Total variation estimates for the TCP process

Jean-Baptiste Bardet[*]   Alejandra Christen[†]   Arnaud Guillin[‡]
Florent Malrieu[§]   Pierre-André Zitt[¶]

Abstract

The TCP window size process appears in the modelling 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. The sample paths are piecewise linear deterministic and the whole randomness of the dynamics comes from the jump mechanism. The aim of the present paper is to provide quantitative estimates for the exponential convergence to equilibrium, in terms of the total variation and Wasserstein distances.

1 Introduction and main results

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 divided by 2 (detection of a congestion). As shown in [9, 12, 26], a correct scaling of this process leads to a continuous time Markov process, called the 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

$$Lf(x) = f'(x) + x(f(x/2) - f(x)). \tag{1.1}$$

[*]Université de Rouen, France. E-mail: jean-baptiste.bardet@univ-rouen.fr
[†]Pontificia Universidad Católica de Valparaíso, Chile. E-mail: achristen@dim.uchile.cl
[‡]Université Blaise Pascal, France. E-mail: guillin@math.univ-bpclermont.fr
[§]Université de Rennes 1, France. E-mail: florent.malrieu@univ-rennes1.fr
[¶]Université de Bourgogne, France. E-mail: pierre-andre.zitt@u-bourgogne.fr

The semigroup $(P_t)_{t \geqslant 0}$ associated to $(X_t)_{t \geqslant 0}$ is classically defined by

$$P_t f(x) = \mathbb{E}(f(X_t)|X_0 = x),$$

for any smooth function f. Moreover, for any probability measure ν on $[0, \infty)$, νP_t stands for the law of X_t when X_0 is distributed according to ν.

The process X_t increases linearly between jump times that occur with a state-dependent rate X_t. The first jump time of X starting at $x \geqslant 0$ has the law of $\sqrt{2E + x^2} - x$ where E is a random variable with exponential law of parameter 1. In other words, the law of this jump time has a density with respect to the Lebesgue measure on $(0, +\infty)$ given by

$$f_x : t \in (0, +\infty) \mapsto (x + t)e^{-t^2/2 - xt}, \tag{1.2}$$

see [6] for further details.

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). In this direction, [23] gives a generalisation of the scaling procedure to interpret various PDMPs as limits of discrete time Markov chains. In the real world (Internet), the AIMD mechanism allows a good compromise between the minimisation of network congestion time and the maximisation of mean throughput. One could consider more general processes (introducing a random multiplicative factor or modifying the jump rate) but their study is essentially the same than the one of the present process.

The TCP window size process X is ergodic and admits a unique invariant law μ, as can be checked using a suitable Lyapunov function (for instance $V(x) = 1 + x$, see e.g. [2, 8, 24, 14] for the Meyn-Tweedie-Foster-Lyapunov technique). It can also be shown that μ has a density on $(0, +\infty)$ given by

$$x \in (0, +\infty) \mapsto \frac{\sqrt{2/\pi}}{\prod_{n \geqslant 0} \left(1 - 2^{-(2n+1)}\right)} \sum_{n \geqslant 0} \frac{(-1)^n 2^{2n}}{\prod_{k=1}^{n} \left(2^{2k} - 1\right)} e^{-2^{2n-1}x^2}, \tag{1.3}$$

(this explicit formula is derived in [9, Prop. 10], see also [12, 23, 22, 11] for further details). In particular, one can notice that the density of μ has a Gaussian tail at $+\infty$ and that all its derivatives are null at the origin. Nevertheless, this process is irreversible since time reversed sample paths are not sample paths and it has infinite support (see [20] for the description of the reversed process). In [29], explicit bounds for the exponential rate of convergence to equilibrium in total variation distance are provided for generic Markov processes in terms of a suitable Lyapunov function but these estimates are very poor even for classical examples as the Ornstein-Uhlenbeck process. They can be improved following [30] if the process under study is *stochastically monotone* that is if its semigroup $(P_t)_{t \geqslant 0}$ is such that $x \mapsto P_t f(x)$ is non decreasing as soon as f is nondecreasing. Unfortunately, due to the fact that the jump rate is an nondecreasing function of the position, the TCP process is not stochastically monotone. Moreover, we will see that our coupling provides better estimates for the example studied in [30].

The work [6] was a first attempt to use the specific dynamics of the TCP process to get explicit rates of convergence of the law of X_t to the invariant measure μ. The answer was partial and a bit disappointing since the authors did not succeed in proving explicit exponential rates.

Our aim in the present paper is to go one step further providing exponential rate of convergence for several classical distances: Wasserstein distance of order $p \geqslant 1$ and total variation distance. Let us recall briefly some definitions.

Definition 1.1. *If ν and $\tilde{\nu}$ are two probability measures on $\mathbb{R}$, we will call a* coupling *of ν and $\tilde{\nu}$ any probability measure on $\mathbb{R} \times \mathbb{R}$ such that the two marginals are ν and $\tilde{\nu}$. Let us denote by $\Gamma(\nu, \tilde{\nu})$ the set of all the couplings of ν and $\tilde{\nu}$.*

Definition 1.2. *For every $p \geqslant 1$, the* Wasserstein distance W_p *of order p between two probability measures ν and $\tilde{\nu}$ on $\mathbb{R}$ with finite p^{th} moment is defined by*

$$W_p(\nu, \tilde{\nu}) = \left(\inf_{\Pi \in \Gamma(\nu, \tilde{\nu})} \int_{\mathbb{R}^2} |x - y|^p \, \Pi(dx, dy) \right)^{1/p}. \tag{1.4}$$

Definition 1.3. *The* total variation distance *between two probability measures ν and $\tilde{\nu}$ on $\mathbb{R}$ is given by*

$$\|\nu - \tilde{\nu}\|_{\text{TV}} = \inf_{\Pi \in \Gamma(\nu, \tilde{\nu})} \int_{\mathbb{R}^2} \mathbb{1}_{\{x \neq y\}} \, \Pi(dx, dy).$$

It is well known that, for any $p \geqslant 1$, the convergence in Wasserstein distance of order p is equivalent to weak convergence together with convergence of all moments up to order p, see e.g. [28, 31]. A sequence of probability measures $(\nu_n)_{n \geqslant 1}$ bounded in L^p which converges to ν in total variation norm converges also for the W_p metrics. The converse is false: if $\nu_n = \delta_{1/n}$ then $(\nu_n)_{n \geqslant 1}$ converges to δ_0 for the distance W_p whereas $\|\nu_n - \delta_0\|_{\text{TV}}$ is equal to 1 for any $n \geqslant 1$.

Any coupling $(X, \tilde{X})$ of $(\nu, \tilde{\nu})$ provides an upper bound for these distances. One can find in [18] a lot of efficient ways to construct smart couplings in many cases. In the present work, we essentially use the coupling that was introduced in [6]. Firstly, we improve the estimate for its rate of convergence in Wasserstein distances from polynomial to exponential bounds.

Theorem 1.4. *Let us define*

$$M = \frac{\sqrt{2}(3 + \sqrt{3})}{8} \sim 0.84 \quad and \quad \lambda = \sqrt{2}(1 - \sqrt{M}) \sim 0.12. \tag{1.5}$$

For any $\tilde{\lambda} < \lambda$, any $p \geqslant 1$ and any $t_0 > 0$, there is a constant $C = C(p, \tilde{\lambda}, t_0)$ such that, for any initial probability measures ν and $\tilde{\nu}$ and any $t \geqslant t_0$,

$$W_p(\nu P_t, \tilde{\nu} P_t) \leqslant C \exp\left(-\frac{\tilde{\lambda}}{p} t \right).$$

Secondly, we introduce a modified coupling to get total variation estimates.

Theorem 1.5. *For any $\tilde{\lambda} < \lambda$ and any $t_0 > 0$, there exists C such that, for any initial probability measures ν and $\tilde{\nu}$ and any $t \geqslant t_0$,*

$$\|\nu P_t - \tilde{\nu} P_t\|_{\text{TV}} \leqslant C \exp\left(-\frac{2\tilde{\lambda}}{3} t \right),$$

where λ is given by (1.5).

Remark 1.6. *In both Theorems 1.4 and 1.5, no assumption is required on the moments nor regularity of the initial measures. Note however that following Remark 3.4, one can obtain contraction's type bounds when the initial measures ν and $\tilde{\nu}$ have initial moments of sufficient orders. In particular they hold uniformly over the Dirac measures. If $\tilde{\nu}$ is chosen to be the invariant measure μ, these theorems provide exponential convergence to equilibrium.*

The remainder of the paper is organised as follows. We derive in Section 2 precise upper bounds for the moments of the invariant measure μ and the law of X_t. Section 3 and Section 4 are respectively devoted to the proofs of Theorem 1.4 and Theorem 1.5. Unlike the classical approach "à la" Meyn-Tweedie, our total variation estimate is obtained by applying a Wasserstein coupling for most of the time, then trying to couple the two paths in one attempt. This idea is then adapted to others processes: Section 5 deals with two simpler PDMPs already studied in [27, 17, 6, 30] and Section 6 is dedicated to diffusion processes.

