Numerical methods in molecular dynamics, or what is a metastable stochastic process ?

T. Lelièvre

CERMICS - Ecole des Ponts ParisTech & Matherials project-team - INRIA

Joint work with D. Aristoff, C. Chipot, N. Chopin, C. Le Bris, M. Luskin, D. Perez, M. Rousset, G. Simpson and G. Stoltz.

ICMS, Computational methods for statistical mechanics, June 2014

Introduction

The aim of molecular dynamics simulations is to understand the relationships between the macroscopic properties of a molecular system and its atomistic features. In particular, one would like to evaluate numerically macroscopic quantities from models at the microscopic scale.

Many applications in various fields: biology, physics, chemistry, materials science.

Various models: discrete state space (kinetic Monte Carlo, Markov State Model) or continuous state space (Langevin).

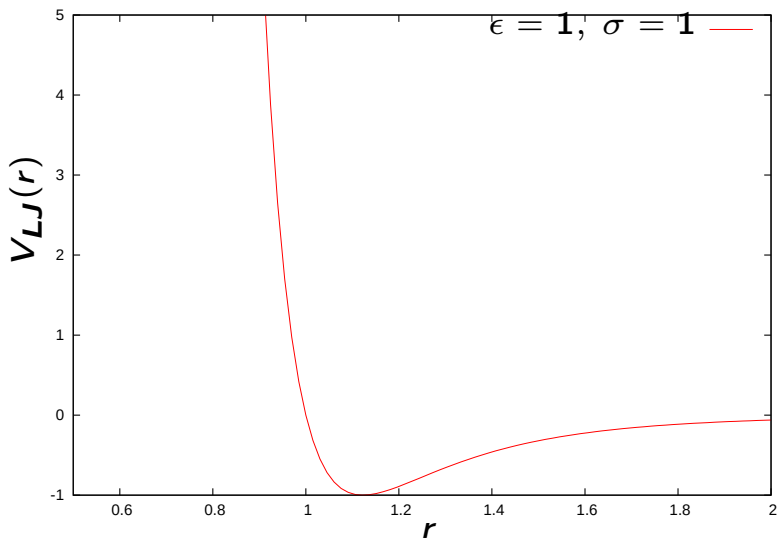
The basic ingredient: a potential V which associates to a configuration $(x_1, ..., x_N) = x \in \mathbb{R}^{3N}$ an energy $V(x_1, ..., x_N)$.

Introduction

Typically, V is a sum of potentials modelling interaction between two particles, three particles and four particles:

$$V = \sum_{i<j} V_1(x_i, x_j) + \sum_{i<j<k} V_2(x_i, x_j, x_k) + \sum_{i<j<k<l} V_3(x_i, x_j, x_k, x_l).$$

For example,
$V_1(x_i, x_j) = V_{LJ}(|x_i - x_j|)$
where
$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)$ is the Lennard-Jones potential.



Introduction

Newton equations of motion:

$$\begin{cases} d\boldsymbol{X}_t = M^{-1}\boldsymbol{P}_t\, dt, \\ d\boldsymbol{P}_t = -\nabla V(\boldsymbol{X}_t)\, dt, \end{cases}$$

Introduction

Newton equations of motion + thermostat: Langevin dynamics:

$$\begin{cases} d\boldsymbol{X}_t = M^{-1}\boldsymbol{P}_t\,dt, \\ d\boldsymbol{P}_t = -\nabla V(\boldsymbol{X}_t)\,dt - \gamma M^{-1}\boldsymbol{P}_t\,dt + \sqrt{2\gamma\beta^{-1}}d\boldsymbol{W}_t, \end{cases}$$

where $\gamma > 0$. Langevin dynamics is ergodic wrt $\mu(d\boldsymbol{x}) \otimes Z_p^{-1} \exp\left(-\beta \frac{\boldsymbol{p}^t M^{-1}\boldsymbol{p}}{2}\right) d\boldsymbol{p}$ with

$$d\mu = Z^{-1}\exp(-\beta V(\boldsymbol{x}))\,d\boldsymbol{x},$$

where $Z = \int \exp(-\beta V(\boldsymbol{x}))\,d\boldsymbol{x}$ is the partition function and $\beta = (k_B T)^{-1}$ is proportional to the inverse of the temperature.

Introduction

Newton equations of motion + thermostat: Langevin dynamics:

$$\begin{cases} d\boldsymbol{X}_t = M^{-1}\boldsymbol{P}_t\,dt, \\ d\boldsymbol{P}_t = -\nabla V(\boldsymbol{X}_t)\,dt - \gamma M^{-1}\boldsymbol{P}_t\,dt + \sqrt{2\gamma\beta^{-1}}d\boldsymbol{W}_t, \end{cases}$$

where $\gamma > 0$. Langevin dynamics is ergodic wrt $\mu(d\boldsymbol{x}) \otimes Z_p^{-1} \exp\left(-\beta \frac{\boldsymbol{p}^t M^{-1}\boldsymbol{p}}{2}\right) d\boldsymbol{p}$ with

$$d\mu = Z^{-1}\exp(-\beta V(\boldsymbol{x}))\,d\boldsymbol{x},$$

where $Z = \int \exp(-\beta V(\boldsymbol{x}))\,d\boldsymbol{x}$ is the partition function and $\beta = (k_B T)^{-1}$ is proportional to the inverse of the temperature.

In the following, we focus on the *over-damped Langevin* (or gradient) dynamics

$$d\boldsymbol{X}_t = -\nabla V(\boldsymbol{X}_t)\,dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}_t,$$

which is also ergodic wrt μ.

Introduction

These dynamics are used to compute macroscopic quantities:

(i) Thermodynamic quantities (averages wrt μ of some observables): stress, heat capacity, free energy,...

$$\mathbb{E}_\mu(\varphi(\boldsymbol{X})) = \int_{\mathbb{R}^d} \varphi(x)\,\mu(dx) \simeq \frac{1}{T}\int_0^T \varphi(\boldsymbol{X}_t)\,dt.$$

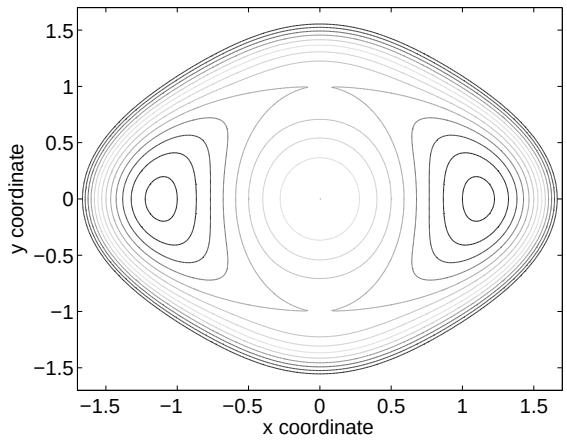
(ii) Dynamical quantities (averages over trajectories): diffusion coefficients, viscosity, transition rates,...

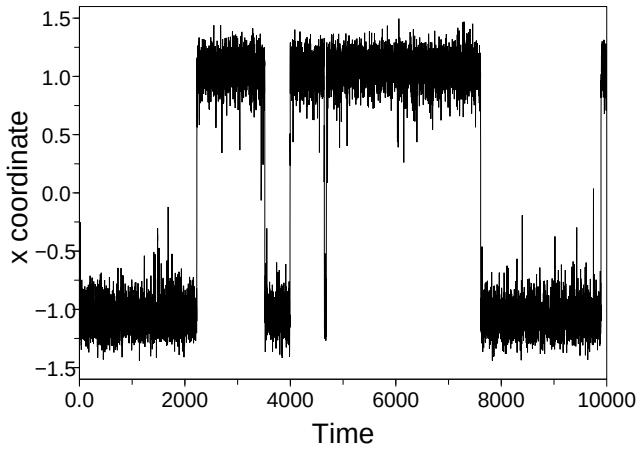
$$\mathbb{E}(\mathcal{F}((\boldsymbol{X}_t)_{t\geq 0})) \simeq \frac{1}{N}\sum_{m=1}^N \mathcal{F}((\boldsymbol{X}_t^m)_{t\geq 0}).$$

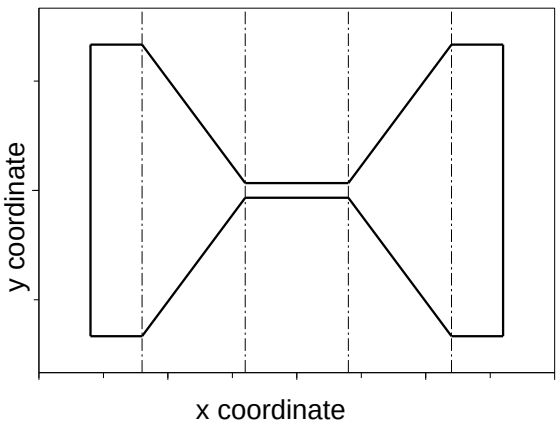
Difficulties: (i) high-dimensional problem ($N \gg 1$); (ii) $\boldsymbol{X}_t$ is a metastable process and μ is a multimodal measure.

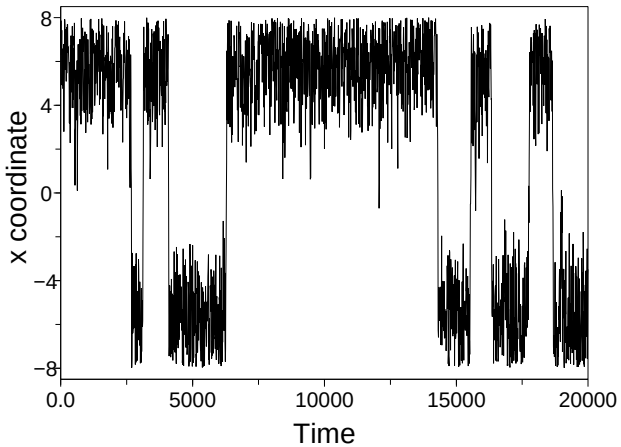
Metastability: energetic and entropic barriers

A two-dimensional schematic picture







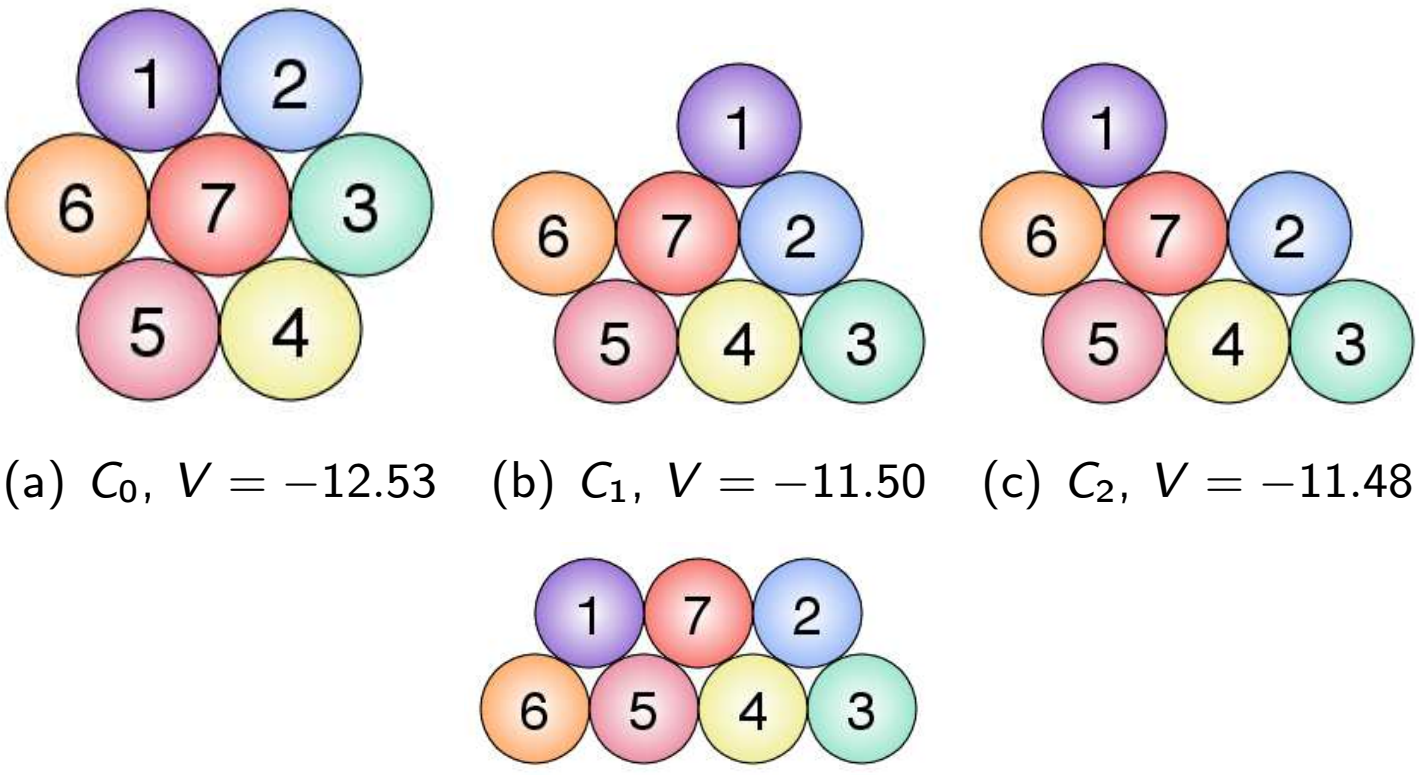


⟶ • Slow convergence of trajectorial averages

• Transitions between metastable states are rare events

A toy example in material sciences

The 7 atoms Lennard Jones cluster in 2D.



(a) C_0, $V = -12.53$ (b) C_1, $V = -11.50$ (c) C_2, $V = -11.48$

(d) C_3, $V = -11.40$

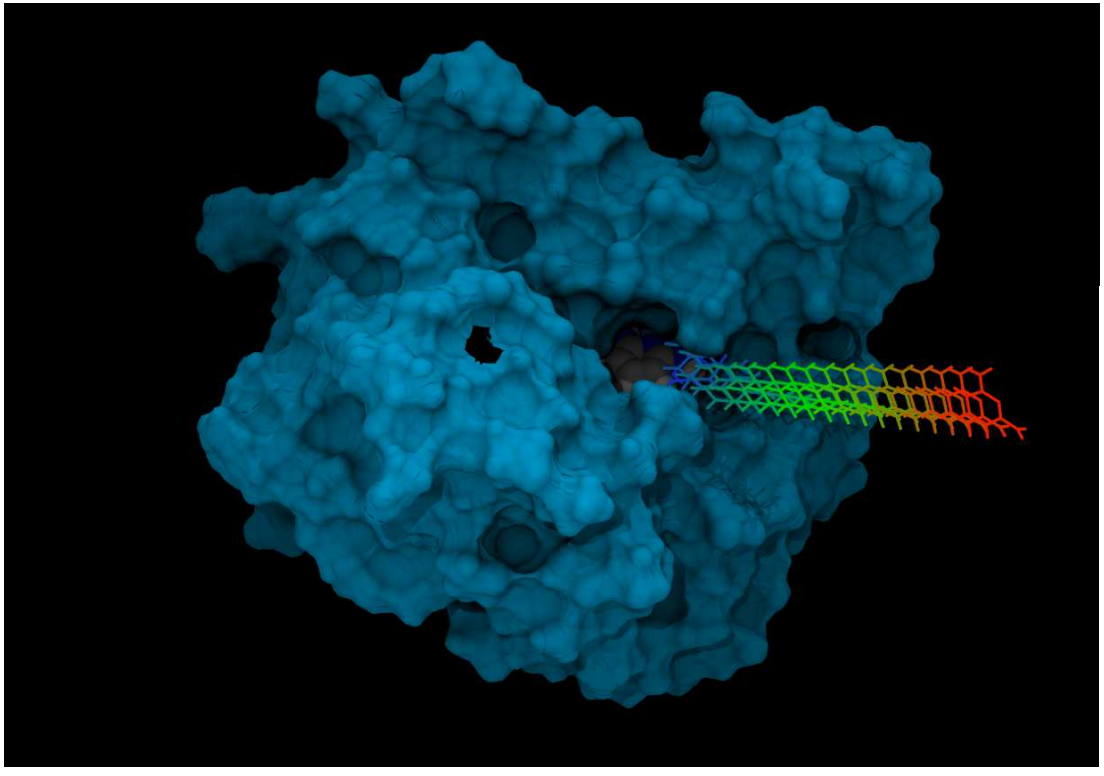
Figure: Low energy conformations of the Lennard-Jones cluster.

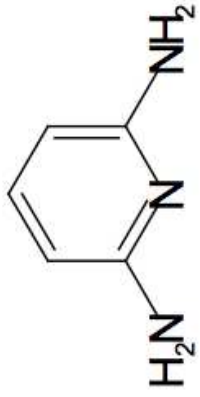
$\longrightarrow$ simulation

Simulations of biological systems

Unbinding of a ligand from a protein

(Diaminopyridine-HSP90, Courtesy of SANOFI)





Elementary time-step for the molecular dynamics $= 10^{-15}\,\mathrm{s}$

Dissociation time $= 0.5\,\mathrm{s}$

Challenge: bridge the gap between timescales

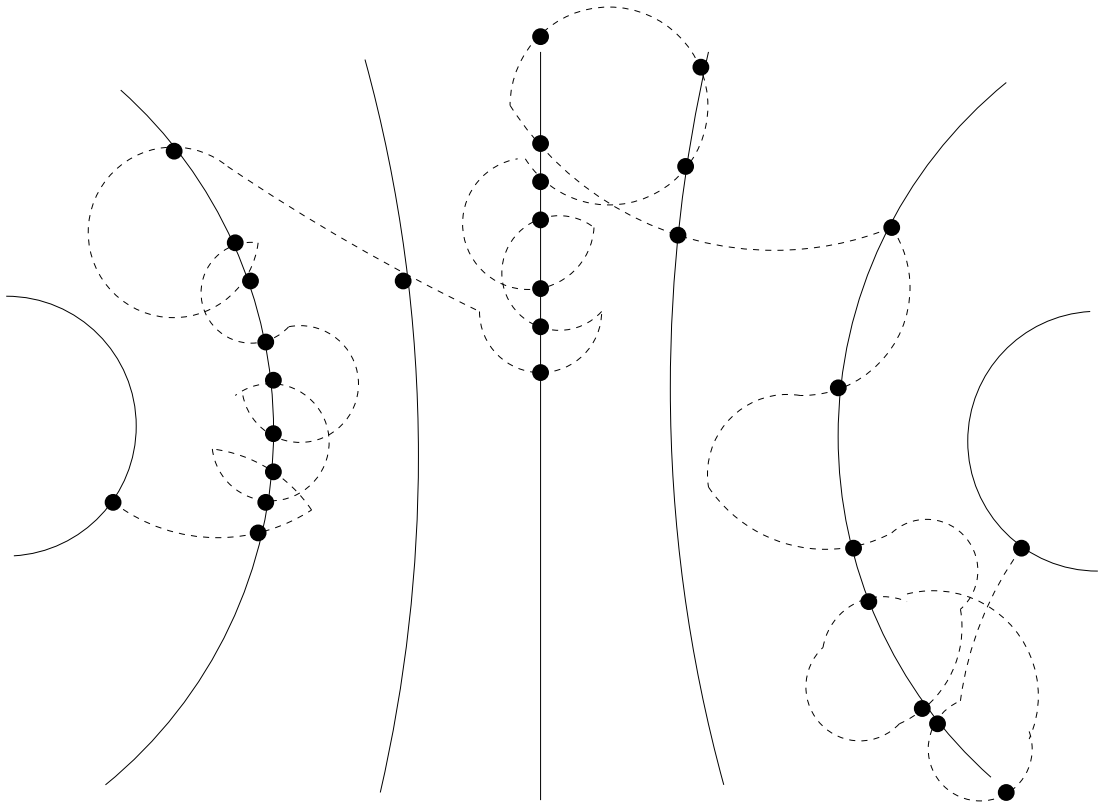
Introduction

Outline of the talk:

1. Adaptive biasing techniques: These are numerical methods to compute thermodynamic quantities, and in particular free energy differences.
   Mathematical tool: Entropy techniques and Logarithmic Sobolev Inequalities.
2. The Parallel Replica dynamics: This is one instance of an algorithm to generate efficiently metastable dynamics.
   Mathematical tool: Quasi Stationary Distributions.

