



On the long time behavior of the TCP window size process

Djalil Chafai, Florent Malrieu, Katy Paroux

► To cite this version:

Djalil Chafai, Florent Malrieu, Katy Paroux. On the long time behavior of the TCP window size process. 2008. hal-00338640v1

HAL Id: hal-00338640

<https://hal.science/hal-00338640v1>

Preprint submitted on 13 Nov 2008 (v1), last revised 15 Dec 2009 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the long time behavior of the TCP window size process

Djalil CHAFAÏ, Florent MALRIEU, Katy PAROUX

Preprint – November 2008

Abstract

The TCP window size process appears in the modeling of the famous Transmission Control Protocol used over the Internet. This continuous time Markov process takes its values in $[0, \infty)$, is ergodic and irreversible. It belongs to the Additive Increase Multiplicative Decrease class of processes. The sample paths are piecewise linear deterministic and the whole randomness of the dynamics comes from the jump mechanism. Several aspects of this process have been already investigated in the literature. In the present paper, we mainly get quantitative estimates for the exponential convergence to equilibrium, in terms of the W_1 Wasserstein coupling distance, for the process and also for its embedded chain.

Keywords. Network Protocols ; Queuing Theory ; Additive Increase Multiplicative Decrease Processes (AIMD) ; Piecewise Deterministic Markov Processes (PDMP) ; Exponential Ergodicity; Coupling.

AMS-MS. 68M12 ; 60K30 ; 60K25 ; 90B18

1 Introduction

Our aim is to investigate the trend to equilibrium, in Wasserstein coupling distance, of the so called general TCP window size process and its embedded Markov chain. The general TCP window size process is a continuous time Markov process $X = (X_t)_{t \geq 0}$ with state space $[0, \infty)$ and infinitesimal generator given, for any smooth function $f : [0, \infty) \rightarrow \mathbb{R}$, by

$$L(f)(x) = f'(x) + x \int_0^1 (f(hx) - f(x))H(dh) \quad (1)$$

for some probability measure H supported in $[0, 1)$. In practice, the probability measure H is a Dirac mass at point $\delta \in [0, 1)$ (and δ is often equal to $1/2$). This process is ergodic and admits a unique invariant law, as can be checked using a suitable Lyapunov function (for instance $V(x) = 1 + x$, see e.g. [13, 1] for the Meyn-Tweedie-Lyapunov technique).

The process X appears naturally as a scaling limit in the modeling of the window size of the Transmission Control Protocol (TCP) used over wide computer networks such as Internet (see e.g. [14, 7, 3] for more details). Recall that the window size of the TCP stack of a single machine is the maximum number of packets that can be transmitted before receiving any acknowledgment from the destination. Basically, X_t is the window size at time t of the TCP stack of a typical machine in a wide network. This window size increases linearly along the time (this is the f' part of L) until the reception of a congestion signal which forces the reduction of the window size by a multiplicative factor of law H or equal to δ in the simplest case (this is the jump part of L). The sample paths of X are

deterministic between jumps, the jumps are multiplicative, and the whole randomness of the dynamics relies on the jump mechanism. Of course, the randomness of X may also come from a random initial value. A typical sample path of X is given by Figure 1. The process $(X_t)_{t \geq 0}$ appears as an Additive Increase Multiplicative Decrease process (AIMD), but also as a very special Piecewise Deterministic Markov Process (PDMP). In real world (Internet), the AIMD mechanism allows a good compromise between the minimization of network congestion time and the maximization of mean throughput. Note that the real world TCP window size process is in discrete space and time and also has a maximal window size threshold (see e.g. [7, 3] and references therein for more technical details).

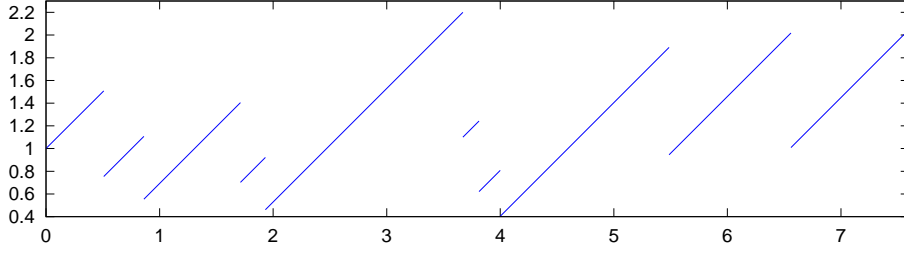


Figure 1: Typical sample path of X generated by (1), with $H = \delta_{1/2}$ and $X_0 = 1$.

Despite the apparent simplicity of the dynamics (1), the study of the long time behavior of X is not easy, mainly because the jump rate depends on the position x of the process. The sample paths of X are well defined on $[0, \infty)$ (absence of explosion), are càdlàg and piecewise linear deterministic with slope 1. The process X is irreversible since time reversed sample paths are not sample paths. This makes Meyn-Tweedie-Lyapunov techniques inefficient for the derivation of quantitative Wasserstein exponential ergodicity.

Let us explain how the trajectories of the process X may be constructed. The *jump rate* (or *jump intensity*) of X is given by $\lambda(x) = x$ for every $x \in [0, \infty)$. If $X_0 = x$ then the process will experience its first jump at a random time T solution of

$$\int_0^T \lambda(X_s) ds = E,$$

where E is an exponential random variable of unit mean. Since the trajectories of X are piecewise deterministic with slope 1, this is nothing else but

$$\int_0^T \lambda(x + s) ds = E,$$

which leads to $T = \sqrt{x^2 + 2E} - x$. Then, the sample paths of the process X generated by (1) may be constructed recursively as follows. Let X_0 be its non-negative random initial position, $(E_n)_{n \geq 1}$ be a sequence of i.i.d. exponential random variables of unit mean, and $(Q_n)_{n \geq 1}$ be a sequence of i.i.d. random variables of law H . Assume that X_0 , $(E_n)_{n \geq 1}$ and $(Q_n)_{n \geq 1}$ are independent. We define by induction the jump times $(T_n)_{n \geq 1}$ and the

positions just after the jumps $(X_{T_n})_{n \geq 1}$ as

$$T_n = T_{n-1} + \sqrt{X_{T_{n-1}}^2 + 2E_n} - X_{T_{n-1}} \quad \text{and} \quad X_{T_n} = Q_n \sqrt{X_{T_{n-1}}^2 + 2E_n}. \quad (2)$$

If we set $T_0 = 0$, then for every $n \geq 0$ and $t \in [T_n, T_{n+1})$, we have $X_t = X_{T_n} + t - T_n$ and in particular, $X_{T_{n+1}} = Q_n X_{T_n}$. For every $t \geq 0$, one can also write the series representation

$$X_t = \sum_{n=0}^{\infty} (X_{T_n} + t - T_n) \mathbb{1}_{[T_n, T_{n+1})}(t).$$

The sequence $\hat{X} = (X_{T_n})_{n \geq 0}$ is the *embedded chain* of X . According to (2), this discrete time Markov chain with state space $[0, \infty)$ satisfies the recursion

$$\hat{X}_{n+1}^2 = Q_{n+1}^2 (\hat{X}_n^2 + 2E_{n+1}). \quad (3)$$

Thus, the embedded chain \hat{X} is the square root of a first order auto-regressive process with non-Gaussian innovations $(2Q_n^2 E_n)_{n \geq 1}$ and random coefficients $(Q_n^2)_{n \geq 1}$, as already observed in [3]. The embedded chain \hat{X} is homogeneous, and its transition kernel K is given, for any $x \geq 0$ and every bounded measurable $f : [0, \infty) \rightarrow \mathbb{R}$, by the formula

$$K(f)(x) = \int_0^\infty f(y) K(x, dy) = \mathbb{E} \left[f \left(Q \sqrt{x^2 + 2E} \right) \right] \quad (4)$$

where E is an exponential random variable of unit mean and Q is a random variable of law H independent of E . We show in Section 3 that the embedded Markov chain \hat{X} admits a unique invariant probability measure ν , and converges in law to ν given any (random) initial value \hat{X}_0 . Similarly, the continuous time process X admits a unique invariant probability measure μ , and converges in law to μ , for any (random) initial value X_0 . Some properties of μ and ν are considered in Section 3.

Our aim is to study the long time behavior of X . We provide quantitative bounds in terms of Wasserstein coupling distance. Recall that for every $p \geq 1$, the W_p Wasserstein distance between two laws μ and ν on \mathbb{R} with finite p^{th} moment is defined by

$$W_p(\mu, \nu) = \left(\inf_{\Pi} \int_{\mathbb{R}^2} |x - y|^p \Pi(dx, dy) \right)^{p^{-1}} \quad (5)$$

where the infimum runs over all coupling of μ and ν . In other words, Π runs over the convex set of laws on \mathbb{R}^2 with marginals μ and ν , see e.g. [15, 16]. There is a dual formulation of Wasserstein distances. If $p = 1$, this writes as:

$$W_1(\mu, \nu) = \sup_{\|f\|_{\text{Lip}} \leq 1} \left(\int f d\mu - \int f d\nu \right) \quad \text{where} \quad \|f\|_{\text{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|}. \quad (6)$$

The jump part of L ensures that the process will remain essentially in a compact set. The jumps act in a way like a confining potential. On the other hand, the jump rate is small when the process is close to the origin. This prevents the decay of the Wasserstein distance to be exponential for small times (see the remark below for more details).

Remark 1.1 (Role of the drift term). *Let us consider the continuous time Markov process $Z = (Z_t)_{t \geq 0}$ with state space $[0, \infty)$, initial condition $Z_0 = 1$, and infinitesimal generator*

$$Af(z) = z(f(z/2) - f(z)).$$

We denote by $(T_n)_{n \geq 0}$ the sequence of jump times of Z , with the convention $T_0 = 0$. The holding times are given by $\tau_n = T_n - T_{n-1}$ for every $n \geq 1$. The process Z admits a unique invariant law which is δ_0 . For any $n \geq 1$, we have $Z_{T_n} = 2^{-n}$ and τ_n follows an exponential law of mean 2^{n-1} . Thus, for any $n \geq 1$ and $t > 0$, we can write

$$\mathbb{P}(Z_t \geq 2^{-n}) = \mathbb{P}(T_n > t) \geq \mathbb{P}(\tau_n \geq t) = e^{-2^{-(n-1)}t}.$$

By choosing $t = 2^{n-1}$ and by using Markov's inequality, we obtain

$$\lim_{n \rightarrow \infty} 2^{n-1} \mathbb{E}(Z_{2^{n-1}}) \geq (2e)^{-1} > 0.$$

As a consequence, $t \mapsto t\mathbb{E}(Z_t)$ does not converge to 0 and Z_t converges to 0 as t tends to ∞ with an at most linear speed. Of course, this dynamics without drift is very different from the TCP dynamics (1) since it is neither irreducible nor recurrent. This elementary example without drift shows that the drift term in (1) is crucial for the exponential decay phenomenon (after some “warmup”). The drift maintains the process away from zero, i.e. away from a slow zone (rare jumps).

Our main result is the following quantitative estimate.

