# Bouncing with velocity jump processes

Pierre Monmarché

June 4 2019, HIM Bonn

1. Kinetic MCMC and bounces

2. Factorization

# Kinetic Markov Chain Monte-Carlo

MCMC principle : given a target probability law $\pi \propto e^{-U(x)}\mathrm{d}x$ on $\mathbb{R}^d$, define a $\pi$-ergodic Markov process :

$$\forall \varphi\,, \qquad \frac{1}{t}\int_0^t \varphi(X_s)\mathrm{d}s \underset{t\to+\infty}{\longrightarrow} \int \varphi(x)\pi(\mathrm{d}x)\,. \tag{1}$$

Kinetic MCMC : consider an auxiliary law $\nu$, $\mu = \pi \otimes \nu$ and define a $\mu$-ergodic kinetic Markov process $(X_t, V_t)_{t\geqslant 0}$, so that (1) still holds.

Advantages :

- Velocity = instantaneous memory ; inertia = less going back = ballistic rather than diffusive behaviour. Better convergence expected = better exploration.
- Exact simulation in some cases.
- Sometimes physically relevant (ex : molecular dynamics)

# Kinetic samplers

Specifications :

- $(X, V)$ Markov on $\mathbb{R}^d \times \mathbb{R}^d$
- $\partial_t X = V$
- Equilibrium $\mu \propto e^{-U(x)} \mathrm{d}x e^{-\frac{1}{2}|v|^2} \mathrm{d}v = e^{-H(x,v)} \mathrm{d}x \mathrm{d}v$

# Kinetic samplers

Specifications :

- $(X, V)$ Markov on $\mathbb{R}^d \times \mathbb{R}^d$
- $\partial_t X = V$
- Equilibrium $\mu \propto e^{-U(x)} \mathrm{d}x e^{-\frac{1}{2}|v|^2} \mathrm{d}v = e^{-H(x,v)} \mathrm{d}x \mathrm{d}v$

Clasical examples (denoting $\rho_t(x, v)$ the density of particles) :

$$
\begin{aligned}
\partial_t \rho + v \cdot \nabla_x \rho &= -\nabla U(x) \cdot \nabla_v \rho \\
\partial_t \rho + v \cdot \nabla_x \rho &= -\nabla U(x) \cdot \nabla_v \rho + \nabla_v \cdot (-v\rho + \nabla_v \rho) \\
\partial_t \rho + v \cdot \nabla_x \rho &= -\nabla U(x) \cdot \nabla_v \rho + \lambda \left( M(v) \int_{\mathbb{R}^d} \rho(x, w) \mathrm{d}w - \rho \right)
\end{aligned}
$$

# Kinetic samplers

Specifications :

- $(X, V)$ Markov on $\mathbb{R}^d \times \mathbb{R}^d$
- $\partial_t X = V$
- Equilibrium $\mu \propto e^{-U(x)}\mathrm{d}x e^{-\frac{1}{2}|v|^2}\mathrm{d}v = e^{-H(x,v)}\mathrm{d}x\mathrm{d}v$

Clasical examples (denoting $\rho_t(x, v)$ the density of particles) :

$$\begin{aligned}
\partial_t \rho + v \cdot \nabla_x \rho &= -\nabla U(x) \cdot \nabla_v \rho \\
\partial_t \rho + v \cdot \nabla_x \rho &= -\nabla U(x) \cdot \nabla_v \rho + \nabla_v \cdot (-v\rho + \nabla_v \rho) \\
\partial_t \rho + v \cdot \nabla_x \rho &= -\nabla U(x) \cdot \nabla_v \rho + \lambda \left( M(v) \int_{\mathbb{R}^d} \rho(x, w)\mathrm{d}w - \rho \right)
\end{aligned}$$

Velocity jump processes : $V$ piecewise constant (jump rate + kernel).

$$\partial_t \rho + v \cdot \nabla_x \rho = \int_{\mathbb{R}^d \times \mathbb{R}^d} \lambda(y, w) q(y, w, x, v) \rho(y, w)\mathrm{d}y\mathrm{d}w - \lambda(x, v)\rho(x, v)$$

# Bounce mechanism

- jump rate $\lambda(x,v) = (v \cdot \nabla U(x))_+$
- Jump kernel $\delta_{R(x,v)}$ with

