Wasserstein convergence of penalized Markov processes

Nicolas Champagnat, Édouard Strickler, Denis Villemonais

Conference "A random walk in the land of stochastic analysis and numerical probability", CIRM, 5 Septembre 2023

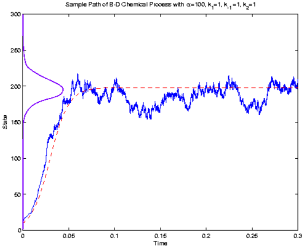
Processes with absorption

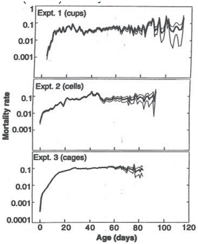
Population dynamics with extinction:

- Stationary behavior $\rightsquigarrow$ usually extinction
- Observed populations often exhibit a stationary behavior
- Quasi-stationary distributions (QSD) are stationary distributions conditionally on non-extinction (Seneta and Vere-Jones, 1966)

Other applications:

- Metastable dynamics (Bovier, den Hollander 2015)
- Molecular simulation: parallel replica algorithm (Le Bris, Lelièvre, Luskin, Perez, 2012)
- Mortality plateau (Carey et al., 1992)





Bibliography

Large time convergence in total variation of absorbed Markov processes conditioned to non-extinction:
Birkhoff (1957); van Doorn (1991); Del Moral (2004); C. and Villemonais (2016, 2023); Bansaye, Cloez and Gabriel (2020); Guillin, Nectoux and Wu (2020); Ferré, Rousset and Stoltz (2021); Benaïm, C., Oçafrain and Villemonais (2022), Bansaye, Cloez, Garbiel and Marguet (2022)...

In Wasserstein distance:
Villemonais (2020); Oçafrain (2020, 2021); Del Moral and Horton (2021); Journel and Monmarché (2022).

Killed Markov process (discrete time)

$(Y_n)_{n\in\mathbb{N}}$ a Markov chain with values in a measurable space $(E, \mathcal{E})$.
$\partial \notin E$ a cemetery point, $p : E \mapsto [0, 1]$ a survival function

We define the killed chain X evolving in $E \cup \{\partial\}$ as:

- If $X_n \in E$, then

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{with probability} \quad p(X_n) \\ \partial & \text{with probability} \quad 1 - p(X_n) \end{cases}$$

- if $X_n = \partial$, then $X_{n+1} = \partial$

Under good assumptions on p;

$$\tau_\partial := \inf\{n \geq 0 : \ X_n = \partial\} < +\infty$$

Killed Markov process (continuous time)

$(Y_t)_{t\geq 0}$ a Markov process with values in a measurable space $(E, \mathcal{E})$.
$\partial \notin E$ a cemetery point, $\rho : E \mapsto \mathbb{R}_+$ a killing rate

We define the killed process X evolving in $E \cup \{\partial\}$:

$$X_t = \begin{cases} Y_t & \text{if} \quad t < \tau_\partial \\ \partial & \text{otherwise} \end{cases}$$

where

$$\tau_\partial := \inf \left\{ t \geq 0 \; : \int_0^t \rho(Y_s) ds > \theta \right\}$$

with θ a random variable with exponential law of parameter 1, independent of $(Y_t)_{t\geq 0}$.

Quasi-stationary distributions (QSD)

The only stationary state of X is the cemetery point ∂.
How to describe X before extinction?

Definition

A probability measure α on E is a *Quasi-Stationary Distribution* if

$$X_0 \sim \alpha \quad \Rightarrow \quad \mathbb{P}(X_t \in \cdot \mid t < \tau_\partial) = \alpha, \quad \forall t$$

Note: if $X_0 \sim \mu$, we write

$$\mathbb{P}_\mu(X_t \in \cdot \mid t < \tau_\partial) = \frac{\mathbb{P}_\mu(X_t \in \cdot,\ t < \tau_\partial)}{\mathbb{P}_\mu(t < \tau_\partial)} \neq \int \mathbb{P}_x(X_t \in \cdot \mid t < \tau_\partial) d\mu(x)$$

Questions : existence, uniqueness, convergence of $\mathbb{P}_\mu(X_t \in \cdot | t < T_\partial)$ towards α ?

