Vitesse de convergence vers le régime stationnaire pour certains processus de Markov déterministes par morceaux

Michel Benaïm, Stéphane Le Borgne,
Florent Malrieu et Pierre-André Zitt

28 mai 2013

Résumé

Nous étudions une classe de processus de Markov déterministes par morceaux à espaces d'états $\mathbb{R}^d \times E$ où E est un ensemble fini. La partie continue évolue suivant une équation différentielle déterminée par la partie discrète E et qui est modifiée à des temps aléatoires. Le taux déterminant les probabilités d'occurrence de ces changements d'équation différentielle peuvent dépendre de la position du processus. Nous établissons que, sous certaines conditions de régularité sur les taux de saut et de stabilité des équations différentielles impliquées, la mesure sur $\mathbb{R}^d$ définie comme la loi de la coordonnées continue du processus converge exponentiellement vite vers l'unique mesure stationnaire en distance de Wasserstein.

Keywords. Couplage ; Ergodicité ; Équations différentielles ; Processus de Markoff déterministe par morceaux ; distance de Wasserstein.

AMS-MSC. 60J75 ; 60J25 ; 93E15 ; 34D23

1 Introduction

Les processus de Markov déterministes par morceaux (abrégé dans la suite en PMDM) sont utiles dans différents domaines (biologie moléculaire [29], gestion de stocks [6], du trafique internet [18, 22, 23], activité neuronale [27, 7], modèles de croissance de populations [19]...). Un processus de Markov est un PMDM si son caractère aléatoire ne tient qu'à un mécanisme de saut : en particulier un PMDM n'a pas de comportement diffusif. Davis (see [13, 14]) est le premier à avoir étudié cette classe de processus en tant que telle. Plusieurs travaux [10, 17, 11] sont consacrés aux comportements des PMDM en temps longs (existence d'une mesure de probabilité invariante, récurrence au sens de Harris, convergence à vitesse exponentielle vers le régime stationnaire...). On peut montrer([12]) que le comportement d'un PMDM est lié au processus en temps discret constitué des positions aux temps de sauts associé à un processus de poisson indépendant. Cette approche ne semble pas fournir d'information quantitative sur la vitesse de convergence à l'équilibre. De récente travaux ont établi de telles estimations pour certains exemples ([9, 20, 4]) ou pour les chaîne de Markov en temps continu ([8]).

Ici, nous étudions le comportement en temps long d'une classe de PMDM (voir [29, 7]). Ce sont des PMDM à valeurs dans $\mathbb{R}^d \times E$ où E un ensemble fini. La première coordonnée évolue dans $\mathbb{R}^d$ suivant les trajectoires d'un champs de vecteurs régulier défini par la seconde coordonnée. La deuxième coordonnée saute d'une valeur à l'autre avec des taux qui peuvent dépendre de la première. On peut rapprocher ces systèmes de ce qu'on appelle, en temps discret, l'itération de fonctions aléatoires (sur ce sujet voir [16]).

Soit E un ensemble fini, $(\lambda(\cdot, i))_{i \in E}$ n fonctions positives ou nulles sur $\mathbb{R}^d$, P une matrice stochastique irréductible et, pour tout $i \in E$, $F^i : \mathbb{R}^d \mapsto \mathbb{R}^d$ un champ de vecteur C^∞ tel que l'équation différentielle ordinaire

$$\begin{cases} x'_t = F^i(x_t), \quad t > 0; \\ x_0 = x, \end{cases}$$

a une solution unique définie pour tout temps et toute condition initiale $x \in \mathbb{R}^d$, notée $t \mapsto \varphi^i_t(x)$. Nous considérons le processus de Markov

$$(Z_t)_{t \geqslant 0} = ((X_t, I_t))_{t \geqslant 0} \text{ on } \mathbb{R}^d \times E$$

défini par son générateur infinitésimal L comme suit :

$$Lf(x,i) = \left\langle F^i(x), \nabla_x f(x,i) \right\rangle + \lambda(x,i) \sum_{j \in E} P(i,j)(f(x,j) - f(x,i)) \tag{1}$$

pour toute fonction C^∞, $f : \mathbb{R}^d \times E \to \mathbb{R}$ (voir [14] pour plus de détails sur L). Les trajectoires de ce processus peuvent être décrites de la façon suivante. Supposons qu'à l'instant 0 on ait $(X_0, I_0) = (x, i) \in \mathbb{R}^d \times E$. Jusqu'à l'instant du premier saut T_1 de la composante discrète I, la première coordonnée X est régie par le champ de vecteur F^i autrement dit $X_t = \varphi^i_t(x)$. L'instant T_1 est donné par :

$$T_1 = \inf \left\{ t > 0 \ : \ \int_0^t \lambda(X_s, i)\, ds \geqslant E_1 \right\},$$

où E_1 est une variable exponentielle de paramètre 1. Comme les trajectoires de X sont déterministes entre les instants de saut I, c'est E_1 qui donne à T_1 son caractère aléatoire

$$T_1 = \inf \left\{ t > 0 \ : \ \int_0^t \lambda(\varphi^i_s(x), i)\, ds \geqslant E_1 \right\}.$$

Remarque 1.1. *La probabilité que le saut ne se produise jamais,* $\mathbb{P}_{(x,i)}(T_1 = +\infty)$, *peut être positive si et seulement si*

$$\int_0^{+\infty} \lambda(\varphi^i_s(x), i)\, ds < +\infty.$$

Par conséquent si $\underline{\lambda} := \inf_{(x,i)} \lambda(x,i) > 0$ *alors la deuxième coordonnée* I *saute une infinité de fois.*

À l'instant T_1, la coordonnée I change de position suivant la loi $P(i, \cdot)$ et le champ de vecteur déterminant l'évolution de la coordonnée X change...

