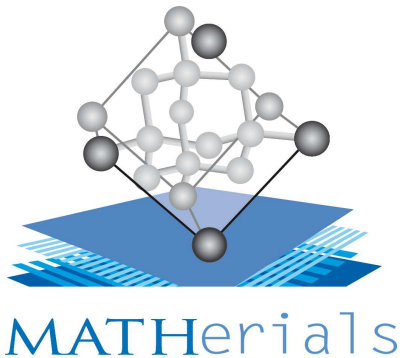
Model reduction techniques for stochastic dynamics

T. Lelièvre

CERMICS - Ecole des Ponts ParisTech & Equipe Matherials - INRIA



Co-authors: D. Aristoff, G. Di Gesu, C. Le Bris, D. Le Peutrec, F. Legoll, M. Luskin, B. Nectoux, F. Nier, D. Perez, G. Simpson, A. Voter.

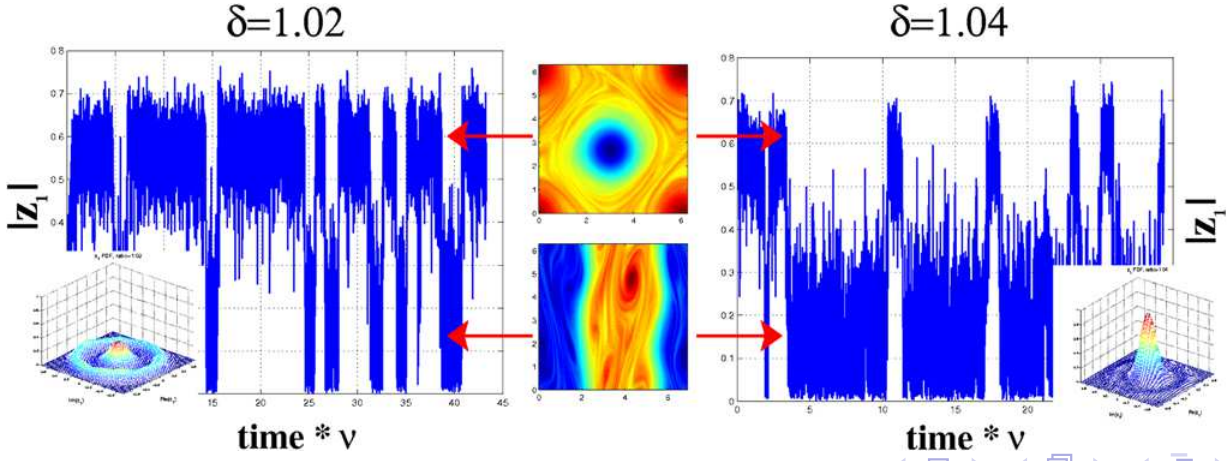
Ecole GDR EGRIN, 25-26 Mai 2016

Introduction

The objective of these lectures is to discuss dynamical coarse-graining techniques for stochastic dynamics. PDE tools are useful to derive an effective dynamics, and analyze the error. We will use the fact that the original dynamics are metastable.

In the context of geophysical flows, metastable stochastic processes are used to model bifurcations in the flow topology. [Bouchet, Simonnet]
Example: 2d periodic stochastic Navier-Stokes equations.

$$\begin{cases} \partial_t \omega + v \cdot \nabla \omega = -\alpha\omega + \nu\Delta\omega + \sqrt{\sigma}\eta \\ v = e_z \times \nabla\psi, \qquad \Delta\psi = \omega \end{cases}$$



Introduction

Starting from a dynamics in high dimension ($(X_t)_{t\geq 0}$ with values in $\mathbb{R}^d$), and a coarse-graining map ($\mathcal{C} : \mathbb{R}^d \to \mathbb{R}$ or $\mathcal{C} : \mathbb{R}^d \to \mathbb{N}$), we would like to derive an effective Markov dynamics close to $(\mathcal{C}(X_t))_{t\geq 0}$.

- Theoretically: this yields simpler and hopefully insightful models.
- Numerically: use these coarse-grained dynamics as predictors in predictor-corrector schemes.

Molecular dynamics

The motivation for these works is molecular dynamics computations.

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.

The basic ingredient: a potential V which associates to a configuration $(\boldsymbol{x}_1, ..., \boldsymbol{x}_{N_{atom}}) = \boldsymbol{x} \in \mathbb{R}^{3N_{atom}}$ an energy $V(\boldsymbol{x}_1, ..., \boldsymbol{x}_{N_{atom}}) \in \mathbb{R}$. The dimension $d = 3N_{atom}$ is large (a few hundred thousand to millions).

Molecular dynamics

Newton equations of motion:

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

Molecular dynamics

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.

Molecular dynamics

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.

Molecular dynamics

These dynamics are used to compute macroscopic quantities:

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

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

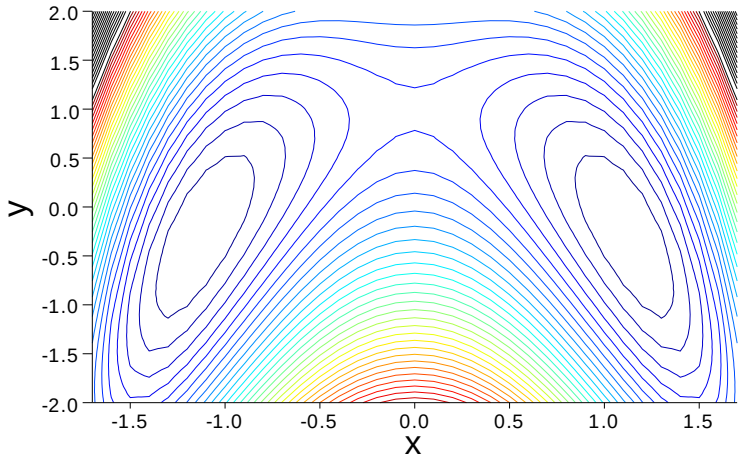
(ii) Dynamical quantities (averages over trajectories): diffusion coefficients, viscosity, transition rates,...

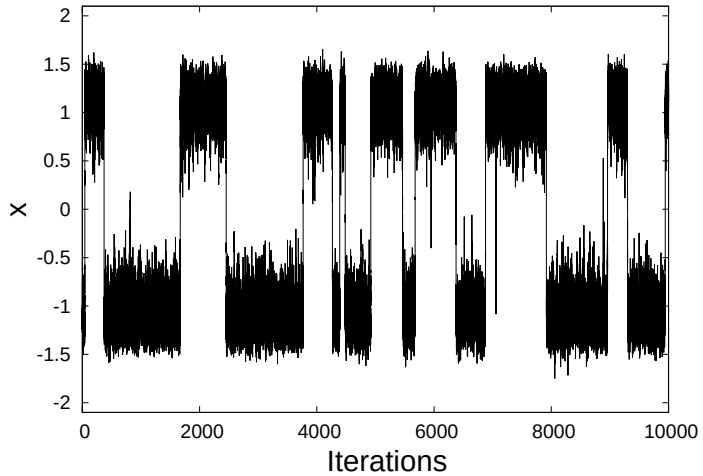
$$\mathbb{E}(\mathcal{F}((\boldsymbol{X}_t)_{t\geq 0})) \simeq \frac{1}{M} \sum_{m=1}^{M} \mathcal{F}((\boldsymbol{X}_t^m)_{t\geq 0}).$$

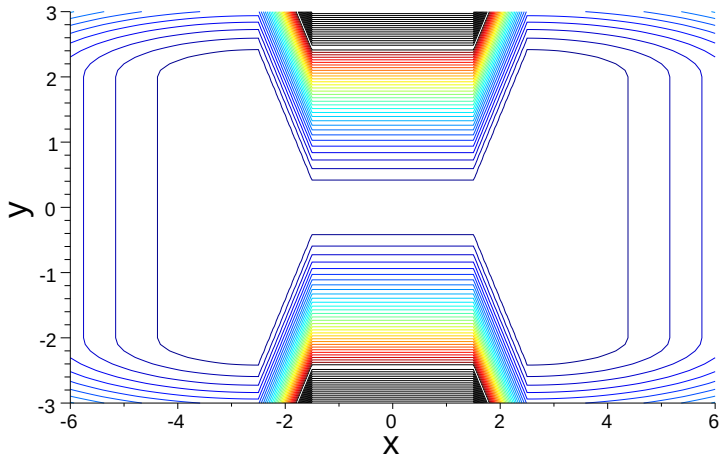
Difficulty: In practice, $\boldsymbol{X}_t$ is a metastable process.

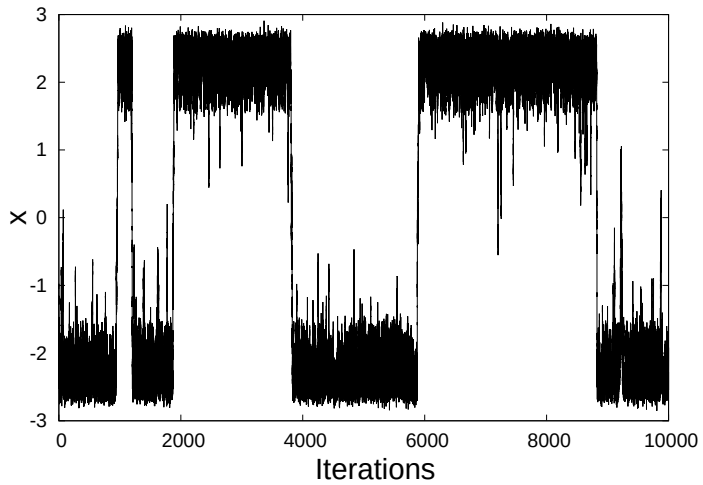
Metastability: energetic and entropic barriers

A two-dimensional schematic picture







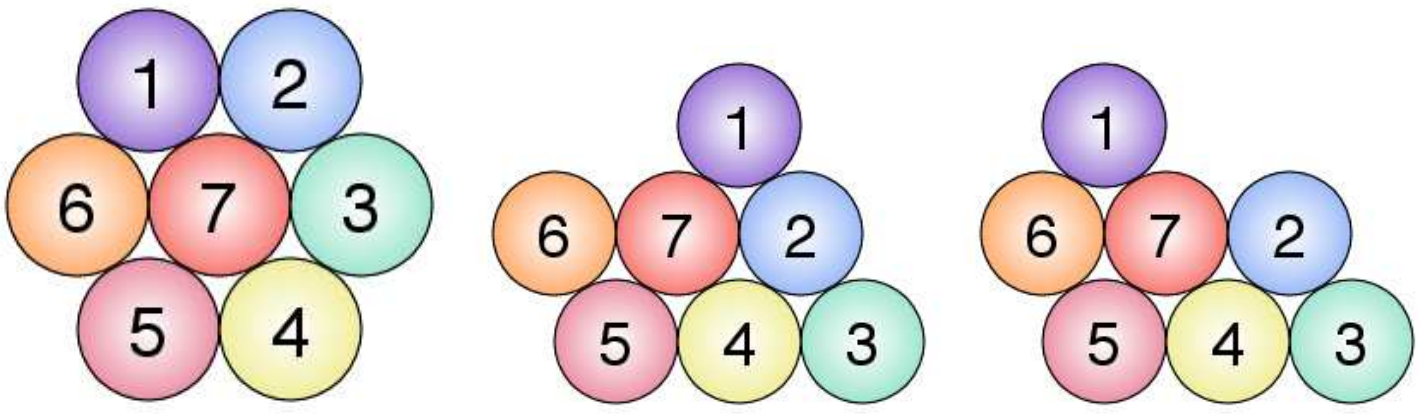


$\longrightarrow$ • Slow convergence of trajectorial averages

• 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, $V = -11.40$

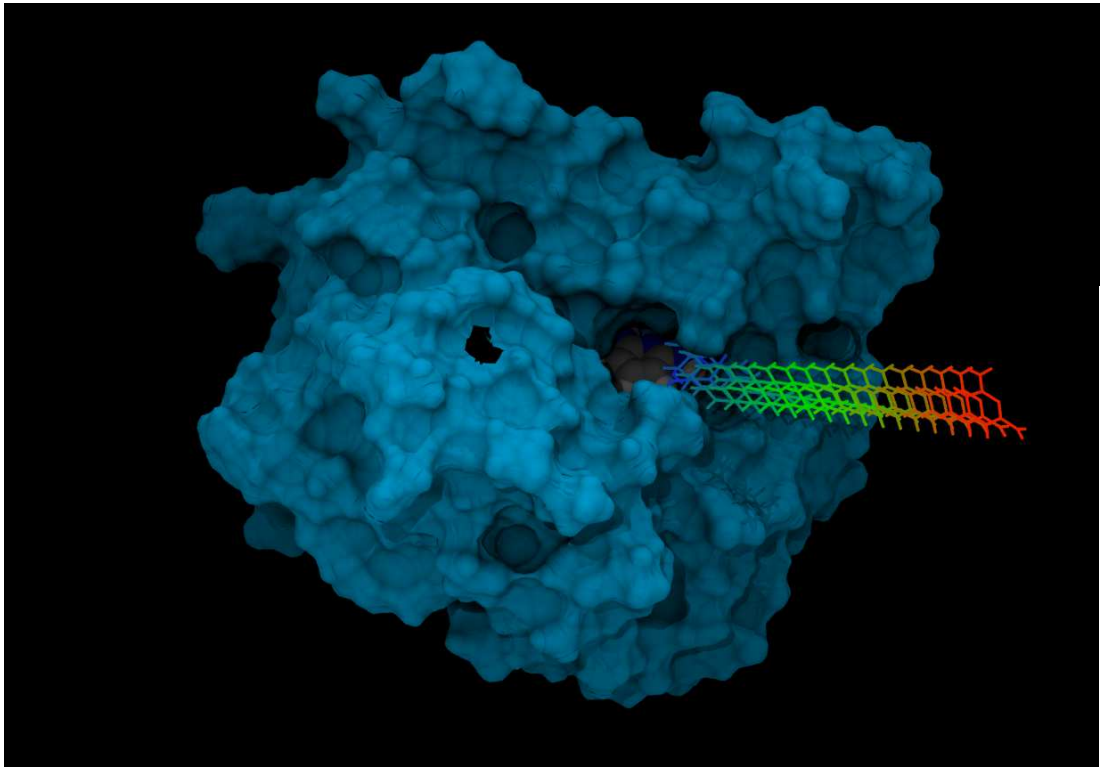
Figure: Low energy conformations of the Lennard-Jones cluster.

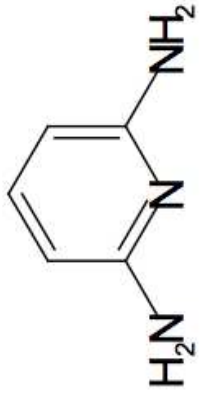
$\longrightarrow$ simulation

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

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.

The aim of this talk is to discuss coarse-graining techniques to efficiently generate dynamical quantities.

- First, for a coarse-graining map with continuous values (justification of Mori-Zwanzig approaches). *Mathematical tool: entropy techniques and logarithmic Sobolev inequalities.*
- Second, for a coarse-graining map with discrete values (justification of Markov state models). *Mathematical tool: quasi-stationary distribution.*

The continuous coarse-graining map

Recall the original dynamics

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

We are given a smooth one dimensional function $\xi : \mathbb{R}^d \to \mathbb{R}$.

Problem: propose a Markovian dynamics (say on $z_t \in \mathbb{R}$) that approximates the dynamics $(\xi(\boldsymbol{X}_t))_{t \geq 0}$.

Two preliminary tools: entropy technique and free energy.

Longtime convergence and entropy (1)

Recall the original gradient dynamics:

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

The associated Fokker-Planck equation writes:

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

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

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

A classical PDE approach: use entropy techniques.

Longtime convergence and entropy (2)

Notice that the Fokker-Planck equation rewrites

$$\partial_t \psi = \beta^{-1} \mathrm{div} \left(\psi_\infty \nabla \left(\frac{\psi}{\psi_\infty} \right) \right)$$

where $\psi_\infty = Z^{-1} \exp(-\beta V)$.
Let us introduce the entropy:

$$H(\psi(t,\cdot)|\psi_\infty) = \int \ln \left(\frac{\psi}{\psi_\infty} \right) \psi.$$

We have (Csiszár-Kullback inequality):

$$\|\psi(t,\cdot) - \psi_\infty\|_{L^1} \leq \sqrt{2H(\psi(t,\cdot)|\psi_\infty)}.$$

Longtime convergence and entropy (3)

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

Definition: The meas $\psi_\infty(x)\,dx$ satisfies a Logarithmic Sobolev inequality (LSI(R)) iff: $\forall\phi$ pdf,

$$H(\phi|\psi_\infty) \le \frac{1}{2R}I(\phi|\psi_\infty)$$

Lemma: ψ_∞ satisfies LSI(R) $\iff$ for all IC $\psi(0,\cdot)$, for all $t \ge 0$, $H(\psi(t,\cdot)|\psi_\infty) \le H(\psi(0,\cdot)|\psi_\infty)\exp(-2\beta^{-1}Rt)$.

Free energy

The free energy $A : \mathbb{R} \to \mathbb{R}$ is defined by:

$$\exp(-\beta A(z)) = Z^{-1} \int_{\{x, \xi(x)=z\}} \exp(-\beta V(x))\, \delta_{\xi(x)-z}(dx).$$

By definition,

$$\int \varphi \circ \xi d\mu = \int \varphi(z) \exp(-\beta A(z))\, dz.$$

Question: What is the dynamical content of the free energy A ? Is the effective dynamics

$$dz_t = -A'(z_t)\, dt + \sqrt{2\beta^{-1}} dB_t$$

close to $(\xi(\boldsymbol{X}_t))_{t\geq 0}$? It is thermodynamically consistent ($\xi * \mu = \exp(-\beta A(z))\, dz$) but is it dynamically consistent ?