Underlying question: how to properly define and quantify metastability ? Various answers: (i) rate of convergence to equilibrium; (ii) exit time from metastable states; (iii) decorrelation time; (iv) asymptotic variance of estimators.

Free energy and adaptive biasing techniques

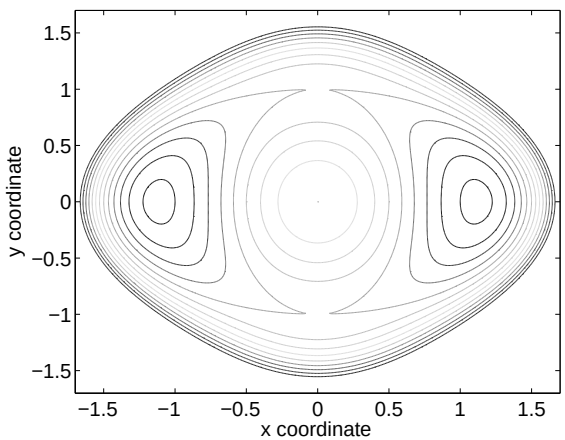


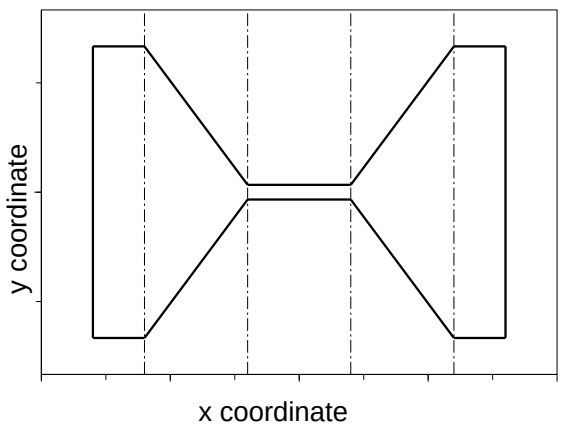
Adaptive biasing techniques

We suppose in this part that we know a slow variable of dimension 1: $\xi(\boldsymbol{X}_t)$, where $\xi : \mathbb{R}^d \to \mathbb{T}$ is a so-called reaction coordinate.

This reaction coordinate will be used to bias the dynamics (adaptive importance sampling technique).

For example, in the 2D simple examples: $\xi(x, y) = x$.





Adaptive biasing techniques

Let us introduce two probability measures associated to μ and ξ:

- The image of the measure μ by ξ:

$$\xi_* \mu\,(dz) = \exp(-\beta A(z))\,dz$$

where the free energy A is defined by:

$$A(z) = -\beta^{-1} \ln \left(\int_{\Sigma(z)} e^{-\beta V} \delta_{\xi(x)-z}(dx) \right),$$

with $\Sigma(z) = \{x,\, \xi(x) = z\}$ is a (smooth) submanifold of $\mathbb{R}^d$, and $\delta_{\xi(x)-z}(dx)\,dz = dx$.

- The probability measure μ conditioned to $\xi(x) = z$:

$$\mu_{\Sigma(z)}(dx) = \frac{\exp(-\beta V(x))\,\delta_{\xi(x)-z}(dx)}{\exp(-\beta A(z))}.$$

Adaptive biasing techniques

In the simple case $\xi(x, y) = x$, we have:

- The image of the measure μ by ξ:

$$\xi_* \mu\,(dx) = \exp(-\beta A(x))\, dx$$

where the free energy A is defined by:

$$A(x) = -\beta^{-1} \ln \left(\int e^{-\beta V(x,y)} dy \right),$$

and $\Sigma(x) = \{(x, y),\ y \in \mathbb{R}\}$.

- The probability measure μ conditioned to $\xi(x, y) = x$:

$$\mu_{\Sigma(x)}(dy) = \frac{\exp(-\beta V(x, y))\, dy}{\exp(-\beta A(x))}.$$

Adaptive biasing techniques

The bottom line of adaptive methods is the following: for "well chosen" ξ the potential $V - A \circ \xi$ is less rugged than V. Indeed, by construction $\xi_* \exp(-\beta(V - A \circ \xi)) = 1_{\mathbb{T}}$.

Problem: A is unknown ! Idea: use a time dependent potential of the form

$$\mathcal{V}_t(x) = V(x) - A_t(\xi(x))$$

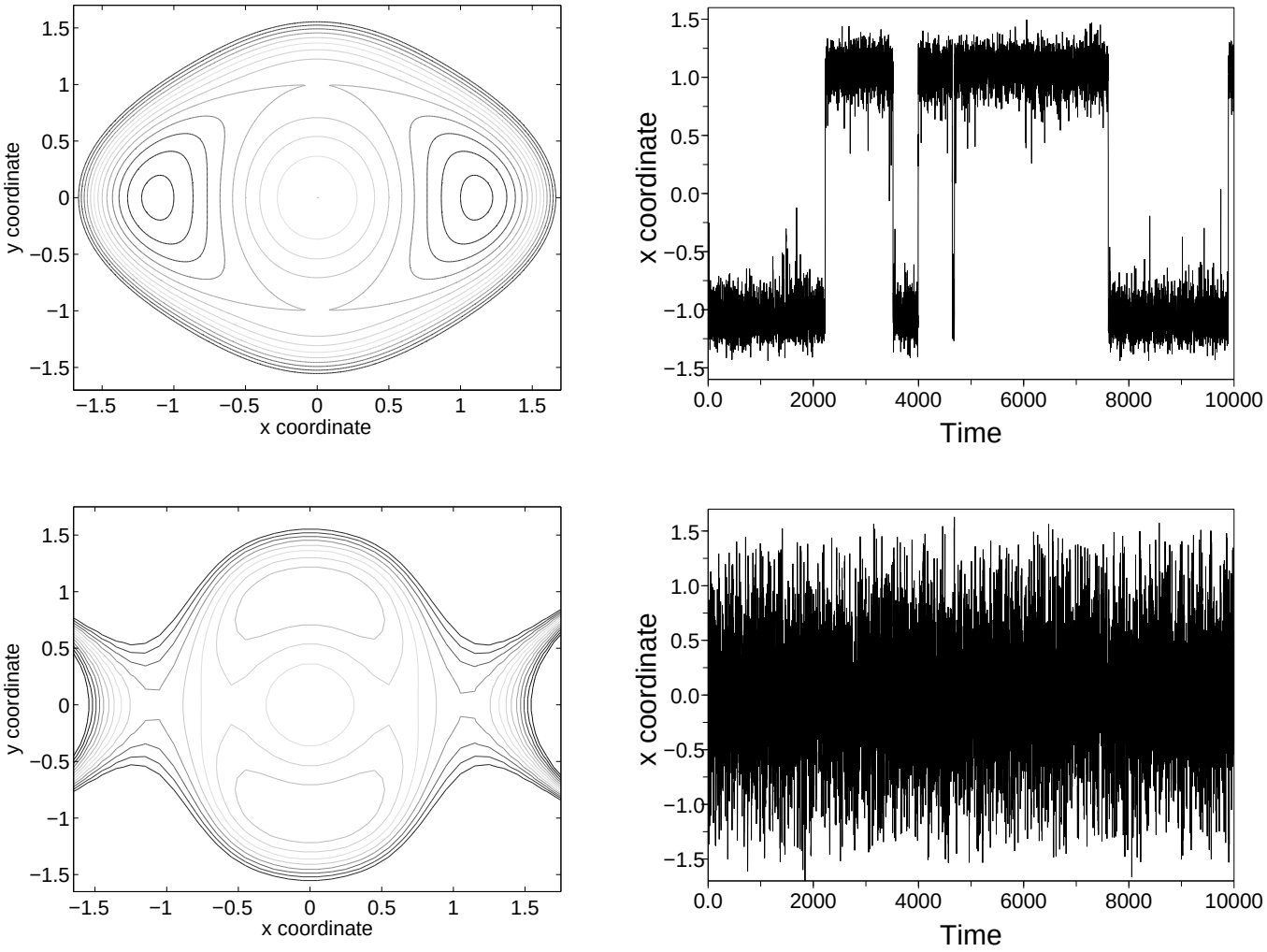
where A_t is an approximation at time t of A, given the configurations visited so far.

Hopes:

- build a dynamics which goes quickly to equilibrium,
- compute free energy profiles.

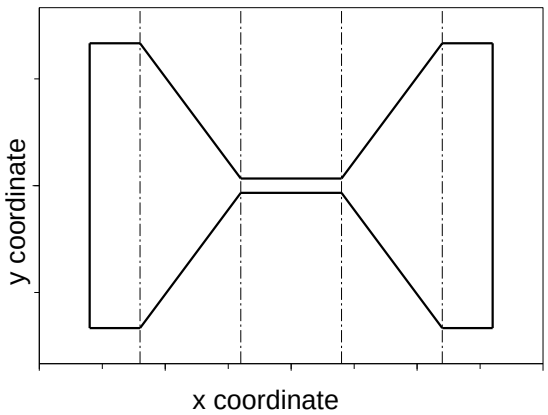
Wang-Landau, ABF, metadynamics: *Darve, Pohorille, Hénin, Chipot, Laio, Parrinello, Wang, Landau,...*

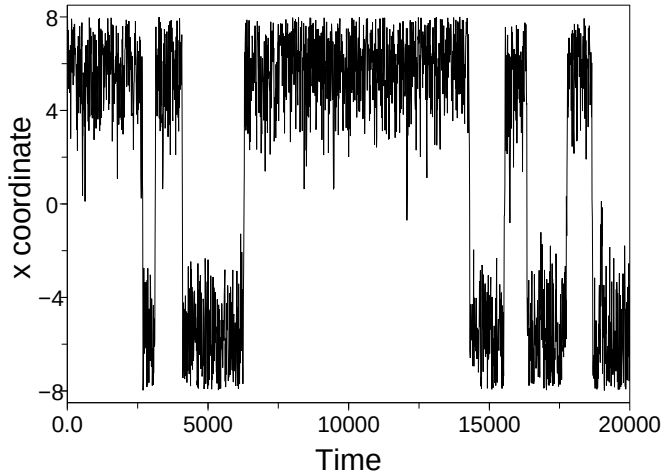
Adaptive biasing techniques

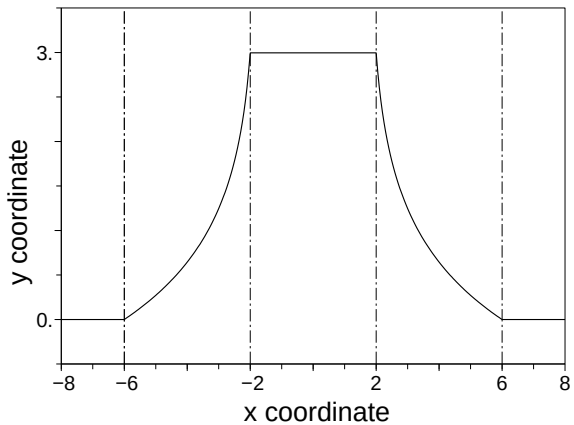


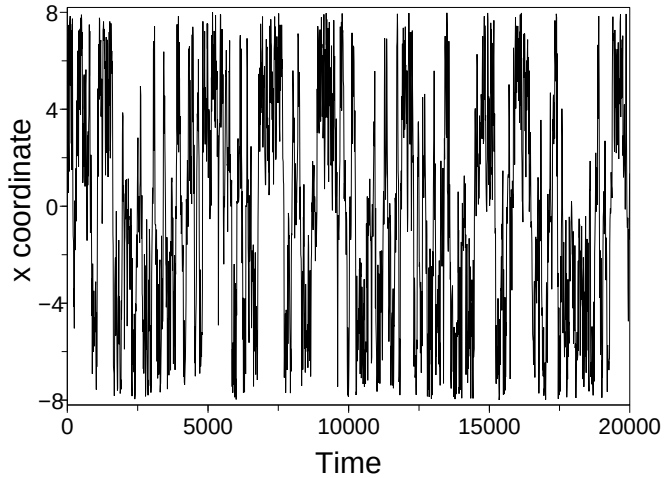
A 2D example of a free energy biased trajectory: energetic barrier.

Adaptive biasing techniques









A 2D example of a free energy biased trajectory: entropic barrier.

The ABF method

How to update A_t ? Two methods depending on wether A'_t (Adaptive Biasing Force) or A_t (Adaptive Biasing Potential) is approximated.

For the Adaptive Biasing Force (ABF) method, the idea is to use the formula

$$A'(z) = \frac{\int \left(\frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right) e^{-\beta V} \, \delta_{\xi(x)-z}(dx)}{\int e^{-\beta V} \, \delta_{\xi(x)-z}(dx)}$$

$$= \int f \, d\mu_{\Sigma(z)} = \mathbb{E}_\mu(f(\boldsymbol{X})|\xi(\boldsymbol{X}) = z).$$

The mean force $A'(z)$ is the mean of f with respect to $\mu_{\Sigma(z)}$.

The ABF method

In the simple case $\xi(x,y) = x$, remember that

$$A(x) = -\beta^{-1} \ln \left(\int e^{-\beta V(x,y)} dy \right),$$

so that

$$A'(x) = \frac{\int \partial_x V \, e^{-\beta V(x,y)} \, dy}{\int e^{-\beta V(x,y)} \, dy}$$

$$= \int \partial_x V \, d\mu_{\Sigma(x)}.$$

Notice that actually, whatever A_t is,

$$A'(z) = \frac{\int f \, e^{-\beta(V - A_t \circ \xi)} \, \delta_{\xi(x)-z}(dx)}{\int e^{-\beta(V - A_t \circ \xi)} \, \delta_{\xi(x)-z}(dx)}.$$

The ABF method

Thus, we would like to simulate:

$$\begin{cases} d\boldsymbol{X}_t = -\nabla(V - A \circ \xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t, \\ A'(z) = \mathbb{E}_\mu \left(f(\boldsymbol{X}) | \xi(\boldsymbol{X}) = z \right) \end{cases}$$

but A is unknown...

The ABF method

The ABF dynamics is then:

$$\begin{cases} d\boldsymbol{X}_t = -\nabla(V - A_t \circ \xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t, \\ A_t'(z) = \mathbb{E}\left(f(\boldsymbol{X}_t) | \xi(\boldsymbol{X}_t) = z\right). \end{cases}$$

The ABF method

The ABF dynamics is then:

$$\begin{cases} d\boldsymbol{X}_t = -\nabla(V - A_t \circ \xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}}\, d\boldsymbol{W}_t, \\ A_t'(z) = \mathbb{E}\left(f(\boldsymbol{X}_t) | \xi(\boldsymbol{X}_t) = z\right). \end{cases}$$

The associated (nonlinear) Fokker-Planck equation writes:

$$\begin{cases} \partial_t \psi = \operatorname{div}\left(\nabla(V - A_t \circ \xi)\psi + \beta^{-1}\nabla\psi\right), \\ A_t'(z) = \dfrac{\displaystyle\int f\, \psi\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})}{\displaystyle\int \psi\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})}, \end{cases}$$

where $\boldsymbol{X}_t \sim \psi(t, \boldsymbol{x})\, d\boldsymbol{x}$. $\longrightarrow$ simulation

The ABF method

The ABF dynamics is then:

$$\begin{cases} d\boldsymbol{X}_t = -\nabla(V - A_t \circ \xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t, \\ A_t'(z) = \mathbb{E}\left(f(\boldsymbol{X}_t) | \xi(\boldsymbol{X}_t) = z\right). \end{cases}$$

The associated (nonlinear) Fokker-Planck equation writes:

$$\begin{cases} \partial_t \psi = \operatorname{div}\left(\nabla(V - A_t \circ \xi)\psi + \beta^{-1}\nabla\psi\right), \\ A_t'(z) = \dfrac{\displaystyle\int f\, \psi\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})}{\displaystyle\int \psi\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})}, \end{cases}$$

where $\boldsymbol{X}_t \sim \psi(t, \boldsymbol{x})\, d\boldsymbol{x}$. $\longrightarrow$ simulation

Questions: Does A_t' converge to A' ? What did we gain compared to the original gradient dynamics ?

Longtime convergence and entropy (1)

Recall the original gradient dynamics:

$$d\boldsymbol{Q}_t = -\nabla V(\boldsymbol{Q}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t.$$

The associated (linear) Fokker-Planck equation writes:

$$\partial_t \phi = \mathrm{div}\, \left(\nabla V \phi + \beta^{-1} \nabla \phi \right).$$

where $\boldsymbol{Q}_t \sim \phi(t, \boldsymbol{q})\, d\boldsymbol{q}$.

The metastable behaviour of $\boldsymbol{Q}_t$ is related to the multimodality of μ, which can be quantified through the rate of convergence of ϕ to $\phi_\infty = Z^{-1} \exp(-\beta V)$.

A classical approach for partial differential equations (PDEs): entropy techniques.

Longtime convergence and entropy (2)

Notice that the Fokker-Planck equation rewrites

$$\partial_t \phi = \beta^{-1} \mathrm{div} \left(\phi_\infty \nabla \left(\frac{\phi}{\phi_\infty} \right) \right).$$

Let us introduce the entropy:

$$E(t) = H(\phi(t,\cdot)|\phi_\infty) = \int \ln \left(\frac{\phi}{\phi_\infty} \right) \phi.$$

We have (Csiszár-Kullback inequality):

$$\|\phi(t,\cdot) - \phi_\infty\|_{L^1} \leq \sqrt{2E(t)}.$$

Longtime convergence and entropy (3)

$$\begin{aligned}\frac{dE}{dt} &= \int \ln\left(\frac{\phi}{\phi_\infty}\right)\partial_t\phi \\ &= \beta^{-1}\int \ln\left(\frac{\phi}{\phi_\infty}\right)\operatorname{div}\left(\phi_\infty\nabla\left(\frac{\phi}{\phi_\infty}\right)\right) \\ &= -\beta^{-1}\int\left|\nabla\ln\left(\frac{\phi}{\phi_\infty}\right)\right|^2\phi =: -\beta^{-1}I(\phi(t,\cdot)|\phi_\infty).\end{aligned}$$

If V is such that the following Logarithmic Sobolev inequality (LSI(R)) holds: $\forall\phi$ pdf,

$$H(\phi|\phi_\infty) \leq \frac{1}{2R}I(\phi|\phi_\infty)$$

then $E(t) \leq E(0)\exp(-2\beta^{-1}Rt)$ and thus ϕ converges to ϕ_∞ exponentially fast with rate $\beta^{-1}R$.

