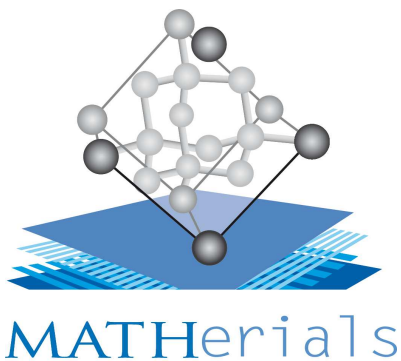
SDEs in large dimension and numerical methods
Part 2: Sampling metastable dynamics

T. Lelièvre

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



RICAM Winterschool, December 2016

Introduction

Remember the dynamics:

- *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 $\gamma > 0$ and $\beta = (k_B T)^{-1}$.

- *overdamped Langevin* (or gradient) dynamics:

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

Introduction

These dynamics are used to compute macroscopic quantities:

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

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

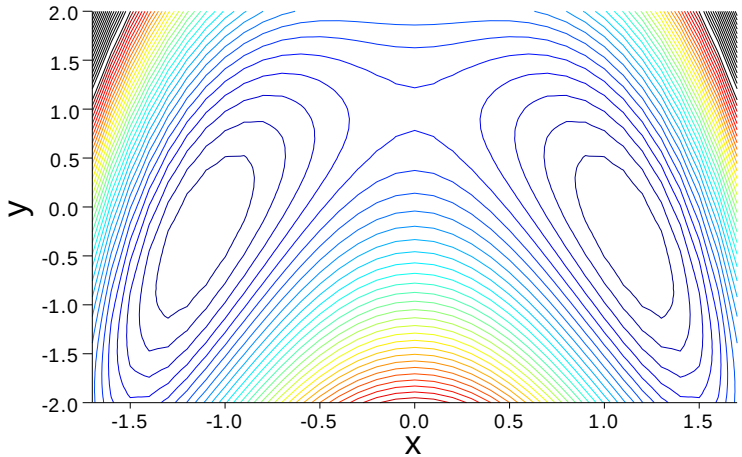
(ii) Dynamical quantities (averages over trajectories): diffusion coefficients, viscosity, transition rates,...

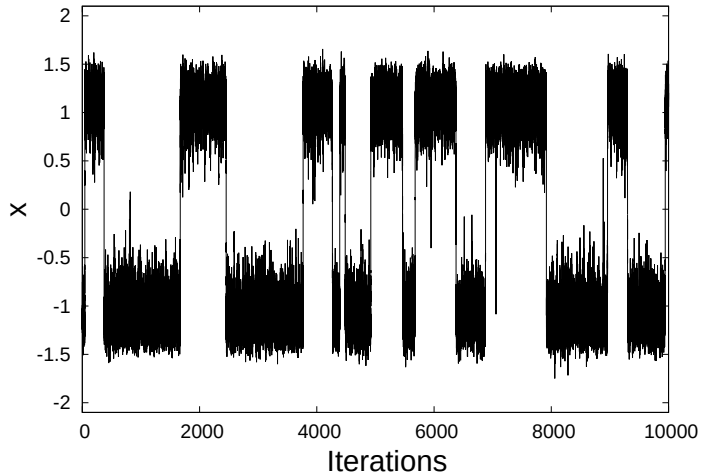
$$\mathbb{E}(\mathcal{F}((\boldsymbol{X}_t)_{t\geq 0})) \simeq \frac{1}{M} \sum_{m=1}^{M} \mathcal{F}((\boldsymbol{X}_t^m)_{t\geq 0}).$$

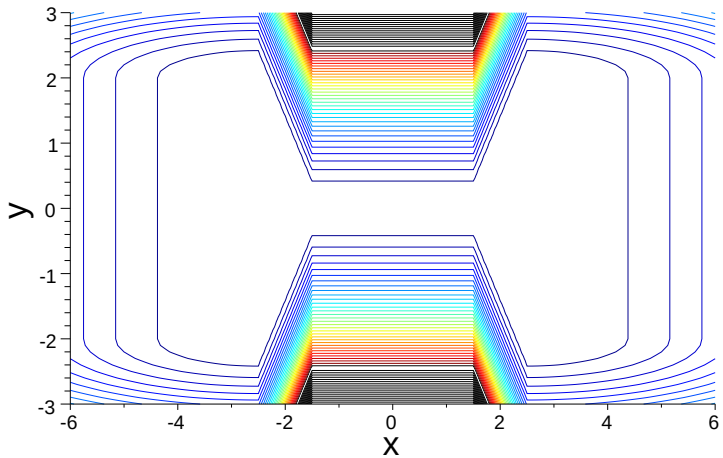
Difficulties: (i) high-dimensional problem ($N \gg 1$); (ii) $\boldsymbol{X}_t$ is a metastable process and μ is a multimodal measure.

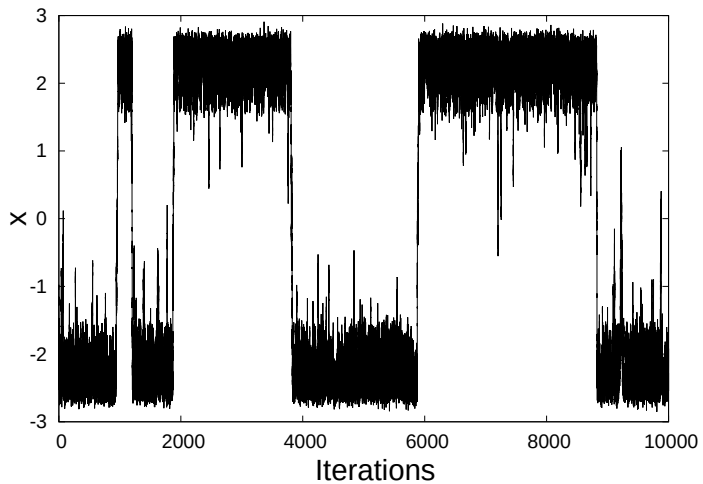
Metastability: energetic and entropic barriers

A two-dimensional schematic picture







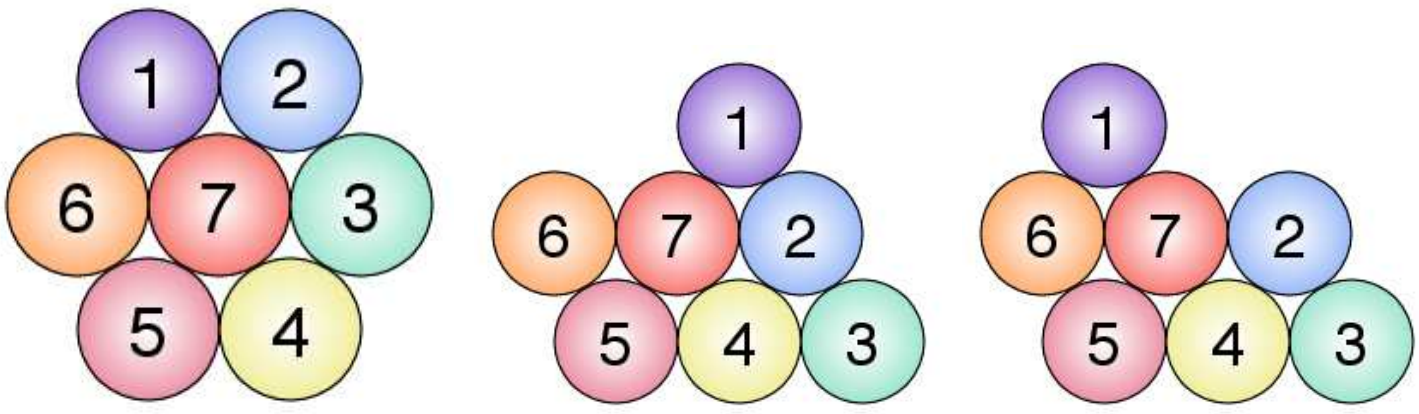


$\longrightarrow$ • Slow convergence of trajectorial averages

• 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

Introduction

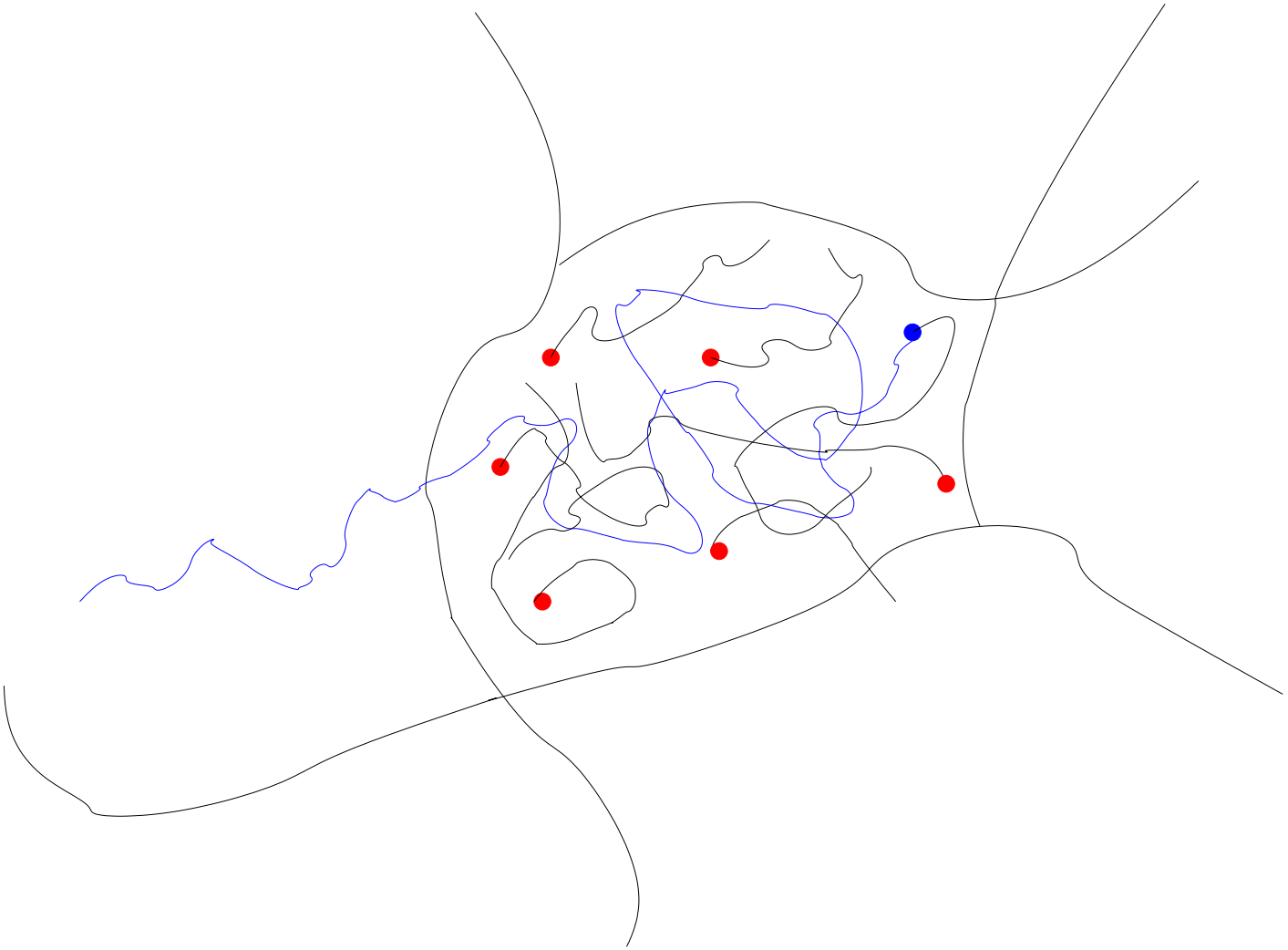
For computing thermodynamics quantities, there is a clear classification of available methods, and the difficulties are now well understood (in particular for free energy computations, see for example [TL, Rousset, Stoltz, 2010]). On the opposite, computing efficiently dynamical quantities remains a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

Outline of the talk:

1. Accelerated dynamics: These methods have been proposed by A.F. Voter to generate efficiently metastable dynamics. *Mathematical tool: Quasi Stationary Distributions.*
2. Adaptive Multilevel Splitting methods: Towards efficient sampling of reactive paths. *Rare event simulation.*

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.

Accelerated dynamics



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

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

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.

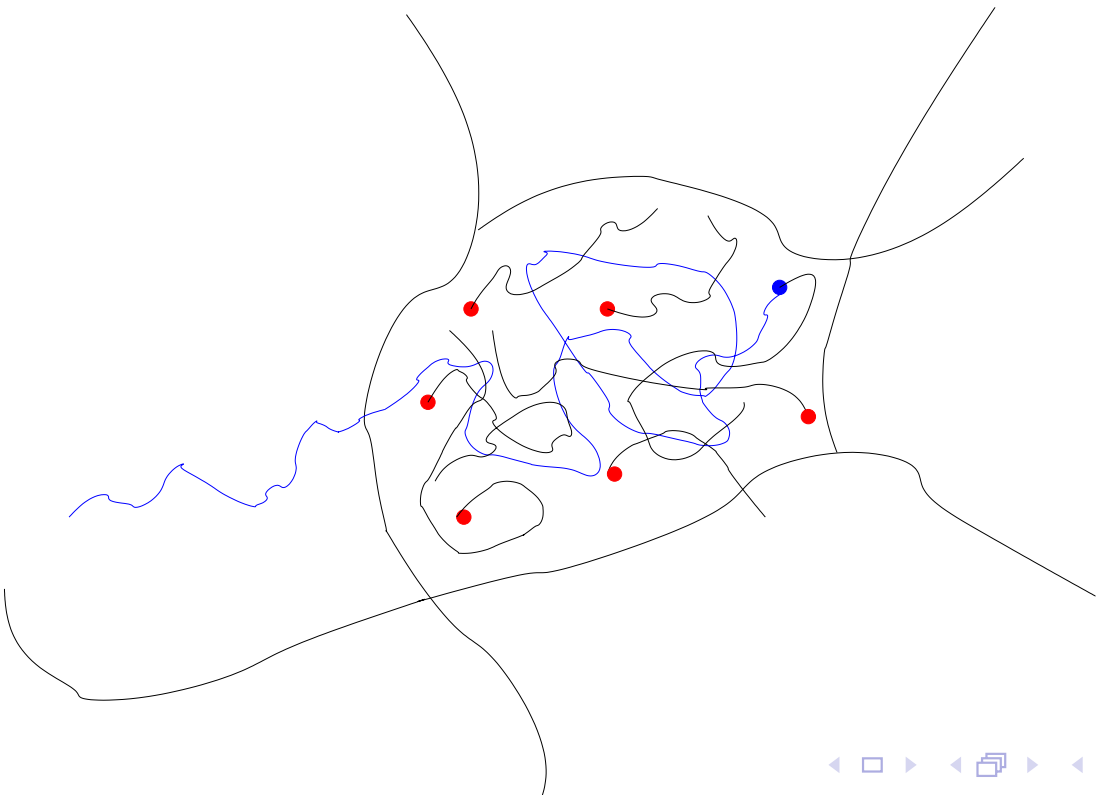
Remark: All the algorithms below equally apply to the Langevin dynamics but the extensions of the mathematical results to the Langevin dynamics are not straightforward...

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

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

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

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

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

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,

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

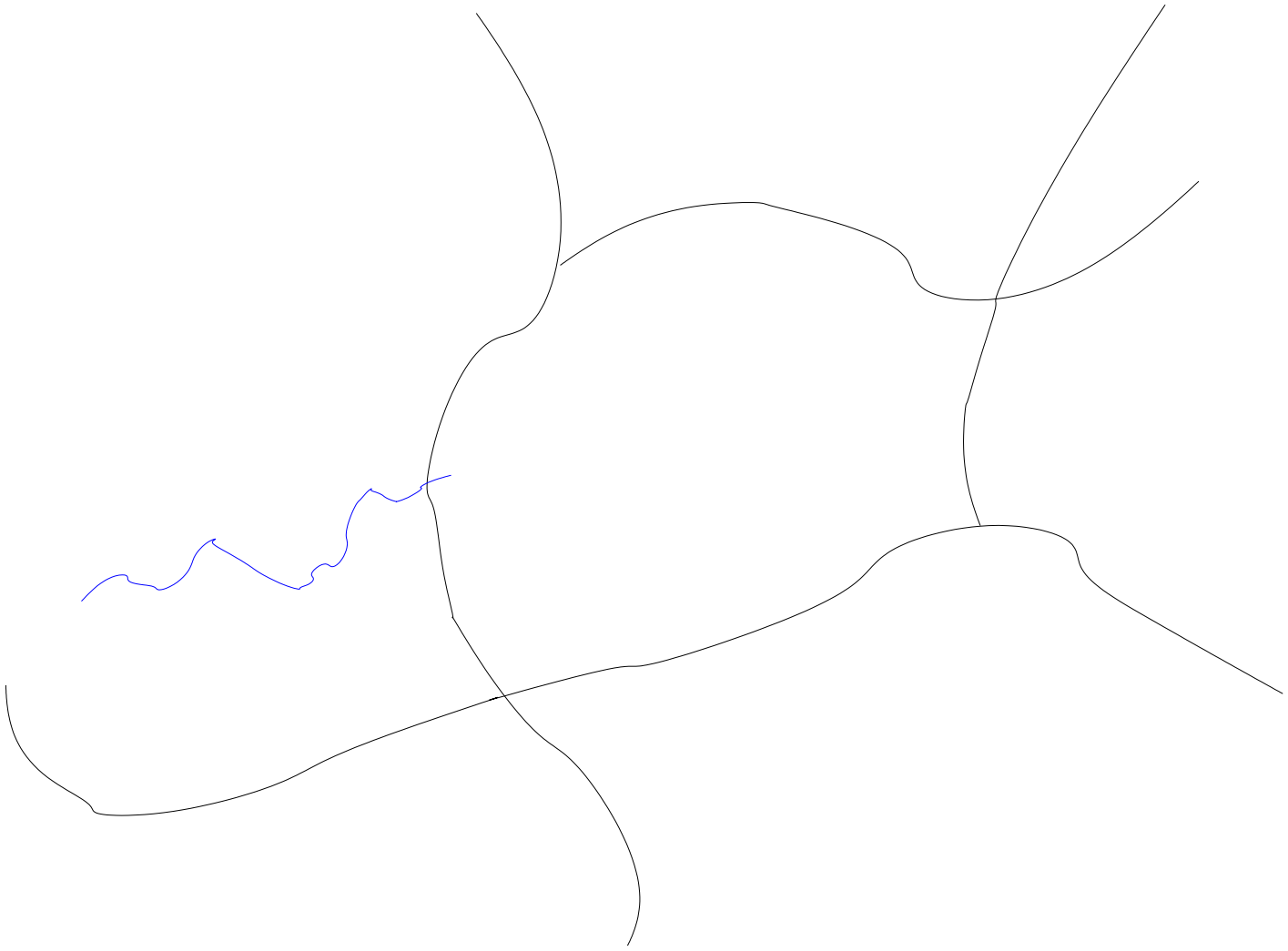
The Parallel Replica Algorithm

The full algorithm is in three steps:

- Decorrelation step
- Dephasing step
- Parallel step

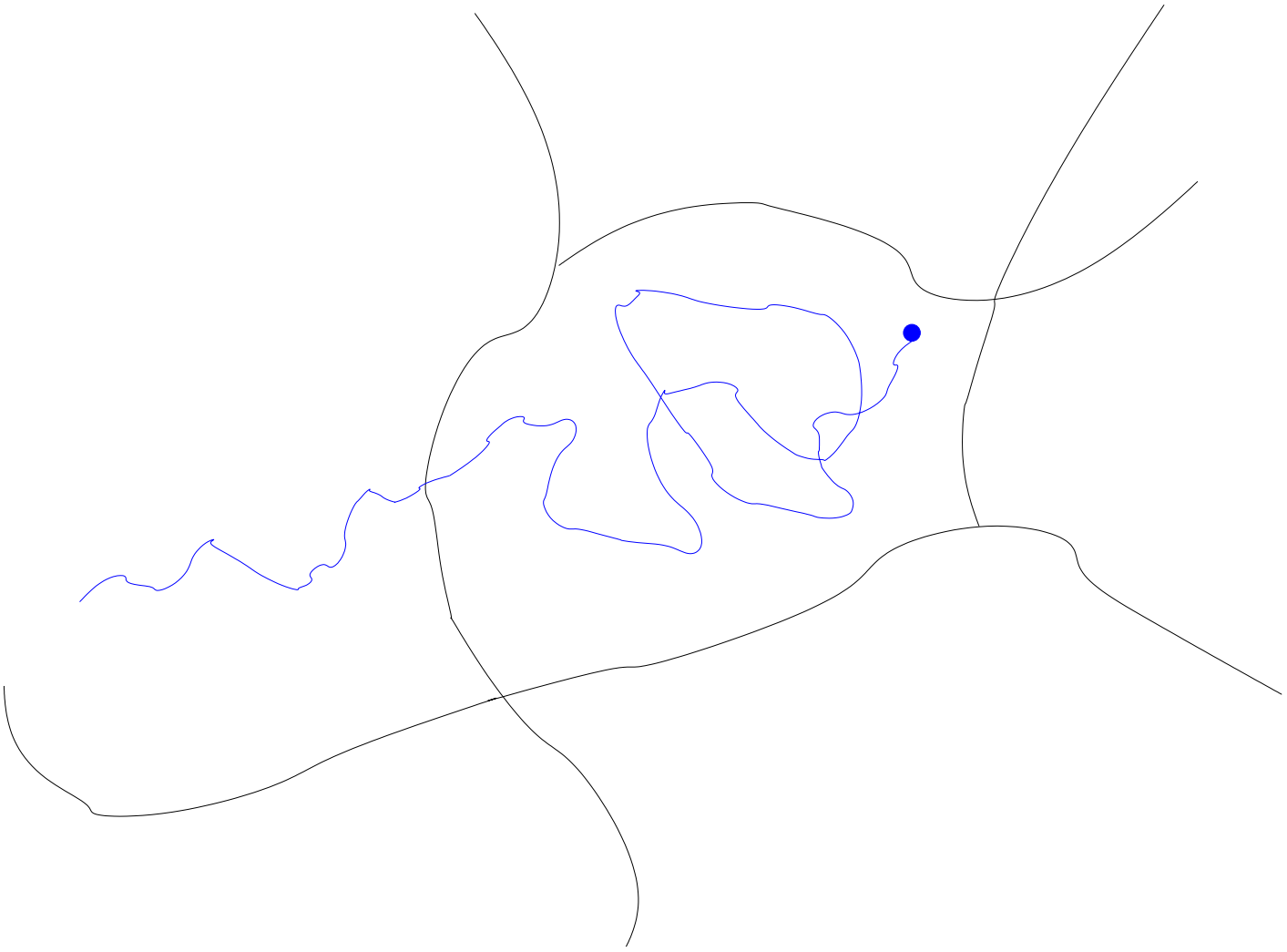
The Parallel Replica Algorithm

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



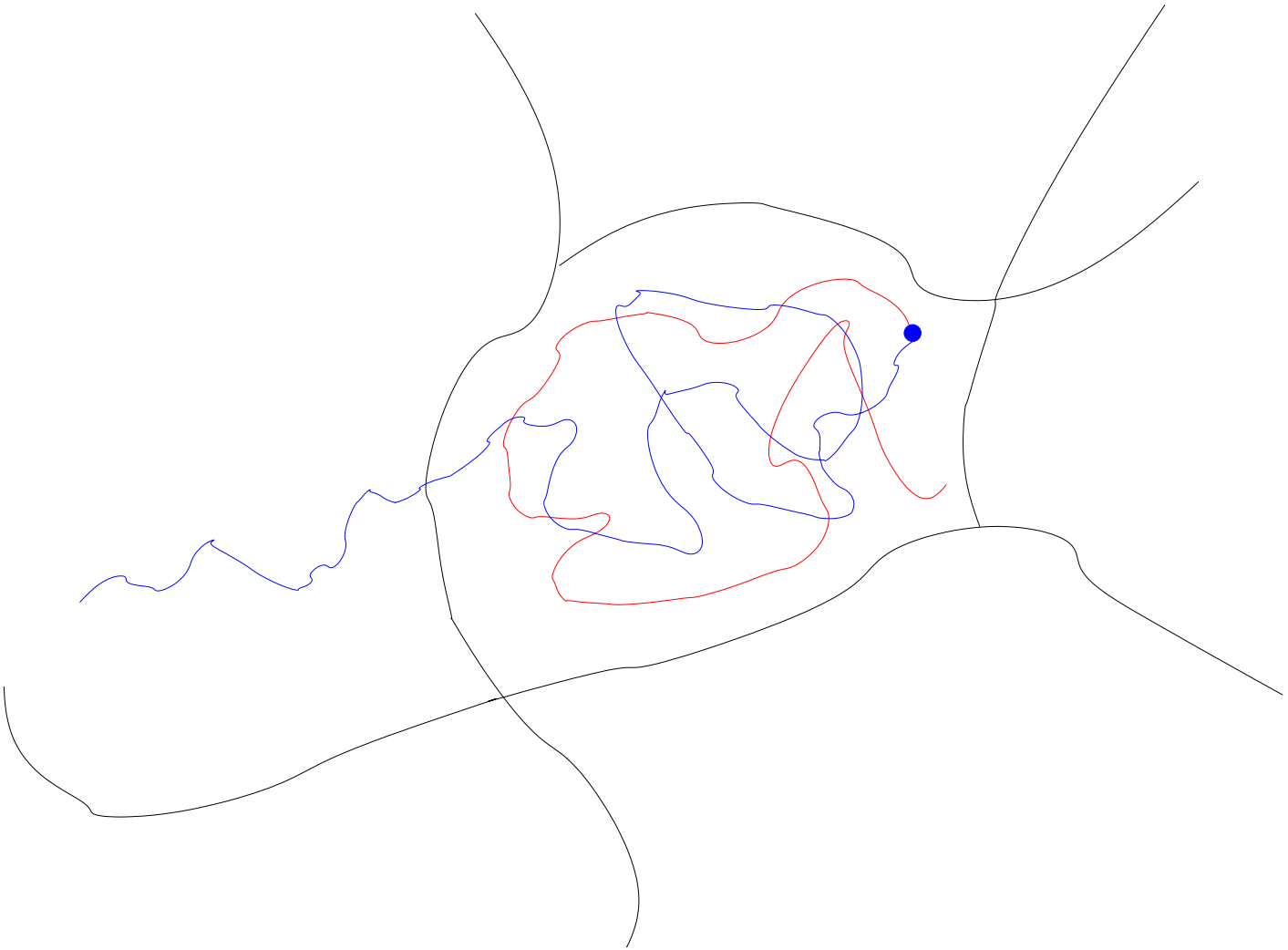
The Parallel Replica Algorithm

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



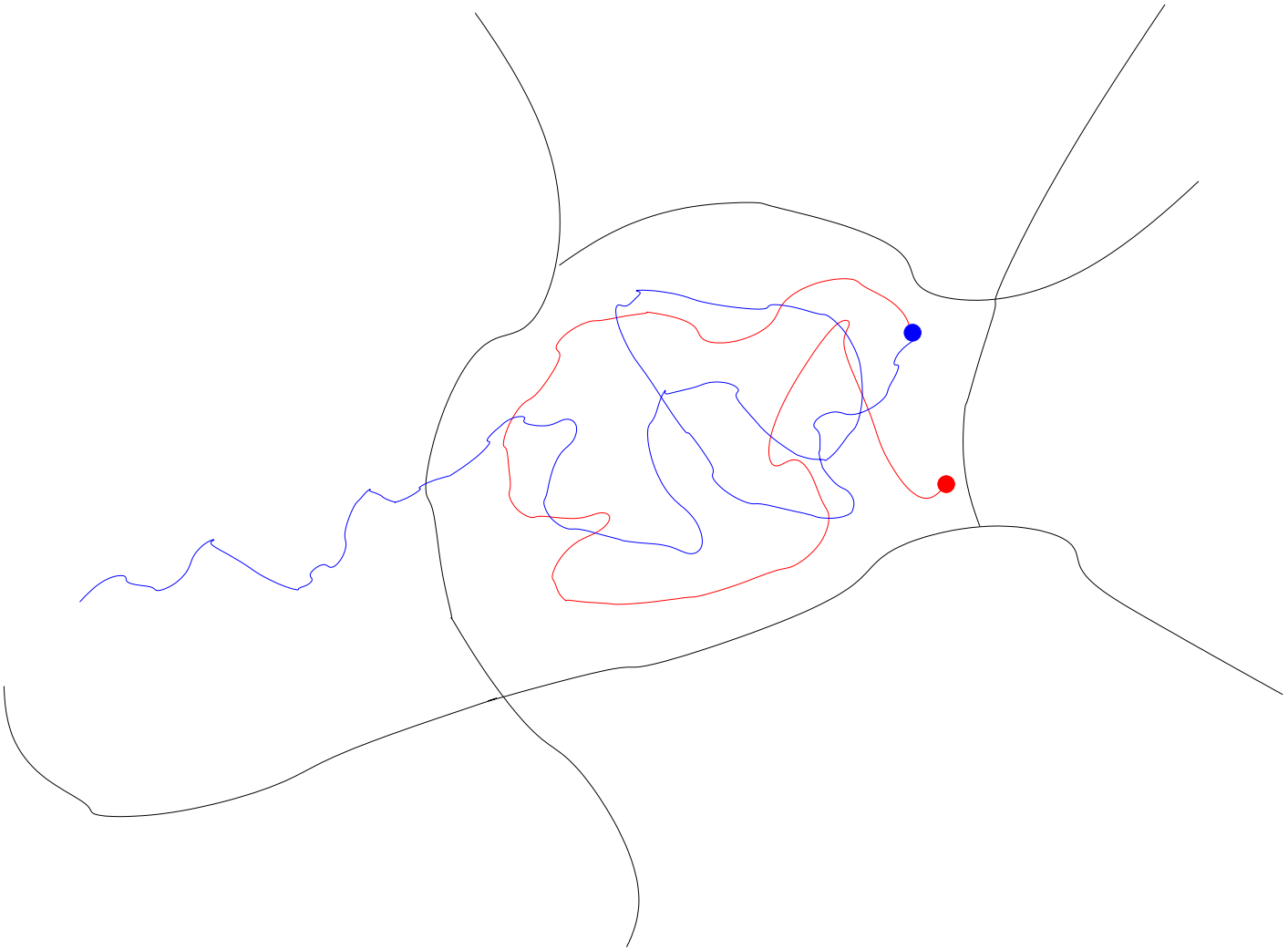
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



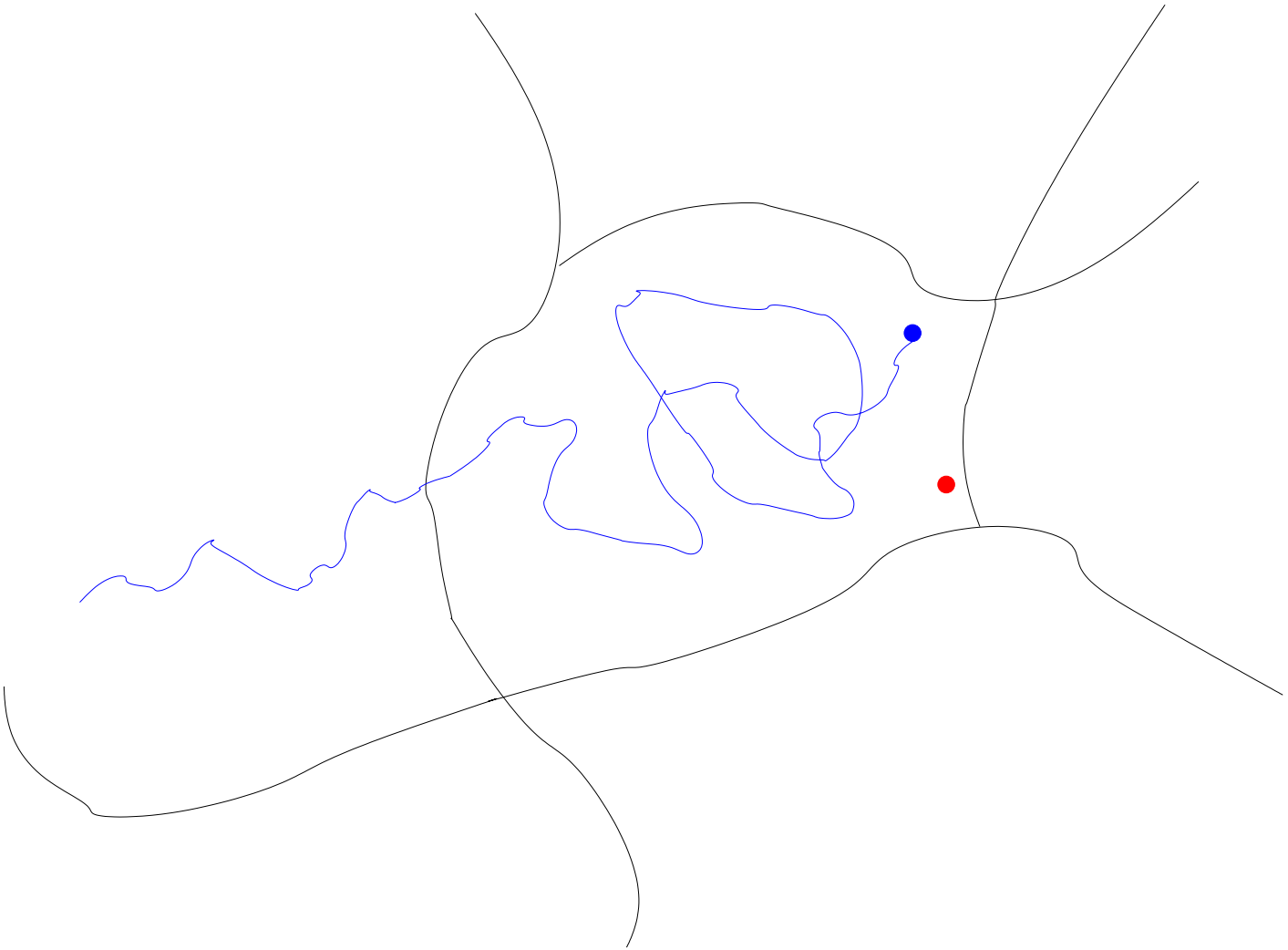
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



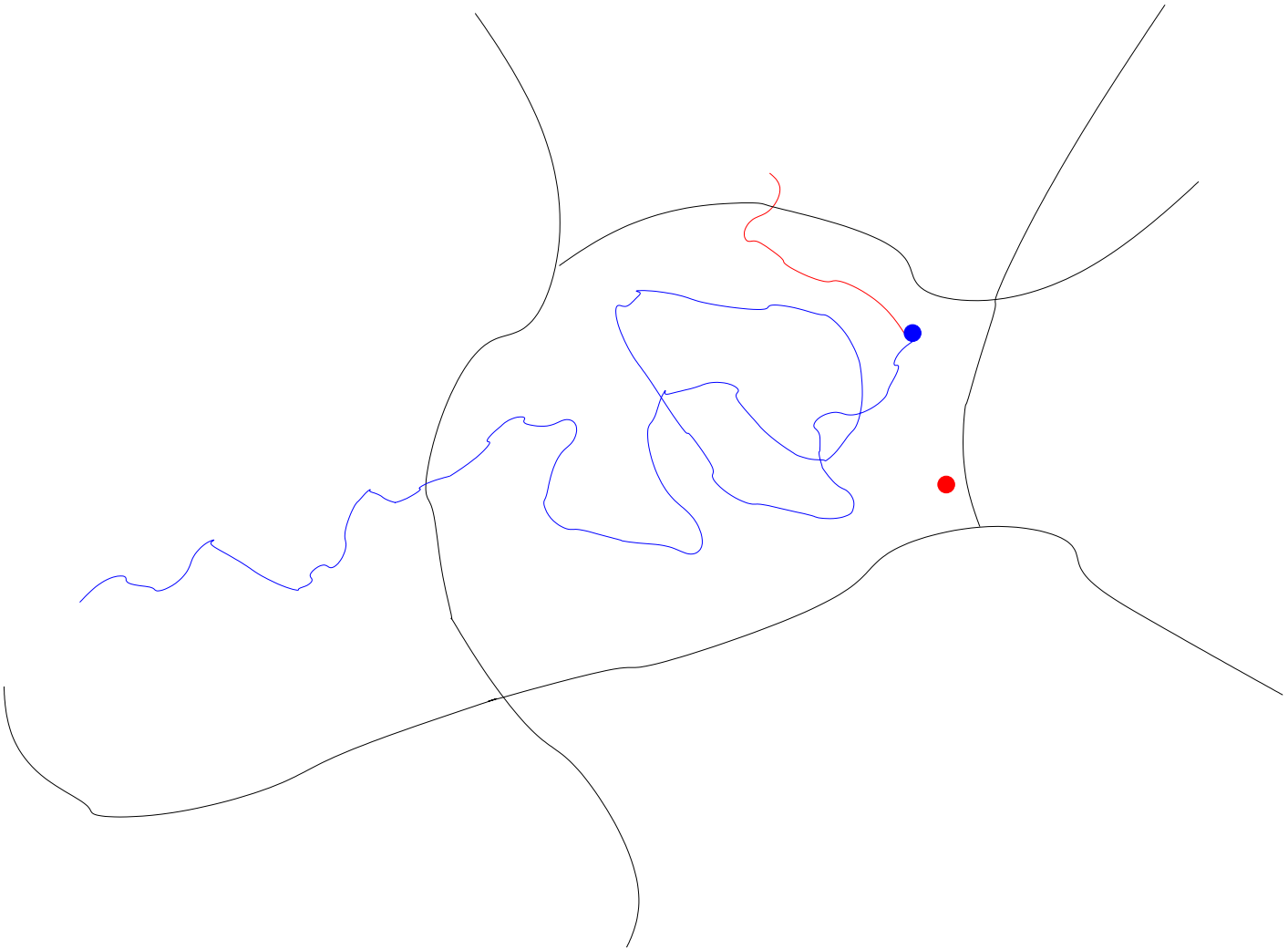
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