Construction of the effective dynamics

By Itô, one has

$$d\xi(\boldsymbol{X}_t) = (-\nabla V \cdot \nabla\xi + \beta^{-1}\Delta\xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}}|\nabla\xi(\boldsymbol{X}_t)|\frac{\nabla\xi(\boldsymbol{X}_t)}{|\nabla\xi(\boldsymbol{X}_t)|}\cdot dW_t$$

First attempt:

$$d\tilde{z}_t = \tilde{b}(t,\tilde{z}_t)\, dt + \sqrt{2\beta^{-1}}\tilde{\sigma}(t,\tilde{z}_t)\, dB_t$$

with

$$\tilde{b}(t,\tilde{z}) = \mathbb{E}\left((-\nabla V \cdot \nabla\xi + \beta^{-1}\Delta\xi)(\boldsymbol{X}_t)\middle|\xi(\boldsymbol{X}_t) = \tilde{z}\right)$$

$$\tilde{\sigma}^2(t,\tilde{z}) = \mathbb{E}\left(|\nabla\xi|^2(\boldsymbol{X}_t)\middle|\xi(\boldsymbol{X}_t) = \tilde{z}\right).$$

Then, for all time $t \geq 0$, $\mathcal{L}(\xi(\boldsymbol{X}_t)) = \mathcal{L}(\tilde{z}_t)$! But $\tilde{b}$ and $\tilde{\sigma}$ are untractable numerically...

Construction of the effective dynamics

By Itô, one has

$$d\xi(\boldsymbol{X}_t) = (-\nabla V \cdot \nabla\xi + \beta^{-1}\Delta\xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}}|\nabla\xi(\boldsymbol{X}_t)|\frac{\nabla\xi(\boldsymbol{X}_t)}{|\nabla\xi(\boldsymbol{X}_t)|}\cdot dW_t$$

The effective dynamics:

$$dz_t = b(z_t)\, dt + \sqrt{2\beta^{-1}}\sigma(z_t)\, dB_t$$

with

$$b(z) = \mathbb{E}_\mu\left((-\nabla V \cdot \nabla\xi + \beta^{-1}\Delta\xi)(\boldsymbol{X})\middle|\xi(\boldsymbol{X}) = z\right)$$

$$\sigma^2(z) = \mathbb{E}_\mu\left(|\nabla\xi|^2(\boldsymbol{X})\middle|\xi(\boldsymbol{X}) = z\right).$$

Related approaches: Mori-Zwanzig and projection operator formalism [E/Vanden-Eijnden, ...], asymptotic approaches [Papanicolaou, Freidlin, Pavliotis/Stuart, ...].

Link with the free energy

- The effective dynamics is:

$$dz_t = b(z_t)\,dt + \sqrt{2\beta^{-1}}\sigma(z_t)\,dB_t$$

with $\sigma^2(z) = \mathbb{E}_\mu\left(|\nabla\xi|^2(\boldsymbol{X})\middle|\xi(\boldsymbol{X}) = z\right)$ and

$$b = -\sigma^2 A' + \beta^{-1}\partial_z(\sigma^2).$$

In particular (i) the effective dynamics is reversible wrt $\xi * \mu$ and (ii) if $|\nabla\xi(x)| = 1$, then $\sigma(z) = 1$, and $b(z) = -A'(z)$.

- It is possible to reparameterize the foliation of the configurational space and work with

$$\zeta(x) = h(\xi(x)),$$

where $h'(z) = 1/\sigma(z)$. Then, the effective dynamics associated to $\zeta(x)$ is

$$dz_t = -\mathcal{A}'(z_t)\,dt + \sqrt{2\beta^{-1}}\,dB_t$$

where $\mathcal{A}$ is the free energy associated to ζ.

Error analysis: time marginals

The effective dynamics is reversible wrt $\xi * \mu$.
Moreover, under the assumptions ($\xi(x_1, \dots, x_n) = x_1$ for simplicity):

(H1) The conditional probability measures $\mu(\cdot|\xi(\boldsymbol{x}) = z)$ satisfy a Logarithmic Sobolev Inequality with constant ρ,

(H2) Bounded coupling assumption: $\|\partial_1 \partial_{2,\dots,n} V\|_{L^\infty} \leq \kappa$.

Then, $\exists C > 0$, $\forall t \geq 0$,

$$H(\mathcal{L}(\xi(\boldsymbol{X}_t)), \mathcal{L}(z_t)) \leq C \frac{\kappa}{\rho} \Big(H(\mathcal{L}(\boldsymbol{X}_0)|\mu) - H(\mathcal{L}(\boldsymbol{X}_t)|\mu) \Big).$$

If ρ is large (timescale decoupling assumption), the error is small.
The proof [Legoll, TL] is inspired by a decomposition of the entropy proposed in [Grunewald/Otto/Villani/Westdickenberg], and entropy estimates.

Proof (1/5)

For simplicity: 2d case and $\xi(x, y) = x$.

Since $\sigma = 1$, the effective dynamics is

$$dz_t = -A'(z_t)\, dt + \sqrt{2/\beta}\, dB_t$$

where the free energy A is defined by

$$\exp(-\beta A(x)) = \int_{\mathbb{R}} \psi_\infty(x, y)\, dy = Z^{-1} \int_{\mathbb{R}} \exp(-\beta V(x, y))\, dy$$

so that

$$A'(x) = \frac{\int_{\mathbb{R}} \partial_x V(x, y) \exp(-\beta V(x, y))\, dy}{\int_{\mathbb{R}} \exp(-\beta V(x, y))\, dy} = \frac{\int_{\mathbb{R}} \partial_x V(x, y) \psi_\infty(x, y)\, dy}{\int_{\mathbb{R}} \psi_\infty(x, y)\, dy}$$

$$= \mathbb{E}_\mu \left[\partial_x V(X) \,|\, \xi(X) = x\right]$$

Proof (2/5)

The FP equation associated with the gradient dynamics $dX_t = -\nabla V(X_t)\,dt + \sqrt{2\beta^{-1}}dW_t$ writes

$$\partial_t \psi = \mathrm{div}\,(\nabla V \psi) + \beta^{-1}\Delta\psi$$

Let $\overline{\psi}(t,x) = \int_{\mathbb{R}} \psi(t,x,y)\,dy$ be the law of $\xi(X_t)$. We have

$$\partial_t \overline{\psi} = \partial_x \left(\tilde{b}(t,x)\overline{\psi}\right) + \beta^{-1}\partial_{x,x}\overline{\psi}$$

where

$$\tilde{b}(t,x) = \frac{\int_{\mathbb{R}} \partial_x V(x,y)\psi(t,x,y)\,dy}{\int_{\mathbb{R}} \psi(t,x,y)\,dy} = \frac{\int_{\mathbb{R}} \partial_x V(x,y)\psi(t,x,y)\,dy}{\overline{\psi}(t,x)}.$$

On the other hand, the density $\phi(t,x)$ of z_t satisfies

$$\partial_t \phi = \partial_x \left(A'(x)\phi\right) + \beta^{-1}\partial_{x,x}\phi.$$

Question: is ϕ close to $\overline{\psi}$?

Proof (3/5)

Consider

$$E(t) = H(\overline{\psi}|\phi) = \int_{\mathbb{R}} \ln\left(\frac{\overline{\psi}(t,x)}{\phi(t,x)}\right) \overline{\psi}(t,x)\, dx.$$

$$\begin{aligned}
\frac{dE}{dt} &= -\int_{\mathbb{R}} \frac{\overline{\psi}}{\phi}\, \partial_t \phi + \int_{\mathbb{R}} \ln\left(\frac{\overline{\psi}}{\phi}\right) \partial_t \overline{\psi} \\
&= -\beta^{-1} I(\overline{\psi}|\phi) + \int_{\mathbb{R}} \overline{\psi}\, \partial_x \left(\ln \frac{\overline{\psi}}{\phi}\right) \left(A' - \tilde{b}\right) \\
&\leq -\beta^{-1} I(\overline{\psi}|\phi) + \frac{1}{2\alpha} \int_{\mathbb{R}} \overline{\psi} \left(\partial_x \left(\ln \frac{\overline{\psi}}{\phi}\right)\right)^2 + \frac{\alpha}{2} \int_{\mathbb{R}} \overline{\psi} \left(A' - \tilde{b}\right)^2 \\
&= \left(\frac{1}{2\alpha} - \beta^{-1}\right) I(\overline{\psi}|\phi) + \frac{\alpha}{2} \int_{\mathbb{R}} \overline{\psi} \left(A' - \tilde{b}\right)^2 = \frac{\beta}{4} \int_{\mathbb{R}} \overline{\psi} \left(A' - \tilde{b}\right)^2,
\end{aligned}$$

by choosing $\alpha = \beta/2$.

Proof (4/5)

Let $\pi^x(dy_1, dy_2)$ be a coupling measure with marginals $\nu_t^x = \frac{\psi(t,x,y)}{\psi(t,x)}$ and $\nu_\infty^x = \frac{\psi_\infty(x,y)}{\psi_\infty(x)}$. We have, using (H2),

$$\left|A'(x) - \tilde{b}(t,x)\right| = \left|\int_{\mathbb{R}^2} \left(\partial_x V(x,y_1) - \partial_x V(x,y_2)\right) \pi^x(dy_1, dy_2)\right|$$
$$\leq \|\partial_{xy} V\|_{L^\infty} \int_{\mathbb{R}^2} |y_1 - y_2| \, \pi^x(dy_1, dy_2).$$

Taking the infimum on $\pi^x \in \Pi\left(\nu_t^x, \nu_\infty^x\right)$,

$$\left|A'(x) - \tilde{b}(t,x)\right| \leq \|\partial_{xy} V\|_{L^\infty} \; W_1(\nu_t^x, \nu_\infty^x).$$

We now use the LSI on ν_∞^x (H1) and Talagrand inequality to get

$$\left|A'(x) - \tilde{b}(t,x)\right| \leq \frac{\|\partial_{xy} V\|_{L^\infty}}{\rho} \sqrt{I(\nu_t^x | \nu_\infty^x)}$$

Proof (5/5)

We thus have

$$\int_{\mathbb{R}} \overline{\psi}(t,x)\left(A'(x)-\tilde{b}(t,x)\right)^2 dx \leq \frac{\|\partial_{xy}V\|_{L^\infty}^2}{\rho^2}\int_{\mathbb{R}} \overline{\psi} I(\nu_t^x|\nu_\infty^x)$$
$$\leq \frac{\|\partial_{xy}V\|_{L^\infty}^2}{\rho^2} I(\psi|\psi_\infty).$$

Plugging this in the entropy estimate, we get

$$\frac{dE}{dt} \leq \frac{\beta}{4}\frac{\|\partial_{xy}V\|_{L^\infty}^2}{\rho^2} I(\psi|\psi_\infty)$$
$$= -\frac{\beta^2\|\partial_{xy}V\|_{L^\infty}^2}{4\rho^2}\frac{dH(\psi|\psi_\infty)}{dt}.$$

Integrating in time (since $E(0)=0$):

$$\forall t \geq 0,\ E(t) \leq \frac{\beta^2\|\partial_{xy}V\|_{L^\infty}^2}{4\rho^2}\left(H(\psi(0)|\psi_\infty) - H(\psi(t)|\psi_\infty)\right).$$

Entropy techniques

Other results based on this set of assumptions:

- [TL, JFA 2008] LSI for the cond. meas. $\mu_{\Sigma(z)}$
  + LSI for the marginal $\overline{\mu}(dz) = \xi * \mu(dz)$
  + bdd coupling ($\|\nabla_{\Sigma(z)} f\|_{L^\infty} < \infty$) $\Longrightarrow$ LSI for μ.
- [TL, Rousset, Stoltz Nonlinearity, 2008] Analysis of the adaptive biasing force method ($\xi(x_1, \ldots, x_n) = x_1$):

$$\begin{cases} dX_t = -\nabla(V - A_t \circ \xi)(X_t)\, dt + \sqrt{2\beta^{-1}} dW_t\,, \\ A_t'(z) = \mathbb{E}(\partial_1 V(X_t) | \xi(X_t) = z). \end{cases}$$

Error analysis: trajectories

Under the assumptions ($\xi(x_1, \dots, x_n) = x_1$ for simplicity):

(H1') The conditional probability measures $\mu(\cdot|\xi(\boldsymbol{x}) = z)$ satisfy a Poincaré inequality with constant ρ,

(H2') Bounded coupling assumption: $\|\partial_1 \partial_{2,\dots,n} V\|_{L^2(\mu)} \leq \kappa$,

(H3) b is one-sided Lipschitz ($-b' \leq L_b$) and such that

$$\int_{\mathbb{R}^d} \left(\sup_{y \in [-|x|,|x|]} |b'(y)| \right)^2 \mu(dx) < \infty.$$

Then, if $z_0 = \xi(\boldsymbol{X}_0)$ is distributed according to a measure μ_0 such that $\frac{d\mu_0}{d\mu} \in L^\infty$,

$$\mathbb{E}\left(\sup_{t \in [0,T]} |\xi(X_t) - z_t| \right) \leq C \frac{\kappa}{\rho}$$

The proof [Legoll, TL, Olla] uses a probabilistic arguments (Poisson equations, and Doob's martingale inequality).

Remark: Application to averaging principle

These techiques can be used to obtain quantitative results for averaging principles. Let us consider

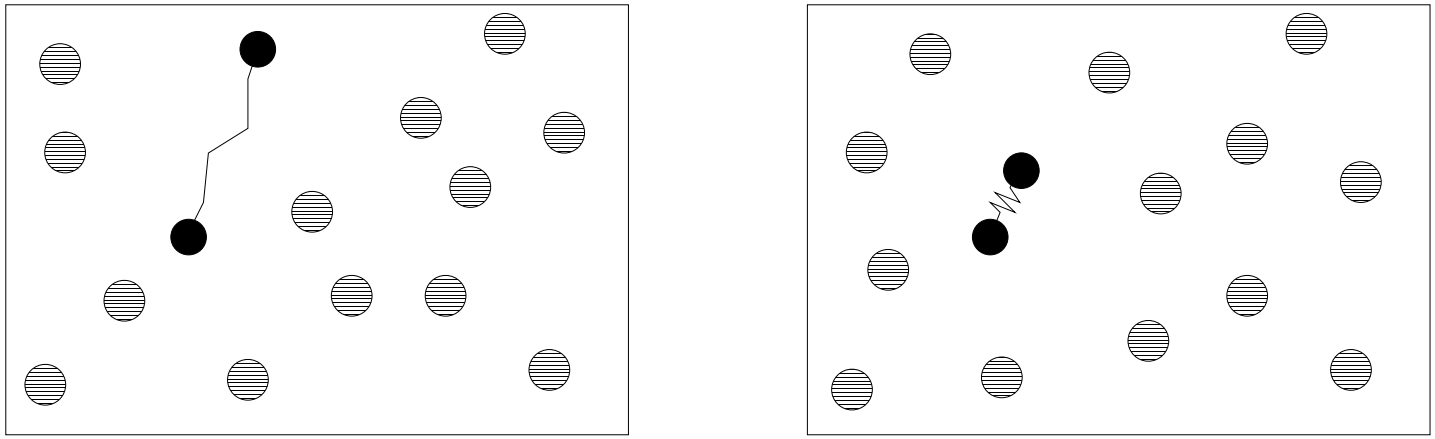
$$\begin{cases} dX_t^{1,\varepsilon} = -\partial_1 V(X_t^\varepsilon)\, dt + \sqrt{2\beta^{-1}} dW_t^1 \\ dX_t^{i,\varepsilon} = -\dfrac{\partial_i V(X_t^\varepsilon)}{\varepsilon}\, dt + \sqrt{\dfrac{2\beta^{-1}}{\varepsilon}} dW_t^i \qquad \text{for } i = 2, \dots, n \end{cases}$$

Then, under the assumptions of the former result:

$$\mathbb{E}\left(\sup_{0 \le t \le T} \left| X_t^{\varepsilon,1} - \xi_t \right| \right) \le C \sqrt{\beta\varepsilon}\, \frac{\kappa}{\rho}.$$

Notice that we do not assume b globally Lipschitz.

Numerical illustration: dimer in a solvant



- Solvent-solvent, solvent-monomer: truncated LJ for $\|x_i - x_j\|$,
- Monomer-monomer: double well potential for $\|x_1 - x_2\|$.

ξ is the distance between the two monomers: $\xi(x) = \|x_1 - x_2\|$.

Transition times from the compact to the stretched state:

β	Reference	Eff. dyn.	Dyn. based on A
0.5	262 ± 6	245 ± 5	504 ± 11
0.25	1.81 ± 0.04	1.68 ± 0.04	3.47 ± 0.08

A multiscale parareal algorithm: the main idea

How to use the effective dynamics ?

One idea: an algorithm which consists in coupling a macroscopic model with a microscopic model in order to efficiently generate the microscopic dynamics. The macroscopic model is the predictor, and the microscopic model is the corrector.

A multiscale parareal algorithm

The basic algorithm is very much inspired by the parareal algorithm [Lions, Maday Turinici].

Let us present it in a very simple setting: Consider the fine dynamics:

$$\begin{cases} \dot{x} = \alpha x + p^T y, \\ \dot{y} = \dfrac{1}{\epsilon}\left(qx - Ay\right). \end{cases}$$

The variable $u = (x, y)$ contains a slow variable $x \in \mathbb{R}$ and a fast variable $y \in \mathbb{R}^{d-1}$. Associated to that dynamics, we have an expensive fine propagator $F_{\Delta t}$ over the time range Δt.

Under appropriate assumption, a natural coarse dynamics ($\epsilon \to 0$) on x is:

$$\dot{X} = (\alpha + p^T A^{-1} q) X.$$

Associated to that dynamics, we have a cheap coarse propagator $G_{\Delta t}$ over the time range Δt.

The algorithm

Let us now present the algorithm

Let $u(0) = u_0$ be the initial condition.