Convergence in total variation

Theorem (C. and Villemonais, 2016)

In the general framework of absorbed process, the conditions

1. (conditional Doeblin) there exist $t_1, c_1 > 0$ and a distribution ν such that for all $x \in E$,
$$\mathbb{P}_x\left(X_{t_1} \in \cdot \mid t_1 < \tau_\partial\right) \geq c_1 \nu(\cdot)$$
2. (Harnack inequality) there exists $c_2 > 0$ such that for all $x \in E$ and $t \geq 0$,
$$\mathbb{P}_\nu(t < \tau_\partial) \geq c_2 \mathbb{P}_x(t < \tau_\partial)$$

are equivalent to the existence of $C, \gamma > 0$ and a unique QSD α such that for all initial distribution μ

$$\|\mathbb{P}_\mu(X_t \in \cdot \mid t < \tau_\partial) - \alpha\|_{TV} \leq C e^{-\gamma t}$$

What about Bernoulli convolutions?

We consider the chain Y_n on $E = [-2, 2]$ defined by

$$Y_{n+1} = \frac{1}{2} Y_n + \theta_{n+1}$$

where $(\theta_n)_{n\geq 0}$ is a sequence of i.i.d. variables such that

$$\mathbb{P}(\theta_n = 1) = \mathbb{P}(\theta_n = -1) = \frac{1}{2}.$$

Let X_n be the killed chain. Then;

$$\mathbb{P}_x(X_n \in \mathbb{Q} \mid n < \tau_\partial) = \mathbb{1}_{\mathbb{Q}}(x)$$

$\rightsquigarrow$ convergence in total variation is not possible

However...

For $x \in [-2, 2]$,

$$\begin{cases} Y^x_{n+1} = \frac{1}{2} Y^x_n + \theta_{n+1} \\ Y^x_0 = x \end{cases}$$

For all x and $y \in [-2, 2]$,

$$|Y^x_n - Y^y_n| = 2^{-n}|x - y|$$

$\rightsquigarrow$ gives hope for convergence towards a QSD

Goal

Definition

(E, d) a Polish space. The Wasserstein distance between two distributions μ and ν on E is

$$\mathcal{W}_d(\mu,\nu) = \inf_{U\sim\mu, V\sim\nu} \mathbb{E}(d(U,V)) = \sup_{f\in \mathrm{Lip}_1(d)} \left| \int f\, d\mu - \int f\, d\nu \right|.$$

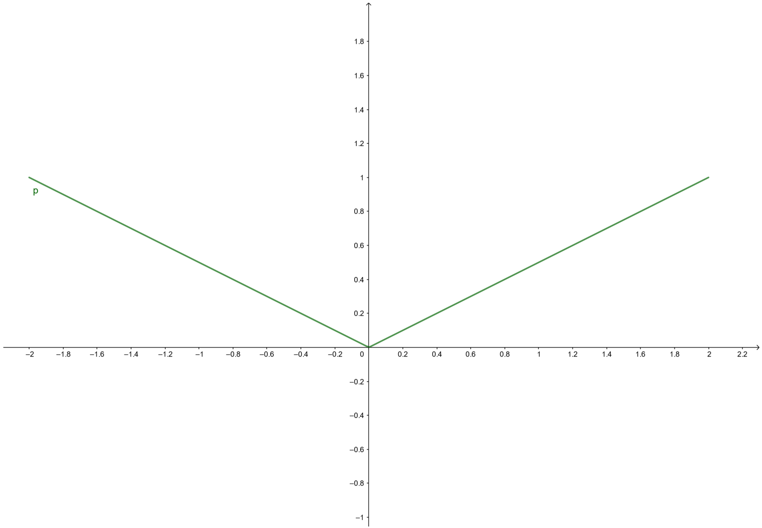
$\rightsquigarrow$ super easy:

$$\mathcal{W}_d\left(\mathbb{P}_\mu(Y_n \in \cdot), \mathbb{P}_\nu(Y_n \in \cdot)\right) \leq 2^{-n}\mathcal{W}_d(\mu,\nu).$$

