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

\[
\mathbb{E}(f(X)) = \frac{\int f(x) e^{-\frac{1}{\varepsilon} U(x)} dx}{\int e^{-\frac{1}{\varepsilon} U(x)} dx} = \int f d\mu
\]

with

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

- Problem: estimate

\[ \mathbb{E}(f(X)) = \frac{\int f(x) e^{-\frac{1}{\epsilon} U(x)} dx}{\int e^{-\frac{1}{\epsilon} U(x)} dx} = \int f \, d\mu \]

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

- MCMC basic ingredient: an ergodic (Markov) process \((X_t)_{t \geq 0}\), i.e.

\[ \frac{1}{t} \int_0^t f(X_s) \, ds \quad \xrightarrow{t \to \infty} \quad \int f \, d\mu \]
Many available dynamics

- (reversible) overdamped Langevin diffusion:

\[ dX_t = -\nabla U(X_t)dt + \sqrt{2\epsilon}dB_t, \]

- kinetic Langevin equation:

\[
\begin{align*}
    dX_t &= Y_t dt \\
    dY_t &= -\nabla U(X_t)dt - Y_t dt + \sqrt{2\epsilon}dB_t,
\end{align*}
\]

- Metropolis-Hastings algorithm (propose, accept/reject),

- Hamiltonian Monte-Carlo.
How to compare them?

Some criteria:

- asymptotic variance in a Central Limit Theorem
  \[
  \sqrt{t} \left( \frac{1}{t} \int_0^t f(X_s) \, ds - \int f \, d\mu \right) \xrightarrow{t \to \infty} \mathcal{N}(0, \sigma_f).
  \]

- Relaxation speed toward equilibrium
  \[
  \mathcal{L}(X_t \mid X_0 \sim \nu) \xrightarrow{t \to \infty} \mu.
  \]

- mixing (decorrelation) time

- scaling of the chain (diffusive, ballistic...)

Pierre Monmarché (CERMICS)
How to compare them?

Some criteria:

- asymptotic variance in a Central Limit Theorem

\[
\sqrt{t} \left( \frac{1}{t} \int_0^t f(X_s) \, ds - \int f \, d\mu \right) \xrightarrow{t \to \infty} \mathcal{N}(0, \sigma_f).
\]

- Relaxation speed toward equilibrium

\[
\mathcal{L}(X_t \mid X_0 \sim \nu) \xrightarrow{t \to \infty} \mu.
\]

- 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.
Problem: minimize a function
- in large dimension (or large finite set),
- with many local minima.

The overdamped Langevin diffusion

\[ dX_t = -\nabla U(X_t) dt + \sqrt{2\varepsilon} dB_t \]

will eventually escape from any local minima.

\[ \mathcal{L}(X_t) \xrightarrow{t \to \infty} e^{-\frac{U(x)}{\varepsilon}} dx \]
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

\[
dX_t = -\nabla U(X_t)dt + \sqrt{2\varepsilon}dB_t
\]

will eventually escape from any local minima.

\[
\mathcal{L}(X_t) \xrightarrow{t \to \infty} e^{-\frac{U(x)}{\varepsilon}}dx \xrightarrow{\varepsilon \to 0} \delta_{\text{argmin}U}.
\]
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

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

will eventually escape from any local minima.

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

How fast can one cool down the system?

⇔ At a given temperature, how fast is the convergence to equilibrium?

Pierre Monmarché (CERMICS)

05/05/2017 7 / 28
Metastability

For the overdamped Langevin diffusion:

\[
dX_t = -\nabla U(X_t)dt + \sqrt{2\varepsilon}dB_t,
\]

at small temperature \(\varepsilon \to 0\):

- escape time from minima \(\gtrsim e^{1/\varepsilon}\Delta U\)
- relaxation rate to equilibrium \(\gtrsim e^{-1/\varepsilon}E\)
- condition on the cooling schedule \(\varepsilon_t \gtrsim \frac{E}{\ln t}\).
Metastability

For the overdamped Langevin diffusion:

\[ dX_t = -\nabla U(X_t)dt + \sqrt{2\varepsilon}dB_t, \]

at small temperature \( \varepsilon \to 0 \):

- escape time from minima \( \approx e^{\frac{1}{\varepsilon} \Delta U} \)
- relaxation rate to equilibrium \( \approx e^{-\frac{1}{\varepsilon}E} \)
- condition on the cooling schedule \( \varepsilon_t \) \( \gtrsim \frac{E}{\ln t} \).