1. Initialization:
   a) Compute $\{X_0^n\}_{0 \leq n \leq N}$ sequentially by using the coarse propagator:

$$X_0^0 = \mathcal{R}(u_0), \qquad X_0^{n+1} = G_{\Delta t}(X_0^n).$$

   b) Lift the macroscopic approximation to the microscopic level:

$$u_0^0 = u_0 \quad \text{and, for all } 1 \leq n \leq N, \quad u_0^n = \mathcal{L}(X_0^n).$$

2. Assume that, for some $k \geq 0$, the sequences $\{u_k^n\}_{0\leq n\leq N}$ and $\{X_k^n\}_{0\leq n\leq N}$ are known. Then, at the iteration $k+1$:
   a) For all $0 \leq n \leq N-1$, compute (in parallel) using the coarse and the fine-scale propagators

$$\overline{X}_k^{n+1} = G_{\Delta t}(X_k^n), \quad \overline{u}_k^{n+1} = F_{\Delta t}(u_k^n).$$

   b) For all $0 \leq n \leq N-1$, evaluate the jumps (the difference between the two propagated values) *at the macroscopic level*:

$$J_k^{n+1} = \mathcal{R}(\overline{u}_k^{n+1}) - \overline{X}_k^{n+1}.$$

   c) Compute $\left\{X_{k+1}^n\right\}_{0\leq n\leq N}$ sequentially by

$$X_{k+1}^0 = \mathcal{R}(u_0), \qquad X_{k+1}^{n+1} = G_{\Delta t}(X_{k+1}^n) + J_k^{n+1}.$$

   d) Compute $\left\{u_{k+1}^{n+1}\right\}_{0\leq n\leq N-1}$ by matching the result of the local microscopic computation, $\overline{u}_k^{n+1}$, on the corrected macroscopic state X_{k+1}^{n+1}:

$$u_{k+1}^0 = u_0 \quad \text{and, for all } 0 \leq n \leq N-1, \quad u_{k+1}^{n+1} = \mathcal{P}(X_{k+1}^{n+1}, \overline{u}_k^{n+1}).$$

Error analysis

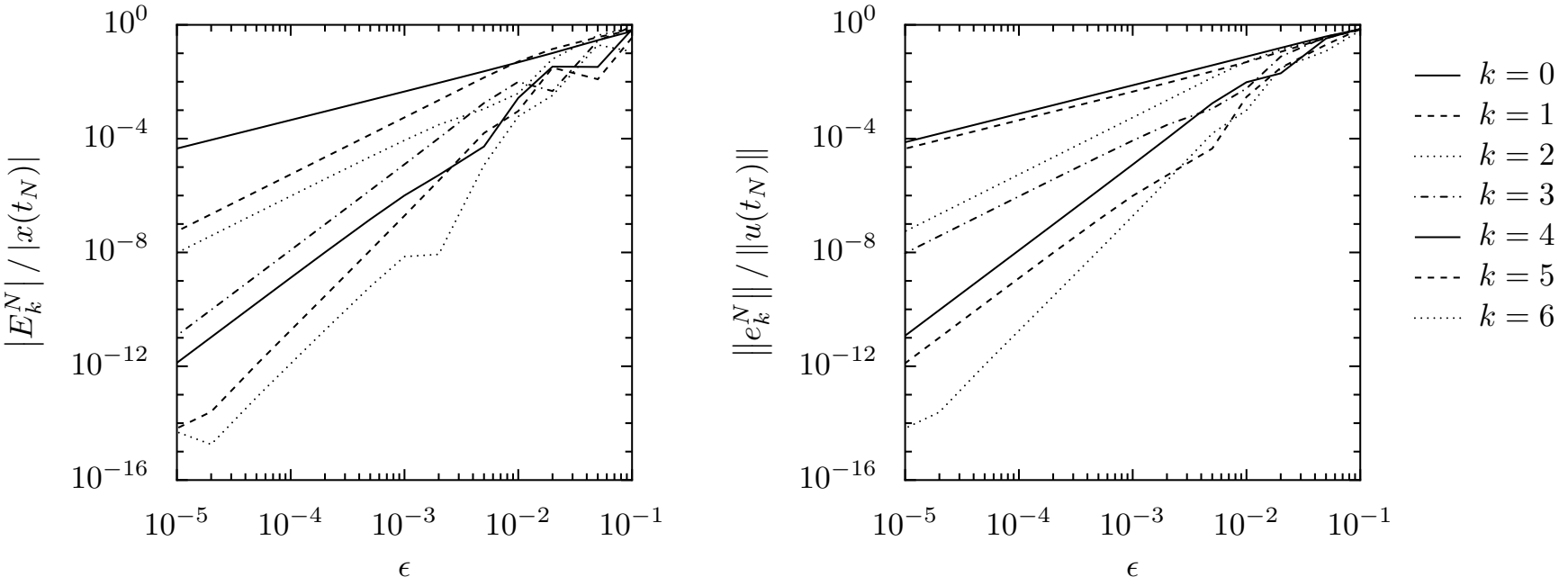
In the simple setting above, the restriction, lifting and matching operators we use are: $\mathcal{R}(x,y) = x$, $\mathcal{L}(X) = (X, (A^{-1}q)X)$ and $\mathcal{P}(X,(x,y)) = (X,y)$.

One can prove the following [Legoll, TL, Samaey]: Let us assume that $F_{\Delta t}$ and $G_{\Delta t}$ are exact (no time-discretization error). The errors on the macroscopic variable and the microscopic variable are:

$$\text{for all } k \geq 0, \quad \sup_{0 \leq n \leq N} |X_k^n - \mathcal{R}u(t_n)| \leq C\epsilon^{1+\lceil k/2 \rceil},$$

$$\text{for all } k \geq 0, \quad \sup_{0 \leq n \leq N} \|u_k^n - u(t_n)\| \leq C\epsilon^{1+\lfloor k/2 \rfloor},$$

Numerical results



Similar results on non linear ODEs such as the Brusselator problem.

Work in progress: extensions to PDEs and stochastic problems.

References

Some papers I mentioned:

- F. Legoll and T. Lelièvre, *Effective dynamics using conditional expectations*, Nonlinearity, 2010.
- F. Legoll, T. Lelièvre and G. Samaey, *A micro-macro parareal algorithm: application to singularly perturbed ordinary differential equations*, SIAM Journal on Scientific Computing, 2013.
- F. Legoll, T. Lelièvre and S. Olla, *Pathwise estimates for an effective dynamics*, `https://hal.archives-ouvertes.fr/hal-01314221` .

Accelerated dynamics

Let us consider the overdamped Langevin dynamics:

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

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 $(\mathcal{S}(\boldsymbol{X}_t))_{t\geq 0}$.

This is 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.

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: Quantitative 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 $L^* = \operatorname{div}(\nabla V + \beta^{-1}\nabla)$ 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 $\varphi : \partial W \to \mathbb{R}$,

$$\mathbb{E}^{\nu}(\varphi(\boldsymbol{X}_{T_W})) = -\frac{\displaystyle\int_{\partial W} \varphi\, \partial_n u_1 \, d\sigma}{\beta\lambda_1 \displaystyle\int_W u_1(x)\, dx}.$$

Link with kinetic Monte Carlo models (1/2)

Starting from the QSD in W, the exit event from W is Markovian: it can be rewritten as one step of a Markov jump process (kinetic Monte Carlo or Markov state model):

Link with kinetic Monte Carlo models (2/2)

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

$$p(i) = \mathbb{P}(\boldsymbol{X}_{T_W} \in \partial W_i) = -\frac{\displaystyle\int_{\partial W_i} \partial_n u_1 \, d\sigma}{\beta\lambda_1 \displaystyle\int_W u_1(x)\, dx}.$$

To each possible exit region ∂W_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 region is $\arg\min(\tau_1, \ldots, \tau_I)$.

Escaping from a metastable state

How to use these properties to design 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 ?

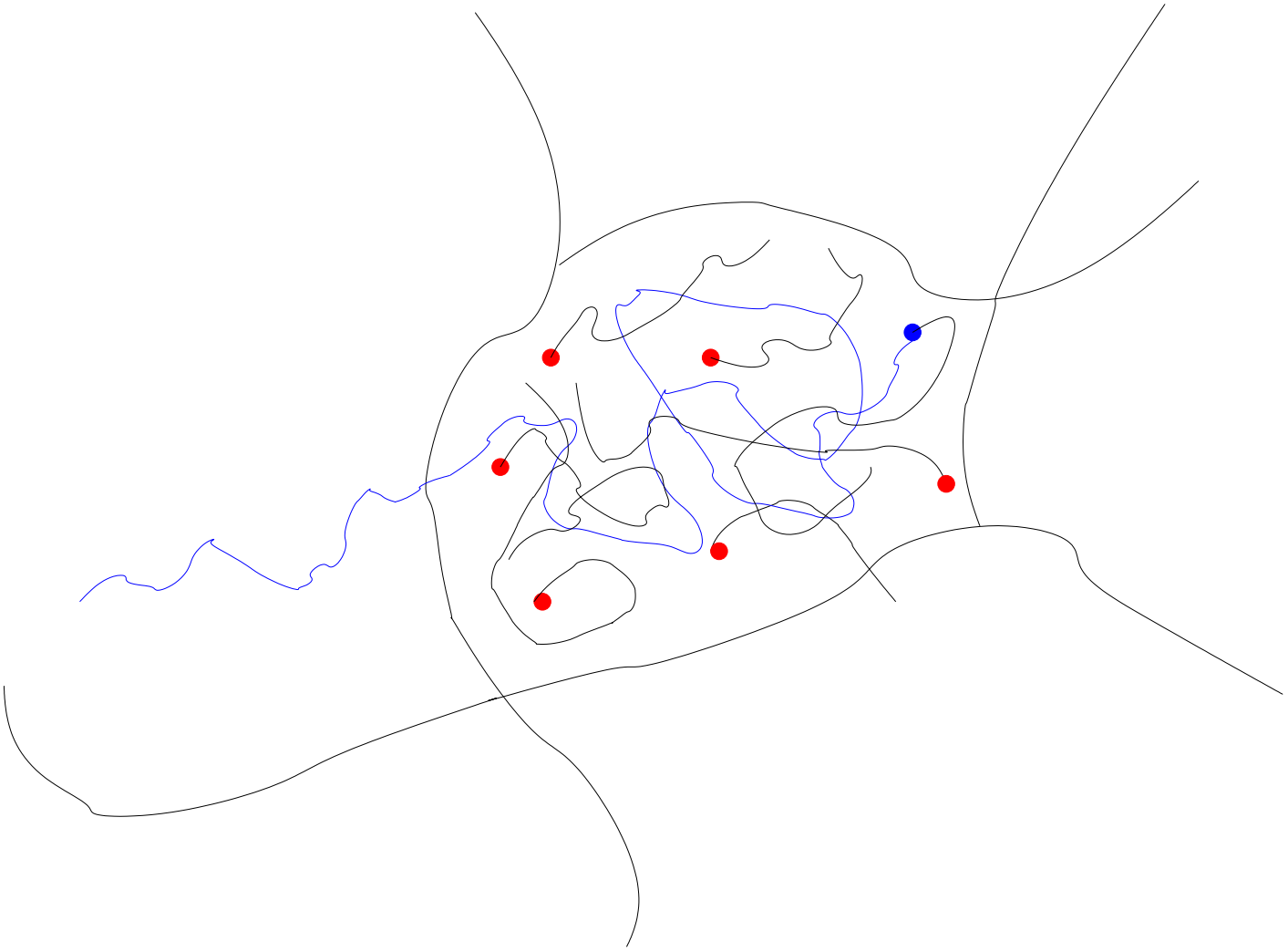
Remark: In practice, one needs to:

- 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

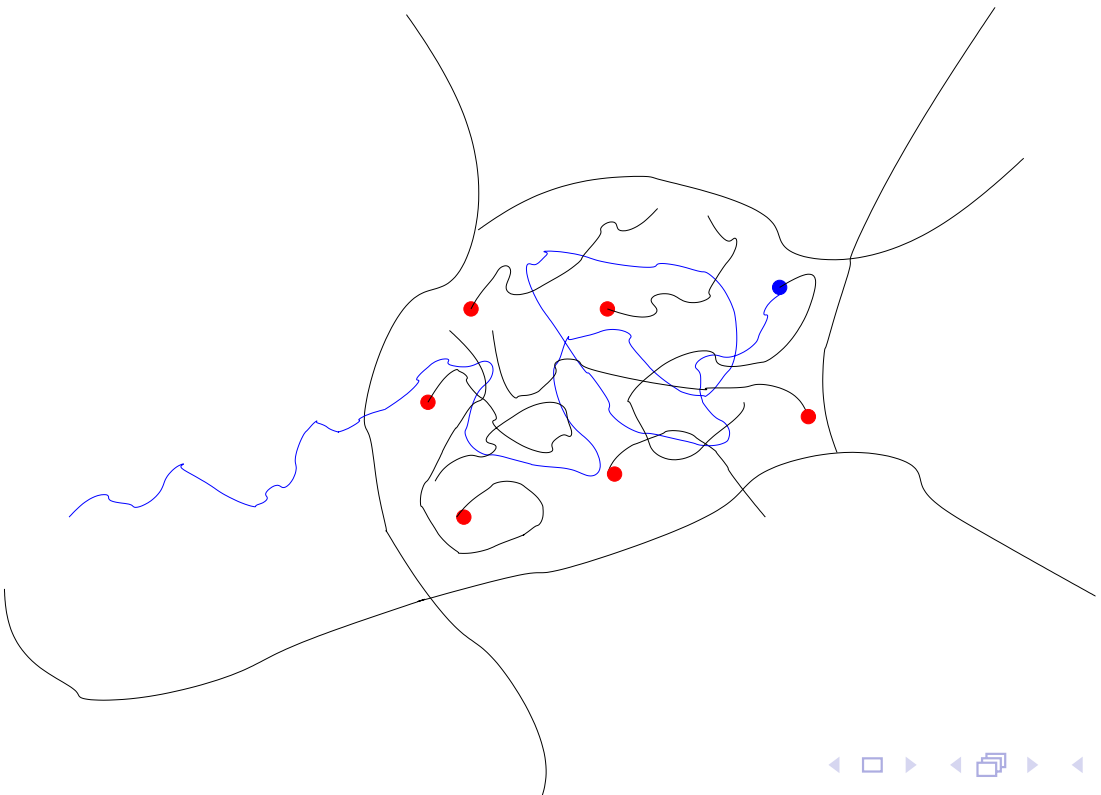


The Parallel Replica Algorithm

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

$$\boldsymbol{X}^{I_0}_{T^{I_0}_W} \stackrel{\mathcal{L}}{=} \boldsymbol{X}^{1}_{T^{1}_W},$$

  where $I_0 = \arg\min_i(T^i_W)$;

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

$$N \min(T^1_W, \ldots, T^N_W) \stackrel{\mathcal{L}}{=} T^1_W.$$

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,

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

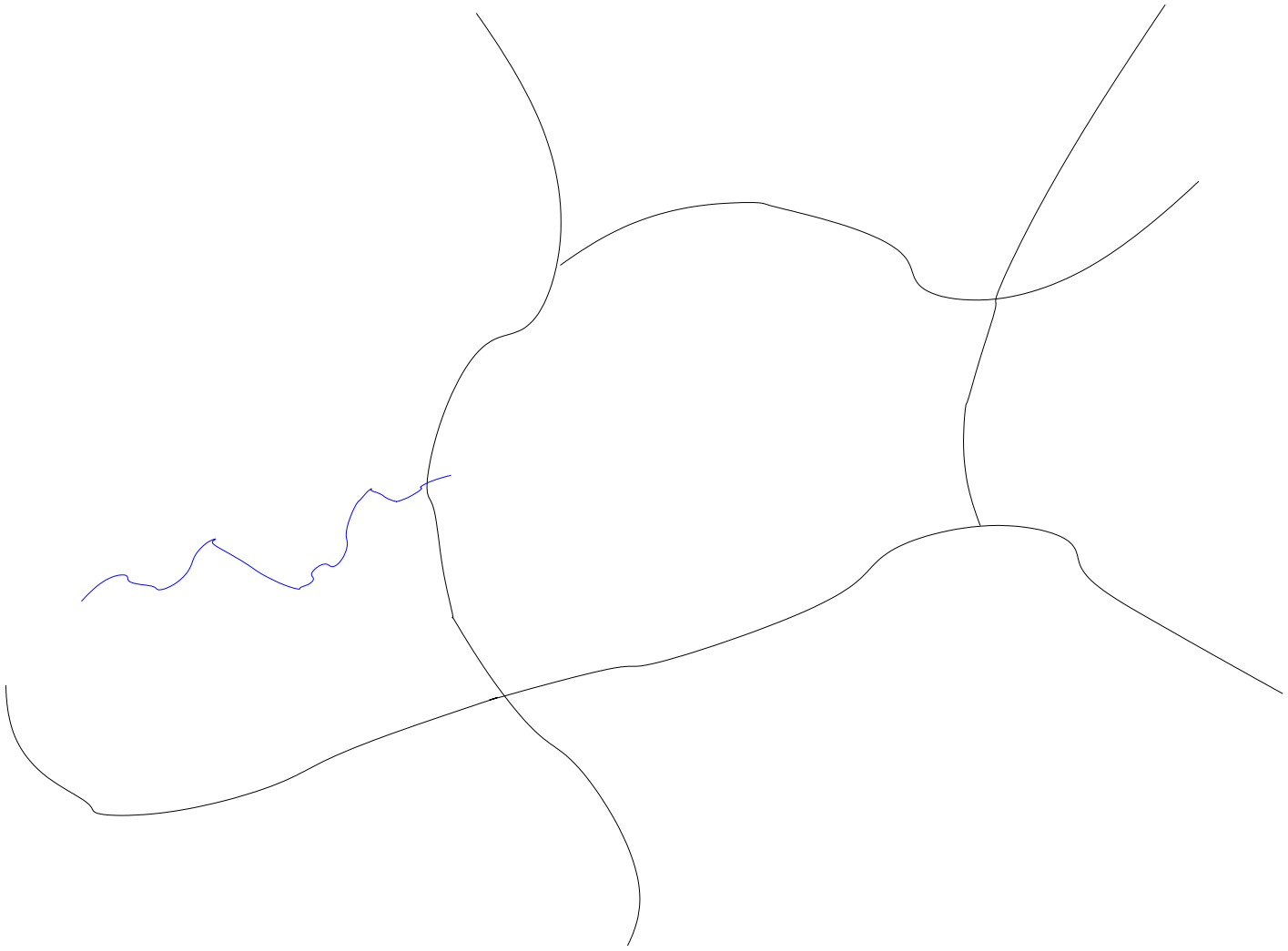
The Parallel Replica Algorithm

The full algorithm is in three steps:

- Decorrelation step
- Dephasing step
- Parallel step

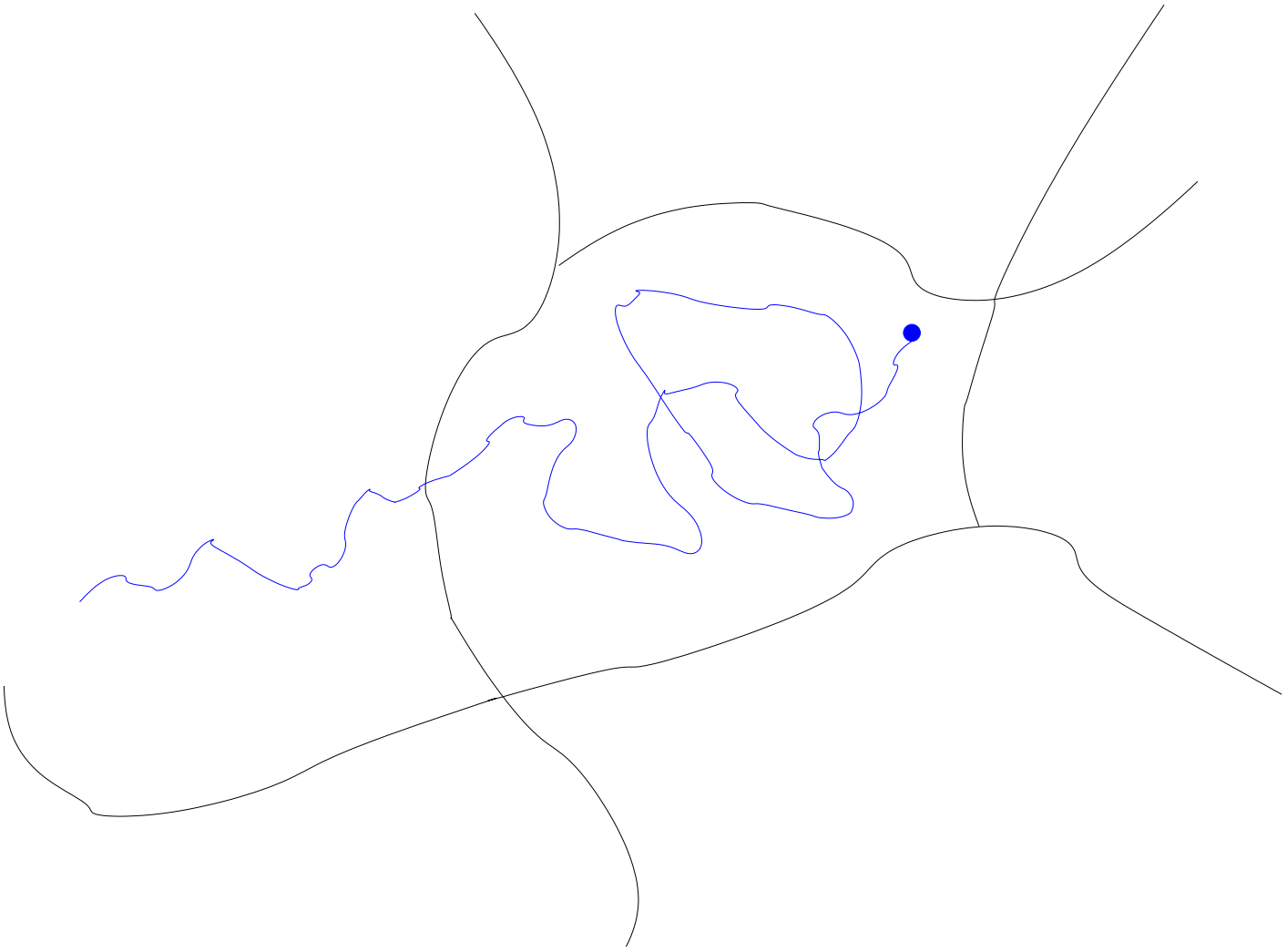
The Parallel Replica Algorithm

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



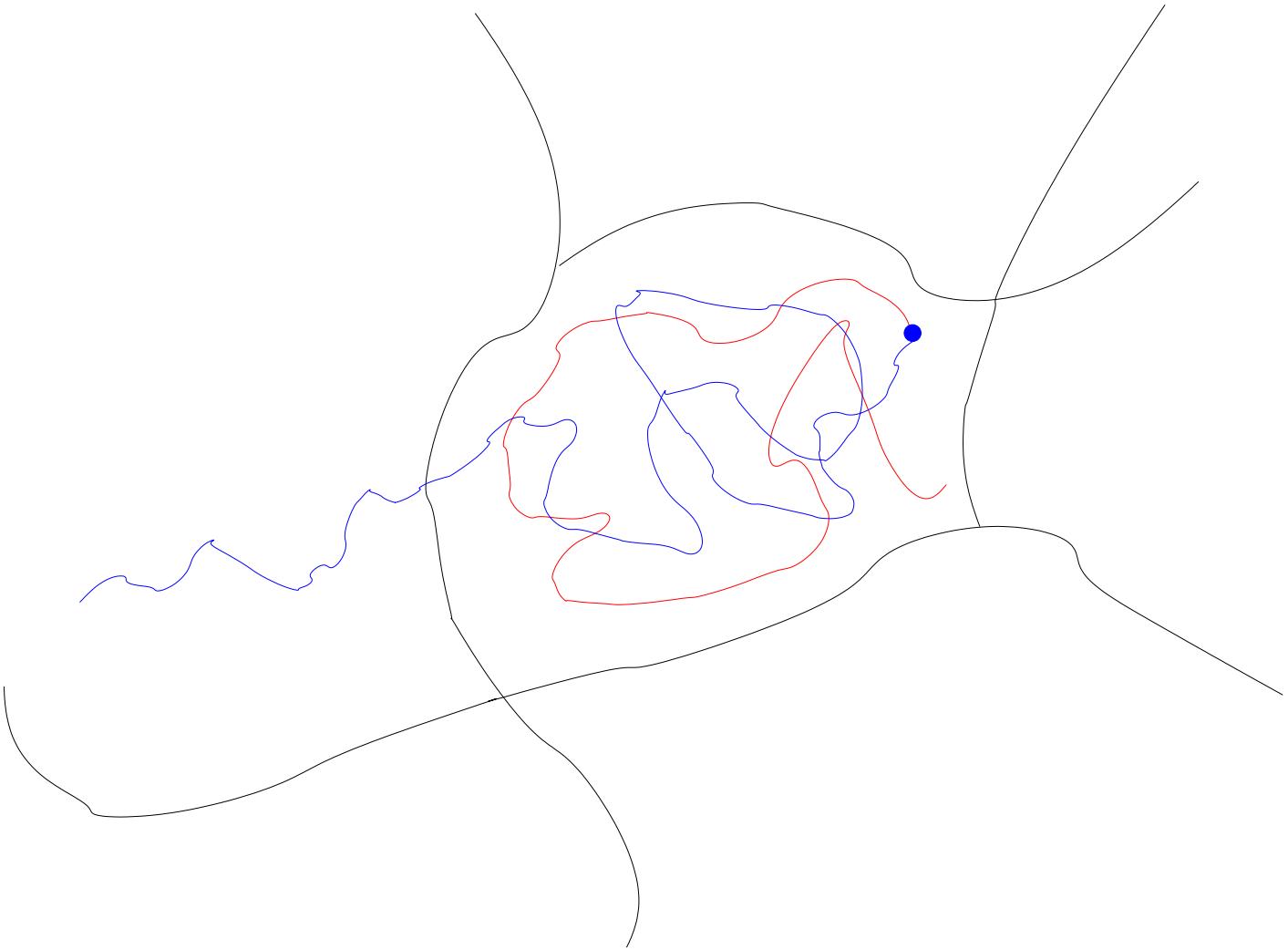
The Parallel Replica Algorithm

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



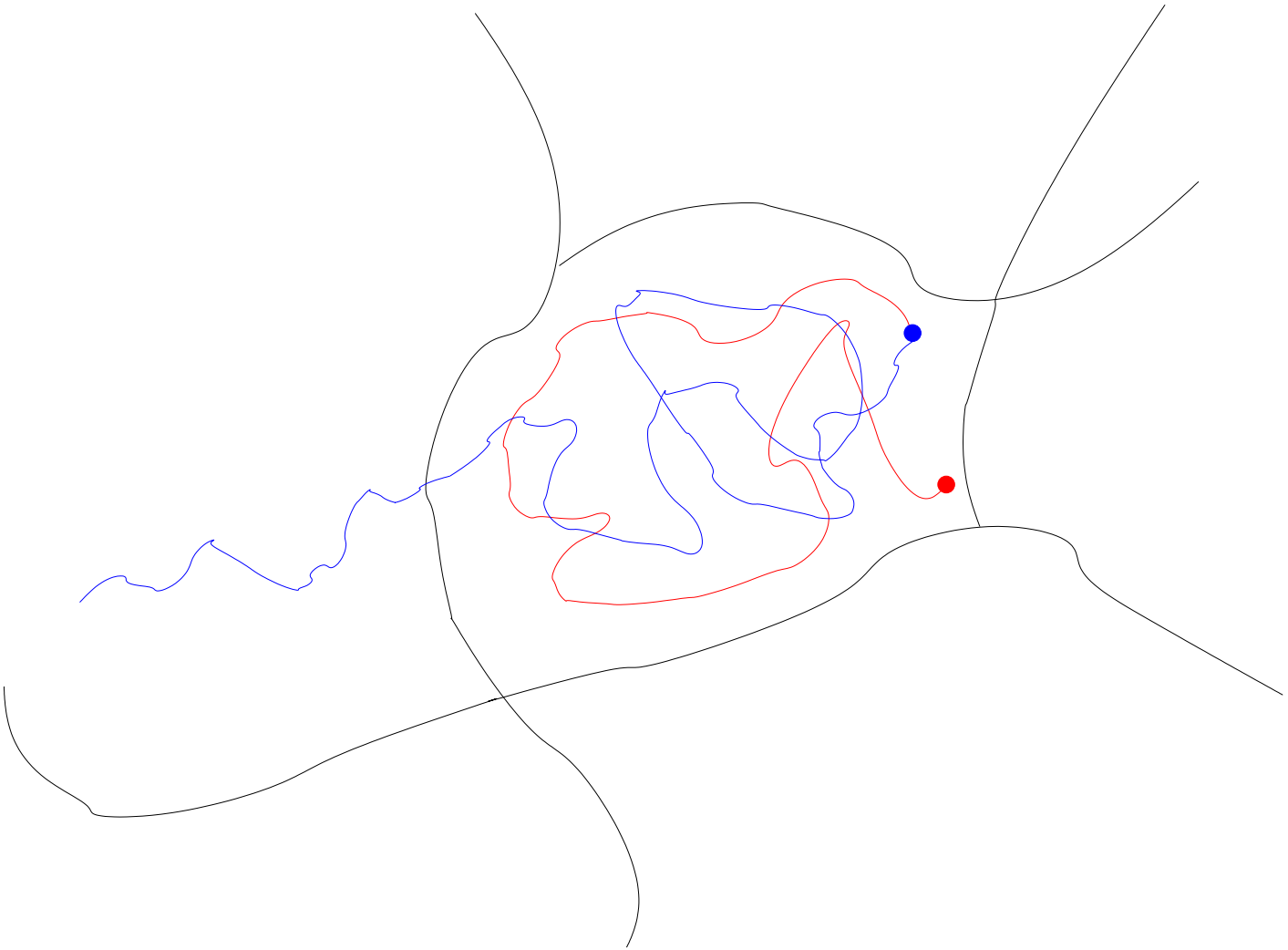
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



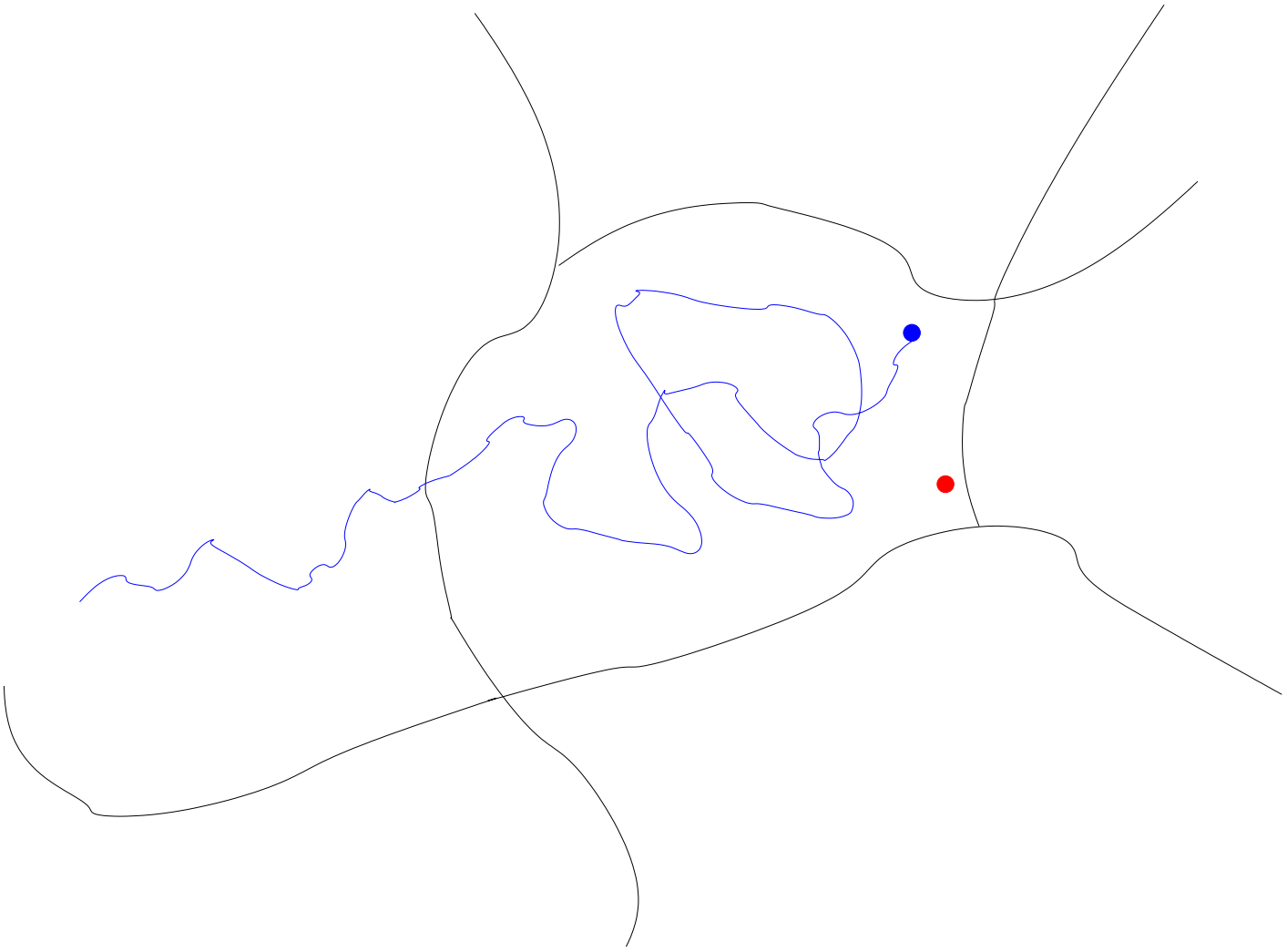
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



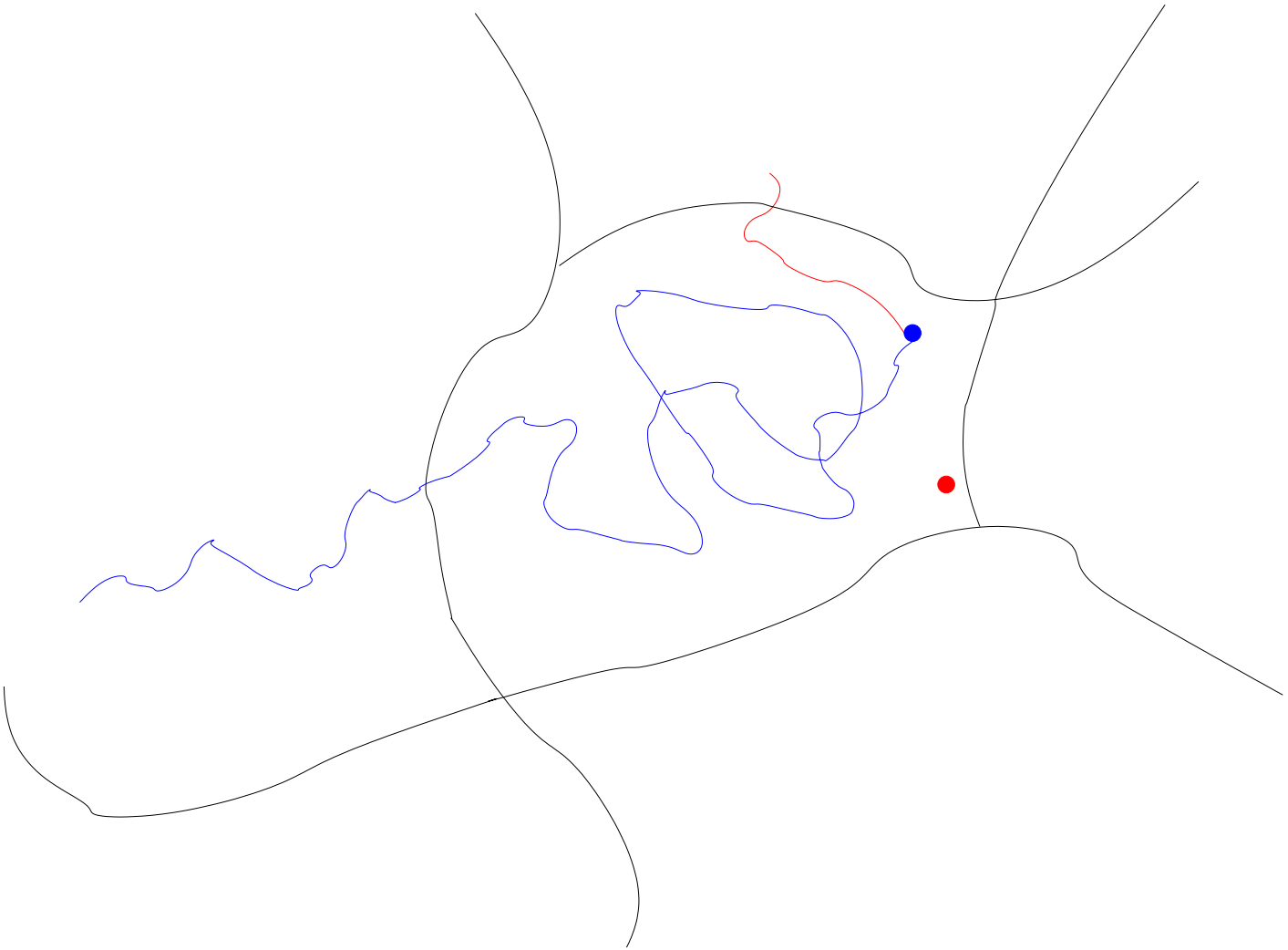
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



