Rates of convergence in total variation norm for Markov processes

Joint work with
A. Christen, A. Guillin, F. Malrieu and P.-A. Zitt

Jean-Baptiste Bardet

LMRS, Université de Rouen

May 15, 2013, Rennes

Setup and notations

We consider a Markov process $(X_t)_{t\geq 0}$, with state space E, assumed to be ergodic with unique invariant probability measure μ.

Let L be its generator, and $(P_t)_{t\geq 0}$ the associated semigroup, acting on

- functions:

$$P_t f(x) = \mathbf{E}(f(X_t) \,|\, X_0 = x) = \mathbf{E}_x(f(X_t)) \,;$$

- measures:

$$\nu P_t = \mathcal{L}(X_t \,|\, \mathcal{L}(X_0) = \nu) \,.$$

Our aim

Get quantitative rates of convergence to equilibrium for X_t, quantify the convergence

$$\nu P_t \xrightarrow[t \to +\infty]{} \mu \, .$$

One needs to choose a distance between probability measures.

Wasserstein distances

Definition

If (E,d) is a metric space, for $p \geq 1$, the Wasserstein distance of order p between two probability measures $\nu, \tilde{\nu} \in \mathcal{P}(E)$ is

$$\begin{aligned} W_p(\nu, \tilde{\nu}) &= \inf \left\{ \left(\int_E d(x,y)^p \, d\pi(x,y) \right)^{1/p} \, : \, \pi \in \Pi(\nu, \tilde{\nu}) \right\} \\ &= \inf \left\{ \mathbf{E}(|X - Y|^p)^{1/p} \, : \, X \sim \nu, Y \sim \tilde{\nu} \right\} , \end{aligned}$$

where $\Pi(\nu, \tilde{\nu})$ is the set of couplings of ν and $\tilde{\nu}$, *i.e.* of probability measures on E^2 with first (resp. second) marginal equal to ν (resp. to $\tilde{\nu}$).

Wasserstein distances

Interest

Among the many interesting features of Wasserstein distances

- "almost" metrizes the weak convergence (and more: Kantorovitch duality theorem);
- "easy" to bound from above: any coupling gives an upper bound (one has to find an efficient one).

See C. Villani, *Optimal transport. Old and new* for (much) more information...

More specifically, for Markov processes, the typical coupling approach is to write

$$W_p(\nu P_t, \tilde{\nu} P_t)^p \leq \mathbf{E}(d(X_t, Y_t)^p) ,$$

where (X_t, Y_t) is a "trajectorial coupling" of the dynamics with initial measures ν and $\tilde{\nu}$:

- $(X_t, Y_t)_{t \geq 0}$ is a Markov process on E^2;
- both marginals are Markov processes with generator L;
- $X_0 \sim \nu$ and $Y_0 \sim \tilde{\nu}$.

Total variation distance

Definition

The total variation distance between two probability measures $\nu, \tilde{\nu} \in \mathcal{P}(E)$ is

$$\|\nu - \tilde{\nu}\|_{\mathsf{TV}} = \inf\left\{\left(\int_E \mathbf{1}_{x \neq y}\, d\pi(x,y)\right) \,:\, \pi \in \Pi(\nu, \tilde{\nu})\right\}$$
$$= \inf\left\{\mathbf{P}(X \neq Y) \,:\, X \sim \nu, Y \sim \tilde{\nu}\right\} ,$$

- Total variation is (essentially) stronger than Wasserstein distances.
- Also defined via an infimum on couplings.

Total variation distance

Coupling method for Markov processes

As before, for Markov processes,

$$\|\nu P_t - \tilde{\nu} P_t\|_{\mathrm{TV}} \leq \mathbf{P}(X_t \neq Y_t) \leq \mathbf{P}(T \geq t)\,,$$

for any trajectorial coupling (X_t, Y_t), where
$T = \inf\{t \geq 0 \,:\, X_t = Y_t\}$ is the coupling (coalescing) time.

General idea

Construct an efficient coalescent coupling to derive quantitative estimates for the rate of convergence in total variation is difficult.

The approach “à la” Meyn & Tweedie hardly gives reasonable quantitative rates of convergence. One needs to develop alternative approaches.

Ours is the following:

1. First apply for a given time a Wasserstein coupling, *i.e.* a coupling which is adapted to the control of convergence in Wasserstein distance.
2. Once the two processes are “close enough”, find an adapted coalescent coupling.

The TCP process

Transmission Control Protocol window size process

Markov process on $E = \mathbf{R}_+$ with generator

$$Lf(x) = f'(x) + x(f(x/2) - f(x)) \,.$$

Modelization of the rate of transmission on the internet, to avoid congestions.

- non reversible;
- not stochastically monotone;
- not much randomness: all randomness comes from the jump times;
- not too complicated...

The constant rate TCP process

Definition

Markov process on $E = \mathbf{R}_+$ with generator

$$Lf(x) = f'(x) + \lambda(f(x/2) - f(x)),$$

λ being a positive constant.

A simplified version of the TCP process.

Constant jump rate: the jump times $(T_n)_{n\geq 1}$ form a Poisson process of intensity λ.

The constant rate TCP process

Wasserstein distance

The best choice is to take the same jump times for both processes.

After n jumps, one has

$$|X_{T_n} - Y_{T_n}| = \frac{|X_0 - Y_0|}{2^n}, \text{ hence } |X_t - Y_t| = \frac{|X_0 - Y_0|}{2^{N_t}}.$$

By a simple computation this gives

$$W_p(\nu P_t, \tilde{\nu} P_t) \leq e^{-\lambda_p t} W_p(\nu, \tilde{\nu}),$$

with $\lambda_p = \lambda(1 - 2^{-p})/p$ (optimal rate).