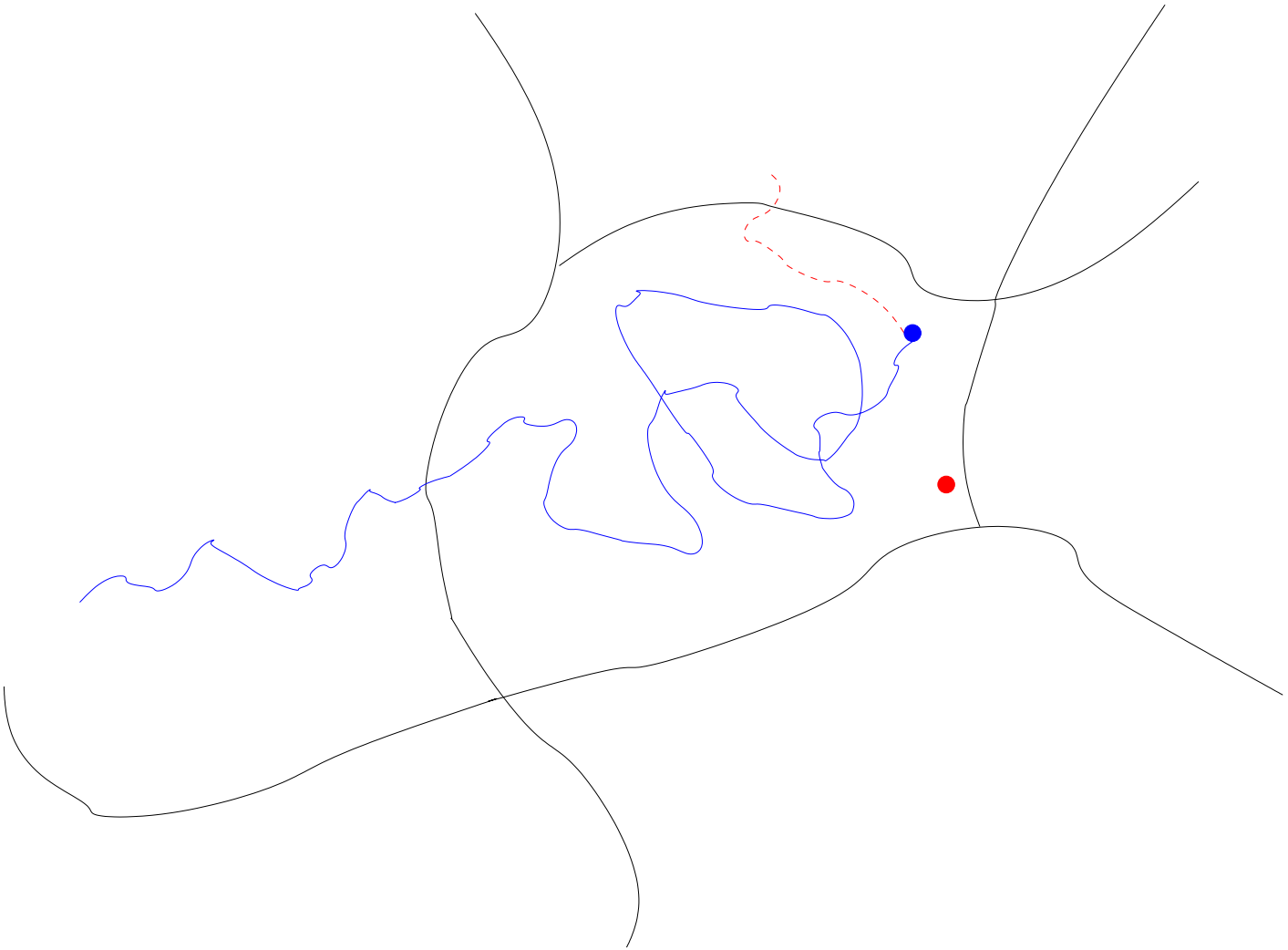
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



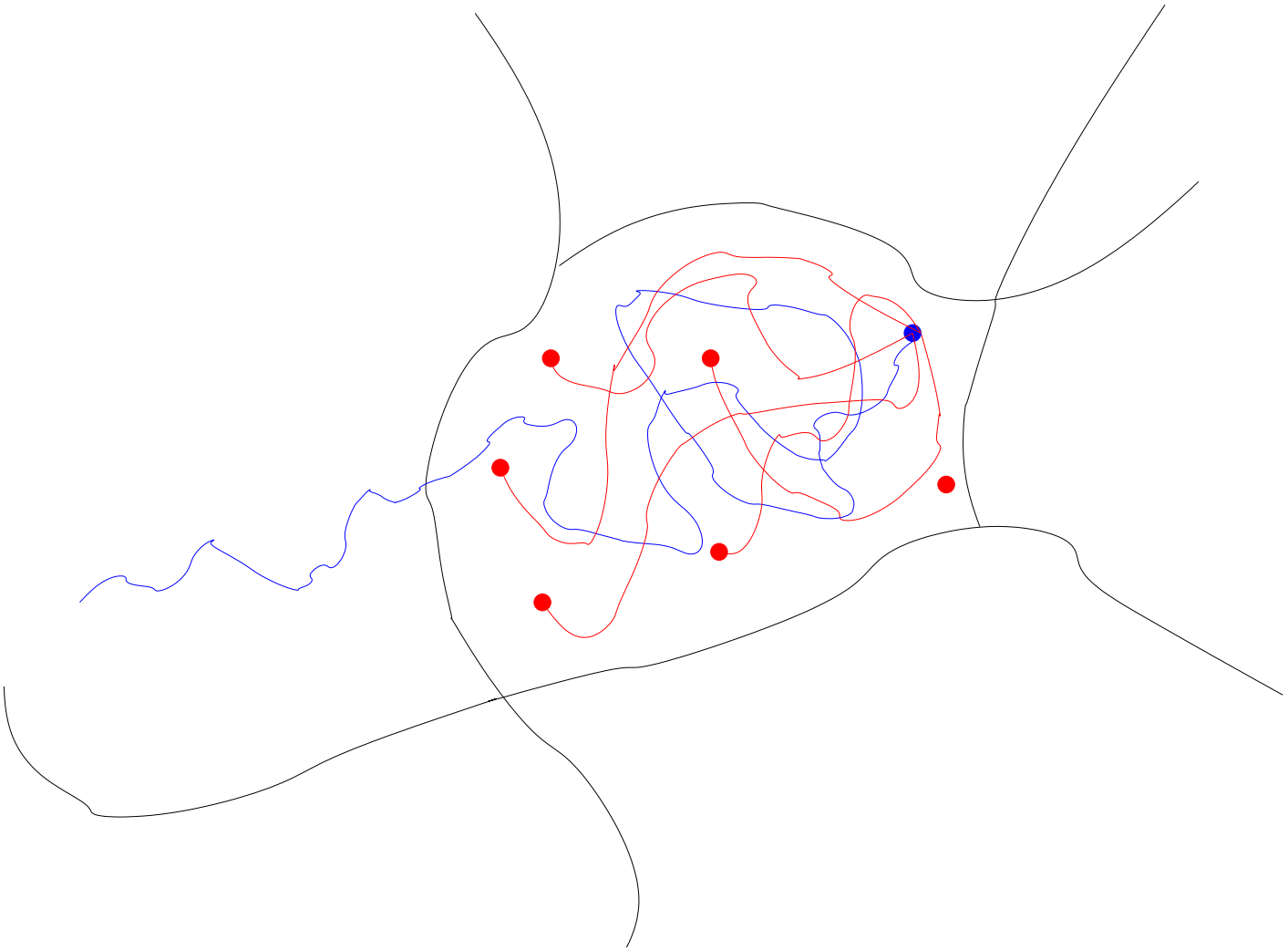
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



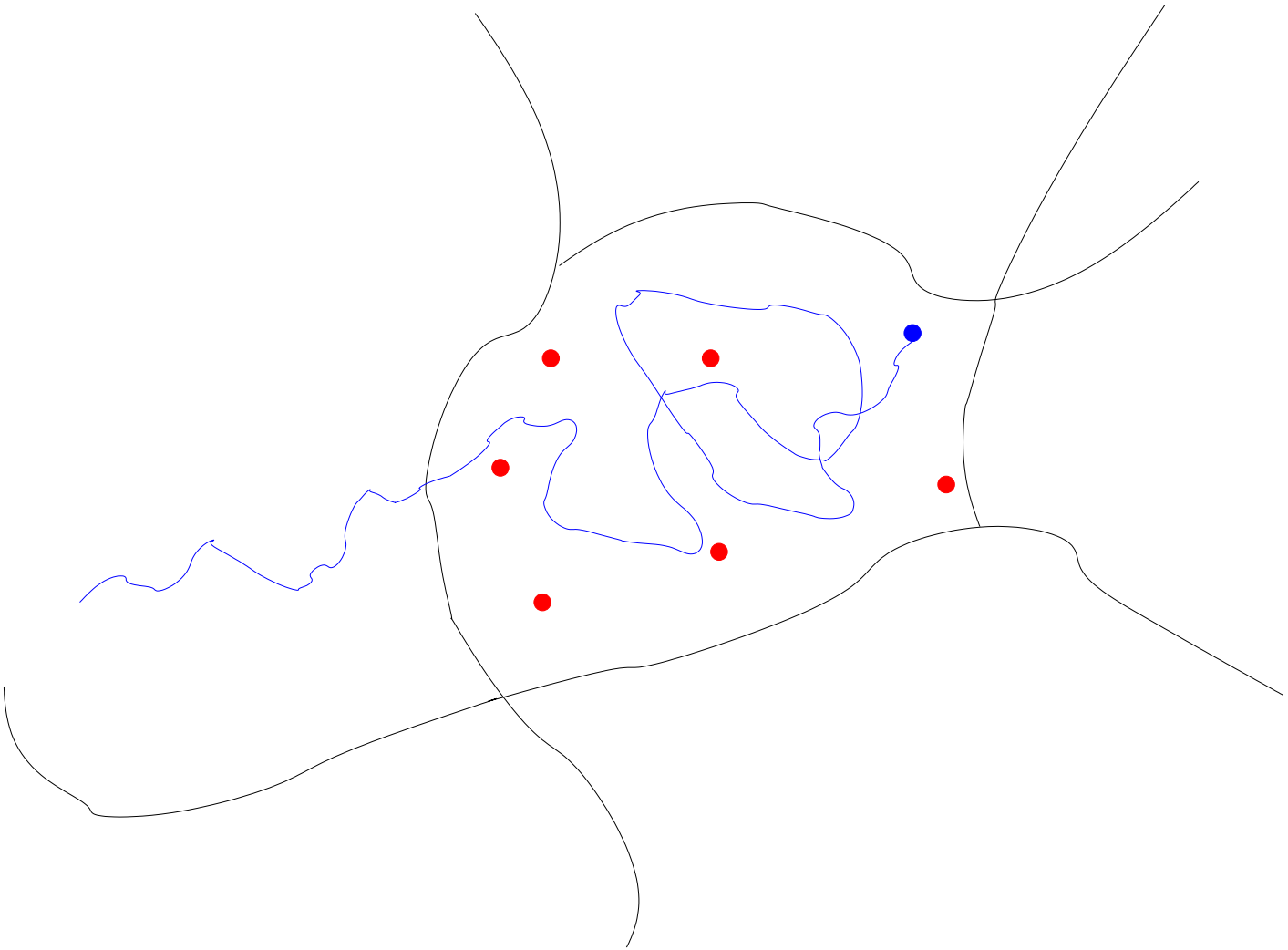
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



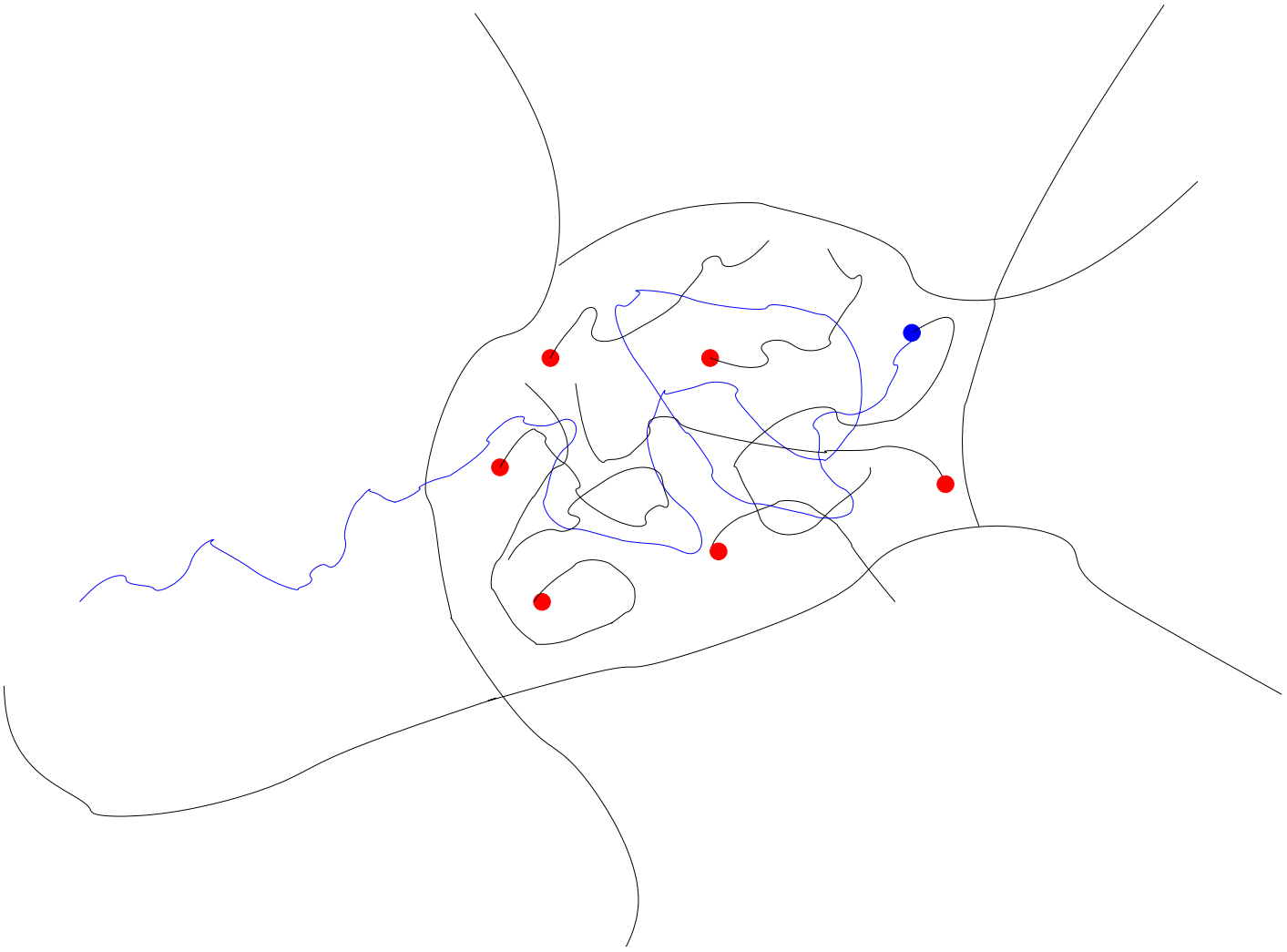
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



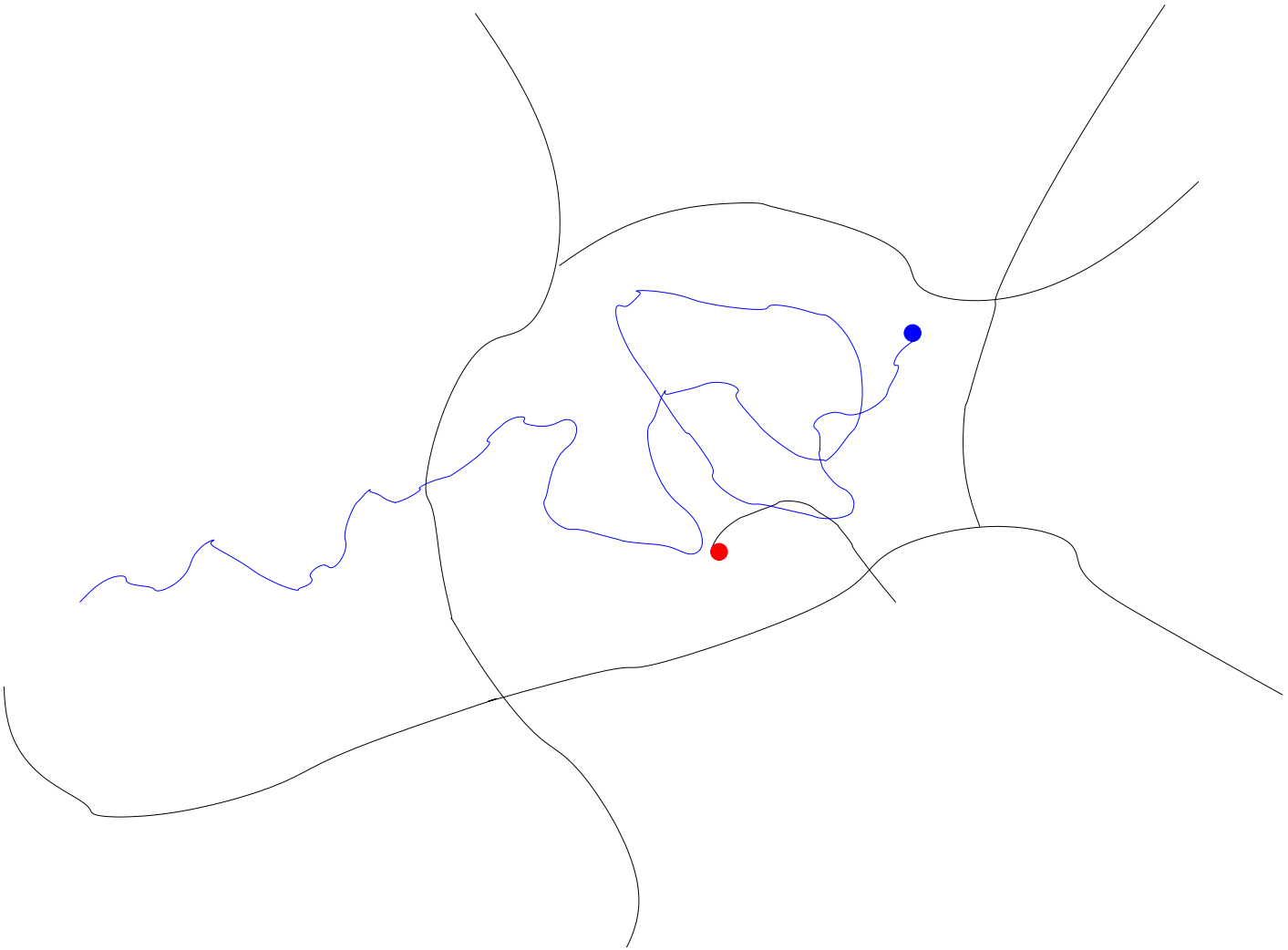
The Parallel Replica Algorithm

Parallel step: run independent trajectories in parallel...