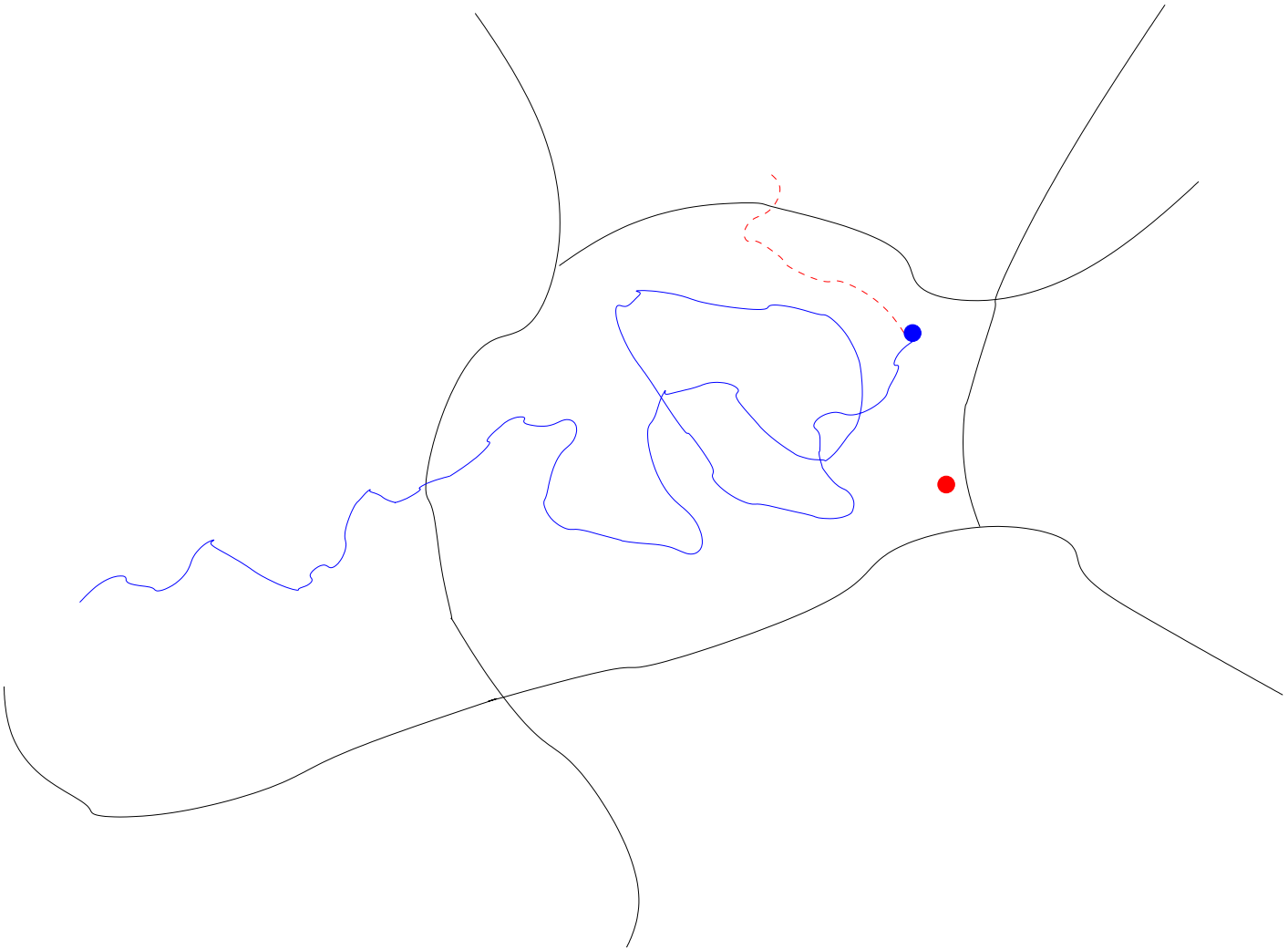
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



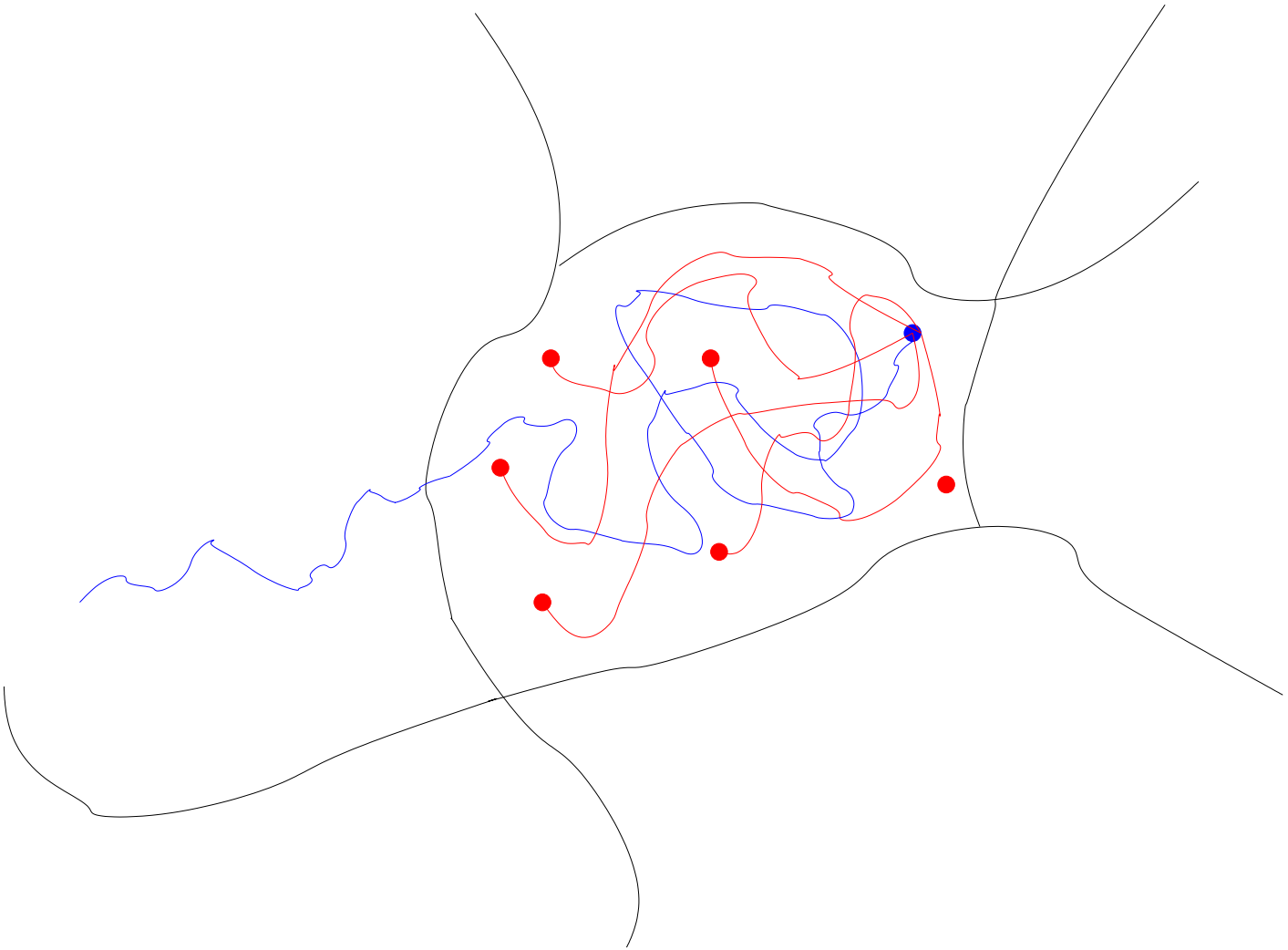
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



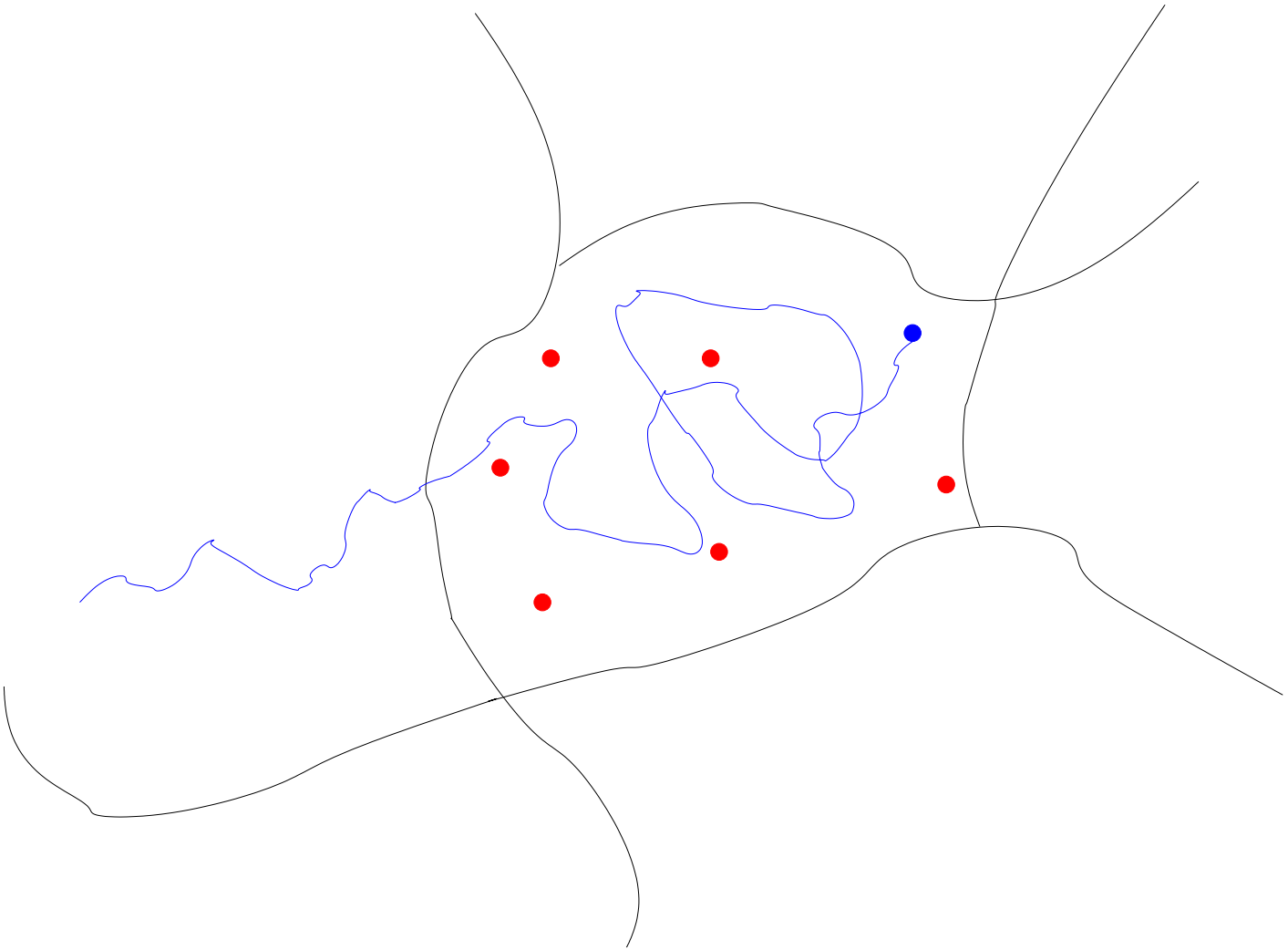
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



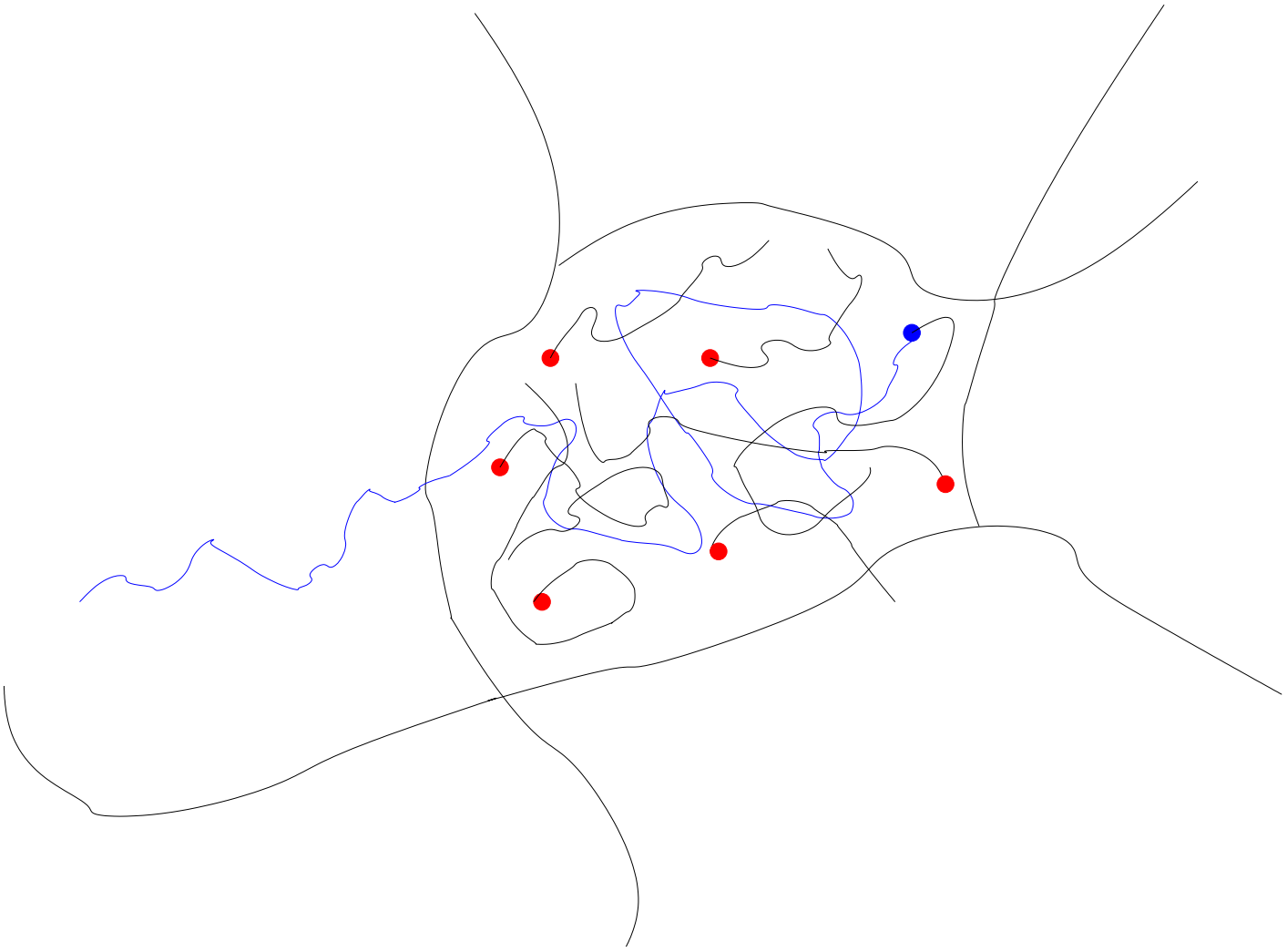
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



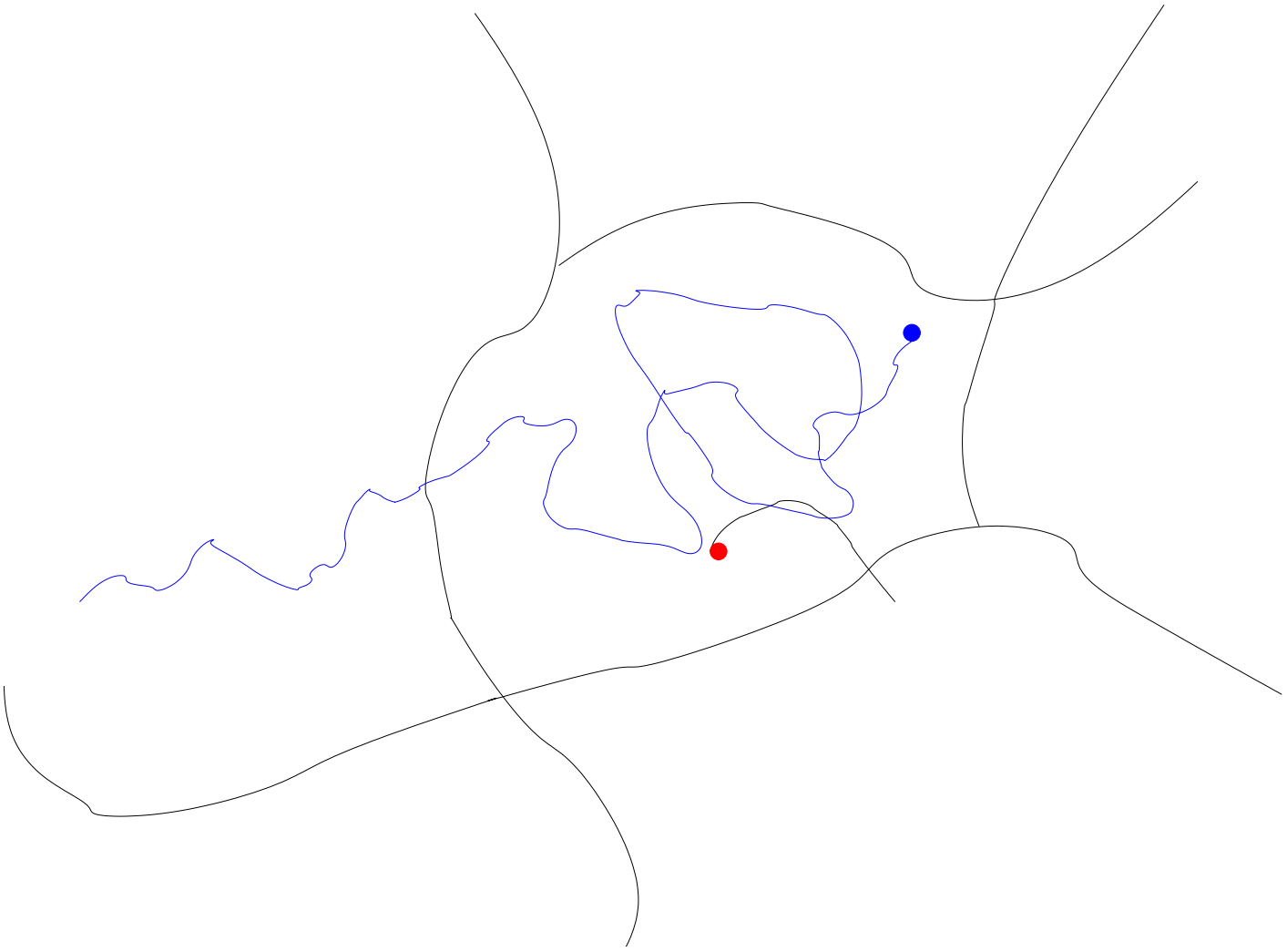
The Parallel Replica Algorithm

Parallel step: run independent trajectories in parallel...