Remarque 1.2. *En général, la coordonnée I n'est pas un processus de Markov car elle dépend de X.*

En construisant des couplages explicites, nous allons donner des informations quantitatives sur le comportement en temps long du processus défini par (1). Pour la première composante X, nous utiliserons la distance de Wasserstein car la distance en variation totale n'est pas adaptée ici. Rappelons que, pour tout $p \geqslant 1$, la distance de Wasserstein W_p entre deux probabilités μ et $\tilde{\mu}$ sur $\mathbb{R}^d$ ayant des moments d'ordre p finis est donnée par

$$W_p(\mu, \tilde{\mu}) = \left(\inf_{\Pi} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - \tilde{x}|^p \, \Pi(dx, d\tilde{x}) \right)^{1/p} \tag{2}$$

où la borne inférieure est prise sur l'ensemble des mesures de probabilités sur $\mathbb{R}^d \times \mathbb{R}^d$ dont les marginales sont μ et $\tilde{\mu}$ (de telles mesures sont appelées couplages de μ et $\tilde{\mu}$). Pout tout $p \geqslant 1$, la convergence en distance W_p de Wasserstein est équivalente à la convergence faible plus celle des moments jusqu'au degré p. Il faut remarquer que deux mesures de probabilités peuvent être proches l'une de l'autre au sens de W_p tout en étant singulières : si par exemple $\mu = \delta_0$ et $\tilde{\mu} = \delta_\varepsilon$, alors

$$W_p(\mu, \tilde{\mu}) = \varepsilon \quad \text{and} \quad \|\mu - \tilde{\mu}\|_{\mathrm{TV}} = 1.$$

On trouvera plus d'informations sur les distances de Wasserstein dans [28, 32].

Pour évaluer la distance de Wasserstein, il n'est pas nécessaire de disposer d'information sur le support de la mesure invariante (l'ensemble des points récurrents). Cet ensemble peut être difficile à décrire et, de plus, il se peut que la loi de X_t et la mesure invariante soient singulières pour tout t. L'exemple suivant l'illustre :

$$E = \{0, 1\}, \quad \lambda(x, i) = 1, \quad F^i(x) = -(x - ia) \quad \text{with } a = (1, 0).$$

Le processus (X, I) est ergodique et la première marginale μ de sa mesure invariante est portée par le segment $\{\rho x \ ; \ \rho \in [0, 1]\}$ (dans [6] les auteurs montrent que μ est une loi Beta). Si $X_0 = (0, 1)$alors la loi de X_t et la mesure invariante sont singulères ; en particulier, $\|\mathcal{L}(X_t) - \mu\|_{\mathrm{TV}}$ vaut 1 pour tout $t \geqslant 0$, en revanche nous verrons plus bas que$W_p(\mathcal{L}(X_t), \mu)$ tend vers 0 exponentiellement vite. Obtenir une vitesse pour la distance en variation totale nécessite donc d'autres hypothèses (on peut penser aux hypothèses portant sur les crochets de champs de vecteurs considérées dans [2] ou [5] ; il est à noter que de telles hypothèses suffisent à assurer un résultat de convergence vers l'unique mesure stationnaire mais il nous semble nécessaire d'y ajouter d'autres conditions pour obtenir des informations sur la vitesse de convergence).

Dans ce travail, nous établissons l'existence de taux de croissance exponentiels pour dans deux situations. Premièrement, si les taux de sauts sur I ne dépendent pas de X, alors une condition de contraction en moyenne pour les champs de vecteurs $(F^i)_{i \in E}$ nous suffit. Deuxièmement, si les taux de sauts sur I dépendent de X, alors nous considérons le cas où cette dépendance et lipschitzienne et nous faisons une hypothèse plus forte sur les champs $(F^i)_{i \in E}$: ils contractent tous exponentiellement vite la distance euclidienne.

Dans la suite, μ_t désigne la première marginale de la loi de $Z_t = (X_t, I_t)$.

1.1 Taux de sauts constants

Lorsque les taux de saut ne dépendent de X, $(I_t)_{t\geqslant 0}$ est un processus de Markov à espace d'états fini E et $(X_s)_{0\leqslant s\leqslant t}$ est une fonction (déterministe) de $(I_s)_{0\leqslant s\leqslant t}$. On dispose alors de nombreux résultats aussi bien en temps discret (voir [26, 31, 21, 25, 1]) qu'en temps continu (voir [24, 15, 3]). On trouvera dans [5] un exemple de PMDM construit à partir de flots stables ayant un comportement explosif (lorsque le taux de saut est suffisamment élevé).

Hypothèse 1.3. *Les taux de sauts $(\lambda(\cdot,i))_{i\in E}$ ne dépendent pas de x et I est un processus de Markov irréductible sur E dont la mesure de probabilité invariante est notée ν.*

Hypothèse 1.4. *Pour tout $i\in E$, il existe $\alpha(i)\in\mathbb{R}$ tel que,*

$$\left\langle x-\tilde{x}, F^i(x)-F^i(\tilde{x})\right\rangle \leqslant -\alpha(i)|x-\tilde{x}|^2, \quad x,\tilde{x}\in\mathbb{R}^d,$$

et

$$\sum_{i\in E}\alpha(i)\nu(i)>0$$

où ν est définie dans l'hypothèse 1.3 précédente.

Alors le processus X est borné en norme L^p, pour un certain p.

Lemme 1.5. *Sous les hypothèses 1.3 et 1.4, il existe $\kappa>0$ tel que, pour tout $q<\kappa$, la fonction $t\mapsto\mathbb{E}(|X_t|^q)$ est bornée dès que $\mathbb{E}(|X_0|^q)$ est finie. Il existe $M(q,m)$ tel que*

$$\sup_{t\geqslant 0}\mathbb{E}(|X_t|^q)\leqslant M(q,m),$$

dès que $\mathbb{E}(|X_0|^q)\leqslant m$.

Le comportement en temps long est précisé par le résultat suivant.

Théorème 1.6. *Supposons sue les hypothèses 1.3 et 1.4 sont satisfaites. Soit $p<q<\kappa$. Notons s le conjugué de q : $q^{-1}+s^{-1}=1$. Si μ_0 et $\tilde{\mu}_0$ ont un moment d'ordre q plus petit que m, alors*

$$W_p(\mu_t,\tilde{\mu}_t)\leqslant 2^{p+1}M(q,m)^{p/q}C_2(p)\exp\left(-\frac{\eta_p}{1+s\eta_p/\rho}t\right),$$

où ρ et η_p sont des constantes positives dépendant seulement de la chaîne de Markov I.

Les constantes ρ, η_p et $C_2(p)$ sont définies en (6), (7) et (8).

