On the long time behavior of the TCP window size process

Djalil Chafai, Florent Malrieu, Katy Paroux

On the long time behavior of the TCP window size process

Djalil Chafaï, Florent Malrieu, Katy Paroux

Preprint 2009

Abstract

The TCP window size process appears in the modeling of the famous Transmission Control Protocol used for data transmission 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 already been investigated in the literature. In the present paper, we mainly get quantitative estimates for the convergence to equilibrium, in terms of the W_1 Wasserstein coupling distance, for the process and also for its embedded chain.

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

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

1 Introduction

The TCP protocol is one of the main data transmission protocols of the Internet. It has been designed to adapt to the various traffic conditions of the actual network. For a connection, the maximum number of packets that can be sent at each round is given by a variable W, called the *congestion window size*. If all the W packets are successfully transmitted, then W is increased by 1, otherwise it is multiplied by $\delta \in [0, 1)$ (detection of a congestion). As shown in [5, 10, 18], a correct scaling of this process leads to a continuous time Markov process, called general TCP window size process. This process $X = (X_t)_{t\geqslant 0}$ has $[0, \infty)$ as state space and its infinitesimal generator is given, for any smooth function $f : [0, \infty) \to \mathbb{R}$, by

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

for some probability measure H supported in $[0, 1)$. This window size $(X_t)_t$ increases linearly (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. The process $(X_t)_{t\geqslant 0}$ appears as an Additive Increase Multiplicative Decrease process (AIMD), but also as a very special Piecewise Deterministic Markov

Process (PDMP) initially introduced in [4]. In this direction, [14] gives a generalization of the scaling procedure to interpret various PDMPs as the limit of discrete time Markov chains. In the real world (Internet), the AIMD mechanism allows a good compromise between the minimization of network congestion time and the maximization of mean throughput.

Our aim in this paper is to get quantitative estimates for the convergence to equilibrium of this general TCP window size process.This process X 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. [2, 3, 16] for the Meyn-Tweedie-Foster-Lyapunov technique). Nevertheless, this process is irreversible since time reversed sample paths are not sample paths and it has infinite support. This makes Meyn-Tweedie-Foster-Lyapunov techniques inefficient for the derivation of quantitative exponential ergodicity.

The *embedded chain* $\hat{X}$ of the process X is the sequence of its positions just after a jump. It is an homogeneous discrete time Markov chain with state space $[0, \infty)$. As already observed in [5], it is also the square root of a first order auto-regressive process with non-Gaussian innovations and random coefficients. We obtain the following results concerning $\hat{X}$. We show first that it admits a unique invariant probability measure ν, and that it converges in law to ν given any (random) initial value $\hat{X}_0$. More precisely, using a coupling technique on trajectories, we prove an ergodic theorem of geometric convergence to equilibrium with respect to any Wasserstein distance. Then we provide non asymptotic concentration bounds, thanks to Gross's logarithmic Sobolev inequalities.

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. The reader may find explicit series for the moments of μ and ν in [10, 14, 15]. Nevertheless, quantitative convergence to equlibium have not yet been obtained. We will adress this question for a slight generalization of the TCP process given by its infinitesimal genrerator:

$$L_a(f)(x) = f'(x) + (x + a) \int_0^1 (f(hx) - f(x))H(dh) \tag{2}$$

where $a \geqslant 0$. We obtain a good answer if $a > 0$. In this case we first show the existence of a coupling with exponential decay. We use this result to prove an exponential ergodic theorem in term of Wasserstein distance. Eventually, we provide a uniform bound over the starting law that implies strong ergodicity. This kind of uniform estimates, though classical for processes on a compact set, is rather unusual for real valued processes. Nevertheless, if $a = 0$, we are not able to derive exponential bounds.

The remainder of the paper is organized as follows. In the next preliminary section, we introduce some notations and give the statements of the main results. In section 3, we focus on the embedded chain $\hat{X}$ and establish its convergence to equilibrium. The last section is devoted to the study of the continuous time process X and its generalization and contains the proof of the results announced in section 2.