Metastability $\Longleftrightarrow$ small R

Convergence of ABF (1)

A convergence result [TL, M. Rousset, G. Stoltz, Nonlinearity 2008]: Recall the ABF Fokker-Planck equation:

$$\begin{cases} \partial_t \psi = \mathrm{div}\, \left(\nabla (V - A_t \circ \xi) \psi + \beta^{-1} \nabla \psi \right), \\ A_t'(z) = \dfrac{\int f\, \psi\, \delta_{\xi(x)-z}(dx)}{\int \psi\, \delta_{\xi(x)-z}(dx)}. \end{cases}$$

Suppose:

(H1) "Ergodicity" of the microscopic variables: the conditional probability measures $\mu_{\Sigma(z)}$ satisfy a LSI(ρ),

(H2) Bounded coupling: $\left\| \nabla_{\Sigma(z)} f \right\|_{L^\infty} < \infty$,

then

$$\| A_t' - A' \|_{L^2} \leq C \exp(-\beta^{-1} \min(\rho, r) t).$$

The rate of convergence is limited by:

- the rate r of convergence of $\overline{\psi} = \int \psi\, \delta_{\xi(x)-z}(dx)$ to $\overline{\psi_\infty}$,
- the LSI constant ρ (the real limitation).

Convergence of ABF (2)

In summary:

- Original gradient dynamics: $\exp(-\beta^{-1}Rt)$ where R is the LSI constant for μ;
- ABF dynamics: $\exp(-\beta^{-1}\rho t)$ where ρ is the LSI constant for the conditioned probability measures $\mu_{\Sigma(z)}$.

If ξ is well chosen, $\rho \gg R$: the free energy can be computed very efficiently.

Two ingredients of the proof:

(1) The marginal $\overline{\psi}(t,z) = \int \psi(t,\boldsymbol{x})\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})$ satisfies a closed PDE:

$$\partial_t \overline{\psi} = \beta^{-1} \partial_{z,z} \overline{\psi} \text{ on } \mathbb{T},$$

and thus, $\overline{\psi}$ converges towards $\overline{\psi_\infty} \equiv 1$, with exponential speed $C\exp(-4\pi^2\beta^{-1}t)$. (Here, $r = 4\pi^2$).

Convergence of ABF (3)

(2) The total entropy can be decomposed as [N. Grunewald, F. Otto, C. Villani, M. Westdickenberg, Ann. IHP, 2009]:

$$E = E_M + E_m$$

where

The total entropy is $E = H(\psi|\psi_\infty)$,

The macroscopic entropy is $E_M = H(\overline{\psi}|\overline{\psi_\infty})$,

The microscopic entropy is

$$E_m = \int H\Big(\psi(\cdot|\xi(x) = z)\Big|\psi_\infty(\cdot|\xi(x) = z)\Big)\,\overline{\psi}(z)\,dz.$$

We already know that E_M goes to zero: it remains only to consider E_m...

Discretization of ABF

Discretization of adaptive methods can be done using two (complementary) approaches:

- Use empirical means over many replicas (interacting particle system):

$$\mathbb{E}(f(X_t)|\xi(X_t)=z) \simeq \frac{\sum_{m=1}^N f(X_t^{m,N})\,\delta^\alpha(\xi(X_t^{m,N})-z)}{\sum_{m=1}^N \delta^\alpha(\xi(X_t^{m,N})-z)}.$$

  This approach is easy to parallelize, flexible (selection mechanisms) and efficient in cases with multiple reactive paths. [TL, M. Rousset, G. Stoltz, 2007; C. Chipot, TL, K. Minoukadeh, 2010 ; TL, K. Minoukadeh, 2010]

- Use trajectorial averages along a single path:

$$\mathbb{E}(f(X_t)|\xi(X_t)=z) \simeq \frac{\int_0^t f(X_s)\,\delta^\alpha(\xi(X_s)-z)\,ds}{\int_0^t \delta^\alpha(\xi(X_s)-z)\,ds}.$$

  The longtime behavior is much more difficult to analyze [Talk by Fort]

Back to the original problem

How to use free energy to compute canonical averages $\int \varphi d\mu = Z^{-1} \int \varphi e^{-\beta V}$?

- Importance sampling:

$$\int \varphi \, d\mu = \frac{\int \varphi e^{-\beta A \circ \xi} \, Z_A^{-1} e^{-\beta(V - A\circ\xi)}}{\int e^{-\beta A \circ \xi} \, Z_A^{-1} e^{-\beta(V - A\circ\xi)}}.$$

- Conditioning:

$$\int \varphi \, d\mu = \frac{\int_z \left(\int_{\Sigma(z)} \varphi \, d\mu_{\Sigma(z)} \right) e^{-\beta A(z)} \, dz}{\int_z e^{-\beta A(z)} \, dz}.$$

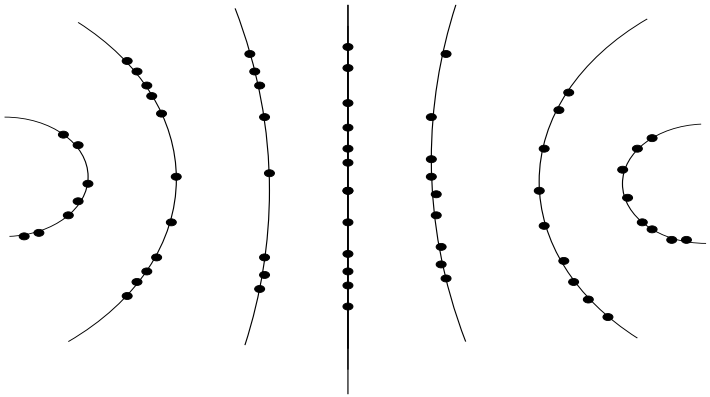
This requires the sampling of the conditional probability measure $\mu_{\Sigma(z)}$ which can be done using projected Langevin dynamics [TL, M. Rousset, G. Stoltz, 2011].

Other techniques to compute thermodynamic quantities

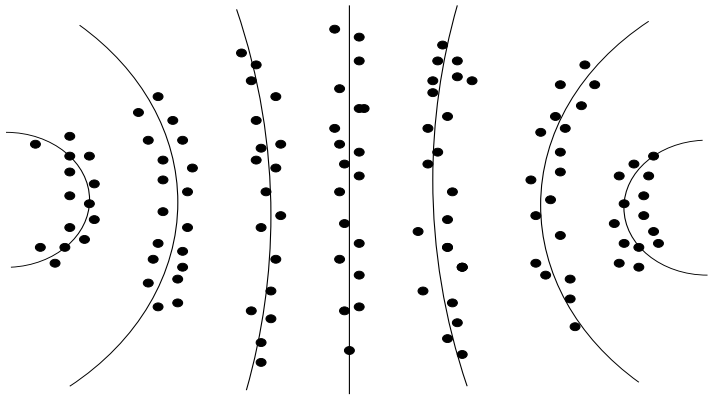
Other algorithms which are used in MD to sample efficiently μ:

- Modify the dynamics: Metropolis Hastings algorithms with well-chosen proposals [Talks by Jourdain and Ottobre], non-gradient forces,...
- Interacting replicas techniques: Parallel tempering
- Conditioning and constrained sampling: Thermodynamic integration
- Umbrella sampling and statistical reconstruction: Histogram methods
- Out of equilibirum methods: fluctuation relations *à la* Jarzynski-Crooks

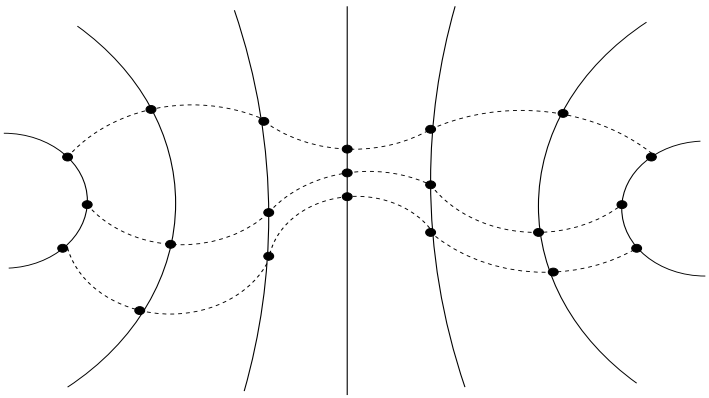
Free energy calculation methods



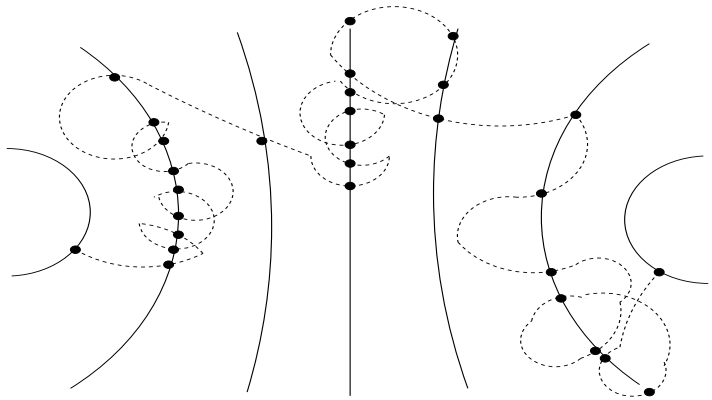
(a) Thermodynamic integration.



(b) Histogram method.



(c) Out of equilibrium dynamics.



(d) Adaptive dynamics.

Towards computational statistics

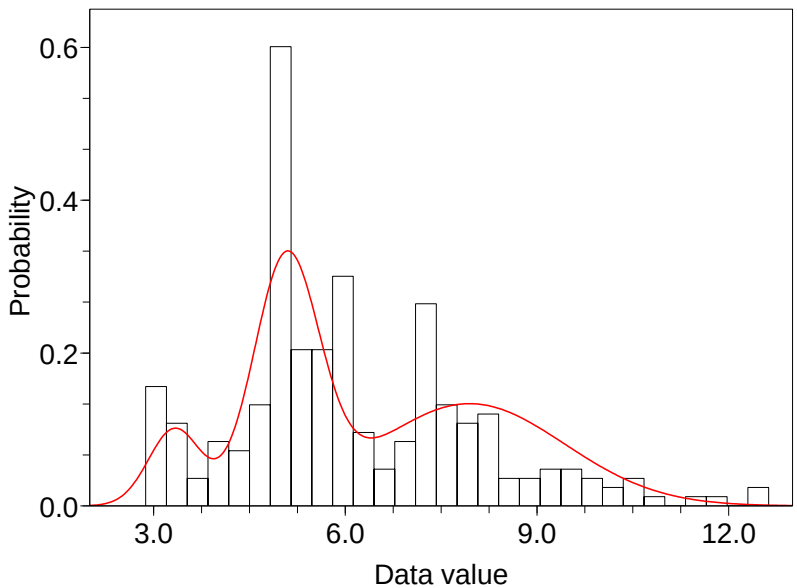
Adaptive biasing techniques can be used whenever the sampling of a multimodal measure is involved, for example for statistical inference in Bayesian statistics [N. Chopin, TL, G. Stoltz, 2011]: Sampling of posterior distributions using a MCMC ABF algorithm.

Question: What is ξ in this context ?

Test on *the fishery problem*

- The size of $N_{\mathrm{data}} = 256$ fishes are measured, and the corresponding histogram is approximated by a mixture of K Gaussians:

$$f(y \mid x) = \sum_{i=1}^{K} q_i \sqrt{\frac{v_i}{2\pi}} \exp\left(-\frac{v_i}{2}(y-\mu_i)^2\right).$$



Application to Bayesian statistics

- Parameters describing the mixture $(q_K = 1 - \sum_{i=1}^{K-1} q_i)$:
  $x = (q_1, \ldots, q_{K-1}, \mu_1, \ldots, \mu_K, v_1, \ldots, v_K) \in$
  $\mathcal{S}_{K-1} \times [\mu_{\min}, \mu_{\max}]^K \times [v_{\min}, +\infty) \subset \mathbb{R}^{3K-1}$, where
  $\mathcal{S}_{K-1} = \left\{(q_1, \ldots, q_{K-1}) \;\middle|\; 0 \leq q_i \leq 1,\ \sum_{i=1}^{K-1} q_i \leq 1\right\}$.
- Given the parameters, the likelihood of observing the data $\{y_i, 1 \leq i \leq N_{\text{data}}\}$ is

$$\Pi(y \,|\, x) = \prod_{d=1}^{N_{\text{data}}} f(y_d \,|\, x).$$

- The prior on the parameters is: $\mu_i \sim \mathcal{N}(M, R^2/4)$, $v_i \sim \text{Gamma}(a, \beta)$ with $\beta \sim \text{Gamma}(g, h)$ and $(q_1, \ldots, q_K) \sim \text{Dirichlet}_K(1, \ldots, 1)$ for fixed values (M, R, a, g, h) (random beta model).

So actually $x = (q_1, \ldots, q_{K-1}, \mu_1, \ldots, \mu_K, v_1, \ldots, v_K, \beta)$.

Application to Bayesian statistics

Objective: sample the posterior distribution (distribution of the parameters given the observations):

$$\Pi(x|y) = \frac{\Pi(y|x)\,\mathrm{Prior}(x)}{\int \Pi(y|x)\,\mathrm{Prior}(x)\,dx} = Z^{-1}\exp(-V)$$

where

$$V = V_{\text{likelihood}} + V_{\text{prior}}$$

with $V_{\text{likelihood}} = \sum_{d=1}^{N_{\text{data}}} \ln\left[\sum_{i=1}^{K} q_i\sqrt{v_i}\exp\left(-\frac{v_i}{2}(y_d-\mu_i)^2\right)\right]$ and

$V_{\text{prior}} = \frac{2}{R^2}\sum_{i=1}^{K}(\mu_i - M)^2 - K\alpha\ln\beta + \beta\sum_{i=1}^{K} v_i - (a-1)\sum_{i=1}^{K}\ln v_i + h\beta - (g-1)\ln\beta.$

The posterior distribution is a multimodal measure (due to, but not only, invariance by permutation of the Gaussians).

Idea: Use ABF within a Metropolis Hastings algorithm. The biasing potential modifies the target probability measure in the acception-rejection step.

Algorithm: Metropolis Hastings-ABF

Iterate on $n \geq 0$:

1. Update the biasing potential by computing and then integrating $(A^{n+1})'$ (the conditional expectation of f at a fixed value of ξ).
2. Propose a move from x^n to $\bar{x}^{n+1}$ according to $T(x^n, \bar{x}^{n+1})$.
3. Acceptance ratio

$$r^n = \min\left(\frac{\pi_{A^{n+1}}(\bar{x}^{n+1})\, T(\bar{x}^{n+1}, x^n)}{\pi_{A^{n+1}}(x^n)\, T(x^n, \bar{x}^{n+1})}, 1\right),$$

   where the biased probability is
   $\pi_{A^{n+1}}(x) \propto \pi(x) \exp(A^{n+1}(\xi(x)))$.
4. Draw a random variable U^n uniformly distributed in $[0,1]$ $(U^n \sim \mathcal{U}[0,1])$.
   - 4.1 if $U^n \leq r^n$, accept the move and set $x^{n+1} = \bar{x}^{n+1}$;
   - 4.2 if $U^n > r^n$, reject the move and set $x^{n+1} = x^n$.

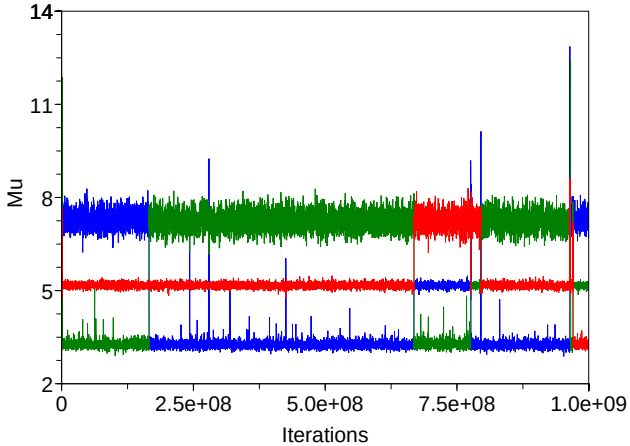
Numerical parameters

- The proposal density kernel $T(x, x')$ is a fixed Gaussian centered on x.
- Binning procedure and trajectorial average: mean force and bias in bin (z_i, z_{i+1})

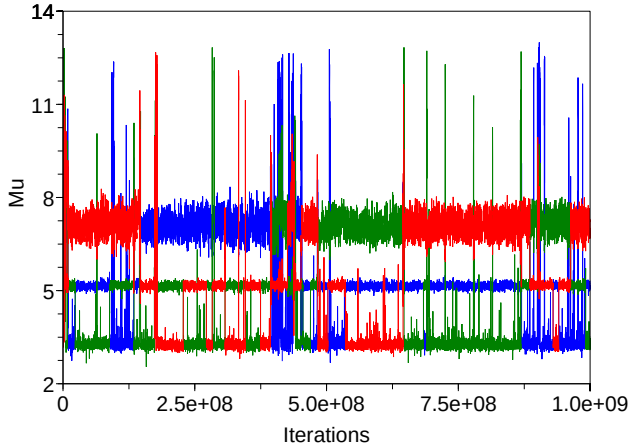
$$\Gamma_n^{\Delta z}(z) = \frac{\sum_{j=0}^{n} f(x_j)\, \mathbf{1}_{z_i \leq \xi(x^j) \leq z_{i+1}}}{\sum_{j=0}^{n} \mathbf{1}_{z_i \leq \xi(x^j) \leq z_{i+1}}}, \quad A_n^{\Delta z}(z) = \sum_{k=0}^{i-1} \Delta z\, \Gamma_n^{\Delta z}\left(k + \frac{1}{2}\Delta z\right)$$

- M is the mean of the data, R is the range of the data, $\alpha = 2$, $g = 0.2$ and $h = 100g/(\alpha R^2)$. We choose $K = 3$.

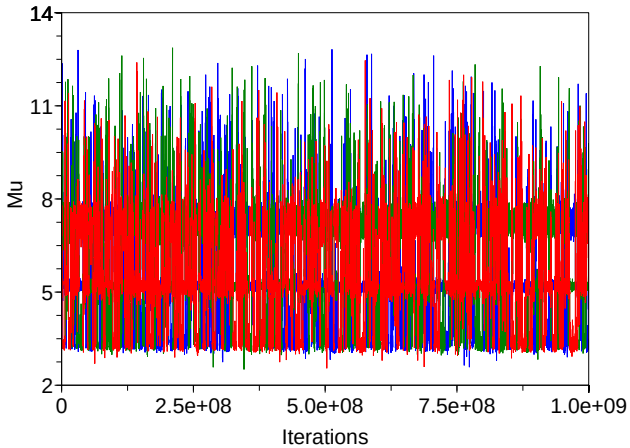
Question: Is there a good "reaction coordinate" $\xi(x)$?
Criteria: (i) How rapid is the convergence to equilibrium (convergence of the free energy and switching of the modes) ? ; (ii) How representative are the points simulated from the biased distribution (effective sample size) ?