2 Moment estimates

The aim of this section is to provide accurate bounds for the moments of X_t. In particular, we establish below that any moment of X_t is bounded uniformly over the initial value X_0. Let $p > 0$ and $\alpha_p(t) = \mathbb{E}(X_t^p)$. Then one has by direct computation

$$\alpha_p'(t) = p\alpha_{p-1}(t) - \left(1 - 2^{-p}\right)\alpha_{p+1}(t). \tag{2.1}$$

2.1 Moments of the invariant measure

Equation (2.1) implies in particular that, if m_p denotes the p-th moment of the invariant measure μ of the TCP process ($m_p = \int x^p \mu(dx)$), then for any $p > 0$

$$m_{p+1} = \frac{p}{1 - 2^{-p}} m_{p-1}.$$

It gives all even moments of μ: $m_2 = 2$, $m_4 = \frac{48}{7}$, ... and all the odd moments in terms of the mean. Nevertheless, the mean itself cannot be explicitly determined. Applying the same technique to $\log X_t$, one gets the relation $\log(2)m_1 = m_{-1}$. With Jensen's inequality, this implies that $1/\sqrt{\log 2} \leqslant m_1 \leqslant \sqrt{2}$.

2.2 Uniform bounds for the moments at finite times

The fact that the jump rate goes to infinity at infinity gives bounds on the moments at any positive time that are uniform over the initial distribution.

Lemma 2.1. *For any $p \geqslant 1$ and $t > 0$*

$$M_{p,t} := \sup_{x \geqslant 0} \mathbb{E}_x(X_t^p) \leqslant \left(\sqrt{2p} + \frac{2p}{t}\right)^p.$$

Proof. One deduces from (2.1) and Jensen's inequality that

$$\alpha_p'(t) \leqslant p\alpha_p(t)^{1-1/p} - (1 - 2^{-p})\alpha_p(t)^{1+1/p}.$$

Let $\beta_p(t) = \alpha_p(t)^{1/p} - \sqrt{2p}$. Then, using the fact that $1 - 2^{-p} \geqslant \frac{1}{2}$,

$$\beta_p'(t) = \frac{1}{p}\alpha_p(t)^{1/p-1}\alpha_p'(t) \leqslant 1 - \frac{1}{2p}\alpha_p(t)^{2/p} = -\sqrt{\frac{2}{p}}\beta_p(t) - \frac{\beta_p(t)^2}{2p}.$$

In particular, $\beta_p'(t) < 0$ as soon as $\beta_p(t) > 0$.

Let us now fix $t > 0$. If $\beta_p(t) \leqslant 0$, then $\alpha_p(t) \leqslant (2p)^{p/2}$ and the lemma is proven. We assume now that $\beta_p(t) > 0$. By the previous remark, this implies that the function $s \mapsto \beta_p(s)$ is strictly decreasing, hence positive, on the interval $[0, t]$. Consequently, for any $s \in [0, t]$,

$$\beta_p'(s) \leqslant -\frac{\beta_p(s)^2}{2p}.$$

Integrating this last inequality gives

$$\frac{1}{\beta_p(t)} \geqslant \frac{1}{\beta_p(0)} + \frac{t}{2p} \geqslant \frac{t}{2p},$$

hence the lemma. $\qquad\square$

Let us derive from Lemma 2.1 some upper bounds for the right tails of $\delta_x P_t$ and μ.

Corollary 2.2. *For any $t > 0$ and $r \geqslant 2e(1 + 1/t)$, one has*

$$\mathbb{P}_x(X_t \geqslant r) \leqslant \exp\left(-\frac{t}{2e(t+1)}r\right). \tag{2.2}$$

Moreover, if X is distributed according to the invariant measure μ then, for any $r \geqslant \sqrt{2e}$, one has

$$\mathbb{P}(X \geqslant r) \leqslant \exp\left(-\frac{r^2}{4e}\right).$$

Proof. Let $t > 0$ and $a = 2(1 + 1/t)$. Notice that, for any $p \geqslant 1$, $\mathbb{E}_x(X_t^p)$ is smaller than $(ap)^p$. As a consequence, for any $p \geqslant 1$ and $r \geqslant 0$,

$$\mathbb{P}_x(X_t \geqslant r) \leqslant \exp\left(p \log(ap/r)\right).$$

Assuming that $r \geqslant ea$, we let $p = r/(ea)$ to get:

$$\mathbb{P}_x(X_t \geqslant r) \leqslant e^{-r/(ea)}.$$

For the invariant measure, the upper bound is better: $\mathbb{E}(X^p) \leqslant (2p)^{p/2}$. Then, the Markov inequality provides that, for any $p \geqslant 1$,

$$\mathbb{P}(X \geqslant r) \leqslant \exp\left(p \log \frac{\sqrt{2p}}{r}\right).$$

As above, if $r^2 \geqslant 2e$, one can choose $p = r^2/(2e)$ to get the desired bound. $\qquad\square$

Remark 2.3. *A better deviation bound should be expected from the expression* (1.3) *of the density of μ. Indeed, one can get a sharp result (see [6]). However, in the sequel we only need the deviation bound* (2.2).

3 Exponential convergence in Wasserstein distance

This section is devoted to the proof of Theorem 1.4. We use the coupling introduced in [6]. Let us briefly recall the construction and the dynamics of this stochastic process on $\mathbb{R}_+^2$ whose marginals are two TCP processes. It is defined by the following generator

$$\mathfrak{L}f(x,y) = (\partial_x + \partial_y)f(x,y) + y\big(f(x/2, y/2) - f(x,y)\big) + (x - y)\big(f(x/2, y) - f(x,y)\big)$$

when $x \geqslant y$ and symmetric expression for $x < y$. We will call the dynamical coupling defined by this generator the Wasserstein coupling of the TCP process (see Figure 1 for a graphical illustration of this coupling). This coupling is the only one such that the lower component never jumps alone. Let us give the pathwise interpretation of this coupling. Between two jump times, the two coordinates increase linearly with rate 1. Moreover, two "jump processes" are simultaneously in action:

1. with a rate equal to the minimum of the two coordinates, they jump (*i.e.* they are divided by 2) simultaneously,
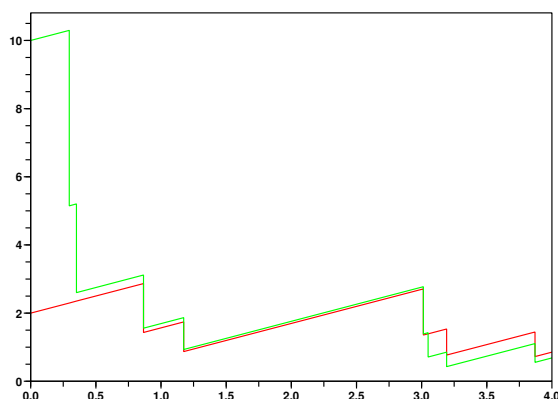
Figure 1: Two trajectories following the Wasserstein coupling; the bigger jumping alone can be good, making the distance between both trajectories smaller, or bad.

2. with a rate equal to the distance between the two coordinates (which is constant between two jump times), the bigger one jumps alone.

The existence of non-simultaneous jumps implies that the generator does not act in a good way on functions directly associated to Wasserstein distances. To see this, let us define $V_p(x, y) = |x - y|^p$. When we compute $\mathfrak{L}V_p$, the term coming from the deterministic drift disappears (since V_p is unchanged under the flow), and we get for $x/2 \leqslant y \leqslant x$:

$$\mathfrak{L}V_p(x, y) = -y(1 - 2^{-p})V_p(x, y) + (x - y)((y - x/2)^p - (x - y)^p).$$

For example, choosing $p = 1$ gives:

$$\mathfrak{L}V_1(x, y) = -V_1(x, y)(y/2 - (y - x/2) + (x - y)) = -(3/2)V_1(x, y)^2.$$

This shows that $\mathbb{E}[|X_t - Y_t|]$ decreases, but only gives a polynomial bound: the problem comes from the region where $x - y$ is already very small.

Choosing $p = 2$, we get

$$\mathfrak{L}V_2(x, y) = -(3/4)yV_2(x, y) + (x - y)(-(3/4)x^2 + xy).$$

The effect of the jumps of X is even worse: if $x = 1$ and $x - y$ is small, $\mathfrak{L}V_2(x, y)$ is positive ($\approx (1/4)(x - y)$).

For $p = 1/2$, the situation is in fact more favourable: indeed, for $0 < y \leqslant x$,

$$\mathfrak{L}V_{1/2}(x, y) = -y\left(1 - \frac{\sqrt{2}}{2}\right)V_{1/2}(x, y) + (x - y)(\sqrt{|y - x/2|} - \sqrt{x - y})$$

$$= -V_{1/2}(x, y)\left[x - \frac{\sqrt{2}}{2}y - \sqrt{(x - y)(y - x/2)}\right]$$

$$= -x\left[1 - \varphi\left(\frac{y}{x}\right)\right]V_{1/2}(x, y),$$

with

$$\varphi(u) = \frac{\sqrt{2}}{2}u + \sqrt{(1 - u)|u - 1/2|} \quad \text{for} \quad u \in [0, 1].$$

By a direct computation, one gets that

$$M := \max_{0 \leqslant u \leqslant 1} \varphi(u) = \varphi\left(\frac{9 + \sqrt{3}}{12}\right) = \frac{\sqrt{2}(3 + \sqrt{3})}{8} \sim 0.8365.$$

Hence, when $0 < y \leqslant x$,

$$\mathfrak{L}V_{1/2}(x,y) \leqslant -x\lambda V_{1/2}(x,y), \tag{3.1}$$

with $\lambda = 1 - M \sim 0.1635$. This would give an exponential decay for $V_{1/2}$ if x was bounded below: the problem comes from the region where x is small and the process does not jump.