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

In other words $(X_t, Y_t) = (x_0 + tv_0, v_0)$ up to a random time $T$ with law

$$\mathbb{P}\left(T > t\right) = \exp\left(-\int_0^t \left(v_0 \cdot \nabla U(x_0 + sv_0)\right)_+\right)$$

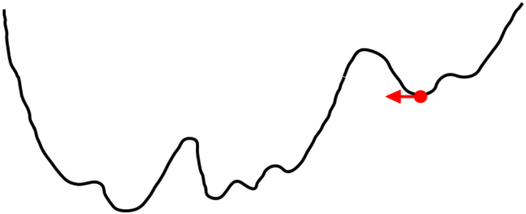
or equivalently, if $E$ is a standard exponential random variable,

$$T \;\overset{law}{=}\; \inf\left\{t > 0,\ E > \int_0^t \lambda(X_s, Y_s)\mathrm{d}s\right\}.$$

# Interpretation

$\lambda(x,v) = (v \cdot \nabla U(x))_+$ ; and since $v = x'$,

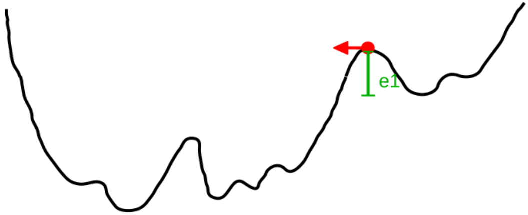
$$\begin{aligned} \int_0^t \lambda(X_s, Y_s)\mathrm{d}s &= U(X_t) - U(X_0) && \text{when climbing up} \\ &= 0 && \text{while going down.} \end{aligned}$$



# Interpretation

$\lambda(x,v) = (v\cdot\nabla U(x))_+$ ; and since $v = x'$,

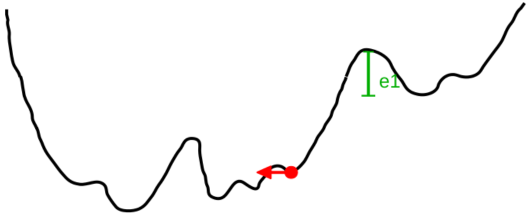
$$\int_0^t \lambda(X_s,Y_s)\mathrm{d}s = U(X_t) - U(X_0) \quad \text{when climbing up}$$
$$= 0 \quad \text{while going down.}$$



# Interpretation

$\lambda(x,v) = (v \cdot \nabla U(x))_+$ ; and since $v = x'$,

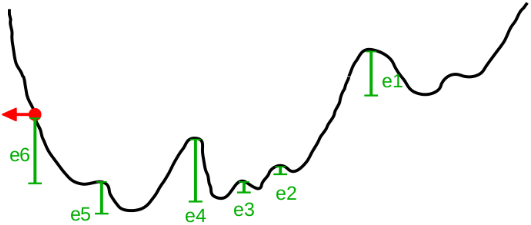
$$\int_0^t \lambda(X_s, Y_s)\mathrm{d}s = U(X_t) - U(X_0) \quad \text{when climbing up}$$
$$= 0 \quad \text{while going down.}$$



# Interpretation

$\lambda(x,v) = (v \cdot \nabla U(x))_+$ ; and since $v = x'$,

$$\int_0^t \lambda(X_s, Y_s)\mathrm{d}s = U(X_t) - U(X_0) \quad \text{when climbing up}$$
$$= 0 \quad \text{while going down.}$$

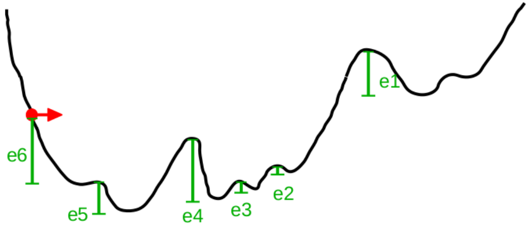


# Interpretation

$\lambda(x,v) = (v \cdot \nabla U(x))_+$ ; and since $v = x'$,

$$\int_0^t \lambda(X_s, Y_s)\mathrm{d}s \;=\; U(X_t) - U(X_0) \quad \text{when climbing up}$$

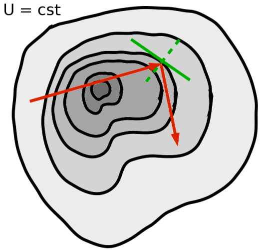
$$\;=\; 0 \quad \text{while going down.}$$



