds \stackrel{p.s.}{\leq} 0 \quad i.e. \quad \limsup_t \frac{1}{t} \int_0^t V^\theta(X_s^x) ds \leq \frac{\tilde{\beta}}{\tilde{\alpha}} < +\infty.$$

Merci à Laurence Marsalle pour ses contributions multiples à la mise au point du manuscrit à partir des transparents des journées MAS 2000 de Rennes. Merci à Jean-Claude Fort et à Damien Lambertson, pour tout.

RÉFÉRENCES

- [1] S. Aida, S. Kusuoka et D.W. Stroock, On the support of Wiener functionals, dans *Asymptotic problems in Probability Theory: Wiener Functionals and Asymptotics*, édité par K.D. El Worthy et N. Ikeda. Longman Scient. and Tech., New-York, *Pitman Res. Notes Math. Ser.* **284** (1993) 3-34.
- [2] D.G. Aronson, Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.* **73** (1967) 890-903.
- [3] J.G. Attali, *Méthodes de stabilité pour les chaînes de Markov non féllériennes*, Thèse de l'Université Paris I (1999).
- [4] G. Basak et R. Bhattacharya, Stability in distributions for a class of singular diffusions. *Ann. Probab.* **20** (1992) 312-321.
- [5] G.K. Basak, I. Hu et C.-Z. Wei, Weak convergence of recursions. *Stochastic Process. Appl.* **68** (1997) 65-82.
- [6] M. Benaïm, Recursive Algorithms, Urn process and Chaining Number of Chain Recurrent sets. *Ergodic Theory Dynam. Systems* **18** (1997) 53-87.
- [7] M. Benaïm, Dynamics of Stochastic Approximation Algorithms, *Séminaire de Probabilités XXXIII*, édité par J. Azéma, M. Émery, M. Ledoux et M. Yor. Springer, *Lecture Notes in Math.* **1709** (1999) 1-68.
- [8] G. Ben Arous et R. Léandre, Décroissance exponentielle du noyau de la chaleur sur la diagonale (II). *Probab. Theory Related Fields* **90** (1991) 377-402.
- [9] A. Benveniste, M. Métivier et P. Priouret, *Algorithmes adaptatifs et approximations stochastiques*. Masson, Paris (1987) 367p.
- [10] S. Borovkov, *Ergodicity and Stability of Stochastic Processes*. Wiley Chichester (England), *Wiley Ser. Probab. Stat.* (1998) 585p.
- [11] C. Bouton, Approximation gaussienne d'algorithmes à dynamique markovienne. *Ann. Inst. H. Poincaré B* **24** (1988) 131-155.
- [12] O. Brandière et M. Dufflo, Les algorithmes stochastiques contournent-ils les pièges ? *Ann. Inst. H. Poincaré* **32** (1996) 395-477.
- [13] G.A. Brosamler, An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.* **104** (1988) 561-574.
- [14] I. Berkes, E. Csáki, A universal result in almost sure central limit theory. *Stochastic Process. Appl.* **94** (2001) 105-134.
- [15] F. Chaâbane, F. Maâouia et A. Touati, *Versions fortes associées aux théorèmes limites en loi pour les martingales vectorielles*. Pré-pub. de l'Université de Bizerte, Tunisie (1996).
- [16] S. Cheng et L. Peng, Almost sure convergence in extreme value theory. *Math. Nachr.* **190** (1998) 43-50.
- [17] M. Dufflo, *Random Iterative systems*. Springer, Berlin (1998).
- [18] N. Dunford et J.T. Schwartz, *Linear Operators*. Wiley-Interscience, New-York (1958).
- [19] S. Ethier et T. Kurtz, *Markov Processes, characterization and convergence*. Wiley, New-York, *Wiley Ser. Probab. Math. Statist.* (1986) 534p.
- [20] J.C. Fort et G. Pagès, Asymptotic behaviour of a Markov constant step stochastic algorithm. *SIAM J. Control Optim.* **37** (1999) 1456-1482.
- [21] J.C. Fort et G. Pagès, *Stochastic algorithms with non constant step: a.s. behaviour of weighted empirical measures*. Pré-pub. Université Paris 12 Val-de-Marne (1998, soumis).
- [22] A. Fisher, Convex invariant means and a pathwise central limit theorem. *Adv. Math.* **63** (1987) 213-246.
- [23] H. Ganidis, B. Roynette et F. Simonot, Convergence rate of some semi-groups to their invariant probability. *Stochastic Process. Appl.* **79** (1999) 243-264.
- [24] P. Hall et C.C. Heyde, *Martingale Limit Theory and its Application*. Academic Press, New-York (1980) 308p.
- [25] R.Z. Has'minskii, *Stochastic stability of differential equations*. Sijthoff & Noordhoff, Alphen aan den Rijn (The Netherlands) (1980) 344p.
- [26] I. Karatzas et S. Shreve, *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New-York (1988) (2nd Ed., 1992) 470p.

