Sampling with non-reversible processes

Pierre Monmarché

Journée algorithmes stochastiques - Dauphine

1 Reversible sampling

2 Lifted chains and kinetic jump processes

3 Non-reversible diffusions

1 Reversible sampling

2 Lifted chains and kinetic jump processes

3 Non-reversible diffusions

MCMC method

- Target law $\mu \propto e^{-U(x)}\mathrm{d}x$
- Markov process $(X_t)_{t\geqslant 0}$, ergodic with equilibrium μ, so that

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

Choice for X ? Criterion ?

MCMC method

- Target law $\mu \propto e^{-U(x)}\mathrm{d}x$
- Markov process $(X_t)_{t\geqslant 0}$, ergodic with equilibrium μ, so that

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

Choice for X ? Criterion ?

- Asymptotical variance :

$$\sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s - \int f\mathrm{d}\mu\right) \underset{t\to\infty}{\longrightarrow} \mathcal{N}\left(0,\sigma(f)\right)$$

MCMC method

- Target law $\mu \propto e^{-U(x)}\mathrm{d}x$
- Markov process $(X_t)_{t\geqslant 0}$, ergodic with equilibrium μ, so that

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

Choice for X ? Criterion ?

- Asymptotical variance :

$$\sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s - \int f\mathrm{d}\mu\right) \underset{t\to\infty}{\longrightarrow} \mathcal{N}\left(0,\sigma(f)\right)$$

- Speed of relaxation to equilibrium (L^2,$\|\cdot\|_1$,$\mathcal{W}$, Ent...)

$$\mathcal{L}(X_t) \underset{t\to\infty}{\longrightarrow} \mu$$

MCMC method

- Target law $\mu \propto e^{-U(x)}\mathrm{d}x$
- Markov process $(X_t)_{t\geqslant 0}$, ergodic with equilibrium μ, so that

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

Choice for X ? Criterion ?

- Asymptotical variance :

$$\sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s - \int f\mathrm{d}\mu\right) \underset{t\to\infty}{\longrightarrow} \mathcal{N}\left(0,\sigma(f)\right)$$

- Speed of relaxation to equilibrium (L^2,$\|\cdot\|_1$,$\mathcal{W}$, Ent...)

$$\mathcal{L}(X_t) \underset{t\to\infty}{\longrightarrow} \mu$$

- self-correlation...

General remarks

- Changing time $(X_{\lambda t})_{t \geqslant 0}$ is cheating: discretization, amount of alea in the system.

General remarks

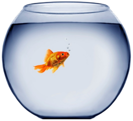
- Changing time $(X_{\lambda t})_{t \geqslant 0}$ is cheating: discretization, amount of alea in the system.
- For any criterion, the goal is to explore efficiently the space.

General remarks

- Changing time $(X_{\lambda t})_{t \geqslant 0}$ is cheating: discretization, amount of alea in the system.
- For any criterion, the goal is to explore efficiently the space.
- (And the explorer is amnesic)

General remarks

- Changing time $(X_{\lambda t})_{t \geqslant 0}$ is cheating: discretization, amount of alea in the system.
- For any criterion, the goal is to explore efficiently the space.
- (And the explorer is amnesic)
- General principle: avoid places which have already been visited.



Theorem (Peskun, 73)

Let p and q two transition kernels on E (finite), irreducible, reversible w.r.t. the same law μ, and such that

$$\forall x \in E, \qquad p(x,x) \leqslant q(x,x).$$

Then $\sigma_p(f) \leqslant \sigma_q(f)$ for any test function f.

Reversible sampling

Definition

The transition kernel p is μ-reversible if for all $x \in E$,

$$\mu(x)p(x,y) \;=\; \mu(y)p(y,x)$$

or, equivalently, denoting $P = (p(x,y))_{x,y\in E^2}$, if

$$L \;=\; P - I$$

is self-adjoint in $L^2(\mu)$

Reversible sampling

Definition

The transition kernel p is μ-reversible if for all $x \in E$,

$$\mu(x)p(x,y) \quad = \quad \mu(y)p(y,x)$$

or, equivalently, denoting $P = (p(x,y))_{x,y \in E^2}$, if

$$L \quad = \quad P - I$$

is self-adjoint in $L^2(\mu)$

- Always the possibility to backtrack.

Reversible sampling

Definition

The transition kernel p is μ-reversible if for all $x \in E$,

$$\mu(x)p(x,y) \;=\; \mu(y)p(y,x)$$

or, equivalently, denoting $P = (p(x,y))_{x,y\in E^2}$, if

$$L \;=\; P - I$$

is self-adjoint in $L^2(\mu)$

- Always the possibility to backtrack.
- Diffusive behaviour : N^2 step to cover a distance N.

A spectral argument

In continuous time,

$$\mathbb{E}_x\left(f(X_t)\right) - \int f \mathrm{d}\mu \;=\; (e^{tL} - \mu) f .$$

Reversible case $\Rightarrow$ L self-adjoint in $L^2(\mu)$ $\Rightarrow$ ON eigenbasis, and

$$\|e^{tL} - \mu\|_{L^2(\mu)} \;=\; \sup_{f \neq 0} \frac{\|(e^{tL} - \mu) f\|}{\|f\|} \;=\; e^{-t\lambda_1}$$

with $\lambda_1 = \min \sigma(-L) \setminus \{0\}$.

