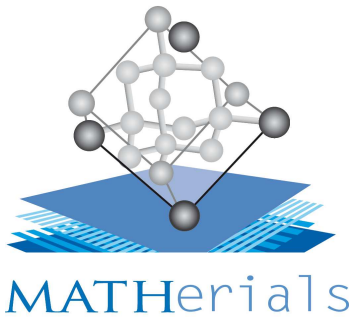
Sampling problems in computational statistical physics

2- Splitting methods for rare event simulations

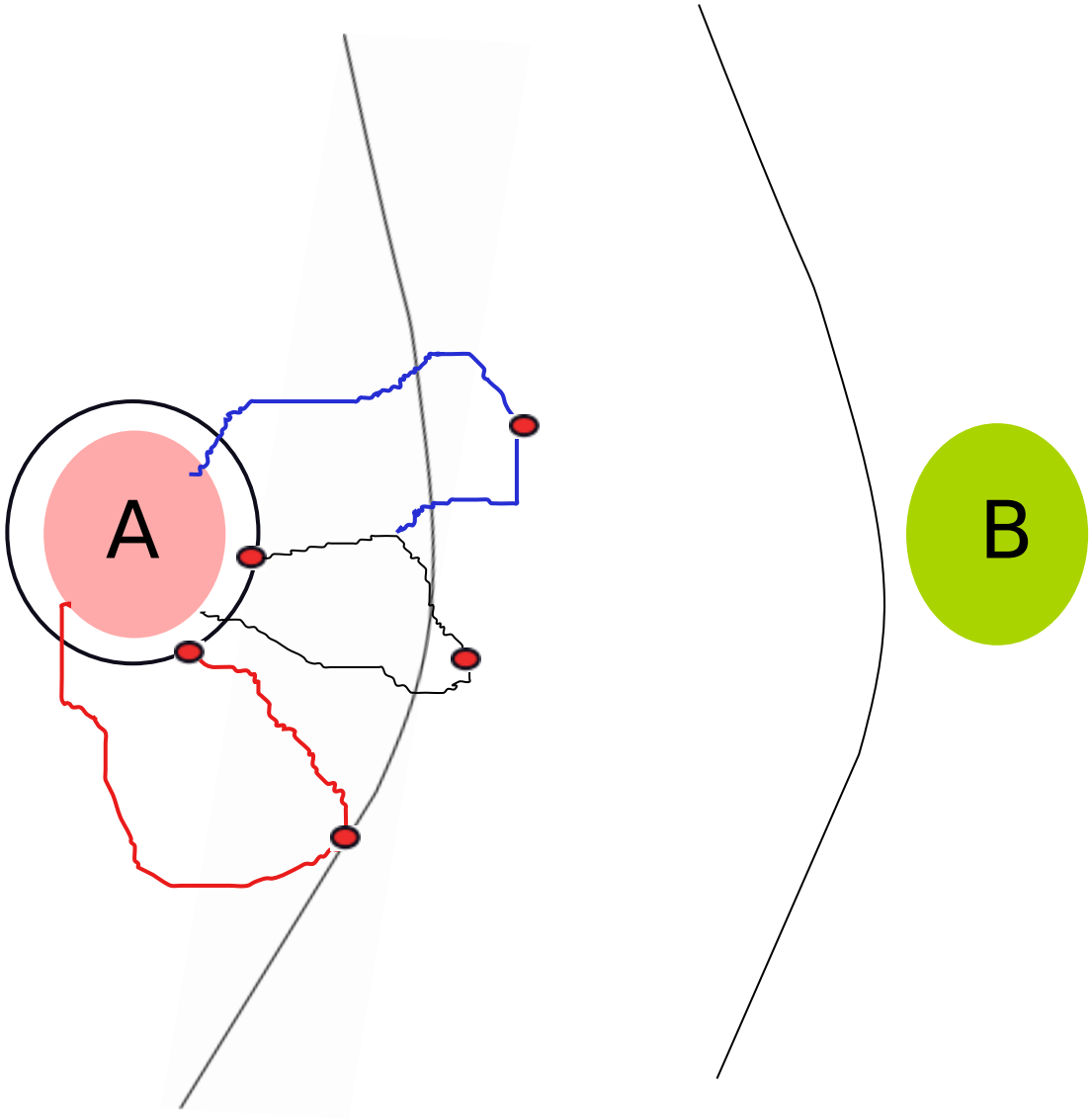
T. Lelièvre

CERMICS - Ecole des Ponts ParisTech & Equipe Matherials - INRIA



Brummer & Partners MathDataLab, KTH, 19/01/2021

The Adaptive Multilevel Splitting algorithm



Multilevel splitting

General setting: Let $(\boldsymbol{X}_t)_{t\geq 0}$ be a Markovian dynamics, and τ_B and τ_A two associated stopping times.

Objective: efficiently compute quantities of the form $\mathbb{E}[F((\boldsymbol{X}_t)_{0\leq t\leq \tau_A\wedge\tau_B})1_{\tau_B<\tau_A}]$ in the rare event setting:

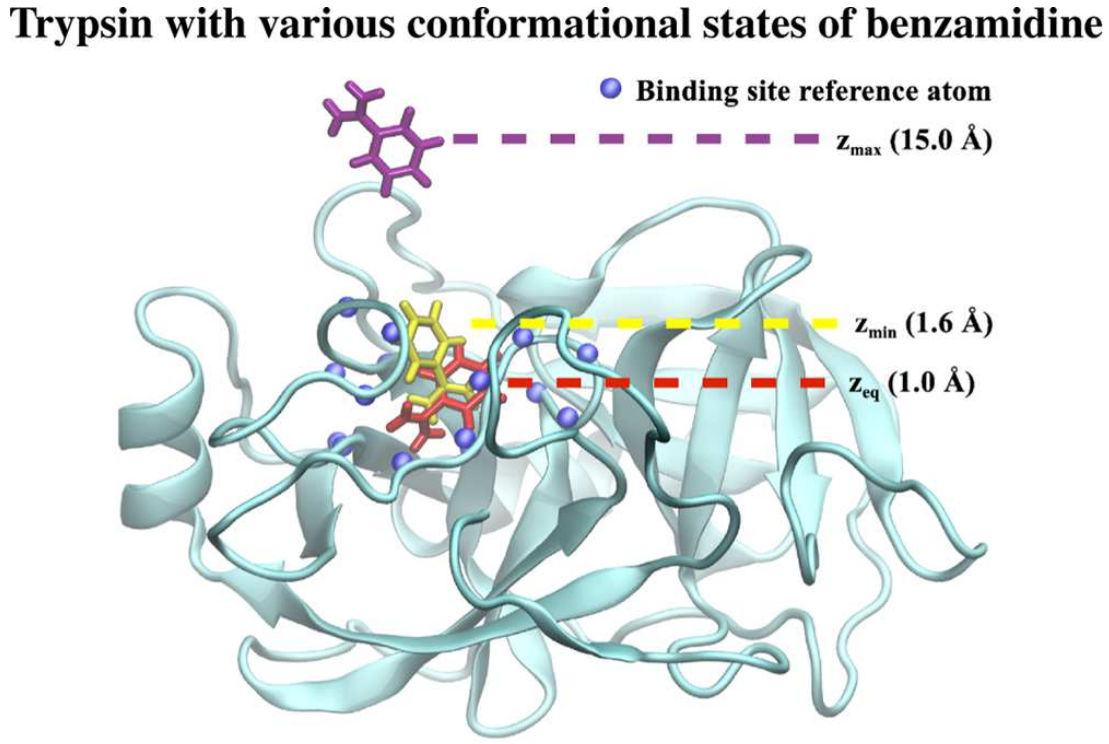
$$\mathbb{P}(\tau_B < \tau_A) \ll 1.$$

Two examples:

- Reactive trajectories: A and B are two metastable states, τ_A and τ_B are the first hitting time of A and B.
- Killed process: τ_A is a killing time, τ_B is the first hitting time of a domain B.

Motivation 1: Simulations of biological systems

Unbinding of a ligand from a protein

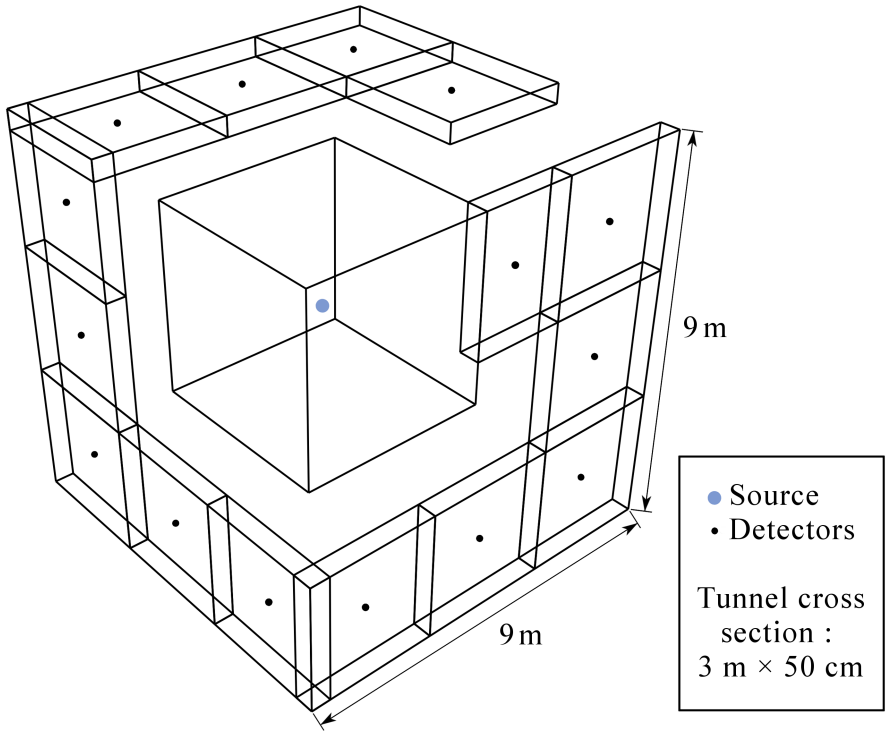


Elementary time-step for the molecular dynamics $= 10^{-15}\,\mathrm{s}$
Dissociation time $\simeq 0.02\,\mathrm{s}$

Challenge: bridge the gap between timescales

Motivation 2: Radiation protection

Monte Carlo particle transport



Concrete tunnel with a neutron source

How to compute the neutron flux at the detector ?

Challenge: the flux is very small

Multilevel splitting: the reactive trajectory setting

We would like to sample trajectories between two given metastable states A and B. The main assumption is that we are given a smooth one dimensional function $\xi : \mathbb{R}^d \to \mathbb{R}$ 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 is a reference configuration in A.

We are interested in the event $\{\tau_A < \tau_B\}$, starting from an initial condition with support in $\{x \in \mathbb{R}^d, \xi(x) < z_{\min}\}$, where

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

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 present 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, RESTART, ...) is to write the rare event

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

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

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

where $\tau_z = \inf\{t > 0, \xi(\boldsymbol{X}_t) > z\}$ and to simulate the successive *conditional events*: for $q = 1, \ldots, Q-1$,

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

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 $(z_q)_{q\geq 1}$?

In an ideal setting, 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 an adaptive version (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 such that $\tau_{z_{\max}} < \tau_A$. Thus

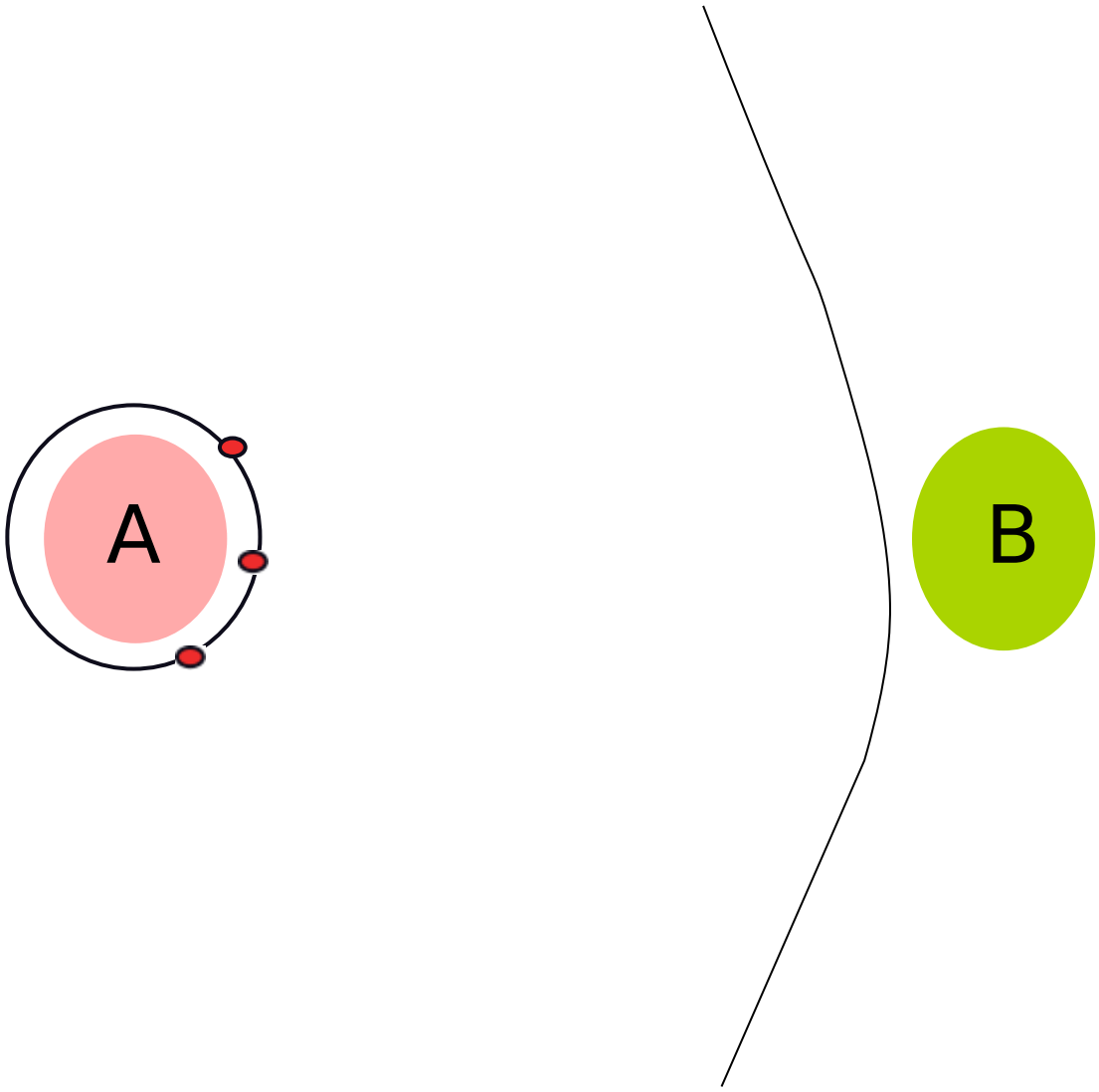
$$\hat{p}_{\text{corr}} := \frac{1}{n} \sum_{\ell=1}^{n} 1_{\{T_B(\boldsymbol{X}^{\ell,Q_{\text{iter}}}) < T_A(\boldsymbol{X}^{\ell,Q_{\text{iter}}})\}} \simeq \mathbb{P}(\tau_B < \tau_A | \tau_{z_{\max}} < \tau_A).$$

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

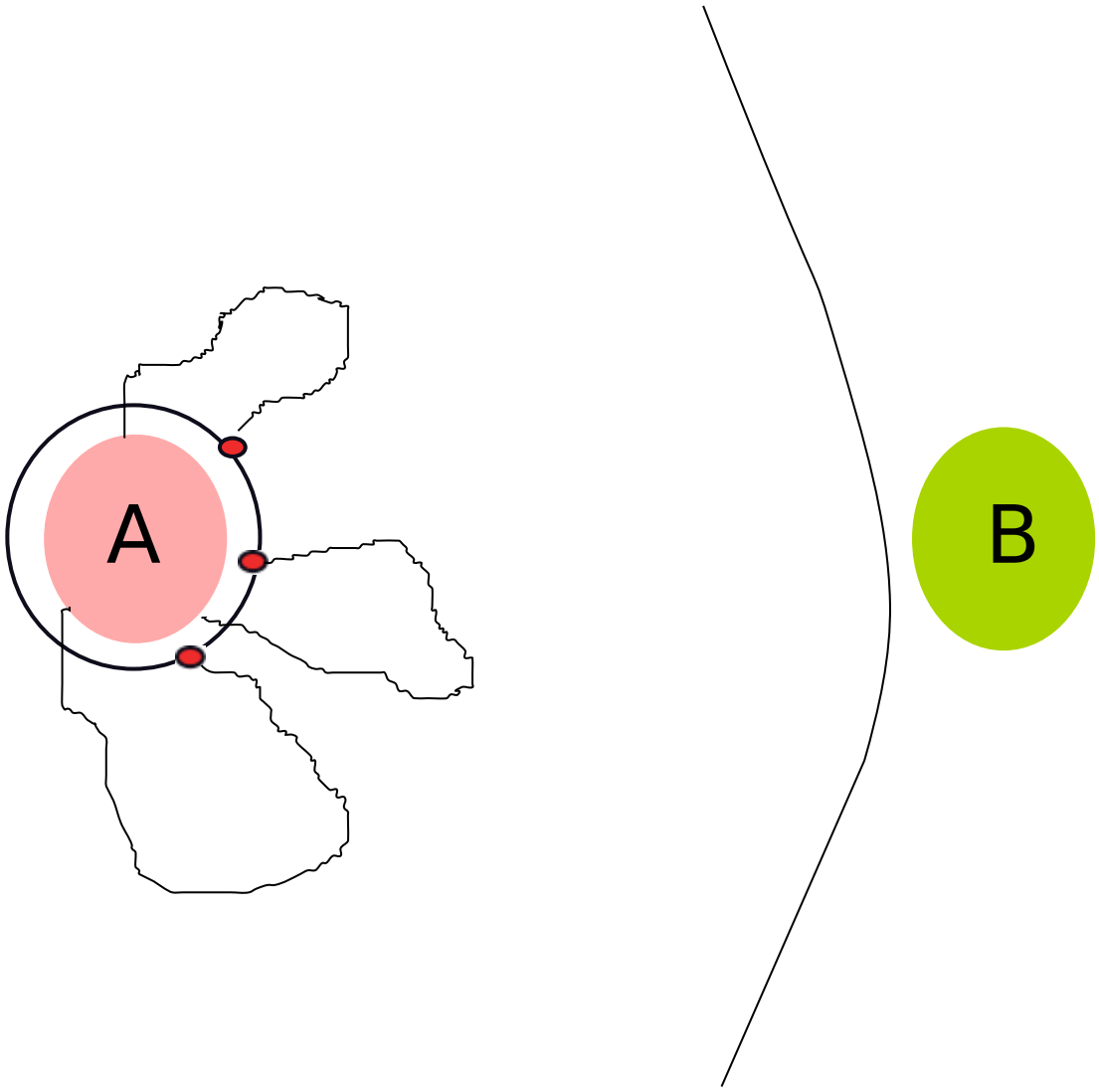
Therefore, an estimator of $\mathbb{P}(\tau_B < \tau_A)$ is

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

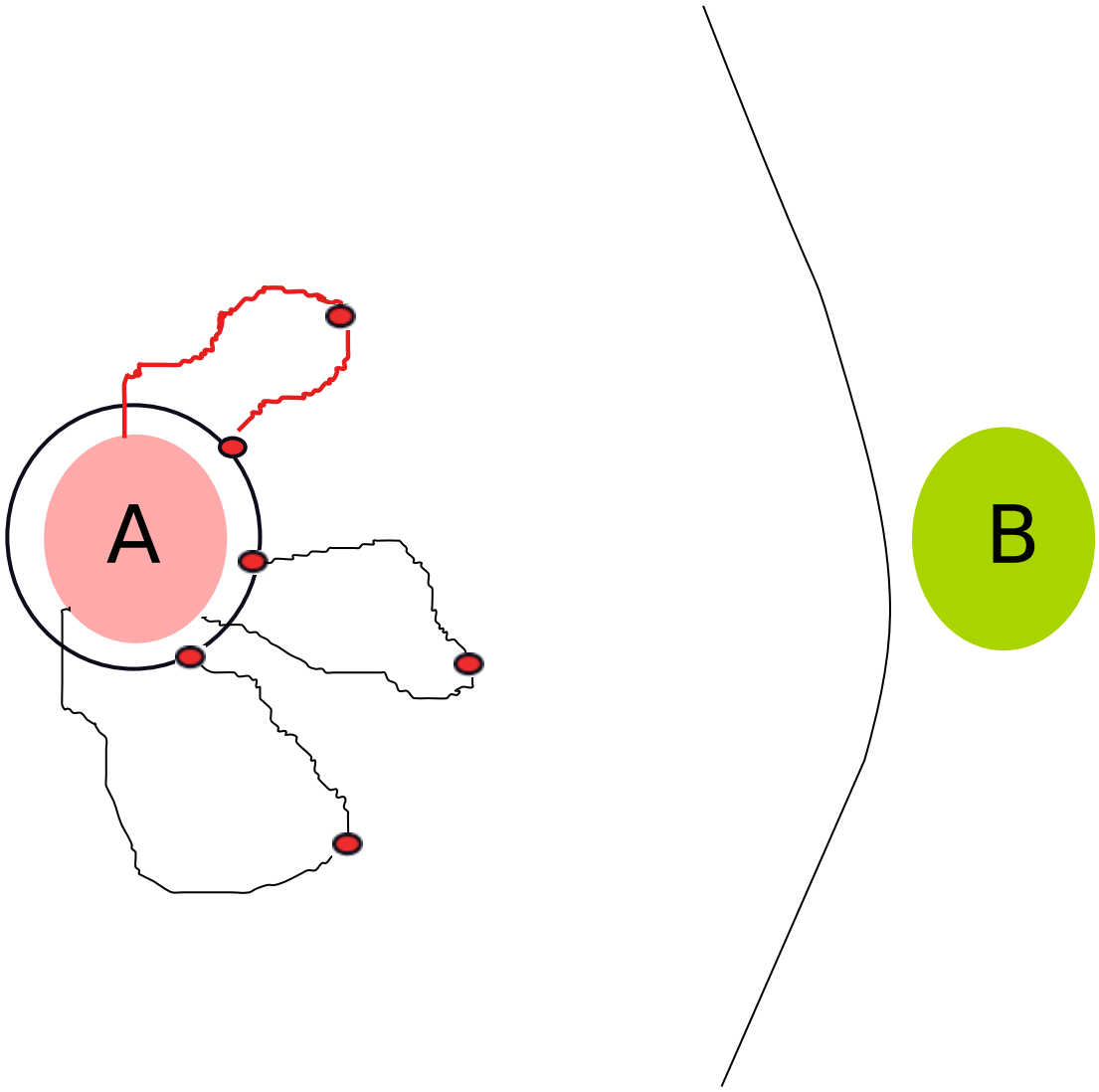
AMS Algorithm



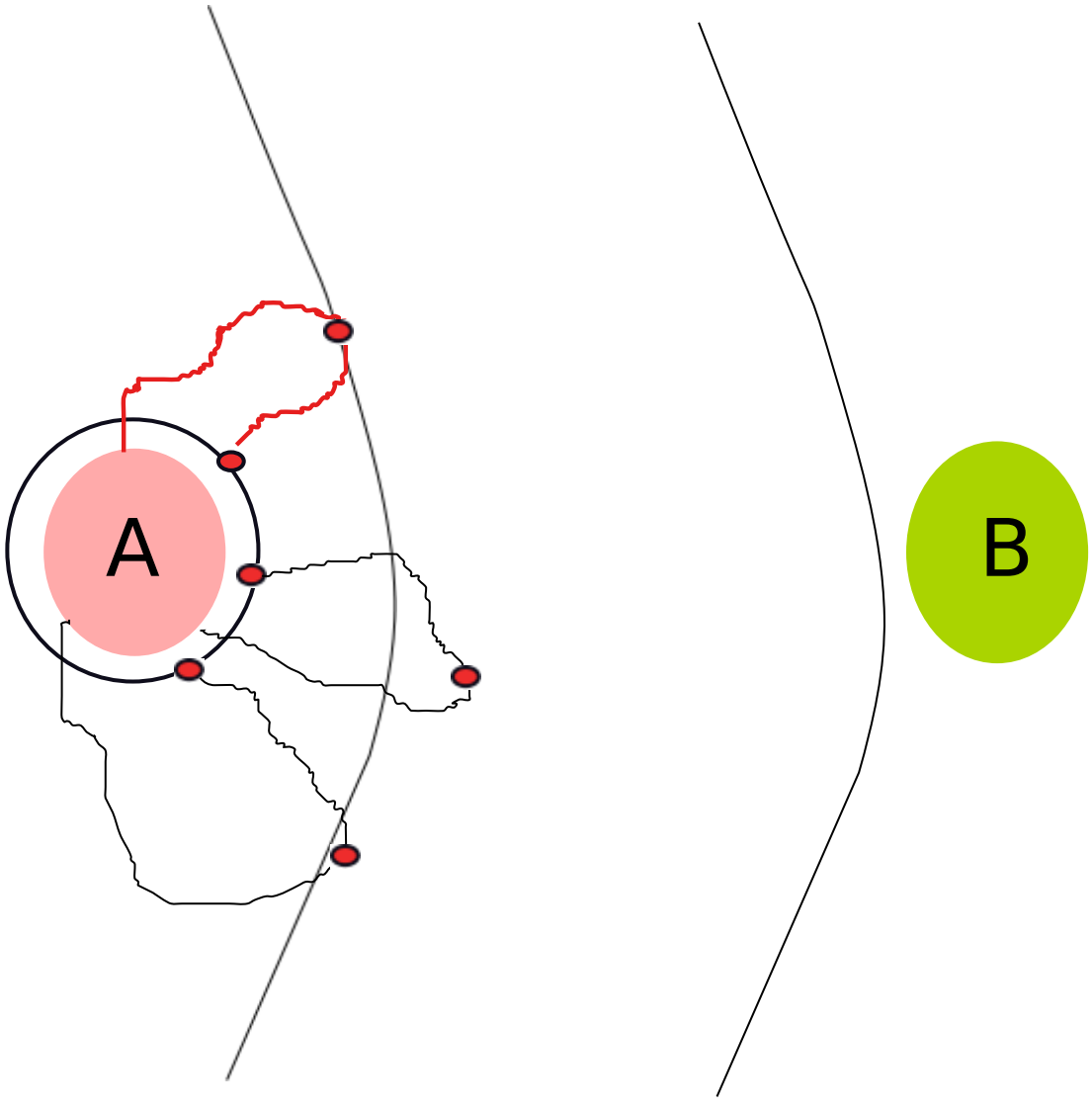
AMS Algorithm



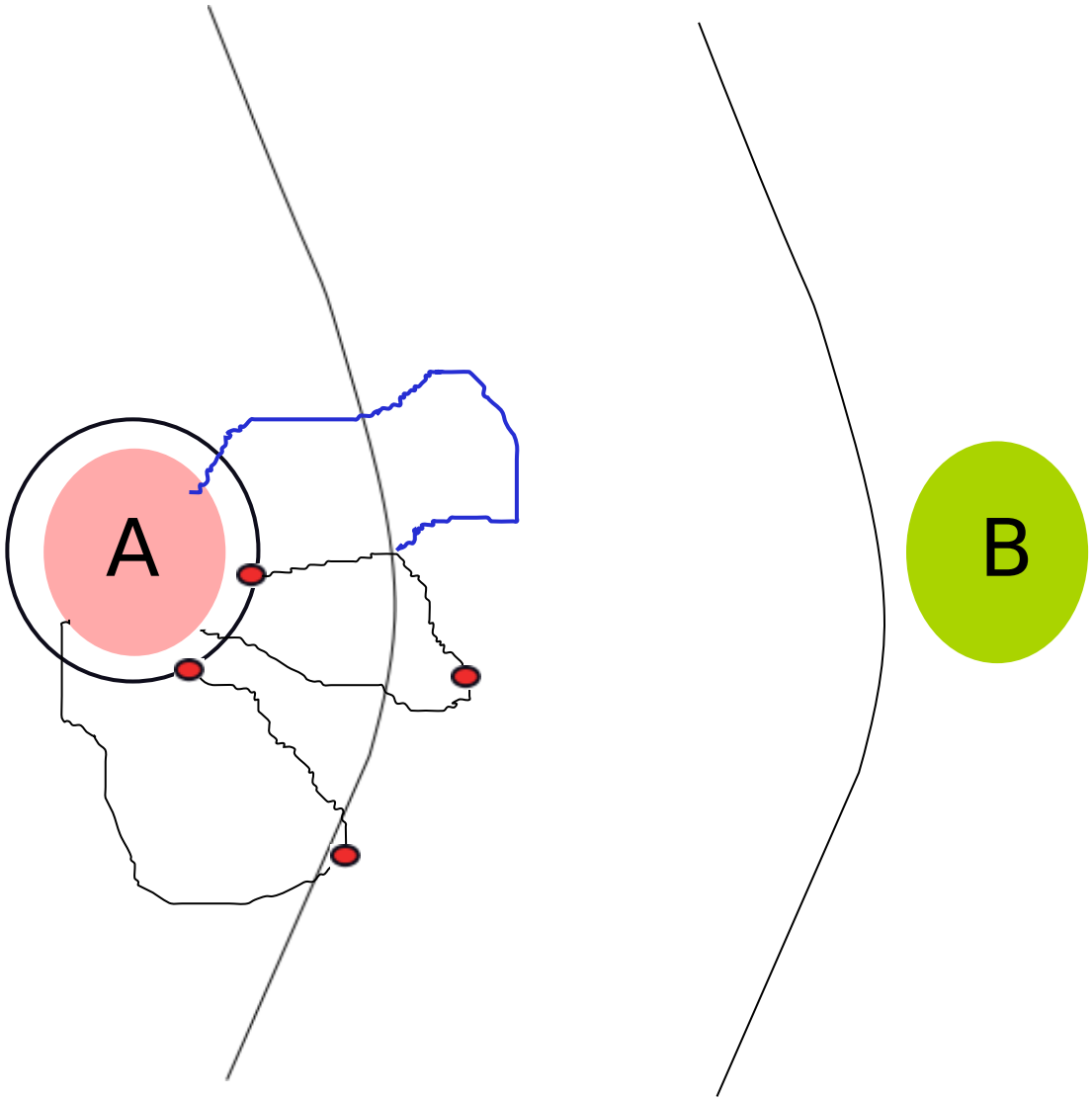
AMS Algorithm



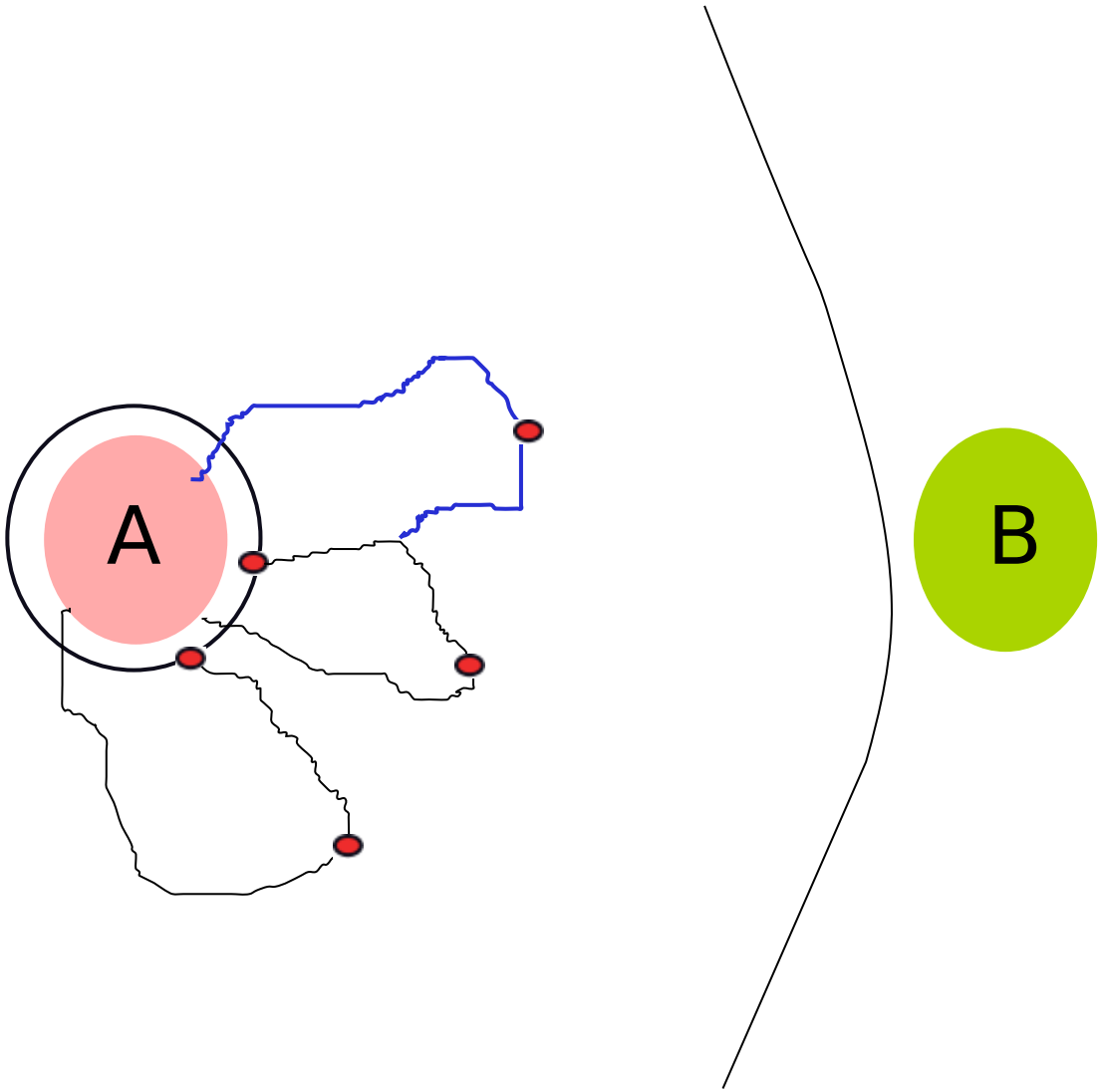
AMS Algorithm



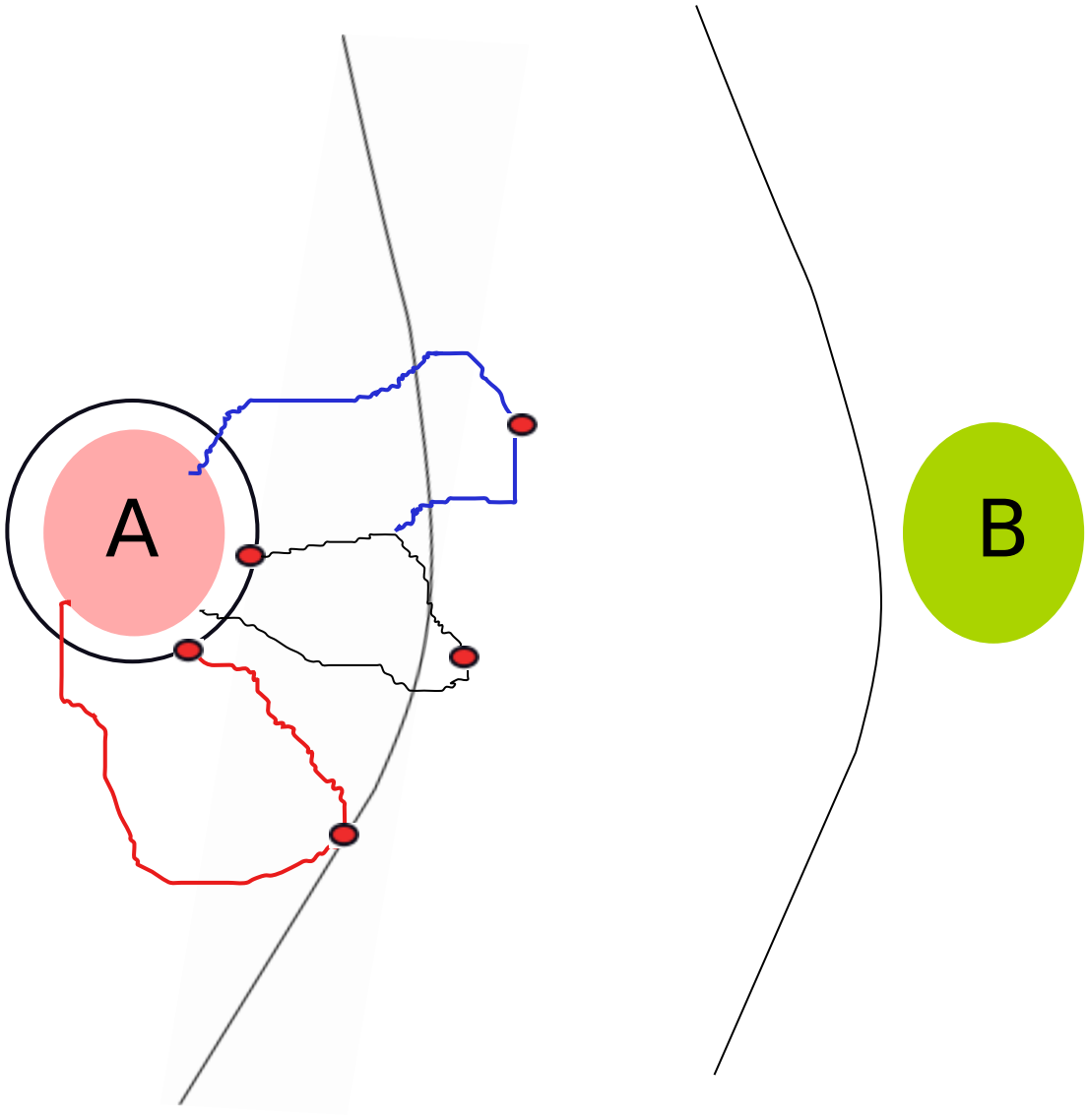
AMS Algorithm



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, 2016]: 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 applied to 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.

Computing transition times

To use the algorithm to compute transition times, we split a transition path from A to B into: excursions from ∂A to $\Sigma_{z_{\min}}$ and then back to ∂A, and finally an excursion from ∂A to $\Sigma_{z_{\min}}$ and then to B. Assuming that A is metastable ($p \ll 1$), it can be shown that the equilibrium mean transition time can be approximated by (see the second part of this talk):

$$\left(\frac{1}{p} - 1\right) \Delta_{\mathrm{Loop}} + \Delta_{React}$$

where:

- p is the probability, starting from $\Sigma_{z_{\min}}$ "at equilibrium", to go to B rather than A (approximated by $\hat{p}$) ;
- Δ_{Loop} is the mean time for an excursion from ∂A to $\Sigma_{z_{\min}}$ and then back to ∂A (approximated by brute force) ;
- Δ_{React} is the mean time for an excursion from ∂A to $\Sigma_{z_{\min}}$ and then to B (approximated by the AMS algorithm).

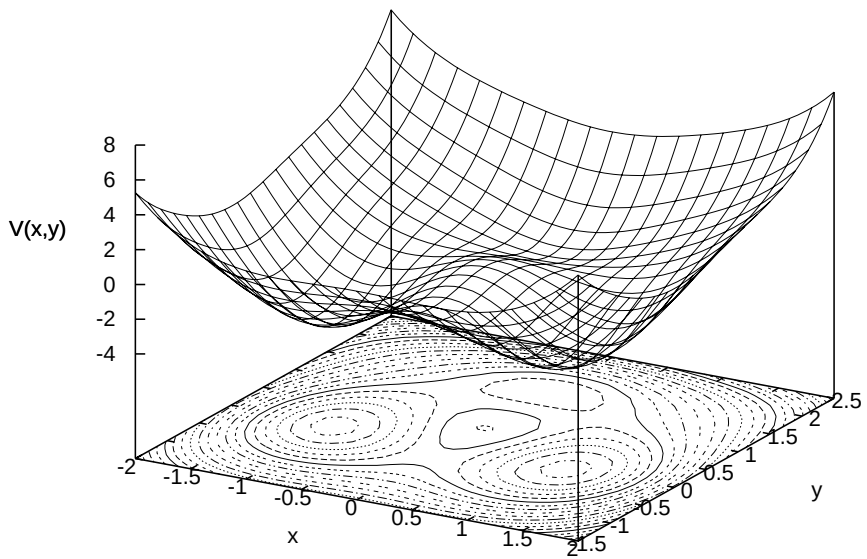
Numerical results: a 2D example

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 $\xi^3(x, y) = x$.

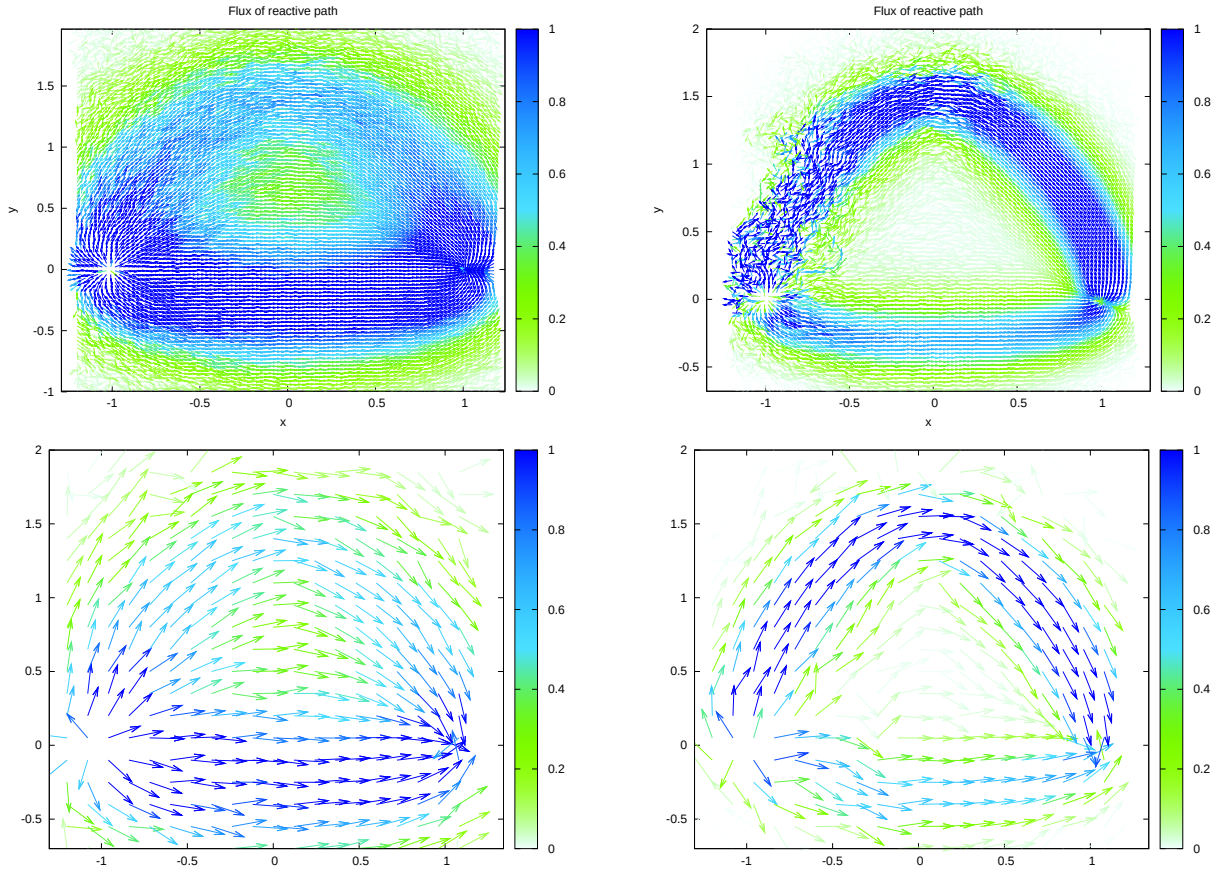
We plot as a function of the number N of independent realizations of AMS, the empirical average

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

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

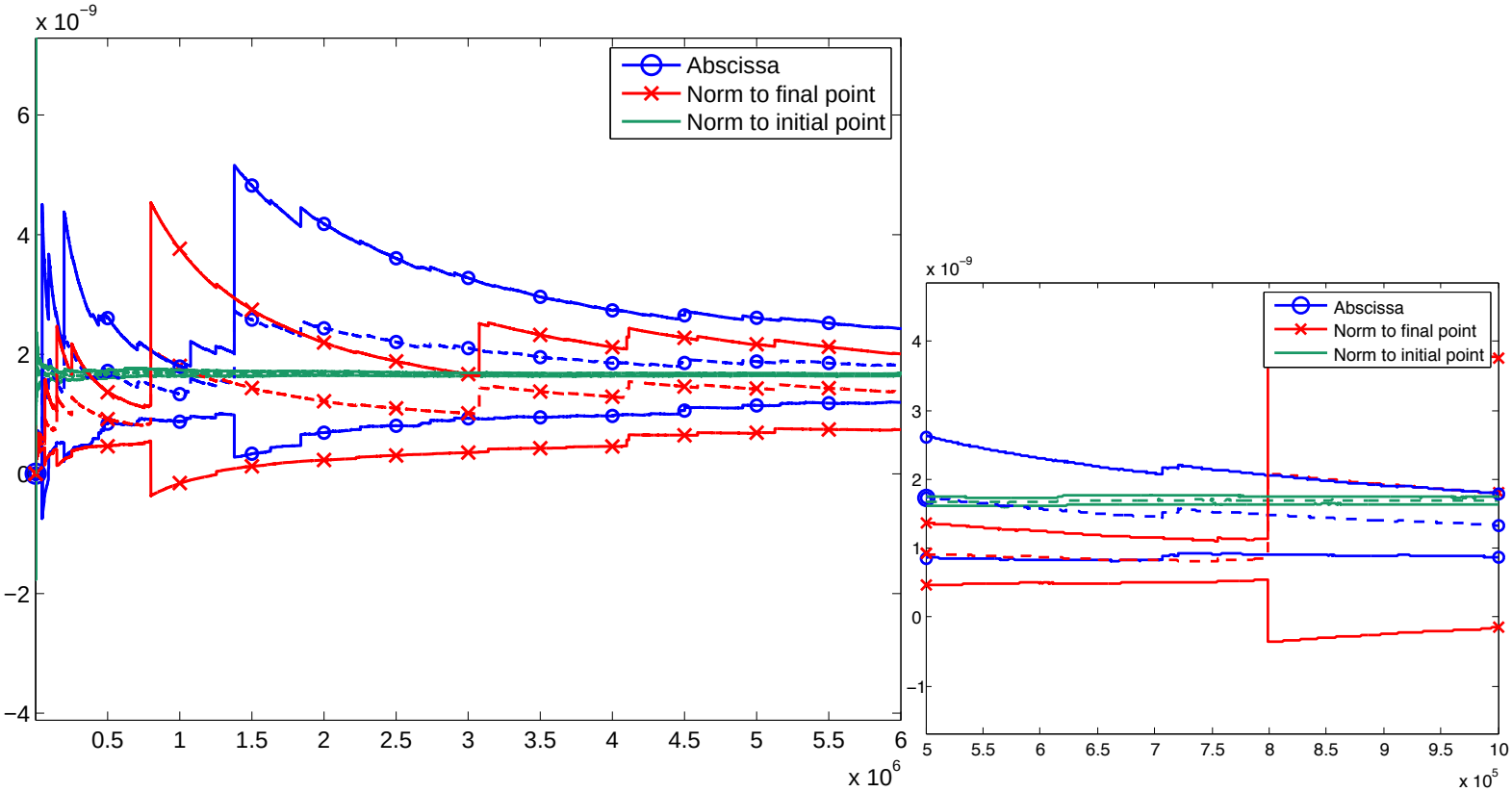
$$\delta_N = 2 \frac{1.96}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{m=1}^{N} (\hat{p}_m)^2 - (\overline{p}_N)^2}$$

A 2D example: flux of reactive trajectories

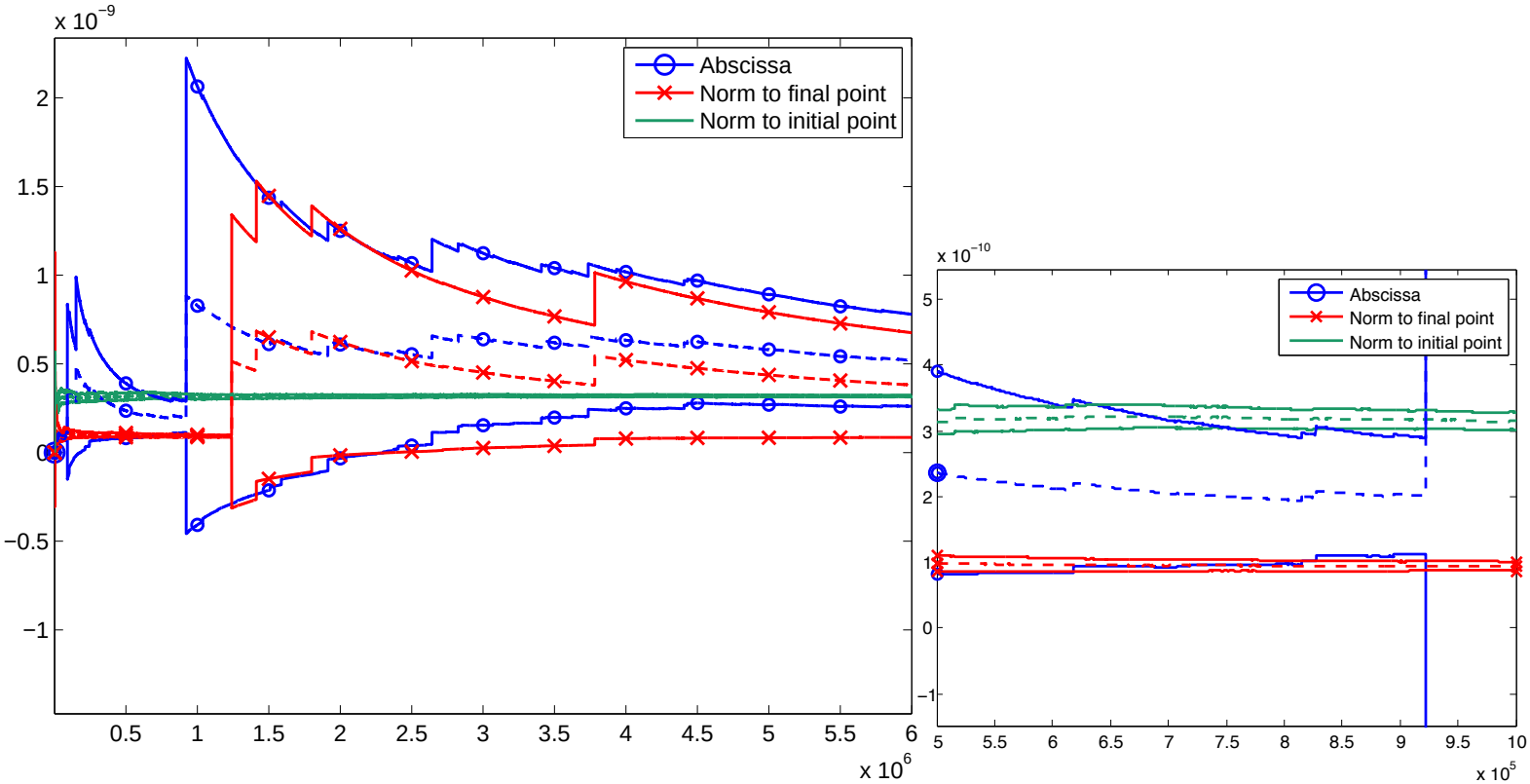


Flux of reactive trajectories, at $\beta = 1.67$ on the left, and $\beta = 6.67$ on the right.

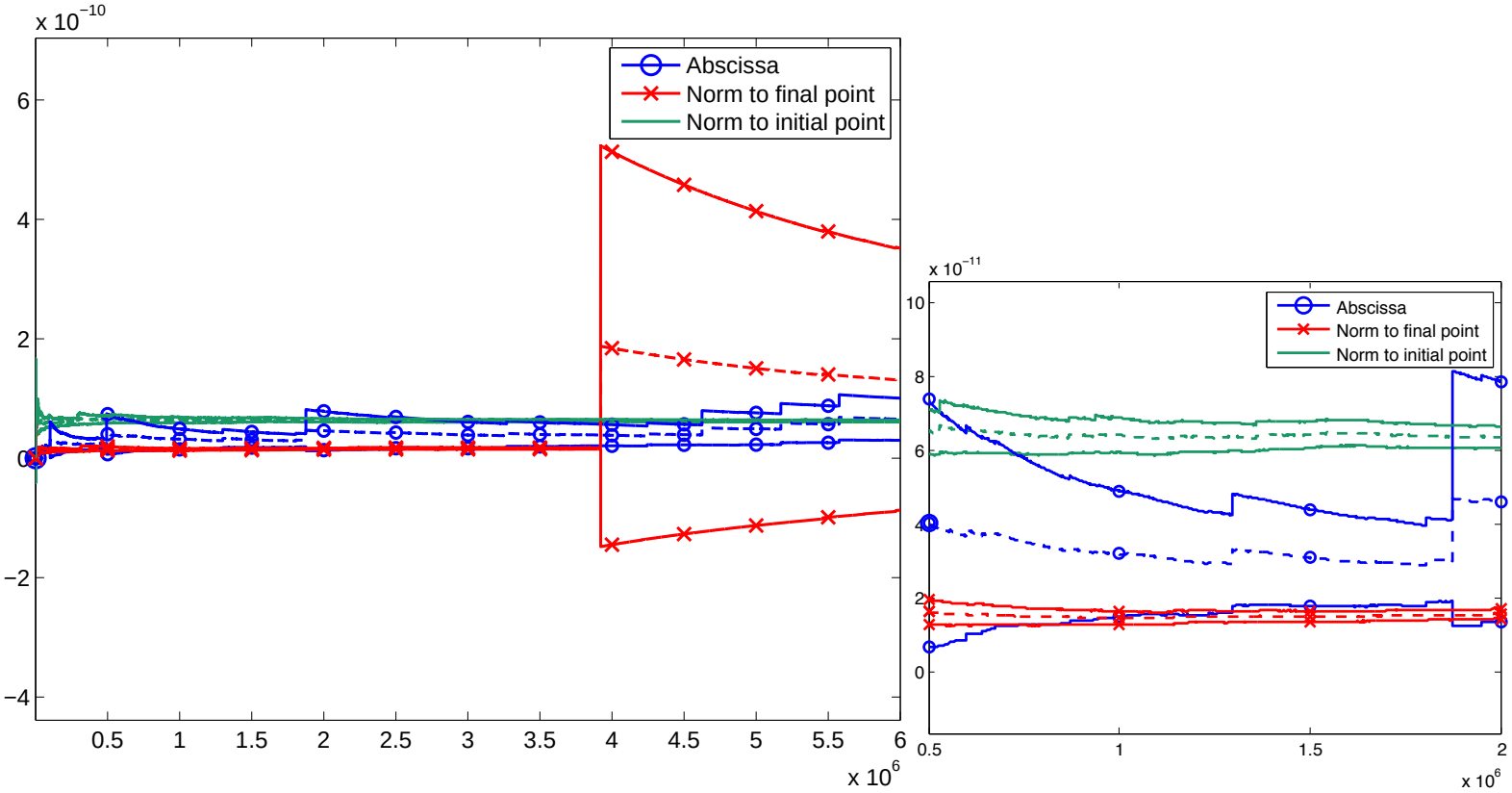
A 2D example: $k = 1$, $n = 100$, $\beta = 8.67$



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.

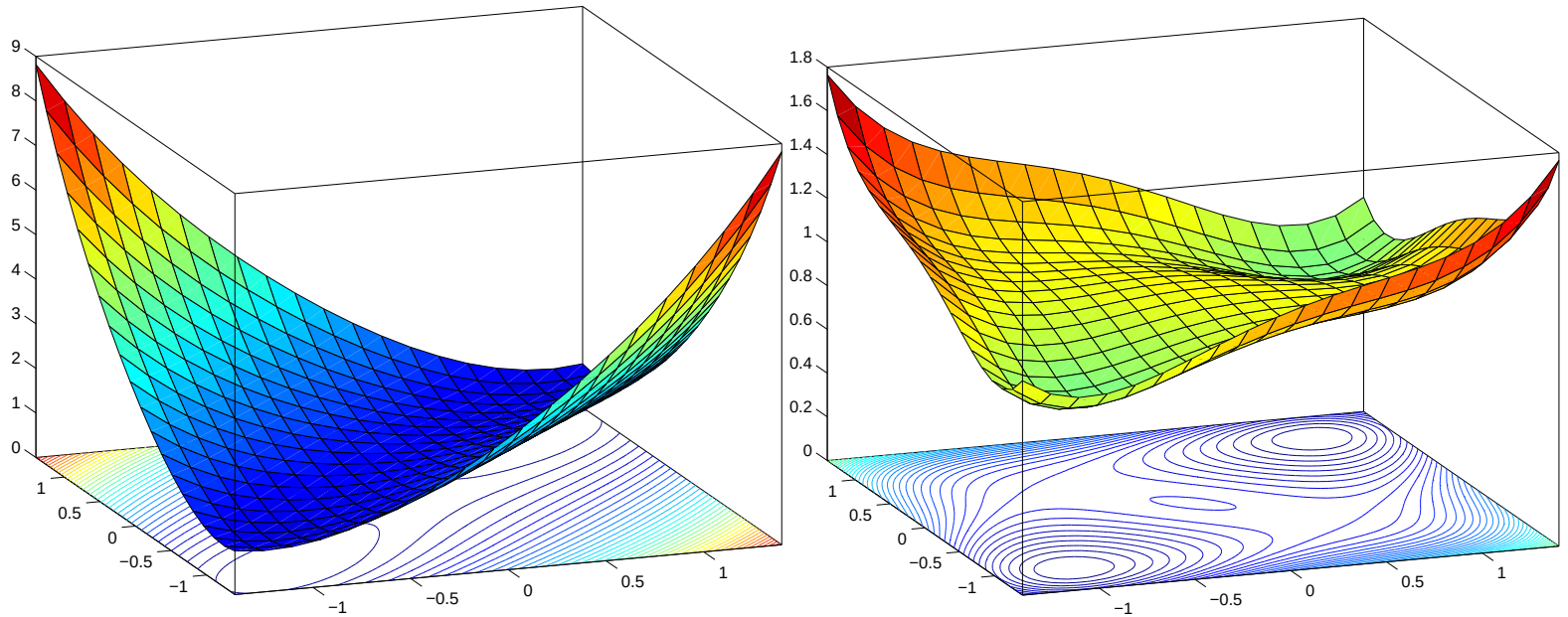
On our example, for small n, we indeed observe that (for ξ^3):

- 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.

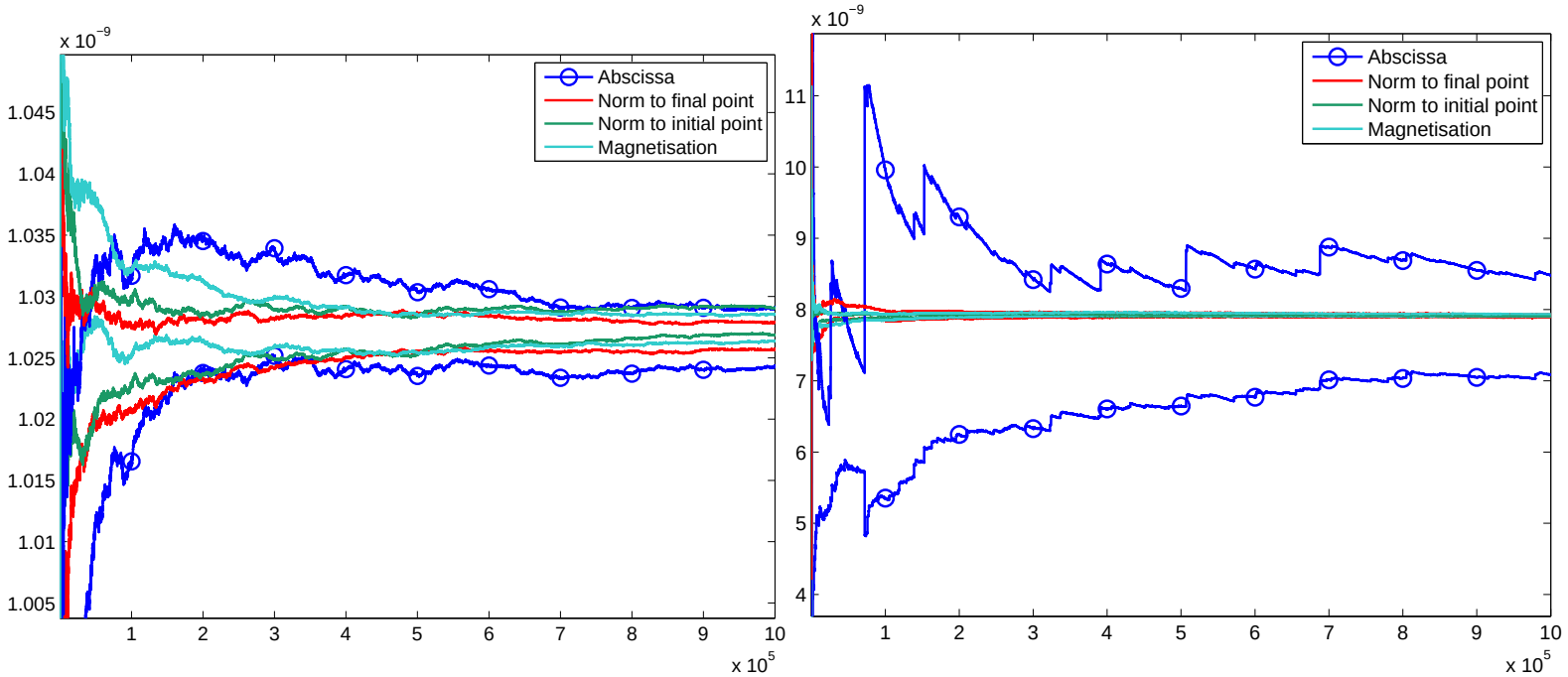
"Apparent bias" phenomenon

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).

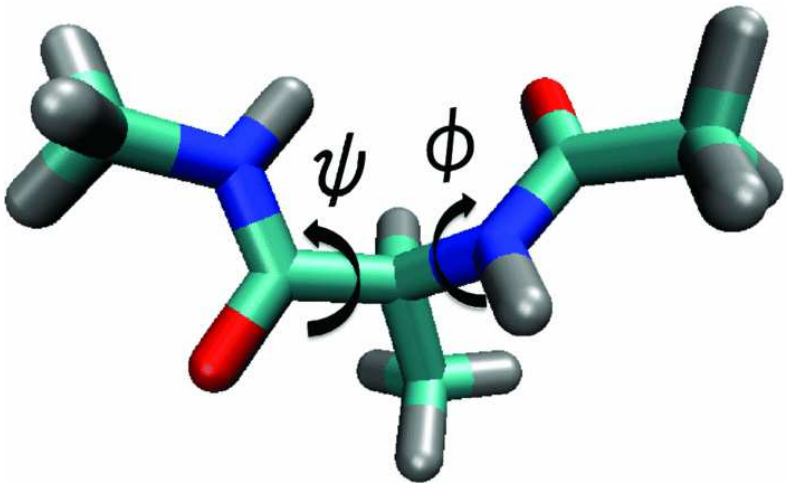
Results on larger test cases

AMS is now implemented in the NAMD software (collaboration with SANOFI, C. Mayne and I. Teo, PhD of L. Silva Lopes with J. Hénin).

Three test cases:

- Alanine di-peptide (test case)
- benzamidine-trypsin dissociation rate
- β-cyclodextrin (in progress)

Alanine di-peptide (1/6)

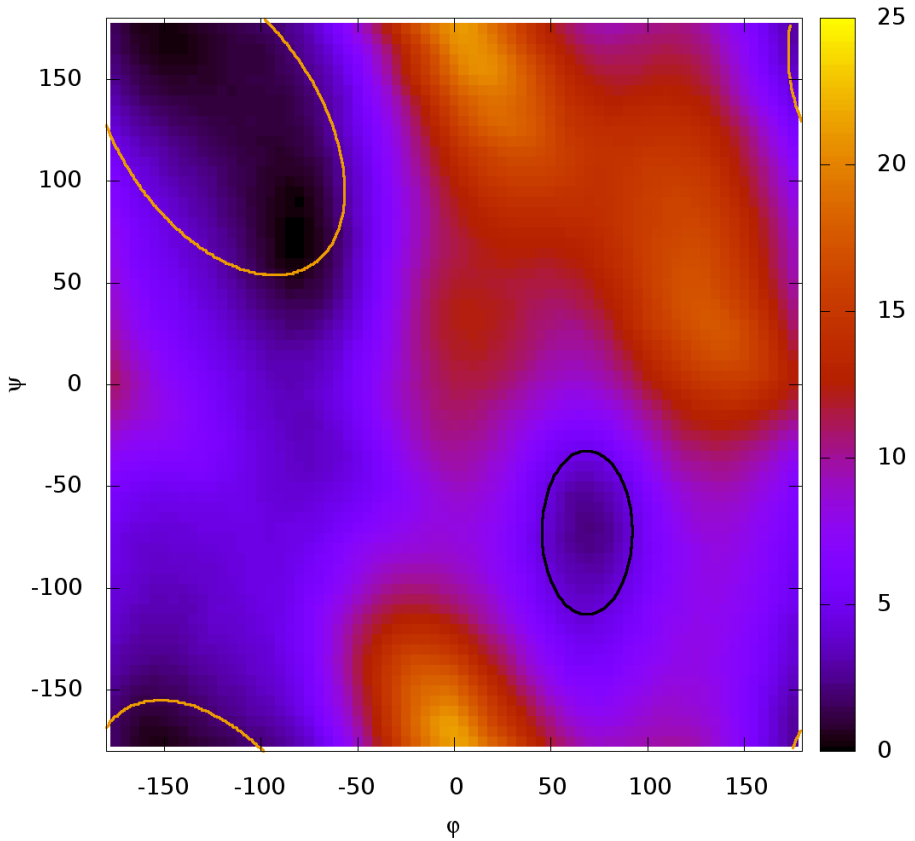


Two reaction coordinates:

- ξ_1 is a continuous piecewise affine function of φ
- $\xi_2(\varphi, \psi) = \min(d_A(\varphi, \psi), 6.4) - \min(d_B(\varphi, \psi), 3.8)$

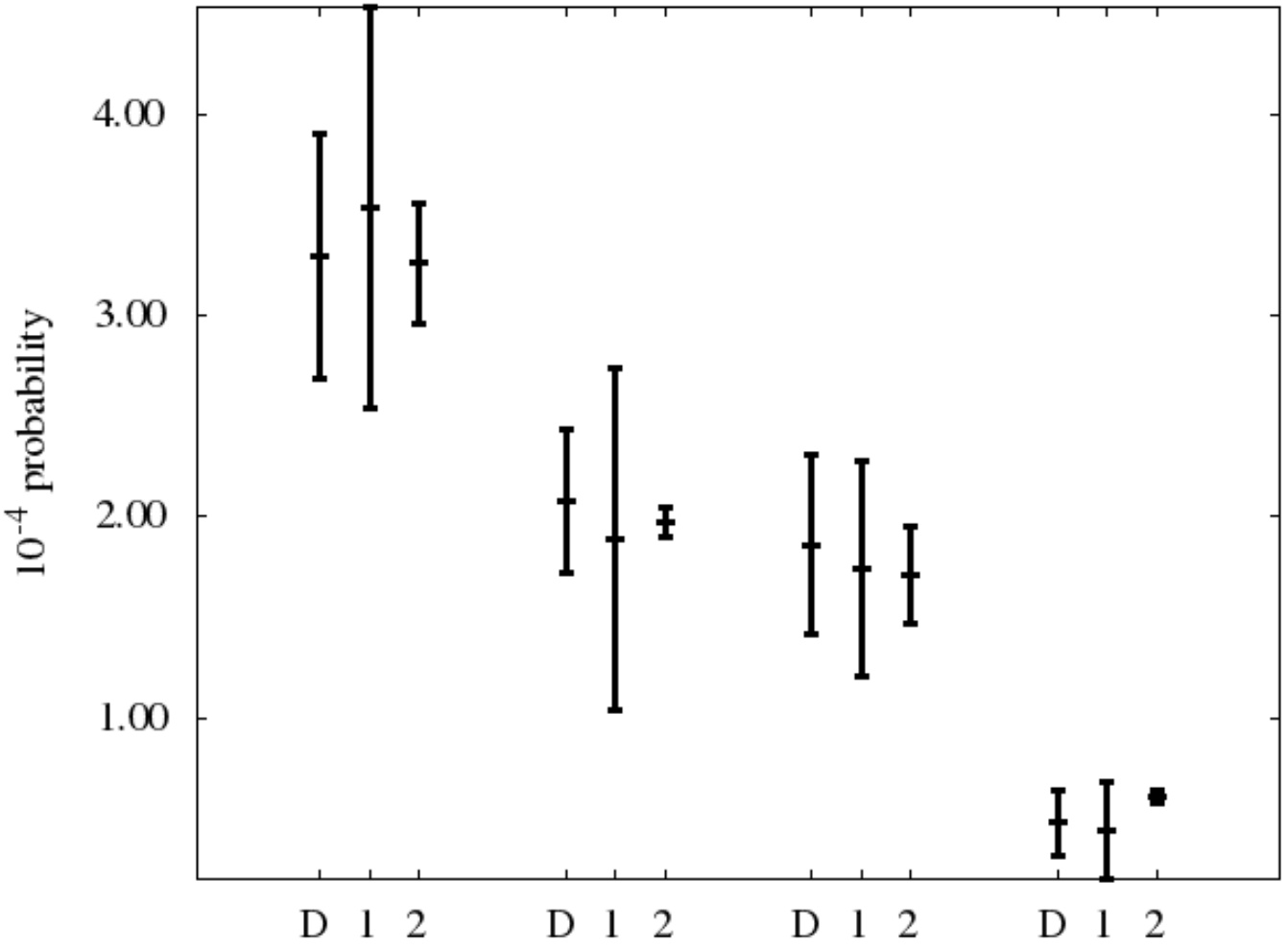
Computational setting: no solvent, force field: CHARMM27. AMS with $n = 500$ to 1000 replicas and $k = 1$.