2 Notations and main results

Let us first 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, $(E_n)_{n\geqslant 1}$ be a sequence of i.i.d. exponential random variables of unit mean, and $(Q_n)_{n\geqslant 1}$ be a sequence of i.i.d. random variables of law H. Assume that X_0, $(E_n)_{n\geqslant 1}$ and $(Q_n)_{n\geqslant 1}$ are independent. We define by induction the jump times $(T_n)_{n\geqslant 1}$ and the positions just after the jumps $(X_{T_n})_{n\geqslant 1}$ as

$$T_n = T_{n-1} + \sqrt{X_{T_{n-1}}^2 + 2E_n} - X_{T_{n-1}} \quad \text{and} \quad X_{T_n} = Q_n\sqrt{X_{T_{n-1}}^2 + 2E_n}. \tag{3}$$

If we set $T_0 = 0$, then for every $n \geqslant 0$ and $t \in [T_n, T_{n+1})$, we have $X_t = X_{T_n} + t - T_n$ and in particular, $X_{T_n} = Q_n X_{T_n^-}$. For every $t \geqslant 0$, 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 $\hat{X} = (X_{T_n})_{n\geqslant 0}$ is the *embedded chain* of X. According to (3), this discrete time Markov chain with state space $[0,\infty)$ satisfies the recursion

$$\hat{X}_{n+1}^2 = Q_{n+1}^2(\hat{X}_n^2 + 2E_{n+1}). \tag{4}$$

Thus, the embedded chain $\hat{X}$ is the square root of a first order auto-regressive process with non-Gaussian innovations $(2Q_n^2 E_n)_{n\geqslant 1}$ and random coefficients $(Q_n^2)_{n\geqslant 1}$, as already observed in [5]. The embedded chain $\hat{X}$ is homogeneous, and its transition kernel K is given, for any $x \geqslant 0$ and every bounded measurable $f : [0,\infty) \to \mathbb{R}$, by the formula

$$K(f)(x) = \int_0^\infty f(y)\,K(x,dy) = \mathbb{E}\Big[f\Big(Q\sqrt{x^2+2E}\Big)\Big] \tag{5}$$

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. We recall that explicit series for the moments of μ and ν can be found in [10, 14, 15].

Despite the apparent simplicity of the dynamics (1), the quantitative study of the long time behavior of X is not easy, mainly because the jump rate depends on the position x of the process. Our strategy is to couple two trajectories starting at two different points in such a way that they get closer and closer. It seems difficult to stick the two trajectories in order to get total variation estimates since the sample paths are parallel between jump times. Thus, we provide quantitative bounds in terms of the Wasserstein coupling distance. Recall that for every $p \geqslant 1$, the W_p Wasserstein distance between two laws μ and ν on $\mathbb{R}$ with finite p^{th} moment is defined by

$$W_p(\mu,\nu) = \left(\inf_{\Pi} \int_{\mathbb{R}^2} |x-y|^p \,\Pi(dx,dy)\right)^{p^{-1}} \tag{6}$$

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. [19, 20]. 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.

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.

In section 3 we first establish the following geometric convergence to equilibrium of the embedded Markov chain $\hat{X}$ for any Wasserstein distance.

Theorem 2.1 (Wasserstein exponential ergodicity for the generic embedded chain)**.** *Let $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ 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 $p \geqslant 1$. Let $\hat{X}$ and $\hat{Y}$ be the embedded chains of X and Y. Then, for any $n \geqslant 0$, with a random variable $Q \sim H$,*

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

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

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

We also establish in section 3 non asymptotic concentration bounds in the ergodic theorem by using Gross logarithmic Sobolev inequalities:

Theorem 2.2 (Gaussian deviations for the ergodic theorem for the embedded chain)**.** *Let $\hat{X}$ be the embedded chain associated to* (1) *and starting from $\hat{X}_0 = x \geqslant 0$. Assume that H is the Dirac mass at point $\delta \in (0,1)$. Then for any $u \geqslant 0$ and any 1-Lipschitz function $f : [0,\infty) \to \mathbb{R}$,*

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

The convergence to equilibrium of the continuous time process X with generator (2) is addressed in section 4. The idea is to exhibit a *coupling*, *i.e.* a Markov process on $[0,\infty)^2$ for which the marginal components are generated by (2), with prescribed initial laws. The infinitesimal generator $\overline{L}$ of this coupling is defined for every smooth $f : [0,\infty)^2 \to \mathbb{R}$ by

$$\overline{L}(f)(x,y) = \mathrm{div}(f)(x,y) + (x+a)\int_0^1 \left(f(hx,hy)\frac{y+a}{x+a} + f(hx,y)\frac{x-y}{x+a} - f(x,y)\right) H(dh) \tag{7}$$

if $x \geqslant y$ and

$$\overline{L}(f)(x,y) = \operatorname{div}(f)(x,y) + (y+a)\int_0^1 \left(f(hx,hy)\frac{x+a}{y+a} + f(x,hy)\frac{y-x}{y+a} - f(x,y) \right) H(dh)$$

if $x \leqslant y$, where $\operatorname{div}(f) = \partial_1 f + \partial_2 f$. 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 are made more precise hereafter. The positions of both "components" increase linearly with slope 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 as follows:

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 bounds.