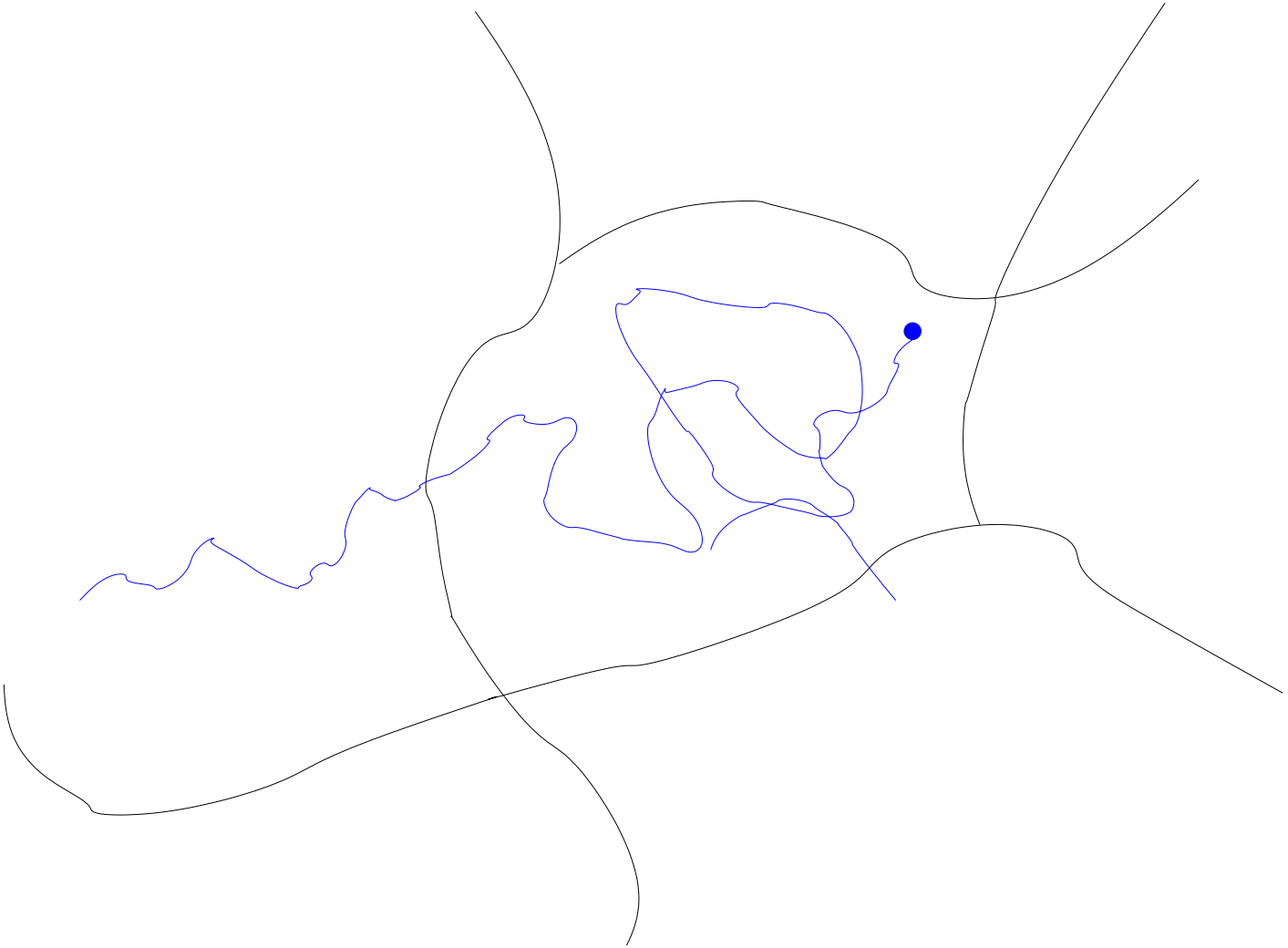
The Parallel Replica Algorithm

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



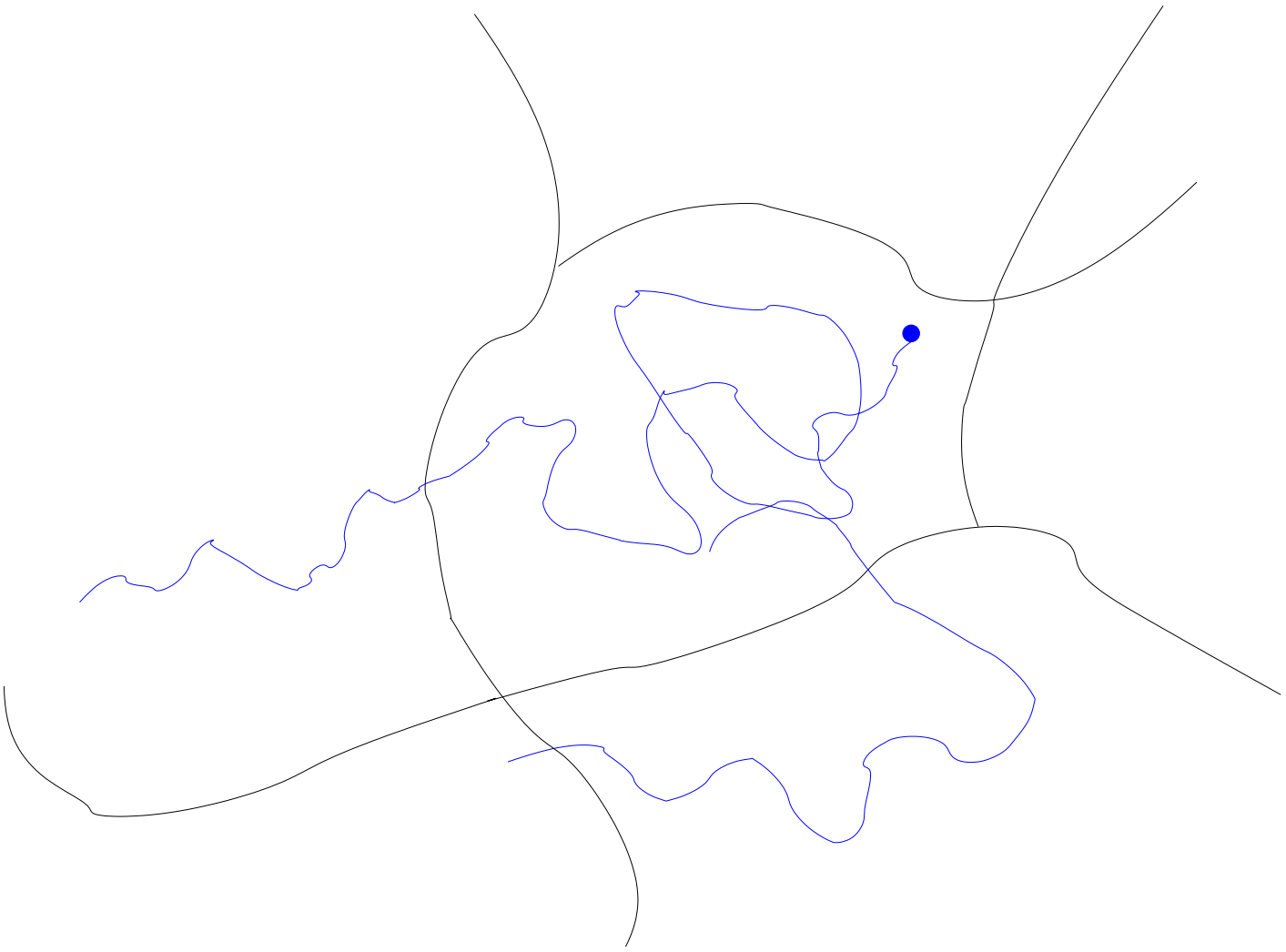
The Parallel Replica Algorithm

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



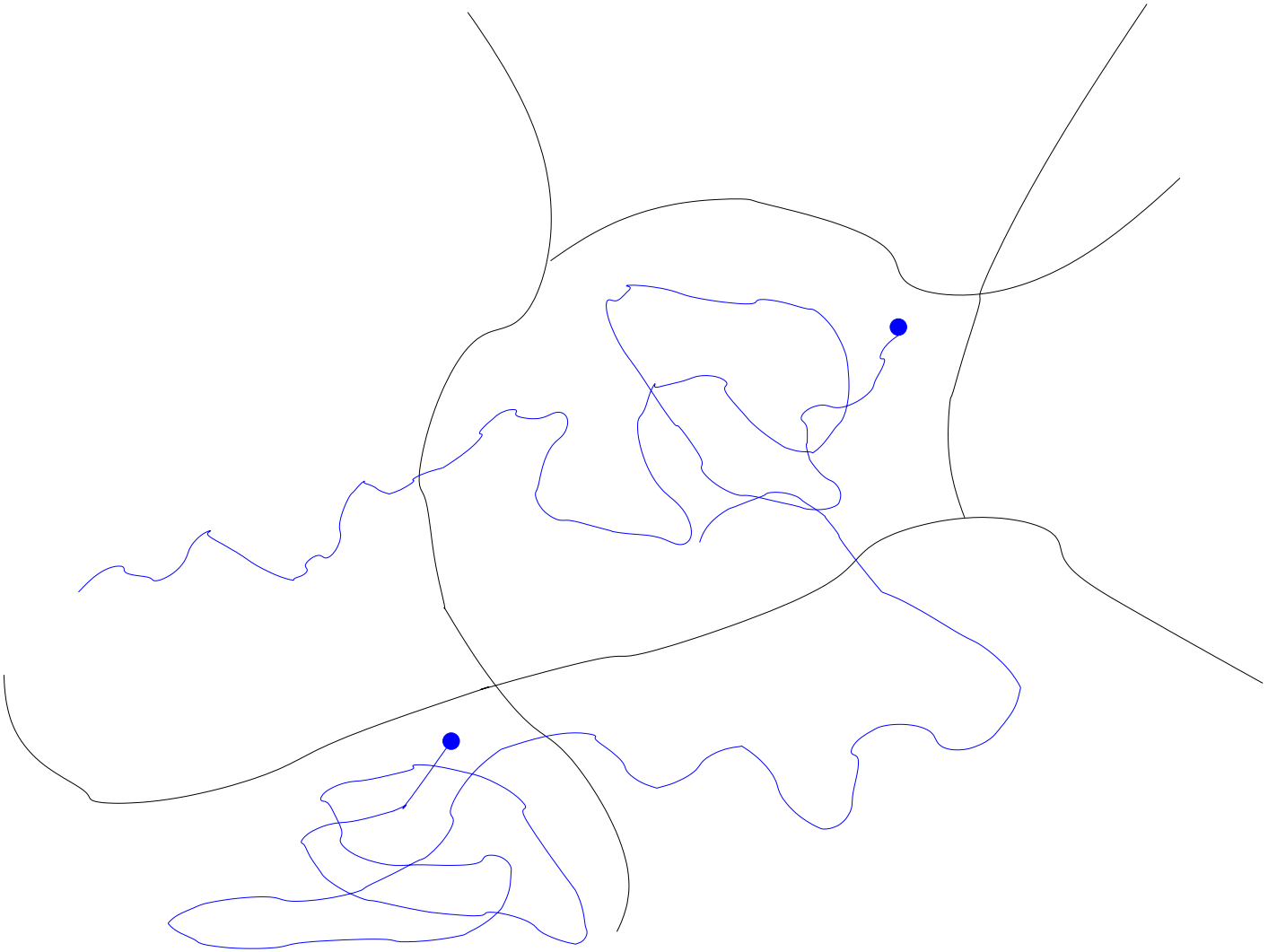
The Parallel Replica Algorithm

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 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:

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

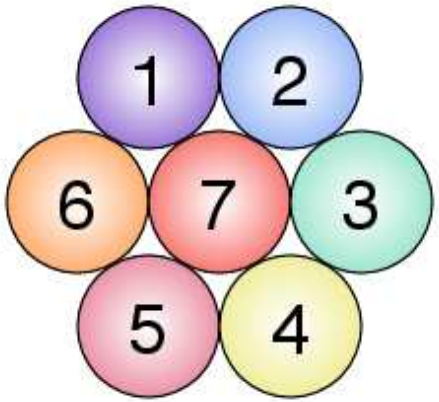
The generalized Parallel Replica algorithm

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.

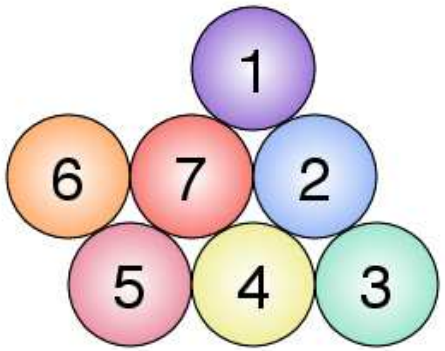
If the Fleming Viot particle process reaches stationarity before the reference walker, go to the parallel step. Otherwise, restart a new decorrelation / dephasing step.

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

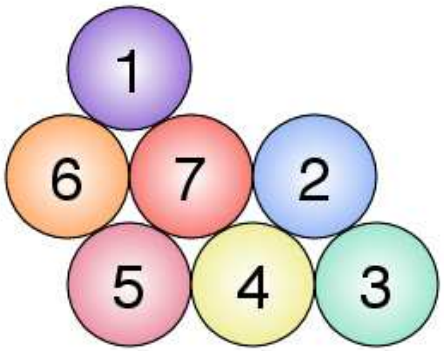
Numerical test case: the 7 atoms LJ cluster



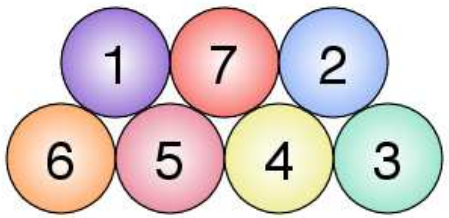
(a) C_0, $V = -12.53$



(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

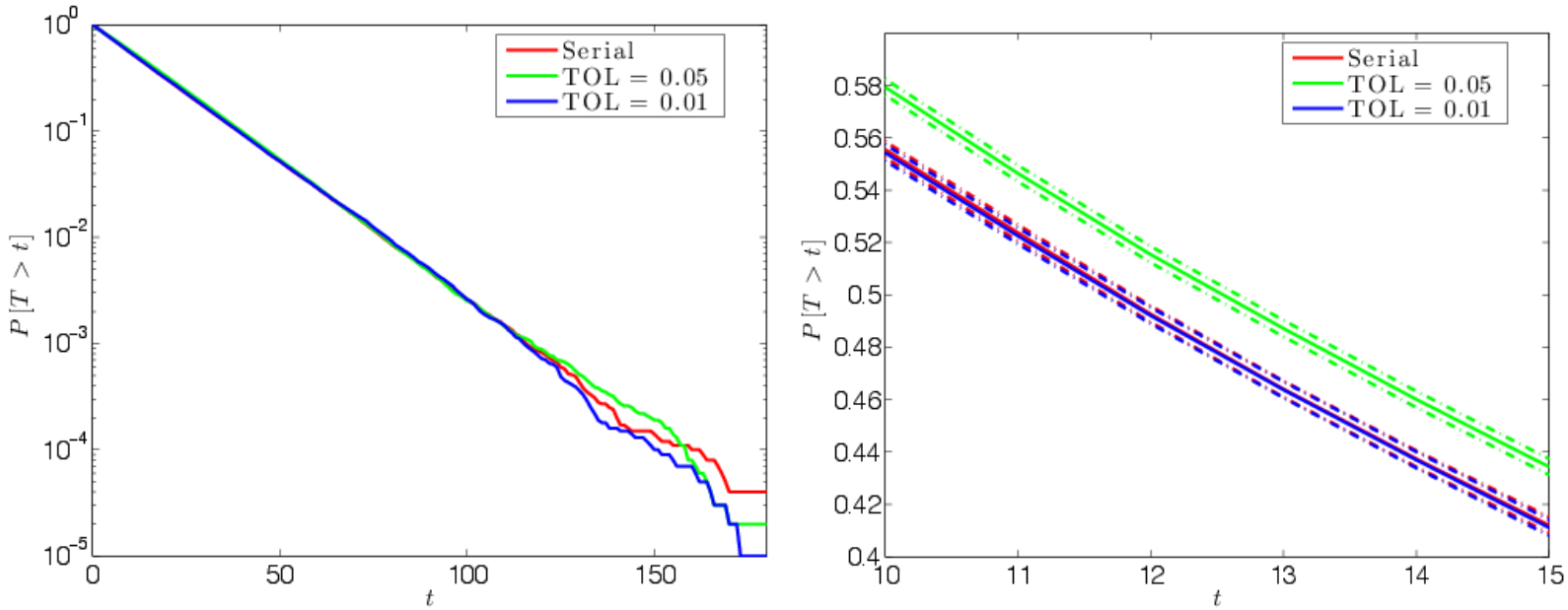


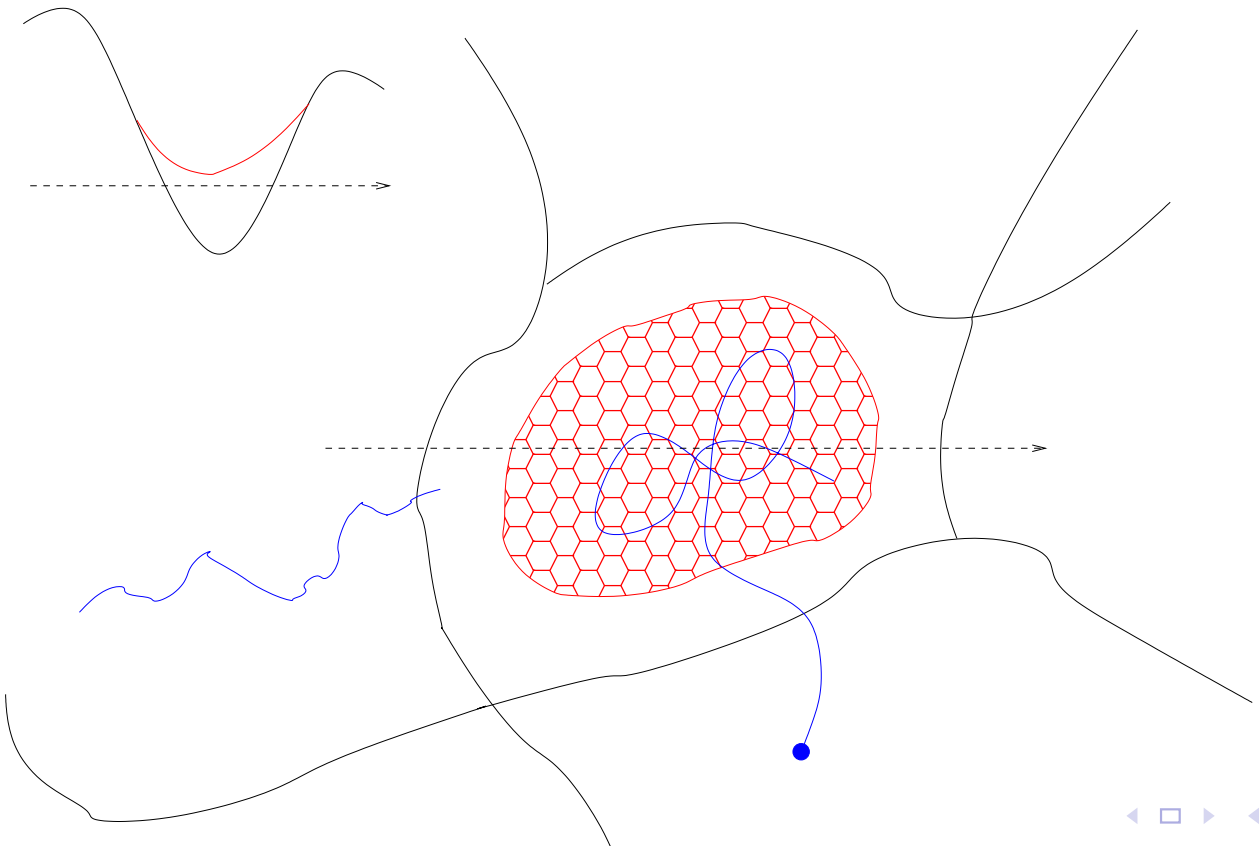
Figure: $\mathrm{LJ}_7^{\mathrm{2D}}$: Cumulative distribution function of the escape time from C_0.