To overcome this problem, we replace $V_{1/2}$ with the function

$$\widetilde{V}(x,y) = \psi(x \vee y)|x-y|^{1/2} = \psi(x \vee y)V_{1/2}(x,y),$$

where

$$\psi(x) = \begin{cases} 1 + \alpha(1 - x/x_0)^2, & x \leqslant x_0, \\ 1, & x \geqslant x_0, \end{cases}$$

the two positive parameters α and x_0 being chosen below. The negative slope of ψ for x small will be able to compensate the fact that the bound from (3.1) tends to 0 with x, hence to give a uniform bound. Indeed, for $0 < y \leqslant x$,

$$
\begin{aligned}
\mathfrak{L}\widetilde{V}(x,y) &= \psi'(x)(x-y)^p + y\psi(x/2)\frac{(x-y)^p}{2^p} + (x-y)\psi(x/2 \vee y)|x/2-y|^p - x\widetilde{V}(x,y) \\
&= -\widetilde{V}(x,y)\left[-\frac{\psi'(x)}{\psi(x)} + x - \frac{\psi(x/2)}{\psi(x)}\frac{y}{2^p} - \frac{\psi(x/2 \vee y)}{\psi(x)}(x-y)^{1-p}|x/2-y|^p\right] \\
&\leqslant -\widetilde{V}(x,y)\left[-\frac{\psi'(x)}{\psi(x)} + x\left(1 - \frac{\psi(x/2)}{\psi(x)}\varphi\left(\frac{y}{x}\right)\right)\right] \\
&\leqslant -\widetilde{V}(x,y)\left[-\frac{\psi'(x)}{(1+\alpha)} + x\left(1 - (1+\alpha)M\right)\right] \\
&= \begin{cases} -\widetilde{V}(x,y)\left[\frac{2\alpha}{x_0(1+\alpha)} + x\left(1 - (1+\alpha)M - \frac{2\alpha}{x_0^2(1+\alpha)}\right)\right] & \text{when } x \leqslant x_0, \\ -\widetilde{V}(x,y)x\left(1 - (1+\alpha)M\right) & \text{when } x \geqslant x_0. \end{cases}
\end{aligned}
$$

Finally, as soon as $(1+\alpha)M < 1$, one has

$$\mathfrak{L}\widetilde{V}(x,y) \leqslant -\lambda_{\alpha,x_0}\widetilde{V}(x,y) \qquad \forall x, y > 0$$

with

$$\lambda_{\alpha,x_0} = \min\left(\frac{2\alpha}{x_0(1+\alpha)}, x_0(1 - (1+\alpha)M)\right).$$

By direct computations, one gets that the best possible choice of parameters α and x_0 is $\alpha = 1/\sqrt{M} - 1 \sim 0.0934$ and $x_0 = \sqrt{2}$. We obtain finally, for any $x, y > 0$,

$$\mathfrak{L}\widetilde{V}(x,y) \leqslant -\lambda\widetilde{V}(x,y), \tag{3.2}$$

with $\lambda = \lambda_{\alpha,x_0} = \sqrt{2}(1 - \sqrt{M}) \sim 0.1208$. Hence, directly from (3.2), for any $x, y > 0$

$$\mathbb{E}_{x,y}[\widetilde{V}(X_t, Y_t)] \leqslant e^{-\lambda t}\widetilde{V}(x,y). \tag{3.3}$$

Immediate manipulations lead to the following estimate.

Proposition 3.1. *Let*

$$M = \frac{\sqrt{2}(3 + \sqrt{3})}{8} \sim 0.84 \quad \text{and} \quad \lambda = \sqrt{2}(1 - \sqrt{M}) \sim 0.12. \tag{3.4}$$

Then, for any $x, y > 0$, one has

$$\mathbb{E}_{x,y}\left(|X_t - Y_t|^{1/2}\right) \leqslant \frac{1}{\sqrt{M}}e^{-\lambda t}|x-y|^{1/2}. \tag{3.5}$$

Moreover, for any initial conditions (X_0, Y_0) and any $t \geqslant t_0 > 0$,

$$\mathbb{E}\left(|X_t - Y_t|^{1/2}\right) \leqslant \sqrt{\frac{2\sqrt{2} + 4t_0^{-1}}{M}}e^{-\lambda(t - t_0)}. \tag{3.6}$$

Proof. Equation (3.5) is a straightforward consequence of Equation (3.2) since $\psi(\mathbb{R}^+)$ is the interval $[1, 1 + \alpha]$. As a consequence, for any $t \geqslant t_0 > 0$, one has

$$\mathbb{E}\left(|X_t - Y_t|^{1/2}\right) \leqslant \frac{1}{\sqrt{M}}e^{-\lambda(t - t_0)}\mathbb{E}\left(|X_{t_0} - Y_{t_0}|^{1/2}\right)$$
$$\leqslant \frac{e^{\lambda t_0}}{\sqrt{M}}\sqrt{\mathbb{E}(X_{t_0}) + \mathbb{E}(Y_{t_0})}e^{-\lambda t}.$$
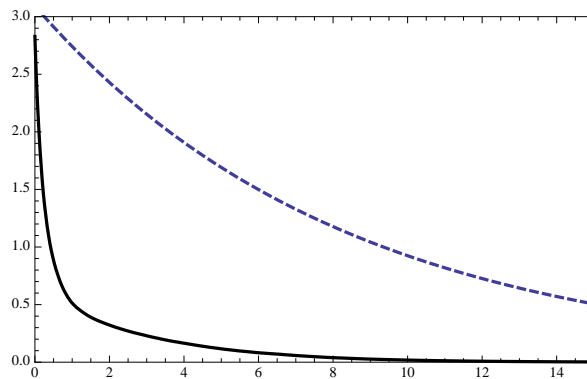
Then, Lemma 2.1 provides the estimate (3.6). $\qquad\square$

Figure 2: The "true" function $\mathbb{E}_{x,y}\left(|X_t - Y_t|^{1/2}\right)$ (solid, black; by Monte Carlo method with $m = 10\,000$ copies) and the bound given in Equation (3.5) (dashed, blue), for $x = 2$ and $y = 10$.

Remark 3.2. *The upper bound obtained in Proposition 3.1 is compared graphically to the "true" function $t \mapsto \mathbb{E}_{x,y}\left(|X_t - Y_t|^{1/2}\right)$ in Figure 2. By linear regression on this data, one gets that the exponential speed of convergence of this function is on the order of 0.4.*
Note also that this method can be adapted to any V_p with $0 < p < 1$, giving even better (but less explicit) speed of convergence for some $p \neq \frac{1}{2}$: we estimated numerically that the best value for λ would be approximately 0.1326, obtained for p close to $2/3$.

We may now deduce from Proposition 3.1 estimates for the Wasserstein distance between the laws of X_t and Y_t.

Theorem 3.3. *Let $p \geqslant 1$. Then, for any $t_0 > 0$ and any $\theta \in (0, 1)$, there exists a finite constant $C(p, t_0, \theta)$ such that, for any initial conditions (X_0, Y_0) and for all $t \geqslant t_0$,*

$$\mathbb{E}(|X_t - Y_t|^p) \leqslant C(p, t_0, \theta)\exp\left(-\lambda\theta t\right)$$

where λ is defined by (3.4).

Proof of Theorem 3.3. Let $p \geqslant 1$. For any $0 < \theta < 1$, Hölder's inequality gives, for any $t \geqslant 0$,

$$\mathbb{E}(|X_t - Y_t|^p) \leqslant \left[\mathbb{E}\left(|X_t - Y_t|^{\frac{2p - \theta}{2(1 - \theta)}}\right)\right]^{1 - \theta}\left[\mathbb{E}\left(|X_t - Y_t|^{1/2}\right)\right]^{\theta}.$$

Thanks to Lemma 2.1 and the inequality $(a+b)^q \leqslant 2^{q-1}(a^q + b^q)$, when $q \geqslant 1$, one gets

$$\left[\mathbb{E}\left(|X_t - Y_t|^{\frac{2p-\theta}{2(1-\theta)}}\right)\right]^{1-\theta} \leqslant \left(2^{\frac{2p-\theta}{2(1-\theta)}} M_{\frac{2p-\theta}{2(1-\theta)},t}\right)^{1-\theta}$$

$$\leqslant 2^{p-\theta/2}\left(\sqrt{\frac{2p-\theta}{1-\theta}} + \frac{2p-\theta}{(1-\theta)t}\right)^{p-\theta/2}.$$

Then, it suffices to use Equation (3.6) to conclude the proof with

$$C(p, t_0, \theta) = \frac{2^p}{M^{\theta/2}}\left(\sqrt{\frac{2p-\theta}{1-\theta}} + \frac{2p-\theta}{(1-\theta)t_0}\right)^{p-\theta/2}\left(\sqrt{2} + \frac{2}{t_0}\right)^{\theta/2} e^{\lambda\theta t_0}.$$

$\square$

Since $W_p(\mathcal{L}(X_t), \mathcal{L}(Y_t))^p \leqslant \mathbb{E}(|X_t - Y_t|^p)$, Theorem 1.4 is a direct consequence of this result.

