Piecewise deterministic sampling and annealing

Pierre Monmarché

CERMICS - École des Ponts Paritech and INRIA Paris

Seminar of Statistics, Oxford

1 Introduction

2 Persistent walk and kinetic PDMP

3 Some results

1 Introduction

2 Persistent walk and kinetic PDMP

3 Some results

MCMC algorithms

- Problem: estimate

$$\mathbb{E}\left(f(X)\right) \;=\; \frac{\int f(x)e^{-\frac{1}{\varepsilon}U(x)}\mathrm{d}x}{\int e^{-\frac{1}{\varepsilon}U(x)}\mathrm{d}x} \;=\; \int f\mathrm{d}\mu$$

with

- $x \in \mathbb{R}^d$ the microscopic configuration/parameters, d large
- U the potential/log-likelihood
- $\varepsilon > 0$ the temperature
- $\mu \propto e^{-\frac{1}{\varepsilon}U(x)}\mathrm{d}x$ the target measure
- f an observable/test function.

MCMC algorithms

- Problem: estimate

$$\mathbb{E}\left(f(X)\right) = \frac{\int f(x)e^{-\frac{1}{\varepsilon}U(x)}\mathrm{d}x}{\int e^{-\frac{1}{\varepsilon}U(x)}\mathrm{d}x} = \int f\mathrm{d}\mu$$

  with
  - $x \in \mathbb{R}^d$ the microscopic configuration/parameters, d large
  - U the potential/log-likelihood
  - $\varepsilon > 0$ the temperature
  - $\mu \propto e^{-\frac{1}{\varepsilon}U(x)}\mathrm{d}x$ the target measure
  - f an observable/test function.
- MCMC basic ingredient: an ergodic (Markov) process $(X_t)_{t\geqslant 0}$, i.e.

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

Many available dynamics

- (reversible) overdamped Langevin diffusion:

$$\mathrm{d}X_t = -\nabla U(X_t)\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}B_t,$$

- kinetic Langevin equation:

$$\begin{aligned} \mathrm{d}X_t &= Y_t\mathrm{d}t \\ \mathrm{d}Y_t &= -\nabla U(X_t)\mathrm{d}t - Y_t\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}B_t, \end{aligned}$$

- Metropolis-Hastings algorithm (propose, accept/reject),
- Hamiltonian Monte-Carlo.

How to compare them ?

Some criteria:

- asymptotic variance in a Central Limit Theorem

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

- Relaxation speed toward equilibrium

$$\mathcal{L}\left(X_t \mid X_0 \sim \nu\right) \underset{t\to\infty}{\longrightarrow} \mu.$$

- mixing (decorrelation) time
- scaling of the chain (diffusive, ballistic. . .)

How to compare them ?

Some criteria:

- asymptotic variance in a Central Limit Theorem

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

- Relaxation speed toward equilibrium

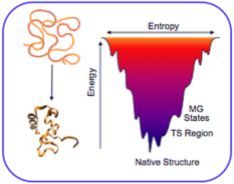
$$\mathcal{L}\left(X_t \mid X_0 \sim \nu\right) \underset{t\to\infty}{\longrightarrow} \mu.$$

- mixing (decorrelation) time
- scaling of the chain (diffusive, ballistic. . .)

Different kind of results:

- empirical results
- theoretical results for toy models ($d = 1$, Gaussian or uniform laws)
- asymptotic theoretical results (more or less impossible to compare)

Stochastic optimization: the simulated annealing algorithm



Problem : minimize a function

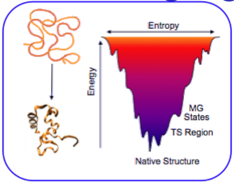
- in large dimension (or large finite set),
- with many local minima.

The gradient descent

$$dX_t = -\nabla U(X_t)dt$$

ends up in a local minima.

Stochastic optimization: the simulated annealing algorithm



Problem : minimize a function

- in large dimension (or large finite set),
- with many local minima.

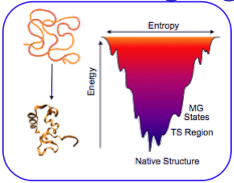
The overdamped Langevin diffusion

$$\mathrm{d}X_t = -\nabla U(X_t)\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}B_t$$

will eventually escape from any local minima.

$$\mathcal{L}(X_t) \underset{t\to\infty}{\longrightarrow} e^{-\frac{U(x)}{\varepsilon}}dx$$

Stochastic optimization: the simulated annealing algorithm



Problem : minimize a function

- in large dimension (or large finite set),
- with many local minima.