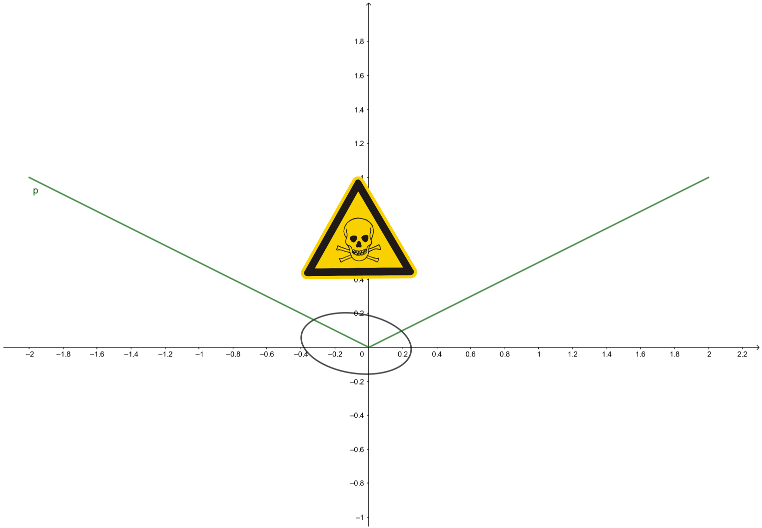
$\rightsquigarrow$ less easy:

$$\mathcal{W}_d\left(\mathbb{P}_\mu(X_n \in \cdot | n < T_\partial), \mathbb{P}_\nu(X_n \in \cdot | n < T_\partial)\right) \leq ??$$

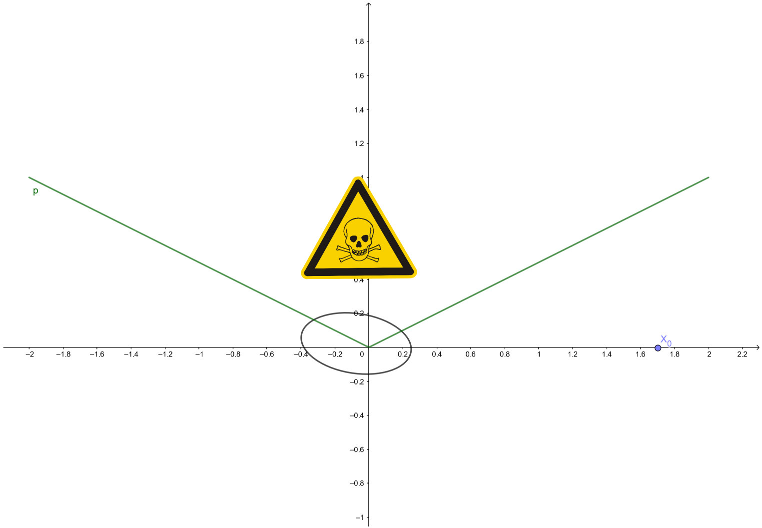
Death Valley $Y_{n+1} = Y_n/2 + \theta_{n+1}$ and $p(x) = |x|/2$



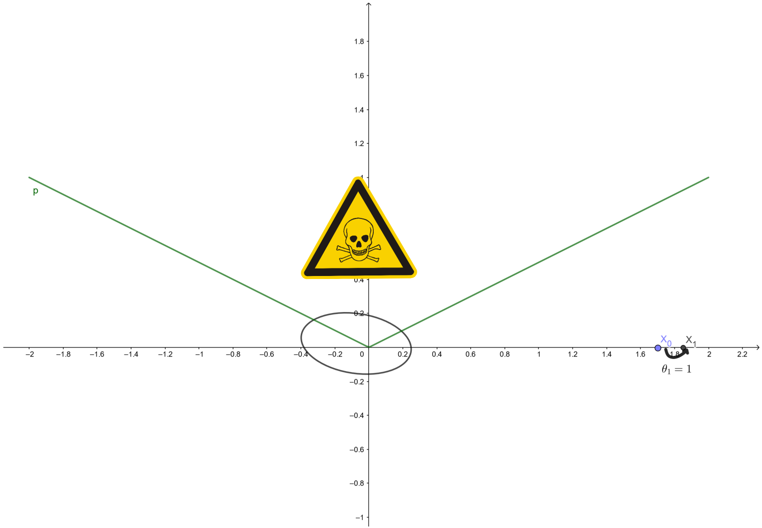
Death Valley $Y_{n+1} = Y_n/2 + \theta_{n+1}$ and $p(x) = |x|/2$