Remark 3.4. *Let us remark that we can obtain "contraction's type bounds" using Equation (3.5) instead of (3.6) : for any $p \geqslant 1$, any $0 < \theta < 1$, any $t \geqslant 0$ and any $x, y > 0$,*

$$\mathbb{E}_{x,y}(|X_t - Y_t|^p) \leqslant \frac{2^{p-\theta/2}}{M^{\theta/2}}\left(\sqrt{\frac{2p-\theta}{1-\theta}} + \frac{2p-\theta}{(1-\theta)t}\right)^{p-\theta/2} e^{-\lambda\theta t} |x - y|^{\theta/2}.$$

We then obtain that if ν and $\tilde{\nu}$ have finite $\theta/2$-moments then for $p \geqslant 1$ and $t \geqslant 0$,

$$W_p(\nu P_t, \tilde{\nu} P_t) \leqslant \left(\frac{2^{p-\theta/2}}{M^{\theta/2}}\left(\sqrt{\frac{2p-\theta}{1-\theta}} + \frac{2p-\theta}{(1-\theta)t}\right)^{p-\theta/2}\right)^{\frac{1}{p}} e^{-\frac{\lambda\theta}{p}t} W_{\frac{\theta}{2}}(\nu, \tilde{\nu})^{\frac{2}{p\theta}}.$$

which still allows a control by some Wasserstein "distance" (in fact, this is not a distance, since $\theta/2 < 1$) of the initial measures.

Remark 3.5. *We estimated numerically the exponential speed of convergence:*

- *of the function $t \mapsto \mathbb{E}_{x,y}(|X_t - Y_t|)$ for the Wasserstein coupling (by Monte Carlo method and linear regression). It seems to be on the order of 0.5 (we obtained 0.48 for $x = 2$, $y = 10$, $m = 10\,000$ copies, and linear regression between times 2 and 10);*
- *of the Wasserstein distance $t \mapsto W_1(\delta_x P_t, \delta_y P_t)$, using the explicit representation of this distance for measures on $\mathbb{R}$ to approximate it by the L^1-distance between the two empirical quantile functions. It is on the order of 1.6 (we get 1.67 for $x = 2$, $y = 0.5$, $m = 1\,000\,000$ copies, and linear regression on 20 points until time 4).*

In conclusion, our bound from Theorem 3.3 seems reasonable (at least when compared to those given by [29], see section 4.2 below), but is still off by a factor of 4 from the true exponential speed of convergence. Since the coupling itself seems to converge approximately 3 times slower than the true Wasserstein distance, one probably needs to find another coupling to get better bounds.

Remark 3.6. *Let us end this section by advertising on a parallel work by B. Cloez [7] who uses a completely different approach, based on a particular Feynman-Kac formulation, to get some related results.*

4 Exponential convergence in total variation distance

In this section, we provide the proof of Theorem 1.5 and we compare our estimate to the ones that can be deduced from [29].

4.1 From Wasserstein to total variation estimate

The fact that the evolution is partially deterministic makes the study of convergence in total variation quite subtle. Indeed, the law $\delta_x P_t$ can be written as a mixture of a Dirac mass at $x + t$ (the first jump time has not occurred) and an absolutely continuous measure (the process jumped at some random time in $(0, t)$):

$$\delta_x P_t = p_t(x)\delta_{x+t} + (1 - p_t(x))\mathcal{L}(X_t | X_0 = x, T_1 \leqslant t)$$

where, according to Equation (1.2),

$$p_t(x) := \mathbb{P}_x(T_1 > t) = e^{-t^2/2 - xt}. \tag{4.1}$$

This implies that the map $y \mapsto \|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}}$ is not continuous at point x since one has, for any $y \neq x$,

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \geqslant p_t(x) \vee p_t(y) = e^{-t^2/2 - (x \wedge y)t}.$$

Nevertheless, one can hope that

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \underset{y \to x}{\sim} p_t(x).$$

The lemma below makes this intuition more precise.

Lemma 4.1. *Let $t \geqslant \varepsilon \geqslant x - y > 0$. There exists a coupling $\left((X_t)_{t \geqslant 0}, (Y_t)_{t \geqslant 0}\right)$ of two processes driven by (1.1) starting at (x, y) such that the probability $\mathbb{P}(X_s = Y_s, \, s \geqslant t)$ is larger than*

$$q_t(x, y) = \left(\int_{x-y}^t (f_x(s) \wedge f_y(s - x + y)) \, ds\right) p_{x-y}\left(\frac{x+t}{2}\right)$$

$$\geqslant (p_\varepsilon(x) - p_t(x) - 2\varepsilon\alpha(x)) p_\varepsilon\left(\frac{x+t}{2}\right), \tag{4.2}$$

where f_x is defined in Equation (1.2) and $\alpha(x) := \int_0^\infty e^{-u^2/2 - ux} \, du$.

Moreover, for any $x_0 > 0$ and $\varepsilon > 0$, let us define

$$A_{x_0, \varepsilon} = \{(x, y) \, : \, 0 \leqslant x, y \leqslant x_0, \, |x - y| \leqslant \varepsilon\}. \tag{4.3}$$

Then,

$$\inf_{(x,y) \in A_{x_0, \varepsilon}} q_t(x, y) \geqslant \exp\left(-\varepsilon^2 - \frac{3x_0 + t}{2}\varepsilon\right) - e^{-t^2/2} - \sqrt{2\pi}\varepsilon.$$

Proof. The idea is to construct an explicit coalescent coupling starting from x and y. The main difficulty comes from the fact that the jump are deterministic. Assume for simplicity that $y < x$. Let us denote by $(T_k^x)_{k \geqslant 1}$ and $(T_k^y)_{k \geqslant 1}$ the jump times of the two processes. If

$$T_1^x = T_1^y + x - y \quad \text{and} \quad T_2^y - T_1^y \geqslant x - y \tag{4.4}$$

then the two paths are at the same place at time T_1^x since in this case

$$X_{T_1^x} = \frac{x + T_1^x}{2} = \frac{y + T_1^y}{2} + T_1^x - T_1^y = Y_{T_1^x}.$$

The law of T_1^x has a density f_x given by (1.2). As a consequence, the density of $T_y^1 + x - y$ is given by $s \mapsto f_y(s - x + y)$. The best way of obtaining the first equality in (4.4) before a time $t \geqslant x - y$ is to realise an optimal coupling of the two continuous laws of T_1^x and $T_1^y + x - y$. This makes these random variables equal with probability

$$I(x, y, t) = \int_{x-y}^t (f_x(s) \wedge f_y(s - x + y)) \, ds.$$

Assume now that $0 \leqslant x - y \leqslant \varepsilon \leqslant t$. For any $s \geqslant x - y$, one has

$$
\begin{aligned}
f_y(s - x + y) &= (2y - x + s) \exp\left(-\frac{1}{2}(s - x + y)^2 - y(s - x + y)\right) \\
&= (s + x - 2(x - y)) \exp\left(-\frac{s^2}{2} - xs + (2s + x - 3(x - y)/2)(x - y)\right) \\
&\geqslant (s + x - 2\varepsilon) \exp\left(-\frac{s^2}{2} - xs\right).
\end{aligned}
$$

As a consequence, if $0 \leqslant x - y \leqslant \varepsilon \leqslant t$,

$$
\begin{aligned}
I(x, y, t) &\geqslant \int_{x-y}^t (s + x - 2\varepsilon) \exp\left(-\frac{s^2}{2} - xs\right) ds \\
&\geqslant p_\varepsilon(x) - p_t(x) - 2\varepsilon\alpha(x)
\end{aligned}
$$

where $p_t(x)$ is defined in (4.1) and $\alpha(x) = \int_0^\infty e^{-u^2/2 - ux} \, du$.

Finally, one has to get a lower bound for the probability of the set $\{T_2^y - T_1^y \geqslant x - y\}$. For this, we notice that

$$z \in [0, +\infty) \mapsto \mathbb{P}(T_1^z \geqslant s) = p_s(z) = e^{-s^2/2 - sz}$$

is decreasing and that $Y_{T_1^y} \leqslant (y + t)/2$ as soon as $T_1^y \leqslant t$. As a consequence,

$$\inf_{0 \leqslant z \leqslant (x+t)/2} \mathbb{P}(T_1^z \geqslant s) \geqslant p_{x-y}\left(\frac{x+t}{2}\right) \geqslant p_\varepsilon\left(\frac{x+t}{2}\right).$$

This provides the bound (4.2). The uniform lower bound on the set $A_{x_0, \varepsilon}$ is a direct consequence of the previous one. □

Let us now turn to the proof of Theorem 1.5.