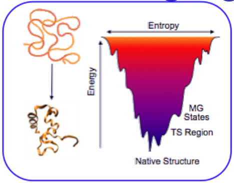
The overdamped Langevin diffusion

$$\mathrm{d}X_t = -\nabla U(X_t)\mathrm{d}t \; + \; \sqrt{2\varepsilon}\mathrm{d}B_t$$

will eventually escape from any local minima.

$$\mathcal{L}(X_t) \quad \underset{t\to\infty}{\longrightarrow} \quad e^{-\frac{U(x)}{\varepsilon}}dx \quad \underset{\varepsilon\to 0}{\longrightarrow} \quad \delta_{argminU}.$$

Stochastic optimization: the simulated annealing algorithm



Problem : minimize a function

- in large dimension (or large finite set),
- with many local minima.

The overdamped Langevin diffusion

$$\mathrm{d}X_t = -\nabla U(X_t)\mathrm{d}t \ + \ \sqrt{2\varepsilon}\mathrm{d}B_t$$

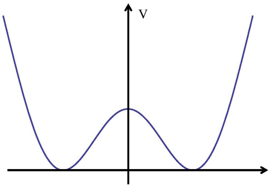
will eventually escape from any local minima.

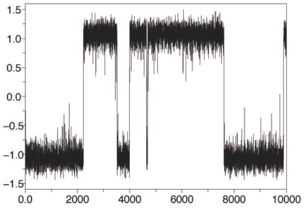
$$\mathcal{L}(X_t) \quad \underset{t\to\infty}{\longrightarrow} \quad e^{-\frac{U(x)}{\varepsilon}}dx \quad \underset{\varepsilon\to 0}{\longrightarrow} \quad \delta_{argminU}.$$

How fast can one cool down the system ?

$\Leftrightarrow$ *At a given temperature, how fast is the convergence to equilibirum ?*

Metastability





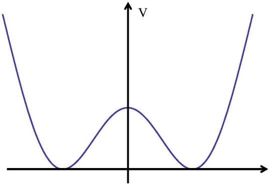
For the overdamped Langevin diffusion:

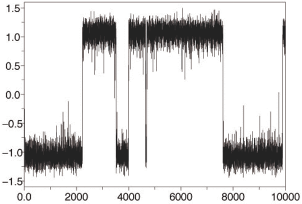
$$\mathrm{d}X_t \;=\; -\nabla U(X_t)\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}B_t,$$

at small temperature $\varepsilon \to 0$:

$$\begin{array}{ccc} \text{escape time from minima} & \simeq & e^{\frac{1}{\varepsilon}\Delta U} \\ \text{relaxation rate to equilibrium} & \simeq & e^{-\frac{1}{\varepsilon}E} \\ \text{condition on the cooling schedule } \varepsilon_t & \gtrsim & \frac{E}{\ln t}. \end{array}$$

Metastability





For the overdamped Langevin diffusion:

$$\mathrm{d}X_t \;=\; -\nabla U(X_t)\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}B_t,$$

at small temperature $\varepsilon \to 0$:

$$\begin{array}{ccc} \text{escape time from minima} & \simeq & e^{\frac{1}{\varepsilon}\Delta U} \\ \text{relaxation rate to equilibrium} & \simeq & e^{-\frac{1}{\varepsilon}E} \\ \text{condition on the cooling schedule } \varepsilon_t & \gtrsim & \frac{E}{\ln t}. \end{array}$$

Morally, all Markov processes with local moves (i.e. continuous trajectories or small jumps) and a finite mean speed should scale the same.

We should be less Markov

- Ergodicity: the time spent between two saddle point crossing should be equal to the ratio of the probabilities.

We should be less Markov

- Ergodicity: the time spent between two saddle point crossing should be equal to the ratio of the probabilities.

- We are exploring a complex space with an amnesic explorer.

We should be less Markov

- Ergodicity: the time spent between two saddle point crossing should be equal to the ratio of the probabilities.
- We are exploring a complex space with an amnesic explorer.

- Ideas: add some global knowledge, some memory.

We should be less Markov

- Ergodicity: the time spent between two saddle point crossing should be equal to the ratio of the probabilities.
- We are exploring a complex space with an amnesic explorer.

- Ideas: add some global knowledge, some memory.
- Problem: high-dimensional memory (or particles) is numerically expensive/unmanageable ($\Rightarrow$ reaction coordinates, collective variables, coarse-grained model).

We should be less Markov

- Ergodicity: the time spent between two saddle point crossing should be equal to the ratio of the probabilities.

- We are exploring a complex space with an amnesic explorer.

- Ideas: add some global knowledge, some memory.

- Problem: high-dimensional memory (or particles) is numerically expensive/unmanageable ($\Rightarrow$ reaction coordinates, collective variables, coarse-grained model).