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 ($f = V$, $h = 2/\beta$).

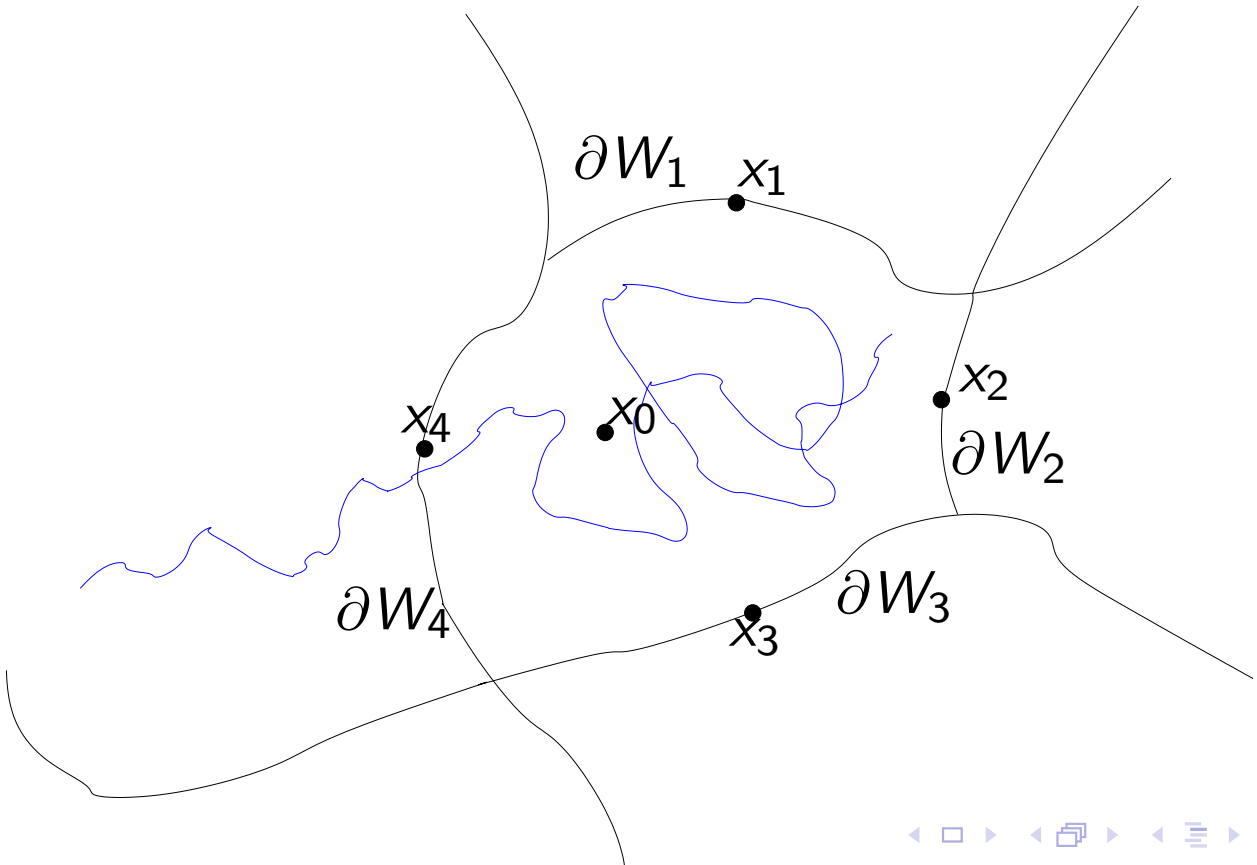
- Build quasimodes for $\Delta_{f,h}^{D,(p)}(W)$ $(p = 0, 1)$ using eigenvectors of $\Delta_{f,h}^{N,(p)}(W^-)$ $(p = 0, 1)$ and of $\Delta_{f,h}^{D,(1)}(W \setminus \overline{W^-})$.
- Analyze the asymptotics of the singular values of the restricted differential ($\nu(h) \leq h$ and $\lim_{h\to 0} h \log(\nu(h)) = 0$)
  $d_{f,h} : F^{(0)} \to F^{(1)}$ 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.

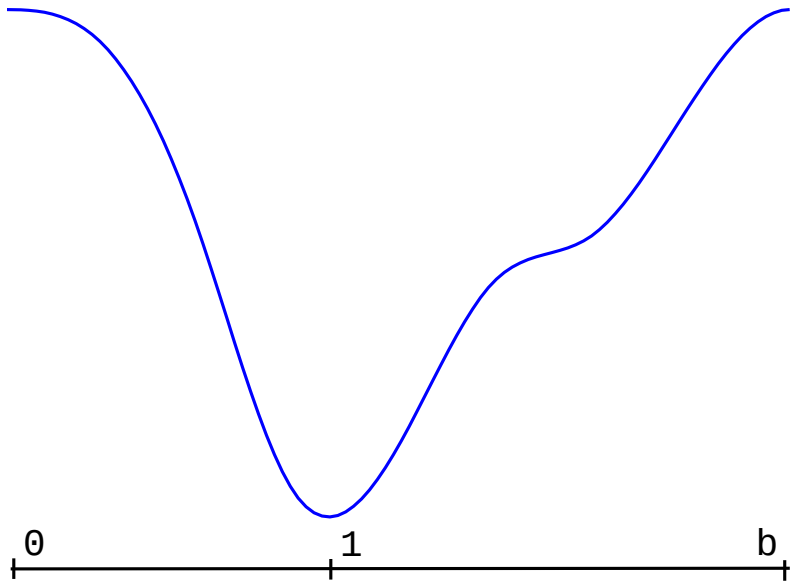
Remark: TAD can be seen as a smart saddle point search method.

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]:

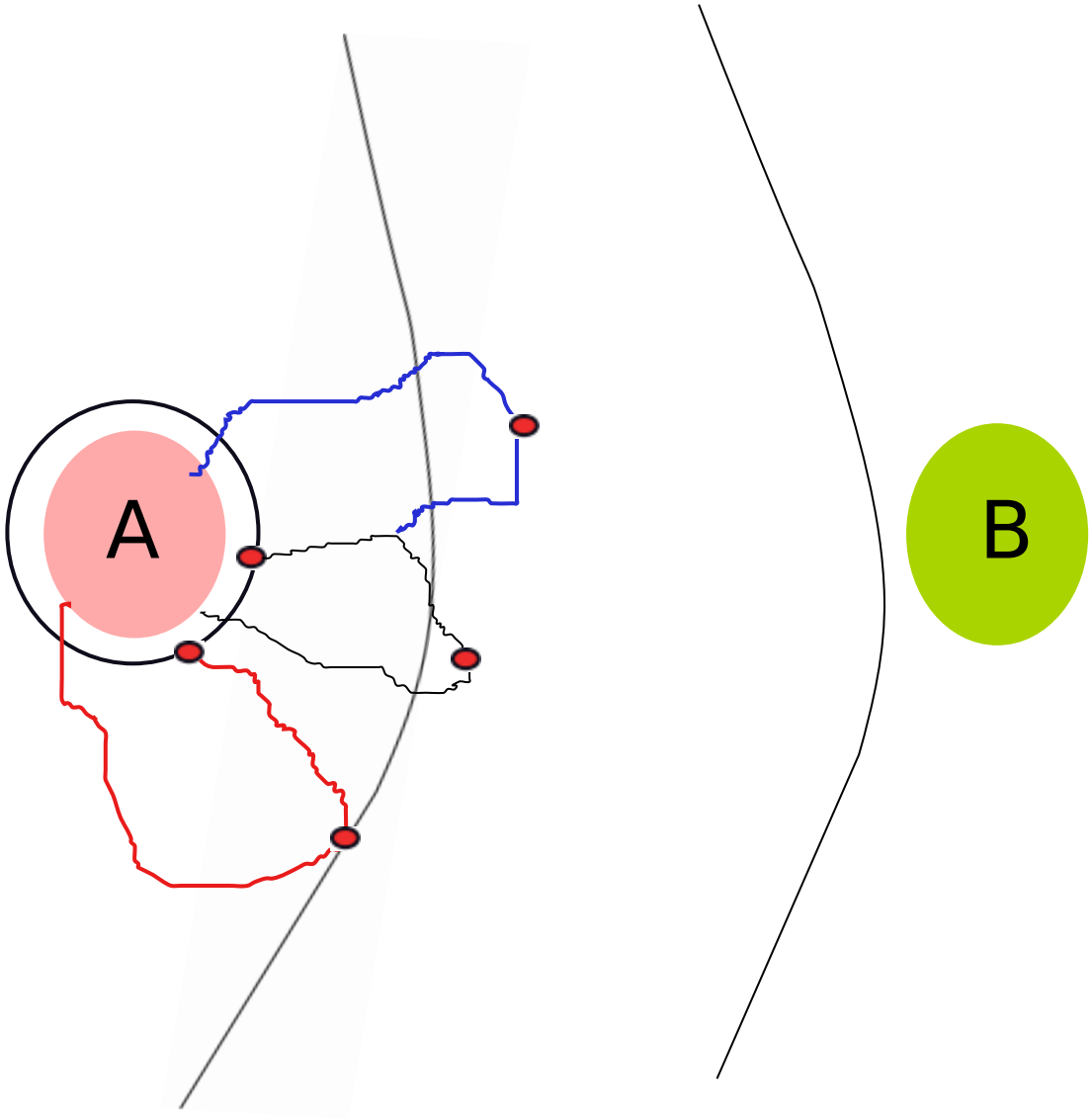


$$\frac{\lambda^{hi} p_i^{hi}}{\lambda^{lo} p_i^{lo}} = e^{-(\beta^{hi}-\beta^{lo})(V(x_i)-V(x_0))} \left(1 + O\left(\frac{1}{\beta^{hi}} - \frac{1}{\beta^{lo}}\right)\right)$$

Concluding 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.

Splitting strategies



Multilevel splitting

We would like to sample trajectories between two given metastable states A and B. The main assumption in this section is that we are given a smooth one dimensional function $\xi : \mathbb{R}^d \to \mathbb{R}$ (s.t. $|\nabla \xi| \neq 0$) which "indexes" the transition from A to B in the following sense:

$$A \subset \{x \in \mathbb{R}^d, \xi(x) < z_{\min}\} \text{ and } B \subset \{x \in \mathbb{R}^d, \xi(x) > z_{\max}\},$$

where $z_{\min} < z_{\max}$, and $\Sigma_{z_{\min}}$ (resp. $\Sigma_{z_{\max}}$) is "close" to ∂A (resp. ∂B).

Example: $\xi(x) = \|x - x_A\|$ where $x_A \in A$ is a reference configuration in A. We are interesting in the event $\{\tau_A < \tau_B\}$, starting from an initial condition on $\Sigma_{z_{\min}}$, where

$$\tau_A = \inf\{t > 0, \boldsymbol{X}_t \in A\}, \quad \tau_B = \inf\{t > 0, \boldsymbol{X}_t \in B\}$$

and

$$\tau_z = \inf\{t > 0, \xi(\boldsymbol{X}_t) > z\}.$$

Multilevel splitting

Objective: Simulate efficiently trajectories which reach B before A and estimate $\mathbb{P}(\tau_B < \tau_A)$. This then gives dynamical information: reactive trajectories from A to B, transition times from A to B, ...

We propose a multilevel splitting approach [Kahn, Harris, 1951] [Rosenbluth, 1955] to discard failed trajectories and branch trajectories approaching the rare set. We focus on an adaptive variant [Cérou, Guyader, 2007] [Cérou, Guyader, TL, Pommier, 2010]: the Adaptive Multilevel Splitting (AMS) algorithm.

Remark: The algorithm can be seen as a kind of adaptive Forward Flux Sampling [Allen, Valeriani, Ten Wolde, 2009]. It is also related to the Interface Sampling Method [Bolhuis, van Erp, Moroni 2003] and the Milestoning method [Elber, Faradjian 2004]. See the review paper [Bolhuis, Dellago, 2009]

Reactive trajectory

A reactive trajectory between two metastable sets A and B is a piece of equilibrium trajectory that leaves A and goes to B without going back to A in the meantime [Hummer,2004] [Metzner, Schütte, Vanden-Eijnden, 2006].

Difficulty: A trajectory leaving A is more likely to go back to A than to reach B.

Splitting algorithm: basic idea

The idea of splitting algorithms (FFS, TIS, RESTART, ...) is to write the rare event

$$\{\tau_B < \tau_A\}$$

as a sequence of nested events: for $z_{\min} < z_1 < \ldots < z_{\max}$,

$$\{\tau_{z_1} < \tau_A\} \subset \{\tau_{z_2} < \tau_A\} \subset \ldots \subset \{\tau_{z_{\max}} < \tau_A\} \subset \{\tau_B < \tau_A\}$$

and to simulate the successive *conditional events*: for $q = 1, 2, \ldots,$

$$\{\tau_{z_{q+1}} < \tau_A\} \text{ knowing that } \{\tau_{z_q} < \tau_A\}.$$

It is then easy to build an unbiased estimator of

$$\mathbb{P}(\tau_B < \tau_A) = \mathbb{P}(\tau_{z_1} < \tau_A)\mathbb{P}(\tau_{z_2} < \tau_A | \tau_{z_1} < \tau_A) \ldots \mathbb{P}(\tau_B < \tau_A | \tau_{z_{\max}} < \tau_A)$$

Splitting algorithm: adaptive level computation

Problem: How to choose the intermediate levels $(z_q)_{q\geq 1}$?

It is easy to check, for a given number of intermediate levels, the optimum in terms of variance is attained if

$$\forall q \geq 1,\ \mathbb{P}(\tau_{z_q} < \tau_A | \tau_{z_{q-1}} < \tau_A) = \mathbb{P}(\tau_{z_2} < \tau_A | \tau_{z_1} < \tau_A).$$

This naturally leads to adaptive versions (AMS, nested sampling) where the levels are determined by using *empirical quantiles*: Fix $k < n$; at iteration $q \geq 1$, given n trajectories $(\boldsymbol{X}^{\ell}_{t\wedge\tau_A})_{t>0,\ell=1,\ldots,n}$ in the event $\{\tau_{z_{q-1}} < \tau_A\}$, choose z_q so that

$$\mathbb{P}(\tau_{z_q} < \tau_A | \tau_{z_{q-1}} < \tau_A) \simeq \left(1 - \frac{k}{n}\right).$$

The level z_q is the k-th order statistics of $\sup_{t\geq 0} \xi(\boldsymbol{X}^{\ell}_{t\wedge\tau_A})$:

$$\sup_{t\geq 0} \xi(\boldsymbol{X}^{(1)}_{t\wedge\tau_A}) < \ldots < \sup_{t\geq 0} \xi(\boldsymbol{X}^{(k)}_{t\wedge\tau_A}) =: z_q < \ldots < \sup_{t\geq 0} \xi(\boldsymbol{X}^{(n)}_{t\wedge\tau_A}).$$

AMS: estimator of the rare event probability (1/2)

Let Q_{iter} be the number of iterations to reach the level $z_{\max}$:

$$Q_{\mathrm{iter}} = \min\{q \geq 0, z_q > z_{\max}\}$$

(where z_0 is the k-th order statistics of the n initial trajectories). Then, one obtains the estimator:

$$\left(1 - \frac{k}{n}\right)^{Q_{\mathrm{iter}}} \simeq \mathbb{P}(\tau_{z_{\max}} < \tau_A).$$

AMS: estimator of the rare event probability (2/2)

At iteration Q_{iter}, one has an ensemble of n trajectories starting from $\Sigma_{z_{\min}}$ and such that $\tau_{z_{\max}} < \tau_A$. Thus