Theorem 1.2 (Wasserstein exponential ergodicity). *There exist explicit constants $c \geq 1$ and $\rho > 0$, see (9) or (10-11), such that for any processes $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ generated by (1) with finite first moment at initial time, and for any $t > 0$, we have*

$$W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leq \min(1, ce^{-\rho t}) W_1(\mathcal{L}(X_0), \mathcal{L}(Y_0)).$$

In particular, when Y_0 follows the invariant law μ of (1), we get for every $t \geq 0$

$$W_1(\mathcal{L}(X_t), \mu) \leq \min(1, ce^{-\rho t}) W_1(\mathcal{L}(X_0), \mu).$$

The jump rate λ goes to infinity at infinity, and the following result shows that the convergence to equilibrium is in fact uniform over the starting laws, as it could be for a process living in a compact set.

Theorem 1.3 (Strong ergodicity). *Let $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ be two processes generated by (1) with arbitrary initial laws $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$. Let c and ρ be the constants given by (9) or (10-11), and let $d = \sqrt{1 - \mathbb{E}(Q)}$ where Q is a random variable of law H . Then for every $t > s > 0$, one has*

$$W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leq \frac{2ce^{\rho s}}{d \tanh(ds)} e^{-\rho t}.$$

Structure of the paper

In Section 2, we consider the simpler process generated by

$$L(f)(x) = f'(x) + \lambda \int_0^1 (f(hx) - f(x)) H(dh) \quad (7)$$

where λ is this time a positive real number. The main difference with the original dynamics (1) is that the jump rate is constant: jumps occur at Poissonian times. For this simpler process, we derive easily the following Theorem, which states an exponential ergodicity in all Wasserstein distances.

Theorem 1.4 (Wasserstein Exponential Ergodicity for constant jump rate). *Let $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ be two processes generated by (7). Assume that $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$ have finite p^{th} moment for some real $p \geq 1$. If one defines $\theta_p = \lambda(1 - \mathbb{E}(Q^p))$ with $Q \sim H$ then for every $t \geq 0$,*

$$W_p(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leq W_p(\mathcal{L}(X_0), \mathcal{L}(Y_0)) e^{-p^{-1}\theta_p t}.$$

In Section 3 we establish the geometric convergence to equilibrium of the embedded Markov chain \hat{X} for any Wasserstein distance. We also obtain non-asymptotic and Gaussian concentration bounds for the ergodic theorem when H is a Dirac mass.

In Section 4, in order to get the exponential convergence to equilibrium, we construct and study a *coupling*, i.e. a Markov process on $[0, \infty)^2$ for which the marginal components are generated by (1), with prescribed initial laws. The infinitesimal generator \bar{L} of this coupling is defined for every smooth $f : [0, \infty)^2 \rightarrow \mathbb{R}$ by

$$\bar{L}(f)(x, y) = \text{div}(f)(x, y) + x \int_0^1 (f(hx, hy)p(x, y) + f(hx, y)q(x, y) - f(x, y)) H(dh) \quad (8)$$

if $x \geq y$ and

$$\bar{L}(f)(x, y) = \text{div}(f)(x, y) + y \int_0^1 (f(hx, hy)p(x, y) + f(x, hy)q(x, y) - f(x, y)) H(dh)$$

if $x \leq y$, where $\text{div}(f) = \partial_1 f + \partial_2 f$ and where $p(x, y)$ and $q(x, y) = 1 - p(x, y)$ are chosen to recover the prescribed marginals (see Section 4). This coupling is the only one such that the lower component never jump alone. Let us give the pathwise interpretation of this coupling. All the heuristic statements below will be made more precise in Section 4. The positions of both “components” increase linearly with speed 1. The jump rate is an increasing function of the position. Thus, “the higher a component is, the sooner it will jump”. The dynamics of the couple of components is the following:

1. After an “appropriate” time which depends only on the initial position of the upper component, this one jumps.
2. Simultaneously, the other one “tosses an appropriate coin” whose probability of success depends on the positions on the two components to decide whether or not it jumps too.
3. In the case of joint jumps, both components use the same multiplicative factor.

4. Then, we repeat these three first steps again and again...

This coupling provides the following quantitative exponential upper bound.

Theorem 1.5 (Existence of a coupling with exponential decay). *Given $x, y \geq 0$, there exists a Markov process $(X, Y) = ((X_t, Y_t))_{t \geq 0}$ on $[0, +\infty)^2$ such that $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ are generated by (1), with $X_0 = x$ and $Y_0 = y$, and for every $t \geq 0$,*

$$\mathbb{E}(|X_t - Y_t|) \leq \min(1, c|x - y|e^{-\rho t}).$$

In the general case, these constants can be chosen as

$$\rho = \frac{1}{2} \mathbb{E} \left(\frac{1 - Q}{1 + Q} \right) \quad \text{and} \quad c = 2 \quad (9)$$

where $Q \sim H$. In the particular case when H is the Dirac mass at point $\delta \in [0, 1)$, these constants can be improved as follows:

$$\rho = \frac{2 \sinh(1/2) \sqrt{1 - \delta}}{\sqrt{(e - 1)\delta + 1} + 1} \geq \frac{1 - \delta}{2} \quad (10)$$

and

$$c = \frac{e - 1}{\sqrt{(e - 1)\delta + 1} + 1} + 1 \leq \frac{e + 1}{2}. \quad (11)$$

Remark 1.6 (Numerical values for the standard case). *If H is the Dirac mass at point $1/2$ then (10) and (11) become $\rho \simeq 0.31$ and $c \simeq 1.73$.*

Theorem 1.2 is a straightforward consequence of Theorem 1.5. Indeed, if $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ are two processes generated by (1) and if Π is a coupling of $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$, we deduce from Theorem 1.5 that, for every $t \geq 0$,

$$\begin{aligned} W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) &\leq \int_{[0, \infty)^2} \mathbb{E}(|X_t - Y_t| \mid X_0 = x, Y_0 = y) \Pi(dx, dy) \\ &\leq ce^{-\rho t} \int_{[0, \infty)^2} |x - y| \Pi(dx, dy). \end{aligned}$$

Then we conclude the proof of Theorem 1.2 by taking the infimum over Π . The proofs of Theorem 1.5 and 1.3 are given in Section 4.

Open questions and further remarks

The determination of the best constants and/or rates in Theorem 1.5 remains an interesting open problem. More generally, one may study their behavior for the process generated with a jump intensity of the form $\lambda(x) = \beta x$ for every $x \geq 0$, where $\beta > 0$ is an additional parameter. The incorporation of a maximal window size is also of interest (recall that the real world TCP features such a tunable maximal window size). Even more generally, and beyond the AIMD dynamics, one may investigate the long time behavior of jump processes over vector fields, which are still in the PDMP class. One may also study the long time behavior of interacting processes associated to (1) or (7), for instance Mac Kean-Vlasov mean field interactions as in [4]. The statistical estimation of λ from the observation of a sample path of the process may be also of interest.

2 When the jump intensity is constant

As already mentioned in the introduction, the analysis of the TCP window size dynamics (1) is tricky because the jump intensity is not constant and depends on the position of the process. In this section, we consider the case of the following simpler dynamics, considered recently in [11]. The window size is modeled by a Markov process $X = (X_t)_{t \geq 0}$ that increases linearly with rate one. Congestion signals arrive according to a Poisson process with constant rate $\lambda > 0$, and upon receipt of the k^{th} signal, the window size is reduced by multiplication with a random variable Q_k . We assume that $(Q_k)_{k \geq 0}$ is a sequence of i.i.d. random variables of law H with support in $[0, 1)$. In other words, the process X is generated by (7) with a constant jump intensity λ . One of the main results of [11] is an explicit formula for the transient moments of X . Namely, [11, Theorem 4 p. 172] states that for every $n \geq 0$, every $x \geq 0$, and every $t \geq 0$,

$$\mathbb{E}((X_t)^n | X_0 = x) = \frac{n!}{\prod_{k=1}^n \theta_k} + n! \sum_{m=1}^n \left(\sum_{k=0}^m \frac{x^k}{k!} \prod_{\substack{j=k \\ j \neq m}}^n \frac{1}{\theta_j - \theta_m} \right) e^{-\theta_m t} \quad (12)$$

where for every real or integer $p \geq 1$ the quantity θ_p is as in our Theorem 1.4. This formula appears also in the literature on Lévy processes and their exponential functionals, see for instance [2]. In fact, this formula may be actually obtained by the following elementary induction. If we define $f_n : x \mapsto x^n$ and $\alpha_{x,n}(t) = \mathbb{E}((X_t)^n | X_0 = x)$ then we have

$$L f_n(x) = n f_{n-1}(x) + \lambda f_n(x) \int_0^1 (y^n - 1) H(dy) = n f_{n-1}(x) - \theta_n f_n(x)$$

and thus, by definition of L and $\alpha_{x,n}$,

$$\alpha'_{x,n}(t) = \mathbb{E}(L(f_n)(X_t) | X_0 = x) = -\theta_n \alpha_{x,n}(t) + n \alpha_{x,n-1}(t).$$

This is equivalent to the recursion formula

$$\alpha_{x,n}(t) e^{\theta_n t} = x^n + n \int_0^t \alpha_{x,n-1}(s) e^{\theta_n s} ds$$

which leads to the formula for the transient moments (12).

Let us now establish Theorem 1.4.

Proof of Theorem 1.4. Let $N = (N_t)_{t \geq 0}$ be a Poisson process with constant intensity λ and $Q = (Q_k)_{k \geq 1}$ be i.i.d. random variables with law H , independent of N . For any $x, y \geq 0$, let us consider the processes $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ starting respectively at x and y at time 0, that jump when N does, with a multiplicative factor Q_k for the k^{th} jump, and increase linearly with slope one between these jumps. It is quite clear that these processes are generated by (7). Moreover, between jumps, $|X_t - Y_t|$ remains constant and

at the k^{th} jump this quantity is multiplied by Q_k . Thus for every $t \geq 0$ and $p \geq 0$,

$$\begin{aligned}\mathbb{E}(|X_t - Y_t|^p) &= \sum_{k=0}^{\infty} \mathbb{E}(|X_t - Y_t|^p \mathbf{1}_{\{N_t=k\}}) \\ &= |x - y|^p \sum_{k=0}^{\infty} \mathbb{E}(Q^p)^k \mathbb{P}(N_t = k) \\ &= |x - y|^p e^{-\lambda t(1 - \mathbb{E}(Q^p))}.\end{aligned}$$

As a consequence, if $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ are now two processes generated by (7) with a constant jump intensity λ and arbitrary initial laws, we obtain that, for any coupling Π of their initial law $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$, any $t \geq 0$, and any $p \geq 1$,

$$W_p(\mathcal{L}(X_t), \mathcal{L}(Y_t))^p \leq e^{-\theta_p t} \int_{[0, \infty)^2} |x - y|^p \Pi(dx, dy).$$

Taking the infimum over Π concludes the proof. \square

3 Embedded chain

It is shown in [3, Proposition 8], by Laplace transform inversion, that if H is a Dirac mass at point $\delta \in (0, 1)$, the invariant measure of the embedded chain $\nu = \nu_\delta$ has Lebesgue density

$$x \geq 0 \mapsto \frac{1}{\prod_{n=1}^{\infty} (1 - \delta^{2n})} \sum_{n=1}^{\infty} \frac{(-1)^{n-1} \delta^{-2n}}{\prod_{k=1}^{n-1} |1 - \delta^{-2k}|} x e^{-\delta^{-2n} x^2 / 2}. \quad (13)$$