- Another possibility: only keep an instantaneous memory ($=$ inertia).

1 Introduction

2 Persistent walk and kinetic PDMP

3 Some results

A second order Markov chain: the persistent walk

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

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

A second order Markov chain: the persistent walk

Diaconis et al. (2000, 2009): to sample 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, $(X_n)_{n\geq 0}$ is not Markov, but (X_n, X_{n-1}) is, or (X_n, Y_n).

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

A second order Markov chain: the persistent walk

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

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

Alone, $(X_n)_{n\geq 0}$ is 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 symmetric walk: $\alpha = 0$. Optimal speed for $\alpha = \alpha_{opt} > 0$.

Spectral study

The transition matrix Q is no more symmetric (i.e. no detailed balance); its spectrum may not be real anymore, its eigenvectors are not orthogonal anymore. Nevertheless, explicit computation:

$$\|e^{t(Q-I)} - \mu\|_{\mathcal{L}^2} \quad = \quad C_\alpha(t) e^{-\rho_\alpha t}.$$

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

$$\rho_{\alpha_{opt}} \;=\; 1 - \sqrt{\alpha_{opt}} \;\simeq\; \frac{\pi}{2N}.$$

Spectral study

The transition matrix Q is no more symmetric (i.e. no detailed balance); its spectrum may not be real anymore, its eigenvectors are not orthogonal anymore. Nevertheless, explicit computation:

$$\|e^{t(Q-I)} - \mu\|_{\mathcal{L}^2} \;=\; C_\alpha(t) e^{-\rho_\alpha t}.$$

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

$$\rho_{\alpha_{opt}} \;=\; 1 - \sqrt{\alpha_{opt}} \;\simeq\; \frac{\pi}{2N}.$$

For the symmetric walk,

$$\rho_0 \;=\; 1 - \cos\frac{\pi}{N} \;\simeq\; \frac{\pi^2}{2N^2}.$$

It took $\mathcal{O}(N^2)$ steps to mix, and now only $\mathcal{O}(N)$ (Nota: the determinisitc computation of an integral is done in exactly N steps).

Scaling limit

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

Scaling limit

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

- (X, Y) 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 $-Y$ at rate $a > 0$ (PDMP, piecewise deterministic Markov process)

Scaling limit

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

- (X, Y) 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 $-Y$ at rate $a > 0$ (PDMP, piecewise deterministic Markov process)

Uniform equilibrium μ, and generator

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

Again a spectral study is possible; for instance, for $a_{opt} = 1$,

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

Scaling limit

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

- (X, Y) 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 $-Y$ at rate $a > 0$ (PDMP, piecewise deterministic Markov process)

Uniform equilibrium μ, and generator

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

Again a spectral study is possible; for instance, for $a_{opt} = 1$,

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

Scaling limit

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

- (X, Y) 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 $-Y$ at rate $a > 0$ (PDMP, piecewise deterministic Markov process)

Uniform equilibrium μ, and generator

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

Again a spectral study is possible; for instance, for $a_{opt} = 1$,

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

Remark: $a = 0 \;\Rightarrow\;$ no cv, but $\left|\frac{1}{t}\int_0^t f(x+s)\mathrm{d}s - \int f\mathrm{d}\mu\right| \leqslant \frac{c}{t}$.

With a potential

Specifications:

- (X, Y) Markov on $\mathbb{R} \times \{\pm 1\}$
- $\mathrm{d}X = Y \mathrm{d}t$
- equilibrium $\mu = e^{-U(x)} \mathrm{d}x \mathrm{d}y$

With a potential

Specifications:

- (X, Y) Markov on $\mathbb{R} \times \{\pm 1\}$
- $\mathrm{d}X = Y\mathrm{d}t$
- equilibrium $\mu = e^{-U(x)}\mathrm{d}x\mathrm{d}y$

Only choice: the jump rate. Solution: $x \mapsto a(x) \geqslant 0$ arbitrary,

$$\lambda(x, y) \quad = \quad \left(yU'(x)\right)_+ + a(x).$$

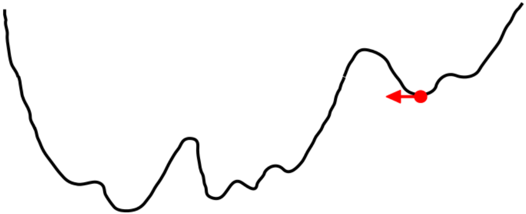
In other words, if E is a standard exponential r.v., next jump at

