Quantitative ergodicity for some switched dynamical systems

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

Quantitative ergodicity for some switched dynamical systems

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

April 9, 2012

Abstract

We provides quantitative bounds for the long time behavior of a class of Piecewise Deterministic Markov Processes with state space $\mathbb{R}^d \times E$ where E is a finite set. The continous component evolves according to a smooth vector field that it switched at the jump times of the discrete coordinate. The jump rates may depend on the whole position of the process. Under regularity assumptions on the jump rates and stability conditions for the vector fields we provide explicit exponential upper bounds for the convergence to equilibrium in terms of Wasserstein distances.

Keywords. Coupling; Ergodicity; Linear Differential Equations; Piecewise Deterministic Markov Process; Switched dynamical systems; Wasserstein distance.

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

1 Introduction and main results

Piecewise deterministic Markov processes (PDMPs in short) are intensively used in many applied areas (molecular biology [29], storage modelling [7], Internet traffic [19, 22, 23], neuronal activity [8, 27],...). Roughly speaking, a Markov process is a PDMP if its randomness is only given by the jump mechanism: in particular, it admits no diffusive dynamics. This huge class of processes has been introduced by Davis (see [14, 15]) in a general framework. Several works [18, 11, 12] deal with their long time behavior (existence of an invariant probability measure, Harris recurrence, exponential ergodicity...). In particular, it is shown in [13] that the behavior of a general PDMP can be related to the one of the discrete time Markov chain made of the positions at the jump times of the process and and of an additional independent Poisson process. Nevertheless, this general approach does not seem to provide quantitative bounds for the convergence to equilibrium. Recent papers have tried to establish such estimates for some specific PDMPs (see [10, 20, 4]) or continuous time Markov chains (see [9]).

In the present paper, we investigate the long time behavior of an interesting subclass of PDMPs that plays a role in molecular biology (see [29, 8]). We consider a PDMP on $\mathbb{R}^d \times E$ where E is a finite set. The first coordinate moves continuously on $\mathbb{R}^d$ according to a smooth vector field that depends on the second coordinate whereas the second coordinate jumps with a rate that may depend on the first one. This class of Markov processes is reminiscent of the so-called iterated random functions in the discrete time setting (see [17] for a good review of this topic).

Let E be a finite set, $(\lambda(\cdot, i))_{i\in E}$ be n nonnegative continuous functions on $\mathbb{R}^d$, P be an irreducible stochastic matrix and, for any $i \in E$, $F^i : \mathbb{R}^d \mapsto \mathbb{R}^d$ be a smooth vector field such that the ordinary differential equation

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

has a unique and global solution $t \mapsto \varphi_t^i(x)$ on $[0, \infty)$ for any initial condition $x \in \mathbb{R}^d$. Let us consider the Markov process

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

defined by its extended generator L as follows:

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

for any smooth function $f : \mathbb{R}^d \times E \to \mathbb{R}$ (see [15] for full details on the domain of L). Let us describe the dynamics of this process. Assume that $(X_0, I_0) = (x, i) \in \mathbb{R}^d \times E$. Before the first jump time T_1 of I, the first component X is driven by the vector field F^i and then $X_t = \varphi_t^i(x)$. The time T_1 can be defined by:

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

where E_1 is an exponential random variable with parameter 1. Since the paths of X are deterministic between the jump times of I, the randomness of T_1 comes from the one of E_1 and

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

Remark 1.1. *Notice that* $\mathbb{P}_{(x,i)}(T_1 = +\infty) > 0$ *if and only if*

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

If we assume that $\underline{\lambda} := \inf_{(x,i)} \lambda(x,i) > 0$ *then the process* I *jumps infinitely often.*

At time T_1, the second coordinate I performs a jump with the law $P(i, \cdot)$ and the vector field that drives the evolution of X is switched...