A spectral argument

In continuous time,

$$\mathbb{E}_x\left(f(X_t)\right) - \int f \mathrm{d}\mu \quad = \quad (e^{tL} - \mu)f.$$

Reversible case $\Rightarrow$ L self-adjoint in $L^2(\mu)$ $\Rightarrow$ ON eigenbasis, and

$$\|e^{tL} - \mu\|_{L^2(\mu)} \quad = \quad \sup_{f \neq 0} \frac{\|(e^{tL} - \mu)f\|}{\|f\|} \quad = \quad e^{-t\lambda_1}$$

with $\lambda_1 = \min \sigma(-L) \setminus \{0\}$. Setting $\widetilde{L} = L + L'$ with L' anti-adjoint,

$$\partial_t \left(\|(e^{t\widetilde{L}} - \mu)f\|^2 \right) \quad = \quad 2 \left\langle f_t, \widetilde{L} f_t \right\rangle$$

A spectral argument

In continuous time,

$$\mathbb{E}_x\left(f(X_t)\right) - \int f \mathrm{d}\mu \quad = \quad (e^{tL} - \mu) f .$$

Reversible case $\Rightarrow$ L self-adjoint in $L^2(\mu)$ $\Rightarrow$ ON eigenbasis, and

$$\|e^{tL} - \mu\|_{L^2(\mu)} \quad = \quad \sup_{f \neq 0} \frac{\|(e^{tL} - \mu)f\|}{\|f\|} \quad = \quad e^{-t\lambda_1}$$

with $\lambda_1 = \min \sigma(-L) \setminus \{0\}$. Setting $\widetilde{L} = L + L'$ with L' anti-adjoint,

$$\begin{aligned} \partial_t \left(\|(e^{t\widetilde{L}} - \mu) f\|^2 \right) \quad &= \quad 2 \left\langle f_t, \widetilde{L} f_t \right\rangle \\ &= \quad 2 \left\langle f_t, L f_t \right\rangle \end{aligned}$$

A spectral argument

In continuous time,

$$\mathbb{E}_x\left(f(X_t)\right) - \int f \mathrm{d}\mu \quad = \quad (e^{tL} - \mu) f .$$

Reversible case $\Rightarrow$ L self-adjoint in $L^2(\mu)$ $\Rightarrow$ ON eigenbasis, and

$$\|e^{tL} - \mu\|_{L^2(\mu)} \quad = \quad \sup_{f \neq 0} \frac{\|(e^{tL} - \mu) f\|}{\|f\|} \quad = \quad e^{-t\lambda_1}$$

with $\lambda_1 = \min \sigma(-L) \setminus \{0\}$. Setting $\widetilde{L} = L + L'$ with L' anti-adjoint,

$$\begin{aligned} \partial_t \left(\|(e^{t\widetilde{L}} - \mu) f\|^2 \right) \quad &= \quad 2 \left\langle f_t, \widetilde{L} f_t \right\rangle \\ &= \quad 2 \left\langle f_t, L f_t \right\rangle \\ &\leqslant \quad -2\lambda_1 \|f_t\|^2 . \end{aligned}$$

A spectral argument

In continuous time,

$$\mathbb{E}_x\left(f(X_t)\right) - \int f \mathrm{d}\mu \quad = \quad (e^{tL} - \mu) f.$$

Reversible case $\Rightarrow$ L self-adjoint in $L^2(\mu) \Rightarrow$ ON eigenbasis, and

$$\|e^{tL} - \mu\|_{L^2(\mu)} \quad = \quad \sup_{f \neq 0} \frac{\|(e^{tL} - \mu) f\|}{\|f\|} \quad = \quad e^{-t\lambda_1}$$

with $\lambda_1 = \min \sigma(-L) \setminus \{0\}$. Setting $\widetilde{L} = L + L'$ with L' anti-adjoint,

$$\begin{aligned} \partial_t \left(\|(e^{t\widetilde{L}} - \mu) f\|^2 \right) \quad &= \quad 2 \left\langle f_t, \widetilde{L} f_t \right\rangle \\ &= \quad 2 \left\langle f_t, L f_t \right\rangle \\ &\leqslant \quad -2\lambda_1 \|f_t\|^2. \end{aligned}$$

$$\Rightarrow \|e^{t\widetilde{L}} - \mu\| \quad \leqslant \quad e^{-\lambda_1 t}.$$

Advantages of reversibility

- Metropolis-Hastings : given a kernel q, define a μ-reversible kernel, by accepting a new proposal y of law $q(x,y)$ with probability

$$1 \wedge \frac{\mu(y)q(y,x)}{\mu(x)q(x,y)}.$$

- Theoric study : spectral tools, functional inequalities, ellipticity for diffusions...