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

The minimal jump rate

If $a = 0$, $\lambda(x,y) = (yU'(x))_+$; since $y = x'$,

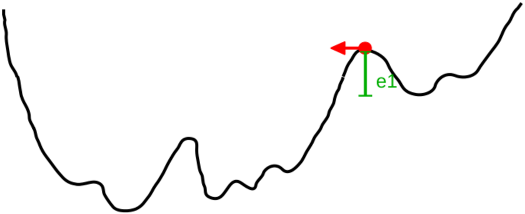
$$\int_0^t \lambda(X_s,Y_s)\mathrm{d}s \quad = \quad U(X_t) - U(X_0) \qquad \text{as long as we climp up}$$
$$= \quad 0 \qquad \text{as long as we fall down.}$$



The minimal jump rate

If $a = 0$, $\lambda(x,y) = (yU'(x))_+$; since $y = x'$,

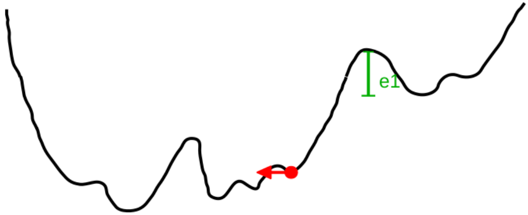
$$\int_0^t \lambda(X_s,Y_s)\mathrm{d}s = U(X_t) - U(X_0) \quad \text{as long as we climp up}$$
$$= 0 \quad \text{as long as we fall down.}$$



The minimal jump rate

If $a = 0$, $\lambda(x,y) = (yU'(x))_+$; since $y = x'$,

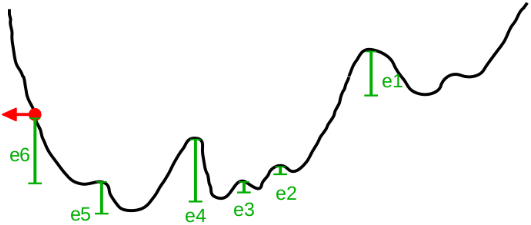
$$\int_0^t \lambda(X_s,Y_s)\mathrm{d}s \;=\; U(X_t) - U(X_0) \quad \text{as long as we climp up}$$
$$\;=\; 0 \quad \text{as long as we fall down.}$$



The minimal jump rate

If $a = 0$, $\lambda(x,y) = (yU'(x))_+$; since $y = x'$,

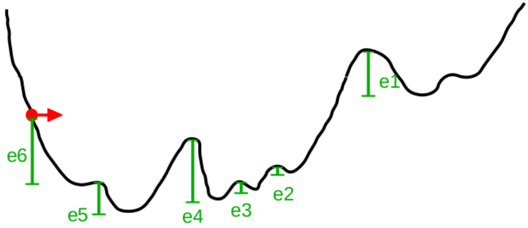
$$\begin{aligned} \int_0^t \lambda(X_s, Y_s)\mathrm{d}s &= U(X_t) - U(X_0) && \text{as long as we climp up} \\ &= 0 && \text{as long as we fall down.} \end{aligned}$$



The minimal jump rate

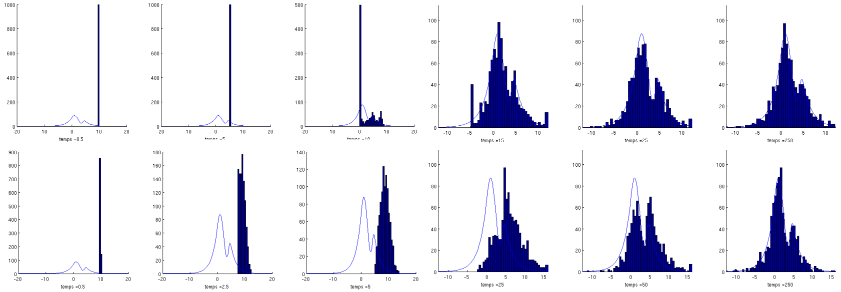
If $a = 0$, $\lambda(x,y) = (yU'(x))_+$; since $y = x'$,

$$\int_0^t \lambda(X_s, Y_s)\mathrm{d}s \quad = \quad U(X_t) - U(X_0) \qquad \text{as long as we climp up}$$
$$= \quad 0 \qquad \text{as long as we fall down.}$$



With a supplementary rate

For $a \neq 0$, it's the same, except that random jumps are added which do not depend on the velocity.