Remark 1.2. *In general,* I *is not a Markov process on its own since its jump rates depend on* X*. In this paper, we will study the two cases of jump rates that depend (or not) on* X*.*

The main goal of the present work is to provide quantitative bounds for the long time behavior of ergodic processes driven by (1). We will provide quantitative bounds in terms of the Wasserstein coupling distance rather than total variation one. Recall that for every $p \geqslant 1$, the Wasserstein distance W_p between two laws μ and $\tilde{\mu}$ on $\mathbb{R}^d$ with finite p^{th} moment is defined by

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

where the infimum runs over all the probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and $\tilde{\mu}$ (such measures are called couplings of μ and $\tilde{\mu}$). It is well known that for any $p \geqslant 1$, the convergence in W_p Wasserstein distance is equivalent to weak convergence together with convergence of all moments up to order p. However, two probability measures can be both very close in the W_p sense and singular. Choose for example $\mu = \delta_0$ and $\tilde{\mu} = \delta_\varepsilon$. In this case,

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

See e.g. [28, 32] for further details and properties for Wasserstein distances.

Estimates for the Wasserstein distances do not require (and provide) any information about the support of the invariant measure (which is the set of the recurrent points). This set may be difficult to determine and if the initial distribution of X is not supported by this set, the law of X_t and the invariant measure may be singular. To illustrate this fact, one can consider the following trivial example:

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

The process (X, I) is ergodic and its invariant measure μ is supported by the segment $\{\rho x \ ; \ \rho \in [0, 1]\}$. The invariant measure is a Beta distribution (see [7]). Despite this process is very simple, if $X_0 = (0, 1)$ then the law of X_t is singular with the invariant measure for any $t \geqslant 0$. In particular, $\|\mathcal{L}(X_t) - \mu\|_{\mathrm{TV}}$ is equal to 1 for any $t \geqslant 0$ whereas $W_p(\mathcal{L}(X_t), \mu)$ goes to 0 exponentially fast (see below).

In [2], the authors provide Hörmander-like conditions on the vectors fields $(F^i)_{i \in E}$ that ensure, for constant jump rate $(\lambda_i)_{i \in E}$, the uniqueness and the absolute continuity of the invariant measure provided that it exists. The main drawback of this result is that these regularity assumptions have to be checked at a point that can be reached starting from any other point. It can be hard to determine the set of such points and it can be empty. Several examples are studied in [6] that underline this fact. In particular, even if the process evolves in a compact set, the process (X, I) may admits one or several recurrent classes (and invariant measures) depending only on the values of jump rates $(\lambda_i)_{i \in E}$.

We are able to get explicit rates of convergence in two situations. Firstly, if the jump rates of I does not depend on X, then the vector fields $(F^i)_{i \in E}$ will be assumed to satisfy an averaged exponential stability. Secondly, if the jump rates of I are assumed to be lipschitz functions of X, then the vector fields $(F^i)_{i \in E}$ will be assumed to satisfy a uniform exponential stability.

1.1 Constant jump rates

If the jump rates of I do not depend on X, then $(I_t)_{t \geqslant 0}$ is a Markov process on the finite space E and $(X_s)_{0 \leqslant s \leqslant t}$ is a deterministic function of $(I_s)_{0 \leqslant s \leqslant t}$. Many results are available both in the discrete time setting (see [26, 31, 21, 25, 1]) and in the continuous time setting (see [24, 16, 3]). Morevover [5] provides a simple example of surprising phase transition for a switching of two exponentially stable flows that can be explosive (when the jump rates are sufficiently large).

Assumption 1.3. *Assume that the jump rates $(\lambda(\cdot, i))_{i \in E}$ do not depend on x and that I is an irreducible Markov process on E. Let us denote by ν its invariant probability measure.*

Assumption 1.4. *Assume that for any $i \in E$, there exists $\alpha(i) \in \mathbb{R}$ such that,*

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

and that

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

where ν is defined in Assumption 1.3.

Firstly, one can establish that the process X is bounded in some L^p space.

Lemma 1.5. *Under Assumptions 1.3 and 1.4, there exists $\kappa > 0$ such that, for any $q < \kappa$, the function $t \mapsto \mathbb{E}(|X_t|^q)$ is bounded as soon as $\mathbb{E}(|X_0|^q)$ is finite. More precisely, there exists $M(q,m)$ such that*

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

as soon as $\mathbb{E}(|X_0|^q) \leqslant m$.

Let us now turn to the long time behavior estimate.

Theorem 1.6. *Assume that Assumptions 1.3 and 1.4 hold. Let $p < q < \kappa$ and denote by s the conjugate of q: $q^{-1} + s^{-1} = 1$. Assume that μ_0 and $\tilde{\mu}_0$ admit a finite q^{th} moment smaller than m. Then,*

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

where ρ and η_p are positive constants depending only the Markov chain I. These constants are made explicit in the proof.