Theorem 2.3 (Wasserstein exponential ergodicity)**.** *Assume that $a > 0$. For any processes $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$ generated by* (2) *with finite first moment at initial time, and for any $t > 0$, we have*

$$W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leqslant e^{-a\kappa_1 t} W_1(\mathcal{L}(X_0), \mathcal{L}(Y_0)),$$

where $\kappa_1 = 1 - \int_0^1 h\, H(dh)$. In particular, when Y_0 follows the invariant law μ of (2)*, we get for every $t \geqslant 0$*

$$W_1(\mathcal{L}(X_t), \mu) \leqslant e^{-a\kappa_1 t}\, W_1(\mathcal{L}(X_0), \mu).$$

The following theorem, proved in section 4, 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 2.4 (Strong ergodicity)**.** *Assume that $a > 0$. For two processes $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ generated by* (2) *with arbitrary initial laws $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$ and for every t and s such that $t > s > 0$, one has*

$$W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) \leqslant \frac{2e^{a\kappa_1 s}}{d\tanh(ds)} e^{-a\kappa_1 t}.$$

Theorem 2.4 provides in particular a uniform bound in $N \in (0,\infty)$ if $X_0 = 0$ and $Y_0 = N$. This kind of uniform estimates are classical for processes on a compact set but rather unusual for real valued ones.

Theorem 2.5. *Assume that $a = 0$ and that $H = \delta_h$ with $h \in (0,1)$. Then the process (X, Y) driven by the infinitesimal generator $\overline{L}$ defined in (7) satisfies*

$$\frac{d}{dt}\mathbb{E}_{(x,y)}(|X_t - Y_t|) \leqslant -(1+h)\mathbb{E}_{(x,y)}\Big(|X_t - Y_t|^2\Big) \tag{8}$$

for any $x, y \in \mathbb{R}$. In particular, for any $t \geqslant 0$ and $X_0, Y_0 \geqslant 0$, we have

$$\mathbb{E}(|X_t - Y_t|) \leqslant \frac{\mathbb{E}(|X_0 - Y_0|)}{1 + (1+h)\mathbb{E}(|X_0 - Y_0|)t}. \tag{9}$$

Open questions and further remarks

The inequality (8) should provide a better bound than (9). As pointed out in Lemma 4.2, one can actually expect an exponential rate, but this remains an open problem. One may also ask for a version involving W_p for any $p \geqslant 1$ or even the total variation distance.

Beyond the TCP window size dynamics, one may ask about the speed of convergence of ergodic PDMPs, for which necessary and sufficient ergodicity criteria are already known, see e.g. [3]. One may also study the long time behavior of interacting processes associated to (1) or (13), for instance Mac Kean-Vlasov mean field interactions as in [8].

3 Embedded chain

It is shown in [5, 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

$$x \geqslant 0 \mapsto \frac{1}{\prod_{n=1}^{\infty}(1-\delta^{2n})}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}\delta^{-2n}}{\prod_{k=1}^{n-1}|1-\delta^{-2k}|}xe^{-\delta^{-2n}x^2/2}. \tag{10}$$

It is unimodal, of order $O(x\exp(-\delta^2x^2/2))$ when $x \to \infty$, and all its derivatives vanish at $x = 0$.

If H is not a Dirac mass, the invariant measure ν of the embedded Markov chain is no longer explicit. Nevertheless, the recursion formula (4) provides the following result, see [6, 7], which establish the existence of an invariant measure with sub-Gaussian tails.

Theorem 3.1 (Convergence of the embedded chain, [6, 7])**.** *Given any $\hat{X}_0$, the embedded Markov chain $\hat{X} = (\hat{X}_n)_{n\geqslant 0}$ associated to the dynamics (1) converges in distribution to the law of the random variable*

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

which is a.s. finite, where $E_1, E_2, \ldots$ and $Q_1, Q_2, \ldots$ 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

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

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

Let us now turn to our quantitative estimate for the convergence to equilibrium for the embedded chain.

Proof of Theorem 2.1. It is sufficient to provide a good coupling. Let $x \geqslant 0$ and $y \geqslant 0$ be two non-negative real numbers, and let $(E_n)_{n\geqslant 1}$ and $(Q_n)_{n\geqslant 1}$ 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

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

By the analogue of (3) for (13), 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 $p \geqslant 1$, since $x \mapsto \sqrt{x^2+a}$ is a 1-Lipschitz function on $[0,\infty)$ for any $a \geqslant 0$, we get

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

A straightforward recurrence leads to

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

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. ☐

Let us now investigate some properties of the kernel K defined by (5) that will be used to provide concentration bounds for the ergodic theorem. The key point is that K^n and ν satisfy a Gross (or logarithmic Sobolev) inequality.