It is unimodal, of order $O(x \exp(-\delta^2 x^2 / 2))$ when $x \rightarrow \infty$, and all its derivatives vanish at $x = 0$. Following [3, Section 4.1.3], the invariant law ν of \hat{X} is linked to the invariant law μ of X by the following formula, valid for any measurable function $f : [0, \infty) \rightarrow [0, \infty)$,

$$\int_0^{\infty} f(x) \mu(dx) = \frac{1}{\mathbb{E}(\sqrt{Z^2 + 2E} - Z)} \mathbb{E} \left(\int_Z^{\sqrt{Z^2 + 2E}} f(x) dx \right) \quad (14)$$

where E is an exponential random variable of unit mean and Z is a random variable of law ν independent of E . The formula (14) is a classical consequence of the Birkhoff-von Neumann ergodic theorem and (2). It allows for instance to “compute” the mean of μ as

$$\int_0^{\infty} x \mu(dx) = \frac{\mathbb{E}(\frac{1}{2}(Z^2 + 2E) - \frac{1}{2}Z^2)}{\mathbb{E}(\sqrt{Z^2 + 2E} - Z)} = \frac{1}{\mathbb{E}(\sqrt{Z^2 + 2E} - Z)}.$$

The formula (14) is also used in [3, Proposition 10] in the case where H is a Dirac mass at point δ to obtain the Lebesgue density of μ_δ from the one of ν_δ , namely

$$x \geq 0 \mapsto \frac{\sqrt{2/\pi}}{\prod_{n=0}^{\infty} (1 - \delta^{2n+1})} \sum_{n=0}^{\infty} \frac{(-1)^n \delta^{-2n}}{\prod_{k=1}^n |1 - \delta^{-2k}|} e^{-\delta^{-2n} x^2 / 2}.$$

It is unimodal, of order $O(\exp(-x^2/2))$ when $x \rightarrow \infty$, and all its derivatives vanish at $x = 0$. Figure 2 gives an idea of the shape of the Lebesgue densities of μ_δ and ν_δ .

What is the behavior of the tails of ν and μ when H is no longer a Dirac mass? We will see that ν is more sensitive than μ to this generalization.

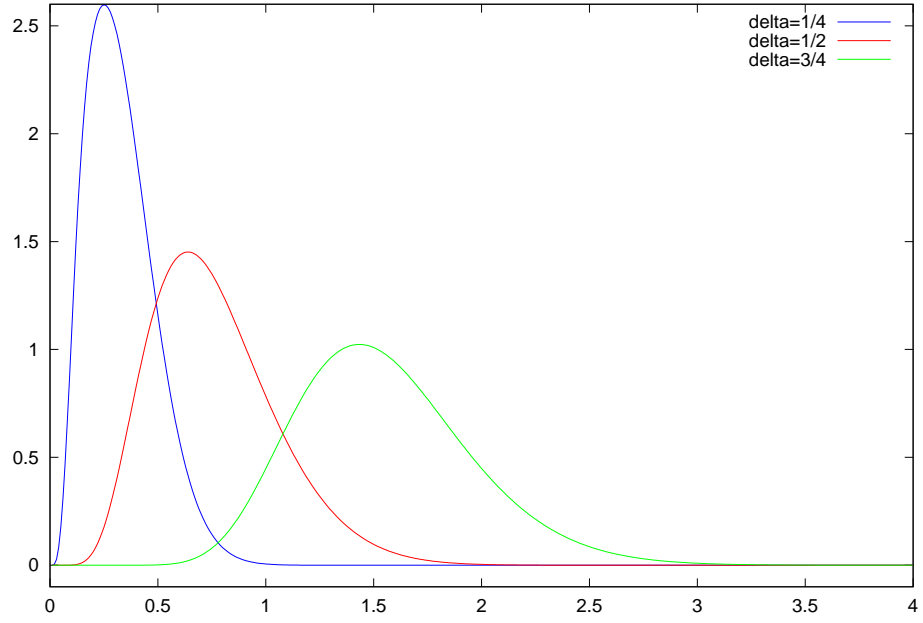
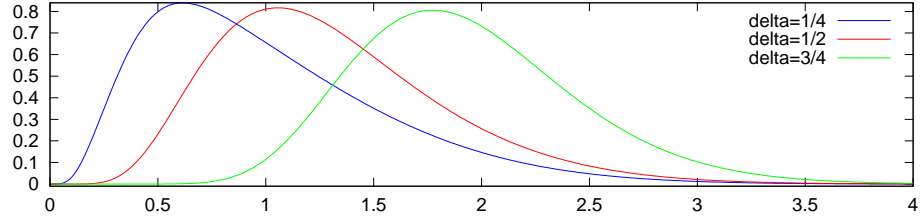


Figure 2: Densities of μ (top) and ν (bottom) when H is a Dirac mass at point δ .

Theorem 3.1 (Convergence of the Embedded chain). *Given any \hat{X}_0 , the embedded Markov chain $\hat{X} = (\hat{X}_n)_{n \geq 0}$ associated to the dynamics (1) converges in distribution to the law of the random variable*

$$\left(2 \sum_{n=1}^{\infty} Q_1^2 \cdots Q_n^2 E_n\right)^{1/2}$$

which is a.s. finite, where E_1, E_2, \dots and Q_1, Q_2, \dots are independent sequences of i.i.d. random variables following respectively the exponential law of unit mean and the law H which appear in (1). In particular, ν is the unique invariant law of \hat{X} and

$$\int e^{sx^2} \nu(dx) = \mathbb{E} \left(\frac{1}{\prod_{n=1}^{\infty} (1 - 2sQ_1^2 \cdots Q_n^2)} \right),$$

which is finite if $2sq^2 < 1$ and infinite if $2sq^2 > 1$, where $q = \inf \{x, \mathbb{P}(Q > x) = 1\} \leq 1$.

Proof. An induction based on (3) gives for every $n \geq 0$,

$$\hat{X}_{n+1}^2 = Q_1^2 \cdots Q_{n+1}^2 \hat{X}_0^2 + 2 \sum_{k=1}^{n+1} Q_k^2 \cdots Q_{n+1}^2 E_k$$

Since the common law H of the random variables Q_1, Q_2, \dots is supported in $[0, 1)$, the sequence $(Q_1^2 \cdots Q_n^2)_{n \geq 1}$ is decreasing and its limit $Z = \lim_{n \rightarrow \infty} Q_1^2 \cdots Q_n^2$ is a well defined random variable on $[0, 1)$. Moreover, by the monotone convergence theorem, $\mathbb{E}(Q_1^2 \cdots Q_n^2)$ decreases to $\mathbb{E}(Z)$. Additionally, $\mathbb{E}(Q_1^2 \cdots Q_n^2) = \mathbb{E}(Q_1^2)^{n+1}$ goes to zero and therefore $\mathbb{E}(Z) = 0$. Thus, $Z = 0$ a.s., and $Q_1^2 \cdots Q_{n+1}^2 \hat{X}_0^2$ goes to zero a.s. This shows in some sense that the embedded chain \hat{X} forgets its starting point X_0 a.s.

Let us show now that the sequence $(B_n)_{n \geq 1}$ defined by $B_{n+1} = 2 \sum_{k=1}^{n+1} Q_k^2 \cdots Q_{n+1}^2 E_k$ converges in distribution. The independence of the i.i.d. sequences (E_k) and (Q_k) gives

$$B_{n+1} \stackrel{d}{=} 2 \sum_{k=1}^{n+1} Q_1^2 \cdots Q_k^2 E_k.$$

Since the common law H of Q_1, Q_2, \dots is supported in $[0, 1)$, the monotone convergence theorem implies that the random variable $2 \sum_{n=1}^{\infty} Q_1^2 \cdots Q_n^2 E_n$ is a.s. finite and is the a.s. limit of $2 \sum_{k=1}^{n+1} Q_1^2 \cdots Q_k^2 E_k$ as n grows to infinity. Consequently, we obtain by Slutsky's lemma that $\hat{X} = (\hat{X}_n)_{n \geq 0}$ converges in distribution to the law ν of $(2 \sum_{n=1}^{\infty} Q_1^2 \cdots Q_n^2 E_n)^{1/2}$. Beware that $(B_n)_{n \geq 1}$ does not converge a.s.! Note that since $\mathbb{E}(Q_1^2) < 1$, we have, if $\mathbb{E}(X_0^2) < \infty$,

$$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{X}_n^2) = \lim_{n \rightarrow \infty} \mathbb{E}(B_{n+1}) = \sum_{k=1}^{\infty} \mathbb{E}(Q_1^2)^k = \frac{\mathbb{E}(Q_1^2)}{1 - \mathbb{E}(Q_1^2)}.$$

Note also that since $B_{n+1} = Q_{n+1}^2 (B_n + 2E_{n+1})$, we can write, for every real $s \geq 0$,

$$\varphi_{n+1}(s) = \mathbb{E}(e^{-sB_{n+1}}) = \mathbb{E} \left(\frac{e^{-sQ_{n+1}^2 B_n}}{1 + 2sQ_{n+1}^2} \right) = \mathbb{E} \left(\frac{\varphi_n(sQ_1^2)}{1 + 2sQ_1^2} \right).$$

Also, $\varphi_{n+1} = \mathcal{A}\varphi_n$ where \mathcal{A} is the linear operator which maps any bounded continuous $\varphi : [0, \infty) \rightarrow \mathbb{R}$ to the bounded continuous $\mathcal{A}\varphi : [0, \infty) \rightarrow \mathbb{R}$ defined for every $s \geq 0$ by

$$(\mathcal{A}\varphi)(s) = \mathbb{E}\left(\frac{\varphi(sQ_1^2)}{1 + 2sQ_1^2}\right).$$

A simple recursion gives, for every $s \geq 0$ and $n \geq 1$,

$$\varphi_{n+1}(s) = (\mathcal{A}^n \varphi_1)(s) = \mathbb{E}\left(\frac{\varphi_1(sQ_1^2 \cdots Q_n^2)}{(1 + 2sQ_1^2) \cdots (1 + 2sQ_n^2)}\right).$$

Note that the function $t \mapsto \varphi_1(t) = \mathbb{E}(e^{-tQ_1^2 E_1})$ is bounded and $\varphi_1(0) = 1$. Now, we have already seen that $Q_1^2 \cdots Q_n^2 \rightarrow 0$ almost surely. On the other hand, for every $s \geq 0$, the infinite product $\prod_{n=1}^{\infty} (1 + 2sQ_n^2)$ converges almost surely iff the series $\sum_{n=1}^{\infty} Q_n^2$ converges almost surely, which is actually the case, as one can check by using the monotone convergence theorem and the fact that the law H of Q_1, Q_2, \dots is supported in $[0, 1]$. Consequently, by Lebesgue's dominated convergence theorem, for every $s \geq 0$,

$$\lim_{n \rightarrow \infty} \varphi_{n+1}(s) = \mathbb{E}\left(\frac{1}{\prod_{n=1}^{\infty} (1 + 2sQ_n^2)}\right) = \varphi(s).$$

At last, one has to notice that

$$\int e^{sx^2} \nu(dx) = \varphi(-s).$$

□

The previous results suggest that the tail of the invariant law of the embedded chain looks like the tail of a Gaussian law with variance $1/q^2$. This is in accordance with the case when H is a Dirac mass. What about the invariant law μ of the continuous time process? We have seen in Section 2 how the generator helps to quickly compute the transient moments of X when the jumps intensity is constant and equal to $\lambda > 0$. The method does not work otherwise. However, in the non-constant case (1), one can compute the moments of the invariant law μ of the process.