1. Reversible sampling

2. Lifted chains and kinetic jump processes

3. Non-reversible diffusions

An order 2 chain

Diaconis et al. (2000, 2009): for the uniform law on $\{1, \dots, N\}$,

$$\begin{aligned} \mathbb{P}\left(X_{n+1} - X_n = X_n - X_{n-1}\right) &= \frac{1+\alpha}{2} \\ \mathbb{P}\big(X_{n+1} - X_n = -\left(X_n - X_{n-1}\right)\big) &= \frac{1-\alpha}{2} \end{aligned}$$

An order 2 chain

Diaconis et al. (2000, 2009): for the uniform law on $\{1,\ldots,N\}$,

$$\begin{aligned}\mathbb{P}\left(X_{n+1}-X_n = X_n - X_{n-1}\right) &= \frac{1+\alpha}{2}\\ \mathbb{P}\big(X_{n+1}-X_n = -\underbrace{(X_n - X_{n-1})}_{Y_n}\big) &= \frac{1-\alpha}{2}\end{aligned}$$

Alone, $(X_n)_{n\geq 0}$ not Markov, but (X_n, X_{n-1}) is, or (X_n, Y_n).

$$\begin{aligned}\mathbb{P}\left(Y_{n+1} = Y_n\right) &= \frac{1+\alpha}{2}\\ \mathbb{P}\left(Y_{n+1} = -Y_n\right) &= \frac{1-\alpha}{2}\\ X_{n+1} &= X_n + Y_{n+1}\end{aligned}$$

An order 2 chain

Diaconis et al. (2000, 2009): for the uniform law on $\{1, \ldots, N\}$,

$$\begin{aligned} \mathbb{P}\left(X_{n+1} - X_n = X_n - X_{n-1}\right) &= \frac{1+\alpha}{2} \\ \mathbb{P}(X_{n+1} - X_n = -\underbrace{(X_n - X_{n-1})}_{Y_n}) &= \frac{1-\alpha}{2} \end{aligned}$$

Alone, $(X_n)_{n\geq 0}$ not Markov, but (X_n, X_{n-1}) is, or (X_n, Y_n).

$$\begin{aligned} \mathbb{P}\left(Y_{n+1} = Y_n\right) &= \frac{1+\alpha}{2} \\ \mathbb{P}\left(Y_{n+1} = -Y_n\right) &= \frac{1-\alpha}{2} \\ X_{n+1} &= X_n + Y_{n+1} \end{aligned}$$

Reversible walk : $\alpha = 0$. Optimal speed for $\alpha = \alpha_{opt} > 0$.

Spectral study, nevertheless

The spectrum is no more real. If Q is the transition matrix,

$$\begin{aligned}\lambda_1 &:= 1 - \sup\{\Re(\nu),\ \nu \in \sigma(Q) \setminus \{1\}\} \\ &= -\sup\{\Re(\nu),\ \nu \in \sigma(Q - I) \setminus \{0\}\}\end{aligned}$$

For $\alpha_{opt} = \frac{1-\sin\left(\frac{\pi}{N}\right)}{1+\sin\left(\frac{\pi}{N}\right)}$

$$\lambda_1 = 1 - \sqrt{\alpha_{opt}} \simeq \frac{\pi}{2N}$$

Spectral study, nevertheless

The spectrum is no more real. If Q is the transition matrix,

$$\begin{aligned}\lambda_1 &:= 1-\sup\{\Re(\nu),\ \nu\in\sigma(Q)\setminus\{1\}\}\\ &= -\sup\{\Re(\nu),\ \nu\in\sigma(Q-I)\setminus\{0\}\}\end{aligned}$$

For $\alpha_{opt} = \frac{1-\sin\left(\frac{\pi}{N}\right)}{1+\sin\left(\frac{\pi}{N}\right)}$

$$\lambda_1 = 1-\sqrt{\alpha_{opt}} \simeq \frac{\pi}{2N}$$

For the symetric walk,

$$\lambda_1 = 1-\cos\frac{\pi}{N} \simeq \frac{\pi^2}{2N^2}$$

To mix, $\mathcal{O}(N^2)$ steps were needed, now only $\mathcal{O}(N)$ (N.B. : the deterministic computation of an in integral is N step).

Scaling limit

Limit $N \to \infty$, with a rate of order N and $\frac{1-\alpha}{2}$ of order $\frac{1}{N}$:

- (X, Y) a Markov process, where $X \in \mathbb{T}$ and $Y = \pm 1$
- $dX_t = Y_t dt$ (kinetic process)
- Y jumps to $-Y$ with rate $a > 0$ (piecewise deterministic)

Scaling limit

Limit $N \to \infty$, with a rate of order N and $\frac{1-\alpha}{2}$ of order $\frac{1}{N}$:

- (X, Y) a Markov process, where $X \in \mathbb{T}$ and $Y = \pm 1$
- $\mathrm{d}X_t = Y_t \mathrm{d}t$ (kinetic process)
- Y jumps to $-Y$ with rate $a > 0$ (piecewise deterministic)