Definition 3.2 (Gross inequality). *A law η on $\mathbb{R}^d$ satisfies a Gross (or logarithmic Sobolev [1, 9]) inequality with constant $c > 0$ when for any smooth compactly supported $f : \mathbb{R}^d \to \mathbb{R}$,*

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

We denote by $\mathrm{G{\scriptsize ROSS}}(\eta) \in (0,\infty]$ the smallest constant for which this holds true.

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

$$\mathbb{E}_\eta\Big(e^{\lambda f}\Big) \leqslant e^{C\alpha^2\lambda^2/4} e^{\lambda \mathbb{E}_\eta f} \tag{11}$$

as soon as $C \geqslant \mathrm{G{\scriptsize ROSS}}(\eta)$. This means that η satisfies a *sub-Gaussian concentration of measure* for Lipschitz functions. For more details, see e.g. [11, 20] and references therein.

Theorem 3.3 (Properties of the kernel of the embedded chain)**.** *Let $\hat{X}$ be the embedded chain associated to* (1) *with transition kernel* (5)*. 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 $x \geqslant 0$, the law $K(\cdot)(x)$ satisfies a Gross inequality with constant $2\delta^2$.*

Proof. If $\delta = 0$, then K is the Dirac mass at 0 and the result is trivial. For any smooth function $f : [0,\infty) \to \mathbb{R}$, we have from (5) that

$$\left|(Kf)'\right| = \delta \left| K\left(\frac{x}{\sqrt{x^2+2E}} f'\left(\delta\sqrt{x^2+2E}\right)\right)\right| \leqslant \delta K(|f'|). \tag{12}$$

Let us show now that for every $x \geqslant 0$ 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

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

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

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

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 $x \geqslant 0$. 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 $(2\delta)^2/2 = 2\delta^2$. □

Remark 3.4. *When $\delta = 0$, the embedded chain is the constant Markov chain equal to 0. Moreover, the chain $(Z_n)_{n\geqslant 0}$ defined by $Z_n = X_{T_n^-}$ is also quite simple to study. Indeed, the random variables $(Z_n)_{n\geqslant 1}$ are i.i.d. and have the law ν of $\sqrt{2E}$. The previous proof ensures that ν satisfies a Gross inequality with constant 2. One of the most useful properties of Gross inequality is the tensorization property:* $\mathrm{Gross}(\eta^{\otimes n}) \leqslant \mathrm{Gross}(\eta)$ *for every $n \geqslant 1$, see e.g. [1, Chapter 1]. Using now the concentration property, one has, for any 1-Lipschitz function and any $u \geqslant 0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^n f(Z_k) - \int f\,d\nu\right| \geqslant u\right) \leqslant 2\exp\left(-\frac{Nu^2}{2}\right).$$

In the more general case where δ is positive, $(\hat{X}_n)_{n\geqslant 1}$ is no longer i.i.d. Nevertheless, the Gross inequality holds true for the iterated kernels and for the invariant law ν:

Corollary 3.5 (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 $n \geqslant 0$, let K^n be the iterated transition kernel of $\hat{X}$, defined recursively for every bounded measurable function $f : [0,\infty) \to \mathbb{R}$ by*

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

where K is the kernel of $\hat{X}$ as in (5)*. Then for every integer $n \geqslant 1$ and every real $x \geqslant 0$, the iterated kernel $K^n(x,\cdot)$ of $\hat{X}$ satisfies a Gross inequality and*

$$\textsc{Gross}(K^n(x,\cdot)) \leqslant 2\delta^2 \frac{1-\delta^{2n}}{1-\delta^2}.$$

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

$$\textsc{Gross}(\nu) \leqslant 2\delta^2(1-\delta^2)^{-1}.$$

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

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

To show that K^n satisfies a Gross inequality, we use a semi-group decomposition technique borrowed from [13]. For any $n \geqslant 1$ and any smooth function $f : [0,\infty) \to \mathbb{R}$, 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

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

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

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

Now, by using the commutation (12), we obtain, for all $1 \leqslant i \leqslant n$,

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

Next, the Cauchy–Schwarz inequality

$$\frac{(Kf)^2}{K(g)} \leqslant K\bigg(\frac{f^2}{g}\bigg)$$

gives

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

From these bounds, a straightforward induction gives

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

Consequently, by putting all together, we have

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

This gives

$$\textsc{Gross}(K^n) \leqslant 2\delta^2(1-\delta^{2n})(1-\delta^2)^{-1}.$$

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

$$\textsc{Gross}(\nu) \leqslant \limsup_{n\to\infty} \textsc{Gross}(K^n) \leqslant 2\delta^2(1-\delta^2)^{-1}.$$

□

The Gross inequality for K can also be used to derive Theorem 2.2.