In higher dimension

We want to keep the same rate:

$$\lambda(x,y) \ = \ (y \cdot \nabla U(x))_+ .$$

To target μ, a necessary and sufficient condition is that, at a jump,

$$Y \cdot \nabla U(X) \ \leftarrow \ -Y \cdot \nabla U(X).$$

In higher dimension

We want to keep the same rate:

$$\lambda(x,y) \ = \ (y \cdot \nabla U(x))_+.$$

To target μ, a necessary and sufficient condition is that, at a jump,

$$Y \cdot \nabla U(X) \ \leftarrow \ -Y \cdot \nabla U(X).$$

For instance:

- $Q(x,y,\cdot) = \delta_{-y}$

In higher dimension

We want to keep the same rate:

$$\lambda(x,y) \;=\; (y\cdot\nabla U(x))_+.$$

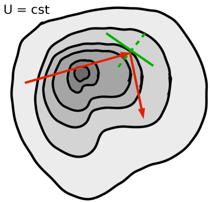
To target μ, a necessary and sufficient condition is that, at a jump,

$$Y\cdot\nabla U(X) \;\leftarrow\; -Y\cdot\nabla U(X).$$

For instance:

- $Q(x,y,\cdot)=\delta_{-y}$
- $Q(x,y,\cdot)=\delta_{y_*}$ with

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



In higher dimension

We want to keep the same rate:

$$\lambda(x,y) \;=\; (y\cdot\nabla U(x))_+.$$

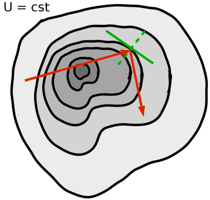
To target μ, a necessary and sufficient condition is that, at a jump,

$$Y\cdot\nabla U(X) \;\leftarrow\; -Y\cdot\nabla U(X).$$

For instance:

- $Q(x,y,\cdot)=\delta_{-y}$
- $Q(x,y,\cdot)=\delta_{y_*}$ with

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



Not ergodic in general!

In higher dimension

At constant rate, the velocity can be (uniformly) refreshed. Ultimately,

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

In higher dimension

At constant rate, the velocity can be (uniformly) refreshed. Ultimately,

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

- ergodic with equilibrium $\mu = e^{-U}\mathrm{d}x\mathrm{d}z$.
- kinetic, non-reversible
- PDMP, no discretization needed thanks to a thining method.

In higher dimension

At constant rate, the velocity can be (uniformly) refreshed. Ultimately,

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

- ergodic with equilibrium $\mu = e^{-U}\mathrm{d}x\mathrm{d}z$.
- kinetic, non-reversible
- PDMP, no discretization needed thanks to a thining method.

In higher dimension

At constant rate, the velocity can be (uniformly) refreshed. Ultimately,

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

- ergodic with equilibrium $\mu = e^{-U}\mathrm{d}x\gamma(\mathrm{d}y)$.
- kinetic, non-reversible
- PDMP, no discretization needed thanks to a thining method.

Velocity jump processes

- Miclo, M. (2012, *Volte-Face*, 2013, 2016)

Velocity jump processes

- Miclo, M. (2012, *Volte-Face*, 2013, 2016)
- Peters, de With (2012, *Rejection-free Metropolis Hastings*, $y \in \mathbb{R}^d$ with Gaussian refreshments; *event-driven MC* in physics litterature, Michel, Kapfer, Krauth 2013 by ex.), Bouchard-Côté, Vollmer, Doucet (2016, 2017, *bouncy particle sampler*)

Velocity jump processes

- Miclo, M. (2012, *Volte-Face*, 2013, 2016)
- Peters, de With (2012, *Rejection-free Metropolis Hastings*, $y \in \mathbb{R}^d$ with Gaussian refreshments; *event-driven MC* in physics 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*, bacterial chemotaxis, non-explicit equilibrium, $y \in [-1, 1]$).

Velocity jump processes

- Miclo, M. (2012, *Volte-Face*, 2013, 2016)
- Peters, de With (2012, *Rejection-free Metropolis Hastings*, $y \in \mathbb{R}^d$ with Gaussian refreshments; *event-driven MC* in physics 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*, bacterial chemotaxis, non-explicit equilibrium, $y \in [-1, 1]$).
- Bierkens, Fearnhead, Roberts (2016, *Zig-zag process*, $y \in \{-1, +1\}^d$)

1 Introduction

2 Persistent walk and kinetic PDMP

3 Some results

Metastability

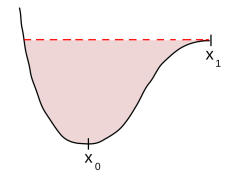
Replace U by $\frac{1}{\varepsilon}U$, with minimal rate $\lambda(x,y) = \frac{1}{\varepsilon}\left(y\nabla U(x)\right)_+$.

Theorem (Eyring-Kramers formula)

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

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

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



Metastability

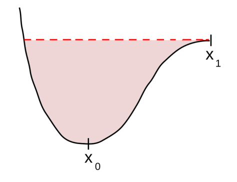
Replace U by $\frac{1}{\varepsilon}U$, with minimal rate $\lambda(x,y) = \frac{1}{\varepsilon}\left(y\nabla U(x)\right)_+$.