Lemma 3.2 (Moments of μ). *All the moments $(\alpha_n)_{n \geq 0}$ of the invariant measure μ of the process X associated to (1) are finite and satisfy*

$$\alpha_{2n} = \frac{(2n)!}{2^n n!} \left(\prod_{k=1}^n \kappa_{2k-1} \right)^{-1} \quad \text{and} \quad \alpha_{2n+1} = \alpha_1 2^n n! \left(\prod_{k=1}^n \kappa_{2k} \right)^{-1}$$

where $\kappa_n = 1 - \int_0^1 h^n H(dh)$. Moreover, $\mathbb{E}_\mu(e^{sX^2})$ is finite if and only if $2s < 1$.

Proof. If $f_n : x \geq 0 \mapsto x^n$ then the n^{th} moment of μ is $\alpha_n = \mathbb{E}_\mu(f_n)$ and for every $n \geq 1$ we have

$$0 = \mathbb{E}_\mu(Lf_n) = n\alpha_{n-1} - \kappa_n \alpha_{n+1}.$$

From this equation, we obtain the explicit formula for α_n by a straightforward recursion. This is in accordance with the expression of the density of μ when H is a Dirac mass. At last, for all $s \in \mathbb{R}$,

$$\mathbb{E}_\mu(e^{sX^2}) = \sum_{n=0}^{\infty} \frac{\alpha_{2n}}{n!} s^n \in [0, +\infty].$$

The convergence radius of this series is $1/2$ since

$$\frac{\alpha_{2n+2}/(n+1)!}{\alpha_{2n}/n!} = 2 \frac{(2n+2)(2n+1)}{4(n+1)^2} \frac{1}{\kappa_{2n+1}} \xrightarrow{n \rightarrow \infty} \frac{2}{1 - H(\{1\})} = 2.$$

Since all the terms of the series are nonnegative, $\mathbb{E}_\mu(e^{X^2/2})$ is infinite. \square

Hence, for any law H on $[0, 1)$, the tail of μ looks like a Gaussian one with unit variance.

Let us now turn to the convergence to equilibrium for the embedded chain. If $\mathbb{E}(\hat{X}_0^2)$ is finite then by induction from (3) we get, for every $n \geq 0$, and for a random variable Q of law H ,

$$\mathbb{E}(\hat{X}_n^2) = \frac{2\mathbb{E}(Q^2)}{1 - \mathbb{E}(Q^2)} + \mathbb{E}(Q^2)^{n+1} \left(\mathbb{E}(\hat{X}_0^2) - \frac{\mathbb{E}(Q^2)}{1 - \mathbb{E}(Q^2)} \right).$$

Thus, the second moment of \hat{X}_n converges to the second moment $2\mathbb{E}(Q^2)(1 - \mathbb{E}(Q^2))^{-1}$ of ν_δ . It is known that the convergence in Wasserstein W_p distance is equivalent to weak convergence together with convergence of moments up to p , see [16]. Also, it is quite natural to ask for a quantitative exponential Wasserstein ergodicity for the embedded chain \hat{X} . The following theorem gives a general answer which complements Theorem 3.1.

Theorem 3.3 (Wasserstein exponential ergodicity for the generic embedded chain). *Let $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ be two processes generated by (1). Assume that $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$ have finite p^{th} moment for some real $p \geq 1$. Let \hat{X} and \hat{Y} be the embedded chains of X and Y . Then, for any $n \geq 0$, with a random variable $Q \sim H$,*

$$W_p(\mathcal{L}(\hat{X}_n), \mathcal{L}(\hat{Y}_n)) \leq \mathbb{E}(Q^p)^{n/p} W_p(\mathcal{L}(X_0), \mathcal{L}(Y_0)).$$

In particular, if ν is the invariant law of \hat{X} then

$$W_p(\mathcal{L}(\hat{X}_n), \nu) \leq \mathbb{E}(Q^p)^{n/p} W_p(\mathcal{L}(X_0), \nu).$$

Proof. It is sufficient to provide a good coupling. Let $x \geq 0$ and $y \geq 0$ be two non-negative real numbers, and let $(E_n)_{n \geq 1}$ and $(Q_n)_{n \geq 1}$ be two independent sequences of i.i.d. random variables with respective laws the exponential law of unit mean and the law H which appears in (1). Let \hat{X} and \hat{Y} be the discrete time Markov chains on $[0, \infty)$ defined by

$$\begin{aligned} \hat{X}_0 &= x \quad \text{and} \quad \hat{X}_{n+1} = Q_{n+1} \sqrt{\hat{X}_n^2 + 2E_{n+1}} \quad \text{for any } n \geq 0 \\ \hat{Y}_0 &= y \quad \text{and} \quad \hat{Y}_{n+1} = Q_{n+1} \sqrt{\hat{Y}_n^2 + 2E_{n+1}} \quad \text{for any } n \geq 0. \end{aligned}$$

By the analogue of (2) for (7), the law of \hat{X} (respectively \hat{Y}) is the law of the embedded chain of a process generated by (1) and starting from x (respectively y). Now, for any $p \geq 1$, since $x \mapsto \sqrt{x^2 + a}$ is a 1-Lipschitz function on $[0, \infty)$ for any $a \geq 0$, we get

$$\begin{aligned} \mathbb{E}\left(\left|\hat{X}_{n+1} - \hat{Y}_{n+1}\right|^p\right) &= \mathbb{E}\left(Q_{n+1}^p \left|\sqrt{\hat{X}_n^2 + 2E_{n+1}} - \sqrt{\hat{Y}_n^2 + 2E_{n+1}}\right|^p\right) \\ &\leq \mathbb{E}\left(Q_{n+1}^p \left|\hat{X}_n - \hat{Y}_n\right|^p\right) = \mathbb{E}(Q_{n+1}^p) \mathbb{E}\left(\left|\hat{X}_n - \hat{Y}_n\right|^p\right) \end{aligned}$$

A straightforward recurrence leads to

$$\mathbb{E}\left(\left|\hat{X}_n - \hat{Y}_n\right|^p\right) \leq \mathbb{E}(Q_1^p)^n |x - y|^p.$$

This gives the desired inequality when the initial laws are Dirac masses. The general case follows by integrating this inequality with respect to couplings of the initial laws. \square

Let us now investigate some properties of the kernel K defined by (4) that will be used to provide concentration bounds for the ergodic theorem. In the sequel, the law H is a Dirac mass at $\delta \in (0, 1)$. Note that the embedded chain associated to the Dirac mass at 0 is the constant Markov chain equal to 0... In order to derive sub-Gaussian concentration bounds, we shall use a powerful tool: the Gross (or logarithmic Sobolev) inequality.

Definition 3.4 (Gross inequality). *A law η on \mathbb{R}^d satisfies a Gross (or logarithmic Sobolev [5, 6]) inequality with constant $c > 0$ when for any smooth compactly supported $f : \mathbb{R}^d \rightarrow \mathbb{R}$,*

$$\int f^2 \log(f^2) d\eta - \int f^2 d\eta \log \int f^2 d\eta \leq c \int |\nabla f|^2 d\eta.$$

We denote by $\text{GROSS}(\eta) \in (0, \infty]$ the least constant for which this holds true.

If $F \cdot \eta$ is the image of η by F then $\text{GROSS}(F \cdot \eta) \leq \text{GROSS}(\eta) \|F\|_{\text{LIP}}^2$. The Gross inequality contains an information on Gaussian concentration of measure: the function $x \mapsto e^{ax^2}$ is η -integrable as soon as $a < 1/\text{GROSS}(\eta)$. Moreover, if η has covariance Σ with spectral radius $\rho(\Sigma)$ then $2\rho(\Sigma) \leq \text{GROSS}(\eta)$ and equality is achieved when η is Gaussian. Furthermore, for any α -Lipschitz function $f : \mathbb{R} \rightarrow \mathbb{R}$ and any $\lambda > 0$,

$$\mathbb{E}_\eta\left(e^{\lambda f}\right) \leq e^{C\alpha^2\lambda^2/4} e^{\lambda \mathbb{E}_\eta f} \quad (15)$$

as soon as $C \geq \text{GROSS}(\eta)$. This means that η satisfies a *sub-Gaussian concentration of measure* for Lipschitz functions. For more details, see e.g. [16, 9] and references therein.

Theorem 3.5 (Properties of the kernel of the embedded chain). *Let \hat{X} be the embedded chain associated to (1) with transition kernel (4). Assume that H is the Dirac mass at point $\delta \in [0, 1)$. If f is a 1-Lipschitz function from $[0, +\infty)$ to \mathbb{R} , then $x \mapsto K(f)(x)$ is a δ -Lipschitz function from $[0, +\infty)$ to \mathbb{R} . Moreover, for any $x \geq 0$, the law $K(\cdot)(x)$ satisfies a Gross inequality with constant $2\delta^2$.*

Proof. For any smooth function $f : [0, \infty) \rightarrow \mathbb{R}$, we have from (4) that

$$|(Kf)'| = \delta \left| K\left(\frac{x}{\sqrt{x^2 + 2E}} f'(\delta\sqrt{x^2 + 2E})\right) \right| \leq \delta K(|f'|). \quad (16)$$

Let us show now that for every $x \geq 0$ the law $K(x, \cdot) = \mathcal{L}(\hat{X}_1 | \hat{X}_0 = x)$ satisfies a Gross inequality with constant $2\delta^2$. Since E is exponential of mean 1, the law η of $\sqrt{E/2}$ is a χ -distribution with probability density and cumulative distribution functions given by

$$g : v \mapsto 4ve^{-2v^2} \mathbf{1}_{\{v>0\}} \quad \text{and} \quad G : v \mapsto (1 - e^{-2v^2}) \mathbf{1}_{\{v>0\}}.$$

On the other hand, $2E \stackrel{d}{=} U_1^2 + U_2^2$ where U_1, U_2 are i.i.d. standard Gaussians, and thus

$$\sqrt{E/2} = \frac{1}{2} \sqrt{2E} \stackrel{d}{=} \frac{1}{2} \sqrt{U_1^2 + U_2^2}.$$

Also, η is the image of the Gaussian law $\mathcal{N}(0, I_2)$ on \mathbb{R}^2 by a $(1/2)$ -Lipschitz function, and this implies that η satisfies a Gross inequality with constant $1/2$. Moreover,

$$\begin{aligned} K(f)(x) &= \int_0^\infty f\left(\delta\sqrt{x^2 + 2u}\right) e^{-u} du \\ &= \int f(2\delta v) \frac{4ve^{-2v^2}}{e^{-x^2/2}} \mathbb{1}_{\{v > x/2\}} dv \\ &= \int f(2\delta v) \frac{g(v)}{1 - G(x/2)} \mathbb{1}_{\{v > x/2\}} dv. \end{aligned}$$

Thus, $K(\cdot)(x)$ is just the image law by the Lipschitz map $v \mapsto 2\delta v$ of the law η conditioned on $(x/2, +\infty)$. This conditional law is in turn the image of η by the function

$$t \mapsto G^{-1}(G(x) + (1 - G(x))G(t)) = G^{-1}(1 - \exp(-t^2 - x^2)) = \sqrt{x^2 + t^2}.$$

This function is 1-Lipschitz for any $x \geq 0$. Consequently, by using twice the stability of Gross inequalities by Lipschitz maps, we obtain that for every $x > 0$, the law $K(x, \cdot)$ satisfies a Gross inequality with constant $(2\delta)^2/2 = 2\delta^2$. \square

These properties of K propagate to the iterated kernels and to the invariant law ν .