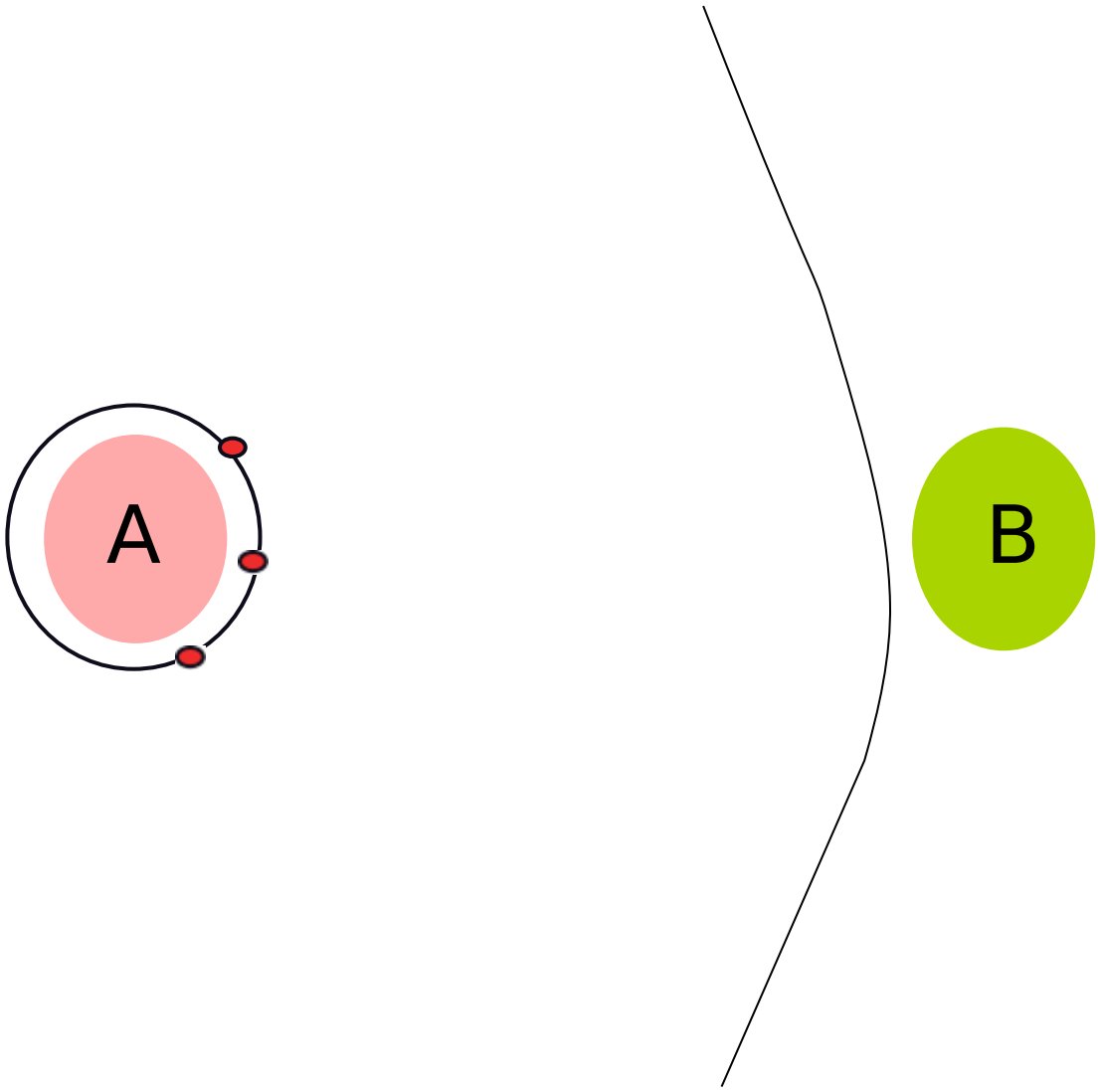
$$\hat{p}_{\text{corr}} := \frac{1}{n} \sum_{\ell=1}^{n} 1_{\{T_B(\boldsymbol{X}^{\ell,Q_{\text{iter}}}) < T_A(\boldsymbol{X}^{\ell,Q_{\text{iter}}})\}} \simeq \mathbb{P}(\tau_B < \tau_A | \tau_{z_{\max}} < \tau_A).$$

$\hat{p}_{\text{corr}}$ is the number of trajectories reaching B before A at the last iteration Q_{iter}.

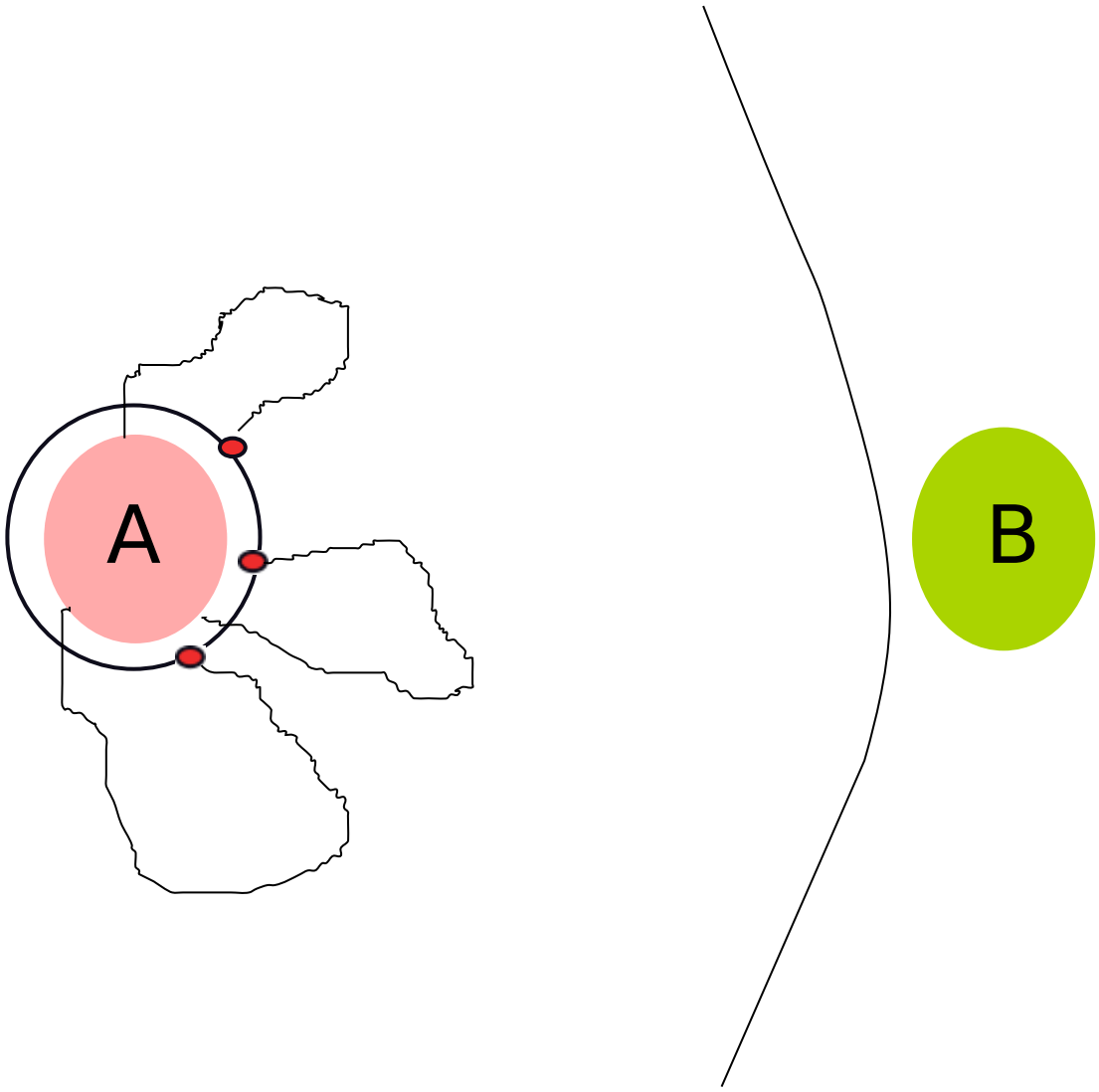
Therefore, an estimator of $\mathbb{P}(\tau_B < \tau_A)$ is