Theorem (Eyring-Kramers formula)

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

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

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



Theorem (annealing)

With a cooling schedule $(\varepsilon_t)_{t\geqslant 0}$, *NSC for the annealing:*

$$\forall\delta > 0 \lim_{t\to\infty}\mathbb{P}\left(U(X_t) < \min_{\mathbb{R}} U + \delta\right) = 1 \quad\Leftrightarrow\quad \int_0^\infty (\varepsilon_s)^{-\frac{1}{2}}\, e^{-\frac{E^*}{\varepsilon_s}} ds = \infty.$$

Sketch of the proof for the EK formula

$$\begin{aligned}\mathbb{E}\left[\tau\right] &= \mathbb{E}\left[\text{duration of a failed attempt to escape}\right] \\ &\quad \times \mathbb{E}\left[\text{number of failure}\right] \times \left(1 + \underset{\varepsilon\rightarrow 0}{o}(1)\right).\end{aligned}$$

As far as the second term is concerned,

$$\mathbb{P}\left(\text{escape in one shot}\right) = \mathbb{P}_{\mathcal{E}(1)}\left(\varepsilon E \geq U(x_1) - U(x_0)\right) = e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}}.$$

For the first one, if $\delta > 0$ is small enough,

$$\int_0^\delta \frac{t}{\varepsilon}\left(-U'(x_0 - t)\right) e^{-\frac{U(x_0-t)-U(x_0)}{\varepsilon}} dt = \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}}\left(1 + \underset{\varepsilon\rightarrow 0}{o}(1)\right).$$

Sketch of the proof for the EK formula

$$\begin{aligned}\mathbb{E}[\tau] &= \mathbb{E}\left[\text{duration of a failed attempt to escape}\right] \\ &\quad \times \mathbb{E}\left[\text{number of failure}\right] \times \left(1 + \underset{\varepsilon\to 0}{o}(1)\right).\end{aligned}$$

As far as the second term is concerned,

$$\mathbb{P}\left(\text{escape in one shot}\right) = \mathbb{P}_{\mathcal{E}(1)}\left(\varepsilon E \geq U(x_1) - U(x_0)\right) = e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}}.$$

For the first one, if $\delta > 0$ is small enough,

$$\int_0^\delta \frac{t}{\varepsilon}\left(-U'(x_0 - t)\right) e^{-\frac{U(x_0-t)-U(x_0)}{\varepsilon}} dt = \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}}\left(1 + \underset{\varepsilon\to 0}{o}(1)\right).$$

Remark: with a supplementary rate $a \neq 0$, one gets

$$\mathbb{P}\left(\text{escape in one shot}\right) = \frac{e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}}}{1 + \int_{x_0}^{x_1} a(z) e^{-\frac{U(x_1)-U(z)}{\varepsilon}} dz}$$

Sketch of the proof for the EK formula

$$\begin{aligned}\mathbb{E}\left[\tau\right] &= \mathbb{E}\left[\text{duration of a failed attempt to escape}\right] \\ &\quad \times \mathbb{E}\left[\text{number of failure}\right] \times \left(1 + \underset{\varepsilon\to 0}{o}(1)\right).\end{aligned}$$

As far as the second term is concerned,

$$\mathbb{P}\left(\text{escape in one shot}\right) = \mathbb{P}_{\mathcal{E}(1)}\left(\varepsilon E \geq U(x_1) - U(x_0)\right) = e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}}.$$

For the first one, if $\delta > 0$ is small enough,

$$\int_0^\delta \frac{t}{\varepsilon}\left(-U'(x_0 - t)\right) e^{-\frac{U(x_0-t)-U(x_0)}{\varepsilon}} dt = \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}}\left(1 + \underset{\varepsilon\to 0}{o}(1)\right).$$

Remark: with a supplementary rate $a \neq 0$, one gets

$$\mathbb{P}\left(\text{escape in one shot}\right) = \frac{e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}}}{1 + \int_{x_0}^{x_1} a(z) e^{-\frac{U(x_1)-U(z)}{\varepsilon}} dz} \underset{\varepsilon\to 0}{\lesssim} e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}}.$$

Sketch of the proof for the annealing algorithm

Regardless of X_0 et t_0, there is a positive probability that the process reaches x_0 after the time t_0. The question is: does it succeed in escaping ?

Suppose the temperature is kept constant during one attempt,

$$\mathbb{P}(\text{success of the } k^{th} \text{ attempt}) \quad = \quad e^{-\frac{E}{\varepsilon_k}}.$$

The result is then mainly the consequence of the Eyring-Kramers and of the Borel-Cantelli Theorem.