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 (⇒ 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 (⇒ reaction coordinates, collective variables, coarse-grained model).

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

Persistent walk and kinetic PDMP

Some results
A second order Markov chain: the persistent walk

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

\[
\mathbb{P}(X_{n+1} - X_n = X_n - X_{n-1}) = \frac{1 + \alpha}{2},
\]

\[
\mathbb{P}(X_{n+1} - X_n = -(X_n - X_{n-1})) = \frac{1 - \alpha}{2}.
\]
A second order Markov chain: the persistent walk

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

\[
\begin{align*}
\mathbb{P}(X_{n+1} - X_n = X_n - X_{n-1}) &= \frac{1 + \alpha}{2} \\
\mathbb{P}(X_{n+1} - X_n = -(X_n - X_{n-1})) &= \frac{1 - \alpha}{2}.
\end{align*}
\]

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

\[
\begin{align*}
\mathbb{P}(Y_{n+1} = Y_n) &= \frac{1 + \alpha}{2} \\
\mathbb{P}(Y_{n+1} = -Y_n) &= \frac{1 - \alpha}{2} \\
X_{n+1} &= X_n + Y_{n+1}.
\end{align*}
\]
A second order Markov chain: the persistent walk

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

\[
\mathbb{P} (X_{n+1} - X_n = X_n - X_{n-1}) = \frac{1 + \alpha}{2},
\]

\[
\mathbb{P} (X_{n+1} - X_n = -(X_n - X_{n-1})) = \frac{1 - \alpha}{2}.
\]

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

\[
\mathbb{P} (Y_{n+1} = Y_n) = \frac{1 + \alpha}{2},
\]

\[
\mathbb{P} (Y_{n+1} = -Y_n) = \frac{1 - \alpha}{2},
\]

\[
X_{n+1} = X_n + Y_{n+1}.
\]

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:

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

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

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

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

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

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

For the symmetric walk,

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

It took $O(N^2)$ steps to mix, and now only $O(N)$ (Nota: the determinisitic 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 $Y = \pm 1$
- $dX_t = Y_t dt$ (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 $Y = \pm 1$
- $dX_t = Y_t \, dt$ (kinetic process)
- $Y$ jumps to $-Y$ at rate $a > 0$ (PDMP, piecewise deterministic Markov process)

Uniform equilibrium $\mu$, 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 $a_{opt} = 1$,

$$\|e^{tL} - \mu\| = e^{-t} \sqrt{1 + \frac{2}{\sqrt{1 + \frac{1}{t^2}} - 1}} \sim_{t \to 0} 1 - \frac{t^3}{3}.$$
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 $Y = \pm 1$
- $dX_t = Y_t dt$ (kinetic process)
- $Y$ jumps to $-Y$ at rate $a > 0$ (PDMP, piecewise deterministic Markov process)

Uniform equilibrium $\mu$, and generator

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

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

$$\|e^{tL} - \mu\| = e^{-t} \sqrt{1 + \frac{2}{\sqrt{1 + \frac{1}{t^2}} - 1}} \overset{t \to \infty}{\sim} 2te^{-t}$$
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 $Y = \pm 1$
- $dX_t = Y_t dt$ (kinetic process)
- $Y$ jumps to $-Y$ at rate $a > 0$ (PDMP, piecewise deterministic Markov process)

Uniform equilibrium $\mu$, and generator

\[
Lf(x, y) = y \partial_x f(x, y) + a (f(x, -y) - f(x, y)) .
\]

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

\[
\| e^{tL} - \mu \| = e^{-t} \sqrt{1 + \frac{2}{t^2}} \sim 2te^{-t} \quad \text{as} \quad t \to \infty.
\]

Remark: $a = 0 \Rightarrow$ no cv, but

\[
\left| \frac{1}{t} \int_0^t f(x + s) ds - \int f d\mu \right| \leq \frac{c}{t}.
\]
With a potential