Corollary 1.7. *Under Assumptions 1.3 and 1.4, the process Z admits a unique invariant measure μ and*

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

1.2 Non constant jump rates

Let us know turn to the case when the jump rates of I depend on X. We will assume that this dependence is smooth and that each vector field F^i has a unique stable point.

Assumption 1.8. *There exist $0 < \underline{\lambda} \leqslant \overline{\lambda}$ and $\kappa > 0$ such that, for any $x, \tilde{x} \in \mathbb{R}^d$ and $i \in E$,*

$$\lambda(x,i) \in [\underline{\lambda}, \overline{\lambda}] \quad \text{and} \quad |\lambda(x,i) - \lambda(\tilde{x},i)| \leqslant \kappa |x - \tilde{x}|,$$

Moreover the matrix P is irreducible.

Assumption 1.9. *Assume that there exists $\alpha > 0$ such that,*

$$\left\langle x - \tilde{x}, F^i(x) - F^i(\tilde{x}) \right\rangle \leqslant -\alpha |x - \tilde{x}|^2, \quad x, \tilde{x} \in \mathbb{R}^d, \ i \in E. \tag{3}$$

Assumption 1.9 ensures that, for any $i \in E$,

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

As a consequence, the vector fields F^i has exactly one critical point $\sigma(i) \in \mathbb{R}^d$. Moreover it is exponentially stable since, for any $x \in \mathbb{R}^d$,

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

In particular, X cannot escape from a sufficiently large ball. More precisely, the following estimate holds.

Lemma 1.10. *Under Assumtions 1.8 and 1.9, the process Z cannot escape from the compact set $\bar{B}(0,r)\times E$ where $\bar{B}(0,r)$ is the (closed) ball centered in $0\in\mathbb{R}^d$ with radius r given by*

$$r = \frac{\max_{i\in E}|F^i(0)|}{\alpha}. \tag{4}$$

Moreover, if $|X_0| > r$ then X hits $\bar{B}(0,r)$ exponentially fast.

Let us now state our main result which establishes the quantitative exponential ergodicity of the process Z under Assumptions 1.8 and 1.9.

Theorem 1.11. *Assume that Assumptions 1.8 and 1.9 hold and that the supports of μ_0 and $\tilde{\mu}_0$ are included in the ball $\bar{B}(0,r)$ where r is given by (4). Then*

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

where

$$\gamma = \frac{(\alpha+2\underline{\lambda}) - \sqrt{(\alpha+2\underline{\lambda})^2 - 8p\alpha\underline{\lambda}}}{2} \quad \text{and} \quad c = \frac{\alpha}{\alpha+\gamma}\frac{2ep\alpha\underline{\lambda}}{\sqrt{(\alpha+2\underline{\lambda})^2-8p\alpha\underline{\lambda}}},$$

with $p = e^{-\kappa/\alpha}$ and $e = \exp(1)$.

Corollary 1.12. *Under Assumptions 1.8 and 1.9, the process Z admits a unique invariant measure μ and*

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

The paper is organized as follows. Section 2 is dedicated to the proof of Theorem 1.6. Theorem 1.11 is established in Section 3.

2 Constant jump rates

The aim of this section is to prove Theorem 1.6. Assumption 1.3 ensures that $(I_t)_{t\geqslant 0}$ is an irreducible Markov process on the finite space E. Its generator A is the matrix defined by $A(i,i) = -\lambda(i)$ and $A(i,j) = \lambda(i)P(i,j)$ for $i\neq j$. Let us denote by ν its unique invariant probability measure. The study of the long time behavior of I is classical: since I takes its values in a finite set, it is quite simple to construct a coalescent coupling of two processes starting from different points.

Lemma 2.1 ([30])**.** *If Assumption 1.3 holds then there exists $\rho > 0$ such that for any $i,j\in E$,*

$$\mathbb{P}(T>t|I_0=i,\ \tilde{I}_0=j) \leqslant e^{-\rho t} \tag{5}$$

where $(I_t)_{t\geqslant 0}$ and $(\tilde{I}_t)_{t\geqslant 0}$ are two independent Markov processes with infinitesimal generator A starting respectively at i and j and $T = \inf\left\{t\geqslant 0\ :\ I_t = \tilde{J}_t\right\}$ is the first intersection time.