# Interpretation

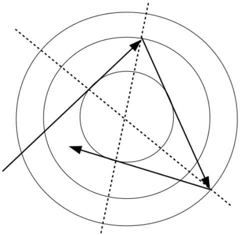
$\lambda(x,v) = (v \cdot \nabla U(x))_+$ ; and since $v = x'$,

$$\int_0^t \lambda(X_s, Y_s) \mathrm{d}s = U(X_t) - U(X_0) \quad \text{when climbing up}$$
$$= 0 \quad \text{while going down.}$$



# Remarks

- This Bouncy Particle Sampler is not ergodic in general :



  $\Rightarrow$ add velocity refreshment at constant rate

- Exact simulation without discretization through a thinning method $\Rightarrow$ no bias on the target law.

- Many variants on the same theme : Zig-zag, randomized bounces, etc.

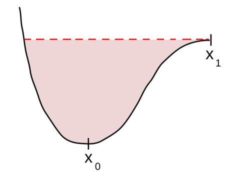
# Some results

## Theorem (Eyring-Kramers formula, M. 2016)

*In dimension 1,* $U = U_0/\varepsilon$*, let* $\tau = \inf\{s > 0,\ X_s = x_1 \mid X_0 = x_0\}$*. Then*

$$\mathbb{E}[\tau] \underset{\varepsilon\to 0}{\simeq} \sqrt{\frac{8\pi\varepsilon}{U''(x_0)}}\, e^{\frac{U(x_1)-U(x_0)}{\varepsilon}}$$

$$\mathbb{P}(\tau \geq t\mathbb{E}[\tau]) \underset{\varepsilon\to 0}{\longrightarrow} e^{-t}.$$




## Theorem (Durmus, Guillin, M. 2018)

*In any dimension, with refreshment, under some conditions on* $U$ *(ex :* $U(x) \simeq |x|^\alpha$*,* $\alpha \geqslant 1$*),*

$$\|\rho_t - \mu\|_{TV} \leqslant Ce^{-rt} \int e^{\kappa H(x,v)} \rho_0(x,v) dx dv\,.$$

*If* $U = U_0/\varepsilon$*,* $r \geqslant e^{-\theta/\varepsilon}$*.*

# Generator of a kinetic process

Decompose the generator $L$ (such that $\partial_t \int \varphi \rho_t = \int L\varphi\rho_t$) as

$$L \;=\; \mathcal{T} + \mathcal{F} + \mathcal{D}$$

where

- $\mathcal{T} = v \cdot \nabla_x$ is the transport (only thing acting on $x$)
- $\mathcal{D}$ is reversible w.r.t. $M$ ($\int f \mathcal{D} g M \mathrm{d}v = \int g \mathcal{D} f \mathrm{d}v$)

The target measure is only taken into account by $\mathcal{F}$. If

$$\forall x, \varphi \qquad \int \mathcal{F}\varphi(x,v) M(v)\mathrm{d}v = -\int \left(v \cdot \nabla U(x)\right) \varphi(x,v) M(v) \mathrm{d}v \,,$$

then $\mu$ is invariant. This condition is linear in $\nabla U$.

# Factorization

If $\nabla U(x) = \sum_{i=1}^N \xi_i(x)$ and if $\mathcal{F}_i$ satisfies

$$\forall x, \varphi \qquad \int \mathcal{F}_i \varphi(x,v) M(v) \mathrm{d}v = - \int \left( v \cdot \xi_i(x) \right) \varphi(x,v) M(v) \mathrm{d}v \,,$$

then

$$L = \mathcal{T} + \sum_{i=1}^N \mathcal{F}_i + \mathcal{D}$$

admits $\mu$ as an invariant measure.

# Factorization

If $\nabla U(x) = \sum_{i=1}^{N} \xi_i(x)$ and if $\mathcal{F}_i$ satisfies

$$\forall x, \varphi \qquad \int \mathcal{F}_i \varphi(x,v) M(v) \mathrm{d}v = -\int \left(v \cdot \xi_i(x)\right) \varphi(x,v) M(v) \mathrm{d}v \,,$$

then

$$L = \mathcal{T} + \sum_{i=1}^{N} \mathcal{F}_i + \mathcal{D}$$

admits $\mu$ as an invariant measure. Examples :

$$\begin{array}{lcll} \mathcal{F}_i \varphi & = & -\xi_i(x) \cdot \nabla_y \varphi & \text{(drift)} \\ \mathcal{F}_i \varphi & = & \left(y \cdot \xi_i(x)\right)_+ \left(\varphi\left(x, R_{\xi_i}(x,v)\right) - \varphi(x,v)\right) & \text{(bounce)} \end{array}$$

# Ex : Zigzag

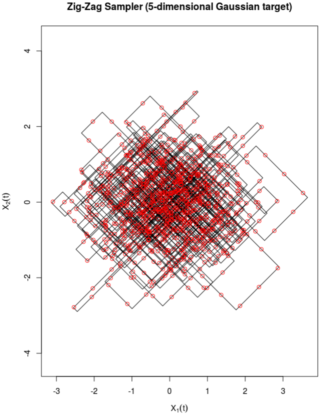
Decomposing over the canonical basis $(e_i)_{1\leqslant i\leqslant d}$

$$\nabla U(x) \; = \; \sum_{i=1}^{d} \partial_{x_i} U(x) e_i := \sum_{i=1}^{d} \xi_i(x)$$

Each $\xi_i$ is dealt with through bounces.

At rate $(y_i \partial_{x_i} U(x))_+$, $v_i$ jumps to $-v_i$

(figure Joris Bierkens)



# Alternative to multi-time-step methods

Suppose $U = U_1 + U_2$ with

- $\nabla U_1(x)$ numerically cheap but with high and fast variation ; possibly singular (ex : short-range interactions).
- $\nabla U_2(x)$ numerically expansive but bounded and with slow variations (ex : long-range interactions).

# Alternative to multi-time-step methods

Suppose $U = U_1 + U_2$ with

- $\nabla U_1(x)$ numerically cheap but with high and fast variation ; possibly singular (ex : short-range interactions).
- $\nabla U_2(x)$ numerically expansive but bounded and with slow variations (ex : long-range interactions).

Example : mean-field Lennard-Jones particles

$$U(x) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \neq i} W(|x_i - x_j|)$$

with $W(r) = 1/r^{12} - 1/r^6$ decomposed as

$$W(r) = W(r)\chi(r) + W(r)(1 - \chi(r))\,.$$

# Decomposition

Considering $A_i$ such that $A_i^T x = x_i \in \mathbb{R}^3$, split

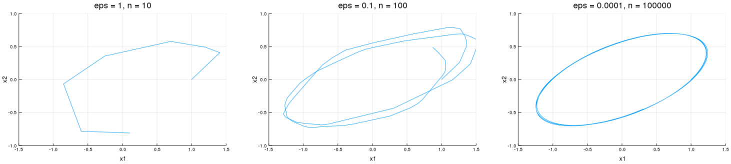
$$\nabla U(x) = \nabla U_{short}(x) + \sum_{i=1}^{N} A_i \sum_{j \neq i} \partial_{x_i} U_{long,j}(x)$$

with $U_{long,j}(x) = W(|x_i - x_j|)(1 - \chi([x_i - x_j|))/N$.
Short-range forces are dealt with by a drift, long-range ones by bounces.
Counting the number of computations of $W'$ :

- For a method only with a drift, $TN^2/\delta$.
- For the factorized method, short-range forces cost $T/\delta \times \mathcal{O}(N)$ (if the number of neighbours is $\mathcal{O}(1)$ and a neighbour list is available). Through a thinning method, for the long-range forces, the average number of computations also scales as $TN$.

# Conclusion

- Implementation in progress for molecular dynamics (promising results)
- Numerical efficiency is problem dependent (how to split, how to bound the jump rates for thinning)
- From bounce to drift (with Pierre-André Zitt and Mathias Rousset)




- Kinetic theory point of view ? Metastability ? Scaling limits ? Etc.

# Références

J. Bierkens, P. Fearnhead, and G. Roberts.
The zig-zag process and super-efficient sampling for Bayesian analysis of big data.
*Ann. Statist.*, 47(3) :1288–1320, 2019.

A. Durmus, A. Guillin, and P. Monmarché.
Geometric ergodicity of the bouncy particle sampler.
*arXiv e-prints*, page arXiv :1807.05401, Jul 2018.

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

E. A. J. F. Peters and G. de With.
Rejection-free monte carlo sampling for general potentials.
*Phys. Rev. E 85, 026703*, 2012.

P. Vanetti, A. Bouchard-Côté, G. Deligiannidis, and A. Doucet.
Piecewise-Deterministic Markov Chain Monte Carlo.
*arXiv e-prints*, page arXiv :1707.05296, Jul 2017.