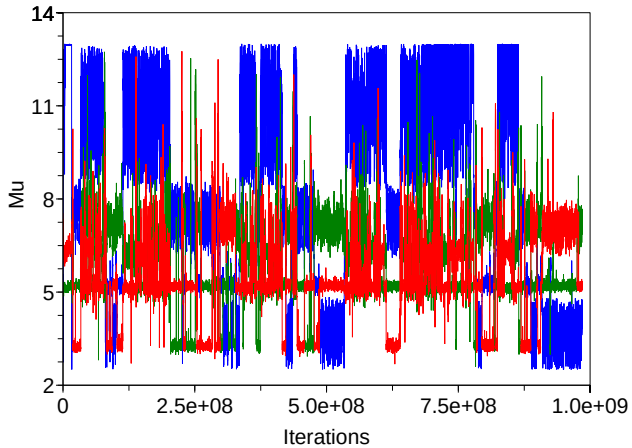
(a) No bias



(b) $\xi = q_1$



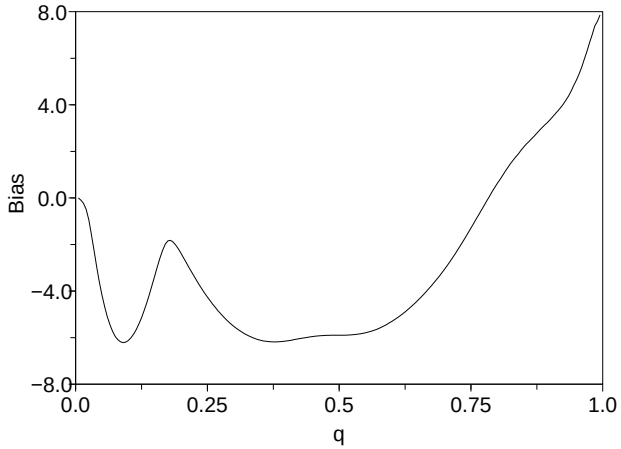
(c) $\xi = \beta$



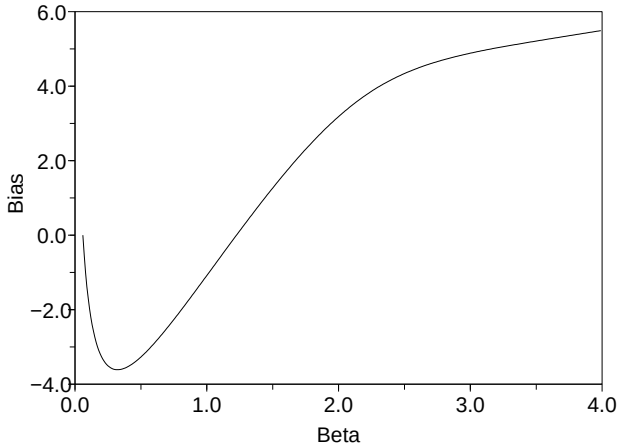
(d) $\xi = \mu$

Evolution of the parameters (μ_1, μ_2, μ_3).

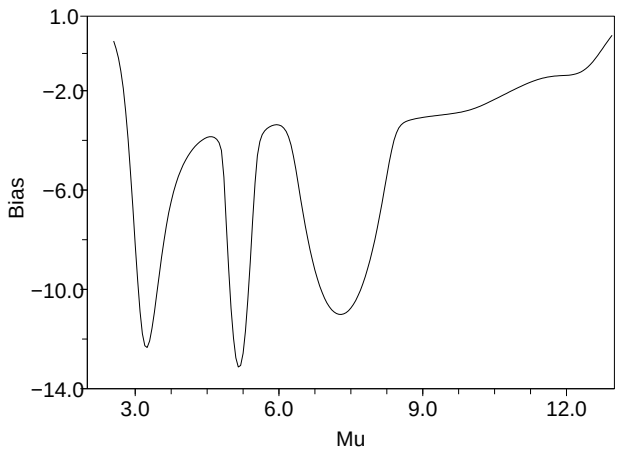
The reaction coordinate $\xi = \beta$ seems to be a good choice.



(a) $\xi = q_1$



(b) $\xi = \beta$

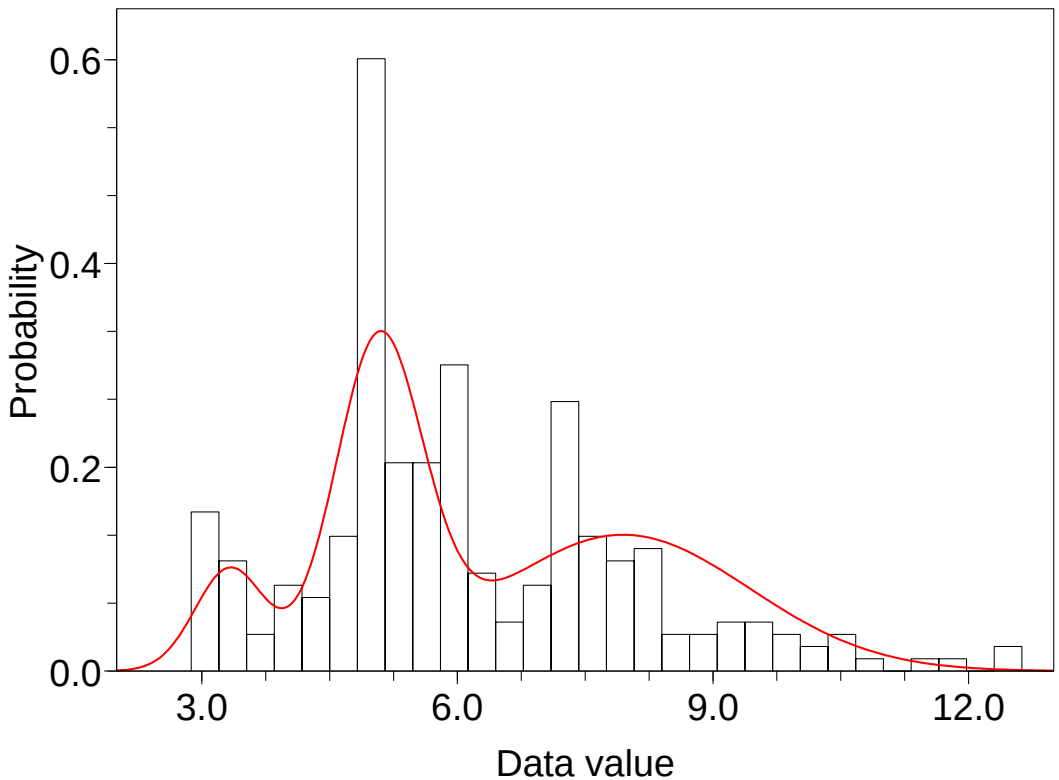


(c) $\xi = \mu$

Converged biases

The effective sample size for $\xi = \beta$ is approximately $0.18N$.

Comparison of the mixture with the datas



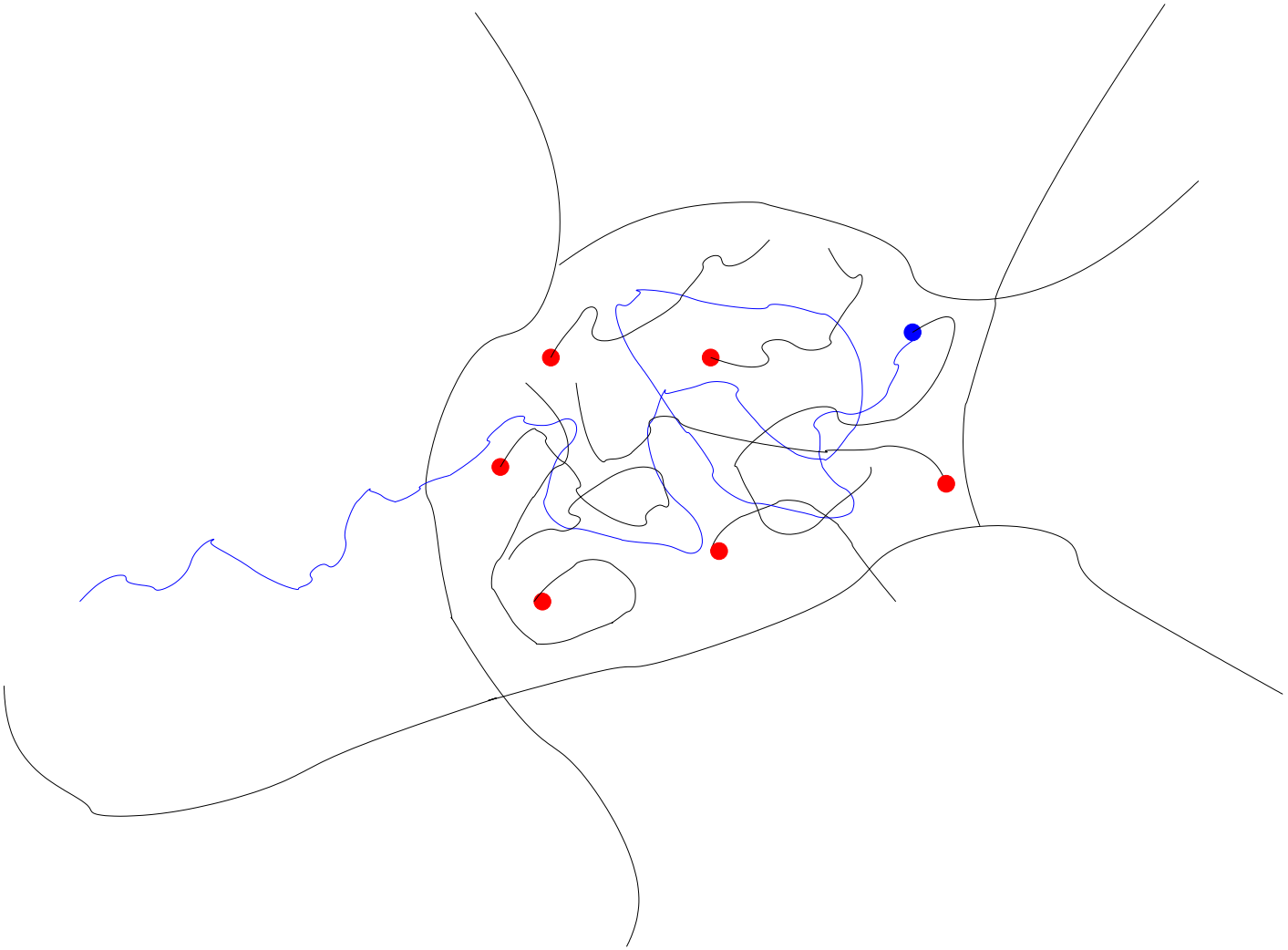
Application to Bayesian statistics: conclusions

For another set of data - the *Hidalgo stamp problem* - $\xi = \beta$ again seems to be a good choice.

Why does it work with $\xi = \beta$? (i) The bias is relatively small; (ii) Forcing large values of β is forcing large values of the variances, which allows for a mixing of the components.

Extension: Bayesian model choice. Look for the best number of components K. It seems that the bias (for $\xi = \beta$) for $K = 3$ is also a good bias for $K = 4$ and $K = 5$.

Accelerated dynamics and the Parallel Replica Algorithm



How to simulate long trajectories of a metastable process ?

For computing thermodynamics quantities, there is a clear classification of available methods, and the difficulties are now well understood (in particular for free energy computations, see for example [TL, Rousset, Stoltz, 2010]). On the opposite, computing efficiently dynamical quantities remains a challenge.

In the following, we again focus on the overdamped Langevin dynamics:

$$d\boldsymbol{X}_t = -\nabla V(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t.$$

Remember that a typical molecular dynamics trajectory is metastable.

How to take advantage of metastability in order to build efficient path sampling techniques ? $\longrightarrow$ simulation

Accelerated dynamics

The bottom line of the accelerated dynamics proposed by A. Voter in the late 90's is to get efficiently the state-to-state dynamics. Three algorithms: Parallel replica, Hyperdynamics, Temperature Accelerated Dynamics.

Let us consider the overdamped Langevin dyanmics:

$$d\boldsymbol{X}_t = -\nabla V(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t$$

and let assume that we are given a mapping

$$\mathcal{S} : \mathbb{R}^d \to \mathbb{N}$$

which to a configuration in $\mathbb{R}^d$ associates a state number. Think of a numbering of the wells of the potential V.

Objective: generate very efficiently a trajectory $(S_t)_{t\geq 0}$ which has (almost) the same law as $(\mathcal{S}(\boldsymbol{X}_t))_{t\geq 0}$.

The Quasi-Stationary Distribution

How to take advantage of metastability to build efficient sampling techniques ?

Let us consider a metastable state W, and

$$T_W = \inf\{t \geq 0, \boldsymbol{X}_t \notin W\}.$$

Lemma: Let $\boldsymbol{X}_t$ start in the well W. Then there exists a probability distribution ν with support W such that

$$\lim_{t\to\infty} \mathcal{L}(\boldsymbol{X}_t | T_W > t) = \nu.$$

Remark: Rigorous definition of a metastable state:

exit time $\gg$ local equilibration time

The Quasi-Stationary Distribution

Property 1: $\forall t > 0$, $\forall A \subset W$,

$$\nu(A) = \frac{\displaystyle\int_W \mathbb{P}(\boldsymbol{X}_t^{\boldsymbol{x}} \in A,\, t < T_W^{\boldsymbol{x}})\, \nu(d\boldsymbol{x})}{\displaystyle\int_W \mathbb{P}(t < T_W^{\boldsymbol{x}})\, \nu(d\boldsymbol{x})}.$$

If $\boldsymbol{X}_0 \sim \nu$ and if $(\boldsymbol{X}_s)_{0 \leq s \leq t}$ has not left the well, then $\boldsymbol{X}_t \sim \nu$.

Property 2: Let $L = -\nabla V \cdot \nabla + \beta^{-1}\Delta$ be the infinitesimal generator of $(\boldsymbol{X}_t)$. Then the density u_1 of ν ($d\nu = u_1(\boldsymbol{x})d\boldsymbol{x}$) is the first eigenfunction of $L^* = \mathrm{div}\,(\nabla V + \beta^{-1}\nabla)$ with absorbing boundary conditions:

$$\begin{cases} L^* u_1 = -\lambda_1 u_1 \text{ on } W, \\ \quad u_1 = 0 \text{ on } \partial W. \end{cases}$$

The Quasi-Stationary Distribution

Property 3: If $\boldsymbol{X}_0 \sim \nu$ then,

- the first exit time T_W from W is exponentially distributed with parameter λ_1 ;
- T_W is independent of the first hitting point $\boldsymbol{X}_{T_W}$ on ∂W ;
- the exit point distribution is proportional to $-\partial_n u_1$: for all smooth test functions $\varphi : \partial W \to \mathbb{R}$,

$$\mathbb{E}^\nu(\varphi(\boldsymbol{X}_{T_W})) = -\frac{\displaystyle\int_{\partial W} \varphi\, \partial_n u_1 \, d\sigma}{\beta\lambda \displaystyle\int_W u_1(x)\, dx}.$$

Remark: This is reminiscent of what is assumed in Transition State Theory (first order kinetics).

back to Hyper

Escaping from a metastable state

How to use these properties to build efficient algorithms ?

Assume that the stochastic process remained trapped for a very long time in a metastable state W. How to accelerate the escape event from W, in a statistically consistent way ?

Remark: In practice, one needs to:

- Choose the partition of the domain into (metastable) states;
- Associate to each state an equilibration time (a.k.a. *decorrelation time*).

These are not easy tasks... we will come back to that.

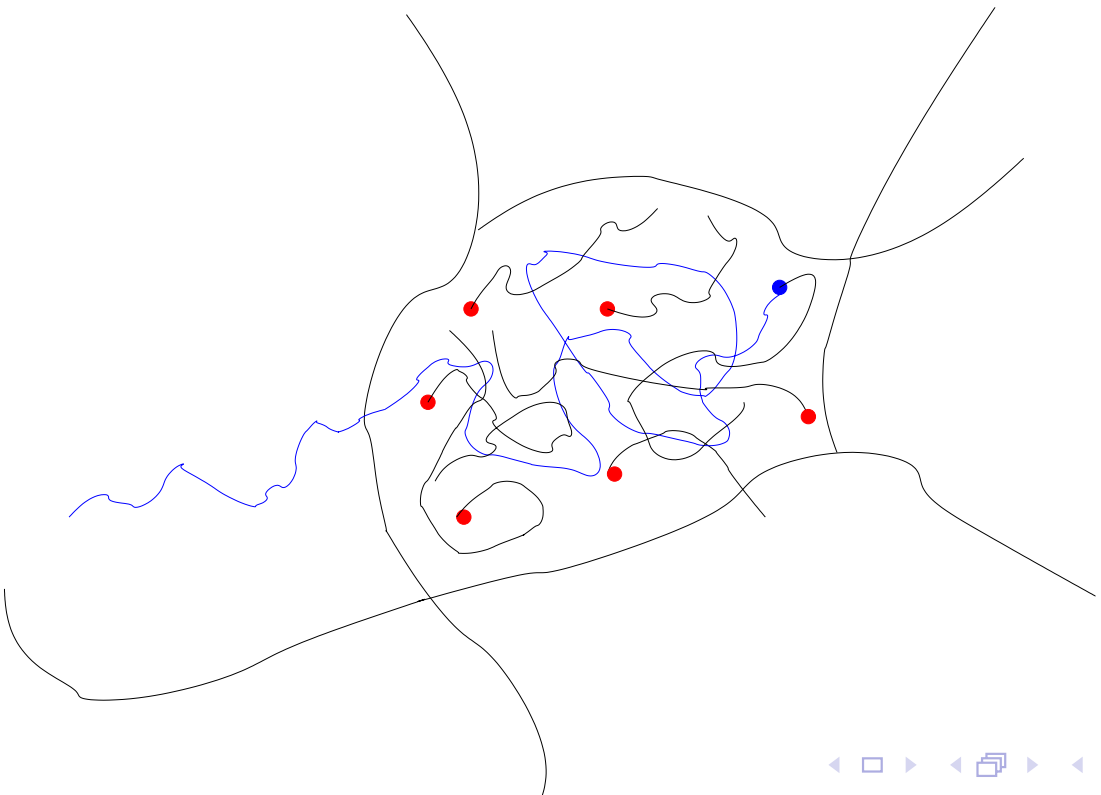
Remark: All the algorithms below equally apply to the Langevin dynamics but the extensions of the mathematical results to the Langevin dynamics are not straightforward...

The Parallel Replica Algorithm

Idea: perform many independent exit events in parallel.

Two steps:

- Distribute N independent initial conditions in W according to the QSD ν ;
- Consider the first exit event, and multiply it by the number of replicas.



The Parallel Replica Algorithm

Why is it consistent ?