Remark 2.2. *If $E=\{1,2\}$, then the first intersection time is distributed as an exponential random variable with parameter $\lambda(1)+\lambda(2)$ and Equation (5) holds with $\rho=\lambda(1)+\lambda(2)$.*

The proof of Theorem 1.6 is made of two steps. Firstly we couple two processes starting respectively from (x,i) and $(\tilde{x},i)$ to get a simple estimate as time goes to infinity. Then we use this estimate and Lemma 2.1 to manage the general case.

2.1 Moments estimates

In this section we prove Lemma 1.5 and get an L^p estimate for $|X_t|$. For any $p \geqslant 2$ and $\varepsilon > 0$,

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

As a consequence,

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

Remark 2.3. *A similar estimate can be obtained for $p \geqslant 1$ using a regularization of the application $x \mapsto |x|^p$.*

We have to investigate the behavior of the matrix $A_{(p,t)}$ on $E \times E$ defined for any $t \geqslant 0$ by

$$A_{(p,t)}(i,j) = \mathbb{E}_i\Big(\exp\Big(-\int_0^t p\alpha(I_u)\,du\Big)\mathbb{1}_{\{I_t=j\}}\Big).$$

This study has been already performed (see [3] for further details). Let us state the precise result. We denote by A_p the matrix $A - pB$ where B is the diagonal matrix with diagonal $(\alpha(1),\ldots,\alpha(n))$ and associate to A_p the quantity

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

The long time behavior of $A_{(p,t)}$ is characterised by η_p as it is recalled below.

Proposition 2.4 ([3])**.** *For any $p > 0$, there exist $0 < C_1(p) < C_2(p) < +\infty$ such that, for any $i, j \in E$ and any $t > 0$,*

$$C_1(p)e^{-\eta_p t} \leqslant A_{(p,t)}(i,j) \leqslant C_2(p)e^{-\eta_p t}. \tag{6}$$

Moreover, the function $p \mapsto \eta_p$ is smooth and concave on $\mathbb{R}_+$. Its derivative at $p = 0$ is equal to

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

and η_p/p tends to $\underline{\alpha} = \min\{\alpha(i) : i \in E\}$ as p goes to infinity.

Remark 2.5. *We have the following dichotomy:*

- *if $\underline{\alpha} \geqslant 0$, then $\eta_p > 0$ for all $p > 0$,*
- *if $\underline{\alpha} < 0$, there is $\kappa \in (0, \min\{-a(i)/\alpha(i) : \alpha(i) < 0\})$ such that $\eta_p > 0$ for $p < \kappa$ and $\eta_p < 0$ for $p > \kappa$.*

Corollary 2.6. *If $p < \kappa$ then $t \mapsto \mathbb{E}(|X_t|^p)$ is bounded as soon as $\mathbb{E}(|X_0|^p)$ is finite.*

2.2 Convergence rate

Let us now get the upper bound for the Wasserstein distance W_p for some $p < \kappa$. Assume firstly that the initial law are two Dirac masses at (x, i) and $(\tilde{x}, i)$. It is easy to construct a good coupling of the two processes (X, I) and $(\tilde{X}, \tilde{I})$: since the jump rates of I do not depend on X, one can choose I and $\tilde{I}$ equal! As a consequence, for any $p \geqslant 2$,

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

As a consequence,

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

Let us now turn to a general initial condition. Choose (x, i) and $(\tilde{x}, j)$ in $\mathbb{R}^d \times E$ and consider the following coupling: the two processes evolve independently until the intersection time T of the second coordinates. Then, I and $\tilde{I}$ are chosen to be equal for ever. Now fix $t > 0$ and $\beta \in (0, 1)$.

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

Choose $q \in (p, \kappa)$ and define $r = q/p$ and s as the conjugate of r. The Hölder inequality ensures that

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

Moreover,

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

At last, one has to optimize over $\beta \in (0, 1)$. With

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

one has

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

This concludes the proof of Theorem 1.6.

3 Exponential convergence for non constant jump rates

Let us now turn to the proof of Theorem 1.11. In this section we do not assume that the jump rates depend only on the discrete component. Thus, the coupling is more subtle since once I and $\tilde{I}$ are equal, they can go appart with positive probability. Nevertheless, the main idea is the following. If I and $\tilde{I}$ are equal, the distance between X and $\tilde{X}$ decreases exponentially fast and then it should be more and more easier to make the processes I and $\tilde{I}$ jump simultaneously (since the jump rates are Lipschitz functions of X). This idea has been used in a different framework in [10, 4].