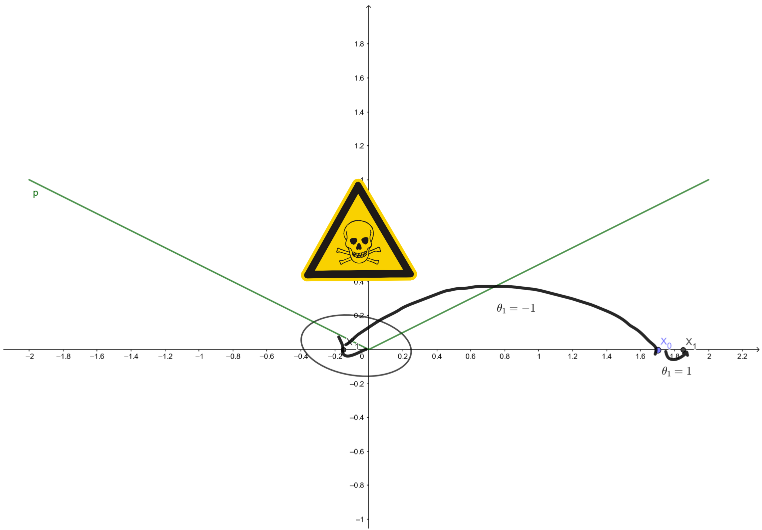
Death Valley $Y_{n+1} = Y_n/2 + \theta_{n+1}$ and $p(x) = |x|/2$



Death Valley $Y_{n+1} = Y_n/2 + \theta_{n+1}$ and $p(x) = |x|/2$



Death Valley $Y_{n+1} = Y_n/2 + \theta_{n+1}$ and $p(x) = |x|/2$



Death Valley: impassable

$$Y_{n+1} = \frac{1}{2} Y_n + \theta_{n+1}$$

Let X_n the chain Y_n killed, survival probability: $p(x) = |x|/2$.

We can show that for all $x, y \in [-2, 2] \setminus \{0\}$ and $n \geq 1$,

$$\mathcal{W}_d \left(\mathbb{P}_x(X_n \in \cdot \mid n < \tau_\partial), \mathbb{P}_y(X_n \in \cdot \mid n < \tau_\partial) \right) \geq \frac{1}{2}|x - y|$$

$\rightsquigarrow$ no contraction for the conditioned chain!

Recall the settings

- $I = \mathbb{N}$ or $\mathbb{R}_+$
- $(Y_t)_{t\in I}$ a Markov process on (E, d)
- $\rho : E \to \mathbb{R}_+$ measurable
- If $I = \mathbb{R}_+$, ρ is the killing rate
- if $I = \mathbb{N}$, $p = e^{-\rho}$ is the survival function

It is more convenient to work in the setting of penalized Markov processes (Del Moral, 2004):

$$\mathbb{P}_x(X_t \in A \mid t < \tau_\partial) = \frac{\mathbb{E}_x[Z_t \mathbb{1}_{Y_t\in A}]}{\mathbb{E}_x[Z_t]},$$

where

$$Z_t = Z_t(Y) = p(Y_0)p(Y_1)\dots p(Y_{t-1}) = \exp\left(-\sum_{i=0}^{n-1} \rho(Y_i)\right) \text{ for } I = \mathbb{N},$$

or

$$Z_t = Z_t(Y) = \exp\left(-\int_0^t \rho(X_s)\,ds\right) \text{ for } I = \mathbb{R}_+.$$