The constant rate TCP process

Total variation norm

A simple geometric idea to coalesce in one jump:
if $X_0 = x > Y_0 = y$ and $T_1^X = T_1^Y + x - y < T_2^Y$, then $X_{T_1^X} = Y_{T_1^X}$.
We will combine this with the simple Wasserstein coupling used before.

The constant rate TCP process

Total variation norm

The complete coalescent coupling starting from x and y before time t:

- sample a Poisson process (T_k) of parameter λ on $[0, t]$
- both X and Y will make exactly $n = N_t$ jumps before time t
- for $1 \leq k \leq n - 1$, $T_k^X = T_k^Y = T_k$; hence

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

- the last jump of each process is chosen to maximize the probability that $T_n^X = T_n^Y + \frac{x-y}{2^{n-1}}$: optimal coupling of two continuous random variables

The constant rate TCP process

Total variation norm

Conditionally on
$\{N_t^X = N_t^Y = n; T_k^X = T_k^Y = T_k \; \forall \, 1 \le k \le n-1\}$,

$$\mathbf{P}\left(T_n^X = T_n^Y + \frac{x-y}{2^{n-1}}\right) \ge 1 - \frac{x-y}{2^{n-1}(t - T_{n-1})}$$

Gives

$$\begin{aligned}\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} &\le 1 - \mathbf{E}\left[\left(1 - \frac{x-y}{2^{N_t-1}(t - T_{N_t-1})}\right)\mathbf{1}_{\{N_t \ge 1\}}\right] \\ &= e^{-\lambda t} + \lambda e^{-\lambda t/2}|x-y|\end{aligned}$$

New proof of a result by Perthame & Ryzhik (2005).

The TCP process

Wasserstein coupling

Defined on $\mathbf{R}_+^2$ by

$$\begin{aligned}\mathcal{L}f(x,y) = \partial_x f(x,y) + \partial_y f(x,y) + y(f(x/2,y/2) - f(x,y)) \\ + (x-y)(f(x/2,y) - f(x,y))\end{aligned}$$

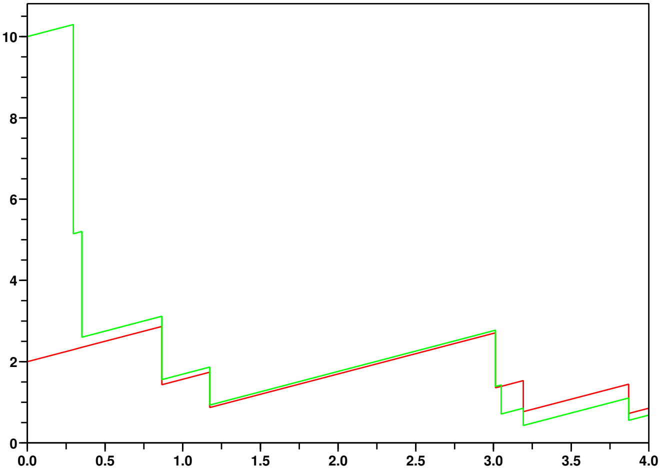
when $x > y$ and symmetric expression when $y < x$.

Introduced by Chafai, Malrieu & Paroux (2010).

The only trajectorial coupling of TCP process where the smallest process never jumps alone.

The TCP process

Wasserstein coupling



The TCP process

Polynomial bound for W_1

For $V_p(x,y) = |x-y|^p$, one would like to get $\mathcal{L} V_p \leq -\lambda_p V_p$ hence $\mathbf{E}(|X_t - Y_t|^p) \leq e^{-\lambda_p t}\mathbf{E}(|X_0 - Y_0|^p)$.
By direct computation

$$\mathcal{L} V_1 = -\frac{3}{2} V_1^2$$

which only gives a polynomial rate of convergence (CMP, 2010).

It is even worse for $p > 1$.

The TCP process

Rate of convergence in Wasserstein distances

Things work better when $p < 1$.
With $p = 1/2$:

$$\mathcal{L}V_{1/2}(x,y) \leq -(x \vee y)(1-M)V_{1/2}(x,y)$$

with $M = \sqrt{2}(3+\sqrt{3})/8 \sim 0.84$.

Proposition

For any $x, y > 0$,

$$\mathbf{E}\left(|X_t - Y_t|^{1/2}\right) \leq \frac{1}{\sqrt{M}} e^{-\lambda t}|x-y|^{1/2},$$

with $\lambda = \sqrt{2}(1-\sqrt{M}) \sim 0.12$.

The TCP process

Rate of convergence in Wasserstein distances

Theorem

For any $\theta \in (0,1)$, $p \geq 1$ and $t_0 > 0$, there exists a constant C such that for any $t \geq t_0$ and for all initial probability measures ν and $\tilde{\nu}$:

$$W_p(\nu P_t, \tilde{\nu} P_t) \leq C e^{-(\theta\lambda/p)t} W_{\theta/2}(\nu, \tilde{\nu})^{2/(p\theta)}.$$

For W_1:

- rate obtained by our method: $\lambda \sim 0.12$
- rate for the Wasserstein coupling by simulation: ~ 0.5
- rate for the true W_1 distance by simulation: ~ 1.6

The TCP process

Rate of convergence in total variation

There exists a constant $K > 0$ such that for any $t, x, y \leq 0$

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

Theorem

For any $\theta \in (0,1)$ and $t_0 > 0$, there exists a constant C such that for any $t \geq t_0$ and for all initial probability measures ν and $\tilde{\nu}$:

$$\|\nu P_t - \tilde{\nu} P_t\|_{\mathrm{TV}} \leq C e^{-\theta(2\lambda/3)t}.$$

The best we could obtain via Meyn-Tweedie method: $\sim 10^{-14}$...