This section is organized as follows. Firstly we prove Lemma 1.10 that ensures that the process X cannot escape from a sufficiently large ball. In particular, the support of the invariant law of X is included in this ball. Then we construct the coupling of two processes (X, I) and $(\tilde{X}, \tilde{I})$ driven by the same infinitesimal generator (1) with different initial condition. At last we compare the distance between X and $\tilde{X}$ to an companion process that goes to 0 exponentially fast.

3.1 A preliminary estimate

Proof of Lemma 1.10. Setting $\tilde{x} = 0$ in (3) ensures that, for $\varepsilon \in (0, \alpha)$,

$$\langle f(x,y), x\rangle \leqslant -\alpha|x|^2 + \langle f(0,y), x\rangle \leqslant -(\alpha-\varepsilon)|x|^2 + C(\varepsilon),$$

if $C(\varepsilon) = \max_{y\in E} |f(0,y)|^2/(4\varepsilon)$. In other words,

$$|X_t|^2 - |X_s|^2 = \int_s^t 2\langle f(X_u, Y_u), X_u\rangle \, du \leqslant -2(\alpha-\varepsilon)\int_s^t |X_u|^2 \, du + 2C(\varepsilon)(t-s).$$

As a consequence,

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

With $\varepsilon = \alpha/2$, one gets that X cannot escape from the centered closed ball with radius $r = \sqrt{2C(\alpha/2)/\alpha}$. With $\varepsilon = \alpha/4$, one gets that if $|X_0| \geqslant r$, then X will hit $\bar{B}(0,r)$ exponentially fast. □

3.2 The coupling

Let us construct a Markov process on $(\mathbb{R}^d \times E)^2$ with marginals driven by (1) starting respectively from (x,i) and $(\tilde{x}, j)$. This is done *via* its infinitesimal generator which is defined as follows:

- if $i \neq j$

$$\begin{aligned} Af(x,i,\tilde{x},j) =& \Big\langle F^i(x), \nabla_x f(x,i,\tilde{x},j)\Big\rangle + \Big\langle F^j(\tilde{x}), \nabla_{\tilde{x}} f(x,i,\tilde{x},j)\Big\rangle \\ &+ \lambda(x,i)\sum_{i'\in E} P(i,i')(f(x,i',\tilde{x},j) - f(x,i,\tilde{x},j)) \\ &+ \lambda(\tilde{x},j)\sum_{j'\in E} P(j,j')(f(x,y,\tilde{x},j') - f(x,y,\tilde{x},j)). \end{aligned}$$

- if $i = j$ and $\lambda(x,i) \geqslant \lambda(\tilde{x},i)$:

$$\begin{aligned} Af(x,i,\tilde{x},j) =& \Big\langle F^i(x), \nabla_x f(x,i,\tilde{x},i)\Big\rangle + \Big\langle F^i(\tilde{x}), \nabla_{\tilde{x}} f(x,i,\tilde{x},i)\Big\rangle \\ &+ \lambda(\tilde{x},i)\sum_{i'\in E} P(i,i')(f(x,i',\tilde{x},i') - f(x,i,\tilde{x},i)) \\ &+ (\lambda(x,i) - \lambda(\tilde{x},i))\sum_{i'\in E} P(i,i')(f(x,i',\tilde{x},i) - f(x,i,\tilde{x},i)), \end{aligned}$$

- if $i=j$ and $\lambda(x,i)<\lambda(\tilde{x},i)$:

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

Notice that if f depends only on (x,i) or on $(\tilde{x},j)$, then $Af=Lf$. Let us explain how this coupling works. When I and $\tilde{I}$ are different, the two processes (X,I) and $(\tilde{X},\tilde{I})$ evolve independently. If $I=\tilde{I}$ then two jump processes are in competition: a single jump *vs* two simultaneous jumps. The rate of arrival of a single jump is given by $|\lambda(x,i)-\lambda(\tilde{x},i)|$. It is bounded above by $\kappa|x-\tilde{x}|$. The rate of arrival of a simultaneous jump is given by $\lambda(x,i)\wedge\lambda(\tilde{x},i)\geqslant\underline{\lambda}$.