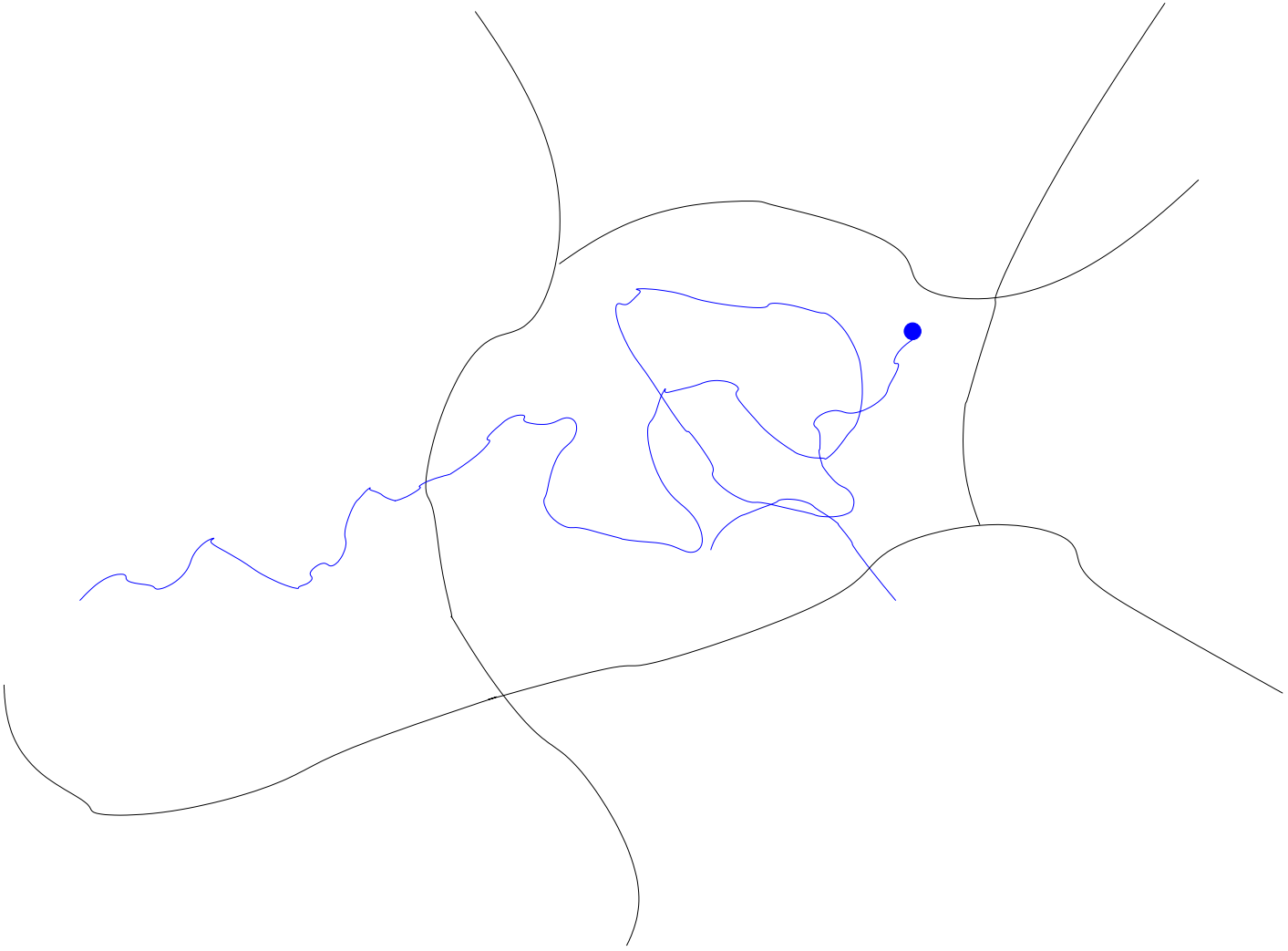
The Parallel Replica Algorithm

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



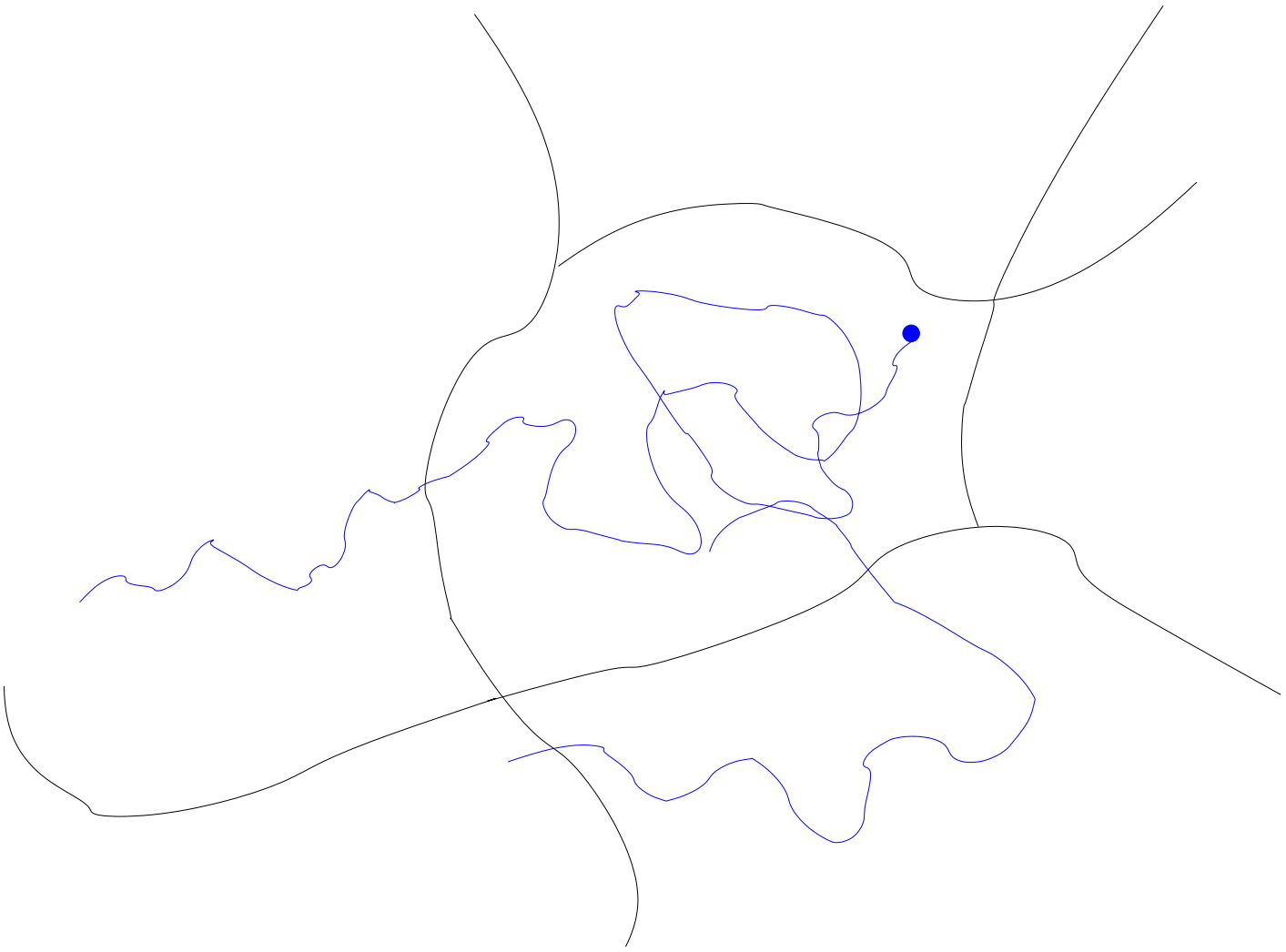
The Parallel Replica Algorithm

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



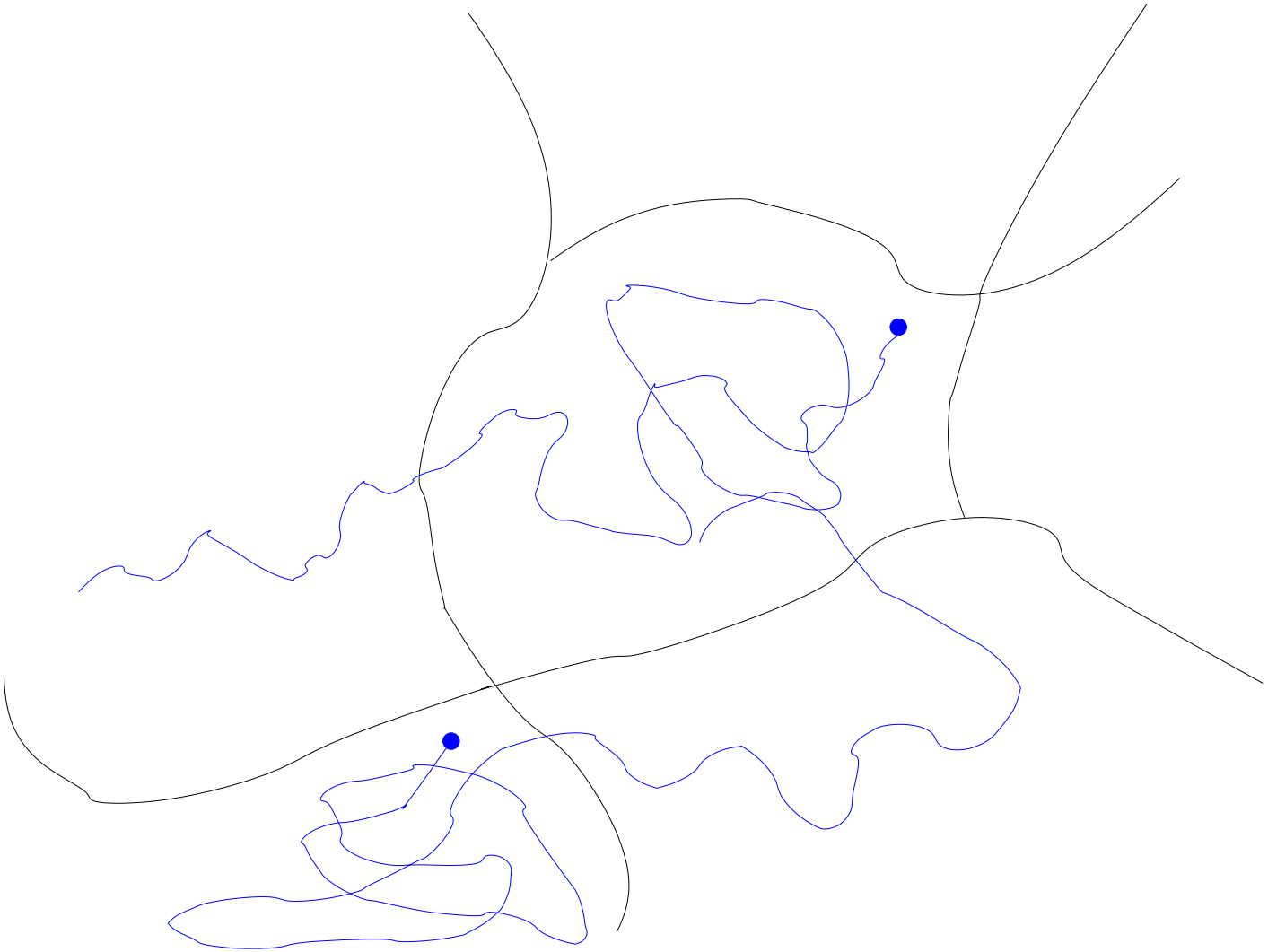
The Parallel Replica Algorithm

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.

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 $t \geq \frac{C}{\lambda_2 - \lambda_1}$,

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

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 recently proposed 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:

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

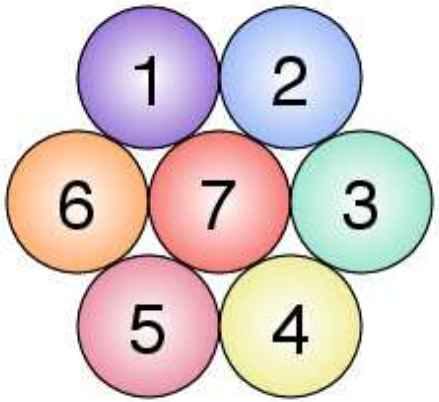
The generalized Parallel Replica algorithm

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.

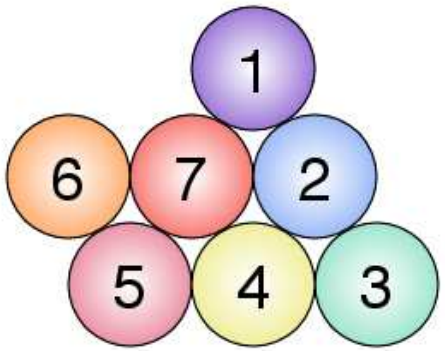
If the Fleming Viot particle process reaches stationarity before the reference walker, go to the parallel step. Otherwise, restart a new decorrelation / dephasing step.

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

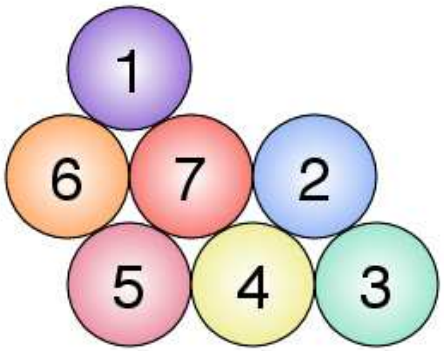
Numerical test case: the 7 atoms LJ cluster



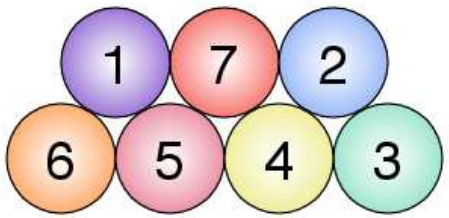
(a) C_0, $V = -12.53$



(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

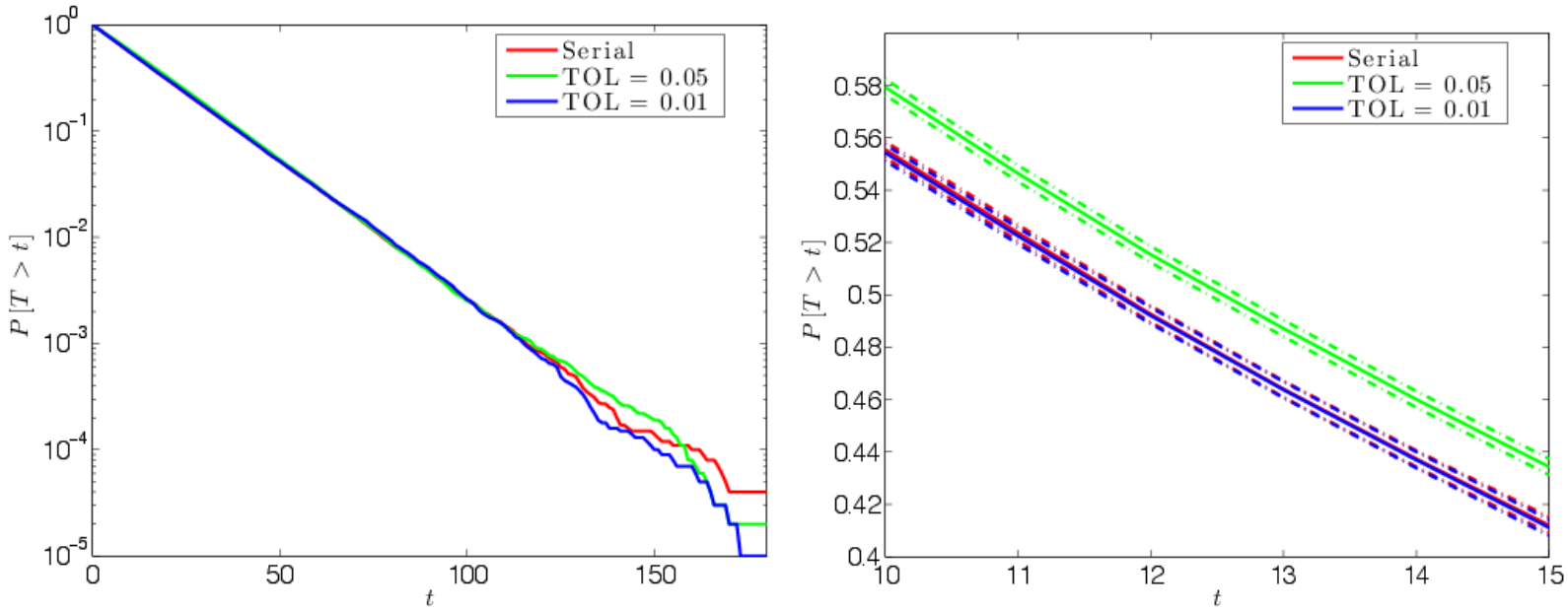
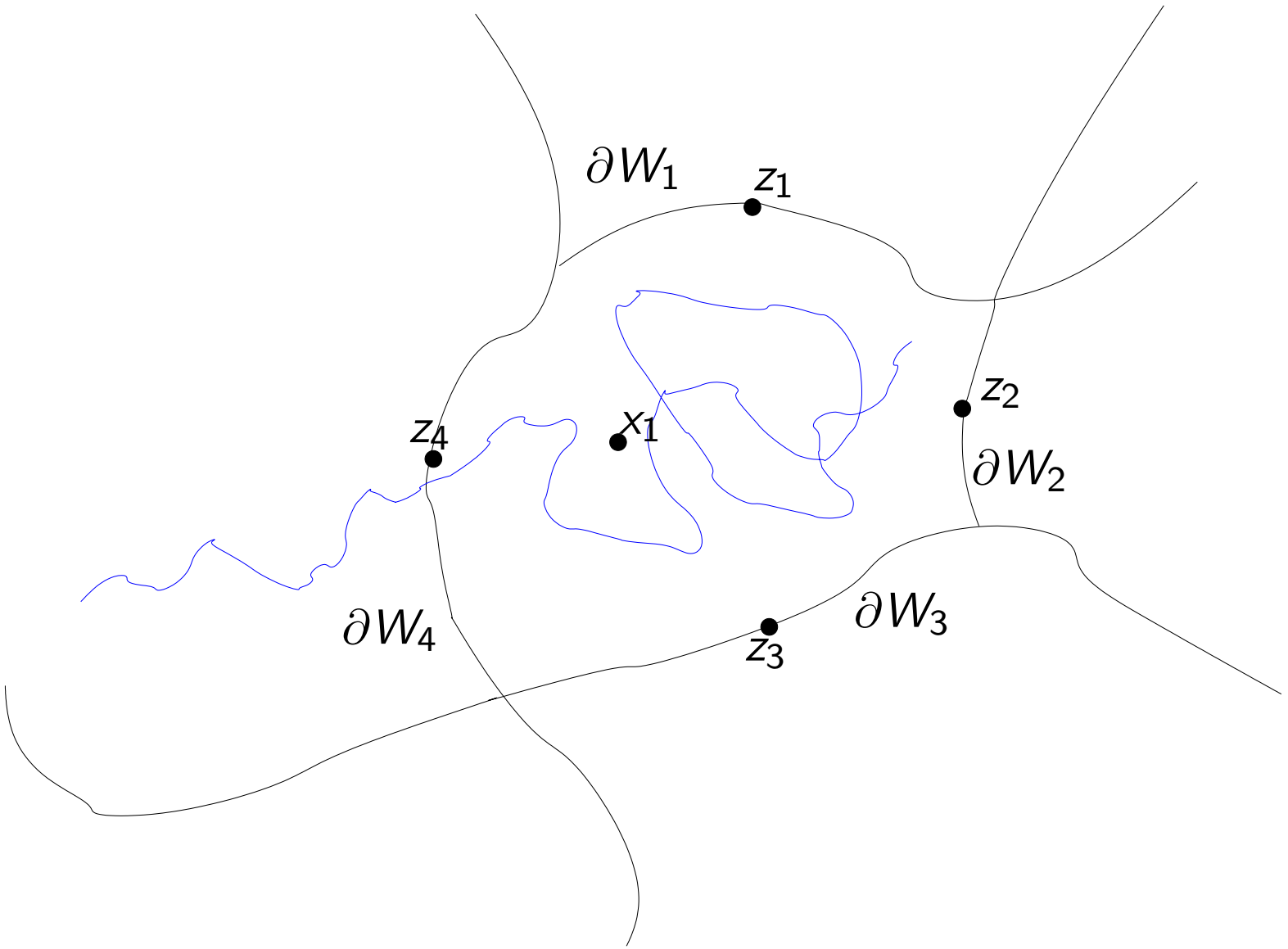