Specifications:

- $(X, Y)$ Markov on $\mathbb{R} \times \{\pm 1\}$
- $dX = Y \, dt$
- equilibrium $\mu = e^{-U(x)} \, dx \, dy$
With a potential

Specifications:

- \((X, Y)\) Markov on \(\mathbb{R} \times \{\pm 1\}\)
- \(dX = Y \, dt\)
- equilibrium \(\mu = e^{-U(x)} \, dx \, dy\)

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

\[
\lambda(x, y) = (y U'(x))_+ + a(x).
\]

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

\[
T = \inf \left\{ \ t > 0, \ E > \int_0^t \lambda(X_s, Y_s) \, ds \right\}.
\]
The minimal jump rate

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

$$\int_0^t \lambda(X_s, Y_s)ds = U(X_t) - U(X_0) = 0$$

as long as we climb up

as long as we fall down.
The minimal jump rate

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

$$\int_0^t \lambda(X_s, Y_s) ds = U(X_t) - U(X_0) = 0$$

as long as we climb up

as long as we fall down.
The minimal jump rate

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

$$\int_0^t \lambda(X_s, Y_s) \, ds = U(X_t) - U(X_0) = 0$$

as long as we climb up

as long as we fall down.
The minimal jump rate

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

$$\int_0^t \lambda(X_s, Y_s)ds = U(X_t) - U(X_0) = 0$$

as long as we climb up

as long as we fall down.
The minimal jump rate

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

$$
\int_0^t \lambda(X_s, Y_s) ds = U(X_t) - U(X_0) = 0
$$

as long as we climb up

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 \( \mu \), 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 $\mu$, 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 $\mu$, 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 \( \mu \), 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,

\[
L f(x, y) = y \nabla_x f(x, y) + (y \cdot \nabla U(x)) + (f(x, y_\ast) - f(x, y))
\]

\[
+ r \left( \int_{S^{d-1}} f(x, z) dz - f(x, y) \right).
\]
In higher dimension

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

\[
Lf(x, y) = y \nabla_x f(x, y) + (y \cdot \nabla U(x)) + (f(x, y^*) - f(x, y)) + r \left( \int_{S^{d-1}} f(x, z) dz - f(x, y) \right).
\]

- ergodic with equilibrium \( \mu = e^{-U} dx dz \).
- kinetic, non-reversible
- PDMP, no discretization needed thanks to a thinning method.
In higher dimension

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

\[ Lf(x, y) = y \nabla_x f(x, y) + (y \cdot \nabla U(x)) + (f(x, y^*) - f(x, y)) + r \left( \int_{S^{d-1}} f(x, z) \, dz - f(x, y) \right). \]

- ergodic with equilibrium \( \mu = e^{-U} \, dx \, dz. \)
- kinetic, non-reversible
- PDMP, no discretization needed thanks to a thinning method.
In higher dimension

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

\[
Lf(x, y) = y\nabla_x f(x, y) + (y \cdot \nabla U(x)) + (f(x, y_*) - f(x, y))
\]
\[
+ r \left( \int_{S^{d-1}} f(x, z) \gamma(dz) - f(x, y) \right).
\]

- ergodic with equilibrium \( \mu = e^{-U} dx \gamma(dy) \).
- kinetic, non-reversible
- PDMP, no discretization needed thanks to a thinning method.
Velocity jump processes

Velocity jump processes


Velocity jump processes

- Fontbona, Guérin, Malrieu (2012, 2016, *integrated telegraph process*)
Velocity jump processes

- Fontbona, Guérin, Malrieu (2012, 2016, *integrated telegraph process*)
- 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 $\lambda(x, y) = \frac{1}{\varepsilon} (y\nabla U(x))_+$. 