Proof of Theorem 2.2. We shall establish that for any $u \geqslant 0$ and any 1-Lipschitz function $f : [0,\infty) \to \mathbb{R}$,

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

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,

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

Now the Markov property ensures that

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

From Theorem 3.3, the kernel $K(x,\cdot)$ of $\hat{X}$ satisfies a Gross inequality with constant $2\delta^2$ for every $x \geqslant 0$. This inequality implies by (11) that for any c-Lipschitz function g,

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

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

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

The commutation relation (12) ensures that $f+Kf$ is $(1+\delta)$-Lipschitz and then

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

A straightforward recurrence ensures that

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

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

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

The right hand side is minimum for $\lambda = u(\delta^{-2}-1)$. At this point, we recall the dual formulation of $W_1(\alpha,\beta)$ for every probability laws α and β:

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

Therefore, by using Theorem 2.1, one gets

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

□

Remark 3.6. *A careful reading of the proof of Theorem 2.2 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* (11).

4 Continuous time process

As an introduction of our coupling method to prove Theorem 2.3, let us consider the following simpler dynamics, studied recently in [12, 17]. The window size is modeled by a Markov process $X=(X_t)_{t\geqslant 0}$ 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 $(Q_k)_{k\geqslant 0}$ 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

$$L(f)(x) = f'(x) + \lambda\int_0^1 (f(hx)-f(x))\,H(dh) \tag{13}$$

where λ is this time a positive real number. In this framework, one can compute explicitly the transient moments of X_t (see [12, 17]): for every $n \geqslant 0$, every $x \geqslant 0$, and every $t \geqslant 0$,

$$\mathbb{E}((X_t)^n \mid X_0 = x) = \frac{n!}{\prod_{k=1}^n \theta_k} + n! \sum_{m=1}^n \left(\sum_{k=0}^m \frac{x^k}{k!} \prod_{\substack{j=k \\ j \neq m}}^n \frac{1}{\theta_j - \theta_m} \right) e^{-\theta_m t} \tag{14}$$

where for every real or integer $p \geqslant 1$ the quantity θ_p is as in our Theorem 4.1. In contrast with the original dynamics (1), the jump rate is constant and thus the jumps occur at Poissonian times. In this framework, we derive easily the following theorem, which states an exponential ergodicity in all Wasserstein distances.

Theorem 4.1 (Wasserstein Exponential Ergodicity for constant jump rate)**.** *Let $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ be two processes generated by* (13)*. Assume that $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$ have finite p^{th} moment for some real $p \geqslant 1$. If one defines $\theta_p = \lambda(1 - \mathbb{E}(Q^p))$ with $Q \sim H$ then for every $t \geqslant 0$,*

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

We ignore if the exponential rate of convergence in Theorem 4.1 is optimal. One may try to get an upper bound from the moments formula (14).

Proof of Theorem 4.1. Let $N = (N_t)_{t\geqslant 0}$ be a Poisson process with constant intensity λ and $Q = (Q_k)_{k\geqslant 1}$ be i.i.d. random variables with law H, independent of N. For any $x, y \geqslant 0$, let us consider the processes $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ 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 (13). 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 $t \geqslant 0$ and $p \geqslant 0$,

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

As a consequence, if $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ are now two processes generated by (13) 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 $t \geqslant 0$, and any $p \geqslant 1$,

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

Taking the infimum over Π concludes the proof. $\square$

Let us now turn to the *generalized* TCP window size process generated by the infinitesimal generator (2). Consider a two dimensional process where both components are generated by (2). 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 is to encourage simultaneous jumps: let us introduce the Markov process $((X_t, Y_t))_{t\geqslant 0}$ on $[0,\infty)^2$ defined by its infinitesimal generator

$$\begin{aligned}\overline{L}f(x,y) =&\partial_1 f(x,y) + \partial_2 f(x,y)\\ &+ (x-y)\int_0^1 (f(hx,y) - f(x,y))\, H(dh)\\ &+ (y+a)\int_0^1 (f(hx,hy) - f(x,y))\, H(dh)\end{aligned}$$

if $x \geqslant y$ (if $y < x$ one has to exchange the variables x and y).

Choosing a test function f of the form $f(x,y) = g(x)$ or $f(x,y) = g(y)$ shows that X and Y are both Markov processes with infinitesimal generator L.