Assume firstly that X_0 and $\tilde{X}_0$ belong to the ball $\bar{B}(0,r)$ where r is given by (4). Let us define D_t as the distance between X_t and $\tilde{X}_t$ for any $t\geqslant 0$. The process $(D_t)_{t\geqslant 0}$ is not Markovian. Nevertheless, as long as $I=\tilde{I}$, D_t decreases with a rate which is greater than α. If it is no longer the case, then D_t can increase. Nevertheless it is still smaller than $d=2r$. After the coalescent time T_c of two independent independent copies of Y, D decreases once again. If $E=\{0,1\}$, then T_c is equal to the minimum of the jump times of the two independent processes which are both stochastically greater than a random variable of law $\mathcal{E}(\underline{\lambda})$. Thus T_c is (stochastically) smaller than $\mathcal{E}(2\underline{\lambda})$. Then $\mathbb{E}(D_t)\leqslant\mathbb{E}(U_t)$ where the Markov process $(U_t)_{t\geqslant 0}$ on $[0,d]\cup\{d+\varepsilon\}$ is driven by the infinitesimal generator

$$Gf(x)=\begin{cases}-\alpha x f'(x)+\kappa x(f(d+\varepsilon)-f(x)) & \text{if } x\in[0,d],\\ 2\underline{\lambda}(f(d)-f(d+\varepsilon)) & \text{if } x=d+\varepsilon.\end{cases}$$

3.3 The companion process

Let us consider the Markov process $V=(V_t)_{t\geqslant 0}$ on $[0,1]\cup\{1+\varepsilon\}$ defined by its infinitesimal generator:

$$Hf(x)=\begin{cases}-\alpha x f'(x)+\kappa x(f(1+\varepsilon)-f(x)) & \text{if } x\in[0,1],\\ b(f(1)-f(1+\varepsilon)) & \text{if } x=1+\varepsilon.\end{cases}$$

Theorem 3.1. *For any $t\geqslant 0$,*

$$\mathbb{E}(V_t|V_0=1)\leqslant\left(1+(1+\varepsilon)\left(\frac{p\alpha be}{\sqrt{(\alpha+b)^2-4p\alpha b}}\right)\frac{\alpha t}{\alpha+\gamma}\right)\exp\left(-\frac{1}{1+\alpha/\gamma}\alpha t\right)\qquad(7)$$

where

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

Remark 3.2. *If α goes to ∞, then γ goes to 1 whereas $\gamma\sim p\alpha/b$ if b goes to ∞.*

Proof. Starting from $1+\varepsilon$, the process V jumps after a random time with law $\mathcal{E}(b)$ to 1 and then goes to zero exponentially fast until it (possibly) goes back to $1+\varepsilon$. The first jump time T starting from 1 can be constructed as follows: let E be a random variable with law $\mathcal{E}(1)$. Then

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

Indeed, conditionally on $\{V_0 = 1\}$,

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

In other words, the cumulative distribution function F_T of T is such that, for any $t \geqslant 0$,

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

Let us define $p = e^{-\kappa/\alpha}$. The law of T is the mixture with respective weights p and $1-p$ of a Dirac mass at $+\infty$ and a probability measure on $\mathbb{R}$ with density

$$f\,:\,t \mapsto f(t) = \frac{\kappa}{1-p}e^{-\alpha t}e^{-\frac{\kappa}{\alpha}(1-e^{-\alpha t})}\mathbb{1}_{(0,+\infty)}(t) \tag{8}$$

and cumulative distribution function

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

Starting at 1, X will return to 1 with probability $1-p$. The Markov property ensures that the number N of returns of X to 1 is a random variable with geometric law with parameter p. The length of a finite loop from 1 to 1 can be written as the sum $S+E$ where the law of S has the density function f given in (8), the law of E is the exponential measure with parameter b and S and E are independent.