Proof of Theorem 1.5. We are looking for an upper bound for the total variation distance between $\delta_x P_t$ and $\delta_y P_t$ for two arbitrary initial conditions x and y. To this end, let us consider the following coupling: during a time $t_1 < t$ we use the Wasserstein coupling. Then during a time $t_2 = t - t_1$ we try to stick the two paths using Lemma 4.1. Let $\varepsilon > 0$ and $x_0 > 0$ be as in Lemma 4.1. If, after the time t_1, one has

$$(X_{t_1}, Y_{t_1}) \in A_{\varepsilon, x_0}$$

where A_{ε, x_0} is defined by (4.3), then the coalescent coupling will work with a probability greater than the one provided by Lemma 4.1. As a consequence, the coalescent coupling occurs before time $t_1 + t_2$ with a probability greater than

$$\mathbb{P}((X_{t_1}, Y_{t_1}) \in A_{\varepsilon, x_0}) \inf_{(x,y) \in A_{\varepsilon, x_0}} q_{t_2}(x, y).$$

Moreover,

$$\mathbb{P}((X_{t_1}, Y_{t_1}) \notin A_{\varepsilon, x_0}) \leqslant \mathbb{P}(X_{t_1} \geqslant x_0) + \mathbb{P}(Y_{t_1} \geqslant x_0) + \mathbb{P}(|X_{t_1} - Y_{t_1}| \geqslant \varepsilon).$$

>From the deviation bound (2.2), we get that, for any $x_0 \geqslant 2e(1 + 1/t_1)$,

$$\mathbb{P}(X_{t_1} \geqslant x_0) \leqslant e^{-a(t_1)x_0} \quad \text{with} \quad a(t_1) = \frac{t_1}{2e(t_1 + 1)}.$$

The estimate of Proposition 3.1 concerning the Wasserstein coupling ensures that

$$\mathbb{P}(|X_{t_1} - Y_{t_1}| \geqslant \varepsilon) \leqslant \frac{C}{\sqrt{\varepsilon}} e^{-\lambda t_1} \quad \text{with} \quad C = \sqrt{\frac{2\sqrt{2} + 4t_0^{-1}}{M}} e^{\lambda t_0},$$

for any $0 < t_0 \leqslant t_1$. As a consequence, the total variation distance between $\delta_x P_t$ and $\delta_y P_t$ is smaller than

$$D := 1 - e^{-\varepsilon^2 - \frac{3x_0 + t_2}{2}\varepsilon} + e^{-t_2^2/2} + \sqrt{2\pi}\varepsilon + 2e^{-a(t_1)x_0} + \frac{C}{\sqrt{\varepsilon}} e^{-\lambda t_1}.$$

In order to get a more tractable estimate, let us assume that $t_2 \leqslant x_0$ and use that $1 - e^{-u} \leqslant u$ to get

$$D \leqslant \varepsilon^2 + 2\varepsilon x_0 + e^{-t_2^2/2} + \sqrt{2\pi}\varepsilon + 2e^{-a(t_0)x_0} + \frac{C}{\sqrt{\varepsilon}} e^{-\lambda t_1}.$$

Finally let us set

$$t_2 = \sqrt{2\log(1/\varepsilon)}, \quad t_1 = \frac{3}{2\lambda}\log(1/\varepsilon), \quad x_0 = \frac{1}{a(t_0)}\log(1/\varepsilon).$$

Obviously, for ε small enough, $x_0 \geqslant \max(t_2, 2e(1 + 1/t_1))$ and $t_1 \geqslant t_0$. Then, one gets that

$$\|\delta_x P_{t_1 + t_2} - \delta_y P_{t_1 + t_2}\|_{\mathrm{TV}} \leqslant \frac{2}{a(t_0)}\varepsilon \log(1/\varepsilon) + (3 + C + \sqrt{2\pi} + \varepsilon)\varepsilon.$$

One can now express ε as a function of t_1 to get that there exists $K = K(t_0) > 0$ such that

$$\|\delta_x P_{t_1 + t_2} - \delta_y P_{t_1 + t_2}\|_{\mathrm{TV}} \leqslant K(1 + t_1)e^{-\frac{2\lambda}{3}t_1}.$$

Since $t_2 = \sqrt{(4\lambda/3)t_1}$, one gets that

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \leqslant K(1 + t)e^{K\sqrt{t}}e^{-\frac{2\lambda}{3}t}.$$

This provides the conclusion of Theorem 1.5 when both initial measures are Dirac masses. The generalisation is straightforward. □

4.2 A bound via small sets

We describe here briefly the approach of [29] and compare it with the hybrid Wasserstein/total variation coupling described above. The idea is once more to build a successful coupling between two copies X and Y of the process. In algorithmic terms, the approach is the following:

- let X and Y evolve independently until they both reach a given set C,
- once they are in C, try to stick them together,
- repeat until the previous step is successful.

To control the time to come back to $C \times C$, [29] advocates an approach via a Lyapunov function. The second step works with positive probability if the set is "pseudo-small", i.e. if one can find a time $t^\star$, an $\alpha > 0$ and probability measures ν_{xy}

$$\forall x, y \in C^2, \quad \mathcal{L}(X_{t^\star}|X_0 = x) \geqslant \alpha\nu_{xy} \text{ and } \mathcal{L}(X_{t^\star}|X_0 = y) \geqslant \alpha\nu_{xy}. \tag{4.5}$$

The convergence result can be stated as follows.

Theorem 4.2 ([29], Theorem 3, Corollary 4 and Theorem 8). [1] *Suppose that there exists a set C, a function $V \geqslant 1$, and positive constants λ, Λ such that*

$$LV \leqslant -\lambda V + \Lambda \mathbb{1}_C. \tag{4.6}$$

Suppose that $\delta = \lambda - \Lambda / \inf_{x \notin C} V > 0$. Suppose additionally that C is pseudo-small, with constant α.

Then for $A = \frac{\Lambda}{\delta} + e^{-\delta t^\star} \sup_{x \in C} V$, $A' = A e^{\delta t^\star}$, and for any $r < 1/t^\star$,

$$\|\mathcal{L}(X_t) - \mu\|_{\mathrm{TV}} \leqslant (1 - \alpha)^{\lfloor rt \rfloor} + e^{-\delta(t - t^\star)} (A')^{\lfloor rt \rfloor - 1} \mathbb{E} V(X_0).$$

If A' is finite, this gives exponential convergence: just choose r small enough so that $(A')^{rt} e^{-\delta t}$ decreases exponentially fast.

To compute explicit bounds we have to make choices for C and V and estimate the corresponding value of α. Our best efforts for the case of the TCP process only give decay rates of the order 10^{-14}. We believe this order cannot be substantially improved even by fine-tuning C and V.

5 Two other models

This section is devoted to the study of two simple PDMPs. The first one is a simplified version of the TCP process where the jump rate is assumed to be constant and equal to λ. It has been studied with different approaches: PDE techniques (see [27, 17]) or probabilistic tools (see [21, 25, 6]). The second one is studied in [30]. It can also be checked that our method gives sharp bounds for the speed of convergence to equilibrium of the PDMP which appears in the study of a penalised bandit algorithm (see Lemma 7 in [16]).

5.1 The TCP model with constant jump rate

In this section we investigate the long time behavior of the TCP process with constant jump rate given by its infinitesimal generator:

$$Lf(x) = f'(x) + \lambda(f(x/2) - f(x)) \quad (x \geqslant 0).$$

The jump times of this process are the ones of a homogeneous Poisson process with intensity λ. The convergence in Wasserstein distance is obvious.

Lemma 5.1 ([27, 6]). *For any $p \geqslant 1$,*

$$W_p(\delta_x P_t, \delta_y P_t) \leqslant |x - y| e^{-\lambda_p t} \quad \text{with} \quad \lambda_p = \frac{\lambda(1 - 2^{-p})}{p}. \tag{5.1}$$

Remark 5.2. *The case $p = 1$ is obtained in [27] by PDEs estimates using the following alternative formulation of the Wasserstein distance on $\mathbb{R}$. If the cumulative distribution functions of the two probability measures ν and $\tilde{\nu}$ are F and $\tilde{F}$ then*

$$W_1(\nu, \tilde{\nu}) = \int_{\mathbb{R}} |F(x) - \tilde{F}(x)| \, dx.$$

The general case $p \geqslant 1$ is obvious from the probabilistic point of view: choosing the same Poisson process $(N_t)_{t \geqslant 0}$ to drive the two processes provides that the two coordinates jump simultaneously and

$$|X_t - Y_t| = |x - y| 2^{-N_t}.$$

---

[1]In fact, in [29], the result is given with A instead of A' in the upper bound. Joaquin Fontbona pointed out to us that Lemma 6 from [29] has to be corrected, adding the exponential term $e^{\delta t^\star}$ to the estimate. We thank him for this remark.

As a consequence, since the law of N_t is the Poisson distribution with parameter λt, one has

$$\mathbb{E}_{x,y}(|X_t - Y_t|^p) = |x - y|^p \mathbb{E}\left(2^{-pN_t}\right) = |x - y|^p e^{-p\lambda_p t}.$$

This coupling turns out to be sharp. Indeed, one can compute explicitly the moments of X_t (see [21, 25]): for every $n \geqslant 0$, every $x \geqslant 0$, and every $t \geqslant 0$,

$$\mathbb{E}_x((X_t)^n) = \frac{n!}{\prod_{k=1}^n \theta_k} + n! \sum_{m=1}^n \left(\sum_{k=0}^m \frac{x^k}{k!} \prod_{\substack{j=k \\ j \neq m}}^n \frac{1}{\theta_j - \theta_m} \right) e^{-\theta_m t}, \tag{5.2}$$

where $\theta_n = \lambda(1 - 2^{-n}) = n\lambda_n$ for any $n \geqslant 1$. Obviously, assuming for example that $x > y$,

$$W_n(\delta_x P_t, \delta_y P_t)^n \geqslant \mathbb{E}_x((X_t)^n) - \mathbb{E}_y((Y_t)^n)$$

$$\underset{t \to \infty}{\sim} n! \left(\sum_{k=0}^n \frac{x^k - y^k}{k!} \prod_{j=k}^{n-1} \frac{1}{\theta_j - \theta_n} \right) e^{-\theta_n t}.$$

As a consequence, the rate of convergence in Equation (5.1) is optimal for any $n \geqslant 1$.