The dynamics of (X,Y) is as follows: if $(X_0, Y_0) = (x,y)$ with for example $x \geqslant y$, then

- the first jump time T has density $t \mapsto (x+t)e^{-t^2/2-xt}\mathbb{1}_{(0,+\infty)}(t)$,
- on the event $\{T = t\}$ we have $(X_s, Y_s) = (x+s, y+s)$ for $s < t$ and

$$(X_t, Y_t) = \begin{cases}\left(\dfrac{x+t}{2}, \dfrac{y+t}{2}\right) & \text{with probability } \dfrac{y+t+a}{x+t+a},\\[2ex] \left(\dfrac{x+t}{2}, y+t\right) & \text{with probability } \dfrac{x-y}{x+t+a}.\end{cases}$$

4.1 The modified TCP process

The first part of this section is dedicated to the proof of Theorem 2.3.

Proof of Theorem 2.3. We have to study the function α: $t \mapsto \mathbb{E}_{(x,y)}(|X_t - Y_t|)$ where (X,Y) evolves according to the generator $\overline{L}$. Assume that $x > y$, then

$$\begin{aligned}\alpha'_{(x,y)}(0) =&(x-y)\int_0^1 (|hx-y| - |x-y|)\, H(dh) + (y+a)(x-y)\int_0^1 (h-1)\, H(dh)\\ =&-(x-y)\int_0^1 \mathbb{1}_{\{hx>y\}}(1-h)(x+y+a)\, H(dh)\\ &-(x-y)\int_0^1 \mathbb{1}_{\{hx\leqslant y\}}((1+h)(x-y) + (1-h)a)\, H(dh)\\ \leqslant&-a(x-y)\int_0^1 (1-h)\, H(dh).\end{aligned}$$

The Markov property ensures that

$$\alpha'_{(x,y)}(t) \leqslant a\kappa_1 \alpha_{(x,y)}(t),$$

where $\kappa_1 = 1 - \int_0^1 h\, H(dh)$. This obviously implies that

$$\mathbb{E}_{(x,y)}(|X_t - Y_t|) \leqslant |x-y|e^{-a\kappa_1 t}.$$

The end of the proof is straightforward. If $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ are two processes generated by (1) and if Π is a coupling of $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$, we hava,for every $t \geqslant 0$,

$$\begin{aligned}
W_1(\mathcal{L}(X_t), \mathcal{L}(Y_t)) &\leqslant \int_{[0,\infty)^2} \mathbb{E}(|X_t - Y_t| \,|\, X_0 = x, Y_0 = y)\, \Pi(dx, dy) \\
&\leqslant e^{-a\kappa_1 t} \int_{[0,\infty)^2} |x - y|\, \Pi(dx, dy).
\end{aligned}$$

Taking the infimum over Π provides the result. $\square$

Let us now turn to the proof of Theorem 2.4.

Proof of theorem 2.4. The function f defined by $f(x) = x$ for every $x \geqslant 0$ satisfies to

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

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

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

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

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

As a consequence, $\alpha_x \leqslant \beta_x$ 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 $x \geqslant 1/d$,

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

and therefore

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

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

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

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

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

Together with Theorem 2.3, this gives the following uniform estimate, for every $t\geqslant s>0$,

$$\begin{aligned} W_1(\mathcal{L}(X_t),\mathcal{L}(Y_t)) &\leqslant W_1(\mathcal{L}(X_s),\mathcal{L}(Y_s))e^{-a\kappa_1(t-s)} \\ &\leqslant \frac{2e^{a\kappa_1 s}}{d\tanh(ds)}e^{-a\kappa_1 t}. \end{aligned}$$

□

4.2 The real TCP process

We end by giving the proof of Theorem 2.5 and making some comments on it.

Proof of Theorem 2.5. We start the proof as in Theorem 2.3:

$$\alpha'_{(x,y)}(0)=\begin{cases} -(1-h)(x+y)(x-y) & \text{if } hx>y, \\ -(1+h)(x-y)^2 & \text{if } hx\leqslant y. \end{cases}$$

The first bound is better. Nevertheless, if D_h is the set $\big\{(x,y),\ hy\leqslant x\leqslant h^{-1}y\big\}$, one has to notice that the process (X,Y) cannot exit D_h. Then, thanks to Markov property, we get the following bound:

$$\frac{d}{dt}\mathbb{E}(|X_t-Y_t|)\leqslant -(1+h)\mathbb{E}\Big(|X_t-Y_t|^2\Big).$$