Corollary 3.6 (Gross inequality for the embedded chain and its invariant law ν). *Let \hat{X} be the embedded chain associated to (1). Assume that H is the Dirac mass at point $\delta \in (0, 1)$. For every $n \geq 0$, let K^n be the iterated transition kernel of \hat{X} , defined recursively for every bounded measurable function $f : [0, \infty) \rightarrow \mathbb{R}$ by*

$$K^0(f) = f \quad \text{and} \quad K^{n+1}(f) = K(K^n(f))$$

where K is the kernel of \hat{X} as in (4). Then for every integer $n \geq 1$ and every real $x \geq 0$, the iterated kernel $K^n(x, \cdot)$ of \hat{X} satisfies a Gross inequality and

$$\text{GROSS}(K^n(x, \cdot)) \leq 2\delta^2 \frac{1 - \delta^{2n}}{1 - \delta^2}.$$

Also, the invariant law ν of \hat{X} (see theorem 3.1) satisfies a Gross inequality and

$$\text{GROSS}(\nu) \leq 2\delta^2(1 - \delta^2)^{-1}.$$

Proof. Recall that for every $n \geq 0$, $x \geq 0$, and bounded measurable $f : [0, \infty) \rightarrow \mathbb{R}$,

$$\mathbb{E}\left(f(\hat{X}_n) \mid \hat{X}_0 = x\right) = (K^n f)(x) = \int_0^\infty f(y) K^n(x, dy)$$

To show that K^n satisfies a Gross inequality, we use a semi-group decomposition technique borrowed from [12]. For any $n \geq 1$ and any smooth function $f : [0, \infty) \rightarrow \mathbb{R}$, the quantity

$$E_n(f) := K^n(f^2 \log f^2) - K^n(f^2) \log K^n(f^2)$$

is equal to the telescopic sum

$$\sum_{i=1}^n \left\{ K^i[K^{n-i}(f^2) \log K^{n-i}(f^2)] - K^{i-1}[K^{n-i+1}(f^2) \log K^{n-i+1}(f^2)] \right\}.$$

Since the measure $K(\cdot)(x)$ satisfies a Gross inequality of constant $2\delta^2$, we get, with $g_{n-i} = \sqrt{K^{n-i}(f^2)}$,

$$E_n(f) = \sum_{i=1}^n K^{i-1}[E_1(g_{n-i})] \leq 2\delta^2 \sum_{i=1}^n K^i(|\nabla g_{n-i}|^2),$$

Now, by using the commutation (16), we obtain, for all $1 \leq i \leq n$,

$$|\nabla g_{n-i}|^2 = \frac{|\nabla K^{n-i}(f^2)|^2}{4K^{n-i}(f^2)} \leq \delta^2 \frac{(K|\nabla K^{n-i-1}(f^2)|)^2}{4KK^{n-i-1}(f^2)}$$

Next, the Cauchy-Schwarz inequality

$$\frac{(Kf)^2}{K(g)} \leq K\left(\frac{f^2}{g}\right)$$

gives

$$\frac{(K|\nabla K^{n-i-1}(f^2)|)^2}{4KK^{n-i-1}(f^2)} \leq K\left(\frac{|\nabla K^{n-i-1}(f^2)|^2}{4K^{n-i-1}(f^2)}\right) = K(|\nabla g_{n-i-1}|^2).$$

From these bounds, a straightforward induction gives

$$|\nabla g_{n-i}|^2 \leq \delta^{2(n-i)} K^{n-i}(|\nabla f|^2).$$

Consequently, by putting all together, we have

$$E_n(f^2) \leq 2\delta^2 \sum_{i=0}^{n-1} \delta^{2i} K^n(|\nabla f|^2) = 2\delta^2 \frac{1 - \delta^{2n}}{1 - \delta^2} K^n(|\nabla f|^2).$$

This gives

$$\text{GROSS}(K^n) \leq 2\delta^2(1 - \delta^{2n})(1 - \delta^2)^{-1}.$$

Finally, from Theorem 3.1, K^n tends weakly to ν as n tends to infinity and thus

$$\text{GROSS}(\nu) \leq \limsup_{n \rightarrow \infty} \text{GROSS}(K^n) \leq 2\delta^2(1 - \delta^2)^{-1}.$$

□

The Gross inequality for K can also be used to derive Gaussian deviations bounds for Lipschitz observables as stated in the following result. This produces the following non asymptotic estimate for the ergodic theorem associated to the embedded chain.

Corollary 3.7 (Gaussian deviations for the ergodic theorem for the embedded chain). *Let \hat{X} be the embedded chain associated to (1) and starting from $x \geq 0$. Assume that H is the Dirac mass at point $\delta \in (0, 1)$. Then for any $u \geq 0$ and any 1-Lipschitz function $f : [0, \infty) \rightarrow \mathbb{R}$,*

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n f(\hat{X}_k) - \int f d\nu\right| \geq u + \frac{\delta}{1 - \delta} W_1(\delta_x, \nu)\right) \leq 2 \exp\left(-\frac{n(1 - \delta^2)u^2}{2\delta^2}\right).$$

Proof. We shall establish that for any $u \geq 0$ and any 1-Lipschitz function $f : [0, \infty) \rightarrow \mathbb{R}$,

$$\mathbb{P}\left(\frac{1}{n} \sum_{k=1}^n f(\hat{X}_k) - \int f d\nu \geq u + \frac{\delta}{1-\delta} W_1(\delta_x, \nu)\right) \leq \exp\left(-\frac{n(1-\delta^2)u^2}{2\delta^2}\right)$$

and the desired result follows immediately from this bound used for f and $-f$. For any 1-Lipschitz function f , any $r > 0$ and $\lambda > 0$, we have,

$$\mathbb{P}\left(\frac{1}{n} \sum_{k=1}^n f(\hat{X}_k) \geq r\right) \leq \mathbb{E}\left(e^{\lambda \sum_{k=1}^n f(\hat{X}_k)}\right) e^{-nr\lambda}.$$

Now the Markov property ensures that

$$\begin{aligned} \mathbb{E}\left(e^{\lambda \sum_{k=1}^n f(\hat{X}_k)}\right) &= \mathbb{E}\left(e^{\lambda \sum_{k=1}^{n-1} f(\hat{X}_k)} \mathbb{E}\left(e^{\lambda f(\hat{X}_n)} | X_{n-1}\right)\right) \\ &= \mathbb{E}\left(e^{\lambda \sum_{k=1}^{n-1} f(\hat{X}_k)} K\left(e^{\lambda f}\right)(X_{n-1})\right). \end{aligned}$$

From Theorem 3.5, the kernel $K(x, \cdot)$ of \hat{X} satisfies a Gross inequality with constant $2\delta^2$ for every $x \geq 0$. This inequality implies by (15) that for any c -Lipschitz function g ,

$$K\left(e^{\lambda g}\right) \leq \exp\left(\lambda K g + \frac{c^2 \delta^2 \lambda^2}{2}\right).$$

Consequently, the Laplace transform of the ergodic mean can be bounded as follows:

$$\mathbb{E}\left(e^{\lambda \sum_{k=1}^n f(\hat{X}_k)}\right) \leq \exp\left(\frac{\delta^2 \lambda^2}{2}\right) \mathbb{E}\left(e^{\lambda \sum_{k=1}^{n-2} f(\hat{X}_k)} \mathbb{E}\left(e^{\lambda(f+Kf)(\hat{X}_{n-1})} | \hat{X}_{n-2}\right)\right).$$

The commutation relation (16) ensures that $f + Kf$ is $(1 + \delta)$ -Lipschitz and then

$$\mathbb{E}\left(e^{\lambda(f+Kf)(\hat{X}_{n-1})} | \hat{X}_{n-2}\right) \leq \exp\left(\frac{(1+\delta)^2 \delta^2 \lambda^2}{2}\right) e^{\lambda(Kf+K^2f)(\hat{X}_{n-2})}.$$

A straightforward recurrence ensures that

$$\mathbb{E}\left(e^{\lambda \sum_{k=1}^n f(\hat{X}_k)}\right) \leq \exp\left(\frac{n\delta^2 \lambda^2}{2(1-\delta^2)}\right) e^{\sum_{k=1}^n K^k f(x)}.$$

Choosing $r = (1/n) \sum_{k=1}^n K^k f(x) + u$ leads to

$$\mathbb{P}\left(\frac{1}{n} \sum_{k=1}^n f(\hat{X}_k) - \frac{1}{n} \sum_{k=1}^n K^k f \geq u\right) \leq \exp\left(\frac{n\delta^2 \lambda^2}{2(1-\delta^2)} - n\lambda u\right).$$

The right hand side is minimum for $\lambda = u(\delta^{-2} - 1)$. At last, from the dual definition the Wasserstein distance (6) and Theorem 3.3, one gets

$$\frac{1}{n} \sum_{k=1}^n K^k f(x) - \int f d\nu \leq \frac{1}{n} \sum_{k=1}^n W_1(K^k(\cdot)(x), \nu) \leq \frac{\delta}{1-\delta} W_1(\delta_x, \nu).$$

□

Remark 3.8. A careful reading of the proof of Theorem 3.7 suggests that one may replace the initial law δ_x by a more general initial law provided that it satisfies a sub-Gaussian concentration for Lipschitz functions (15).

4 Coupling for exponential convergence

Consider a two dimensional process where both components are generated by (1). Since the sample paths of both components have slope 1 between jumps, the distance between them remains constant except at jump times. If the components jump together with the same factor Q , then this distance is also multiplied by Q . Thus, our strategy will be to determine the greatest simultaneous jump probability $p(x, y)$ in \overline{L} .

The section is organized as follows. We first establish elementary considerations about the laws of jump times as a function of the initial point. This allows us to construct rigorously our coupling. Then we provide the proof of Theorem 1.5. It relies on tricky controls of non-simultaneous jumps and on an original use of Gronwall's Lemma. We end up this section by the rather simple proof of Theorem 1.3.