**Theorem (Eyring-Kramers formula)**

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

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

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

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

**Theorem (Eyring-Kramers formula)**

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

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

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

**Theorem (annealing)**

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

$$\forall \delta > 0 \lim_{t \to \infty} \mathbb{P}\left( U(X_t) < \min_{\mathbb{R}} U + \delta \right) = 1 \iff \int_{0}^{\infty} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E^*}{\varepsilon_s}} ds = \infty.$$
Sketch of the proof for the EK formula

\[ \mathbb{E}[\tau] = \mathbb{E}[\text{duration of a failed attempt to escape}] \times \mathbb{E}[\text{number of failure}] \times \left(1 + o(1)\right). \]

As far as the second term is concerned,

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

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

\[ \int_0^\delta \frac{t}{\varepsilon} \left(-U'(x_0 - t)\right) e^{-\frac{U(x_0-t) - U(x_0)}{\varepsilon}} dt = \sqrt{\frac{\pi \varepsilon}{2U''(x_0)}} \left(1 + o(1)\right). \]
Sketch of the proof for the EK formula

\[ \mathbb{E}[\tau] = \mathbb{E}[\text{duration of a failed attempt to escape}] \times \mathbb{E}[\text{number of failure}] \times \left(1 + \frac{o}{\varepsilon} (1)\right). \]

As far as the second term is concerned,

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

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

\[ \int_{0}^{\delta} \frac{t}{\varepsilon} \left(-U'(x_0 - t)\right) e^{-\frac{U(x_0 - t) - U(x_0)}{\varepsilon}} \, dt = \sqrt{\frac{\pi \varepsilon}{2U''(x_0)}} \left(1 + \frac{o}{\varepsilon} (1)\right). \]

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

\[ \mathbb{P}(\text{escape in one shot}) = \frac{e^{-\frac{U(x_1) - U(x_0)}{\varepsilon}}}{1 + \int_{x_0}^{x_1} a(z)e^{-\frac{U(x_1) - U(z)}{\varepsilon}} \, dz}. \]
Sketch of the proof for the EK formula

\[ E \left[ \tau \right] = E \left[ \text{duration of a failed attempt to escape} \right] \times E \left[ \text{number of failure} \right] \times \left( 1 + o \left( \frac{1}{\varepsilon} \right) \right). \]

As far as the second term is concerned,

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

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

\[ \int_{0}^{\delta} \frac{t}{\varepsilon} \left( -U'(x_0 - t) \right) e^{-\frac{U(x_0 - t) - U(x_0)}{\varepsilon}} dt = \sqrt{\frac{\pi \varepsilon}{2U''(x_0)}} \left( 1 + o \left( \frac{1}{\varepsilon} \right) \right). \]

**Remark:** with a supplementary rate \( a \neq 0 \), one gets

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

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

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

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

### 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

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

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

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

**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**

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

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

$$\mathbb{P}(U(X_t) > \min U + \delta) \xrightarrow{t \to \infty} 0.$$
Metastability in higher dimension

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

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

**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**

\[
\| \mathcal{L} (Z_t) - \mathcal{L} (Z_\infty) \|_1 \leq e^{-c} e^{-\theta (t-t_0)} \| \mathcal{L} (Z_0) - \mathcal{L} (Z_\infty) \|_1.
\]

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

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

**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.

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

- 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:

\[ dX_t = Y_t \, dt \]
\[ dY_t = -\nu \nabla U (X_t) \, dt - \frac{1}{\nu} Y_t \, dt + \sqrt{2} dB_t, \]

with equilibrium \( e^{-U(x) - \frac{|y|^2}{2\nu}} \, dx \, dy \). Calibrate \( \nu \) ?
Some remarks

Same question for the kinetic Langevin equation:

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

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

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

\[
\begin{align*}
\mathrm{d}X_t &= -\lambda X_t \mathrm{d}t + \sqrt{2} \, \mathrm{d}B_t
\end{align*}
\]

is \( \lambda \), 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.

V. Calvez, G. Raoul, and C. Schmeiser.
Confinement by biased velocity jumps: aggregation of *escherichia coli*.

M. Michel, S.C. Kapfer, and W. Krauth.
Generalized event-chain monte carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps.

P. Monmarché.
Piecewise deterministic simulated annealing.

E. A. J. F. Peters and G. de With.
Rejection-free monte carlo sampling for general potentials.

A. Scemama, T. Lelièvre, G. Stoltz, and M. Caffarel.
An efficient sampling algorithm for variational monte carlo.