Assumption (A)

1. d is bounded and $\rho : E \to \mathbb{R}_+$ is Lipschitz,
2. There exists $C_A, \gamma > 0$ such that, for all $t \in I$ and $x, y \in E$, there exists a Markov coupling of Y^x and Y^y such that

$$\frac{\mathbb{E}[Z_t(Y^x) d(Y_t^x, Y_t^y)]}{\mathbb{E}_x Z_t} = \mathbb{E}\left(d(Y_t^x, Y_t^y) \,|\, t < T_\partial^x\right) \le C_A e^{-\gamma t} d(x,y)$$

Theorem (with E. Strickler and D. Villemonais, 2023)

Under Assumption (A), there exist C_0 and $\beta > 0$ such that

$$\mathcal{W}_d\left(\mathbb{P}_\mu(X_t \in \cdot \mid t < \tau_\partial), \mathbb{P}_\nu(X_t \in \cdot \mid t < \tau_\partial)\right) \le C_0 e^{-\beta t} \mathcal{W}_d(\mu, \nu)$$

Moreover, there exists a unique QSD α and

$$\mathcal{W}_d\left(\mathbb{P}_\mu(X_t \in \cdot \mid t < \tau_\partial), \alpha\right) \le C_0 e^{-\beta t} \mathcal{W}_d(\mu, \alpha)$$

Note: in the previous example, $\rho(x) = -\log(|x|/2)$ is not Lipschitz

Ideas of the proof

Define for all $s \leq t \leq T$

$$R^T_{s,t} f(x) = R^{T-s}_{0,t-s} f(x) = \frac{\mathbb{E}_x Z_{T-s} f(Y_{t-s})}{\mathbb{E}_x Z_{T-s}}.$$

This is a time-inhomogeneous semigroup: for all $r \leq s \leq t \leq T$

$$R^T_{r,s} R^T_{s,t} = R^T_{r,t}.$$

Note that

$$\mathbb{E}_x[f(X_t) \mid t < \tau_\partial] = \delta_x R^t_{0,t} f$$

but

$$\mathbb{E}_\mu[f(X_t) \mid t < \tau_\partial] \neq \mu R^t_{0,t} f$$

Ideas of the proof

Key property: (A) implies that

(H) There exist $C_H > 0$ such that, for all $x, y \in E$ and $t \in I$,
$\mathbb{E}_x(Z_t) \leq C_H \mathbb{E}_y(Z_t)$.

First consequence: for all $x, y \in E$ and $t \leq T$,

$$\frac{\mathbb{E}\left[Z_T(Y^x) d(Y_t^x, Y_t^y)\right]}{\mathbb{E}_x Z_T} = \frac{\mathbb{E}\left[Z_t(Y^x) d(Y_t^x, Y_t^y) \mathbb{E}_{Y_t^x} Z_{T-t}\right]}{\mathbb{E}_x\left(Z_t \mathbb{E}_{Y_t} Z_{T-t}\right)}$$
$$\leq C_H \frac{\mathbb{E}\left[Z_t(Y^x) d(Y_t^x, Y_t^y)\right]}{\mathbb{E}_x Z^t}$$
$$\leq C_A C_H e^{-\gamma t} d(x, y).$$

Ideas of the proof

Second consequence:

$$\begin{aligned}
\mathbb{E}\,|Z_t(Y^x) - Z_t(Y^y)| &\leq \mathbb{E}\left[\left(\exp\left(\int_0^t \rho(Y_s^x)\,ds\right) + \exp\left(\int_0^t \rho(Y_s^y)\,ds\right)\right)\right. \\
&\qquad\qquad \left.\times \left|\int_0^t (\rho(Y_s^x) - \rho(Y_s^y))\,ds\right|\right] \\
&\leq \|\rho\|_{\mathrm{Lip}} \int_0^t \mathbb{E}\left[(Z_t(Y^x) + Z_t(Y^y))\, d(Y_s^x, Y_s^y)\right] ds \\
&\leq \|\rho\|_{\mathrm{Lip}(d)}\, C_A\, C_H \int_0^t e^{-\gamma s}\,ds\, (\mathbb{E}_x Z_t + \mathbb{E}_y Z_t)\ d(x,y) \\
&\leq \|\rho\|_{\mathrm{Lip}(d)}\, \frac{C_A\, C_H}{\gamma}(1 + C_H)\mathbb{E}_y Z_t\, d(x,y),
\end{aligned}$$