Corollaire 1.7. *Sous les hypothèses 1.3 et 1.4, le processus X admet une unique mesure invariante μ et*

$$W_p(\mu_t,\mu)\leqslant 2^{p+1}M(q,m)^{p/q}C_2(p)\exp\left(-\frac{\eta_p}{1+s\eta_p/\rho}t\right).$$

1.2 Taux de saut non constants

Venons en maintenant au cas où les taux de saut dépendent de X. Pour les applications il est intéressant de faire dépendre aussi la matrice P de X. Nous modifions donc légèrement la définition de notre processus et considérons le processus défini par le générateur infinitésimal suivant :

$$Lf(x,i) = \left\langle F^i(x), \nabla_x f(x,i) \right\rangle + \sum_{j\in E} a(x,i,j)(f(x,j) - f(x,i)) \tag{3}$$

pour les fonctions régulières $f: \mathbb{R}^d \times E \to \mathbb{R}$. Nous allons supposer que les fonctions $a(x,i,j)$ sont régulières et que les champs de vecteurs F^i ont des propriétés de contraction fortes (en particulier ils auront un unique point fixe).

Hypothèse 1.8. *Il existe $\underline{a} > 0$ et $\kappa > 0$ tels que, pour tout $x, \tilde{x} \in \mathbb{R}^d$ et $i, j \in E$,*

$$a(x,i,j) \geqslant \underline{a} \quad \textit{and} \quad \sum_{j\in E} |a(x,i,j) - a(\tilde{x},i,j)| \leqslant \kappa |x - \tilde{x}|,$$

La minoration assure que coordonnée I de Z change suffisamment souvent (de telle sorte que les deuxièmes coordonnées de deux copies indépendantes de Z coïncident suffisamment souvent). Dans certaines condition on eut affaiblir cette hypothèse : par exemple si $a(x,i,j) = \lambda(x,i)P_{ij}$ comme dans (1), une fonction lipschitzienne λ minorée par une constante positive et une matrice apériodique P sont suffisants pour obtenir des vitesses de convergence analogues à ceux donnés plus bas.

Hypothèse 1.9. *Il existe $\alpha > 0$ tel que,*

$$\left\langle x - \tilde{x}, F^i(x) - F^i(\tilde{x}) \right\rangle \leqslant -\alpha |x - \tilde{x}|^2, \quad x, \tilde{x} \in \mathbb{R}^d,\ i \in E. \tag{4}$$

L'hypothèse 1.9 assure que, pour tout $i \in E$,

$$\left| \varphi_t^i(x) - \varphi_t^i(\tilde{x}) \right| \leqslant e^{-\alpha t} |x - \tilde{x}|, \quad x, \tilde{x} \in \mathbb{R}^d.$$

En particulier le champ de vecteurs F^i a un unique point stationnaire $\sigma(i) \in \mathbb{R}^d$, stable : pour tout $x \in \mathbb{R}^d$,

$$\left| \varphi_t^i(x) - \sigma(i) \right| \leqslant e^{-\alpha t} |x - \sigma(i)|.$$

Cela entraîne que X est confiné dans un compact et que les fonctions continues $a(\cdot, i, j)$ sont bornées le long des trajectoires de X. Plus précisément on a

Lemme 1.10. *Sous les hypothèses 1.8 et 1.9, le processus Z ne peut quitter l'ensemble compact $\bar{B}(0,r) \times E$ où $\bar{B}(0,r)$ est la boule fermée de centre $0 \in \mathbb{R}^d$ et de rayon r*

$$r = \frac{\max_{i\in E} |F^i(0)|}{\alpha}. \tag{5}$$

De plus, si $|X_0| > r$ alors

$$\left(|X_t|^2 - r^2 \right)^+ \leqslant e^{-\alpha t} \left(|X_0|^2 - r^2 \right)^+.$$

En particulier le support d'une mesure invariante est inclus dans $\bar{B}(0,r)$.

La vitesse de convergence que nous obtenons sous les hypothèses 1.8 et 1.9 est décrite dans le résultat suivant.

Théorème 1.11. *Sous les hypothèses 1.8 et 1.9 et si les supports de μ_0 et $\tilde{\mu}_0$ sont inclus dans $\bar{B}(0,r)$ où r est défini par (5), alors il existe des constantes positives c, α, p telles que*

$$W_1(\mu_t, \tilde{\mu}_t) \leqslant 2r(1+ct)\exp\left(-\frac{\alpha}{1+\alpha/\gamma}t\right).$$

Les constantes obtenues dans la démonstrations sont les suivantes

$$\gamma = \frac{(\alpha+b) - \sqrt{(\alpha+b)^2 - 4bp\alpha}}{2} \quad \text{and} \quad c = \frac{\alpha}{\alpha+\gamma}\frac{ep\alpha b}{\sqrt{(\alpha+b)^2-4bp\alpha}},$$

avec $p = e^{-2r\kappa/\alpha}$ et $e = \exp(1)$ et où b dépend du temps de coalescence de deux processus sur E, indépendants, définis comme deuxièmes coordonnées de deux copies indépendantes de Z. Nous obtenons ce résultat en comparant le processus à un autre qui peut passer instantanément (mais de moins en moins souvent) de petites à de grandes valeurs (de l'ordre de $2r$). Cela peut paraître assez grossier. Nous pensons qu'il n'est pas possible de faire beaucoup mieux en général : si l'un des deux flots est très fortement contractant, deux trajectoires peuvent s'éloigner très rapidement l'une de l'autre.

Corollaire 1.12. *Sous les hypothèses 1.8 et 1.9, le processus X admet une unique mesure invariante μ et*

$$W_1(\mu_t, \mu) \leqslant 2r(1+ct)\exp\left(-\frac{\alpha}{1+\alpha/\gamma}t\right).$$

La section 2 est consacrée à la démonstration du théorème 1.6, la section 3 à celle du théorème 1.11.

2 Taux de saut constants

L'hypothèse 1.3 assure que $(I_t)_{t\geqslant 0}$ est un processus de Markov irreductible sur l'espace E. Son générateur infinitésimal A est la matrice définie par $A(i,i) = -\lambda(i)$ et $A(i,j) = \lambda(i)P(i,j)$ pour $i \neq j$. Notons ν son unique mesure de probabilité invariante unique. Comme E est fini il est facile de construire un couplage coalescent de deux processus I partant de points différents.