- Exit time is independent of exit point so that

$$\boldsymbol{X}^{I_0}_{T_W^{I_0}} \overset{\mathcal{L}}{=} \boldsymbol{X}^{1}_{T_W^{1}},$$

  where $I_0 = \arg\min_i(T_W^i)$;

- Exit times are i.i.d. exponentially distributed so that, for all N,

$$N \min(T_W^1, \ldots, T_W^N) \overset{\mathcal{L}}{=} T_W^1.$$

Remark: In practice, discrete time processes are used. Exponential laws become geometric, and one can adapt the algorithm by using the identity [Aristoff, TL, Simpson, 2014]: if τ_i i.i.d. with geometric law,

$$N[\min(\tau_1, \ldots, \tau_N) - 1] + \min[i \in \{1, \ldots, N\}, \tau_i = \min(\tau_1, \ldots, \tau_N)] \overset{\mathcal{L}}{=} \tau_1.$$

The Parallel Replica Algorithm

The full algorithm is in three steps:

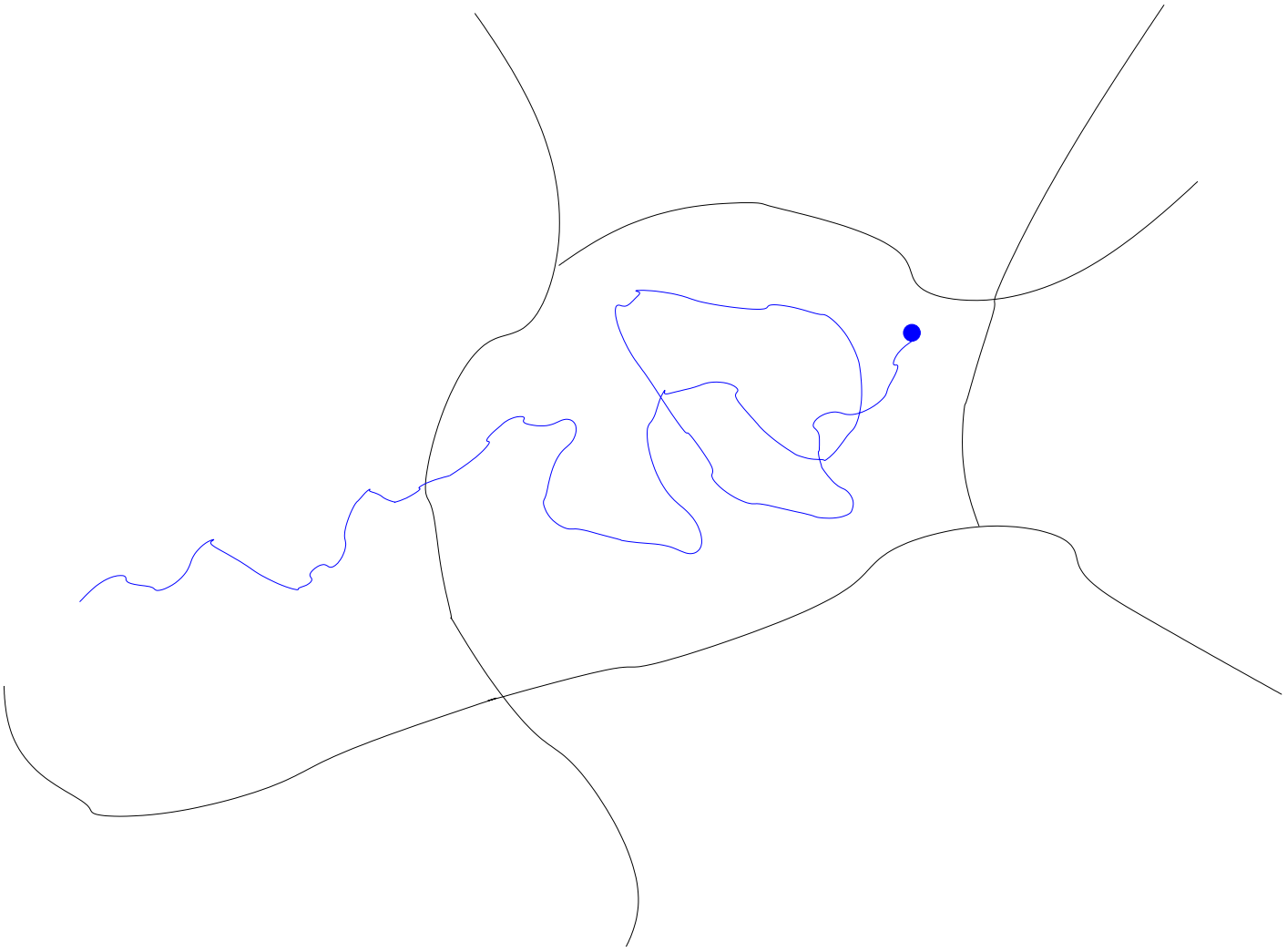
- Decorrelation step
- Dephasing step
- Parallel step

The Parallel Replica Algorithm

Decorrelation step: run the dynamics on a reference walker...

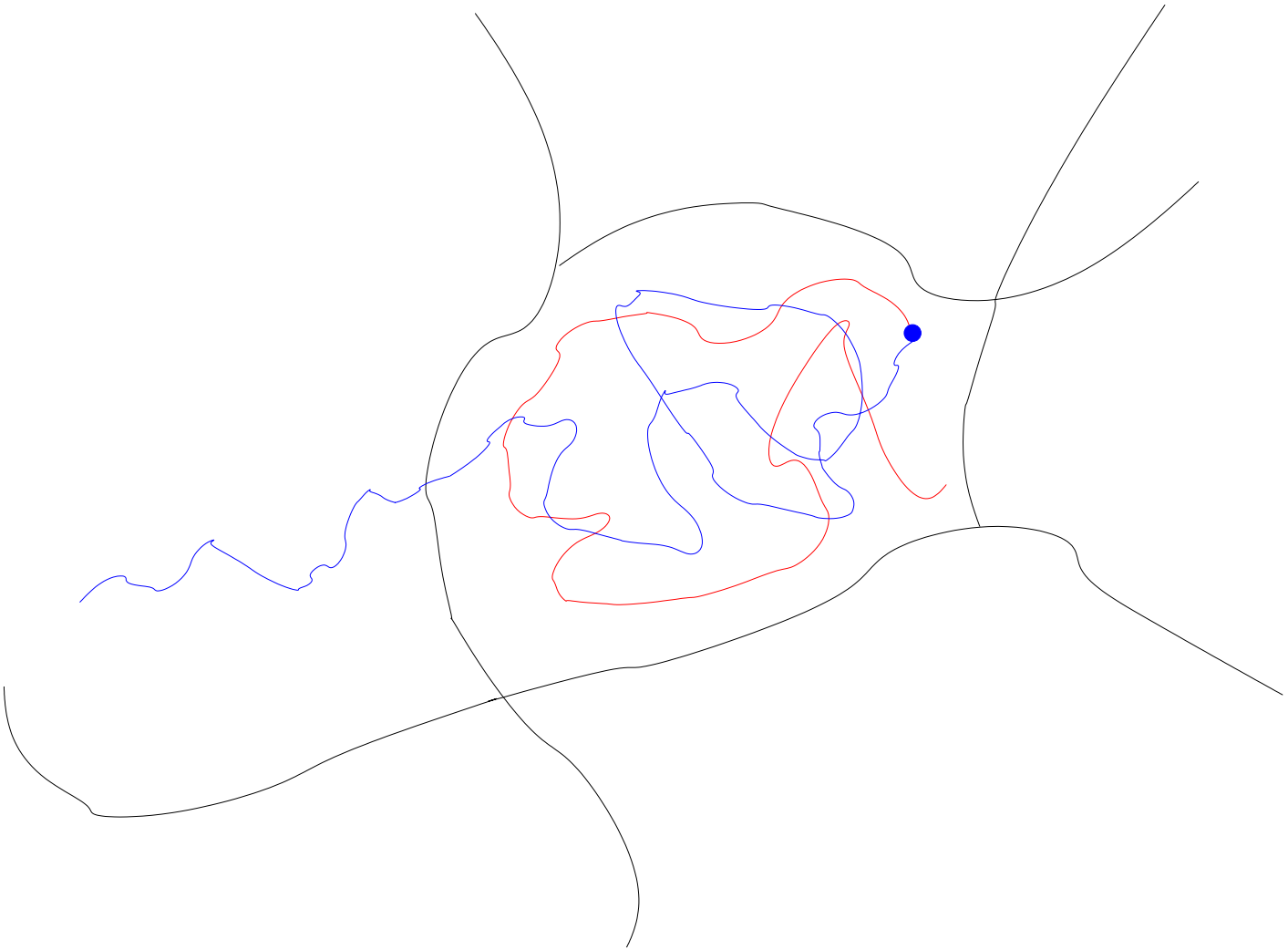
The Parallel Replica Algorithm

Decorrelation step: ... until it remains trapped for a time τ_{corr}.



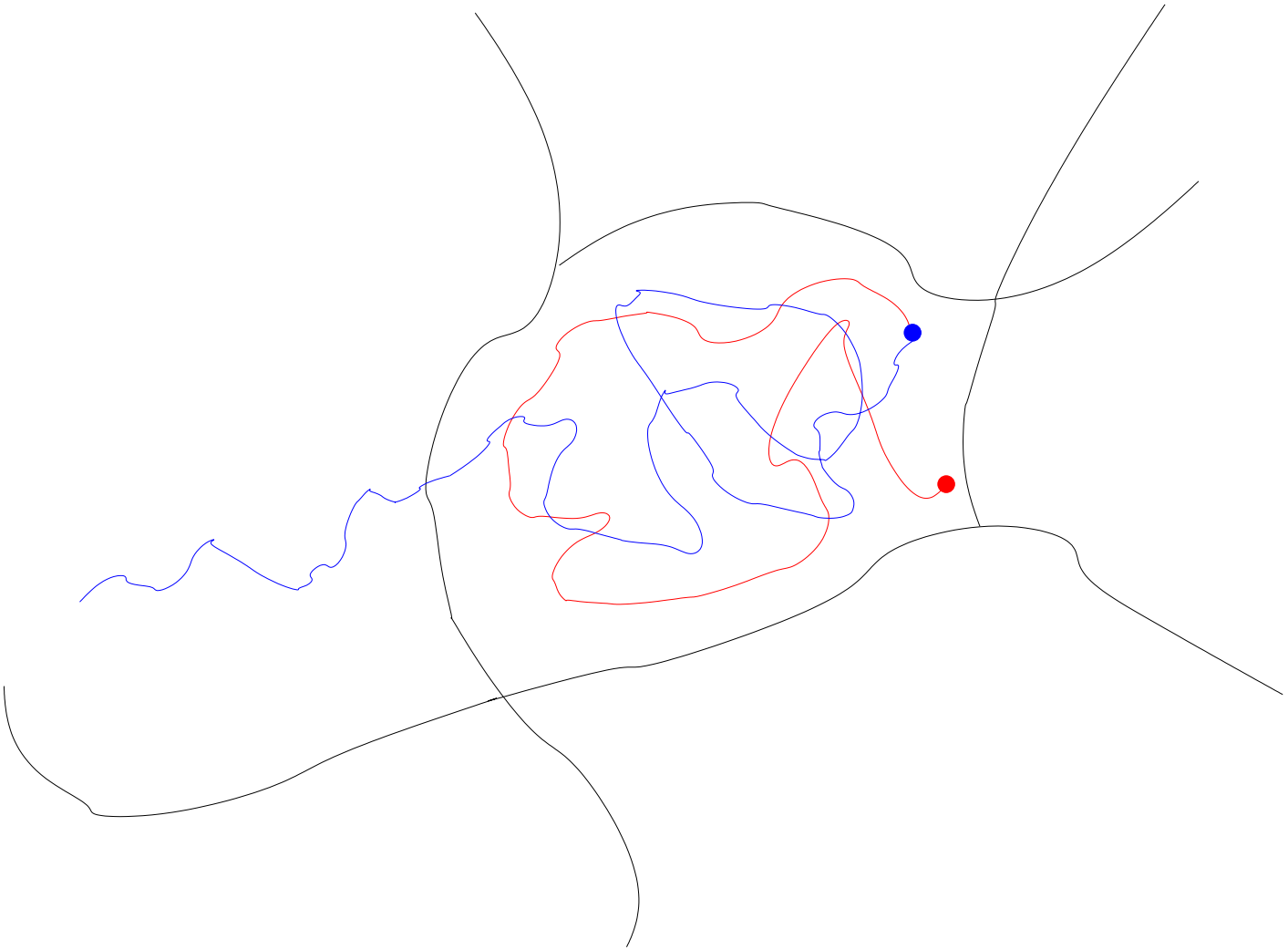
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



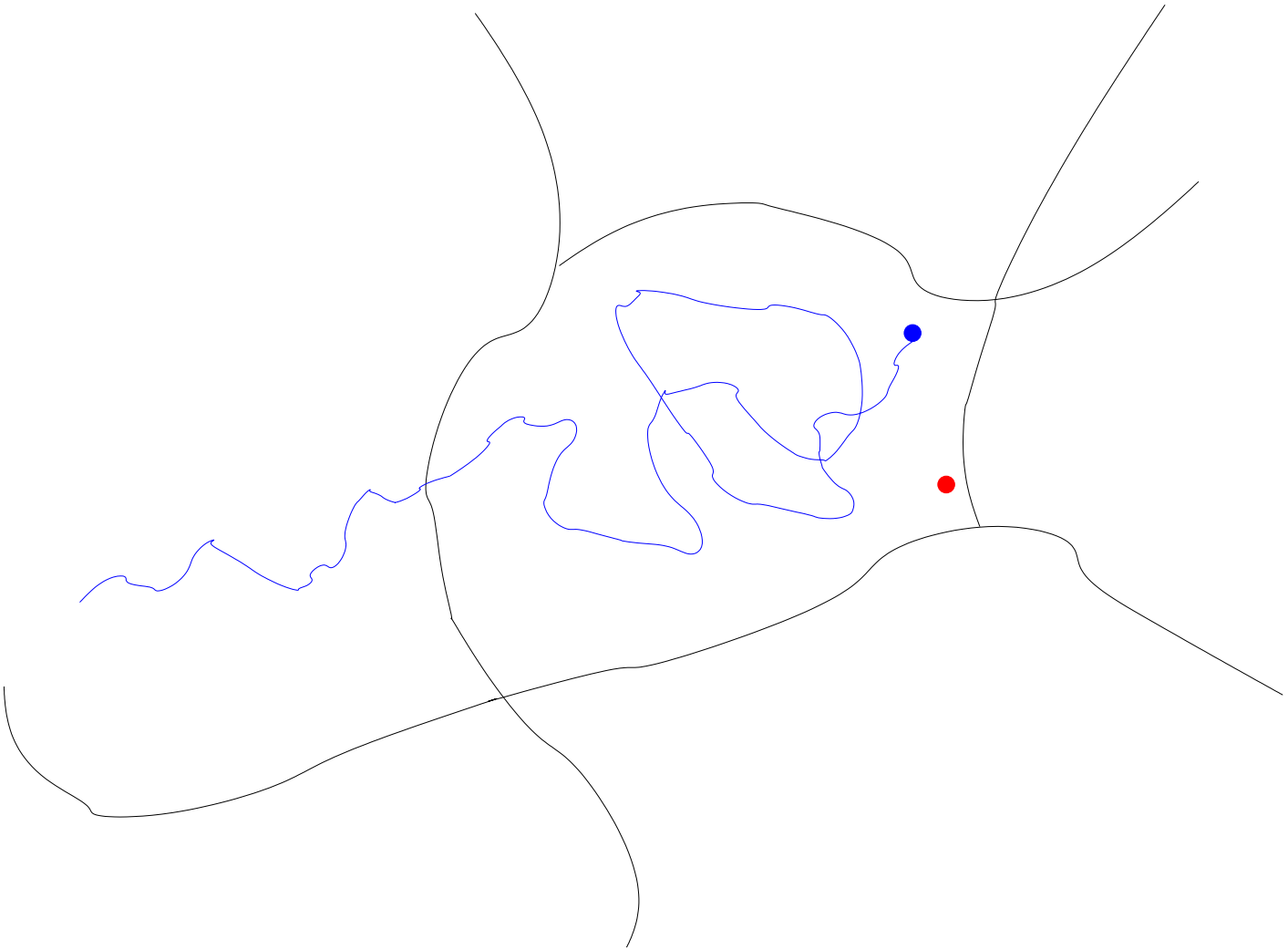
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



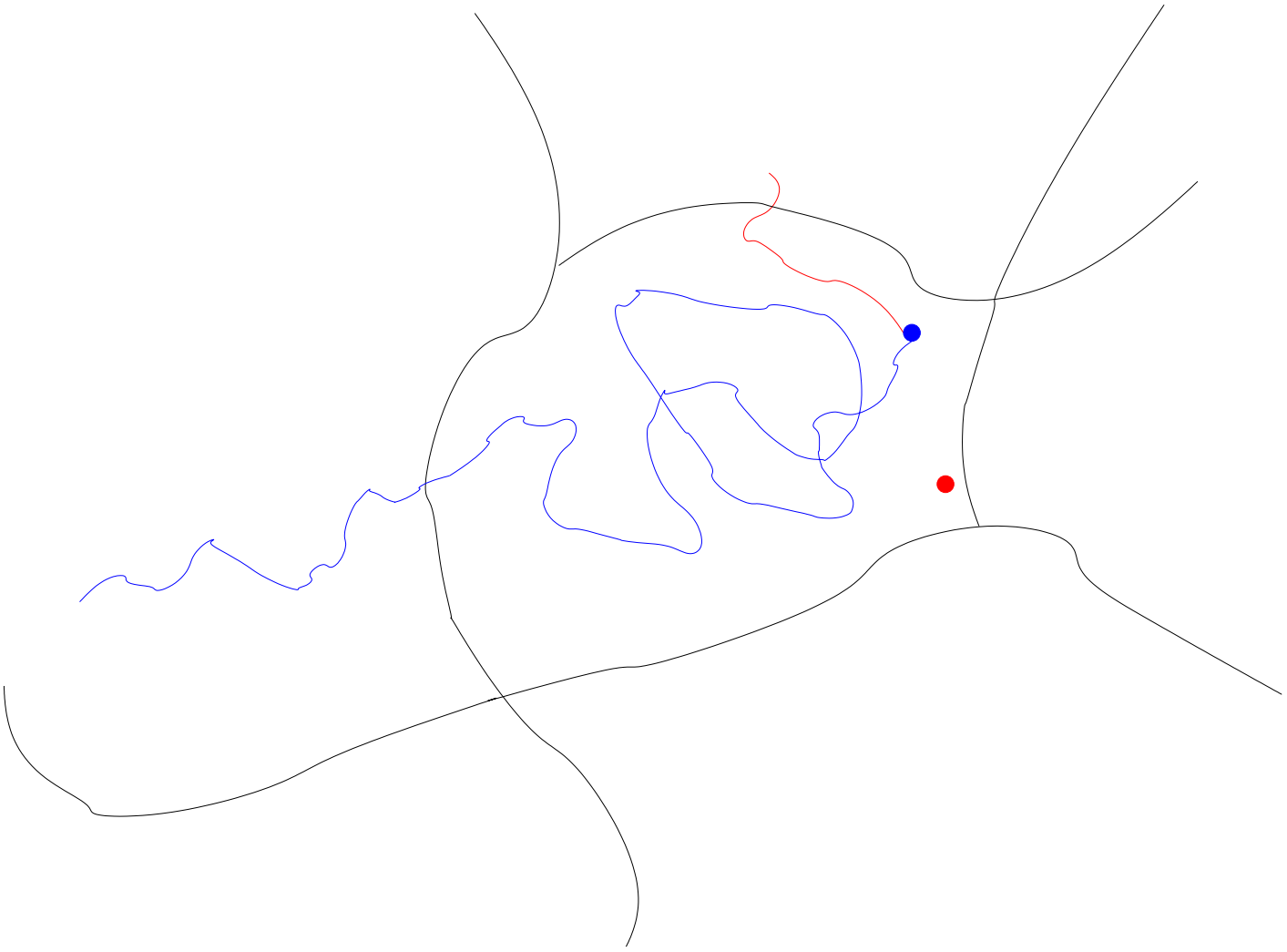
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