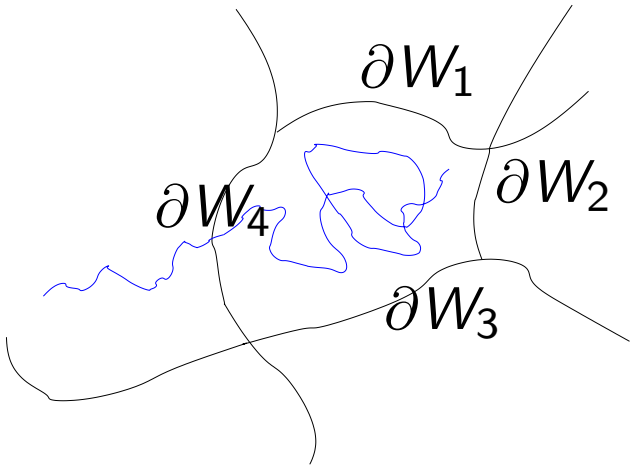
Figure: $\mathrm{LJ}_7^{\mathrm{2D}}$: Cumulative distribution function of the escape time from C_0.

kinetic Monte Carlo and Harmonic Transition State Theory



kMC models

Let us go back to the kinetic Monte Carlo model.



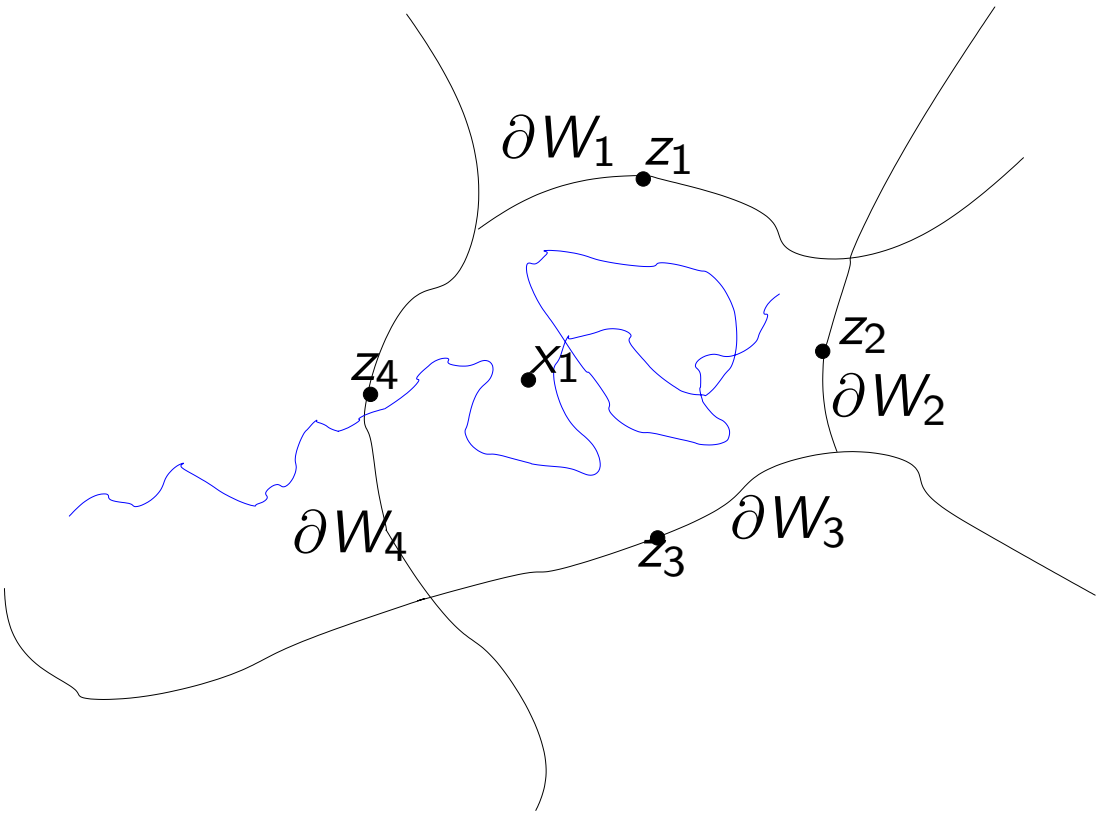
To each exit region ∂W_i is associated a rate $k(i)$. Let $\tau_i \sim \mathcal{E}(k(i))$ be independent exponential random variables. And then,

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

Thus, (i) exit time and exit region are independent r.v. ; (ii) exit time is $\mathcal{E}(k(1) + \ldots + k(I))$; (iii) exit region is i with prob. $\frac{k(i)}{k(1)+\ldots+k(I)}$.

The Eyring Kramers law and HTST

In practice, kMC models are parameterized using HTST.



We assume in the following $V(z_1) < V(z_2) < \ldots < V(z_I)$.

Eyring Kramers law (HTST): $k(i) = A_i \exp\left(-\beta(V(z_i) - V(x_1))\right)$
where A_i is a prefactor depending on V at z_i and x_1.

kMC and HTST

Thus, one obtains the following law for the exit event:

- exit time and exit region are independent r.v.
- exit time is $\mathcal{E}(k(1)+\ldots+k(I))$ and, when β is large

$$k(1)+\ldots+k(I) \simeq k(1) = A_1 \exp\left(-\beta(V(z_1)-V(x_1))\right)$$

- exit region is i with probability $\frac{k(i)}{k(1)+\ldots+k(I)}$ and, when β is large,

$$\frac{k(i)}{k(1)+\ldots+k(I)} \simeq \frac{k(i)}{k(1)} = \frac{A_i}{A_1} \exp\left(-\beta(V(z_i)-V(z_1))\right)$$

Our aim: justify and analyze this method.

Back to overdamped Langevin and the QSD

Starting from the QSD $d\nu = u_1(x)dx$, we already know that

- the exit time T_W and the exit point X_{T_W} are independent r.v.
- the exit time is exponentially distributed with parameter λ_1
- the exit region is ∂W_i with probability

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

We thus need to prove that

$$\lambda_1 \simeq A_1 \exp\left(-\beta(V(z_1) - V(x_1))\right)$$

and

$$-\frac{\displaystyle\int_{\partial W_i} \partial_n u_1 \, d\sigma}{\beta\lambda_1 \displaystyle\int_W u_1(x)\, dx} \simeq \frac{A_i}{A_1} \exp\left(-\beta(V(z_i) - V(z_1))\right).$$

Small temperature regime

The question is thus: consider (λ_1, u_1) such that (first eigenvalue eigenfunction pair)

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

We assume wlg $u_1 > 0$ and $\int u_1^2 e^{\beta V} = 1$.
In the small temperature regime ($\beta \to \infty$), prove that

$$\lambda_1 \simeq A_1 \exp\left(-\beta(V(z_1) - V(x_1))\right)$$

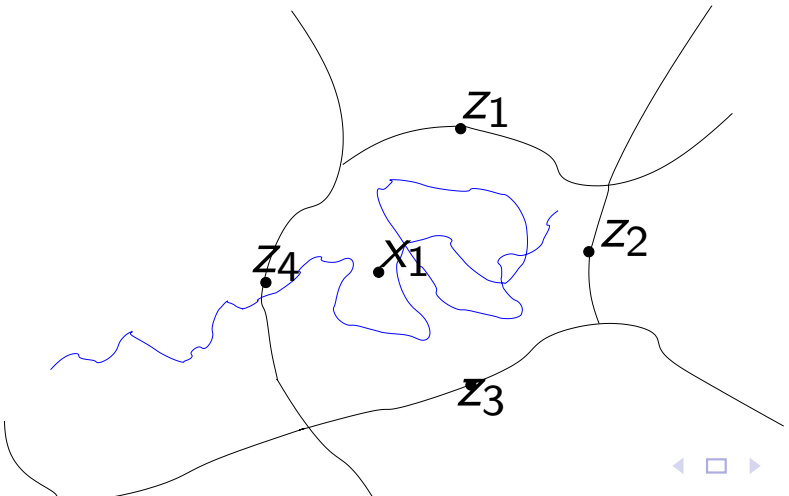
and

$$-\frac{\displaystyle\int_{\partial W_i} \partial_n u_1 \, d\sigma}{\beta\lambda_1 \displaystyle\int_W u_1(x)\, dx} \simeq \frac{A_i}{A_1} \exp\left(-\beta(V(z_i) - V(z_1))\right).$$

Assumptions

- W is an open bounded smooth domain in $\mathbb{R}^d$.
- $V : \overline{W} \to \mathbb{R}$ is a Morse function with a single critical point x_1. Moreover, $x_1 \in W$ and $V(x_1) = \min_{\overline{W}} V$.
- $\partial_n V > 0$ on ∂W and $V|_{\partial W}$ is a Morse function with local minima reached at $z_1, \ldots, z_I$ with $V(z_1) < \ldots < V(z_I)$.
- $V(z_1) - V(x_1) > V(z_I) - V(z_1)$
- $\forall i \in \{1, \ldots I\}$, consider B_{z_i} the basin of attraction for the dynamics $\dot{x} = -\nabla_T V(x)$ and assume that

$$\inf_{z \in B_{z_i}^c} d_a(z, z_i) > V(z_I) - V(z_1)$$



Agmon distance

Here, d_a is the Agmon distance:

$$d_a(x,y) = \inf_{\gamma} \int_0^1 g(\gamma(t))|\gamma'(t)|\,dt$$

where $g = \begin{cases} |\nabla V| \text{ in } W \\ |\nabla_T V| \text{ in } \partial W \end{cases}$, and the infimum is over all Lipschitz paths $\gamma : [0,1] \to \overline{W}$ such that $\gamma(0) = x$ and $\gamma(1) = y$. A few properties:

- One has $\forall x, y \in \overline{W}$, $|V(x) - V(y)| \leq d_a(x,y) \leq C|x-y|$
- On a neighborhood $\mathcal{V}$ of a local minima z_i, the function $x \mapsto d_a(x, z_i)$ satisfies the eikonal equation: $|\nabla \Phi|^2 = |\nabla V|^2$ on $\mathcal{V}$ with boundary conditions $\Phi = V$ on $\mathcal{V} \cap \partial W$, and $\Phi \geq V(z_i)$.

Results

[In preparation with G. Di Gesu, D. Le Peutrec and B. Nectoux] In the limit $\beta \to \infty$, the exit rate is

$$\lambda_1 = \sqrt{\frac{\beta}{2\pi}} \partial_n V(z_1) \frac{\sqrt{\det(Hess V)(x_1)}}{\sqrt{\det(Hess V_{|\partial W})(z_1)}} \mathrm{e}^{-\beta(V(z_1)-V(x_1))}(1+O(\beta^{-1})).$$

Moreover, for all open set Σ_i containing z_i such that $\overline{\Sigma}_i \subset B_{z_i}$,

$$\frac{\int_{\Sigma_i} \partial_n u_1 \, d\sigma}{\int_W u_1} = -C_i(\beta) e^{-\beta(V(z_i)-V(x_1))}(1+O(\beta^{-1})),$$

where $C_i(\beta) = \frac{\beta^{3/2}}{\sqrt{2\pi}} \partial_n V(z_i) \frac{\sqrt{\det(Hess V)(x_1)}}{\sqrt{\det(Hess V|_{\partial W})(z_i)}}$. Therefore,

$$\mathbb{P}^\nu(X_{T_W} \in \Sigma_i) = \frac{\partial_n V(z_i) \sqrt{\det \mathrm{Hess}(V|_{\partial W})(z_1)}}{\partial_n V(z_1) \sqrt{\det \mathrm{Hess}(V|_{\partial W})(z_i)}} e^{-\beta(V(z_i)-V(z_1))}(1+O(\beta^{-1})).$$

Related results in the literature (1/3)

The result on λ_1 is well known and actually holds under weaker assumptions. See for example [Helffer Nier] [Le Peutrec].

Similar formulas are obtained concerning the problem on the whole domain to compute the cascade of timescales down to the global minimum.

- 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]

Warning: The exit rate is (1/2) times the transition rate !

Related results in the literature (2/3)

Another approach to study the exit problem from a domain: Large deviation techniques [Freidlin, Wentzell, Day, Vanden Eijnden, Weare, Touchette,...].

Compared to our approach, the assumptions in LD are much less stringent but LD only provides the exponential rates (not the prefactors) and LD does not provide error bounds. (Moreover the fact that the exit time is exponentially distributed and the independance property between exit time and exit point are only obtained when $\beta = \infty$.)

Typical result [Freidlin, Wentzell, Theorem 5.1]: for all $W' \subset\subset W$, for any $\gamma > 0$, for any $\delta > 0$, there exists $\delta_0 \in (0, \delta]$ and $\beta_0 > 0$ such that for all $\beta \geq \beta_0$, for all $x \in W'$ and for all $y \in \partial W$,

$$\exp(-\beta(V(y) - V(z_1) + \gamma)) \leq \mathbb{P}^x(X_{\tau_W} \in \mathcal{V}_{\delta_0}(y)) \leq \exp(-\beta(V(y) - V(z_1) - \gamma))$$

Related results in the literature (3/3)

Why do we care about prefactors ?

Consider a situation with two local minima on the boundary ($V(z_1) < V(z_2)$). Compare

- the probability to leave through Σ_2 such that $z_2 \in \Sigma_2$, $\overline{\Sigma_2} \subset B_{z_2}$ and
- the probability to leave through Σ such that $\overline{\Sigma} \subset B_{z_1}$ and $\inf_\Sigma V = V(z_2)$.

Then, in the limit $\beta \to \infty$,

$$\frac{\mathbb{P}^\nu(X_{T_W} \in \Sigma)}{\mathbb{P}^\nu(X_{T_W} \in \Sigma_2)} = O(\beta^{-1/2}).$$

Discussion on the assumptions (1/5)

The assumption $V(z_1) - V(x_1) > V(z_I) - V(z_1)$ is probably not needed. It is required by our technique of proof.