Lemme 2.1 ([30])**.** *Sous l'hypothèse 1.3 il existe $\rho > 0$ tel que, pour tout $i, j \in E$,*

$$\mathbb{P}(T > t | I_0 = i,\ \tilde{I}_0 = j) \leqslant e^{-\rho t} \tag{6}$$

où $(I_t)_{t\geqslant 0}$ et $(\tilde{I}_t)_{t\geqslant 0}$ sont deux processus de Markov indépendants de générateur infinitésimal A de positions initiales respectives i et j, et $T = \inf\left\{t \geqslant 0 \,:\, I_t = \tilde{I}_t\right\}$ est le temps de première intersection.

Remarque 2.2. *Si $E = \{1, 2\}$, alors le temps de première intersection suit une loi exponentielle de paramètre $\lambda(1) + \lambda(2)$ et l'équation* (6) *est satisfaite avec $\rho = \lambda(1) + \lambda(2)$.*

La démonstration du théorème 1.6 comporte deux étapes. Dans un premier temps nous donnons un couplage de deux copies du processus de positions initiales (x, i) et $(\tilde{x}, i)$ et en déduisons une majoration. Ensuite nous utilisons cette majoration et le lemme lemma 2.1 pour traiter le cas général.

2.1 Moments

Nous démontrons le lemme 1.5 et obtenons une estimation de la norme L^p de $|X_t|$. Pour tout $p \geqslant 2$ et $\varepsilon > 0$,

$$\begin{aligned}\frac{d}{dt}|X_t|^p &= p|X_t|^{p-2}\Big\langle X_t, F^{I_t}(X_t)\Big\rangle \\ &= p|X_t|^{p-2}\Big\langle X_t, F^{I_t}(X_t) - F^{I_t}(0)\Big\rangle + p|X_t|^{p-2}\Big\langle X_t, F^{I_t}(0)\Big\rangle \\ &\leqslant -(p\alpha(I_t) - \varepsilon)|X_t|^p + C(p, \varepsilon).\end{aligned}$$

Du lemme de Gronwall on déduit :

$$\mathbb{E}(|X_t|^p) \leqslant C(p, \varepsilon)\int_0^t \mathbb{E}\Big(e^{-\int_s^t (p\alpha(I_u)-\varepsilon)\,du}\Big)\,ds + \mathbb{E}(|X_0|^p)\mathbb{E}\Big(e^{-\int_0^t (p\alpha(I_u)-\varepsilon)\,du}\Big).$$

Le membre de droite dépend uniquement de $\mathbb{E}(|X_0|^p)$ et $(I_s)_{0\leqslant s\leqslant t}$. Par conséquent il suffit d'étudier le comportement de $e(p, t)$ défini pour tout $t \geqslant 0$ par

$$e(p, t) = \max_{i\in E} \mathbb{E}_i\Big(\exp\Big(-\int_0^t p\alpha(I_u)\,du\Big)\Big).$$

Nous suivons le raisonnement déveoppé dans [3]. Désignons par A_p la matrice $A-pB$ où A est le générateur infinitésimal de I et B est la matrice diagonale $(\alpha(1), \ldots, \alpha(n))$ et par η_p la quantité

$$\eta_p := -\max_{\gamma\in \mathrm{Spec}(A_p)} \mathrm{Re}\ \gamma. \tag{7}$$

Le nombre η_p caractérise le comportement de $e(p, t)$ en temps long : pour tout $p > 0$, il existe $0 < C_1(p) < 1 < C_2(p) < +\infty$ tels que, pour tout $t > 0$, on ait

$$C_1(p)e^{-\eta_p t} \leqslant e(p, t) \leqslant C_2(p)e^{-\eta_p t}. \tag{8}$$

D'autre part on a la dichotomie suivante :

1. if $\underline{\alpha} \geqslant 0$, alors $\eta_p > 0$ pour tout $p > 0$,
2. if $\underline{\alpha} < 0$, il existe $\kappa \in (0, \min\{A_{ii}/\alpha(i) : \alpha(i) < 0\})$ tel que $\eta_p > 0$ pour $p < \kappa$ et $\eta_p < 0$ pour $p > \kappa$.

Nous renvoyons à [3] pour plus de détails.

Corollaire 2.3. *Si $p < \kappa$ alors $t \mapsto \mathbb{E}(|X_t|^p)$ est bornée dès que $\mathbb{E}(|X_0|^p)$ est finie.*

2.2 Vitesse de convergence

Nous allons maintenant majorer la distance de Wasserstein $W_p(\mu, \mu_t)$ pour un $p < \kappa$. Pour cela nous construisons un couplage de deux copies de notre processus partant de lois initiales distinctes. Supposons d'abord que les lois initiales sont masses de Dirac en (x, i) il $(\tilde{x}, i)$. Comme les taux de saut sont indépendants de X, on peut choisir un couplage de (X, I) et $(\tilde{X}, \tilde{I})$ pour lequel I et $\tilde{I}$ sont égaux. Par conséquent, pout tout $p \geqslant 2$,

$$\begin{aligned}\frac{d}{dt}|X_t - \tilde{X}_t|^p &= p|X_t - \tilde{X}_t|^{p-2}\Big\langle X_t - \tilde{X}_t, F^{I_t}(X_t) - F^{I_t}(\tilde{X}_t)\Big\rangle \\ &\leqslant -p\alpha(I_t)|X_t - \tilde{X}_t|^p;\end{aligned}$$

cela entraîne

$$\begin{aligned}\mathbb{E}\Big(|X_t - \tilde{X}_t|^p\Big) &\leqslant \mathbb{E}_i\Big(\exp\Big(-p\int_0^t \alpha(I_s)\,ds\Big)\Big)|x - \tilde{x}|^p \\ &\leqslant e^{-\eta_p t}|x - \tilde{x}|^p.\end{aligned}$$

Traitons maintenant le cas général. Choisissons (x, i) et $(\tilde{x}, j)$ dans $\mathbb{R}^d \times E$ et considérons le couplage suivant : les deux processus évoluent indépendamment jusqu'au premier temps T pour lequel les deuxième coordonnées coïncident. Alors, I et $\tilde{I}$ sont égales pour $t > T$. Fixons $t > 0$ et $\beta \in (0, 1)$.