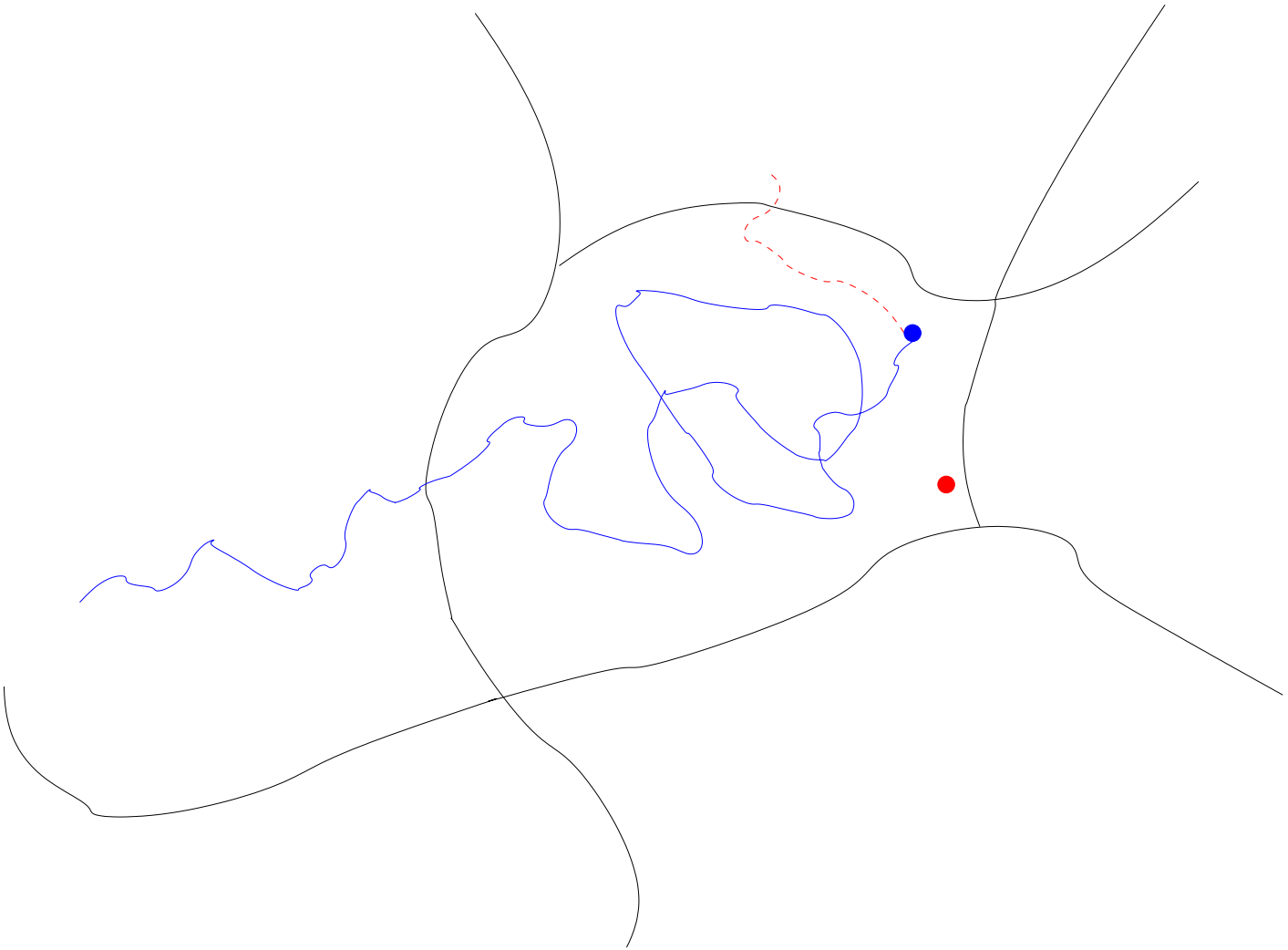
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



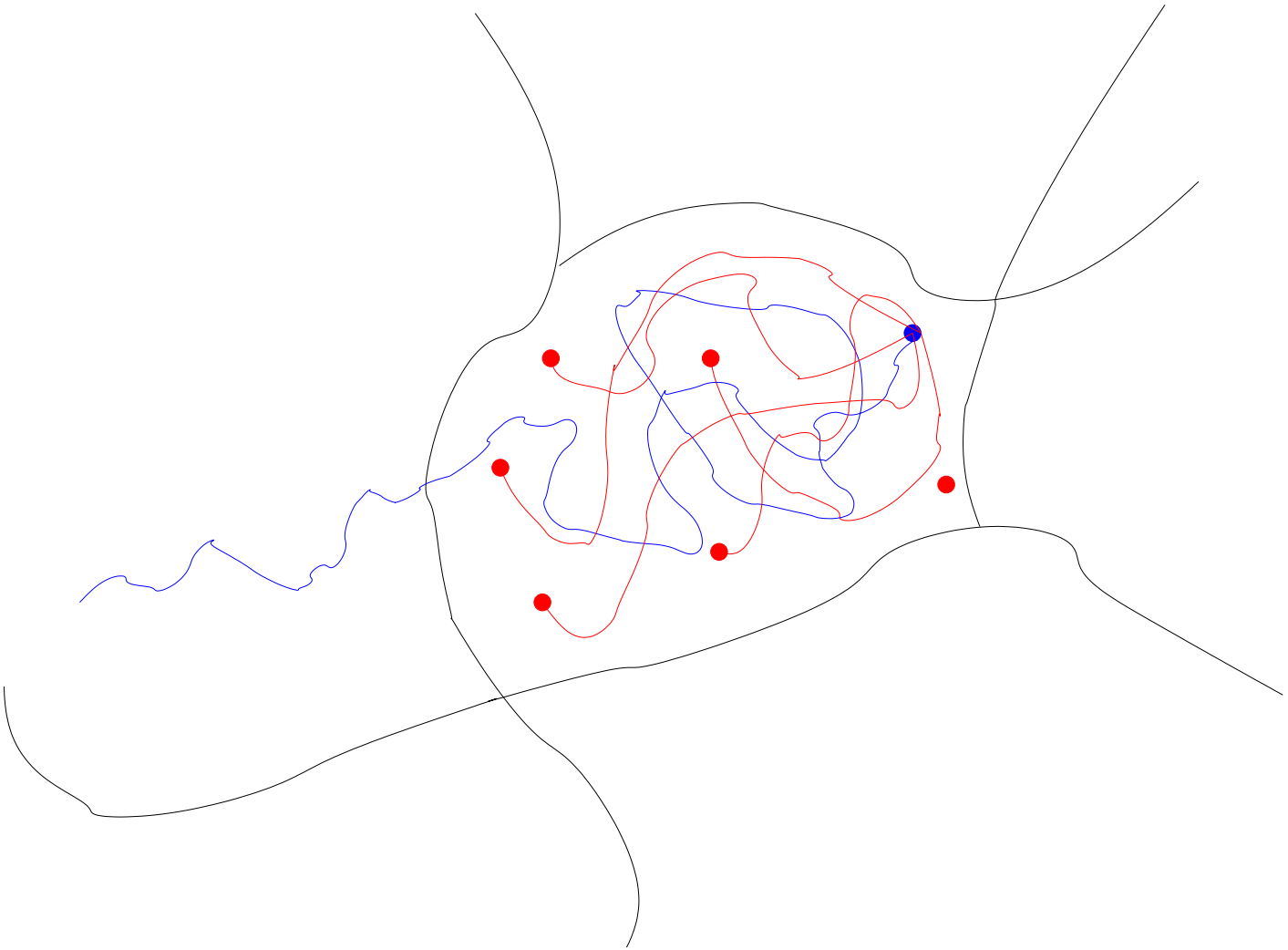
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



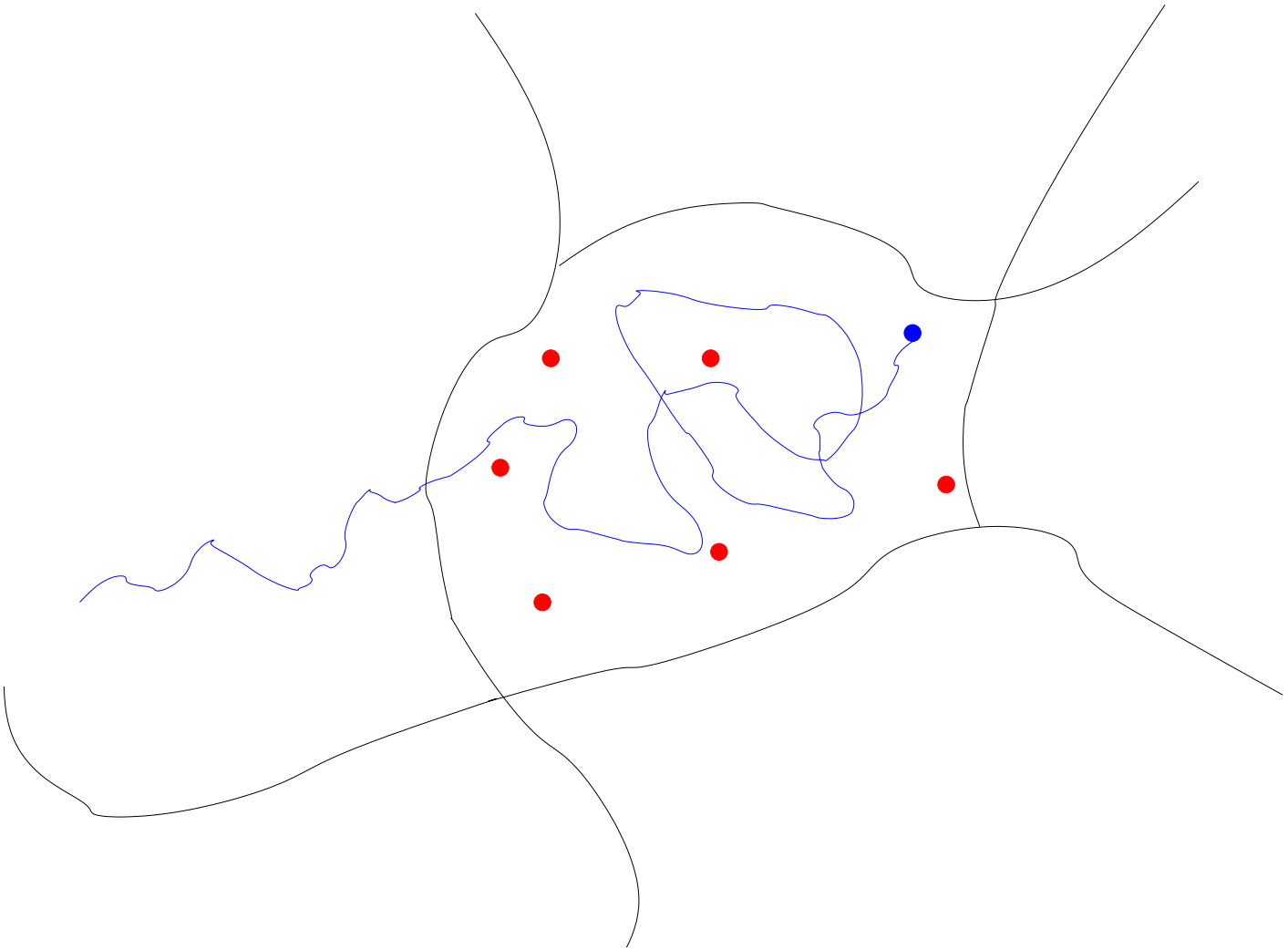
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



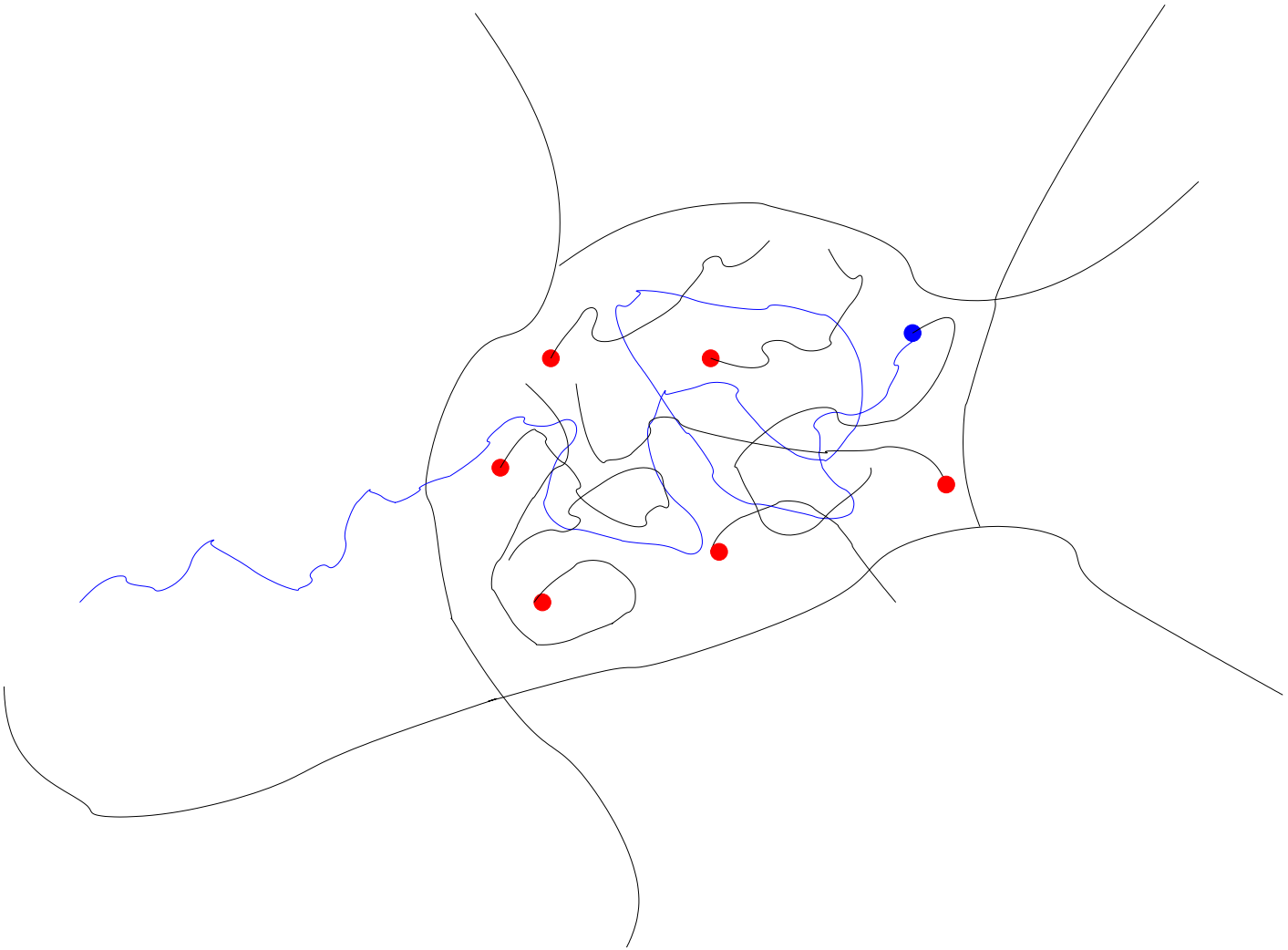
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



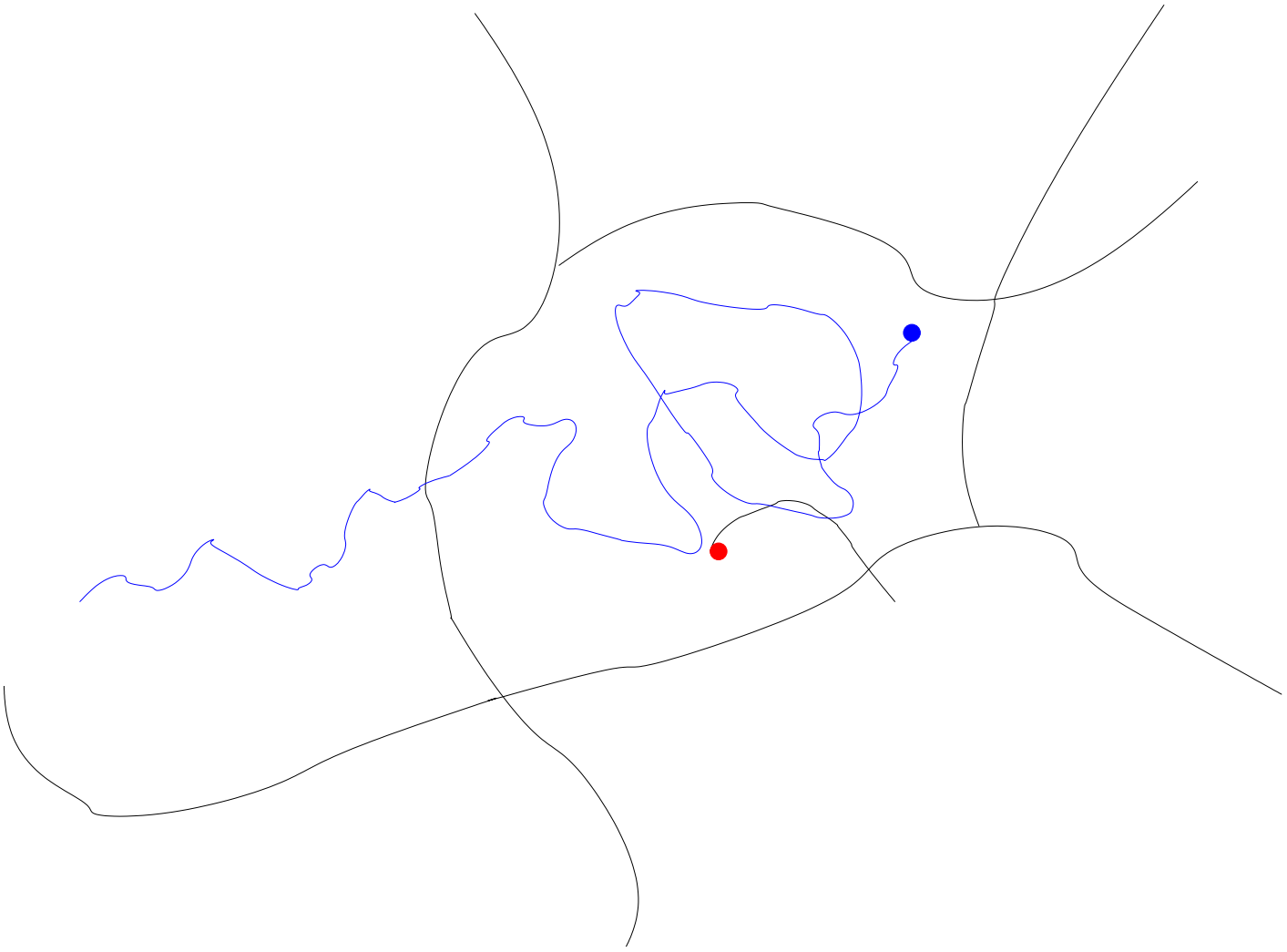
The Parallel Replica Algorithm

Parallel step: run independent trajectories in parallel...