Remark 3.3. *In the general case, E is not distributed as an exponential variable but as the coalescent time of a finite Markov chain. Its Laplace transform is finite on a neighbourhood of the origin.*

Lemma 3.4. *The variable S is stochastically smaller than an exponential random variable with parameter α* i.e. *for any $t \geqslant 0$, $F(t) \geqslant F_\alpha(t)$ where $F_\alpha(t) = (1-e^{-\alpha t})\mathbb{1}_{\{t>0\}}$.*

Proof of Lemma 3.4. Recall that $e^{ux} - 1 \leqslant (e^x - 1)u$ for any $x > 0$ and $u \in [0,1]$. Indeed,

$$e^{ux} - 1 = u\sum_{k\geqslant 1} u^{k-1}\frac{x^k}{k!} \leqslant u\sum_{k\geqslant 1}\frac{x^k}{k!} = (e^x-1)u.$$

As a consequence, for any $t \geqslant 0$,

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

This ensures the stochastic bound. $\square$

As a consequence, the Laplace transform L_S of S with density f is smaller than the one of an exponential variable with parameter α: for any $s < \alpha$,

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

If L_e is the Laplace transform of $S + E$, then, for any $s < \alpha \wedge b$, we have

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

Let us denote by H the last hitting time of 1 *i.e.* the last jump time of X and by L its Laplace transform. Let us introduce $N \sim \mathcal{G}(p)$, $(S_i)_{i\geqslant 1}$ with density f and $(E_i)_{i\geqslant 1}$ with law $\mathcal{E}(b)$. All the random variables are assumed to be independent. Then

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

Classically, for any $s \in \mathbb{R}$ such that $(1-p)L_e(s) < 1$, one has

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

Let us denote by

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

the two roots of $X^2 - (\alpha + b)X + p\alpha b = 0$. Notice that $\gamma < \alpha \wedge b < \tilde{\gamma}$. For any $s < \gamma$, one has $(1-p)L_e(s) < 1$ and

$$L(s) \leqslant \frac{p\alpha b}{(\gamma - s)(\tilde{\gamma} - s)} \leqslant \frac{p\alpha b}{\tilde{\gamma} - s} \frac{1}{\gamma - s}. \tag{9}$$

Let us now turn to the control of $\mathbb{E}(V_t|V_0 = 1)$. The idea is to discuss wether $H > \beta t$ or not for some $\beta \in (0,1)$ (and then to choose β as good as possible):

- if $H < \beta t$, then $V_t \leqslant e^{-(1-\beta)\alpha t}$,
- the event $\{H \geqslant \beta t\}$ has a small probability for large t since H has a finite Laplace transform on a neighbourhood of the origin.

For any $\beta \in (0,1)$ and $s > 0$,

$$\begin{aligned} \mathbb{E}(V_t|X_0 = 1) &= \mathbb{E}\Big(V_t \mathbb{1}_{\{T \leqslant \beta t\}}\Big) + \mathbb{E}\Big(V_t \mathbb{1}_{\{T > \beta t\}}\Big) \\ &\leqslant e^{-(1-\beta)\alpha t} + (1+\varepsilon)L(s)e^{-s\beta t}. \end{aligned} \tag{10}$$

From Equation (9), we get that, for any $s < \gamma$, $\log L(s) - \beta t s \leqslant h(s)$ where

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

The function h reaches its minimun at $s(t) = \gamma - (\beta t)^{-1}$ and

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

For $t > 0$ and $\beta \in (0,1)$, choose $s(t) = \gamma - (\beta t)^{-1}$ in (10) to get

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

At last, one can choose $\beta = \alpha(\alpha+\gamma)^{-1}$ in order to have $(1-\beta)\alpha = \gamma\beta$. This ensures that

$$\mathbb{E}(V_t) \leqslant \left(1 + (1+\varepsilon)\left(\frac{p\alpha b e}{\tilde{\gamma}-\gamma}\right)\frac{\alpha t}{\alpha+\gamma}\right)\exp\left(-\frac{\alpha\gamma}{\alpha+\gamma}t\right).$$

Replacing $\tilde{\gamma} - \gamma$ by its expression as a function of α, b and p provides (7). □

Acknowledgements. FM and PAZ thank MB for his kind hospitality and his coffee breaks. We acknowledge financial support from the Swiss National Foundation Grant FN 200021-138242/1 and the French ANR projects EVOL and ProbaGeo.

References

[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, 2.4

[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

[6] ______, *Qualitative properties of certain piecewise deterministic markov processes*, preprint, 2012. 1

[7] 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

[8] 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

[9] 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

[10] 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

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

[12] 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

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

[14] 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

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

[16] 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

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

[18] 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

[19] 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

[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