$$\left(1 - \frac{k}{n}\right)^{Q_{\text{iter}}} \hat{p}_{\text{corr}}.$$

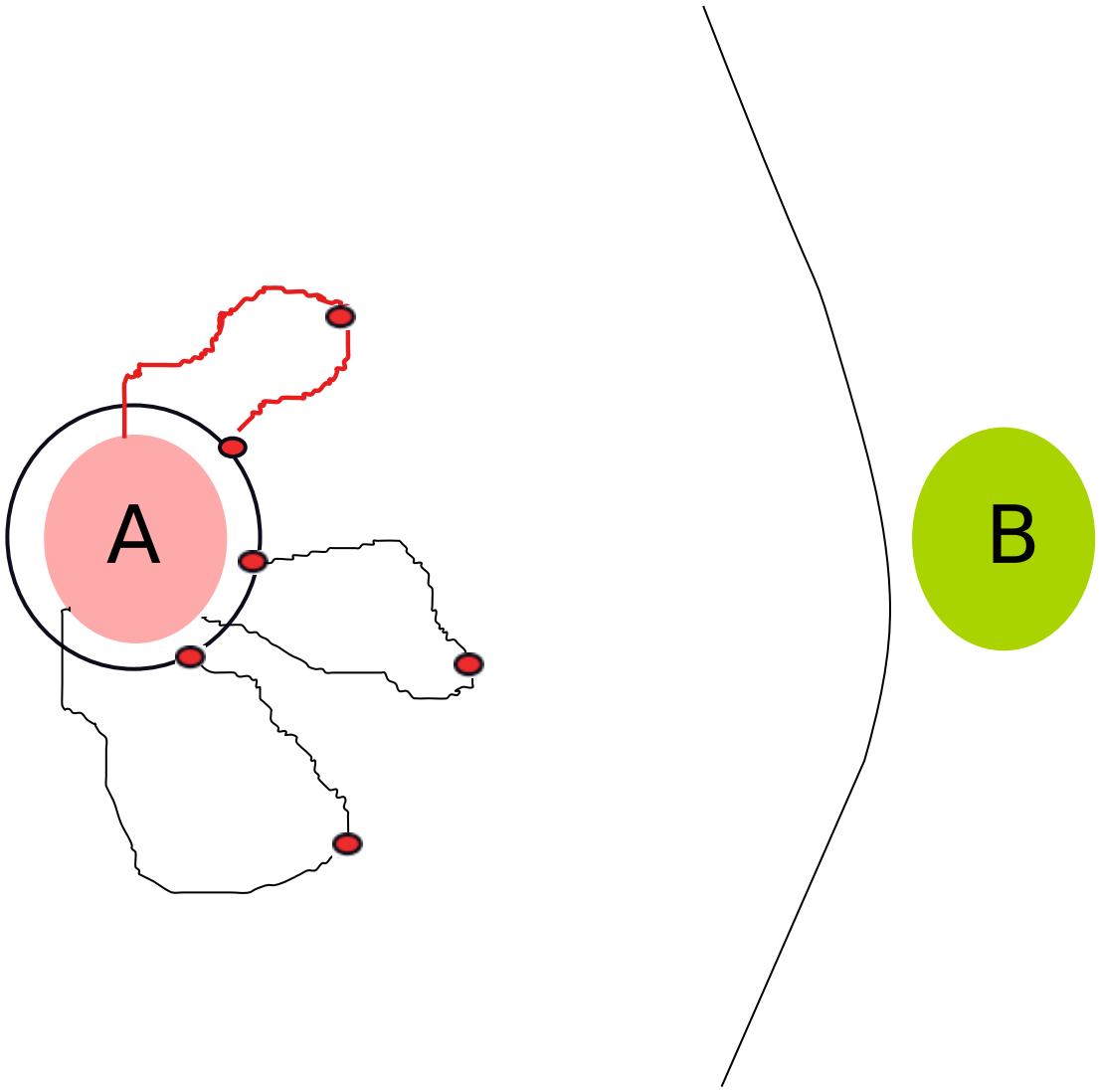
AMS Algorithm



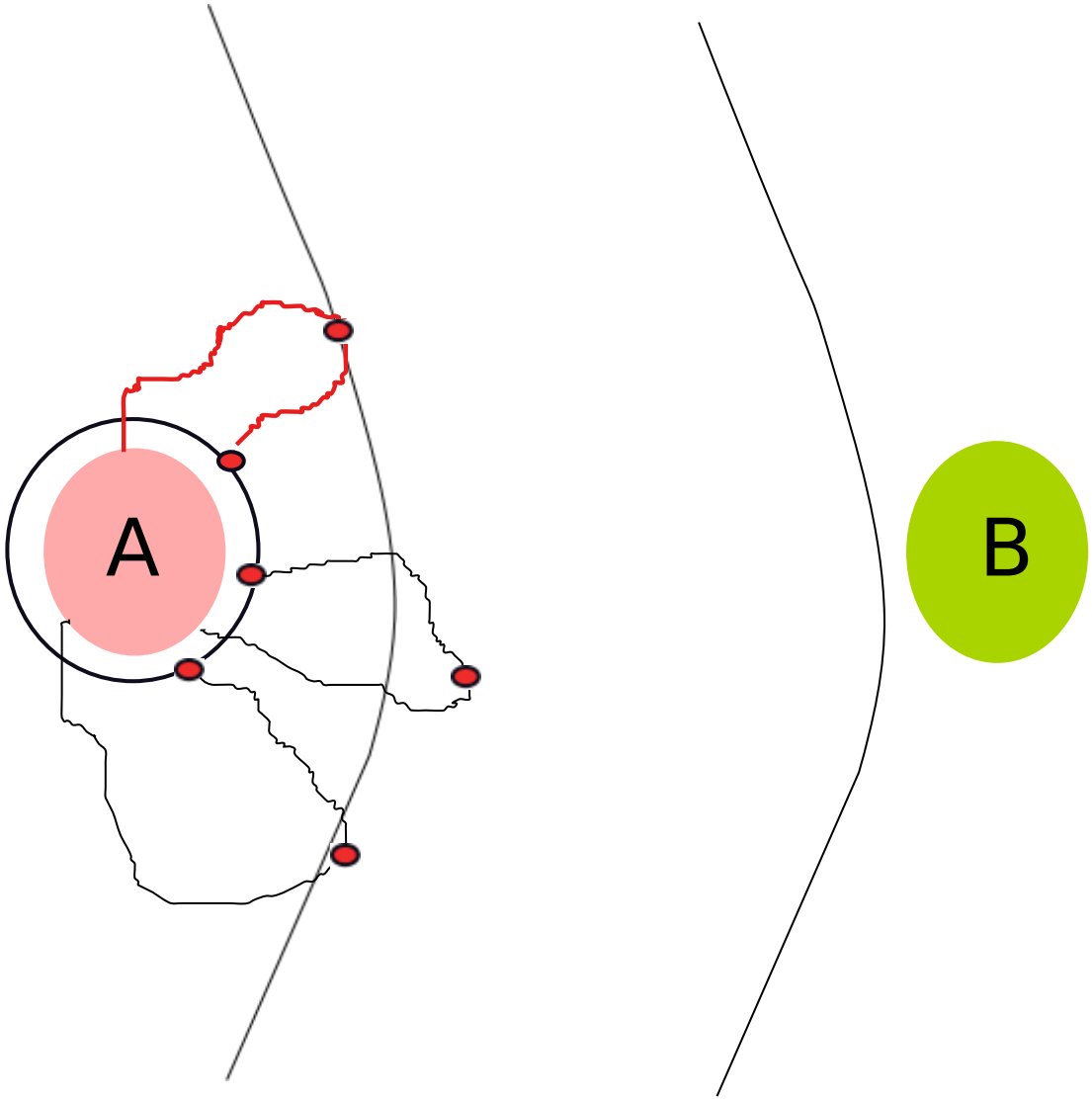
AMS Algorithm



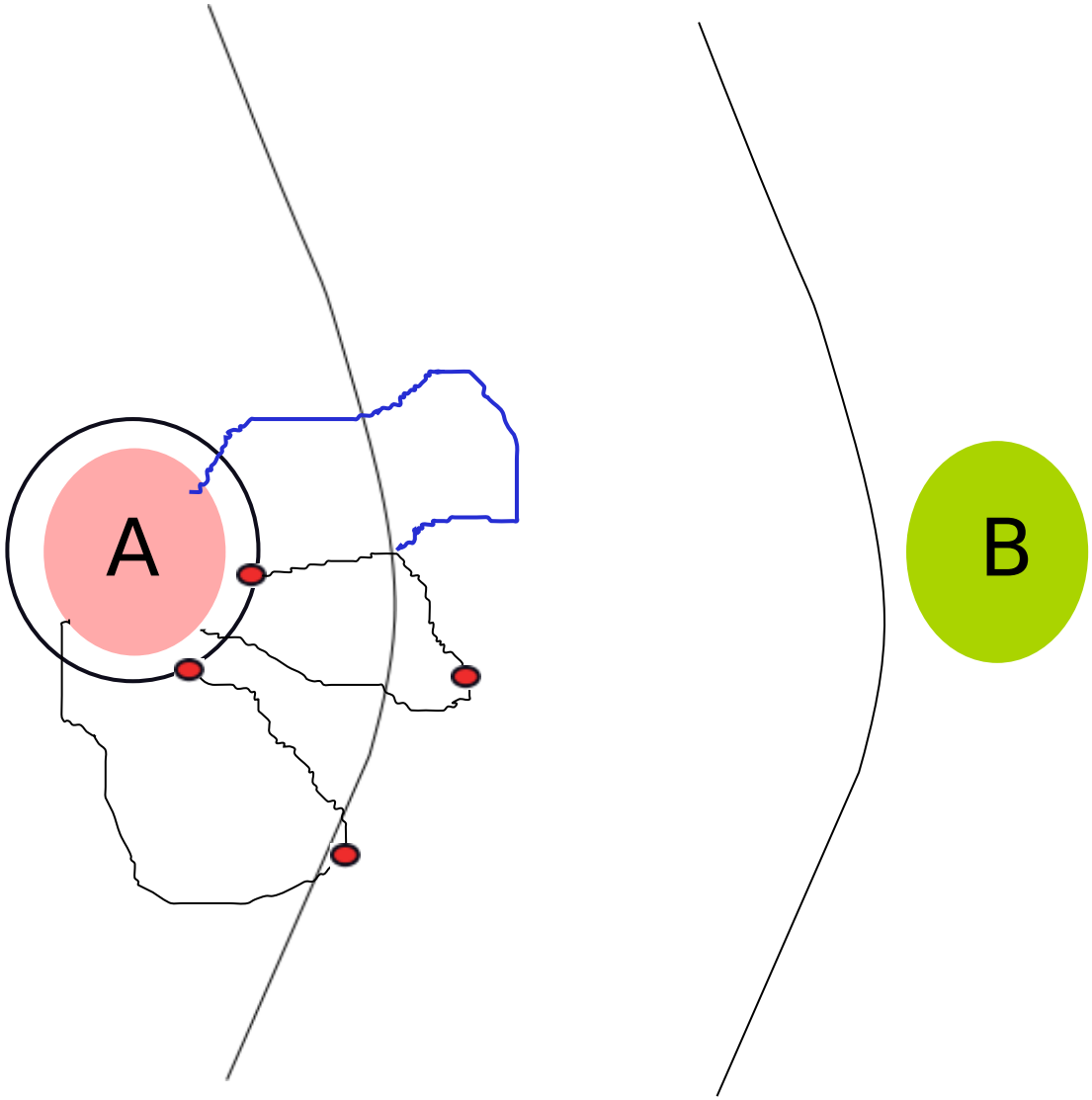
AMS Algorithm



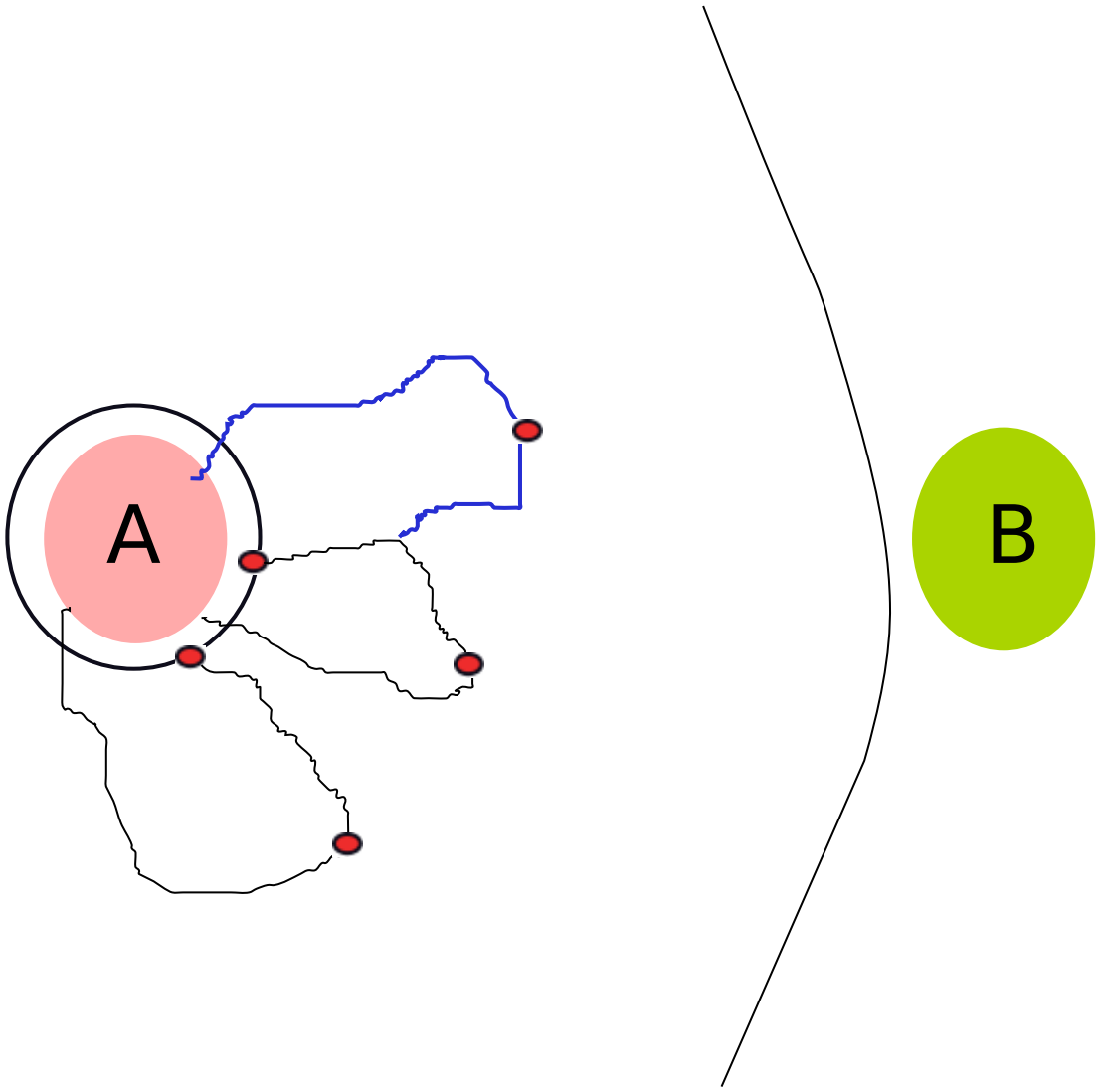
AMS Algorithm



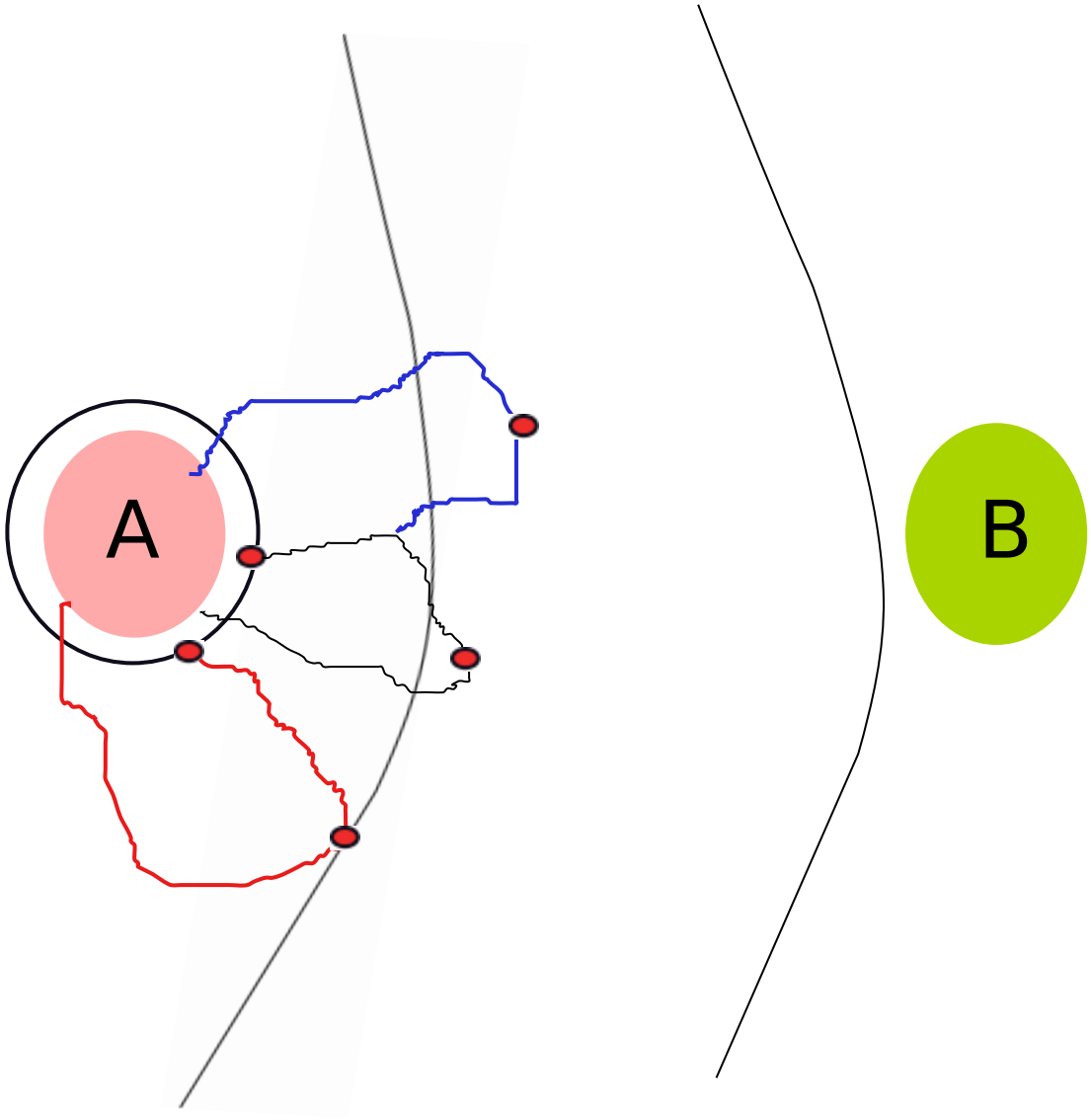
AMS Algorithm



AMS Algorithm



AMS Algorithm



AMS Algorithm: the case of Markov chains

In practice, the dynamics are *discrete in time* and thus, it may happen that more than k trajectories are such that

$$\sup_{t\geq 0} \xi(\boldsymbol{X}^{\ell}_{t\wedge\tau_A}) \leq \sup_{t\geq 0} \xi(\boldsymbol{X}^{(k)}_{t\wedge\tau_A}) =: z_q$$

In this case, all the trajectories with maximum level smaller or equal than z_q should be discarded.

The actual estimator of $\mathbb{P}(\tau_B < \tau_A)$ thus reads:

$$\hat{p} = \left(1 - \frac{K_1}{n}\right) \dots \left(1 - \frac{K_{Q_{\text{iter}}}}{n}\right) \hat{p}_{\text{corr}}$$

instead of $\left(1 - \frac{k}{n}\right)^{Q_{\text{iter}}} \hat{p}_{\text{corr}}$, where $K_q \geq k$ is the effective number of discarded trajectories at iteration q.

AMS Algorithm: unbiasedness

Theorem [C.-E. Bréhier, M. Gazeau, L. Goudenège, TL, M. Rousset, 2015]: For any choice of ξ, n and k,

$$\mathbb{E}(\hat{p}) = \mathbb{P}(\tau_B < \tau_A).$$

The proof is based on Doob's stopping theorem on a martingale built using filtrations indexed by the level sets of ξ. Actually, this result is proved for general path observables and in a much more general setting.

Practical counterparts:

- The algorithm is easy to parallelize.
- One can compare the results obtained with different reaction coordinates ξ to gain confidence in the results.

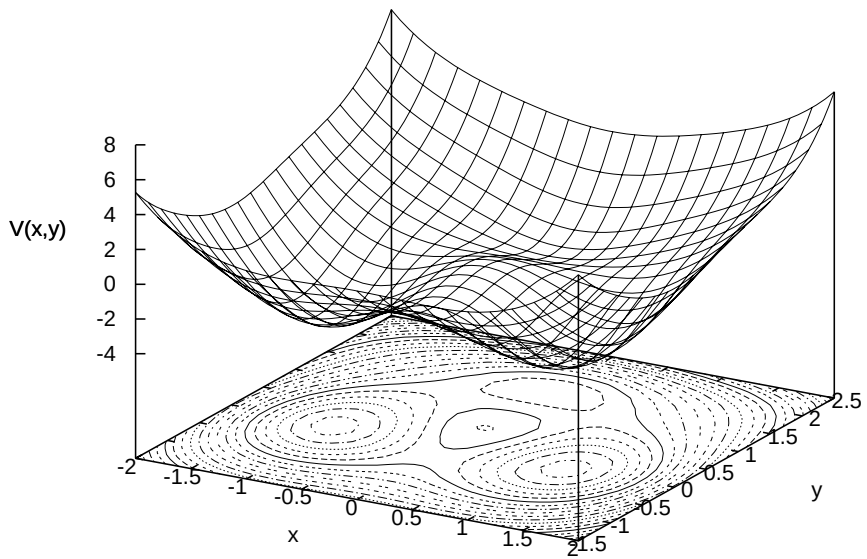
Numerical results: a 2D example

Time-discretization of the overdamped Langevin dynamics:

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

with a deterministic initial condition $\boldsymbol{X}_0 = \boldsymbol{x}_0$ and the 2D potential

[Park, Sener, Lu, Schulten, 2003] [Metzner, Schütte and Vanden-Eijnden, 2006]



$$V(x,y) = 3\mathrm{e}^{-x^2-(y-\frac{1}{3})^2} - 3\mathrm{e}^{-x^2-(y-\frac{5}{3})^2} - 5\mathrm{e}^{-(x-1)^2-y^2}$$
$$- 5\mathrm{e}^{-(x+1)^2-y^2} + 0.2x^4 + 0.2\left(y-\frac{1}{3}\right)^4.$$

A 2D example

The interest of this "bi-channel" potential is that, depending on the temperature, one or the other channel is prefered to go from A (around $H_- = (-1, 0)$) to B (around $H_+ = (1, 0)$).

Three reaction coordinates: $\xi^1(x, y) = \|(x, y) - H_-\|$, $\xi^2(x, y) = C - \|(x, y) - H_+\|$ or $\xi^3(x, y) = x$.

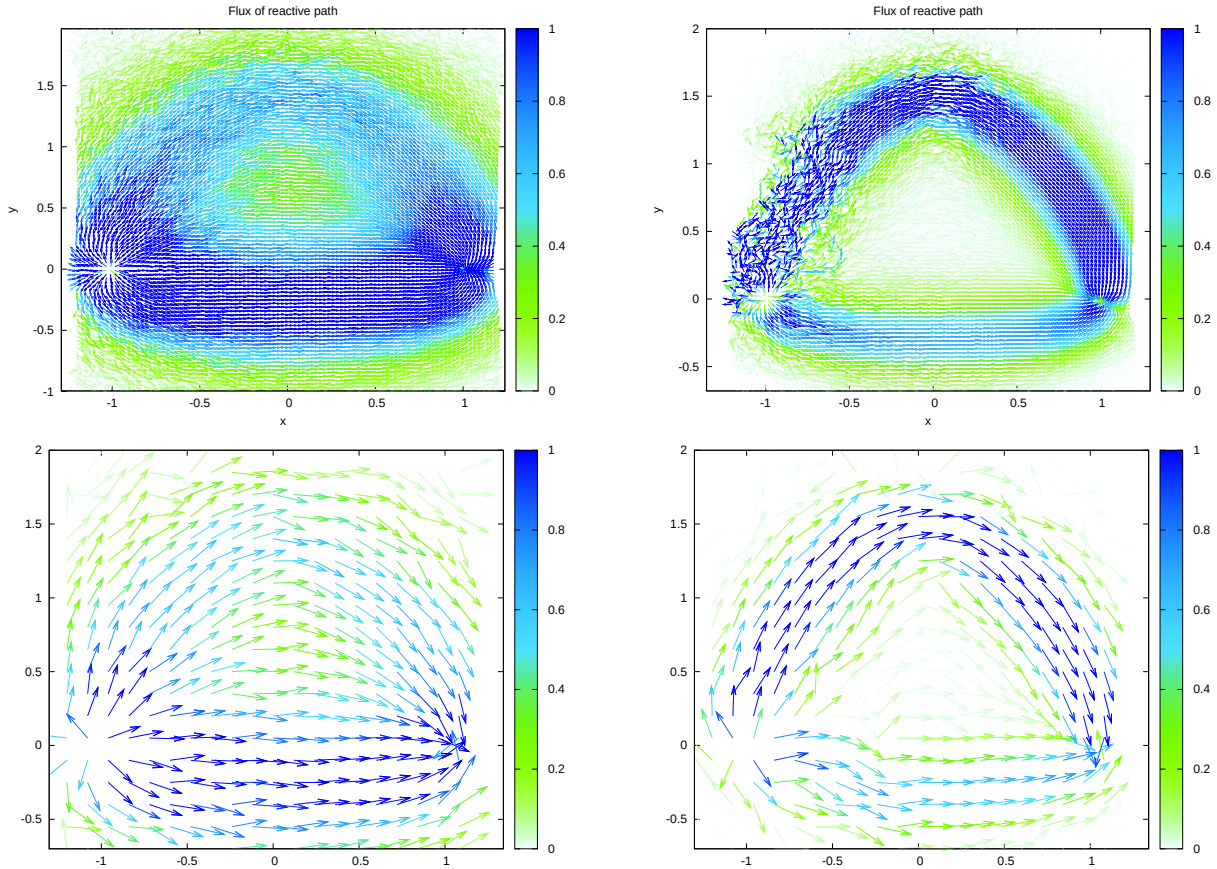
We plot as a function of the number N of independent realizations of AMS, the empirical average

$$\overline{p}_N = \frac{1}{N} \sum_{m=1}^{N} \hat{p}_m$$

together with the associated empirical confidence interval: $[\overline{p}_N - \delta_N/2, \overline{p}_N + \delta_N/2]$ where

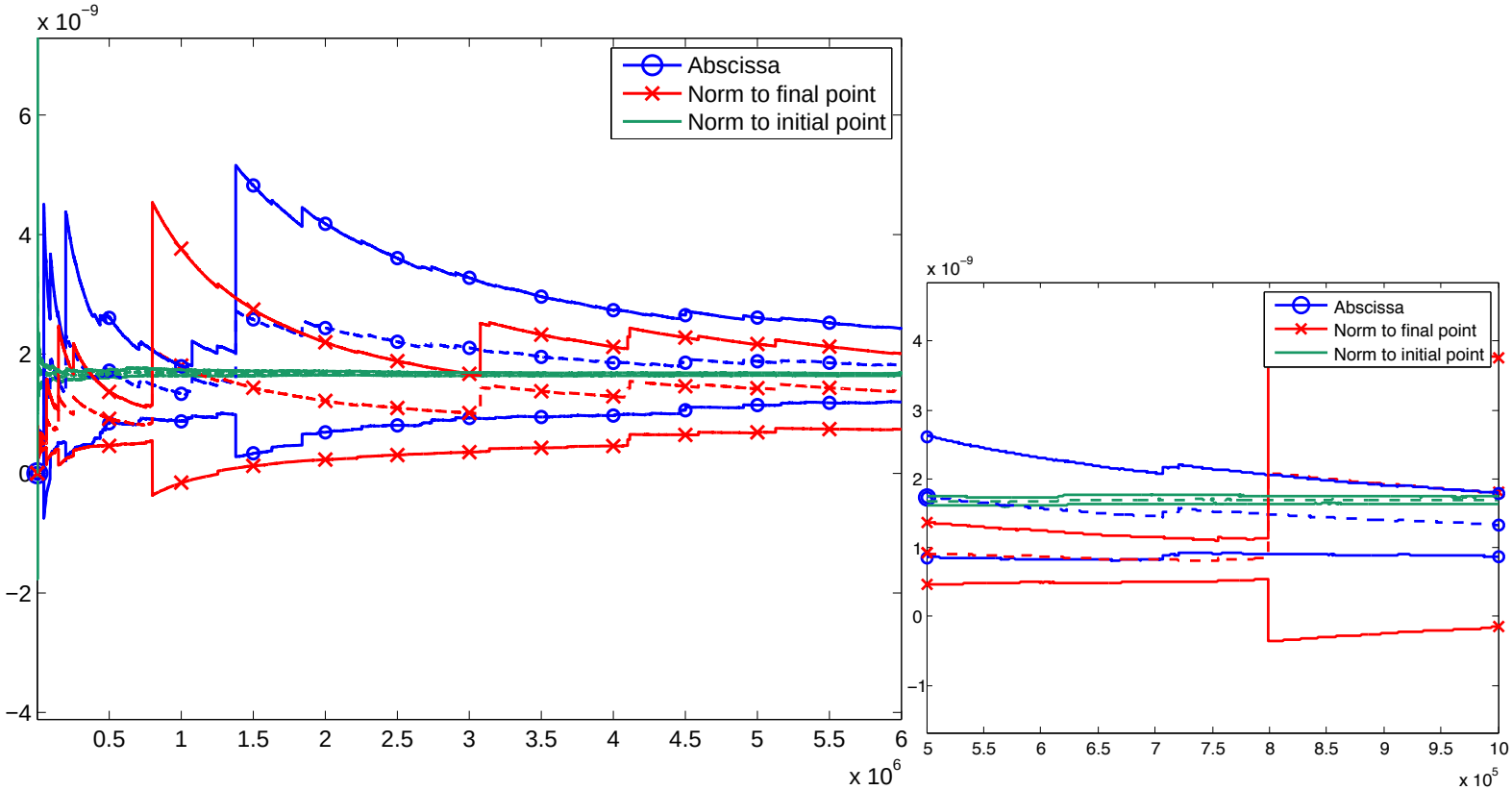
$$\delta_N = 2 \frac{1.96}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{m=1}^{N} (\hat{p}_m)^2 - (\overline{p}_N)^2}$$