$$\mathbb{E}\Big(|X_t - \tilde{X}_t|^p\Big) = \mathbb{E}\Big(|X_t - \tilde{X}_t|^p\mathbb{1}_{\{T>\beta t\}}\Big) + \mathbb{E}\Big(|X_t - \tilde{X}_t|^p\mathbb{1}_{\{T\leqslant\beta t\}}\Big)$$

Choisissons $q \in (p, \kappa)$ et définissons $r = q/p$ et s le conjugué de r. L'inégalité de Hölder assure que

$$\begin{aligned}\mathbb{E}\Big(|X_t - \tilde{X}_t|^p\mathbb{1}_{\{T>\beta t\}}\Big) &\leqslant \mathbb{E}\Big(|X_t - \tilde{X}_t|^q\Big)^{p/q}\mathbb{P}(T \geqslant \beta t)^{1/s} \\ &\leqslant 2^p M(q, m)^{p/q} e^{-(\beta\rho/s)t}.\end{aligned}$$

De plus,

$$\begin{aligned}\mathbb{E}\Big(|X_t - \tilde{X}_t|^p\mathbb{1}_{\{T\leqslant\beta t\}}\Big) &= \mathbb{E}\Big(|X_T - \tilde{X}_T|^p\mathbb{E}_{I_T}\Big(\exp\Big(-p\int_T^t \alpha(I_s)\,ds\Big)\Big)\mathbb{1}_{\{T\leqslant\beta t\}}\Big) \\ &\leqslant 2^p M(p, m) C_2(p) e^{-\eta_p(1-\beta)t}.\end{aligned}$$

Enfin on prend la valeur de $\beta \in (0, 1)$ qui minimise le majorant obtenu. Avec

$$\beta = \frac{\eta_p}{\eta_p + \rho/s},$$

on obtient

$$\mathbb{E}\Big(|X_t - \tilde{X}_t|^p\Big) \leqslant 2^{p+1} M(q, m)^{p/q} C_2(p) \exp\left(-\frac{\rho/s}{\eta_p + \rho/s}\eta_p t\right).$$

C'est ce que nous voulions montrer.

3 Taux de saut variables

Nous allons démontrer le théorème 1.11. Comme les taux de saut sont variables le couplage que nous allons construire est plus subtil car la probabilité que I et $\tilde{I}$ se séparent est positive. Mais lorsque I et $\tilde{I}$ sont égaux, la distance entre X et $\tilde{X}$ décroît exponentiellement vite et comme les taux de saut sont lipschitziens la probabilité de séparation de I et $\tilde{I}$ diminue : il est de plus en plus facile de les faire sauter simultanément. Cette idée a déjà servi dans des contextes différents dans [9, 4].

Nous démontrons d'abord le lemme 1.10 qui assure que le processus X ne peut quitter une boule suffisamment grande. En particulier le support de la mesure invariante est inclus dans cette boule. Ensuite nous construisons un couplage entre deux copies (X, I) et $(\tilde{X}, \tilde{I})$ définies par le même générateur infinitésimal (1) mais partant de différentes conditions initiales. Enfin nous comparons la distance entre X et $\tilde{X}$ avec un processus auxiliaire qui converge vers 0 exponentiellement vite.

Démonstration du lemme 1.10. Prendre $\tilde{x} = 0$ dans (4) assure que, pour $\varepsilon \in (0, \alpha)$,

$$\left\langle F^i(x), x \right\rangle \leqslant -\alpha |x|^2 + \left\langle F^i(0), x \right\rangle \leqslant -(\alpha - \varepsilon)|x|^2 + M/(4\varepsilon),$$

si $M = \max_{i \in E} |F^i(0)|^2$. En d'autres termes,

$$|X_t|^2 - |X_s|^2 = \int_s^t 2\left\langle F^{I_u}(X_u), X_u \right\rangle du \leqslant -2(\alpha - \varepsilon) \int_s^t |X_u|^2 \, du + \frac{M}{2\varepsilon}(t - s).$$

Par conséquent, on a

$$|X_t|^2 \leqslant \frac{M}{4\varepsilon(\alpha - \varepsilon)} (1 - e^{-2(\alpha - \varepsilon)t}) + |X_0|^2 e^{-2(\alpha - \varepsilon)t}.$$

En choisissant $\varepsilon = \alpha/2$ on obtient

$$\left(|X_t|^2 - \frac{M}{\alpha^2} \right)^+ \leqslant e^{-\alpha t} \left(|X_0|^2 - \frac{M}{\alpha^2} \right)^+.$$

En particulier, X ne peut quitter la boule de centre 0 et de rayon $r = \sqrt{M}/\alpha$. $\square$

3.1 Le couplage

Nous construisons un processus de Markov à espace d'états $(\mathbb{R}^d \times E)^2$ dont les lois marginales (1) sont celles du processus que nous étudions partant respectivement de (x, i) et $(\tilde{x}, j)$. Le générateur infinitésimal de ce couplage est le suivant :

1. if $i \neq j$

$$\begin{aligned}
\mathfrak{L} f(x, i, \tilde{x}, j) = & \left\langle F^i(x), \nabla_x f(x, i, \tilde{x}, j) \right\rangle + \left\langle F^j(\tilde{x}), \nabla_{\tilde{x}} f(x, i, \tilde{x}, j) \right\rangle \\
& + \sum_{i' \in E} a(x, i, i')(f(x, i', \tilde{x}, j) - f(x, i, \tilde{x}, j)) \\
& + \sum_{j' \in E} a(\tilde{x}, j, j')(f(x, y, \tilde{x}, j') - f(x, y, \tilde{x}, j)).
\end{aligned}$$