Jensen's inequality ensures that

$$\frac{d}{dt}\mathbb{E}(|X_t-Y_t|)\leqslant -(1+h)\{\mathbb{E}(|X_t-Y_t|)\}^2,$$

and thus, for any $t\geqslant 0$,

$$\mathbb{E}(|X_t-Y_t|)\leqslant \frac{\mathbb{E}(|X_0-Y_0|)}{1+(1+h)\mathbb{E}(|X_0-Y_0|)t}.$$

□

Figure 1 suggests that the convergence rate given by Theorem 2.5 is far from being satisfactory. The non-optimality of the coupling is clear. However, even with such a coupling, one could expect an explicit exponential upper bound. Let us denote $D_t=|X_t-Y_t|$ where (X_t,Y_t) is defined in Theorem 2.5 . We think that $\mathbb{E}(D_t^2)$ is in fact of the order of $\mathbb{E}(D_t)$ (instead of $\mathbb{E}(D_t)^2$). Indeed, with a rate of order D_t a nonsimultaneous jump occurs at time t and then D_t is again of order one. In the following lemma, we introduce a simple Markov chain which captures the essential feature the dynamics of D_t (division by 2 with probability $1-O(D_t)$) and we show that the expected position at time n goes to zeros exponentially fast as n goes to infinity. Additionally the recursive equation (15) plays the role of (8).

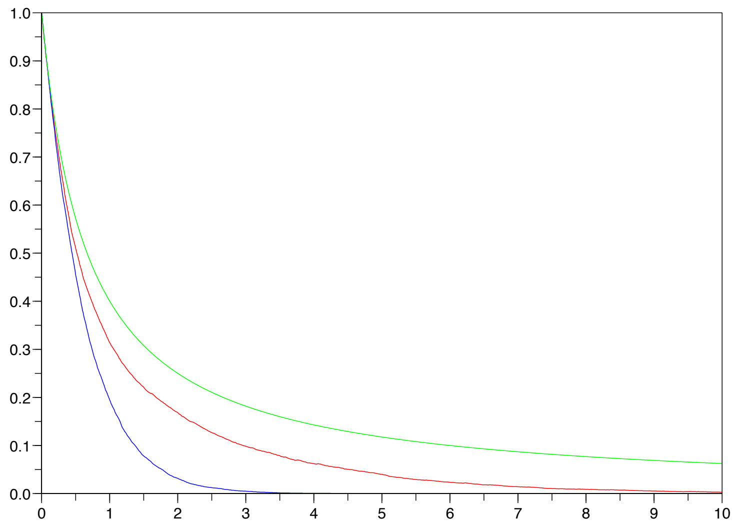

Figure 1: Here $(x, y) = (2, 1)$ and $H = \delta_{1/2}$. This picture presents the three following functions of time: $t \mapsto W_1(\mathcal{L}(X_t^x), \mathcal{L}(X_t^y))$ where X^u is driven by (1) with $X_0^u = u$ (blue curve), $t \mapsto \mathbb{E}(|X_t^{x,y} - Y_t^{x,y}|)$ where $(X^{x,y}, Y^{x,y})$ is driven by (7) with $(X_0^{x,y}, Y_0^{x,y}) = (x, y)$ (red curve), $t \mapsto (x - y)/(1 + 1.5(x - y)t)$ the upper bound (9) of Theorem 2.5 (green curve). The first and second curves have been obtained by Monte-Carlo simulations.

Lemma 4.2. *Consider the homogeneous irreducible Markov chain $X = (X_n)_{n\geqslant 0}$ with state space $E = \{2^{-i},\ i \in \mathbb{N}\}$ such that, for any $n \geqslant 0$ and $x \in E$, on the event $\{X_n = x\}$*

$$X_{n+1} = \begin{cases} 1 & \text{with probability } x/2, \\ x/2 & \text{with probability } 1 - x/2. \end{cases}$$

Denote by $\mathbb{E}^1(X_n)$ the quantity $\mathbb{E}(X_n | X_0 = 1)$. Then, for any $n \geqslant 1$,

$$\mathbb{E}^1(X_{n+1}) = \mathbb{E}^1(X_n) - \frac{1}{4}\mathbb{E}^1\left(X_n^2\right) \tag{15}$$

and there exists a constant $c > 0$ such that for any $n \geqslant 1$,

$$\mathbb{E}^1(X_n) \leqslant \exp(-cn). \tag{16}$$

Proof. The Markov chain X is transient (and converges to 0) since

$$p := \mathbb{P}(\forall n \geqslant 0,\ X_{n+1} = X_n/2 | X_0 = 1) = \prod_{i=1}^{\infty}(1 - 2^{-i}) > 0.$$