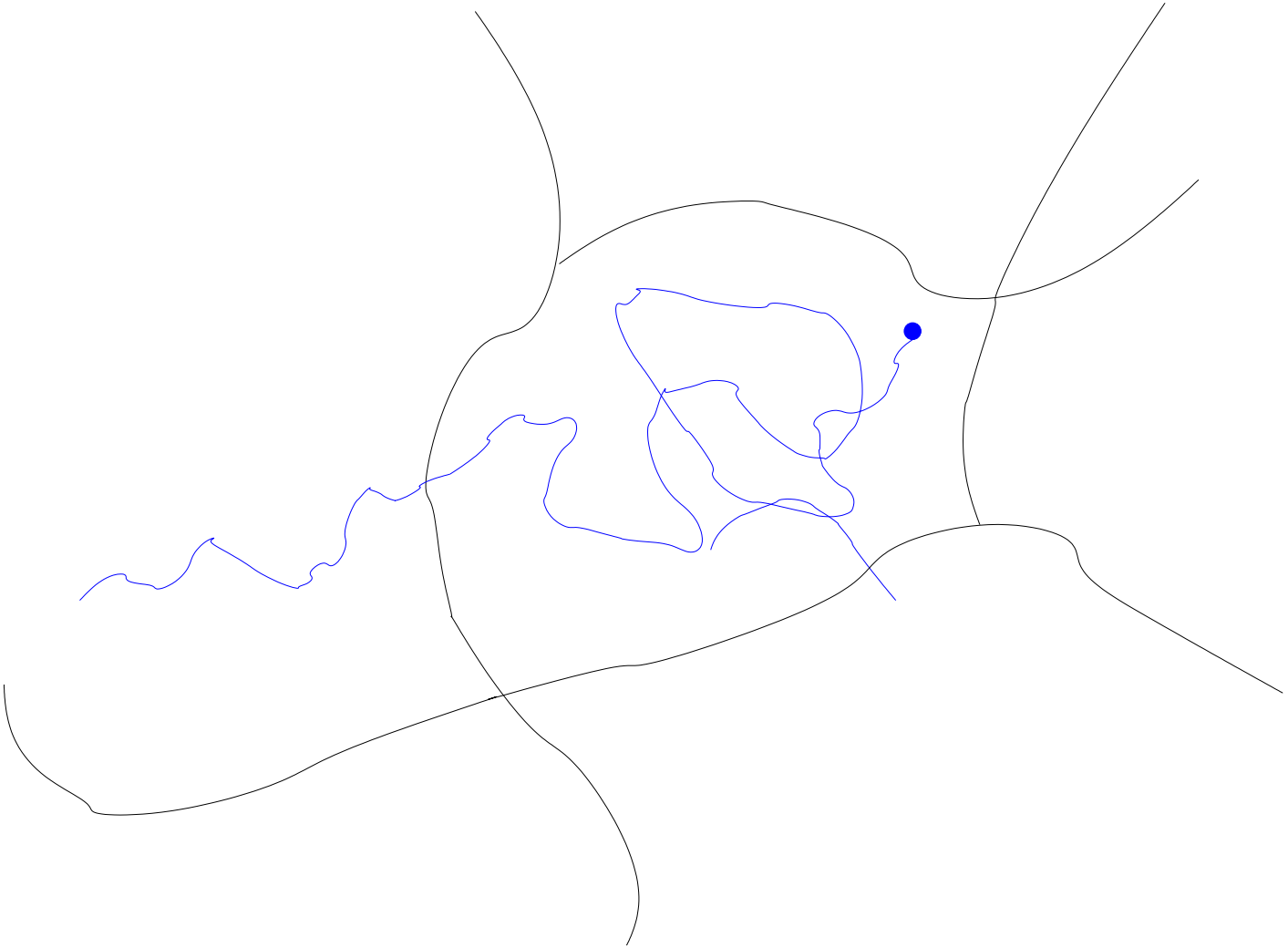
The Parallel Replica Algorithm

Parallel step: ... and detect the first transition event.



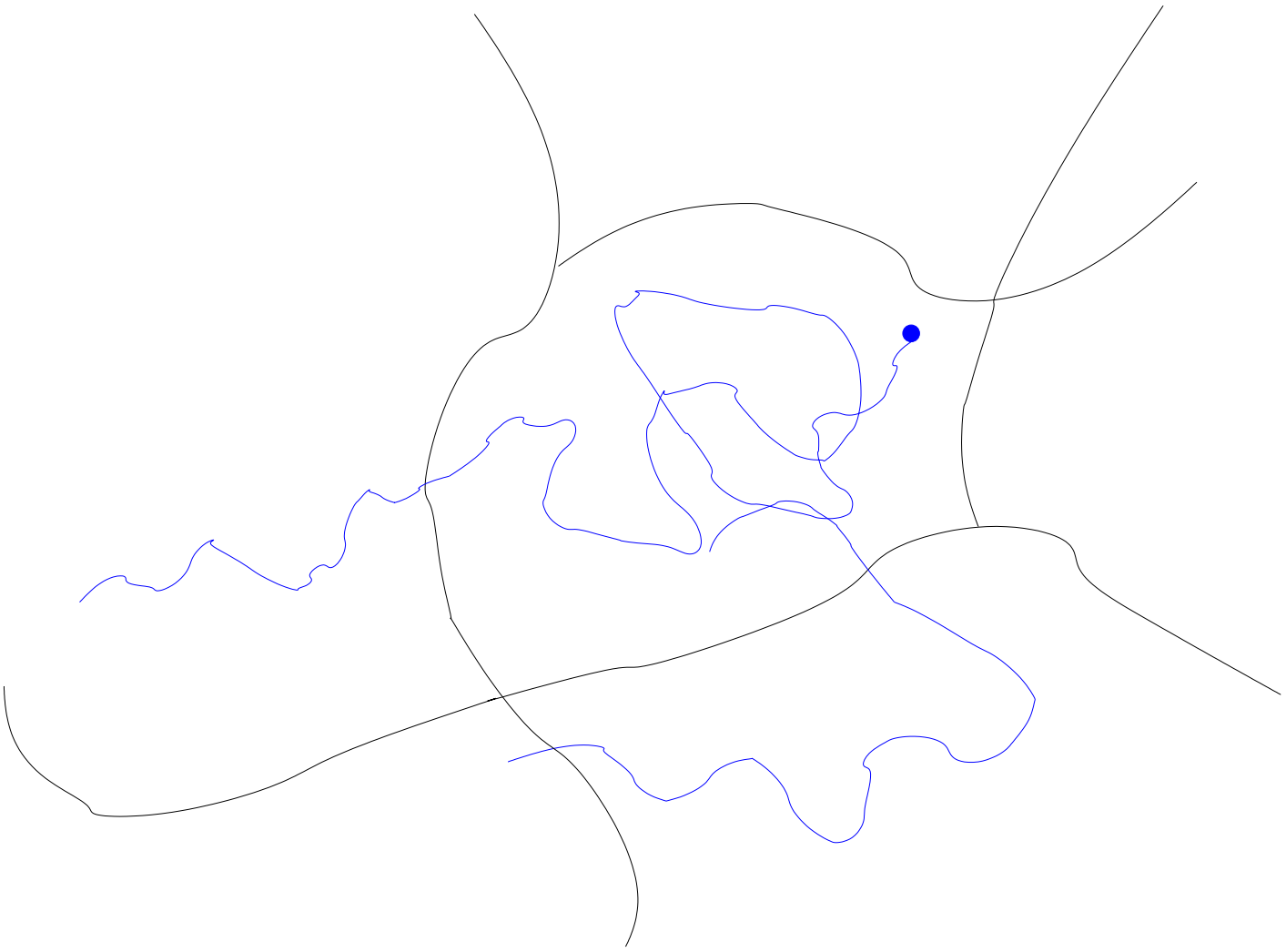
The Parallel Replica Algorithm

Parallel step: update the time clock: $T_{simu} = T_{simu} + NT$.



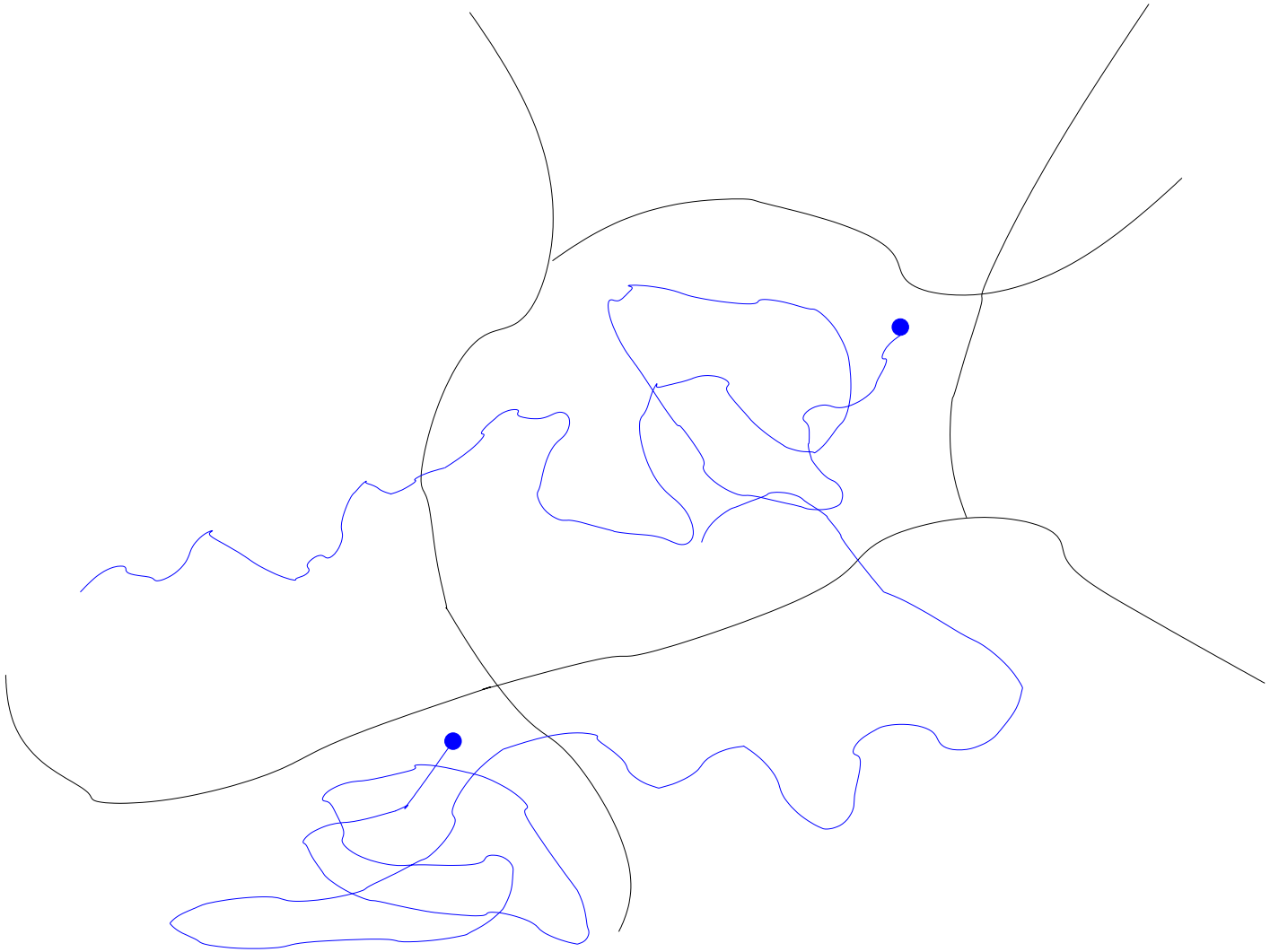
The Parallel Replica Algorithm

A new decorrelation step starts...



The Parallel Replica Algorithm

New decorrelation step



The Parallel Replica Algorithm

The three steps of ParRep:

- Decorrelation step: does the reference walker remain trapped in a set ?
- Dephasing step: prepare many initial conditions in this trapping set.
- Parallel step: detect the first escaping event.

$\longrightarrow$ simulation

The decorrelation step

How to quantify the error introduced by the dephasing and parallel steps, when the decorrelation step is successful ?

When the decorrelation step is successful, it is assumed that the reference walker is distributed according to the QSD : if it was indeed the case, the algorithm would be exact. The decorrelation step can be seen as a way to probe this assumption. What is the error introduced there ?

The decorrelation step

We have the following error estimate in total variation norm: for $t \geq \frac{C}{\lambda_2 - \lambda_1}$,

$$\sup_{f, \|f\|_{L^\infty} \leq 1} \left| \mathbb{E}(f(T_W - t, X_{T_W}) | T_W \geq t) - \mathbb{E}^\nu(f(T_W, X_{T_W})) \right| \leq C \exp(-(\lambda_2 - \lambda_1)t),$$

where $-\lambda_2 < -\lambda_1 < 0$ are the two first eigenvalues of L^* with absorbing boundary conditions on ∂W.

This shows that τ_{corr} should be chosen such that:

$$\tau_{corr} \geq \frac{\overline{C}}{\lambda_2 - \lambda_1}.$$

On the other hand, it should be smaller than the typical time to leave the well, $\mathbb{E}(T_W)$. Since $\mathbb{E}^\nu(T_W) = 1/\lambda_1$, this typically implies the spectral gap requirement,

$$\frac{\overline{C}}{\lambda_2 - \lambda_1} \leq \frac{1}{\lambda_1}.$$

The Parallel Replica Algorithm

This algorithm is very versatile: it works for entropic barriers, and for any partition of the state space into states. But it requires some a priori knowledge on the system: the equilibration time τ_{corr} attached to each state S.

Two questions: How to choose τ_{corr} ? How to sample the QSD ?

We propose a generalized Parallel Replica algorithm [Binder, TL, Simpson, 2014] to solve these issues. It is based on two ingredients:

- the Fleming-Viot particle process
- the Gelman-Rubin statistical test

The Fleming-Viot particle process

Start N processes i.i.d. from μ_0, and iterate the following steps:

1. Integrate (in parallel) N realizations ($k = 1, \ldots, N$)
$$d\boldsymbol{X}_t^k = -\nabla V(\boldsymbol{X}_t^k)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t^k$$
until one of them, say $\boldsymbol{X}_t^1$, exits;
2. Kill the process that exits;
3. With uniform probability $1/(N-1)$, randomly choose one of the survivors, $\boldsymbol{X}_t^2, \ldots, \boldsymbol{X}_t^N$, say $\boldsymbol{X}_t^2$;
4. Branch $\boldsymbol{X}_t^2$, with one copy persisting as $\boldsymbol{X}_t^2$, and the other becoming the new $\boldsymbol{X}_t^1$.

It is known that the empirical distribution

$$\mu_{t,N} \equiv \frac{1}{N} \sum_{k=1}^{N} \delta_{\boldsymbol{X}_t^k}$$

satisfies:

$$\lim_{N\to\infty} \mu_{t,N} = \mathcal{L}(\boldsymbol{X}_t | t < T_W).$$

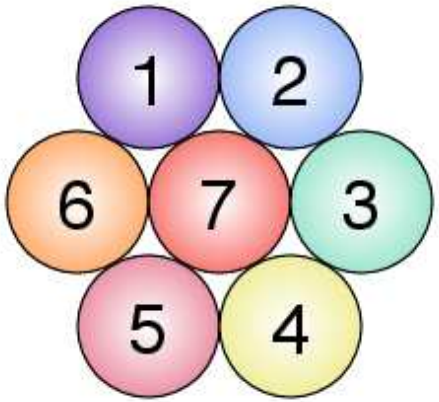
The generalized Parallel Replica algorithm

The generalized Parallel Replica algorithm consists in using a Fleming-Viot particle process for the dephasing step and running in parallel the decorrelation and the dephasing steps.

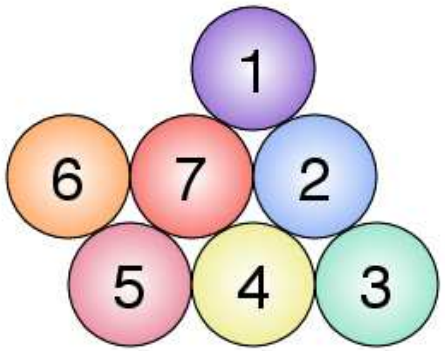
If the Fleming Viot particle process reaches stationarity before the reference walker, go to the parallel step. Otherwise, restart a new decorrelation / dephasing step.

The time at which the Fleming-Viot particle process becomes stationary is determined using the Gelman-Rubin statistical test.

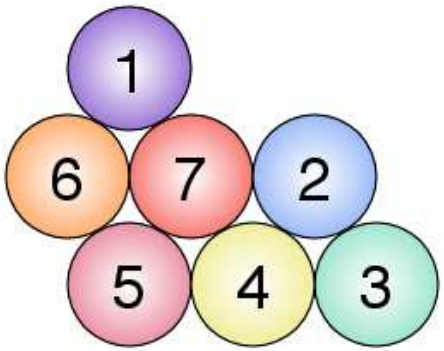
Numerical test case: the 7 atoms LJ cluster



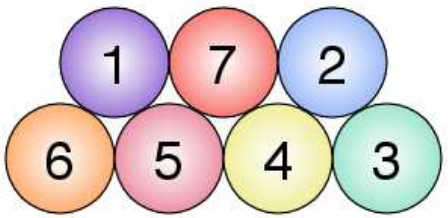
(a) C_0, $V = -12.53$



(b) C_1, $V = -11.50$



(c) C_2, $V = -11.48$



(d) C_3, $V = -11.40$

We study the escape from the configuration C_0 using overdamped Langevin dynamics with $\beta = 6$. The next visited states are C_1 or C_2.

Numerical test case: the 7 atoms LJ cluster

Method	TOL	$\langle T \rangle$	$\mathbb{P}[C_1]$	$\mathbb{P}[C_2]$
Serial	–	17.0	(0.502, 0.508)	(0.491, 0.498)
ParRep	0.2	19.1	(0.508, 0.514)	(0.485, 0.492)
ParRep	0.1	18.0	(0.506, 0.512)	(0.488, 0.494)
ParRep	0.05	17.6	(0.505, 0.512)	(0.488, 0.495)
ParRep	0 .01	17.0	(0.504, 0.510)	(0.490, 0.496)

Method	TOL	$\langle t_{\mathrm{corr}} \rangle$	$\langle$Speedup$\rangle$	% Dephased
Serial	–	–	–	–
ParRep	0.2	0.41	29.3	98.5%
ParRep	0.1	.98	14.9	95.3%
ParRep	0.05	2.1	7.83	90.0%
ParRep	0 .01	11	1.82	52.1%