2. if $i = j$:

$$\begin{aligned}\mathfrak{L}f(x,i,\tilde{x},j) =& \Big\langle F^i(x), \nabla_x f(x,i,\tilde{x},i)\Big\rangle + \Big\langle F^i(\tilde{x}), \nabla_{\tilde{x}} f(x,i,\tilde{x},i)\Big\rangle \\ &+ \sum_{i'\in E} (a(x,i,i') \wedge a(\tilde{x},i,i'))(f(x,i',\tilde{x},i') - f(x,i,\tilde{x},i)) \\ &+ \sum_{i'\in E} (a(x,i,i') - a(\tilde{x},i,i'))^+(f(x,i',\tilde{x},i) - f(x,i,\tilde{x},i)) \\ &+ \sum_{i'\in E} (a(x,i,i') - a(\tilde{x},i,i'))^-(f(x,i,\tilde{x},i') - f(x,i,\tilde{x},i)).\end{aligned}$$

Si f ne dépend que de (x,i) ou de $(\tilde{x},j)$, alors $\mathfrak{L}f = Af$. Le couplage défini ici fonctionne de la façon suivante. Quand I et $\tilde{I}$ sont différents, les deux processus (X,I) et $(\tilde{X},\tilde{I})$ évoluent indépendamment. Si $I = \tilde{I}$ alors deux processus de saut sont en compétition : un saut de I ou $\tilde{I}$ ou deux sauts simultanés. Un saut unique se produit avec le taux $\sum_{i'\in E} |a(x,i,i') - a(\tilde{x},i,i')|$. Ce taux est majoré par $\kappa|x-\tilde{x}|$. Deux sauts simultanés se produisent avec le taux $\sum_{i'\in E}(a(x,i,i') \wedge a(\tilde{x},i,i'))$.

Supposons que X_0 et $\tilde{X}_0$ appartiennent à la boule $\bar{B}(0,r)$ où r est donné par (5). Pour $t \geqslant 0$, notons D_t la distance entre X_t et $\tilde{X}_t$. Le processus $(D_t)_{t\geqslant 0}$ n'est pas markovien. Mais, tant que $I = \tilde{I}$, D_t décroît avec un taux supérieur à α. Si $I \neq \tilde{I}$, D_t peut grandir, mais ne dépasse jamais $d = 2r$. Après l'instant de coalescence T_c de deux copies independantes de I, D décroît de nouveau. Or T_c est stochastiquement dominé par une variable $\mathcal{E}(b)$ de loi exponentielle d'un certain paramètre $b > 0$ (par exemple, si $E = \{0,1\}$, alors T_c est le minimum de deux instants de saut de deux processus indépendants tous deux stochastiquement dominés par une variable aléatoire $\mathcal{E}(\underline{\lambda})$ et T_c est stochastiquement dominés par $\mathcal{E}(2\underline{\lambda})$). On en déduit que $\mathbb{E}(D_t) \leqslant \mathbb{E}(U_t)$ où $(U_t)_{t\geqslant 0}$ est le processus de Markov à espace d'état $[0,d] \cup \{d+\varepsilon\}$ de générateur infinitésimal

$$Gf(x) = \begin{cases} -\alpha x f'(x) + \kappa x(f(d+\varepsilon) - f(x)) & \text{if } x \in [0,d], \\ b(f(d) - f(d+\varepsilon)) & \text{if } x = d+\varepsilon. \end{cases}$$

3.2 Le processus auxiliaire

Théorème 3.1. *Pour tout $t \geqslant 0$,*

$$\mathbb{E}(U_t|U_0 = d) \leqslant \left(d + (d+\varepsilon)\left(\frac{p\alpha b e}{\sqrt{(\alpha+b)^2 - 4p\alpha b}}\right)\frac{\alpha t}{\alpha+\gamma}\right)\exp\left(-\frac{1}{1+\alpha/\gamma}\alpha t\right) \quad (9)$$

où

$$p = e^{-d\kappa/\alpha} \quad \text{et} \quad \gamma = \frac{(\alpha+b) - \sqrt{(\alpha+b)^2 - 4p\alpha b}}{2} = \frac{(\alpha+b) - \sqrt{(\alpha-b)^2 + 4(1-p)\alpha b}}{2}.$$

Remarque 3.2. *Si α tend vers ∞, alors γ tend vers d ; si b tend vers ∞, $\gamma \sim p\alpha/b$.*

Démonstration. Partant de $d+\varepsilon$, le processus U saute à un instant de loi $\mathcal{E}(b)$ en d et va alors exponentiellement vite vers 0 jusqu'à ce qu'il revienne éventuellement en $d+\varepsilon$. L'instant de premier saut T partant de d se définit de la façon suivante : soit E une variable aléatoire de loi $\mathcal{E}(1)$, alors

$$T \stackrel{\mathcal{L}}{=} \begin{cases} -\dfrac{1}{\alpha}\log\left(1-\dfrac{\alpha E}{d\kappa}\right) & \text{si } E < \dfrac{d\kappa}{\alpha}, \\ +\infty & \text{sinon.} \end{cases}$$

En effet, si $\{U_0 = d\}$,

$$\int_0^t \lambda(V_s)ds = \int_0^t d\kappa e^{-\alpha s}\, ds = \frac{d\kappa}{\alpha}(1-e^{-\alpha t}).$$

En d'autres termes, la fonction de répartition F_T de T satisfait, pour tout $t \geqslant 0$,

$$1 - F_T(t) = \mathbb{P}(T > t) = \exp\left(-\frac{d\kappa}{\alpha}(1-e^{-\alpha t})\right).$$

Définissons $p = e^{-d\kappa/\alpha}$. La loi de T est un mélange avec des poids p et $1-p$ de la masse de Dirac en $+\infty$ et de la mesure de probabilité sur $\mathbb{R}$ de densité

$$f\ :\ t \mapsto f(t) = \frac{d\kappa}{1-p} e^{-\alpha t} e^{-\frac{d\kappa}{\alpha}(1-e^{-\alpha t})} \mathbb{1}_{(0,+\infty)}(t) \tag{10}$$

et de fonction de répartition

$$F\ :\ t \mapsto F(t) = \left(\frac{1-e^{-\frac{d\kappa}{\alpha}(1-e^{-\alpha t})}}{1-e^{-\frac{d\kappa}{\alpha}}}\right)\mathbb{1}_{(0,+\infty)}(t).$$

Partant de d, U revient en d avec probabilité $1-p$. La propriété de Markov assure que le nombre N de retours U en d est une variable aléatoire géométrique de paramètre p. La longueur d'une boucle finie de d à ds'écrit comme la somme $S+E$ où la loi de S a pour densité la fonction f donnée dans (10), la loi de E est exponentielle de paramètre b et S et E sont independantes.