Since for any $n \geqslant 0$,

$$\mathbb{E}(X_{n+1}|X_n) = \frac{1}{2}X_n + \frac{1}{2}X_n\left(1 - \frac{1}{2}X_n\right),$$

we get (15). In particular, $\mathbb{E}^1(X_1) = 3/4$ and $n \mapsto \mathbb{E}^1(X_n)$ is decreasing. Similarly,

$$\mathbb{E}^1(X_n^2) = \frac{1}{2}\mathbb{E}^1(X_{n-1}) + \frac{1}{4}\mathbb{E}^1\left(X_{n-1}^2(1 - \frac{1}{2}X_{n-1})\right) \geqslant \frac{1}{2}\mathbb{E}^1(X_{n-1}) \geqslant \frac{1}{2}\mathbb{E}^1(X_n).$$

As a consequence, for any $n \geqslant 1$,

$$\mathbb{E}^1(X_{n+1}) \leqslant \frac{7}{8}\mathbb{E}^1(X_n).$$

and (16) follows since for any $n \geqslant 1$,

$$\mathbb{E}^1(X_n) \leqslant \frac{6}{7}\left(\frac{7}{8}\right)^n.$$

$\square$

Acknowledgments. The authors are grateful to the anonymous referees for their quick reports and fine comments which helped to improve the paper. DC and FM would like to thank Philippe Robert and his team for kind hospitality during their visits to INRIA Rocquencourt in December 2007 and June 2009.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000, With a preface by D. Bakry and M. Ledoux. MR MR1845806 (2002g:46132) 3.2, 3.4

[2] 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

[3] O. L. V. Costa and F. Dufour, *Stability and ergodicity of piecewise deterministic Markov processes*, SIAM J. Control Optim. **47** (2008), no. 2, 1053–1077. MR MR2385873 (2009b:93163) 1, 2

[4] 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

[5] 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, 2, 3

[6] C. M. Goldie and R. Grübel, *Perpetuities with thin tails*, Adv. in Appl. Probab. **28** (1996), no. 2, 463–480. MR MR1387886 (97f:60124) 3, 3.1

[7] C. M. Goldie and R. A. Maller, *Stability of perpetuities*, Ann. Probab. **28** (2000), no. 3, 1195–1218. MR MR1797309 (2003b:60045) 3, 3.1

[8] C. Graham and Ph. Robert, *Interacting multi-class transmissions in large stochastic networks*, preprint `arXiv:0810.0347 [math.PR]`, 2008. 2

[9] L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083. MR MR0420249 (54 #8263) 3.2

[10] 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, 1, 2

[11] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR MR1849347 (2003k:28019) 3

[12] A. H. Löpker and J. S. H. van Leeuwaarden, *Transient moments of the TCP window size process*, J. Appl. Probab. **45** (2008), no. 1, 163–175. 4, 4

[13] 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

[14] K. Maulik and B. Zwart, *An extension of the square root law of TCP*, to appear in Annals of Operations Research, available on `http://www.isical.ac.in/~krishanu/`. 1, 2

[15] ———, *Tail asymptotics for exponential functionals of Lévy processes*, Stochastic Process. Appl. **116** (2006), no. 2, 156–177. MR MR2197972 (2007g:60049) 1, 2

[16] 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

[17] T. J. Ott and J. H. B. Kemperman, *Transient behavior of processes in the TCP paradigm*, Probab. Engrg. Inform. Sci. **22** (2008), no. 3, 431–471. MR MR2426601 4, 4

[18] T. J. Ott, J. H. B. Kemperman, and M. Mathis, *The stationary behavior of ideal TCP congestion avoidance*, unpublished manuscript available at `http://www.teunisott.com/`, 1996. 1

[19] 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) 2

[20] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR MR1964483 (2004e:90003) 2, 3

Djalil Chafaï, `mailto:djalil(AT)chafai(DOT)net`

UMR 8050 CNRS Laboratoire d'Analyse et de Mathématiques Appliquées
Université Paris-Est Marne-la-Vallée
5 bd Descartes, Champs-sur-Marne, F-77454 Cedex 2, France.

Florent Malrieu, corresponding author, `mailto:florent.malrieu(AT)univ-rennes1(DOT)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(DOT)fr`

UMR 6623 CNRS Laboratoire de Mathématiques
Université de Franche-Comté, F-25030 Besançon Cedex, France

Present adress UMR 6625 CNRS Institut de Recherche Mathématique de Rennes (IRMAR)
Université de Rennes I, Campus de Beaulieu, F-35042 Rennes Cedex, France.

`Compiled December 15, 2009.`