Ideas of the proof

For all Lipschitz ϕ,

$$\left|\delta_x R^T_{s,t}\phi - \delta_y R^T_{s,t}\phi\right|$$
$$\leq \frac{|\mathbb{E} Z_{T-s}(Y^x)\phi(Y^x_{t-s}) - \mathbb{E} Z_{T-s}(Y^x)\phi(Y^y_{t-s})|}{\mathbb{E}_x Z_{T-s}} + \|\phi\|_\infty \frac{|\mathbb{E}_x Z_{T-s} - \mathbb{E}_y Z_{T-s}|}{\mathbb{E}_x Z_{T-s}}$$
$$\leq \|\phi\|_{\mathrm{Lip}(d)} \frac{\mathbb{E}\left[Z_{T-s}(Y^x) d(Y^x_{t-s}, Y^y_{t-s})\right]}{\mathbb{E}_x Z_{T-s}} + 2\|\phi\|_\infty \frac{|\mathbb{E}_x Z_{T-s} - \mathbb{E}_y Z_{T-s}|}{\mathbb{E}_x Z_{T-s}}$$
$$\leq C\left(\|\phi\|_\infty + e^{-\gamma(t-s)}\|\phi\|_{\mathrm{Lip}(d)}\right) d(x,y).$$

Introducing the equivalent distance on E

$$d_\kappa(x,y) = (\kappa d(x,y)) \wedge 1,$$

with well-chosen κ and t_0, we deduce that there exists $\beta < 1$ such that, for all $\phi \in \mathrm{Lip}_1(d_\kappa)$ and all k,

$$\left|\delta_x R^T_{kt_0,(k+1)t_0}\phi - \delta_y R^T_{kt_0,(k+1)t_0}\phi\right| \leq \beta d_\kappa(x,y).$$

Iterating with the semigroup property allows to conclude.

Consequence on the survival probability

Corollary

Assume (A). Then, there exists a d-Lipschitz function $\eta : E \to (0, +\infty)$ such that

$$\eta(x) = \lim_{t \to +\infty} e^{\lambda_0 t} \mathbb{E}_x Z_t,$$

where λ_0 is the absorption rate of the QSD α and the convergence holds exponentially fast for the uniform norm.

Consequence on the Q-process

Corollary

Assume (A). There exists a Markov family $(\mathbb{Q}_x)_{x\in E}$ such that

$$\left|\frac{\mathbb{E}_x(\mathbb{1}_A Z_t)}{\mathbb{E}_x Z_t} - \mathbb{Q}_x(A)\right| \leq Ce^{-\beta(t-s)}$$

for all $\mathcal{F}_s$-measurable set A for all $s \leq t \in I$ and the probability measure

$$\nu_Q(dx) = \eta(x)\alpha(dx)$$

is the unique invariant distribution of $(Y_t)_{t\in I}$ under $(\mathbb{Q}_x)_{x\in E}$. In addition, for any initial distributions μ and ν on E,

$$\mathcal{W}_d(\mathbb{Q}_\mu(Y_t \in \cdot), \mathbb{Q}_\nu(Y_t \in \cdot)) \leq Ce^{-\beta s}\mathcal{W}_d(\mu,\nu).$$

Consequence on quasi-ergodicity

For all $x \in E$ and $t > 0$, define

$$\mu_t^x = \frac{\mathbb{E}_x\left[Z_t\left(\frac{1}{t}\int_0^t \delta_{Y_s}\,ds\right)\right]}{\mathbb{E}_x Z_t} \text{ if } I = [0, +\infty)$$

or

$$\mu_t^x = \frac{\mathbb{E}_x\left[Z_t\left(\frac{1}{t}\sum_{s<t} \delta_{Y_s}\right)\right]}{\mathbb{E}_x Z_t} \text{ if } I = \mathbb{N}.$$

We obtain the following quasi-ergodic (*sensu* Breyer and Roberts, 1999) result:

Corollary

Assume (A). Then

$$\mathcal{W}_d\left(\mu_t^x, \nu_Q\right) = \sup_{f \in \mathrm{Lip}_1(d)} \frac{\mathbb{E}_x\left[Z_t\left(\frac{1}{t}\int_0^t f(Y_s)\,ds - \nu_Q(f)\right)\right]}{\mathbb{E}_x Z_t} \leq \frac{C}{t}.$$

When coupling is faster than killing (1)

As expected from Cloez and Thai (2016) and Journel and Monmarché (2022):

Proposition

Assume that, for all $x, y \in E$,

$$\mathbb{E}\left[d(Y_t^x, Y_t^y)\right] \leq Ce^{-\gamma t}\, d(x,y), \;\forall t \geq 0$$

with $\gamma > \operatorname{osc}(\rho)$. Then Condition (A) holds true.

When coupling is faster than killing (2)

We can also obtain a local version, which improves Villemonais (2020), by observing that

$$\frac{\partial}{\partial t}\mathbb{E}\left[\frac{Z_t(Y^x)}{\mathbb{E}_x Z_t} d(Y_t^x, Y_t^y)\right]$$

$$\leq \frac{\mathbb{E}\left[e^{-\int_0^t \rho(Y_s^x)\mathrm{d}s}(-\rho(Y_t^x)d(Y_t^x,Y_t^y)+L^c d(Y_t^x,Y_t^y)+\|\rho\|_\infty d(Y_t^x,Y_t^y))\right]}{\mathbb{E}_x\left[e^{-\int_0^t \rho(Y_s)\,\mathrm{d}s}\right]},$$

where L^c is the generator of the coupling of Y^x and Y^y.
Hence, assuming that

$$\sigma := \sup_{x\neq y} \frac{L^c d(x,y)}{d(x,y)} - \rho(x) + \|\rho\|_\infty < 0,$$

(A) is satisfied with $C_A = 1$ and $\gamma = -\sigma$.

Back to Bernoulli convolutions

For the chain

$$Y_{n+1} = \frac{1}{2} Y_n + \theta_{n+1}$$

on $E = [-2, 2]$, since

$$|Y_n^x - Y_n^y| = 2^{-n}|x - y|$$

(A) is clear (provided p is Lipschitz).

More generally, for the chain

$$Y_{n+1} = f_{\theta_{n+1}}(Y_n),$$

where (θ_n) are i.i.d. and $(f_\theta)_\theta$ is a family of Lipschitz functions such that, for all θ, $\ell_\theta := \|f_\theta\|_{\text{Lip}} \leq 1$.

Proposition

If $p = e^{-\rho}$ with ρ Lipschitz and $\mathbb{P}(l_{\theta_1} = 1) < e^{-\text{osc}(\rho)}$, then (A) holds true.

Switched dynamical systems

We consider the PDMP $(X_t, I_t)_{t\in\mathbb{R}_+}$, where the environment I_t is an irreducible Markov chain on a finite state space S and

$$\dot{X}_t = F_{I_t}(X_t) \quad \text{in } \mathbb{R}^k$$

where, for all $i \in S$,

$$\langle F_i(x) - F_i(y), x - y\rangle \leq -\gamma \|x - y\|^2$$

for some constant $\gamma > 0$.
For R large enough, the ball $\|x\| \leq R$ is invariant for all F_i.
We define the distance

$$d((x,i),(y,j)) = \mathbb{1}_{i\neq j} + \mathbb{1}_{i=j}\frac{\|x-y\|}{2R}, \ \forall \|x\| \leq R, \ \|y\| \leq R, \ i,j \in S.$$

We assume that the killing rate $\rho(x,i)$ is Lipschitz.

Proposition

Using the classical independent Markov coupling, Condition (A) holds true.