Lemme 3.3. *La variable aléatoire S est stochastiquement dominée par une variable exponentielle de paramètre α* i.e. *pour tout $t \geqslant 0$, $F(t) \geqslant F_\alpha(t)$ où $F_\alpha(t) = (1-e^{-\alpha t})\mathbb{1}_{\{t>0\}}$.*

Démonstration. Il suffit de remarquer que, pour tout $t \geqslant 0$,

$$1 - F(t) = \frac{e^{\frac{d\kappa}{\alpha}e^{-\alpha t}} - 1}{e^{\frac{d\kappa}{\alpha}} - 1} \leqslant e^{-\alpha t} = 1 - F_\alpha(t).$$

$\square$

La transformée de Laplace L_S de S de densité f est donc inférieure à celle d'une variable exponentielle de paramètre α : pour tout $s < \alpha$,

$$L_S(s) \leqslant \frac{\alpha}{\alpha - s}.$$

Si L_e est la transformée de Laplace de $S + E$, alors, pour tout $s < \alpha \wedge b$, on a

$$L_e(s) \leqslant \frac{\alpha}{\alpha - s} \frac{b}{b - s}.$$

Désignons par H le dernier temps de passage en d *i.e.* le dernier instant de saut de X et par L sa tranformée de Laplace. Soit N une variable géométrique p, $(S_i)_{i\geqslant 1}$ des variables de loi de densité f et $(E_i)_{i\geqslant 1}$ des variables exponentielles de paramètres b, toutes variables supposées indépendantes. Alors

$$H \stackrel{\mathcal{L}}{=} \sum_{i=1}^{N} (S_i + E_i).$$

Pour tout $s \in \mathbb{R}$ tel que $(1 - p)L_e(s) < 1$, on a

$$L(s) = \mathbb{E}\left(e^{sH}\right) = \frac{pL_e(s)}{1 - (1 - p)L_e(s)} = \frac{p}{1 - p}\left(\frac{1}{1 - (1 - p)L_e(s)} - 1\right).$$

Notons

$$\gamma = \frac{(\alpha + b) - \sqrt{(\alpha + b)^2 - 4p\alpha b}}{2} \quad \text{et} \quad \tilde{\gamma} = \frac{(\alpha + b) + \sqrt{(\alpha + b)^2 - 4p\alpha b}}{2}$$

les deux racines de $\xi^2 - (\alpha + b)\xi + p\alpha b = 0$ (avec $\gamma < \alpha \wedge b < \tilde{\gamma}$). Pour tout $s < \gamma$, on a $(1 - p)L_e(s) < 1$ et

$$L(s) \leqslant \frac{p\alpha b}{(\gamma - s)(\tilde{\gamma} - s)} \leqslant \frac{p\alpha b}{\tilde{\gamma} - s} \frac{1}{\gamma - s}. \tag{11}$$

Reste à contrôler $\mathbb{E}(U_t|U_0 = d)$. Nous distinguons suivant que $H > \beta t$ ou non $\beta \in (0, 1)$ (puis choisissons β au mieux) :

1. si $H < \beta t$, alors $U_t \leqslant e^{-(1-\beta)\alpha t}$,
2. la probabilité de l'événement $\{H \geqslant \beta t\}$ est petite pour les grands valeurs de t car la transformée de Laplace de H est finie sur un voisinage de l'origine.

Pour tous $\beta \in (0, 1)$ et $s > 0$,

$$\begin{aligned}\mathbb{E}(U_t|U_0 = d) &= \mathbb{E}\left(U_t \mathbb{1}_{\{T \leqslant \beta t\}}\right) + \mathbb{E}\left(U_t \mathbb{1}_{\{T > \beta t\}}\right) \\ &\leqslant de^{-(1-\beta)\alpha t} + (d + \varepsilon)L(s)e^{-s\beta t}. \end{aligned} \tag{12}$$

De l'équation (11), on déduit que, pour tout $s < \gamma$, on a $\log L(s) - \beta ts \leqslant h(s)$ où

$$h(s) = \log\left(\frac{p\alpha b}{\tilde{\gamma} - \gamma}\right) - \log(\gamma - s) - \beta ts.$$

La fonction h atteint son minimun en $s(t) = \gamma - (\beta t)^{-1}$ et

$$h(s(t)) = \log\left(\frac{p\alpha b}{\tilde{\gamma} - \gamma}\right) + \log(\beta t) + 1 - \gamma\beta t.$$

Pour $t > 0$ et $\beta \in (0,1)$, choisissons $s(t) = \gamma - (\beta t)^{-1}$ dans (12), nous obtenons

$$\begin{aligned}\mathbb{E}(U_t) &\leqslant de^{-(1-\beta)\alpha t} + (d+\varepsilon)e^{h(\gamma(t))} \\ &\leqslant de^{-(1-\beta)\alpha t} + (d+\varepsilon)\left(\frac{p\alpha be}{\tilde{\gamma}-\gamma}\right)\beta t e^{-\gamma\beta t}.\end{aligned}$$

Enfin, en prenant $\beta = \alpha(\alpha+\gamma)^{-1}$ de sorte que $(1-\beta)\alpha = \gamma\beta$, on en déduit

$$\mathbb{E}(U_t) \leqslant \left(d + (d+\varepsilon)\left(\frac{p\alpha be}{\tilde{\gamma}-\gamma}\right)\frac{\alpha t}{\alpha+\gamma}\right)\exp\left(-\frac{\alpha\gamma}{\alpha+\gamma}t\right).$$

En remplaçant $\tilde{\gamma} - \gamma$ par son expression en fonction de α, b et p on obtient (9). □

Références

[1] G. Alsmeyer, A. Iksanov, and U. Rösler, *On distributional properties of perpetuities*, J. Theoret. Probab. **22** (2009), no. 3, 666–682. 1.1

[2] Y. Bakhtin and T. Hurth, *Invariant densities for dynamical systems with random switching*, Preprint available on arXiv, 2012. 1

[3] J.-B. Bardet, H. Guérin, and F. Malrieu, *Long time behavior of diffusions with Markov switching*, ALEA Lat. Am. J. Probab. Math. Stat. **7** (2010), 151–170. MR 2653702 (2011k :60263) 1.1, 2.1, 2.1

[4] J.B. Bardet, A. Christen, A. Guillin, A. Malrieu, and P.-A. Zitt, *Total variation estimates for the TCP process*, Preprint available on arXiv, 2012. 1, 3

[5] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt, *On the stability of planar randomly switched systems*, preprint, 2012. 1, 1.1

[6] O. Boxma, H. Kaspi, O. Kella, and D. Perry, *On/Off Storage Systems with State-Dependent Inpout, Outpout and Swithching Rates*, Probability en the Engineering and Informational Siences **19** (2005), 1–14. 1, 1

[7] E. Buckwar and M. G. Riedler, *An exact stochastic hybrid model of excitable membranes including spatio-temporal evolution*, J. Math. Biol. **63** (2011), no. 6, 1051–1093. 1

[8] P. Caputo, P. Dai Pra, and G. Posta, *Convex entropy decay via the Bochner-Bakry-Émery approach*, Ann. Inst. Henri Poincaré Probab. Stat. **45** (2009), no. 3, 734–753. 1

[9] D. Chafaï, F. Malrieu, and K. Paroux, *On the long time behavior of the TCP window size process*, Stochastic Process. Appl. **120** (2010), no. 8, 1518–1534. 1, 3

[10] O. L. V. Costa, *Stationary distributions for piecewise-deterministic Markov processes*, J. Appl. Probab. **27** (1990), no. 1, 60–73. 1

[11] O. L. V. Costa and F. Dufour, *Ergodic properties and ergodic decompositions of continuous-time Markov processes*, J. Appl. Probab. **43** (2006), no. 3, 767–781. 1

[12] ———, *Stability and ergodicity of piecewise deterministic Markov processes*, SIAM J. Control Optim. **47** (2008), no. 2, 1053–1077. 1

[13] M. H. A. Davis, *Piecewise-deterministic Markov processes : a general class of nondiffusion stochastic models*, J. Roy. Statist. Soc. Ser. B **46** (1984), no. 3, 353–388, With discussion. MR MR790622 (87g :60062) 1

[14] ———, *Markov models and optimization*, Monographs on Statistics and Applied Probability, vol. 49, Chapman & Hall, London, 1993. 1, 1

[15] B. de Saporta and J.-F. Yao, *Tail of a linear diffusion with Markov switching*, Ann. Appl. Probab. **15** (2005), no. 1B, 992–1018. MR MR2114998 (2005k :60257) 1.1

[16] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Rev. **41** (1999), no. 1, 45–76. MR 1669737 (2000c :60102) 1

[17] F. Dufour and O. L. V. Costa, *Stability of piecewise-deterministic Markov processes*, SIAM J. Control Optim. **37** (1999), no. 5, 1483–1502 (electronic). 1

[18] V. Dumas, F. Guillemin, and Ph. Robert, *A Markovian analysis of additive-increase multiplicative-decrease algorithms*, Adv. in Appl. Probab. **34** (2002), no. 1, 85–111. 1

[19] Kussell E. and Leibler S., *Fluctuating environments phenotypic diversity, population growth, and information in fluctuating environments*, Science **309** (2005), 2075–2078. 1

[20] J. Fontbona, H. Guérin, and F. Malrieu, *Quantitative estimates for the long time behavior of a PDMP describing the movement of bacteria*, arXiv, 2010. 1

[21] C. M. Goldie and R. Grübel, *Perpetuities with thin tails*, Adv. in Appl. Probab. **28** (1996), no. 2, 463–480. 1.1

[22] C. Graham and P. Robert, *Interacting multi-class transmissions in large stochastic networks*, Ann. Appl. Probab. **19** (2009), no. 6, 2334–2361. 1

[23] ———, *Self-adaptive congestion control for multi-class intermittent connections in a communication network*, arXiv, 2010. 1

[24] X. Guyon, S. Iovleff, and J.-F. Yao, *Linear diffusion with stationary switching regime*, ESAIM Probab. Stat. **8** (2004), 25–35 (electronic). MR MR2085603 (2005h :60244) 1.1

[25] P. Hitsczenko and J. Wesołowski, *Perpetuities with thin tails revisited*, Ann. Appl. Probab. **19** (2009), no. 6, 2080–2101. 1.1

[26] H. Kesten, *Random difference equations and renewal theory for products of random matrices*, Acta Math. **131** (1973), 207–248. MR MR0440724 (55 #13595) 1.1

[27] K. Pakdaman, M. Thieullen, and G. Wainrib, *Fluid limit theorems for stochastic hybrid systems with application to neuron models*, Adv. in Appl. Probab. **42** (2010), no. 3, 761–794. 1

[28] S. T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1991. MR MR1105086 (93b :60012) 1

[29] O. Radulescu, A. Muller, and A. Crudu, *Théorèmes limites pour des processus de Markov à sauts. Synthèse des résultats et applications en biologie moléculaire*, Technique et Science Informatiques **26** (2007), no. 3-4, 443–469. 1

[30] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413. 2.1

[31] W. Vervaat, *On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables*, Adv. in Appl. Probab. **11** (1979), no. 4, 750–783. MR MR544194 (81b :60064) 1.1

[32] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR MR1964483 (2004e :90003) 1

Michel Benaïm, e-mail : `michel.benaim(AT)unine.ch`

Institut de Mathématiques, Université de Neuchâtel, 11 rue Émile Argand, 2000 Neuchâtel, Suisse.

Stéphane Le Borgne, e-mail : `stephane.leborgne(AT)univ-rennes1.fr`

IRMAR Université de Rennes 1, Campus de Beaulieu, F-35042 Rennes Cedex, France.

Florent Malrieu, e-mail : `florent.malrieu(AT)univ-rennes1.fr`

IRMAR Université de Rennes 1, Campus de Beaulieu, F-35042 Rennes Cedex, France.

Pierre-André Zitt, e-mail : `pierre-andre.zitt(AT)u-bourgogne.fr`

Institut de Mathématiques de Bourgogne, Université de Bourgogne, 9 rue Alain Savary - BP 47870, 21078 Dijon Cedex, France