Numerical test case: the 7 atoms LJ cluster

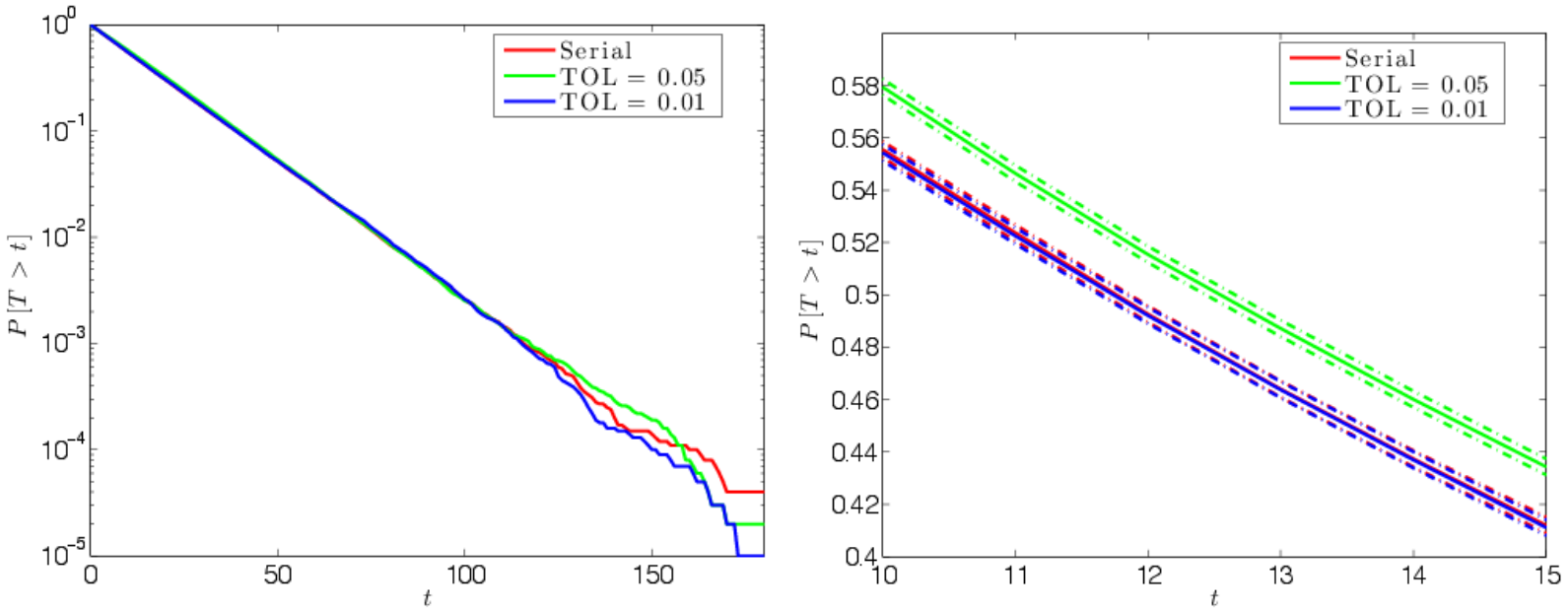


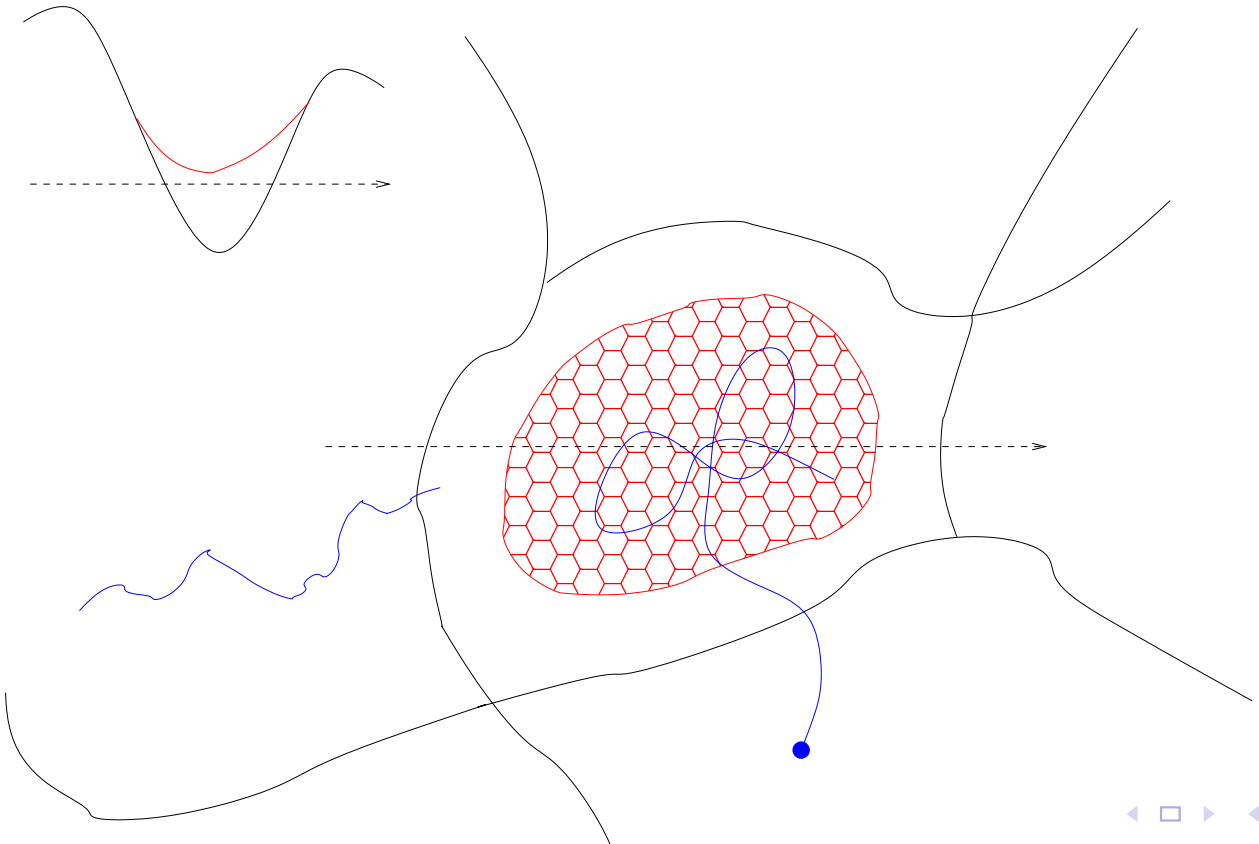
Figure: $\mathrm{LJ}_7^{2\mathrm{D}}$: Cumulative distribution function of the escape time from C_0.

The Hyperdynamics

Idea: raise the potential in W to reduce the exit time.

Two steps:

- Equilibrate on the biased potential $V + \delta V$;
- Wait for an exit and multiply the exit time $T_W^{\delta V}$ by the boost factor $B = \frac{1}{T_W^{\delta V}} \int_0^{T_W^{\delta V}} \exp(\beta\, \delta V(\boldsymbol{X}_t))\, dt$.



The Hyperdynamics

Why is it consistent ?

Recall property 3 go to Prop3. The underlying mathematical question is: how λ_1 and $\partial_n u_1$ are modified when V is changed to $V + \delta V$?

Recall that

$$\begin{cases} \operatorname{div}\left(\nabla V\, u_1 + \beta^{-1}\nabla u_1\right) = -\lambda_1 u_1 \text{ on } W, \\ u_1 = 0 \text{ on } \partial W. \end{cases}$$

Strategy: change u_1 to $u_1 \exp(V/2)$ and use results from semi-classical analysis for boundary Witten Laplacians in order to characterize $(\lambda_1, \partial_n u_1)$ in terms of V.

The Hyperdynamics: mathematical analysis

Assumptions on V. We assume there exists $W^- \subset\subset W$ such that:

- Regularity: V and $V|_{\partial W}$ are Morse functions ;
- Localization of the small eigenvectors in W^-:
  - (i) $|\nabla V| \neq 0$ in $\overline{W} \setminus W^-$,
  - (ii) $\partial_n V > 0$ on ∂W^- ,
  - (iii) $\min_{\partial W} V \geq \min_{\partial W^-} V$,
  - (iv) $\min_{\partial W^-} V - \mathrm{cvmax} > \mathrm{cvmax} - \min_{W^-} V$ where $\mathrm{cvmax} = \max\{V(x), x \text{ s.t. } |\nabla V(x)| = 0\}$;
- Non degeneracy of exponentially small eigenvalues: The critical values of V in W^- are all distinct and the differences $V(y) - V(x)$, where $x \in \mathcal{U}^{(0)}$ ranges over the local minima of $V|_{W^-}$ and $y \in \mathcal{U}^{(1)}$ ranges over the critical points of $V|_{W^-}$ with index 1, are all distinct.

Assumptions on δV.

- $V + \delta V$ satisfies the same assumptions as V ;
- $\delta V = 0$ on $\overline{W} \setminus W^-$.

The Hyperdynamics: mathematical analysis

Result [TL, Nier, 2013]: Under the above assumptions on the potentials V and $(V + \delta V)$, there exists $c > 0$ such that, in the limit $\beta \to \infty$,

$$\frac{\lambda_1(V+\delta V)}{\lambda_1(V)} = \frac{\int_W e^{-\beta V}}{\int_W e^{-\beta(V+\delta V)}}(1+\mathcal{O}(e^{-\beta c}))\,,$$

$$\frac{\partial_n\left[u_1(V+\delta V)\right]\big|_{\partial W}}{\|\partial_n\left[u_1(V+\delta V)\right]\|_{L^1(\partial W)}} = \frac{\partial_n\left[u_1(V)\right]\big|_{\partial W}}{\|\partial_n\left[u_1(V)\right]\|_{L^1(\partial W)}} + \mathcal{O}(e^{-\beta c}) \quad \text{in} \quad L^1(\partial W)\,.$$

Remark: We indeed have

$$\begin{aligned} B &= \frac{1}{T_W^{\delta V}} \int_0^{T_W^{\delta V}} \exp(\beta\,\delta V(\boldsymbol{X}_t))\,dt. \\ &\simeq \frac{\int_W \exp(\beta\delta V)\exp(-\beta(V+\delta V))}{\int_W \exp(-\beta(V+\delta V))} \\ &= \frac{\int_W \exp(-\beta V)}{\int_W \exp(-\beta(V+\delta V))}. \end{aligned}$$

The Hyperdynamics: idea of the proof

Use semi-classical analysis for boundary Witten laplacians ($f = V$, $h = 2/\beta$).

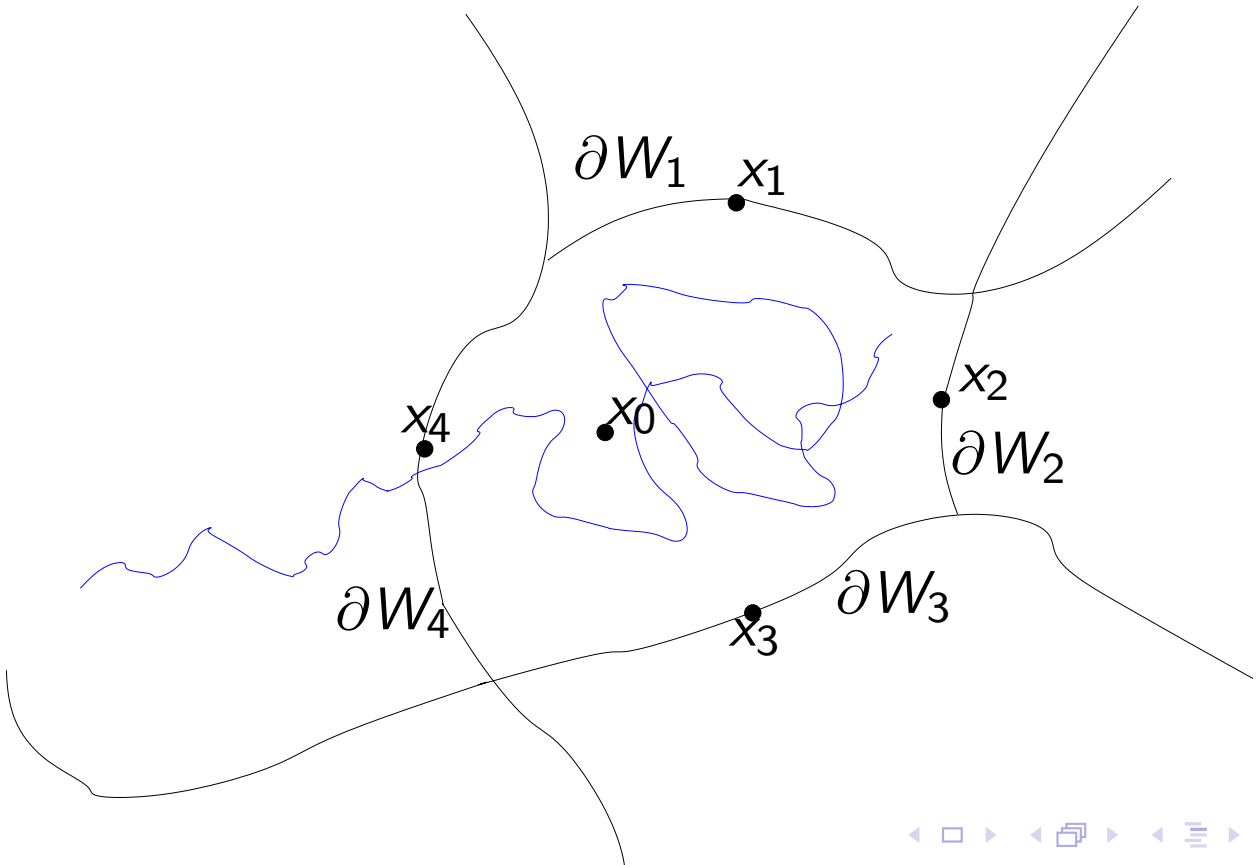
- Build quasimodes for $\Delta_{f,h}^{D,(p)}(W)$ $(p = 0, 1)$ using eigenvectors of $\Delta_{f,h}^{N,(p)}(W^-)$ $(p = 0, 1)$ and of $\Delta_{f,h}^{D,(1)}(W \setminus \overline{W^-})$.
- Analyze the asymptotics of the singular values of the restricted differential ($\nu(h) \leq h$ and $\lim_{h\to 0} h \log(\nu(h)) = 0$)
  $d_{f,h} : F^{(0)} \to F^{(1)}$ where $F^{(p)} = \mathrm{Ran}\left(1_{[0,\nu(h)]}\left(\Delta_{f,h}^{D,(p)}(W)\right)\right)$.
  This is a finite dimensional linear operator.
- Show that, up to exponentially small terms,
  $\lambda_1(V) = \frac{A}{\int_W \exp(-\beta V)}(1 + \mathcal{O}(e^{-\frac{c}{h}}))$ and $\frac{\partial_n u_1}{\|\partial_n u_1\|} = B + \mathcal{O}(e^{-\frac{c}{h}})$
  where A and B only depends on the eigenvectors of $\Delta_{f,h}^{D,(1)}(W \setminus \overline{W^-})$, and are thus not modified when changing V to $V + \delta V$.

The Temperature Accelerated Dynamics

Idea: increase the temperature to reduce the exit time.

Algorithm:

- Observe the exit events from W at high temperature ;
- Extrapolate the high temperature exit events to low temperature exit events.



Extrapolation procedure (1/2)

Rewriting the exit event using a kinetic Monte Carlo model:

Let us introduce $\lambda_1 = 1/\mathbb{E}(T_W)$ and

$$p(i) = \mathbb{P}(\boldsymbol{X}_{T_W} \in \partial W_i) = -\frac{\displaystyle\int_{\partial W_i} \partial_n u_1 \, d\sigma}{\beta\lambda \displaystyle\int_W u_1(x)\, dx}.$$

To each possible exit saddle point i is associated a rate $k(i) = \lambda_1 p(i)$. If $\tau_i \sim \mathcal{E}(k_i)$ are independent, then

- The exit time is $\min(\tau_1, \ldots, \tau_I)$;
- The exit saddle point is $\arg\min(\tau_1, \ldots, \tau_I)$.

Extrapolation procedure (2/2)

Extrapolating from high temperature to low temperature:

The extrapolation procedure is based on the empirical Arrhenius law: for large β,

$$k(i) = \lambda_1 p(i) \simeq \eta_i \exp(-\beta(V(x_i) - V(x_0)))$$

where η_i is independent of β, which yields

$$\frac{k^{lo}(i)}{k^{hi}(i)} = \frac{\lambda_1^{lo} p^{lo}(i)}{\lambda_1^{hi} p^{hi}(i)} \simeq \exp(-(\beta^{lo} - \beta^{hi})(V(x_i) - V(x_0))).$$

Algorithm: observe exit events at high temperature, extrapolate the rates to low temperature, stop when the extrapolated event will not modify anymore the low temperature exit event.

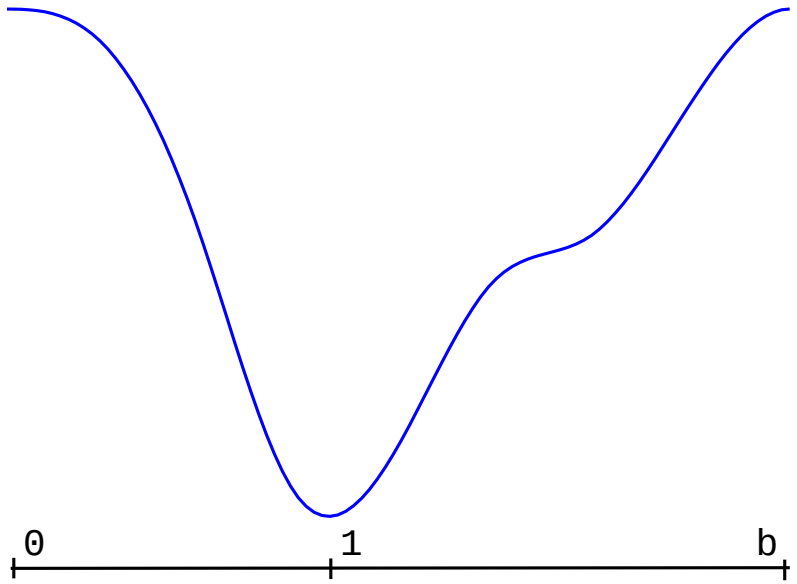
Remark: TAD can be seen as a smart saddle point search method.

Arrhenius law

If the Arrhenius law is exactly satisfied, one can show that the temperature accelerated dynamics method is exact.

Mathematical question: Under which assumptions is the Arrhenius law satisfied ? This is again a semi-classical analysis problem...

In 1D, this can be done. In the limit $\beta^{hi}, \beta^{lo} \to \infty$, $\beta^{lo}/\beta^{hi} = r$, under appropriate assumptions, one has [Aristoff, TL, 2014]:



$$\frac{\lambda^{hi} p_i^{hi}}{\lambda^{lo} p_i^{lo}} = e^{-(\beta^{hi}-\beta^{lo})(V(x_i)-V(x_0))} \left(1 + O\left(\frac{1}{\beta^{hi}} - \frac{1}{\beta^{lo}}\right)\right)$$

Simulating dynamics: conclusions (1/3)

Remarks on accelerated dynamics:

- From ParRep to Hyper to TAD, the underlying assumptions for the algorithms to be correct are more and more stringent. In particular, Hyper and TAD require *energetic barriers* and *small temperature*.
- The QSD is a good intermediate between continuous state dynamics and kMC-like approximations (Markov state models). Transition rates could be defined starting from the QSD.
- It can be used to analyze the validity of the transition state theory and kMC models, in the small temperature regime.

Simulating dynamics: conclusions (2/3)

There are other mathematical settings to characterize / quantify metastability:

- Large deviation techniques [Freidlin, Wentzell, Vanden Eijnden, Weare, Touchette,...] [Talk by Touchette] and Onsager-Machlup functionals [Stuart, Pinsky, Theil] [Talk by Pinsky]
- Potential theoretic approaches [Bovier, Schuette, Hartmann,...]
- Spectral analysis of the Fokker Planck operator on the whole space and semi-classical analysis [Schuette, Helffer, Nier, Pavliotis]

Simulating dynamics: conclusions (3/3)

There are many other numerical techniques:

- Going from state A to state B:
  - *Local search*: the string method [E, Ren, Vanden-Eijnden], max flux [Skeel], transition path sampling methods [Chandler, Bolhuis, Dellago],
  - *Global search, ensemble of trajectories*: AMS [Talks by Cérou and Guyader], transition interface sampling [Bolhuis, van Erp], forward flux sampling [Allen, Valeriani, ten Wolde], milestoning techniques [Elber, Schuette, Vanden-Eijnden]
- Importance sampling approaches on paths, reweighting [Dupuis, Vanden-Einjden, Weare, Schuette, Hartmann]
- Saddle point search techniques [Mousseau, Henkelman] and graph exploration
- Starting from a long trajectory, extract states: clustering, Hidden Markov chain [Schuette]

References

Monographs on numerical methods in MD:

- Allen, M., Tildesley, D.: Computer Simulation of Liquids. Oxford University Press (1989)
- Frenkel, D., Smit, B.: Understanding Molecular Simulation: From Algorithms to Applications. Oxford University Press (2001)
- Hairer, E., Lubich, C., Wanner, G.: Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations, Springer-Verlag (2006).
- Leimkuhler, B., Reich, S.: Simulating Hamiltonian Dynamics. Cambridge University Press (2005).
- TL, Stoltz, G., Rousset, M.: Free Energy Computations: A Mathematical Perspective. World Scientific (2010).
- Rapaport, D.: The Art of Molecular Dynamics Simulation. Cambridge University Press (2004)
- Tuckerman, M.: Statistical Mechanics: Theory and Molecular Simulation. Oxford University Press (2010).

References

Some papers I mentioned:

- D. Aristoff and TL, *Mathematical analysis of Temperature Accelerated Dynamics*, SIAM MMS, 12(1), 290-317, 2014.
- C. Le Bris, TL, M. Luskin and D. Perez, *A mathematical formalization of the parallel replica dynamics*, Monte Carlo Methods and Applications, 18(2), 119-146, 2012.
- TL and F. Nier, *Low temperature asymptotics for Quasi-Stationary Distributions in a bounded domain*, `http://arxiv.org/abs/1309.3898` .
- TL, M. Rousset and G. Stoltz, *Long-time convergence of an Adaptive Biasing Force method*, Nonlinearity, 21, 1155-1181, 2008.