Metastability in higher dimension

The study is restricted to the compact (periodic) case. Denote $Z=(X,Y)$ and

$$\|\nu_1-\nu_2\|_1=\inf_{Z_i\sim\nu_i}\mathbb{P}\left(Z_1\neq Z_2\right).$$

Theorem

1. *There exist $\theta,c,t_0>0$ which depend only on the potential U, the rate r and the dimension d such that*

$$\|\mathcal{L}\left(Z_t\right)-\mathcal{L}\left(Z_\infty\right)\|_1\leq e^{-ce^{\frac{-\theta}{\varepsilon}}(t-t_0)}\|\mathcal{L}\left(Z_0\right)-\mathcal{L}\left(Z_\infty\right)\|_1.$$

Metastability in higher dimension

The study is restricted to the compact (periodic) case. Denote $Z=(X,Y)$ and

$$\|\nu_1-\nu_2\|_1=\inf_{Z_i\sim\nu_i}\mathbb{P}\left(Z_1\neq Z_2\right).$$

Theorem

1. *There exist* $\theta,c,t_0>0$ *which depend only on the potential* U*, the rate* r *and the dimension* d *such that*

$$\|\mathcal{L}\left(Z_t\right)-\mathcal{L}\left(Z_\infty\right)\|_1\leq e^{-ce^{\frac{-\theta}{\varepsilon}}(t-t_0)}\|\mathcal{L}\left(Z_0\right)-\mathcal{L}\left(Z_\infty\right)\|_1.$$

2. *For the annealing, if* $\frac{d}{dt}\left(\frac{1}{\varepsilon_t}\right)\leq\frac{1}{(\theta+\eta)t}$ *with* $\eta>0$ *then* $\forall\delta>0$

$$\mathbb{P}\left(U(X_t)>\min U+\delta\right)\underset{t\to\infty}{\longrightarrow}0.$$

Metastability in higher dimension

The study is restricted to the compact (periodic) case. Denote $Z=(X,Y)$ and

$$\|\nu_1-\nu_2\|_1 = \inf_{Z_i\sim\nu_i} \mathbb{P}\left(Z_1\neq Z_2\right).$$

Theorem

1. *There exist $\theta, c, t_0 > 0$ which depend only on the potential U, the rate r and the dimension d such that*

$$\left\|\mathcal{L}\left(Z_t\right)-\mathcal{L}\left(Z_\infty\right)\right\|_1 \leq e^{-ce^{\frac{-\theta}{\varepsilon}}(t-t_0)}\left\|\mathcal{L}\left(Z_0\right)-\mathcal{L}\left(Z_\infty\right)\right\|_1.$$

2. *For the annealing, if $\frac{d}{dt}\left(\frac{1}{\varepsilon_t}\right) \leq \frac{1}{(\theta+\eta)t}$ with $\eta>0$ then $\forall\delta>0$*

$$\mathbb{P}\left(U\left(X_t\right)>\min U+\delta\right) \underset{t\to\infty}{\longrightarrow} 0.$$

Proof: couplings.

Some remarks

- The NSC in dimension 1 implies
  - if $\varepsilon_t \geq \frac{c}{\ln(1+t)}$ with $c > E^*$, the algorithm converges,
  - if $\varepsilon_t \leq \frac{c}{\ln(1+t)}$ with $c < E^*$, it may fail.

  The SC in higher dimension implies
  - if $\varepsilon_t \geq \frac{c}{\ln(1+t)}$ with $c > \theta$, the algorithm converges.

Some remarks

- The NSC in dimension 1 implies
  - if $\varepsilon_t \geq \frac{c}{\ln(1+t)}$ with $c > E^*$, the algorithm converges,
  - if $\varepsilon_t \leq \frac{c}{\ln(1+t)}$ with $c < E^*$, it may fail.

  The SC in higher dimension implies
  - if $\varepsilon_t \geq \frac{c}{\ln(1+t)}$ with $c > \theta$, the algorithm converges.
- Bad news: not faster than the overdamped Langevin diffusion, same metastability (short memory).
- Good news: as fast as the overdamped Langevin diffusion, and simpler to compute (”exact” computation, no MH-step needed).

Some remarks

- The NSC in dimension 1 implies
  - if $\varepsilon_t \geq \frac{c}{\ln(1+t)}$ with $c > E^*$, the algorithm converges,
  - if $\varepsilon_t \leq \frac{c}{\ln(1+t)}$ with $c < E^*$, it may fail.