Nevertheless this estimate for the Wasserstein rate of convergence does not provide on its own any information about the total variation distance between $\delta_x P_t$ and $\delta_y P_t$. It turns out that this rate of convergence is the one of the W_1 distance. This is established by Theorem 1.1 in [27]. It can be reformulated in our setting as follows.

Theorem 5.3 ([27]). *Let μ be the invariant measure of X. For any measure ν with a finite first moment and $t \geqslant 0$,*

$$\|\nu P_t - \mu\|_{\mathrm{TV}} \leqslant e^{-\lambda t/2} (3\lambda W_1(\nu, \mu) + \|\nu - \mu\|_{\mathrm{TV}}).$$

Let us provide here an improvement of this result by a probabilistic argument.

Proposition 5.4. *For any $x, y \geqslant 0$ and $t \geqslant 0$,*

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \leqslant \lambda e^{-\lambda t/2} |x - y| + e^{-\lambda t}. \tag{5.3}$$

As a consequence, for any measure ν with a finite first moment and $t \geqslant 0$,

$$\|\nu P_t - \mu\|_{\mathrm{TV}} \leqslant \lambda e^{-\lambda t/2} W_1(\nu, \mu) + e^{-\lambda t} \|\nu - \mu\|_{\mathrm{TV}}. \tag{5.4}$$

Remark 5.5. *Note that the upper bound obtained in Equation (5.3) is non-null even for $x = y$. This is due to the persistence of a Dirac mass at any time, which implies that taking y arbitrarily close to x for initial conditions does not make the total variation distance arbitrarily small, even for large times.*

Proof of Proposition 5.4. The coupling is a slight modification of the one used to control Wasserstein distance. The paths of $(X_s)_{0 \leqslant s \leqslant t}$ and $(Y_s)_{0 \leqslant s \leqslant t}$ starting respectively from x and y are determined by their jump times $(T_n^X)_{n \geqslant 0}$ and $(T_n^Y)_{n \geqslant 0}$ up to time t. These sequences have the same distribution than the jump times of a Poisson process with intensity λ.

Let $(N_t)_{t \geqslant 0}$ be a Poisson process with intensity λ and $(T_n)_{n \geqslant 0}$ its jump times with the convention $T_0 = 0$. Let us now construct the jump times of X and Y. Both processes make exactly N_t jumps before time t. If $N_t = 0$, then

$$X_s = x + s \quad \text{and} \quad Y_s = y + s \quad \text{for } 0 \leqslant s \leqslant t.$$

Assume now that $N_t \geqslant 1$. The $N_t - 1$ first jump times of X and Y are the ones of $(N_t)_{t \geqslant 0}$:

$$T_k^X = T_k^Y = T_k \quad 0 \leqslant k \leqslant N_t - 1.$$

In other words, the coupling used to control Wasserstein distance (see Lemma 5.1) acts until the penultimate jump time T_{N_t-1}. At that time, we have

$$X_{T_{N_t-1}} - Y_{T_{N_t-1}} = \frac{x-y}{2^{N_t-1}}.$$

Then we have to define the last jump time for each process. If they are such that

$$T^X_{N_t} = T^Y_{N_t} + X_{T_{N_t-1}} - Y_{T_{N_t-1}}$$

then the paths of X and Y are equal on the interval $(T^X_{N_t}, t)$ and can be chosen to be equal for any time larger than t.

Recall that conditionally on the event $\{N_t = 1\}$, the law of T_1 is the uniform distribution on $(0, t)$. More generally, if $n \geqslant 2$, conditionally on the set $\{N_t = n\}$, the law of the penultimate jump time T_{n-1} has a density $s \mapsto n(n-1)t^{-n}(t-s)s^{n-2}\mathbb{1}_{(0,t)}(s)$ and conditionally on the event $\{N_t = n, T_{n-1} = s\}$, the law of T_n is uniform on the interval (s, t).

Conditionally on $N_t = n \geqslant 1$ and T_{n-1}, T^X_n and T^Y_n are uniformly distributed on (T_{n-1}, t) and can be chosen such that

$$\mathbb{P}\left(T^X_n = T^Y_n + \frac{x-y}{2^{n-1}} \,\Big|\, N^X_t = N^Y_t = n,\, T^X_{n-1} = T^Y_{n-1} = T_{n-1}\right)$$

$$= \left(1 - \frac{|x-y|}{2^{n-1}(t-T_{n-1})}\right) \vee 0 \geqslant 1 - \frac{|x-y|}{2^{n-1}(t-T_{n-1})}.$$

This coupling provides that

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \leqslant 1 - \mathbb{E}\left[\left(1 - \frac{|x-y|}{2^{N_t-1}(t-T_{N_t-1})}\right)\mathbb{1}_{\{N_t \geqslant 1\}}\right]$$

$$\leqslant e^{-\lambda t} + |x-y|\mathbb{E}\left(\frac{2^{-N_t+1}}{(t-T_{N_t-1})}\mathbb{1}_{\{N_t \geqslant 1\}}\right).$$

For any $n \geqslant 2$,

$$\mathbb{E}\left(\frac{1}{t-T_{N_t-1}}\Big|N_t = n\right) = \frac{n(n-1)}{t^n}\int_0^t u^{n-2}\, du = \frac{n}{t}.$$

This equality also holds for $n = 1$. Thus we get that

$$\mathbb{E}\left(\frac{2^{-N_t+1}}{(t-T_{N_t-1})}\mathbb{1}_{\{N_t \geqslant 1\}}\right) = \frac{1}{t}\mathbb{E}\left(N_t 2^{-N_t+1}\right) = \lambda e^{-\lambda t/2},$$

since N_t is distributed according to the Poisson law with parameter λt. This provides the estimate (5.3).

To treat the case of general initial conditions and to get (5.4), we combine the coupling between the dynamics constructed above with the choice of the coupling of the initial measures μ and ν as a function of the underlying Poisson process $(N_t)_{t\geqslant 0}$: the time horizon $t > 0$ being fixed, if $N_t = 0$, one chooses for $\mathcal{L}(X_0, Y_0)$ the optimal total variation coupling of ν and μ; if $N_t \geqslant 1$, one chooses their optimal Wasserstein coupling. One checks easily that this gives an admissible coupling, in the sense that its first (resp. second) marginal is a constant rate TCP process with initial distribution ν (resp. μ). And one gets with this construction, using the same estimates as above in the case where $N_t \geqslant 1$:

$$\mathbb{P}(X_t \neq Y_t) = \mathbb{P}(X_t \neq Y_t, N_t \geqslant 1) + \mathbb{P}(X_t \neq Y_t, N_t = 0)$$

$$\leqslant \mathbb{E}\left(|X_0 - Y_0|\frac{2^{-N_t+1}}{(t-T_{N_t-1})}\mathbb{1}_{\{N_t \geqslant 1\}}\right) + \mathbb{P}(X_0 \neq Y_0, N_t = 0)$$

$$= \lambda e^{-\lambda t/2}W_1(\nu, \mu) + e^{-\lambda t}\|\nu - \mu\|_{\mathrm{TV}},$$

which clearly implies (5.4). $\qquad\square$

5.2 A storage model example

In [30], Roberts and Tweedie improve the approach from [29] via Lyapunov functions and minorization conditions in the specific case of stochastically monotonous processes. They get better results on the speed of convergence to equilibrium in this case. They give the following example of a storage model as a good illustration of the efficiency of their method. The process $(X_t)_{t \geqslant 0}$ on $\mathbb{R}^+$ is driven by the generator

$$Lf(x) = -\beta x f'(x) + \alpha \int_0^\infty (f(x+y) - f(x)) e^{-y} \, dy.$$

In words, the current stock X_t decreases exponentially at rate β, and increases at random exponential times by a random (exponential) amount. Let us introduce a Poisson process $(N_t)_{t \geqslant 0}$ with intensity α and jump times $(T_i)_{i \geqslant 0}$ (with $T_0 = 0$) and a sequence $(E_i)_{i \geqslant 1}$ of independent random variables with law $\mathcal{E}(1)$ independent of $(N_t)_{t \geqslant 0}$. The process $(X_t)_{t \geqslant 0}$ starting from $x \geqslant 0$ can be constructed as follows: for any $i \geqslant 0$,

$$X_t = \begin{cases} e^{-\beta(t-T_i)} X_{T_i} & \text{if } T_i \leqslant t < T_{i+1}, \\ e^{-\beta(T_{i+1}-T_i)} X_{T_i} + E_{i+1} & \text{if } t = T_{i+1}. \end{cases}$$

Proposition 5.6. *For any $x, y \geqslant 0$ and $t \geqslant 0$,*

$$W_p(\delta_x P_t, \delta_y P_t) \leqslant |x - y| e^{-\beta t},$$

and (when $\alpha \neq \beta$)

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \leqslant e^{-\alpha t} + |x - y| \alpha \frac{e^{-\beta t} - e^{-\alpha t}}{\alpha - \beta}. \tag{5.5}$$

Moreover, if μ is the invariant measure of the process X, we have for any probability measure ν with a finite first moment and $t \geqslant 0$,

$$\|\nu P_t - \mu\|_{\mathrm{TV}} \leqslant \|\nu - \mu\|_{\mathrm{TV}} e^{-\alpha t} + W_1(\nu, \mu) \alpha \frac{e^{-\beta t} - e^{-\alpha t}}{\alpha - \beta}.$$

Remark 5.7. *In the case $\alpha = \beta$, the upper bound (5.5) becomes*

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \leqslant (1 + |x - y| \alpha t) e^{-\alpha t}.$$

Remark 5.8 (Optimality). *Applying L to the test function $f(x) = x^n$ allows us to compute recursively the moments of X_t. In particular,*

$$\mathbb{E}_x(X_t) = \frac{\alpha}{\beta} + \left(x - \frac{\alpha}{\beta} \right) e^{-\beta t}.$$

This relation ensures that the rate of convergence for the Wasserstein distance is sharp. Moreover, the coupling of total variation distance requires at least one jump. As a consequence, the exponential rate of convergence is greater than α. Thus, Equation (5.5) provides the optimal rate of convergence $\alpha \wedge \beta$.