- [27] S. Karlin et H. Taylor, *A second course in stochastic processes*. Academic Press, New-York (1981) 542p.
- [28] Y. Kifer, *Random perturbations of Dynamical Systems*. Birkhäuser, *Progr. Probab. Statist.* (1988) 294p.
- [29] U. Krengel, *Ergodic Theorems. de Gruyter Stud. Math.* (1989) 357p.
- [30] H.J. Kushner et D.S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*. Springer, *Appl. Math. Sci.* **26** (1978) 261p.
- [31] H.J. Kushner, *Approximation and weak convergence methods for random processes and applications to stochastic system theory*. MIT Cambridge (1985).
- [32] H.J. Kushner et H. Huang, Rates of convergence for stochastic approximation type algorithms. *SIAM J. Control Optim.* **17** (1979) 607-617.
- [33] D. Lamberton et G. Pagès, Recursive computation of the invariant measure of a diffusion. *Bernoulli* (à paraître).
- [34] M.T. Lacey et W. Philip, A note on the almost sure central limit theorem. *Statist. Probab. Lett.* **9** (1990) 201-205.
- [35] S. Meyn et R. Tweedie, *Markov chains and Stochastic Stability*. Springer (1993) 550p.
- [36] M. Pelletier, Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.* **8** (1998) 10-44.
- [37] M. Pelletier, An almost sure central limit theorem for stochastic algorithms. *J. Multivariate Anal.* **71** (1999) 76-93.
- [38] M. Pelletier, Efficacité asymptotique presque sûre des algorithmes stochastiques moyennés. *C. R. Acad. Sci. Paris Série I* **323** (1996) 813-816 ; développé dans *Asymptotic almost sure efficiency of averaged stochastic algorithms* (soumis).
- [39] B.T. Polyak, New Stochastic Approximation type procedures. *Avtomat. i Telemekh.* **7** (1990), in Russian, *Automat. Remote Control* **51** (1990) 107-118.
- [40] D. Revuz et M. Yor, *Continuous martingales and Brownian Motion*, 2nd Ed. Springer, Berlin (1991) 557p.
- [41] D. Ruppert, *Efficient estimators from a slowly convergent Robbins-Monro Process*, Technical Report, School of Operations Research and Industrial, Engineering. Cornell University, Ithaca, NY, No. 781 (1985).
- [42] P. Schatte, On strong versions of the central limit theorem. *Math. Nachr.* **137** (1988) 249-256.
- [43] D.W. Stroock, *Probability Theory: An analytic view*. Cambridge University Press (revised edition, 1994) 512p.
- [44] D. Talay, Second order discretization of stochastic differential systems for the computation of the invariant law. *Stochastics Stochastics Rep.* **29** (1990) 13-36.
- [45] D. Talay et L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.* **8** (1990) 94-120.
- [46] A. Touati, *Sur les versions fortes du théorème de la limite centrale*. Pré-pub. de l'Université de Marne-la-Vallée (1995).