  The SC in higher dimension implies
  - if $\varepsilon_t \geq \frac{c}{\ln(1+t)}$ with $c > \theta$, the algorithm converges.
- Bad news: not faster than the overdamped Langevin diffusion, same metastability (short memory).
- Good news: as fast as the overdamped Langevin diffusion, and simpler to compute ("exact" computation, no MH-step needed).
- Anyway, at this scale (never reached in practice), all local Markov dynamics perform the same.

Some remarks

- The NSC in dimension 1 implies
  - if $\varepsilon_t \geq \frac{c}{\ln(1+t)}$ with $c > E^*$, the algorithm converges,
  - if $\varepsilon_t \leq \frac{c}{\ln(1+t)}$ with $c < E^*$, it may fail.

  The SC in higher dimension implies
  - if $\varepsilon_t \geq \frac{c}{\ln(1+t)}$ with $c > \theta$, the algorithm converges.
- Bad news: not faster than the overdamped Langevin diffusion, same metastability (short memory).
- Good news: as fast as the overdamped Langevin diffusion, and simpler to compute ("exact" computation, no MH-step needed).
- Anyway, at this scale (never reached in practice), all local Markov dynamics perform the same.

Another question: how do you chose the radius of the ball (i.e. the scalar velocity of the process) ? Or, in the Gaussian case, the variance at equilibrium of the velocity ?

Some remarks

Same question for the kinetic Langevin equation:

$$\begin{aligned} \mathrm{d}X_t &= Y_t \mathrm{d}t \\ \mathrm{d}Y_t &= -\nu \nabla U(X_t)\,\mathrm{d}t - \frac{1}{\nu} Y_t \mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \end{aligned}$$

with equilibrium $e^{-U(x)-\frac{|y|^2}{2\nu}}\mathrm{d}x\mathrm{d}y$. Calibrate ν ?

Some remarks

Same question for the kinetic Langevin equation:

$$\begin{aligned} \mathrm{d}X_t &= Y_t \mathrm{d}t \\ \mathrm{d}Y_t &= -\nu \nabla U\left(X_t\right) \mathrm{d}t - \frac{1}{\nu} Y_t \mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \end{aligned}$$

with equilibrium $e^{-U(x)-\frac{|y|^2}{2\nu}}\mathrm{d}x\mathrm{d}y$. Calibrate ν ?

When $U(x) = \frac{1}{2}\lambda|x|^2$, $\nu_{opt} = (4\lambda)^{-\frac{1}{3}}$ with convergence rate $\left(\frac{1}{2}\lambda\right)^{\frac{1}{3}}$.
By comparison, the rate of convergence of

$$\mathrm{d}X_t = -\lambda X_t \mathrm{d}t + \sqrt{2}\mathrm{d}B_t$$

is λ, which is better than $(\lambda/2)^{\frac{1}{3}}$ if and only if $\lambda > \frac{1}{\sqrt{2}}$.

Some remarks

- Too much inertia kills inertia (example of the kinetic diffusion; or Gadat-Panloup 2012 on long-term memory gradient).
- However, the escape time from local traps may be as small as we want.

Some remarks

- Too much inertia kills inertia (example of the kinetic diffusion; or Gadat-Panloup 2012 on long-term memory gradient).
- However, the escape time from local traps may be as small as we want.
- Problem: entropic barrier.

Some remarks

- Too much inertia kills inertia (example of the kinetic diffusion; or Gadat-Panloup 2012 on long-term memory gradient).
- However, the escape time from local traps may be as small as we want.
- Problem: entropic barrier.
- Short-term memory (and more generally non-reversible sampling) can be used together with global and long-memory methods (Wang-Landau, ABF, metadynamics, etc.)

Some remarks

- Too much inertia kills inertia (example of the kinetic diffusion; or Gadat-Panloup 2012 on long-term memory gradient).
- However, the escape time from local traps may be as small as we want.
- Problem: entropic barrier.
- Short-term memory (and more generally non-reversible sampling) can be used together with global and long-memory methods (Wang-Landau, ABF, metadynamics, etc.)
- Work in progress with Alain Durmus, Ninon Fetique and Arnaud Guillin: long-time convergence of the BPS in $\mathbb{R}^d$.

Some references

A. Bouchard-Côté, S. J. Vollmer, and A. Doucet.
The Bouncy Particle Sampler: A Non-Reversible Rejection-Free Markov Chain Monte Carlo Method.
Journal of the American Statistical Association, 2017.

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

M. Michel, S.C. Kapfer, and W. Krauth.
Generalized event-chain monte carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps.
The Journal of Chemical Physics, 140, 2014.

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.

A. Scemama, T. Lelièvre, G. Stoltz, and M. Caffarel.
An efficient sampling algorithm for variational monte carlo.
Journal of Chemical Physics, 125, September 2006.