A 2D example: flux of reactive trajectories

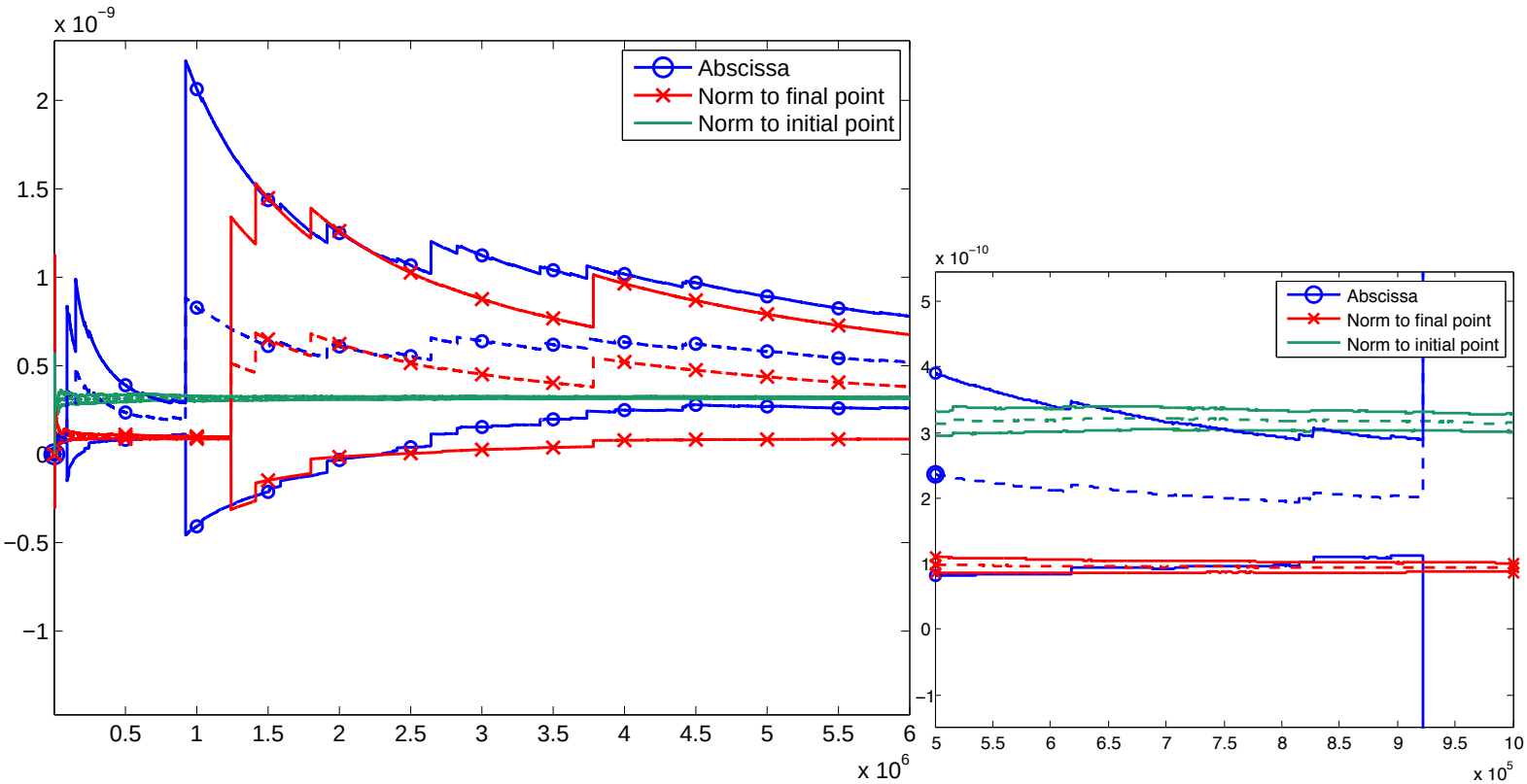


Flux of reactive trajectories, at $\beta = 1.67$ on the left, and $\beta = 6.67$ on the right.

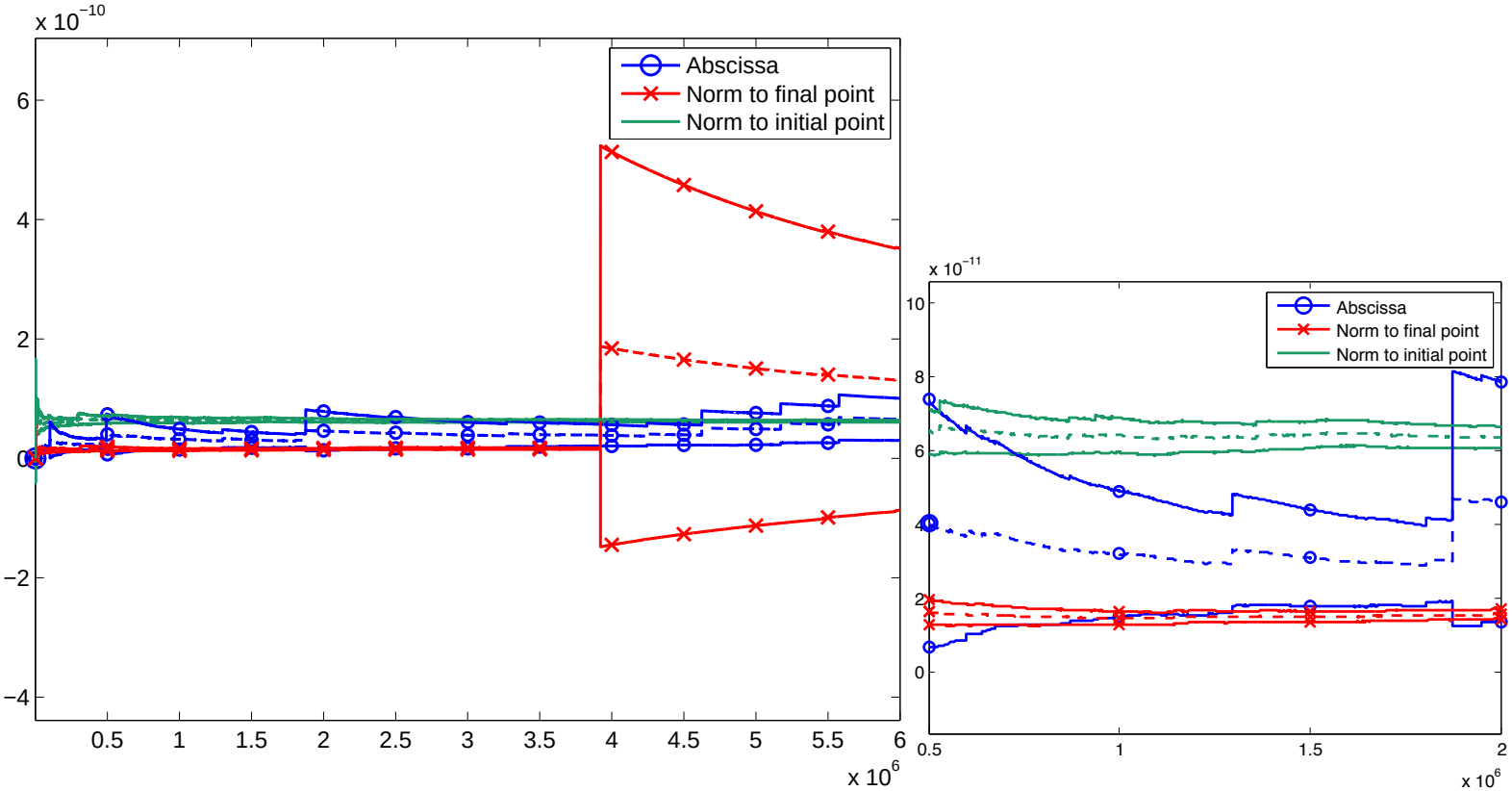
A 2D example: $k = 1$, $n = 100$, $\beta = 8.67$



A 2D example: $k = 1$, $n = 100$, $\beta = 9.33$



A 2D example: $k = 1$, $n = 100$, $\beta = 10$



A 2D example

Observations:

- When N is sufficiently large, confidence intervals overlap.
- For too small values of N, "apparent bias" is observed [Glasserman, Heidelberger, Shahabuddin, Zajic, 1998].
- Fluctuations depend a lot on ξ.

$\longrightarrow$ To gain confidence in the results, check that the estimated quantity is approximately the same for different ξ's.

"Apparent bias" phenomenon

The apparent bias is due to the fact that [Glasserman, Heidelberger, Shahabuddin, Zajic, 1998]:

- Multiple pathways exist to go from A to B.
- Conditionally to reach Σ_z before A, the relative likelihood of each of these pathways depends a lot on z.

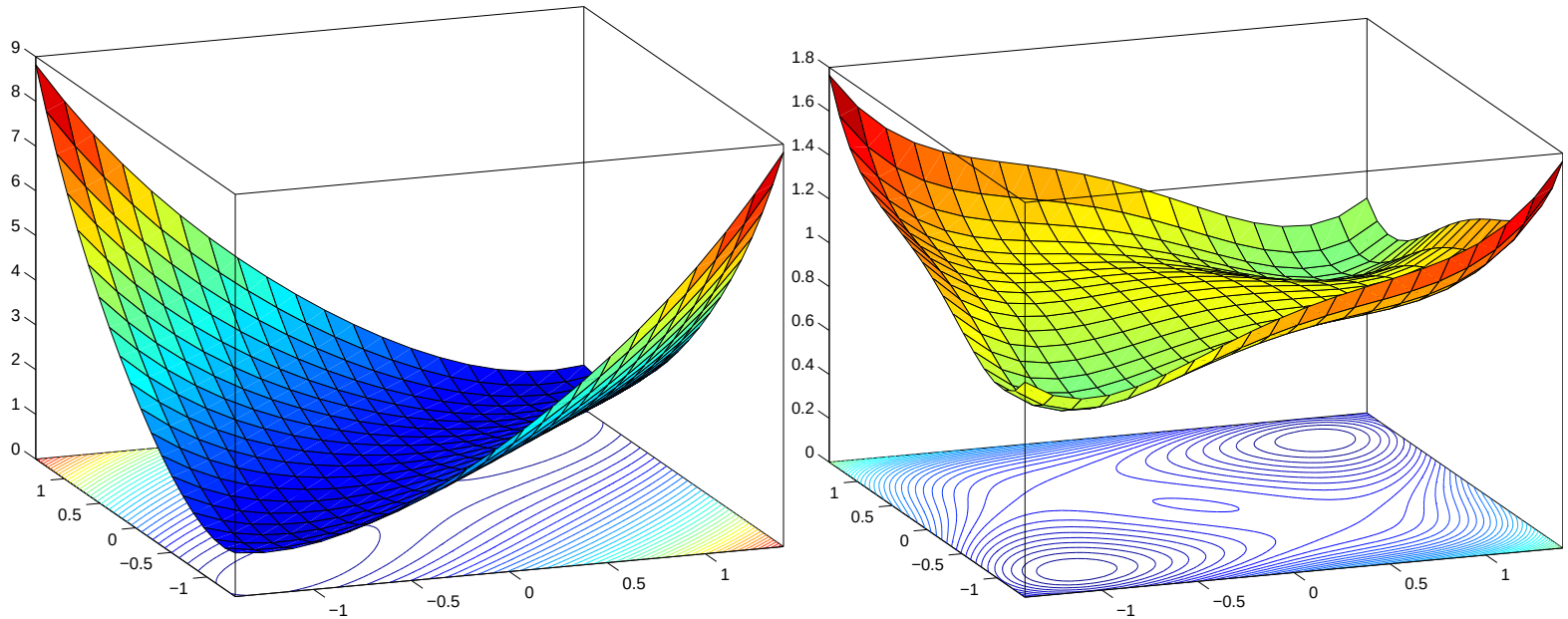
On our example, for small n, we indeed observe that (for ξ^3):

- Most of the time, all replicas at the end go through only one of the two channels (two possible scenarios).
- One of this scenario is rare.
- The values of $\hat{p}$ associated to each of these two scenarios are very different.

This explains the large fluctuations.

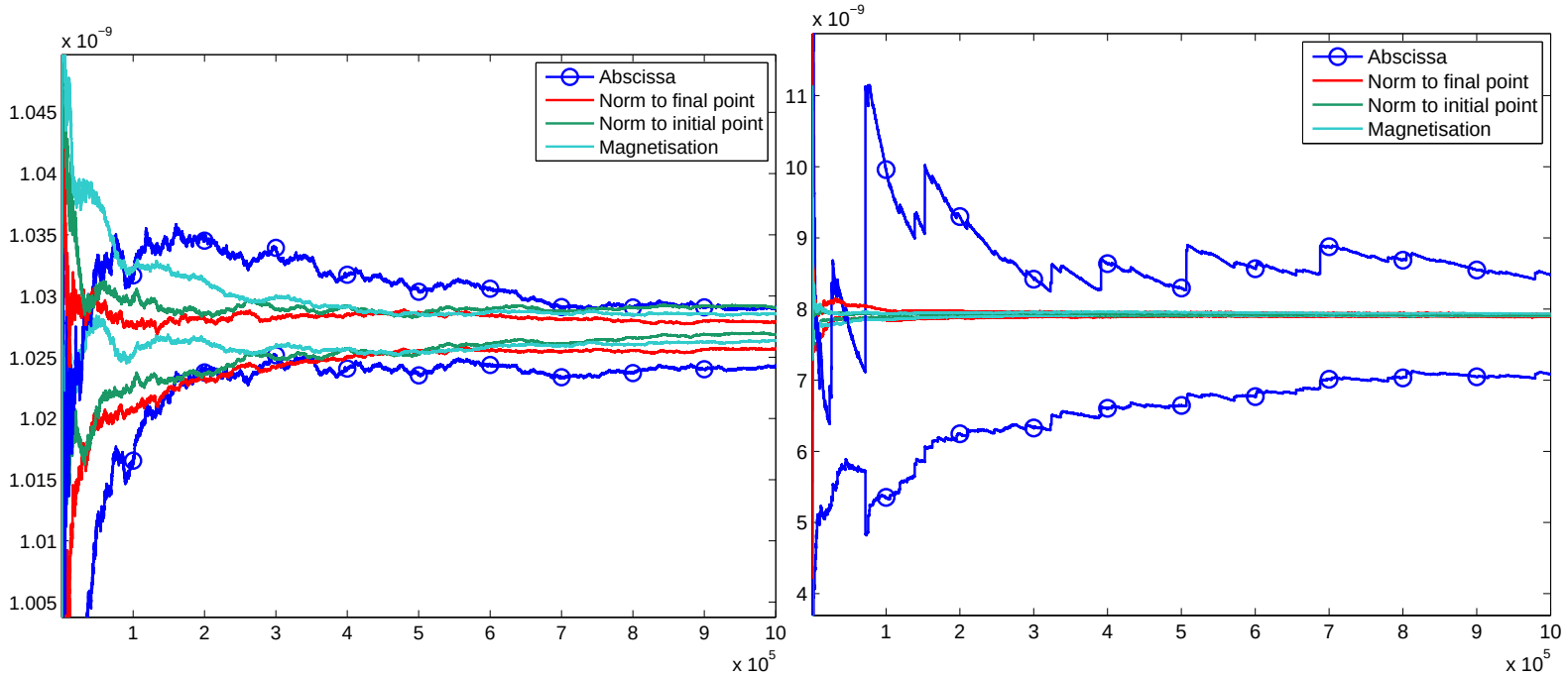
"Apparent bias" phenomenon

Another 2D test case:



Potential $V_\gamma(x, y)$.
Left: $\gamma = 1$ (one channel); right: $\gamma = 0.1$ (two channels).

"Apparent bias" phenomenon



Parameters: $k = 1$, $n = 100$ and $\beta = 80$.

Left: $\gamma = 1$ (one channel). Right: $\gamma = 0.1$ (two channels).

Current developments

The AMS algorithm can be used to study reactive trajectories and estimate transition times. The algorithm is non-intrusive and very versatile.

Works in progress:

- Implementation in the NAMD software (collaboration with SANOFI, C. Mayne and I. Teo), and in TRIPOLI (collaboration with CEA)
- Adaptive computation of better and better ξ.
- Analysis of the efficiency as a function of ξ. For optimal choice of ξ, the cost of AMS is (for n large)

$$\left((\log p)^2 - \log p\right)$$

much better than the cost of naive Monte Carlo: $\frac{1-p}{p}$. How does this degrade when ξ departs from the optimal case ?

Simulating dynamics: conclusions (1/2)

There are other mathematical settings to characterize / quantify metastability:

- Large deviation techniques [Freidlin, Wentzell, Vanden Eijnden, Weare, Touchette,...] and Onsager-Machlup functionals [Stuart, Pinsky, Theil]
- 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 (2/2)

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, 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

Review paper:

- TL and G. Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*, Acta Numerica, 2016.

Accelerated dynamics:

- C. Le Bris, TL, M. Luskin and D. Perez, *A mathematical formalization of the parallel replica dynamics*, MCMA, 2012.
- D. Aristoff and TL, *Mathematical analysis of Temperature Accelerated Dynamics*, SIAM MMS, 2014.
- A. Binder, TL and G. Simpson, *A Generalized Parallel Replica Dynamics*, Journal of Computational Physics, 2015.
- TL and F. Nier, *Low temperature asymptotics for Quasi-Stationary Distributions in a bounded domain*, Analysis & PDE, 2015.
- G. Di Gesù, D. Le Peutrec and B. Nectoux, *Jump Markov models and transition state theory: the Quasi-Stationary Distribution approach*, Faraday Discussion, 2016.

References

Adaptive Multilevel Splitting algorithm:

- C.-E. Bréhier, M. Gazeau, L. Goudenège , TL and M. Rousset, *Unbiasedness of some generalized Adaptive Multilevel Splitting algorithms*, Ann. App. Prob., 2016.
- F. Cérou, A. Guyader, TL and D. Pommier, *A multiple replica approach to simulate reactive trajectories*, J. Chem. Phys. 2011.
- TL, C. Mayne, K. Schulten and I. Teo, *Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time*, J. Chem. Th. and Comput., 2016.