Let $X = (X_t)_{t \geq 0}$ be a process generated by (1) starting at $x \geq 0$. As pointed out in the introduction, the law of the first jump time T_1 is the one of $\sqrt{x^2 + 2E} - x$ where E is an exponential random variable of unit mean. This law ν_x admits the density f_x defined by

$$f_x(v) = (x + v)e^{-(x+v)^2/2}e^{x^2/2}\mathbb{1}_{\{v \geq 0\}} = (x + v)e^{-v^2/2 - xv}\mathbb{1}_{\{v \geq 0\}}. \quad (17)$$

Its cumulative distribution F_x is given by

$$F_x(v) = \left(1 - e^{-(x+v)^2/2 + x^2/2}\right)\mathbb{1}_{\{v \geq 0\}} = \left(1 - e^{-v^2/2 - xv}\right)\mathbb{1}_{\{v \geq 0\}}.$$

Note that the probability density function f_x satisfies to the equation

$$f_x(v) = (x + v)(1 - F_x(v)) = \frac{f_0(x + v)}{1 - F_0(x)}.$$

Let us compare the laws of T_1 as a function of the starting point (see Figure 3).

Lemma 4.1 (Comparison of the jump time densities f_x and f_y). *Assume that $x > y \geq 0$. Then there exists a unique value $\tau(x, y)$ such that for every $v \geq 0$,*

$$f_x(v) > f_y(v) \quad \text{if and only if} \quad v < \tau(x, y).$$

Moreover, the function $x \mapsto \tau(x, y)$ is increasing on $[y, +\infty)$ whereas the function $y \mapsto \tau(x, y)$ is decreasing on $[0, x]$ and, for every $0 \leq y < x$, one has

$$\frac{\sqrt{x^2 + 4} - x}{2} \leq \tau(x, y) \leq \frac{\sqrt{y^2 + 4} - y}{2} \leq 1.$$

At last, if $q(x, y) = 1 - p(x, y)$ is the area of the set (one may take a look at Figure 3)

$$\{(u, v) \in [0, \infty) \times [0, \infty); f_y(u) \leq v \leq f_x(u)\},$$

then, we have, shortening $\tau(x, y)$ in τ ,

$$q(x, y) = e^{-\tau^2/2 - y\tau} - e^{-\tau^2/2 - x\tau}, \quad (18)$$

and

$$\partial_y q(x, y) = -\tau e^{-\tau y - \tau^2/2} < 0. \quad (19)$$

Proof. The function $h_{x,y} : [0, \infty) \rightarrow \mathbb{R}$ defined by

$$h_{x,y}(v) = \log f_x(v) - \log f_y(v) = \log \left(\frac{v+x}{v+y} \right) - (x-y)v \quad (20)$$

is decreasing and convex since

$$h'_{x,y}(v) = -(x-y) \left(\frac{1}{(x+v)(y+v)} + 1 \right) < 0 \quad \text{and} \quad h''_{x,y}(v) = \frac{x-y}{(x+v)(y+v)} > 0.$$

Since $h_{x,y}(0) > 0$ and $h_{x,y}$ goes to $-\infty$ as $v \rightarrow +\infty$, there is a unique $\tau(x,y)$ such that $h_{x,y}(\tau(x,y)) = 0$ or

$$\log \frac{\tau(x,y) + x}{\tau(x,y) + y} - (x-y)\tau(x,y) = 0. \quad (21)$$

For any $x \geq 0$, let us define

$$\tilde{\tau}(x) = (\sqrt{x^2 + 4} - x)/2.$$

The function $g : [0, x] \rightarrow \mathbb{R}$ defined by

$$g(y) = \log \left(\frac{f_y(\tilde{\tau}(x))}{f_x(\tilde{\tau}(x))} \right)$$

is increasing and $g(x) = 0$. Then $f_x(\tilde{\tau}(x)) > f_y(\tilde{\tau}(x))$ and then $\tilde{\tau}(x) \leq \tau(x,y)$. A similar argument leads to the upper bound. Let us differentiate (21) with respect to x :

$$\frac{\partial_x \tau(x,y) + 1}{\tau(x,y) + x} - \frac{\partial_x \tau(x,y)}{\tau(x,y) + y} - \tau(x,y) - (x-y)\partial_x \tau(x,y) = 0,$$

which is nothing else than

$$-\partial_x \tau(x,y)(x-y) \left(\frac{1}{(\tau(x,y) + x)(\tau(x,y) + y)} + 1 \right) = \frac{\tau(x,y)^2 + x\tau(x,y) - 1}{\tau(x,y) + x}.$$

Since $\tau(x,y) \geq \tilde{\tau}$ and $\tilde{\tau}$ is the largest root of the equation $T^2 + xT - 1 = 0$, we have $\tau(x,y)^2 + x\tau(x,y) - 1 \geq 0$ and $\partial_x \tau(x,y) \leq 0$. A similar argument leads to $\partial_y \tau(x,y) \leq 0$.

At last, since $f_y(u) \leq f_x(u)$ if and only if $u \leq \tau$, one has

$$q(x,y) = \int_0^\tau f_x(u) du - \int_0^\tau f_y(u) du = F_x(\tau) - F_y(\tau),$$

which is nothing else but (18). Moreover,

$$\partial_y q(x,y) = \partial_y \tau(f_x(\tau) - f_y(\tau)) - \partial_y F_y(\tau).$$

The fact that $f_x(\tau) = f_y(\tau)$ provides (19). □

Let us consider the functions g_1, g_2, g_3 defined by

$$\begin{aligned} g_1 : v &\mapsto \frac{\min(f_x(v), f_y(v))}{p(x,y)} \mathbb{1}_{\{v>0\}} \\ g_2 : v &\mapsto \frac{f_x(v) - f_y(v)}{q(x,y)} \mathbb{1}_{\{0<v<\tau(x,y)\}} \\ g_3 : v &\mapsto \frac{f_y(v) - f_x(v)}{q(x,y)} \mathbb{1}_{\{v>\tau(x,y)\}}. \end{aligned}$$

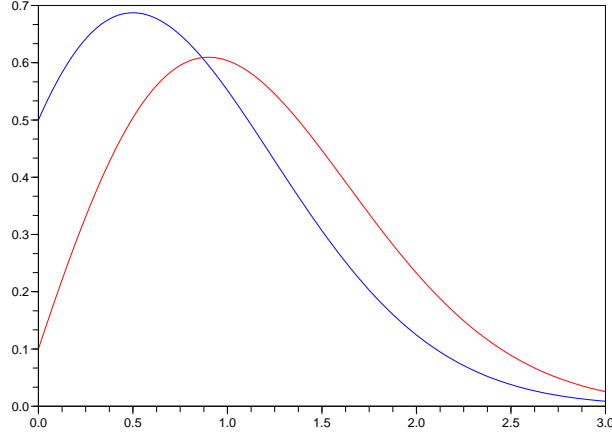


Figure 3: Plot of the densities $f_{0.5}$ and $f_{0.1}$. The abscissa of the intersection is $\tau(0.1, 0.5)$.

From Lemma 4.1, they are three probability density functions and we can write f_x and f_y as two-component mixtures:

$$f_x(v) = p(x, y)g_1(v) + q(x, y)g_2(v) \quad \text{and} \quad f_y(v) = p(x, y)g_1(v) + q(x, y)g_3(v).$$

In other words, let us consider three random variables V_1 , V_2 and V_3 with respective density functions g_1 , g_2 , and g_3 independent of U with Bernoulli law $\mathcal{B}(p(x, y))$. Then, $S = UV_1 + (1 - U)V_2$ and $T = UV_1 + (1 - U)V_3$ have respective density functions f_x and f_y and $\mathbb{P}(S = T) = p(x, y)$. This procedure is a particular case of a well known maximal coupling Lemma (for the total variation distance, see e.g. [10] for a general formulation).

We are now able to construct the couple of processes $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ starting respectively from x and y that we shall use in the proof of Theorem 1.5. Let $(E_n)_{n \geq 1}$ be a sequence of i.i.d. exponential random variables with unit mean, and let $(Q_n)_{n \geq 1}$ be a sequence of i.i.d. random variables of law H where H is as in (1), independent of $(E_n)_{n \geq 1}$. Assume that $x > y$ (since we will be interested in $|X_t - Y_t|$, this does not matter). For any t in the time interval $[0, V_1)$ where $V_1 = \sqrt{x^2 + 2E_1} - x$, we have

$$X_t = x + t \quad \text{and} \quad Y_t = y + t.$$

At time V_1 , the process X jumps to $Q_1 X_{V_1^-}$ while Y jumps to $Q_1 Y_{V_1^-}$ with probability $p(X_{V_1^-}, Y_{V_1^-})$ and stays at $Y_{V_1^-}$ with probability $q(X_{V_1^-}, Y_{V_1^-})$. Next we do the same by replacing x by $\max(X_{V_1}, Y_{V_1})$ and y by $\min(X_{V_1}, Y_{V_1})$ and E_1 by E_2 and Q_1 by Q_2 , etc. This recursive scheme produces two processes X and Y starting from x and y , and following the dynamics (1), and last but not least with maximal joint jump probability.

In other words, we have constructed a Markov process on $[0, \infty)^2$ with infinitesimal generator given by \bar{L} (see (8)). We are now ready for the proof of Theorem 1.5.

Proof of Theorem 1.5. Let $(X_t, Y_t)_{t \geq 0}$ be the couple of processes starting from (x, y) with $x > y$ and constructed as explained before, by maximizing the joint jump probability. We

assume for the moment that H is the Dirac mass at point δ . Let $\varphi : [0, \infty) \rightarrow [0, \infty)$ be a smooth nondecreasing and bounded function such that $\varphi(0) = 1$. The need of φ will be explained in Remark 4.3 below. For every $t \geq 0$ we define

$$\gamma_{x,y}(t) = \mathbb{E} \left(\frac{|X_t - Y_t|}{\varphi(\max(X_t, Y_t))} \right).$$

The idea of the proof is to control the derivative of γ in terms of γ and then to use Gronwall's lemma. Let T be the time of the first jump. Recall that by the coupling construction, the law of T has density f_x defined by (17) and that, at time T , X jumps automatically and Y jumps with probability $p(X_T, Y_T)$. We start with the identity

$$\gamma_{x,y}(t) = \mathbb{E} \left(\frac{|X_t - Y_t|}{\varphi(\max(X_t, Y_t))} (\mathbb{1}_{\{T \leq t, Y \text{ jumps at } T\}} + \mathbb{1}_{\{T \leq t, Y \text{ does not jump at } T\}} + \mathbb{1}_{\{T > t\}}) \right),$$

which gives

$$\begin{aligned} \gamma_{x,y}(t) &= \int_0^t (\gamma_{\delta x, \delta y}(t-s)p(x+s, y+s) + \gamma_{\delta x, y}(t-s)q(x+s, y+s))f_x(s) ds \\ &\quad + \frac{|x-y|}{\varphi(x+t)}F_x(t) \\ &= \frac{\delta(x-y)}{\varphi(\delta x)}p(x, y)xt + \frac{|\delta x - y|}{\varphi(\delta x)}q(x, y)xt + \frac{|x-y|}{\varphi(x)}(1-xt) - \frac{(x-y)\varphi'(x)}{\varphi(x)^2}t + o(t). \end{aligned}$$

Consequently, we obtain

$$\gamma'_{x,y}(0) = \frac{x}{\varphi(\delta x)}(\delta|x-y|p(x, y) + |\delta x - y|q(x, y)) - \frac{x|x-y|}{\varphi(x)} - \frac{|x-y|\varphi'(x)}{\varphi^2(x)}. \quad (22)$$

Of course, this expression can be deduced from the expression (8) of the infinitesimal generator \bar{L} of (X, Y) . At this point, in order to use Gronwall's Lemma, we need to bound $\gamma'_{x,y}$ in terms of $\gamma_{x,y}$. The following result makes this possible. Its proof can be skipped at first sight.

Lemma 4.2 (Technical lemma). *For any $x, y \geq 0$ and $\delta \in [0, 1)$, we have*

$$\beta(x, y) = -(1-\delta)p(x, y) + \left(\frac{|\delta x - y|}{|x-y|} - 1 \right) q(x, y) \leq -a(1-\delta)$$

where

$$a = 1 - e^{-1} \approx 0.63212.$$

Equivalently, for any $x, y \geq 0$, we have

$$\delta|x-y|p(x, y) + |\delta x - y|q(x, y) \leq b|x-y|$$

where

$$b = 1 - (1-\delta)a < 1.$$

Proof of Lemma 4.2. If $0 \leq y \leq \delta x$ then

$$\beta(x, y) = -(1 - \delta) \left(1 - \frac{y}{x - y} q(x, y) \right) \leq -(1 - \delta).$$

Let us assume now that $\delta x \leq y \leq x$. We have then

$$\begin{aligned} \beta(x, y) &= -(1 - \delta)p(x, y) + \frac{2y - (\delta + 1)x}{x - y} q(x, y) \\ &= -(1 - \delta) + \frac{-2\delta x + (1 + \delta)y}{x - y} q(x, y). \end{aligned}$$

Let us study the function $y \mapsto \beta(x, y)$ on $[\delta x, x]$.

$$\partial_y \beta(x, y) = (1 - \delta) \frac{x}{(x - y)^2} q(x, y) + \frac{2\delta x - (1 + \delta)y}{x - y} \partial_y p(x, y)$$

By using the identities (18) and (19) we get

$$\partial_y \beta(x, y) = \frac{\tau e^{-\tau^2/2 - y\tau}}{(x - y)} \left((1 - \delta)x \frac{1 - e^{-(x-y)\tau}}{(x - y)\tau} + 2\delta x - (1 + \delta)y \right).$$

Since $(1 - e^{-u})/u \geq 1 - u/2$ we obtain

$$\partial_y \beta(x, y) \geq \frac{\tau e^{-\tau^2/2 - y\tau}}{(x - y)} \left((1 + \delta) - (1 - \delta) \frac{x\tau}{2} \right).$$

Since $\tau \leq (\sqrt{y^2 + 4} - y)/2 = 2/(\sqrt{y^2 + 4} + y)$ and $\delta x \leq y$, we have

$$\begin{aligned} (1 + \delta) - (1 - \delta) \frac{x\tau}{2} &\geq \frac{(1 + \delta)\sqrt{y^2 + 4} + (1 + \delta)y - (1 - \delta)x}{\sqrt{y^2 + 4} + y} \\ &\geq \frac{(1 + \delta)\delta x - (1 - \delta)x}{\sqrt{y^2 + 4} + y} \geq 0. \end{aligned}$$

As a conclusion, β is increasing on the interval $[\delta x, x]$. Let us now compute the limit of β as $y \rightarrow x$. Since $0 \leq \tau \leq 1$ and $\tau \sim_{y \sim x} \tilde{\tau} = (\sqrt{x^2 + 4} - x)/2$, we have, by using (19),

$$\lim_{y \rightarrow x} \beta(x, y) = -(1 - \delta) + (1 - \delta)x\tilde{\tau}e^{-\tilde{\tau}^2/2 - x\tilde{\tau}}.$$

Now, the identity $\tilde{\tau}^2 + x\tilde{\tau} - 1 = 0$ yields

$$\lim_{y \rightarrow x} \beta(x, y) = -(1 - \delta) + (1 - \delta)x\tilde{\tau}e^{-x\tilde{\tau}/2}e^{-1/2}.$$

Next, by using

$$\frac{x\tilde{\tau}}{2} = \frac{1}{1 + \sqrt{1 + 4/x^2}} \leq \frac{1}{2}$$

and the elementary inequality $ue^{-u} \leq (1/2)e^{-1/2}$ valid for $u \in [0, 1/2]$, we finally obtain

$$\lim_{y \rightarrow x} \beta(x, y) \leq -(1 - \delta)(1 - e^{-1}).$$

□

Thus from (22) and Lemma 4.2 we get

$$\gamma'_{x,y}(0) \leq - \left[\left(1 - b \frac{\varphi(x)}{\varphi(\delta x)} \right) x + \frac{\varphi'(x)}{\varphi(x)} \right] \gamma_{x,y}(0). \quad (23)$$

Remark 4.3 (The need of φ). *At this point, it is clear why we need φ . Indeed, if φ is constant and equal to 1, the previous computations lead to*

$$\frac{d}{dt} \mathbb{E}(|X_t - Y_t|) \leq -(1-b) \mathbb{E}(\max(X_t, Y_t) |X_t - Y_t|) \leq 0$$

which does not provide exponential convergence due to the presence of $\max(X_t, Y_t)$ which can be arbitrarily small. However, this shows that $t \mapsto \mathbb{E}(|X_t - Y_t|)$ is non-increasing. Note that the introduction of the φ function appears as a modification of the norm, a successful idea which is helpful in various situations, see for instance in [17] and [8].

It remains to find a good function φ . The ideal one should be the arg-maximum φ_0 of the functional

$$\varphi \mapsto J(\varphi) = \inf_{x>0} \left(\left(1 - b \frac{\varphi(x)}{\varphi(\delta x)} \right) x + \frac{\varphi'(x)}{\varphi(x)} \right)$$

on the set of smooth positive bounded functions. . . We ignore how to determine such an optimal φ_0 . However, one can seek for some φ which will give a relatively good bound. If for some function φ , $J(\varphi)$ is positive, equal to some $\rho(\delta)$ (since b depends on δ), then, plugging this bound in Equation (23) leads to

$$\gamma'_{x,y}(0) \leq -\rho(\delta) \gamma_{x,y}(0).$$

Then, the Markov property ensures that, for any $t \geq 0$,

$$\gamma'_{x,y}(t) \leq -\rho(\delta) \gamma_{x,y}(t),$$

which leads to

$$\mathbb{E} \left(\frac{|X_t - Y_t|}{\varphi(\max(X_t, Y_t))} \right) \leq \frac{|x - y|}{\varphi(\max(x, y))} e^{-\rho(\delta)t}.$$

Since $1 \leq \varphi \leq \sup \varphi$, we get

$$\mathbb{E}(|X_t - Y_t|) \leq \sup_{u>0} \varphi(u) |x - y| e^{-\rho(\delta)t}.$$

Let us now examine the problem of finding a *good* function φ . Let us define, for any $u > 0$ and $c > 1$, the functions $\varphi_{u,c}$ and $\psi_{u,c}$ as

$$\varphi_{u,c}(x) = \begin{cases} \frac{c-1}{u}x + 1 & \text{for } 0 < x < u, \\ c & \text{for } x \geq u, \end{cases}$$

and

$$\psi_{u,c}(x) = \left(1 - b \frac{\varphi_{u,c}(x)}{\varphi_{u,c}(\delta x)} \right) x - \frac{\varphi'_{u,c}(x)}{\varphi_{u,c}(x)}.$$

Then

$$\psi_{u,c}(x) = \begin{cases} \left(1 - b \frac{(c-1)x + u}{(c-1)\delta x + u}\right)x + \frac{c-1}{(c-1)x + u} & \text{if } 0 < x \leq u, \\ \left(1 - \frac{ubc}{(c-1)\delta x + u}\right)x & \text{if } u < x < u/\delta, \\ (1-b)x & \text{if } x \geq u/\delta. \end{cases}$$

One can check that $\psi_{u,c}$ is increasing on $(u, +\infty)$. Moreover

$$\psi_{u,c}(u^-) = \psi_{u,c}(u^+) + \frac{c-1}{cu}.$$

If one chooses

$$u = \sqrt{(c-1) \frac{(c-1)\delta + 1}{(c-1)\delta + 1 - bc}},$$

which requires that $(c-1)\delta + 1 - bc > 0$ i.e. $c \leq e$, then

$$\psi_{u(c),c}(0) = \psi_{u(c),c}(u^+) = \frac{c-1}{u}.$$

Let us now explain why $\psi_{u(c),c}$ is concave on $[0, u)$ (which implies that the minimum of $\psi_{u(c),c}$ is achieved at 0). For any $x \in [0, u)$, one can write $\psi_{u(c),c}(x)$ as follows:

$$\psi_{u(c),c}(x) = Ax + B - \frac{b(1-\delta)v^2}{\delta^2} \frac{1}{\delta x + v} + \frac{1}{x + v},$$

where A and B are constants and $v = u/(c-1)$. As a consequence, $\psi_{u(c),c}''(x)$ has the sign of

$$-\frac{b(1-\delta)v^2}{(\delta x + v)^3} + \frac{1}{(x + v)^3} = \frac{(\delta x + v)^3 - b(1-\delta)v^2(x + v)^3}{(\delta x + v)^3(x + v)^3}.$$

Then it has the sign of $(\delta - w)x + v(1 - w)$ where $w = (b(1-\delta)v^2)^{1/3}$ since x and w are nonnegative. This sign is negative since $w > 1$. Indeed, using the expressions of u and b , we get

$$w^3 = be \frac{\delta(c-1) + c}{-(c-1)^2 + (e-1)(c-1)}.$$

Then, $w > 1$ if and only if $a \mapsto x^2 + (be\delta - e + 1)x + be$ is positive, which is equivalent to

$$g(\delta) = ((e-1)\delta^2 + \delta + 1 - e)^2 - 4(e - (e-1)(1-\delta)) > 0.$$

This function appears to be decreasing on $[0, \delta_0]$ for some $\delta_0 \in (0, 1)$ and increasing on $(\delta_0, 1]$. Since $g(1) < g(0) < 0$, the function g is negative on $[0, 1]$.

Then, one has to find c that maximizes $\psi_{u(c),c}(0)$. In other words, one has to maximize the function

$$c \in [1, e] \mapsto \sqrt{\frac{-(b-\delta)(c-1)^2 + (1-b)(c-1)}{\delta(c-1) + 1}}.$$

Since $b-\delta = e^{-1}(1-\delta)$ and $1-b = (1-e^{-1})(1-\delta)$, this is equivalent to find the maximum of the function

$$h(x) = \frac{-x^2 + (e-1)x}{\delta x + 1}$$

on $[0, e-1]$. The derivative of h has the sign of $-\delta x^2 - 2x + e - 1$ then

$$\max_{x \in [0, e-1]} h(x) = h\left(\frac{\sqrt{1 + (e-1)\delta} - 1}{\delta}\right) = \frac{1}{\delta^2} \left(2 + (e-1)\delta - 2\sqrt{1 + (e-1)\delta}\right).$$

As a conclusion, one gets, for $\delta \in (0, 1)$,

$$c(\delta) = \frac{\sqrt{1 + (e-1)\delta} - 1}{\delta} + 1 = \frac{e-1}{\sqrt{1 + (e-1)\delta} + 1} + 1$$

and

$$\begin{aligned} \psi_{u(c(\delta)), c(\delta)}(0) &= \sqrt{\frac{1-\delta}{e\delta^2} \left(2 + (e-1)\delta - 2\sqrt{1 + (e-1)\delta}\right)} \\ &= \frac{2 \sinh(1/2) \sqrt{1-\delta}}{\sqrt{(e-1)\delta + 1} + 1}. \end{aligned}$$

Notice that $\delta \mapsto c(\delta)$ is decreasing. In the particular case of $\delta = 1/2$, one can choose

$$c = \sqrt{2e+2} - 1 \sim 1.73 \quad \text{and} \quad u \sim 2.34$$

to get

$$J(\varphi_{u(c), c}) = 3e^{-1} + 1 - 2\sqrt{2e+2} \sim 0.311798.$$

Some elementary but tedious computations (or a numerical graphical analysis) reveal that $J(\varphi_{u(c(\delta)), c(\delta)}) \geq (1-\delta)/2$ for any $\delta \in (0, 1)$. Let us now turn to the case when the probability measure H is not the Dirac mass at δ . In that general model, we have to choose a function φ which does not depend on δ . For sake of simplicity, let us choose $u = c = 2$. As above one can show that $\varphi_{2,2}$ is concave on $(0, 2)$ and increasing on $(2, +\infty)$. Then,

$$J(\varphi_{2,2}) = \min(\psi_{2,2}(0), \psi_{2,2}(2)) = \min\left(\frac{1}{2}, (2 - 4e^{-1})\frac{1-\delta}{1+\delta}\right) \geq \frac{1}{2} \frac{1-\delta}{1+\delta}.$$

Then the same proof (with an additional conditioning with respect to H) provides

$$\gamma'_{x,y}(0) \leq -\rho_G \gamma_{x,y}(0) \quad \text{where} \quad \rho_G = \frac{1}{2} \mathbb{E}\left(\frac{1-Q}{1+Q}\right).$$

As in the deterministic case, we obtain, for every $t \geq 0$,

$$\mathbb{E}(|X_t - Y_t|) \leq 2|x - y|e^{-\rho_G t}.$$

On the other hand, by Remark 4.3, we have

$$\mathbb{E}(|X_t - Y_t|) \leq |x - y|$$

for every $t \geq 0$. The desired result follows then immediately. \square

Remark 4.4. In the special but useful case where H is the Dirac mass at point $\delta = 1/2$, we have $J(\varphi_{0.2,3.5}) \geq 0.46$ where $\varphi_{u,c}(x) = c - (c-1)e^{-ux}$ with $c = 3.5$ and $u = 0.2$. This gives a better exponential rate $\rho = 0.46$ but a greater constant in front of the exponential. Moreover, explicit controls for others values of δ (and uniform ones) are not tractable.

In the proof above, one has in any cases that $\gamma'_{x,y}(t) \leq 0$. This ensures that

$$W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leq \mathbb{E}(|X_t - Y_t| | X_0 = x, Y_0 = y) \leq |x - y|.$$

By the coupling argument, it is still the case for any initial laws: for any $t \geq 0$,

$$W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leq W_1(\mathcal{L}(X_0), \mathcal{L}(Y_0)).$$

We end up with the proof of Theorem 1.3.

Proof of theorem 1.3. The function f defined by $f(x) = x$ for every $x \geq 0$ satisfies to

$$Lf(x) = 1 - \kappa_1 x^2$$

where $\kappa_1 = 1 - \int_0^1 h H(dh) \in (0, 1]$. Now, for every $x \geq 0$ and $t \geq 0$,

$$\begin{aligned} \alpha_x(t) &:= \mathbb{E}(X_t | X_0 = x) \\ &= \alpha_x(0) + \int_0^t \alpha'_x(s) ds \\ &= x + \int_0^t \mathbb{E}((Lf)(X_s) | X_0 = x) ds \\ &= x + \int_0^t (1 - \kappa_1 \mathbb{E}(X_s^2 | X_0 = x)) ds. \end{aligned}$$

Also, since $-\kappa_1$ is negative, we obtain, by using Jensen's inequality,

$$\alpha'_x(t) = 1 - \kappa_1 \mathbb{E}(X_t^2 | X_0 = x) \leq 1 - \kappa_1 \alpha_x(t)^2.$$

As a consequence, $\alpha_x \leq \beta_x$ where β_x is the solution of the Riccati differential equation

$$\begin{cases} \beta_x(0) = x, \\ \beta'_x(t) = 1 - \kappa_1 \beta_x(t)^2 \text{ for } t > 0 \end{cases}$$

Denoting $d = \sqrt{\kappa_1}$, one gets, for $x \geq 1/d$,

$$\beta_x(t) = \frac{1}{d} + \frac{2(x - 1/d)e^{-2dt}}{(dx + 1) - (dx - 1)e^{-2dt}} = \frac{1}{d} \frac{dx \cosh(dt) + \sinh(dt)}{dx \sinh(dt) + \cosh(dt)} \leq \frac{1}{d \tanh(dt)},$$

and therefore

$$\sup_{x \geq 1/d} \alpha_x(t) \leq \frac{1}{d \tanh(dt)}.$$

On the other hand, we have also $\sup_{x \leq 1/d} \alpha_x(t) \leq 1/d$, and thus for every $t > 0$,

$$\sup_{x \geq 0} \alpha_x(t) \leq \frac{1}{d \tanh(dt)}.$$

Consider now two processes $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ generated by (1) with arbitrary initial laws. For any $s > 0$, $\mathbb{E}(|X_s - Y_s|) \leq 2 \sup_x \alpha_x(s)$ and therefore the upper bound above gives

$$W_1(\mathcal{L}(X_s), \mathcal{L}(Y_s)) \leq \frac{2}{d \tanh(ds)}.$$

Together with Theorem 1.5, this gives the following uniform estimate, for every $t \geq s > 0$,

$$\begin{aligned} W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) &\leq cW_1(\mathcal{L}(X_s), \mathcal{L}(Y_s))e^{-\rho(t-s)} \\ &\leq \frac{2ce^{\rho s}}{d \tanh(ds)}e^{-\rho t}. \end{aligned}$$

As an example, when H is the Dirac measure at δ , recall that one can choose $\rho = (1 - \delta)/2$ and $d = \sqrt{1 - \delta}$. The function $s \mapsto e^{\rho s} / \tanh(ds)$ is convex with a minimum at point

$$s_0 = \frac{1}{d} \operatorname{argth} \left(\frac{4}{\sqrt{1 - \delta} + \sqrt{17 - \delta}} \right).$$

In the case when $\delta = 1/2$, one gets, for any $t > 1.72$

$$W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leq 9e^{-.25t}.$$

□

References

- [1] D. Bakry, P. Cattiaux, and A. Guillin, *Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré*, J. Funct. Anal. **254** (2008), no. 3, 727–759. MR MR2381160 1
- [2] Ph. Carmona, F. Petit, and M. Yor, *On the distribution and asymptotic results for exponential functionals of Lévy processes*, Exponential functionals and principal values related to Brownian motion, Bibl. Rev. Mat. Iberoamericana, Rev. Mat. Iberoamericana, Madrid, 1997, pp. 73–130. MR MR1648657 (99h:60144) 2
- [3] 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. MR MR1895332 (2003f:60168) 1, 1, 3, 3, 3
- [4] C. Graham and Ph. Robert, *Interacting multi-class transmissions in large stochastic networks*, preprint [arXiv:0810.0347](https://arxiv.org/abs/0810.0347) [math.PR], 2008. 1
- [5] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083. MR MR0420249 (54 #8263) 3.4
- [6] ———, *Hypercontractivity, logarithmic Sobolev inequalities, and applications: a survey of surveys*, Diffusion, quantum theory, and radically elementary mathematics, Math. Notes, vol. 47, Princeton Univ. Press, Princeton, NJ, 2006, pp. 45–73. MR MR2325763 3.4
- [7] F. Guillemin, Ph. Robert, and B. Zwart, *AIMD algorithms and exponential functionals*, Ann. Appl. Probab. **14** (2004), no. 1, 90–117. MR MR2023017 (2004m:60216) 1
- [8] M. Hairer and J.C. Mattingly, *Yet another look at Harris’ ergodic theorem for markov chains*, preprint [arXiv:0810.2777v1](https://arxiv.org/abs/0810.2777v1) [math.PR], 2008. 4.3
- [9] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR MR1849347 (2003k:28019) 3
- [10] T. Lindvall, *Lectures on the coupling method*, Dover Publications Inc., Mineola, NY, 2002, Corrected reprint of the 1992 original. MR MR1924231 4
- [11] A. H. Löpker and J. S. H. van Leeuwen, *Transient moments of the window size in TCPA*, J. Appl. Probab. **45** (2008), no. 1, 163–175. 2
- [12] F. Malrieu and D. Talay, *Concentration inequalities for Euler schemes*, Monte Carlo and quasi-Monte Carlo methods 2004, Springer, Berlin, 2006, pp. 355–371. MR MR2208718 (2006m:65022) 3
- [13] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Communications and Control Engineering Series, Springer-Verlag London Ltd., London, 1993. MR MR1287609 (95j:60103) 1

- [14] T. J. Ott, J. H. B. Kemperman, and M. Mathis, *The stationary behavior of ideal TCP congestion avoidance*, unpublished manuscript available at www.teunisott.com, 1996. 1
- [15] 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
- [16] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR MR1964483 (2004e:90003) 1, 3, 3
- [17] ———, *Hypocoercivity*, to appear in the series Memoirs of the AMS, available electronically on the author's personal web page, 2008. 4.3

Acknowledgments. DC would like to thank Philippe Robert and his team for kind hospitality during a visit to INRIA Rocquencourt, near Paris, France (December 2007).

Djalil CHAFAÏ, [mailto:chafai\(AT\)math.univ-toulouse.fr](mailto:chafai(AT)math.univ-toulouse.fr)

UMR 181 INRA ENVT, 23 CHEMIN DES CAPELLES F-31076 TOULOUSE CEDEX 3, FRANCE.

UMR 5583 CNRS INSTITUT DE MATHÉMATIQUES DE TOULOUSE (IMT)

UNIVERSITÉ DE TOULOUSE III, 118 ROUTE DE NARBONNE, F-31062, TOULOUSE CEDEX, FRANCE.

Florent MALRIEU, corresponding author, [mailto:florent.malrieu\(AT\)univ-rennes1.fr](mailto:florent.malrieu(AT)univ-rennes1.fr)

UMR 6625 CNRS INSTITUT DE RECHERCHE MATHÉMATIQUE DE RENNES (IRMAR)

UNIVERSITÉ DE RENNES I, CAMPUS DE BEAULIEU, F-35042 RENNES CEDEX, FRANCE.

Katy PAROUX, [mailto:katy.paroux\(AT\)univ-fcomte.fr](mailto:katy.paroux(AT)univ-fcomte.fr)

UMR 6623 CNRS LABORATOIRE DE MATHÉMATIQUES

UNIVERSITÉ DE FRANCHE-COMTÉ, F-25030 BESANÇON CEDEX, FRANCE

Present adress INRIA RENNES-BRETAGNE ATLANTIQUE,

F-35042 CAMPUS DE BEAULIEU, RENNES CEDEX, FRANCE.

Compiled November 13, 2008.