Remark 5.9 (Comparison with previous work). *By way of comparison, the original method of [30] does not seem to give these optimal rates. The case $\alpha = 1$ and $\beta = 2$ is treated in this paper (as an illustration of Theorem 5.1), with explicit choices for the various parameters needed in this method. With these choices, in order to get the convergence rate, one first needs to compute the quantity θ (defined in Theorem 3.1), which turns out to be approximately 5.92. The result that applies is therefore the first part of Theorem 4.1 (Equation (27)), and the convergence rate is given by $\tilde{\beta}$ defined by Equation (22). The computation gives the approximate value 0.05, which is off by a factor 20 from the optimal value $\alpha \wedge \beta = 1$.*

Proof of Proposition 5.6. Firstly, consider two processes X and Y starting respectively at x and y and driven by the same randomness (*i.e.* Poisson process and jumps). Then the distance between X_t and Y_t is deterministic:

$$X_t - Y_t = (x - y)e^{-\beta t}.$$

Obviously, for any $p \geqslant 1$ and $t \geqslant 0$,

$$W_p(\delta_x P_t, \delta_y P_t) \leqslant |x - y|e^{-\beta t}.$$

Let us now construct explicitly a coupling at time t to get the upper bound (5.5) for the total variation distance. The jump times of $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$ are the ones of a Poisson process $(N_t)_{t\geqslant 0}$ with intensity α and jump times $(T_i)_{i\geqslant 0}$. Let us now construct the jump heights $(E_i^X)_{1\leqslant i\leqslant N_t}$ and $(E_i^Y)_{1\leqslant i\leqslant N_t}$ of X and Y until time t. If $N_t = 0$, no jump occurs. If $N_t \geqslant 1$, we choose $E_i^X = E_i^Y$ for $1 \leqslant i \leqslant N_t - 1$ and $E_{N_t}^X$ and $E_{N_t}^Y$ in order to maximise the probability

$$\mathbb{P}\big(X_{T_{N_t}} + E_{N_t}^X = Y_{T_{N_t}} + E_{N_t}^Y \big| X_{T_{N_t}}, Y_{T_{N_t}}\big).$$

This maximal probability of coupling is equal to

$$\exp\big(-|X_{T_{N_t}} - Y_{T_{N_t}}|\big) = \exp\big(-|x - y|e^{-\beta T_{N_t}}\big) \geqslant 1 - |x - y|e^{-\beta T_{N_t}}.$$

As a consequence, we get that

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \leqslant 1 - \mathbb{E}\big[\big(1 - |x - y|e^{-\beta T_{N_t}}\big)\mathbb{1}_{\{N_t \geqslant 1\}}\big]$$
$$\leqslant e^{-\alpha t} + |x - y|\mathbb{E}\big(e^{-\beta T_{N_t}}\mathbb{1}_{\{N_t \geqslant 1\}}\big).$$

The law of T_n conditionally on the event $\{N_t = n\}$ has the density

$$u \mapsto n\frac{u^{n-1}}{t^n}\mathbb{1}_{[0,t]}(u).$$

This ensures that

$$\mathbb{E}\big(e^{-\beta T_{N_t}}\mathbb{1}_{\{N_t \geqslant 1\}}\big) = \int_0^1 e^{-\beta t v}\mathbb{E}\big(N_t v^{N_t - 1}\big)\,dv.$$

Since the law of N_t is the Poisson distribution with parameter λt, one has

$$\mathbb{E}\big(N_t v^{N_t - 1}\big) = \alpha t e^{\alpha t(v-1)}.$$

This ensures that

$$\mathbb{E}\big(e^{-\beta N_t}\mathbb{1}_{\{N_t \geqslant 1\}}\big) = \alpha\frac{e^{-\beta t} - e^{-\alpha t}}{\alpha - \beta}$$

which completes the proof. Finally, to get the last estimate, we proceed as follows: if N_t is equal to 0, a coupling in total variation of the initial measures is done, otherwise, we use the coupling above (the method is exactly the same as for the equivalent result in Proposition 5.4, see its proof for details). $\qquad\square$

6 The case of diffusion processes

Let us consider the process $(X_t)_{t\geqslant 0}$ on $\mathbb{R}^d$ solution of

$$dX_t = A(X_t)\,dt + \sigma(X_t)\,dB_t, \tag{6.1}$$

where $(B_t)_{t\geqslant 0}$ is a standard Brownian motion on $\mathbb{R}^n$, σ is a smooth function from $\mathbb{R}^d$ to $\mathcal{M}_{d,n}(\mathbb{R})$ and A is a smooth function from $\mathbb{R}^d$ to $\mathbb{R}^d$. Let us denote by $(P_t)_{t\geqslant 0}$ the

semigroup associated to $(X_t)_{t \geqslant 0}$. If ν is a probability measure on $\mathbb{R}^d$, νP_t stands for the law of X_t when the law of X_0 is ν.

Under ergodicity assumptions, we are interested in getting quantitative rates of convergence of $\mathcal{L}(X_t)$ to its invariant measure in terms of classical distances (Wasserstein distances, total variation distance, relative entropy,...). Remark that if A is not in gradient form (even if σ is constant), X_t is not reversible and the invariant measure is usually unknown, so that it is quite difficult to use functional inequalities such as Poincaré or logarithmic Sobolev to get a quantitative rate of convergence in total variation or Wasserstein distance (using for example Pinsker's inequality or more generally transportation-information inequality). Therefore the only general tool seems to be Meyn-Tweedie's approach, via small sets and Lyapunov functions, as explained in Section 4.2. However, we have seen that in practical examples the resulting estimate can be quite poor.

The main goal of this short section is to recall the known results establishing the decay in Wasserstein distance and then to propose a strategy to derive control in total variation distance.

6.1 Decay in Wasserstein distance

The coupling approach to estimate the decay in Wasserstein distance was recently put forward, see [5] and [4] or [10]. It is robust enough to deal with nonlinear diffusions or hypoelliptic ones. In [3], the authors approach the problem directly, by differentiating the Wasserstein distance along the flow of the SDE.

Let us gather some of the results in these papers in the following statement.

Proposition 6.1. *1. Assume that there exists $\lambda > 0$ such that for $p > 1$*

$$\frac{p-1}{2}(\sigma(x)-\sigma(y))(\sigma(x)-\sigma(y))^t + (A(x)-A(y)) \cdot (x-y) \leqslant -\lambda|x-y|^2, \quad x,y \in \mathbb{R}^d. \tag{6.2}$$

Then, for any $\nu, \nu' \in \mathcal{P}_p(\mathbb{R}^d)$, one has

$$W_p(\nu P_t, \nu' P_t) \leqslant e^{-\lambda t} W_p(\nu, \nu'). \tag{6.3}$$

2. *Assume that for all x, $\sigma(x) = \alpha \, \mathrm{I}_d$ for some constant α and that A is equal to $-\nabla U$ where U is a $\mathcal{C}^2$ function such that $\mathrm{Hess}(U) \leqslant -K \, \mathrm{I}_d$ outside a ball $B(0, r_0)$ and $\mathrm{Hess}(U) \leqslant \rho \, \mathrm{I}_d$ inside this ball for some positive ρ. Then there exists $c > 1, \alpha > 0$ such that*

$$W_1(\nu P_t, \nu' P_t) \leqslant c \, e^{-\alpha t} W_1(\nu, \nu'). \tag{6.4}$$

3. *Suppose that the diffusion coefficient σ is constant. Assume that A is equal to $-\nabla U$ where U a $\mathcal{C}^2$ convex function such that $\mathrm{Hess}(U) \leqslant \rho \, \mathrm{I}_d$ outside a ball $B(0, r_0)$, with $\rho > 0$. Then there exists an invariant probability measure ν_∞ and $\alpha > 0$ such that*

$$W_2(\nu P_t, \nu_\infty) \leqslant e^{-\alpha t} W_2(\nu, \nu_\infty). \tag{6.5}$$

Proof. The first point is usually proved using a trivial coupling (and it readily extends to $p = 1$ in the case of constant diffusion coefficient), namely considering the same Brownian motion for two different solutions of the SDE starting with different initial measures. Note also that, in the case $p = 2$, the coercivity condition (6.2) is equivalent to the uniform contraction property (6.3) for the W_2 metric (see [32]).

The second point is due in this form to Eberle [10], using reflection coupling as presented by [19], used originally in this form by Chen and Wang to prove spectral gap estimates (see a nice short proof in [15, Prop. 2.8]) in the reversible case.

Finally, the third part was proved by [3] establishing the dissipation of the Wasserstein distance by an adequate functional inequality. $\qquad\square$

Remark 6.2.   • *Let us remark that the contraction property of the two first points ensures the existence of an invariant probability measure ν_∞ and an exponential rate of convergence towards this invariant measure in Wasserstein distance.*

• *Note also that in the hypoelliptic case of a kinetic Fokker-Planck equation:*

$$\begin{cases} dx_t = v_t dt \\ dv_t = dB_t - F(x_t)dt - V(v_t)dt \end{cases}$$

with $F(x) \sim ax$ and $V(v) \sim bv$ for positive a and b, one has

$$W_2(\nu P_t, \nu' P_t) \leq c\, e^{-\alpha t} W_2(\nu, \nu')$$

for $c > 1$ and a positive α (see [5]).

6.2   Total variation estimate

If $\Sigma = 0$ in Equation (6.1), the process $(X_t)_{t \geqslant 0}$ is deterministic and its invariant measure is a Dirac mass at the unique point $\bar{x} \in \mathbb{R}^d$ such that $A(\bar{x}) = 0$. As a consequence, for any $x \neq \bar{x}$,

$$\|\delta_x P_t - \delta_{\bar{x}}\|_{\mathrm{TV}} = 1 \quad \text{and} \quad H(\delta_x P_t | \delta_{\bar{x}}) = +\infty.$$

A non-zero variance is needed to get a convergence estimate in total variation distance. Classically, the Brownian motion creates regularity and density. There are a lot of results giving regularity, in terms of initial points, of semigroup in small time. Let us quote the following result of Wang, which holds for processes living on a manifold.

Lemma 6.3 ([33])**.** *Suppose that σ is constant and denote by η the infimum of its spectrum. If A is a $\mathcal{C}^2$ function such that*

$$\frac{1}{2}(\operatorname{Jac} A + \operatorname{Jac} A^T) \geqslant K\, \mathrm{I}_d$$

then there exists K_η such that, for small $\varepsilon > 0$,

$$\|\delta_x P_\varepsilon - \delta_y P_\varepsilon\|_{\mathrm{TV}} \leqslant K_\eta \frac{|x-y|}{\sqrt{\varepsilon}}.$$

Remark 6.4.

- *There are many proofs leading to this kind of results, see for example Aronson [1] for pioneering works, and [33] using Harnack's and Pinsker's inequalities.*
- *Note that in [13], an equivalent bound was given for the kinetic Fokker-Planck equation but with ε replaced by ε^3.*

Now that we have a decay in Wasserstein distance and a control on the total variation distance after a small time, we can use the same idea as for the TCP process. As a consequence, we get the following result.

Theorem 6.5. *Assume that σ is constant. Under Points 1. or 2. of Proposition 6.1, one has, for any ν and $\tilde{\nu}$ in $\mathcal{P}_1(\mathbb{R}^d)$,*

$$\|\nu P_t - \tilde{\nu} P_t\|_{\mathrm{TV}} \leqslant \frac{Ke^{\lambda \varepsilon}}{\sqrt{\varepsilon}} W_1(\nu, \tilde{\nu}) e^{-\lambda t}.$$

Under Point 3. of Proposition 6.1, one has, for any ν and $\tilde{\nu}$ in $\mathcal{P}_2(\mathbb{R}^d)$,

$$\|\nu P_t - \tilde{\nu} P_t\|_{\mathrm{TV}} \leqslant \frac{Ke^{\lambda \varepsilon}}{\sqrt{\varepsilon}} W_2(\nu, \tilde{\nu}) e^{-\lambda t}.$$

Proof. Using first Lemma 6.3 and then Point 1. of Proposition 6.1, we get

$$
\begin{aligned}
\|\nu P_t - \tilde{\nu} P_t\|_{\mathrm{TV}} &= \|\nu P_{t-\varepsilon} P_\varepsilon - \tilde{\nu} P_{t-\varepsilon} P_\varepsilon\|_{\mathrm{TV}} \\
&\leq \frac{K}{\sqrt{\varepsilon}} W_1(\nu P_{t-\varepsilon}, \tilde{\nu} P_{t-\varepsilon}) \\
&\leq \frac{K e^{\lambda \varepsilon}}{\sqrt{\varepsilon}} W_1(\nu, \tilde{\nu}) e^{-\lambda t}.
\end{aligned}
$$

The proof of the second assertion is similar, except the use of Point 3. of Proposition 6.1 in the second step. □

Remark 6.6. *Once again, one can give in the case of kinetic Fokker-Planck equation estimate in total variation distance, using the previous remarks (see [2] for qualitative results).*

References

[1] D. G. Aronson. Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.*, 73:890–896, 1967. MR-0217444

[2] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008. MR-2381160

[3] F. Bolley, I. Gentil, and A. Guillin. Convergence to equilibrium in Wasserstein distance for Fokker–Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012. MR-2964689

[4] F. Bolley, A. Guillin, and F. Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *M2AN Math. Model. Numer. Anal.*, 44(5):867–884, 2010. MR-2731396

[5] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields*, 140(1-2):19–40, 2008. MR-2357669

[6] D. Chafaï, F. Malrieu, and K. Paroux. On the long time behavior of the TCP window size process. *Stochastic Process. Appl.*, 120(8):1518–1534, 2010. MR-2653264

[7] B. Cloez. Wasserstein decay of one dimensional jump-diffusions. preprint arXiv:1202.1259v1 available at arXiv:1202.1259, 2012.

[8] O. L. V. Costa and F. Dufour. Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, 47(2):1053–1077, 2008. MR-2385873

[9] V. Dumas, F. Guillemin, and Ph. Robert. A Markovian analysis of additive-increase multiplicative-decrease algorithms. *Adv. in Appl. Probab.*, 34(1):85–111, 2002. MR-1895332

[10] A. Eberle. Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris*, 349(19-20):1101–1104, 2011. MR-2843007

[11] I. Grigorescu and M. Kang. Recurrence and ergodicity for a continuous AIMD model. Preprint available at `http://www.math.miami.edu/~igrigore/pp/research.html`, 2009.

[12] F. Guillemin, Ph. Robert, and B. Zwart. AIMD algorithms and exponential functionals. *Ann. Appl. Probab.*, 14(1):90–117, 2004. MR-2023017

[13] A. Guillin and F.-Y. Wang. Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality. *J. Differential Equations*, 253(1):20–40, 2012. MR-2917400

[14] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011. MR-2857021

[15] M. Hairer, A. M. Stuart, and S. Vollmer. Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. preprint available on arXiv:1112.1392, 2011.

[16] D. Lamberton and G. Pagès. A penalized bandit algorithm. *Electron. J. Probab.*, 13:no. 13, 341–373, 2008. MR-2386736

[17] P. Laurençot and B. Perthame. Exponential decay for the growth-fragmentation/cell-division equation. *Commun. Math. Sci.*, 7(2):503–510, 2009. MR-2536450

[18] T. Lindvall. *Lectures on the coupling method.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication. MR-1180522

[19] T. Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986. MR-0841588

[20] A. H. Löpker and Z. Palmowski. On Time Reversal of Piecewise Deterministic Markov Processes. preprint available at arXiv:1110.3813, 2011.

[21] A. H. Löpker and J. S. H. van Leeuwaarden. Transient moments of the TCP window size process. *J. Appl. Probab.*, 45(1):163–175, 2008. MR-2409318

[22] K. Maulik and B. Zwart. Tail asymptotics for exponential functionals of Lévy processes. *Stochastic Process. Appl.*, 116(2):156–177, 2006. MR-2197972

[23] K. Maulik and B. Zwart. An extension of the square root law of TCP. *Ann. Oper. Res.*, 170:217–232, 2009. MR-2506283

[24] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993. MR-1287609

[25] T. J. Ott and J. H. B. Kemperman. Transient behavior of processes in the TCP paradigm. *Probab. Engrg. Inform. Sci.*, 22(3):431–471, 2008. MR-2426601

[26] 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.

[27] B. Perthame and L. Ryzhik. Exponential decay for the fragmentation or cell-division equation. *J. Differential Equations*, 210(1):155–177, 2005. MR-2114128

[28] S. T. Rachev. *Probability metrics and the stability of stochastic models.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1991. MR-1105086

[29] G. O. Roberts and J. S. Rosenthal. Quantitative bounds for convergence rates of continuous time Markov processes. *Electron. J. Probab.*, 1:no. 9, approx. 21 pp. (electronic), 1996. MR-1423462

[30] G. O. Roberts and R. L. Tweedie. Rates of convergence of stochastically monotone and continuous time Markov models. *J. Appl. Probab.*, 37(2):359–373, 2000. MR-1780996

[31] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003. MR-1964483

[32] M.-K. von Renesse and K.-T. Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005. MR-2142879

[33] F.-Y. Wang. Harnack inequalities on manifolds with boundary and applications. *J. Math. Pures Appl. (9)*, 94(3):304–321, 2010. MR-2679029