The assumption

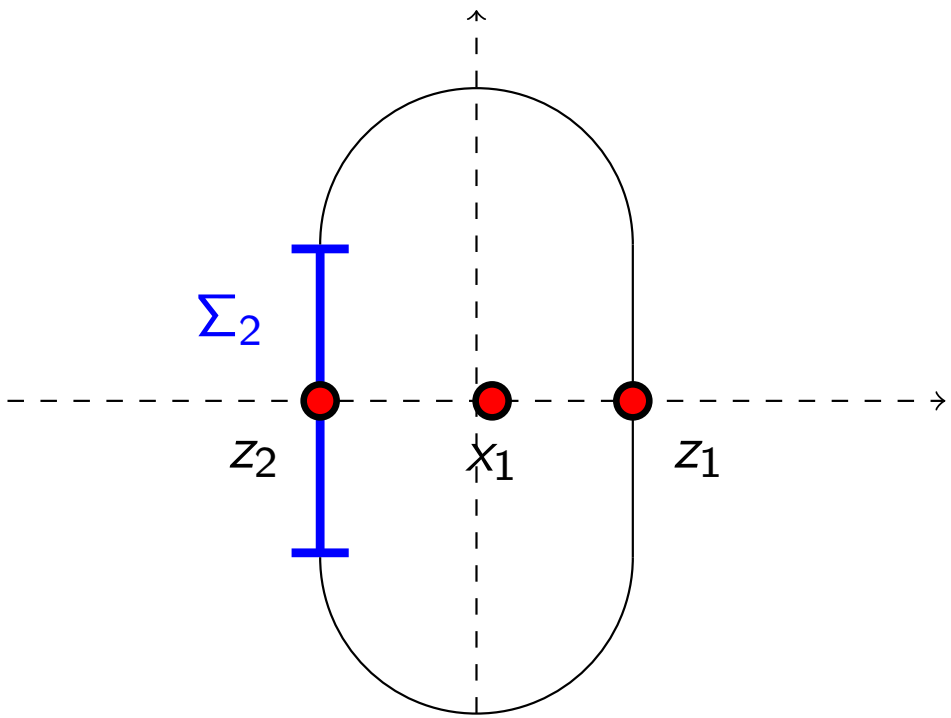
$$\forall i \in \{1, \dots I\}, \inf_{z \in B_{z_i}^c} d_a(z, z_i) > V(z_I) - V(z_1)$$

seems indeed important to get the expected results.

Discussion on the assumptions (2/5)

Let us consider the potential function $V(x, y) = x^2 + y^2 - ax$ with $a \in (0, 1/9)$ on the domain W. Two saddle points: $z_1 = (1, 0)$ and $z_2 = (-1, 0)$ (and $V(z_2) - V(z_1) = 2a$). One can check that the above assumptions are satisfied.

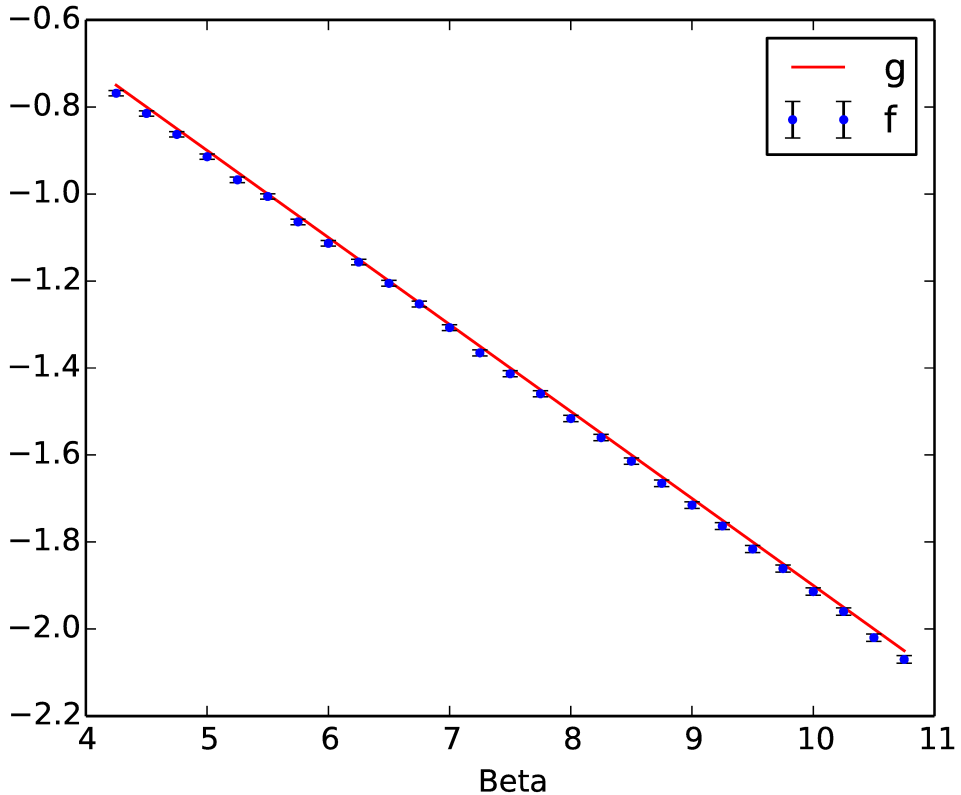
The domain W



Discussion on the assumptions (3/5)

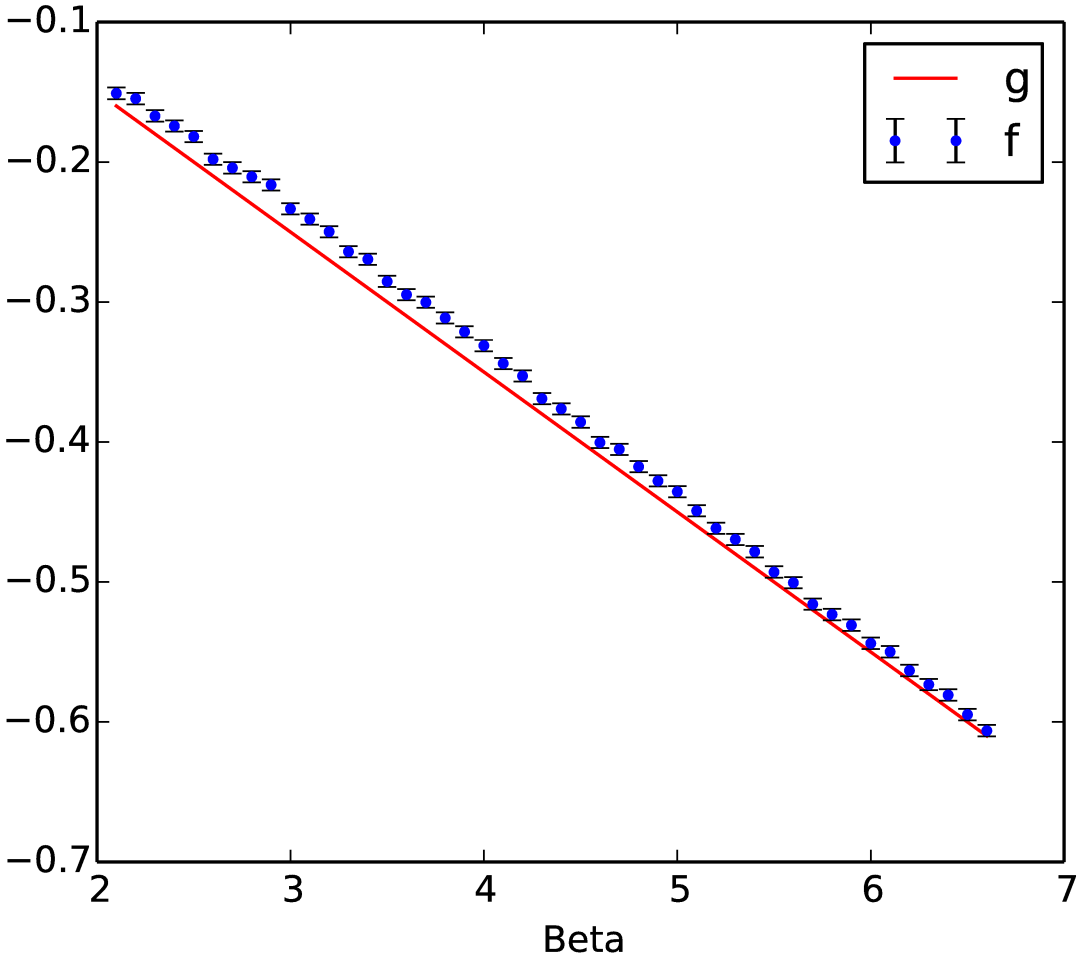
With $a = 1/10$, let us plot

- the numerical results $f : \beta \mapsto \ln\ \mathbb{P}^{\nu}(X_{T_W} \in \Sigma_2)$
- the theoretical result $g : \beta \mapsto \ln B_2 - \beta(V(z_2) - V(z_1))$, where $B_2 = \frac{\partial_n V(z_2)\sqrt{\det\ \mathrm{Hess}(V|_{\partial W})(z_1)}}{\partial_n V(z_1)\sqrt{\det\ \mathrm{Hess}(V|_{\partial W})(z_2)}}$ is the expected prefactor.



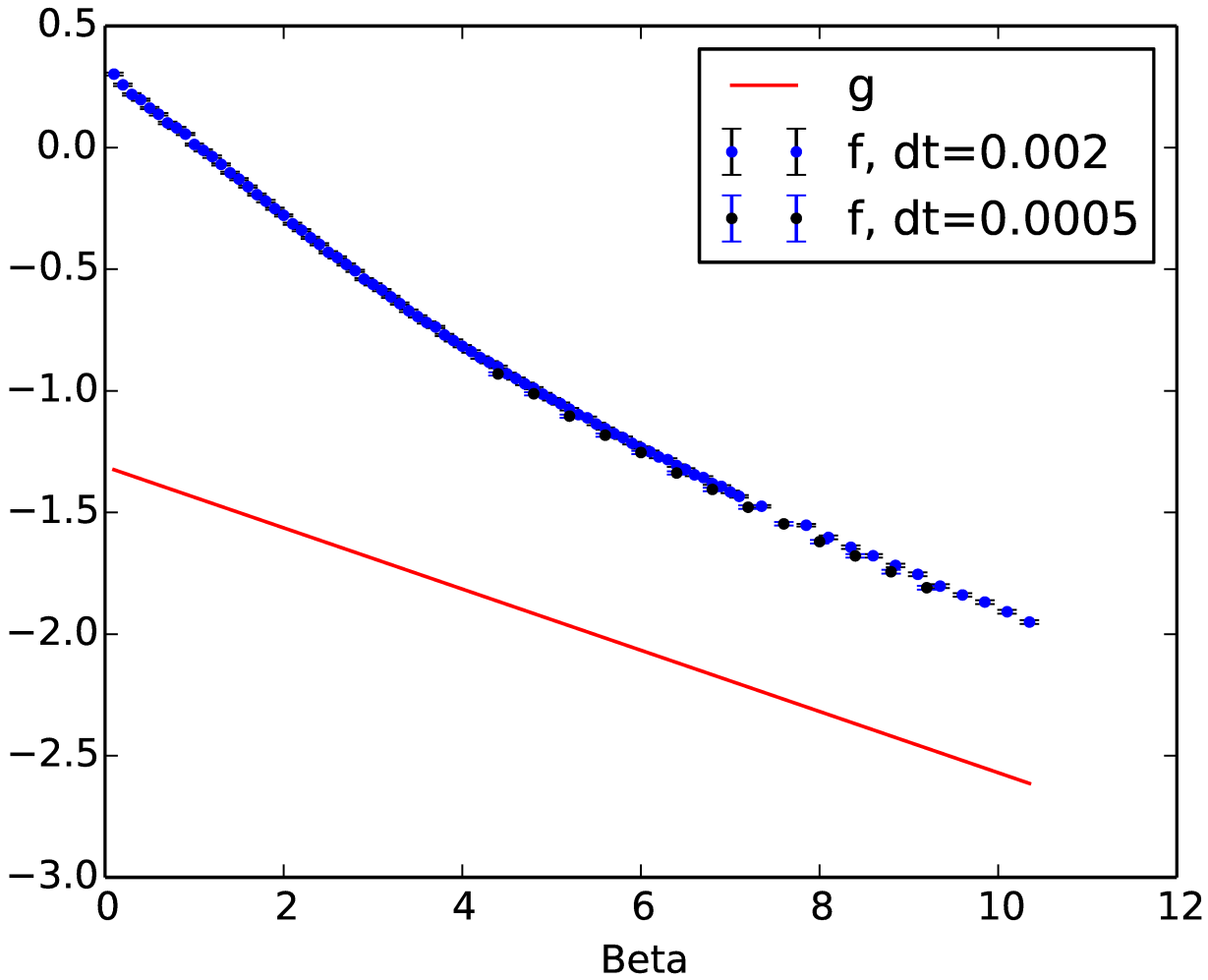
Discussion on the assumptions (4/5)

Same result with $a = 1/20$.



Discussion on the assumptions (5/5)

We now modify the potential such that the assumption on the Agmon distance is not satisfied anymore.



Sketch of the proof (1/4)

Let us consider $v_1 = u_1 \exp(\beta V)$, so that

$$\begin{cases} L^{(0)} v_1 = -\lambda_1 v_1 \text{ on } W, \\ v_1 = 0 \text{ on } \partial W, \end{cases}$$

where $L^{(0)} = \beta^{-1}\Delta - \nabla V \cdot \nabla$ is a self adjoint operator on $L^2(\exp(-\beta V))$. We are interested in $\nabla v_1 \cdot n$, and ∇v_1 satisfies

$$\begin{cases} L^{(1)} \nabla v_1 = -\lambda_1 \nabla v_1 \text{ on } W, \\ \nabla_T v_1 = 0 \text{ on } \partial W, \\ (\beta^{-1}\mathrm{div} \ - \nabla V \cdot) \nabla v_1 = 0 \text{ on } \partial W, \end{cases}$$

where

$$L^{(1)} = \beta^{-1}\Delta - \nabla V \cdot \nabla - \mathrm{Hess}(V).$$

Therefore ∇v_1 is an eigenvector (eigen-1-form) of $-L^{(1)}$ associated with the small eigenvalue λ_1.

Sketch of the proof (2/4)

We build so-called quasi-modes which approximate the eigenvectors of $L^{(0)}$ and $L^{(1)}$ associated with small eigenvalues in the regime $\beta \to \infty$:

- An approximation of v_1 (notice that for β sufficiently large, $\dim(\mathrm{Ran} 1_{[0,\beta^{-1/2}]}(-L^{(0)})) = 1$):

$$\tilde{v} \propto 1_{W'}$$

  where $W' \subset\subset W$.

- An approximation of $\mathrm{Ran}\left[1_{[0,\beta^{-1/2}]}(-L^{(1)})\right]$:

$$\mathrm{Span}(\tilde{\psi}_1, \ldots, \tilde{\psi}_I).$$

The functions $\tilde{\psi}_i$ are built using auxiliary simpler eigenvalue problems and WKB approximations. The support of $\tilde{\psi}_i$ is essentially in a neighborhood of z_i. (Agmon estimates are used to prove exponential decay away z_i.)

Sketch of the proof (3/4)

The last step consists in projecting the approximation of ∇v_1 on the approximation of $\mathrm{Ran}\left[1_{[0,\beta^{-1/2}]}(-L^{(1)})\right]$.

One can check that if $\tilde{v}$ and $(\tilde{\psi}_i)_{i=1\ldots I}$ are such that

- [Normalization] $\tilde{v} \in H^1_0(e^{-\beta V})$ and $\|\tilde{v}\|_{L^2(e^{-\beta V})} = 1$. $\forall i$, $\tilde{\psi}_i \in H^1_T(e^{-\beta V})$ and $\|\tilde{\psi}_i\|_{L^2(e^{-\beta V})} = 1$.
- [Good quasimodes]
  - $\forall \delta > 0$

$$\|\nabla \tilde{v}\|^2_{L^2(e^{-\beta V})} = O(e^{-\beta(V(z_1)-V(x_1)-\delta)}),$$

  - $\exists \varepsilon > 0$, $\forall i$,

$$\|1_{[\beta^{-1/2},\infty)}(-L^{(1)})\tilde{\psi}_i\|^2_{H^1(e^{-\beta V})} = O(e^{-\beta(V(z_I)-V(z_1)+\varepsilon)})$$

- [Orthonomality of quasimodes] $\exists \varepsilon_0 > 0$, $\forall i < j$

$$\langle \tilde{\psi}_i, \tilde{\psi}_j \rangle_{L^2(e^{-\beta V})} = O(\ e^{-\frac{\beta}{2}(V(z_j)-V(z_i)+\varepsilon_0)}\ ).$$

Sketch of the proof (4/4)

- [Decomposition of $\nabla\tilde{v}$] $\exists C_i, p, \forall i,$

$$\langle \nabla\tilde{v}, \tilde{\psi}_i \rangle_{L^2(e^{-\beta V})} = C_i \ \beta^{-p} e^{-\frac{\beta}{2}(V(z_i)-V(x_1))} \ (1 + O(\beta^{-1}) \ ).$$

- [Normal components of the quasimodes] $\exists B_i, m, \forall i, j$

$$\int_{\Sigma_i} \tilde{\psi}_j \cdot n \, e^{-\beta V} d\sigma = \begin{cases} B_i \ \beta^{-m} \ e^{-\frac{\beta}{2}V(z_i)} \ (\ 1 \ + O(\beta^{-1}) \ ) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Then for $i = 1, ..., n$, when $\beta \to \infty$

$$\int_{\Sigma_i} \partial_n v_1 \ e^{-\beta V} d\sigma = C_i B_i \ \beta^{-(p+m)} \ e^{-\frac{\beta}{2}(2V(z_i)-V(x_1))} \ (1 + O(\beta^{-1}))$$

The proof is based on a Gram-Schmidt procedure.

Conclusions

- There are two other accelerated dynamics methods: Hyperdynamics and Temperature Accelerated Dynamics. From ParRep to Hyper to TAD, the underlying assumptions for the algorithms to be correct are more and more stringent.
- 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.
- The QSD can be used to analyze the validity of kMC models and the Eyring-Kramers law, in the small temperature regime.

Simulating dynamics

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*: splitting techniques [Cérou, Guyader, TL, Weare], 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 [Wales]
- Starting from a long trajectory, extract states: clustering, Hidden Markov chain [Schuette]

References

Some papers I mentioned:

- D. Aristoff, TL and G. Simpson, *The parallel replica method for simulating long trajectories of Markov chains*, AMRX, 2, 2014.
- D. Aristoff and TL, *Mathematical analysis of Temperature Accelerated Dynamics*, SIAM MMS, 12(1), 2014.
- A. Binder, TL and G. Simpson, *A Generalized Parallel Replica Dynamics*, Journal of Computational Physics, 284, 2015.
- C. Le Bris, TL, M. Luskin and D. Perez, *A mathematical formalization of the parallel replica dynamics*, Monte Carlo Methods and Applications, 18(2), 2012.
- TL, *Accelerated dynamics: Mathematical foundations and algorithmic improvements*, `http://arxiv.org/abs/1501.01034` .
- TL and F. Nier, *Low temperature asymptotics for Quasi-Stationary Distributions in a bounded domain*, Analysis & PDE, 8(3), 2015.

References

Recent review paper: TL and Gabriel Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*. Acta Numerica, **25**, 681-880, (2016).

References

A book on the mathematics of stochastic MD: