Geometric ergodicity for the Bouncy Particle Sampler

Pierre Monmarché

MCQMC 2018

Joint work with Alain Durmus (ENS Cachan) and Arnaud Guillin (Université Clermont-Ferrand)

MCMC with velocity jump processes

- target distribution $\pi(x) \propto \exp(-U(x))$ on $\mathbb{R}^d$
- Markov process $(X_t, Y_t)_{t\geqslant 0}$ on $\mathbb{R}^d \times \mathcal{V}$ with $\mathcal{V} \subset \mathbb{R}^d$, such that :
  - the position $X_t = X_0 + \int_0^t Y_s \mathrm{d}s$ for all $t \geqslant 0$
  - the velocity Y is piecewise constant.
  - ergodic w.r.t. $\mu = \pi \otimes \nu_v$ for some ν_v on $\mathcal{V}$, so that, for all f,

$$\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s \underset{t\to\infty}{\longrightarrow} \int f(x)\pi(x)\mathrm{d}x$$

MCMC with velocity jump processes

- target distribution $\pi(x) \propto \exp(-U(x))$ on $\mathbb{R}^d$
- Markov process $(X_t, Y_t)_{t\geqslant 0}$ on $\mathbb{R}^d \times \mathcal{V}$ with $\mathcal{V} \subset \mathbb{R}^d$, such that :
  - the position $X_t = X_0 + \int_0^t Y_s \mathrm{d}s$ for all $t \geqslant 0$
  - the velocity Y is piecewise constant.
  - ergodic w.r.t. $\mu = \pi \otimes \nu_v$ for some ν_v on $\mathcal{V}$, so that, for all f,

$$\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s \underset{t\to\infty}{\longrightarrow} \int f(x)\pi(x)\mathrm{d}x$$

Assets:

- kinetic (inertia = short-term memory, prevent backtracking)
- exact simulation by thinning (= rejection-free Metropolization)

MCMC with velocity jump processes

- target distribution $\pi(x) \propto \exp(-U(x))$ on $\mathbb{R}^d$
- Markov process $(X_t, Y_t)_{t\geqslant 0}$ on $\mathbb{R}^d \times \mathcal{V}$ with $\mathcal{V} \subset \mathbb{R}^d$, such that :
  - the position $X_t = X_0 + \int_0^t Y_s \mathrm{d}s$ for all $t \geqslant 0$
  - the velocity Y is piecewise constant.
  - ergodic w.r.t. $\mu = \pi \otimes \nu_v$ for some ν_v on $\mathcal{V}$, so that, for all f,

$$\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s \underset{t\to\infty}{\longrightarrow} \int f(x)\pi(x)\mathrm{d}x$$

Assets:

- kinetic (inertia = short-term memory, prevent backtracking)
- exact simulation by thinning (= rejection-free Metropolization)

Many possibilites : Bouncy Particle Sampler, Zig-Zag process, randomized bounces, etc.

The bouncy particle sampler

The jump rate $\lambda(x,y) = (y \cdot \nabla U(x))_+$ gives the jump time law by

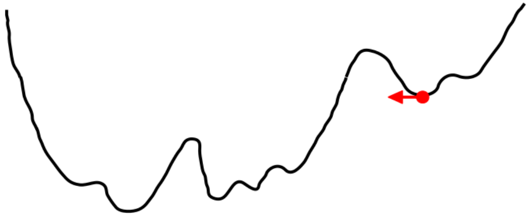
$$\mathbb{P}\,(T > t) \;=\; \exp\left(-\int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right)$$

The bouncy particle sampler

The jump rate $\lambda(x,y) = (y \cdot \nabla U(x))_+$ gives the jump time law by

$$\mathbb{P}(T > t) \;=\; \exp\left(-\int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right) \;=\; \mathbb{P}\left(E > \int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right)$$

$$\begin{aligned}\int_0^t \lambda(X_s, Y_s)\mathrm{d}s &= U(X_t) - U(X_0) && \text{as long as } U(X_t) \text{ increases}\\ &= 0 && \text{as long as it decreases}\end{aligned}$$

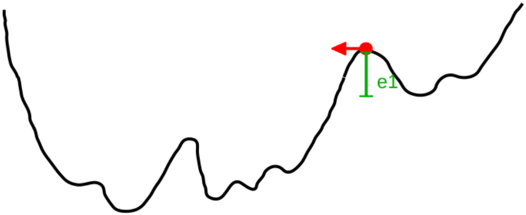


The bouncy particle sampler

The jump rate $\lambda(x,y) = (y \cdot \nabla U(x))_+$ gives the jump time law by

$$\mathbb{P}(T > t) = \exp\left(-\int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right) = \mathbb{P}\left(E > \int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right)$$

$$\begin{aligned} \int_0^t \lambda(X_s, Y_s)\mathrm{d}s &= U(X_t) - U(X_0) && \text{as long as } U(X_t) \text{ increases} \\ &= 0 && \text{as long as it decreases} \end{aligned}$$

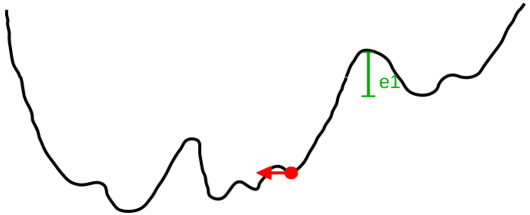


The bouncy particle sampler

The jump rate $\lambda(x,y) = (y \cdot \nabla U(x))_+$ gives the jump time law by

$$\mathbb{P}\left(T > t\right) \ = \ \exp\left(-\int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right) \ = \ \mathbb{P}\left(E > \int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right)$$

$$\begin{aligned} \int_0^t \lambda(X_s, Y_s)\mathrm{d}s \ &= \ U(X_t) - U(X_0) && \text{as long as } U(X_t) \text{ increases} \\ &= \ 0 && \text{as long as it decreases} \end{aligned}$$

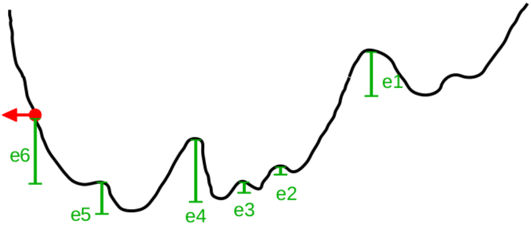


The bouncy particle sampler

The jump rate $\lambda(x,y) = (y \cdot \nabla U(x))_+$ gives the jump time law by

$$\mathbb{P}(T > t) \;=\; \exp\left(-\int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right) \;=\; \mathbb{P}\left(E > \int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right)$$

$$\begin{aligned} \int_0^t \lambda(X_s, Y_s)\mathrm{d}s &= U(X_t) - U(X_0) && \text{as long as } U(X_t) \text{ increases} \\ &= 0 && \text{as long as it decreases} \end{aligned}$$

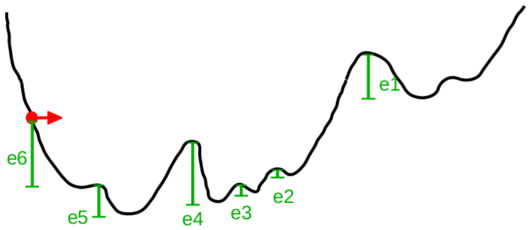


The bouncy particle sampler

The jump rate $\lambda(x,y) = (y \cdot \nabla U(x))_+$ gives the jump time law by

$$\mathbb{P}\left(T > t\right) = \exp\left(-\int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right) = \mathbb{P}\left(E > \int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right)$$

$$\begin{aligned} \int_0^t \lambda(X_s, Y_s)\mathrm{d}s &= U(X_t) - U(X_0) && \text{as long as } U(X_t) \text{ increases} \\ &= 0 && \text{as long as it decreases} \end{aligned}$$

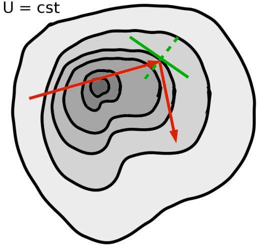


The bouncy particle sampler

The jump rate $\lambda(x,y) = (y \cdot \nabla U(x))_+$ gives the jump time law by

$$\mathbb{P}(T > t) \;=\; \exp\left(-\int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right) \;=\; \mathbb{P}\left(E > \int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right)$$

$$\begin{aligned} \int_0^t \lambda(X_s, Y_s)\mathrm{d}s &= U(X_t) - U(X_0) && \text{as long as } U(X_t) \text{ increases} \\ &= 0 && \text{as long as it decreases} \end{aligned}$$

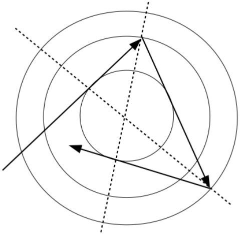


At the jump time: reflection of $y \mathbin{/\!/} \nabla_x U(x)$.

$$y \;\leftarrow\; R(x,y) \;=\; y - 2\frac{y \cdot \nabla U(x)}{|\nabla U(x)|^2}\nabla U(x)$$

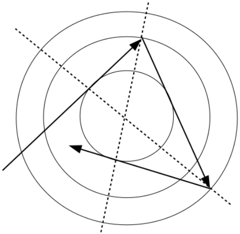
Non irreducibility

Problem:



Non irreducibility

Problem:



Possible solution: at constant rate λ_v, resample the velocity

$$y \leftarrow W \sim \nu_v$$

with ν_v rotation-invariant distribution on $\mathcal{V}$ (uniform on $\mathcal{V} = \mathbb{S}_{d-1}$, standard Gaussian on $\mathcal{V} = \mathbb{R}^d \dots$)

Long-time convergence

Notations:

- semigroup $P_t f(x,y) \;=\; \mathbb{E}_{(x,y)}\left[f(X_t,Y_t)\right]$
- generator $Lf(x,y) = (\partial_t)_{|t=0} P_t f(x,y)$. Here,

$$\begin{aligned} Lf(x,y) \;&=\; y\cdot\nabla_x f(x,y) + \lambda(x,y)\left(f(x,R(x,y)) - f(x,y)\right) \\ &\quad + \lambda_v\left(\int f(x,\cdot)\mathrm{d}\nu_v - f(x,y)\right). \end{aligned}$$

Goal: quantify the convergence of $\mathcal{L}aw(X_t,Y_t)$ toward $\mu = \pi\otimes\nu_v$:

$$\|\delta_{(x,y)}P_t - \mu\|_{TV} \;\leqslant\; Ce^{-\rho t}V(x,y)$$

Previous and similar works

- Many one-dimensional works: Miclo, M. (2013, 2014), Fontbona, Guérin, Malrieu (2012, 2016), Bierkens, Roberts (2017), etc.
- M. (2016), compact periodic space $\mathbb{T}^d$, unitary velocity $\mathcal{V} = \mathbb{S}_{d-1}$
- Bouchard-Côté, Deligiannidis, Doucet (2017), unitary velocity, unbounded space, restrictive assumptions on $U = -\log \pi$ and λ_v.
- Bierkens, Roberts, Zitt (2017) on the Zig-Zag process.

Previous and similar works

- Many one-dimensional works: Miclo, M. (2013, 2014), Fontbona, Guérin, Malrieu (2012, 2016), Bierkens, Roberts (2017), etc.
- M. (2016), compact periodic space $\mathbb{T}^d$, unitary velocity $\mathcal{V} = \mathbb{S}_{d-1}$
- Bouchard-Côté, Deligiannidis, Doucet (2017), unitary velocity, unbounded space, restrictive assumptions on $U = -\log \pi$ and λ_v.
- Bierkens, Roberts, Zitt (2017) on the Zig-Zag process.
- Durmus, Guillin, M. (2018):
  - less restrictive assumptions (thin tails, no restriction on λ_v)
  - explicit coupling

The Meyn-Tweedie approach

1. Find a Lyapunov function $V \geqslant 1$ and $\rho, C > 0$ such that $\forall (x,y) \in \mathbb{R}^d \times \mathcal{V}$

$$LV(x,y) \leqslant -\rho V(x,y) + C,$$

   that is to say such that, for all $t \geqslant 0$

$$\partial_t \mathbb{E}_{(x,y)}\left[V(X_t,Y_t)\right] \leqslant -\rho \mathbb{E}_{(x,y)}\left[V(X_t,Y_t)\right] + C$$

2. A Doeblin condition or, alternatively, a coupling: for all (x,y) and (x',y') in a given compact set of $\mathbb{R}^d \times \mathcal{V}$, for some time t, define two processes $(X_s,Y_s)_{s\geqslant 0}$ and $(X'_s,Y'_s)_{s\geqslant 0}$ with these initial conditions in such a way that

$$\mathbb{P}\left(X_t = X'_t \text{ and } Y_t = Y'_t\right) > 0.$$

The Lyapunov function

Find V such that $LV \leqslant -\lambda V$ outside from a compact. Problem :

$$L = y \cdot \nabla_x + (y \cdot \nabla_x U)_+ \left(\delta_{x,R(x,y)} - \delta_{x,y}\right) + \lambda_v \left(\delta_x \otimes \nu_v - \delta_{x,y}\right)$$

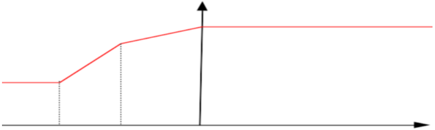
is non-local.

The Lyapunov function

Find V such that $LV \leqslant -\lambda V$ outside from a compact. Problem :

$$L = y \cdot \nabla_x + (y \cdot \nabla_x U)_+ \left(\delta_{x,R(x,y)} - \delta_{x,y}\right) + \lambda_v \left(\delta_x \otimes \nu_v - \delta_{x,y}\right)$$

is non-local. Set $V(x,y) = \exp(\kappa U(x)) \varphi(y \cdot \nabla U(x))$ with φ like:

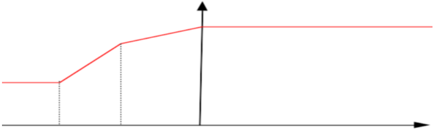


The Lyapunov function

Find V such that $LV \leqslant -\lambda V$ outside from a compact. Problem :

$$L = y \cdot \nabla_x + (y \cdot \nabla_x U)_+ \left(\delta_{x,R(x,y)} - \delta_{x,y}\right) + \lambda_v \left(\delta_x \otimes \nu_v - \delta_{x,y}\right)$$

is non-local. Set $V(x,y) = \exp(\kappa U(x))\varphi(y \cdot \nabla U(x))$ with φ like:



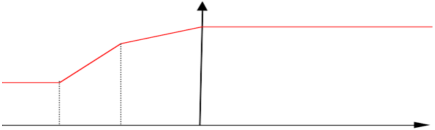
- If $y \cdot \nabla U(x) < 0$: transport $\exp(\kappa U) \searrow$, refreshment $\varphi \nearrow$.
- If $y \cdot \nabla U(x) > 0$: transport $\exp(\kappa U) \nearrow$, bounces $\varphi \searrow$.
- If $y \cdot \nabla U(x) \simeq 0$: refreshment $\varphi \searrow$.

The Lyapunov function

Find V such that $LV \leqslant -\lambda V$ outside from a compact. Problem :

$$L = y \cdot \nabla_x + (y \cdot \nabla_x U)_+ \left(\delta_{x,R(x,y)} - \delta_{x,y}\right) + \lambda_v \left(\delta_x \otimes \nu_v - \delta_{x,y}\right)$$

is non-local. Set $V(x,y) \;=\; \exp\left(\kappa W(x)\right) \varphi\left(y \cdot \nabla W(x)\right)$ with φ like:



$$W(x) \;=\; F(U(x))$$

- If $y \cdot \nabla U(x) < 0$: transport $\exp(\kappa U) \searrow$, refreshment $\varphi \nearrow$.
- If $y \cdot \nabla U(x) > 0$: transport $\exp(\kappa U) \nearrow$, bounces $\varphi \searrow$.
- If $y \cdot \nabla U(x) \simeq 0$: refreshment $\varphi \searrow$.

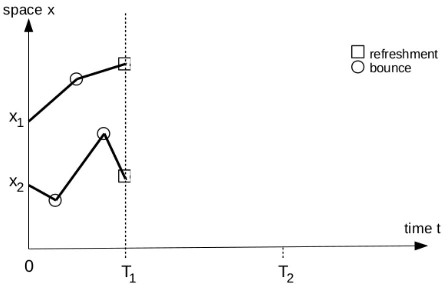
Conditions on U

A bit intricate but, for instance, allows for:

- $U(x) \simeq |x|^\alpha$, $\alpha > 1$, ν_v admits a Gaussian moment.
- $\nabla^2 U$ bounded, $|\nabla U| \to \infty$ at infinity and bounded velocity.
- $|\nabla U| \geqslant c > 0, \nabla^2 U \to 0$ at infinity, bounded velocity and $\lambda_v \leqslant \lambda_0$.

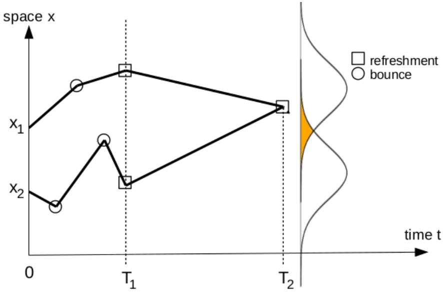
The coupling

- Same Poisson process for the refreshment times of both processes.



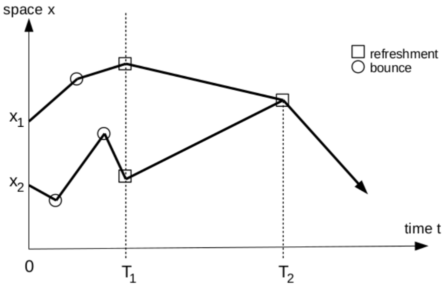
The coupling

- Same Poisson process for the refreshment times of both processes.
- At the first refreshment time T_1, couple the velocities so that, at the next refreshment time T_2, both processes reach the same position.



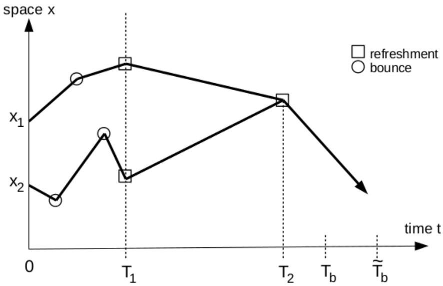
The coupling

- Same Poisson process for the refreshment times of both processes.
- At the first refreshment time T_1, couple the velocities so that, at the next refreshment time T_2, both processes reach the same position.
- At time T_2, both processes take the same velocity.



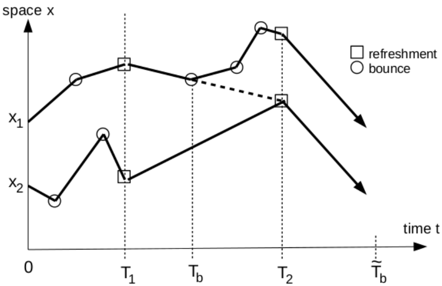
The coupling

- Same Poisson process for the refreshment times of both processes.
- At the first refreshment time T_1, couple the velocities so that, at the next refreshment time T_2, both processes reach the same position.
- At time T_2, both processes take the same velocity.
- Success if no bounce occurs in $[T_1, T_2]$



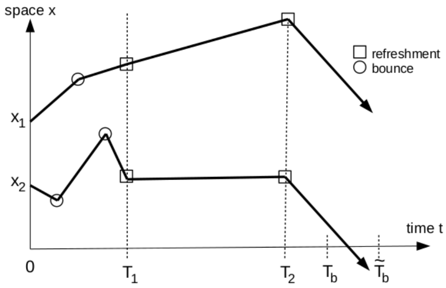
The coupling

- Same Poisson process for the refreshment times of both processes.
- At the first refreshment time T_1, couple the velocities so that, at the next refreshment time T_2, both processes reach the same position.
- At time T_2, both processes take the same velocity.
- Success if no bounce occurs in $[T_1, T_2]$



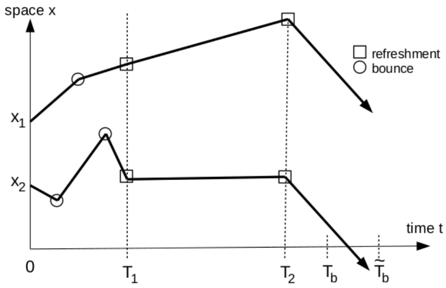
The coupling

- Same Poisson process for the refreshment times of both processes.
- At the first refreshment time T_1, couple the velocities so that, at the next refreshment time T_2, both processes reach the same position.
- At time T_2, both processes take the same velocity.
- Success if no bounce occurs in $[T_1, T_2]$, and velocity coupling suceeds.



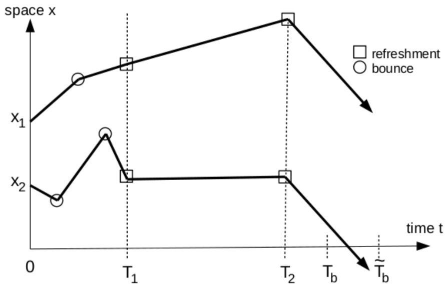
The coupling

- Same Poisson process for the refreshment times of both processes.
- At the first refreshment time T_1, couple the velocities so that, at the next refreshment time T_2, both processes reach the same position.
- At time T_2, both processes take the same velocity.
- Success if no bounce occurs in $[T_1, T_2]$, and velocity coupling suceeds.
- Remark: each process is Markovian, but $(X_t, Y_t, X'_t, Y'_t)_{t\geqslant 0}$ is not.



The coupling

- Same Poisson process for the refreshment times of both processes.
- At the first refreshment time T_1, couple the velocities so that, at the next refreshment time T_2, both processes reach the same position.
- At time T_2, both processes take the same velocity.
- Success if no bounce occurs in $[T_1, T_2]$, and velocity coupling suceeds.
- Remark: each process is Markovian, but $(X_t, Y_t, X'_t, Y'_t)_{t\geqslant 0}$ is not.
- Remark 2: works for all velocity jump process, given refreshment.



A toy model

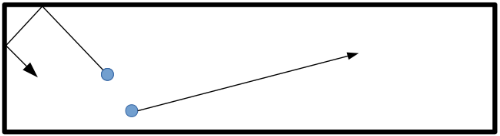
Bounces do not help us in the coupling.

Ex: uniform measure on $[0,1]\times[0,\varepsilon]^{d-1}$

A toy model

Bounces do not help us in the coupling.

Ex: uniform measure on $[0,1] \times [0,\varepsilon]^{d-1}$



A toy model

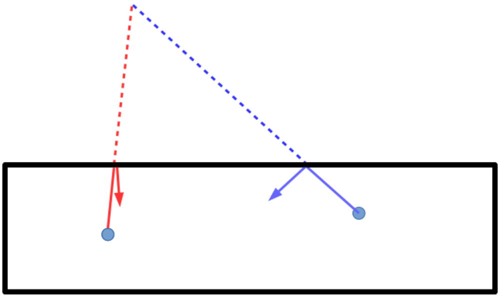
Bounces do not help us in the coupling.