Uniform equilibrium μ, generator

$$Lf(x,y) \;=\; y\partial_x f(x,y) + a\left(f(x,-y) - f(x,y)\right).$$

Again, spectral study, for instance for $a_{opt} = 1$:

$$\|P_t - \mu\| \;=\; e^{-t}\sqrt{1 + \frac{2}{\sqrt{1+\frac{1}{t^2}} - 1}}$$

Scaling limit

Limit $N \to \infty$, with a rate of order N and $\frac{1-\alpha}{2}$ of order $\frac{1}{N}$:

- (X, Y) a Markov process, where $X \in \mathbb{T}$ and $Y = \pm 1$
- $\mathrm{d}X_t = Y_t \mathrm{d}t$ (kinetic process)
- Y jumps to $-Y$ with rate $a > 0$ (piecewise deterministic)

Uniform equilibrium μ, generator

$$Lf(x,y) \;=\; y\partial_x f(x,y) + a\left(f(x,-y) - f(x,y)\right).$$

Again, spectral study, for instance for $a_{opt} = 1$:

$$\|P_t - \mu\| \;=\; e^{-t}\sqrt{1 + \frac{2}{\sqrt{1+\frac{1}{t^2}} - 1}} \;\underset{t\to 0}{\simeq}\; 1 - \frac{t^3}{3}$$

Scaling limit

Limit $N \to \infty$, with a rate of order N and $\frac{1-\alpha}{2}$ of order $\frac{1}{N}$:

- (X, Y) a Markov process, where $X \in \mathbb{T}$ and $Y = \pm 1$
- $dX_t = Y_t dt$ (kinetic process)
- Y jumps to $-Y$ with rate $a > 0$ (piecewise deterministic)

Uniform equilibrium μ, generator

$$Lf(x,y) = y\partial_x f(x,y) + a\left(f(x,-y) - f(x,y)\right).$$

Again, spectral study, for instance for $a_{opt} = 1$:

$$\|P_t - \mu\| = e^{-t}\sqrt{1 + \frac{2}{\sqrt{1+\frac{1}{t^2}} - 1}} \underset{t\to\infty}{\simeq} 2te^{-t}$$

Piecewise deterministic MCMC

Kinetic jump processes:

- (X, Y) Markov on $\mathbb{R}^d \times E$ with $E \subset \mathbb{R}^d$.
- $\mathrm{d}X = Y\mathrm{d}t$.
- Y piecewise constant, jumps at random times.
- At equilibrium, $X \sim \mu$.

Piecewise deterministic MCMC

Kinetic jump processes:

- (X, Y) Markov on $\mathbb{R}^d \times E$ with $E \subset \mathbb{R}^d$.
- $\mathrm{d}X = Y\mathrm{d}t$.
- Y piecewise constant, jumps at random times.
- At equilibrium, $X \sim \mu$.

Two things to chose:

- The jump rate $\lambda(x, y)$, which defines the next jump time by

$$T \quad = \quad \inf\left\{ t \geqslant 0,\ E \leqslant \int_0^t \lambda(X_s, Y_s)\mathrm{d}s \right\}, \qquad \text{où } E \sim \mathcal{E}(1).$$

- The jump kernel $Q(x, y)$, so that $Y_T \sim Q(X_T, Y_{T-})$.

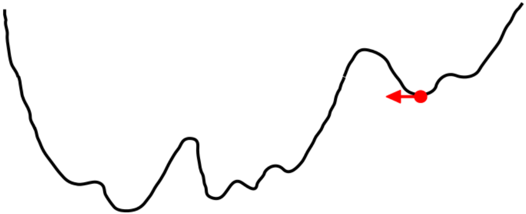
Recent years (re)discovery

- Peters, de With (2012, *Rejection-free Metropolis Hastings*, $y \in \mathbb{R}^d$ gaussian; *event-driven MC* in physic litterature, Michel, Kapfer, Krauth 2013 by ex.), Bouchard-Côté, Vollmer, Doucet (2016, 2017, *bouncy particle sampler*)
- Fontbona, Guérin, Malrieu (2012, 2016, *integrated telegraph process*)
- Calvez, Raoul, Schmeiser (2016, *run-and-tumble process*, bacterium chemotaxis, non-explicit equilibrium, $y \in [-1, 1]$).
- Bierkens, Fearnhead, Roberts (2016, *Zig-zag process*, $y \in \{-1, +1\}^d$)
- Miclo, M. (2012, *Volte-Face*, 2013, 2016)

The bouncy particle sampler

Jump rate $\lambda(x,y) = (y \cdot \nabla_x U(x))_+$; since $y = x'$,

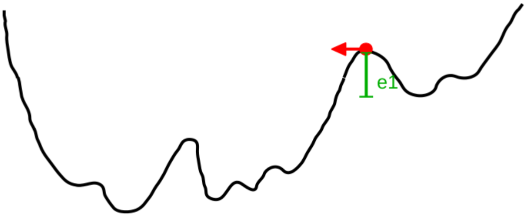
$$\begin{aligned} \int_0^t \lambda(X_s, Y_s) \mathrm{d}s &= U(X_t) - U(X_0) & \text{when going up} \\ &= 0 & \text{when going down} \end{aligned}$$



The bouncy particle sampler

Jump rate $\lambda(x,y) = (y \cdot \nabla_x U(x))_+$; since $y = x'$,

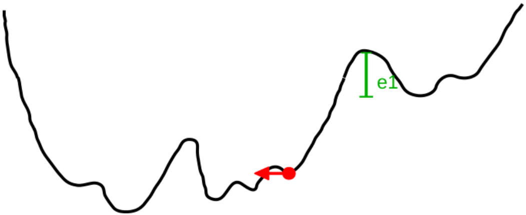
$$\begin{aligned} \int_0^t \lambda(X_s, Y_s) ds &= U(X_t) - U(X_0) \quad \text{when going up} \\ &= 0 \quad \text{when going down} \end{aligned}$$



The bouncy particle sampler

Jump rate $\lambda(x,y) = (y \cdot \nabla_x U(x))_+$; since $y = x'$,

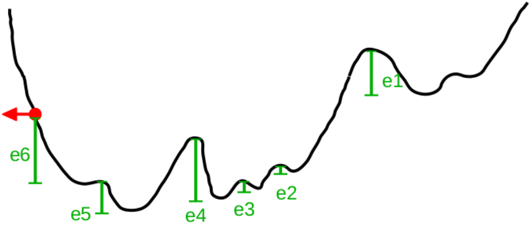
$$\begin{aligned} \int_0^t \lambda(X_s, Y_s) ds &= U(X_t) - U(X_0) & \text{when going up} \\ &= 0 & \text{when going down} \end{aligned}$$



The bouncy particle sampler

Jump rate $\lambda(x,y) = (y \cdot \nabla_x U(x))_+$; since $y = x'$,

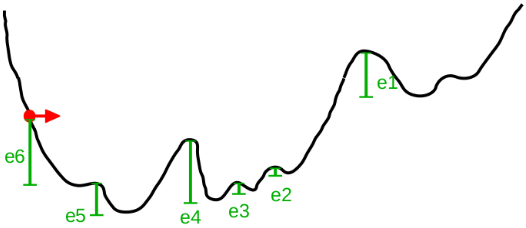
$$\int_0^t \lambda(X_s, Y_s) ds \quad = \quad U(X_t) - U(X_0) \qquad \text{when going up}$$
$$= \quad 0 \qquad \text{when going down}$$



The bouncy particle sampler

Jump rate $\lambda(x,y) = (y \cdot \nabla_x U(x))_+$; since $y = x'$,

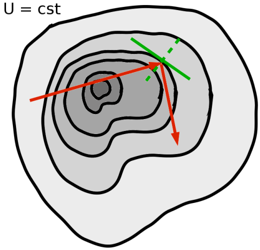
$$\begin{aligned} \int_0^t \lambda(X_s, Y_s) ds &= U(X_t) - U(X_0) && \text{when going up} \\ &= 0 && \text{when going down} \end{aligned}$$



The Bouncy Particle Sampler

Jump kernel $Q(x,y) = \delta_{y^*}$ with

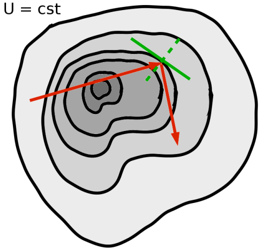
$$y_* = y - 2\frac{y \cdot \nabla U(x)}{|\nabla U(x)|^2}\nabla U(x).$$



The Bouncy Particle Sampler

Jump kernel $Q(x,y) = \delta_{y^*}$ with

$$y_* = y - 2\frac{y \cdot \nabla U(x)}{|\nabla U(x)|^2}\nabla U(x).$$

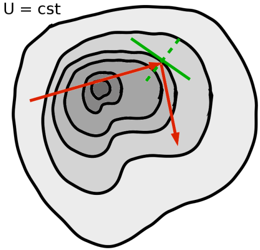


Not necessarily ergodic

The Bouncy Particle Sampler

Jump kernel $Q(x, y) = \delta_{y^*}$ with

$$y_* = y - 2\frac{y \cdot \nabla U(x)}{|\nabla U(x)|^2}\nabla U(x).$$



Not necessarily ergodic $\Rightarrow$ velocity refreshment at constant rate.

The bouncy particle sampler

Finally,

$$\begin{aligned} Lf(x,y) &= y\nabla_x f(x,y) + (y\cdot\nabla U(x))_+ \left(f(x,y_*) - f(x,y)\right) \\ &\quad + r\left(\int_{\mathbb{S}^{d-1}} f(x,z)\mathrm{d}z - f(x,y)\right) \end{aligned}$$

The bouncy particle sampler

Finally,

$$\begin{aligned} Lf(x,y) &= y\nabla_x f(x,y) + (y\cdot\nabla U(x))_+ (f(x,y_*) - f(x,y)) \\ &\quad + r\left(\int_{\mathbb{S}^{d-1}} f(x,z)\mathrm{d}z - f(x,y)\right) \end{aligned}$$

- Ergodic with equilibrium $\mu = e^{-U}\mathrm{d}x\mathrm{d}y$.
- Non-reversible, kinetic (y = inertia = short-term memory)

The bouncy particle sampler

Finally,

$$
\begin{aligned}
Lf(x,y) &= y\nabla_x f(x,y) + (y\cdot\nabla U(x))_+ \left(f(x,y_*) - f(x,y)\right) \\
&\quad + r\left(\int_{\mathbb{S}^{d-1}} f(x,z)\mathrm{d}z - f(x,y)\right)
\end{aligned}
$$

- Ergodic with equilibrium $\mu = e^{-U}\mathrm{d}x\mathrm{d}y$.
- Non-reversible, kinetic (y = inertia = short-term memory)
- Piecewise deterministic Markov process (PDMP) : exat simulation (thining).

The bouncy particle sampler

Finally,

$$\begin{aligned} Lf(x,y) &= y\nabla_x f(x,y) + (y\cdot\nabla U(x))_+ \left(f(x,y_*) - f(x,y)\right) \\ &\quad + r\left(\int_{\mathbb{S}^{d-1}} f(x,z)\mathrm{d}z - f(x,y)\right) \end{aligned}$$

- Ergodic with equilibrium $\mu = e^{-U}\mathrm{d}x\mathrm{d}y$.
- Non-reversible, kinetic (y = inertia = short-term memory)
- Piecewise deterministic Markov process (PDMP) : exat simulation (thining).
- ("Physical", trajectoral reversibility)

Some recent questions

- Empirical studies (choice of the jump rate, of the law of the velocity, of the deterministic flow).
- Adapting existing methods (subsampling, control variates. . .).
- Irreducibility without refreshment (for the Zig-Zag).
- Geometric ergodicity and CLT.
- Diffusive limits.

Some recent questions

- Empirical studies (choice of the jump rate, of the law of the velocity, of the deterministic flow).
- Adapting existing methods (subsampling, control variates. . .).
- Irreducibility without refreshment (for the Zig-Zag).
- Geometric ergodicity and CLT.
- Diffusive limits.

Not easy to compare different dynamics (dimension 1. . .)

Metastability

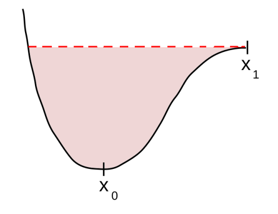
Replace U by $\frac{1}{\varepsilon}U$.

Theorem (Eyring-Kramers formula, M. 2016)

In dimension 1, let $\tau = \inf\{s > 0,\ X_s = x_1\}$. *Then*

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

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



Metastability

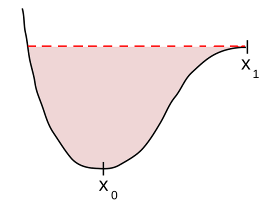
Replace U by $\frac{1}{\varepsilon}U$.

Theorem (Eyring-Kramers formula, M. 2016)

In dimension 1, let $\tau = \inf\{s > 0,\ X_s = x_1\}$. *Then*

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

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



With a temperature scheme $(\varepsilon_t)_{t\geqslant 0}$, NSC for annealing (same as reversible) ; SC in dimension d (same).

1 Reversible sampling

2 Lifted chains and kinetic jump processes

3 Non-reversible diffusions

Non-reversible diffusions

- Fokker-Planck (or overdamped Langevin) diffusion, reversible :

$$dX_t = -\nabla U(X_t) + \sqrt{2} dB_t,$$

with invariant measure e^{-U}.

Non-reversible diffusions

- Fokker-Planck (or overdamped Langevin) diffusion, reversible :

$$dX_t = -\nabla U(X_t) + \sqrt{2}dB_t,$$

with invariant measure e^{-U}.

- Still the case, if $\nabla \cdot \left(be^{-U}\right) = 0$ (ex: $b = A\nabla U$), for

$$dX_t = -\nabla U(X_t) + b(X_t)dt + \sqrt{2}dB_t$$

(Hwang, Hwang-Ma, Sheu 05) and improved spectral gap.

Non-reversible diffusions

- Fokker-Planck (or overdamped Langevin) diffusion, reversible :

$$dX_t = -\nabla U(X_t) + \sqrt{2}dB_t,$$

with invariant measure e^{-U}.

- Still the case, if $\nabla \cdot \left(be^{-U}\right) = 0$ (ex: $b = A\nabla U$), for

$$dX_t = -\nabla U(X_t) + b(X_t)dt + \sqrt{2}dB_t$$

(Hwang, Hwang-Ma, Sheu 05) and improved spectral gap.

- Kinetic Langevin diffusion (inertia),

$$\begin{aligned} dX_t &= Y_t \\ dY_t &= -\nabla U(X_t) - \gamma Y_t + \sqrt{2\gamma}dB_t, \end{aligned}$$

with equilibrium e^{-H}, $H(x,y) = U(x) + \frac{1}{2}|y|^2$.

Some questions

- Empirical studies.
- Scaling limit (large dimension) of a Metropolized discretization.
- Non-reversible discretization.
- Metastability and annealing.

Some questions

- Empirical studies.
- Scaling limit (large dimension) of a Metropolized discretization.
- Non-reversible discretization.
- Metastability and annealing.

Again, hard to compare dynamics and to tune parameters.

The Gaussian case

For A a matrix and D a positive one, let

$$\mathrm{d}X_t = AX_t\mathrm{d}t + \sqrt{2D}\mathrm{d}B_t.$$

The Gaussian case

For A a matrix and D a positive one, let

$$dX_t = AX_t dt + \sqrt{2D} dB_t.$$

Question:

- A target equilibrium $\mu(x) \propto \exp\left(-\frac{1}{2}x \cdot Sx\right)$ being fixed,
- The amount of randomness $\mathrm{Tr}(D) = \mathrm{Tr}(I) = d$ being fixed (Gadat, Miclo 2012),

$\Rightarrow$ Find A and D which maximizes the speed of $\mathcal{L}(X_t) \to \mu$.

Explicit hypocoercive speed

Set :

- $\rho \;=\; \rho(A) \;=\; -\sup\{\Re(\nu),\ \nu \in \sigma(A)\}$
- N the dimension of the largest Jordan block of A with $\{\Re(\nu) = -\rho\}$
- M the number of Lie brackets necessary to satisfy Hörmander's condition, equivalent here to

$$\sum_{k=0}^{M} A^k D (A^T)^k \;>\; 0.$$

Explicit hypocoercive speed

Set :

- $\rho \;=\; \rho(A) \;=\; -\sup\{\Re(\nu),\ \nu \in \sigma(A)\}$
- N the dimension of the largest Jordan block of A with $\{\Re(\nu) = -\rho\}$
- M the number of Lie brackets necessary to satisfy Hörmander's condition, equivalent here to

$$\sum_{k=0}^{M} A^k D (A^T)^k \;>\; 0.$$

Theorem (Arnold, Erb 2014, M. 2015)

There exist $c, \kappa > 0$ *such that*

$$\frac{1}{c}(1 + t^{2(N-1)})e^{-2\rho t} \;\leq\; \|e^{tL} - \mu\|^2 \;\leq\; c(1 + t^{2(N-1)})e^{-2\rho t}$$

and

$$\|e^{tL} - \mu\|^2 \;\leq\; e^{-\kappa t (1-e^{-t})^{2M}} \underset{t \to 0}{\simeq} 1 - \kappa t^{2M+1}$$

Hypoelliptic diffusion

For a covariance matrix S, denote

$$\mathcal{I}(S) \ = \{(A,D), AS^{-1} + S^{-1}A^T \ = \ -2D \text{ et } \mathrm{Tr} D \leqslant d\}.$$

Hypoelliptic diffusion

For a covariance matrix S, denote

$$\mathcal{I}(S) \;= \{(A,D), AS^{-1}+S^{-1}A^T \;=\; -2D \text{ et } \mathrm{Tr}D \leqslant d\}.$$

Theorem (Lelièvre, Nier, Pavliotis, 2012 ; Guillin, M. 2016)

$$\begin{aligned}
\rho(-S) &= \min \sigma(S)\\
\inf_{\mathcal{I}(S),D=I}\rho(A) &= \frac{TrS}{N}\\
\inf_{\mathcal{I}(S),reversible}\rho(A) &= \frac{N}{TrS^{-1}}\\
\inf_{\mathcal{I}(S)}\rho(A) &= \max \sigma(S)
\end{aligned}$$

Hypoelliptic diffusion

For a covariance matrix S, denote

$$\mathcal{I}(S) \ = \{(A,D), AS^{-1} + S^{-1}A^T \ = \ -2D \text{ et } \mathrm{Tr} D \leqslant d\}.$$

Theorem (Lelièvre, Nier, Pavliotis, 2012 ; Guillin, M. 2016)

$$\begin{aligned}
\rho(-S) &= \min \sigma(S) \\
\inf_{\mathcal{I}(S), D=I} \rho(A) &= \frac{TrS}{N} \\
\inf_{\mathcal{I}(S), reversible} \rho(A) &= \frac{N}{TrS^{-1}} \\
\inf_{\mathcal{I}(S)} \rho(A) &= \max \sigma(S)
\end{aligned}$$

No improvement when S is an homothety.

Highly degenerated diffusion

The Brownian noise is concentrated on a single coordinate: slow regularization,

$$\|e^{tL_{opt}} - \mu\|^2 \quad \leqslant \quad Ce^{-\max \sigma(S)t}$$

with an horrible C.

Highly degenerated diffusion

The Brownian noise is concentrated on a single coordinate: slow regularization,

$$\|e^{tL_{opt}} - \mu\|^2 \quad \leqslant \quad Ce^{-\max\sigma(S)t}$$

with an horrible C.

Theorem (Guillin, M. 2016)

One can construct $(A, D) \in \mathcal{I}(S)$ *with* $\sqrt{Tr(A^T A)} \leqslant 4d^2\sqrt{\frac{\max\sigma(S)^3}{\min\sigma(S)}}$ *and*

$$\|e^{(t-t_0)L_{A,D}}e^{t_0 L_{-S,I}} - \mu\| \;\leqslant\; \frac{1}{t_0 \min\sigma(S)} e^{-(t-t_0)\max\sigma(S)}.$$

The bound is optimal for $t_0^{-1} = \max\sigma(S)$, *which yields*

$$\|e^{(t-t_0)L_{A,D}}e^{t_0 L_{-S,I}} - \mu\| \;\leqslant\; \frac{\max\sigma(S)}{\min\sigma(S)} e^{1-t\max\sigma(S)}$$

The kinetic case

$$\begin{cases} \mathrm{d}X_t &= Y_t \mathrm{d}t \\ \mathrm{d}Y_t &= -\nu S X_t \mathrm{d}t - \frac{1}{\nu} Y_t \mathrm{d}t + \sqrt{2} \mathrm{d}B_t \end{cases}$$

Choice for the variance ν ? If λ is an eigenvalue of S,

$$r = \frac{1}{2\nu}\left(1 \pm \sqrt{1 - 4\lambda\nu^3}\right)$$

are eigenvalues of $A = \begin{pmatrix} 0 & I \\ -\nu S & -\frac{1}{\nu} I \end{pmatrix}$

The kinetic case

$$\begin{cases} dX_t &= Y_t dt \\ dY_t &= -\nu S X_t dt - \frac{1}{\nu} Y_t dt + \sqrt{2} dB_t \end{cases}$$

Choice for the variance ν ? If λ is an eigenvalue of S,

$$r = \frac{1}{2\nu}\left(1 \pm \sqrt{1 - 4\lambda\nu^3}\right)$$

are eigenvalues of $A = \begin{pmatrix} 0 & I \\ -\nu S & -\frac{1}{\nu} I \end{pmatrix}$

$$\Rightarrow \rho(A) \underset{\nu \to \infty}{\simeq} \frac{1}{2\nu} \quad \text{and} \quad \rho(A) \underset{\nu \to 0}{\simeq} \nu^2 \min \sigma(S)$$

The kinetic case

$$\left\{ \begin{array}{lcl} dX_t & = & Y_t dt \\ dY_t & = & -\nu S X_t dt - \frac{1}{\nu} Y_t dt + \sqrt{2} dB_t \end{array} \right.$$

Choice for the variance ν ? If λ is an eigenvalue of S,

$$r = \frac{1}{2\nu}\left(1 \pm \sqrt{1 - 4\lambda\nu^3}\right)$$

are eigenvalues of $A = \begin{pmatrix} 0 & I \\ -\nu S & -\frac{1}{\nu} I \end{pmatrix}$

$$\Rightarrow \rho(A) \underset{\nu \to \infty}{\simeq} \frac{1}{2\nu} \quad \text{and} \quad \rho(A) \underset{\nu \to 0}{\simeq} \nu^2 \min \sigma(S)$$

The truth lies in the middle... If $S = \lambda I$, the optimal rate is $(\lambda/2)^{\frac{1}{3}}$, and

$$(\lambda/2)^{\frac{1}{3}} > \lambda \quad \Leftrightarrow \quad \lambda < \frac{1}{\sqrt{2}} \simeq 0.7$$

Conclusions

Non-reversible sampling :

- Not exhaustive presentation.
- Empirically, it seems to work, sometimes.
- Theoretically, we are sometimes able to prove that it is not less efficient than reversible processes.
- Not always theoretical means to compare dynamics or tune parameters
- . . . except for some toy models (uniform law, Gaussian, dimension 1)

Conclusions

Non-reversible sampling :

- Not exhaustive presentation.
- Empirically, it seems to work, sometimes.
- Theoretically, we are sometimes able to prove that it is not less efficient than reversible processes.
- Not always theoretical means to compare dynamics or tune parameters
- ... except for some toy models (uniform law, Gaussian, dimension 1)
- On the other hand, we hadn't better result for reversible processes, to start with.

Conclusions

Non-reversible sampling :

- Not exhaustive presentation.
- Empirically, it seems to work, sometimes.
- Theoretically, we are sometimes able to prove that it is not less efficient than reversible processes.
- Not always theoretical means to compare dynamics or tune parameters
- ... except for some toy models (uniform law, Gaussian, dimension 1)
- On the other hand, we hadn't better result for reversible processes, to start with.
- Inertia and short-term memory: compatible with long-term memory (non Markov) or global (particles) strategies.

Conclusions

Non-reversible sampling :

- Not exhaustive presentation.
- Empirically, it seems to work, sometimes.
- Theoretically, we are sometimes able to prove that it is not less efficient than reversible processes.
- Not always theoretical means to compare dynamics or tune parameters
- ... except for some toy models (uniform law, Gaussian, dimension 1)
- On the other hand, we hadn't better result for reversible processes, to start with.
- Inertia and short-term memory: compatible with long-term memory (non Markov) or global (particles) strategies.

Thanks for your attention!