Ex: uniform measure on $[0,1] \times [0,\varepsilon]^{d-1}$

A toy model

Bounces do not help us in the coupling.

Ex: uniform measure on $[0,1] \times [0,\varepsilon]^{d-1}$

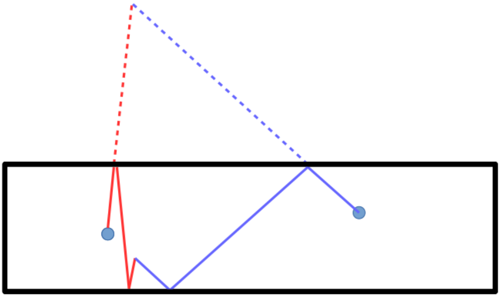


Bound in the proof: $\mathcal{O}\left(\varepsilon^{d-1}\right)$

A toy model

Bounces do not help us in the coupling.

Ex: uniform measure on $[0,1] \times [0,\varepsilon]^{d-1}$



Bound in the proof: $\mathcal{O}\left(\varepsilon^{d-1}\right)$

True relaxation time : $\mathcal{O}\left(\sqrt{1+(d-1)\varepsilon}\right)$

Perturbation result

Two piecewise deterministic processes with generator L_1 and L_2, equilibrium μ_1 and μ_2. Suppose L_1 is V-ergodic with $\mu_2(V) < \infty$. Then

$$\|\mu_1 - \mu_2\|_{TV} \leqslant C \sup_{|h| \leqslant V} \int (L_1 - L_2) h \mathrm{d}\mu_2 .$$

Example: for the BPS, replace the true jump rate by $\lambda(x,y) \wedge \lambda_*$ for some constant $\lambda_* > 0$.

$$\|\mu - \mu(\lambda_*)\|_{TV} \leqslant C \int \left(|\nabla U(x)| - \lambda_*\right)_+ e^{\kappa W(x) - U(x)} \mathrm{d}x .$$

In particular, if π has a Gaussian tail, then

$$\|\mu - \mu(\lambda_*)\|_{TV} \leqslant C e^{-\rho \lambda_*^2} .$$

Some references :

V. Calvez, G. Raoul, and C. Schmeiser.
Confinement by biased velocity jumps: aggregation of *escherichia coli*.
Kinet. Relat. Models, 8(4):651–666, 2015.

J. Fontbona, H. Guérin, and F. Malrieu.
Long time behavior of telegraph processes under convex potentials.
Stochastic Process. Appl., 126(10):3077–3101, 2016.

L. Miclo and P. Monmarché.
Étude spectrale minutieuse de processus moins indécis que les autres.
Lecture Notes in Mathematics, 2078:459, 2013.

P. Monmarché.
Piecewise deterministic simulated annealing.
ALEA Lat. Am. J. Probab. Math. Stat., 13(1):357–398, 2016.

Some references :

V. Calvez, G. Raoul, and C. Schmeiser.
Confinement by biased velocity jumps: aggregation of *escherichia coli*.
Kinet. Relat. Models, 8(4):651–666, 2015.

J. Fontbona, H. Guérin, and F. Malrieu.
Long time behavior of telegraph processes under convex potentials.
Stochastic Process. Appl., 126(10):3077–3101, 2016.

L. Miclo and P. Monmarché.
Étude spectrale minutieuse de processus moins indécis que les autres.
Lecture Notes in Mathematics, 2078:459, 2013.

P. Monmarché.
Piecewise deterministic simulated annealing.
ALEA Lat. Am. J. Probab. Math. Stat., 13(1):357–398, 2016.

Thank you for your attention !