Alanine di-peptide (2/6)



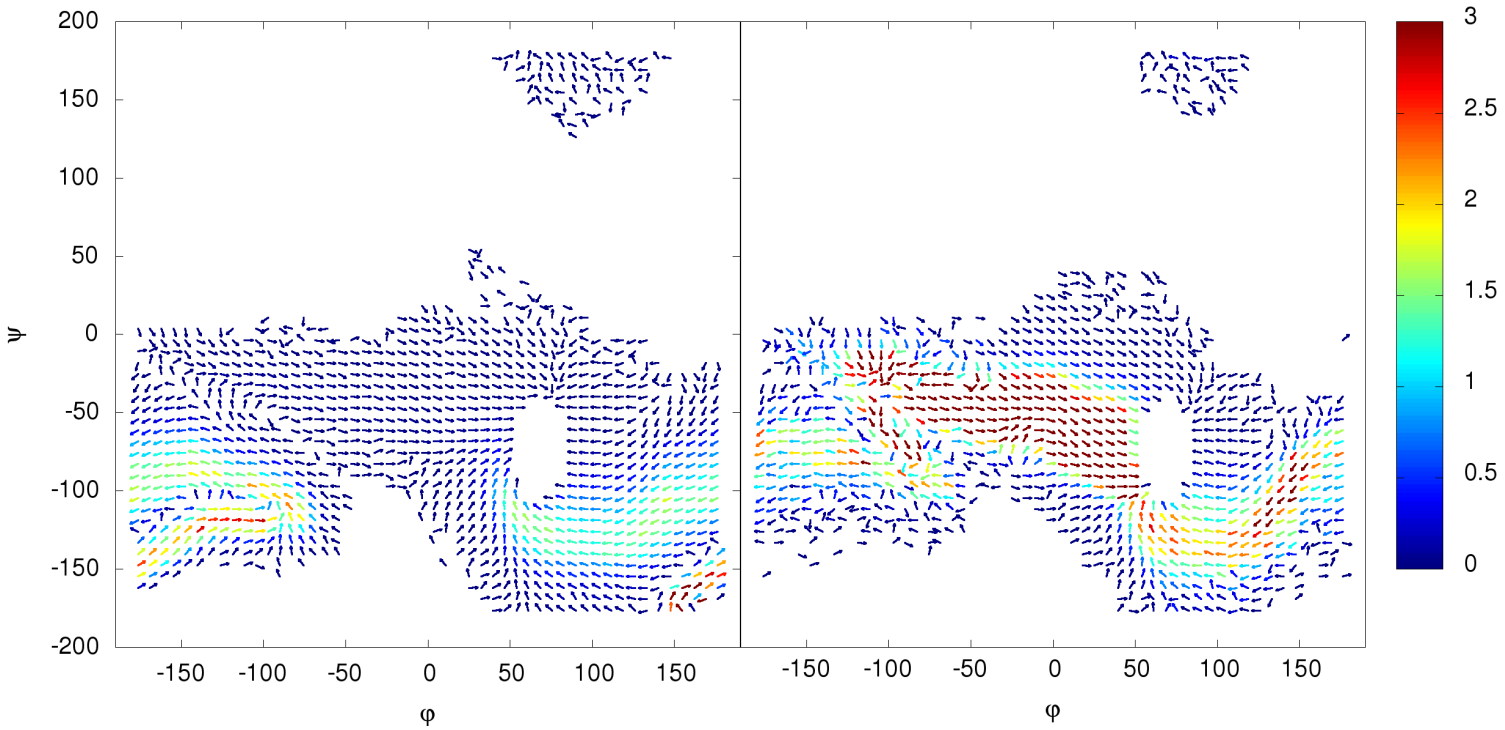
Free energy landscape and zones A (yellow) and B (black).

Alanine di-peptide (3/6)



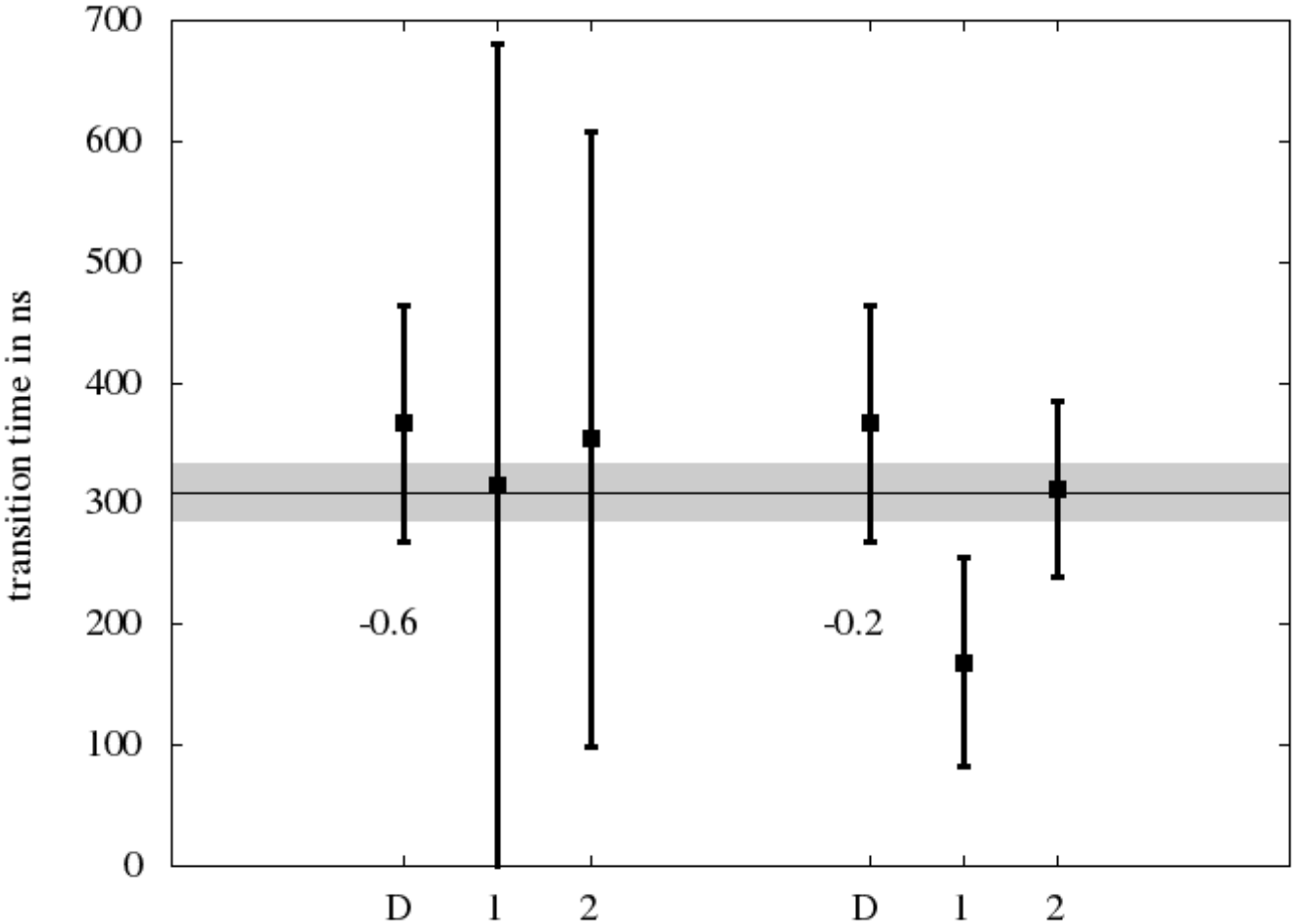
Probability estimations using different initial conditions: D=DNS, 1=ξ_1, 2=ξ_2.

Alanine di-peptide (4/6)



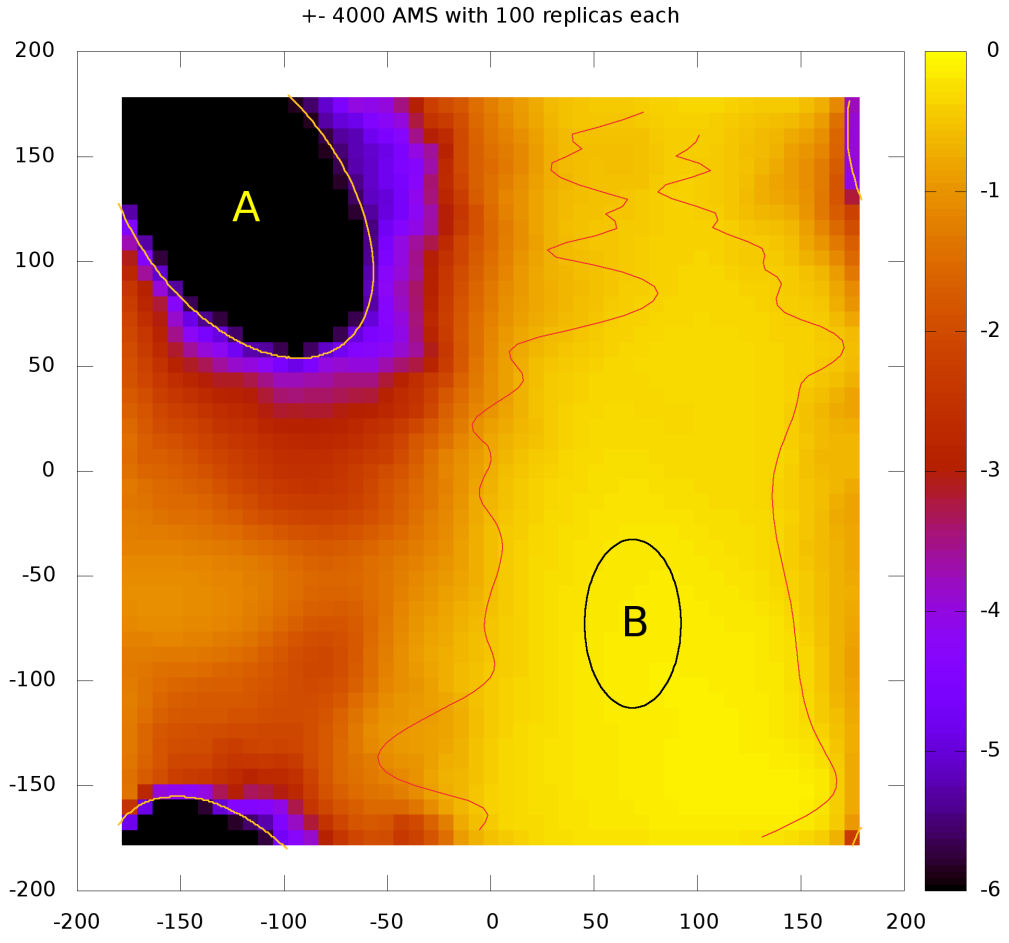
Flux of reactive trajectories, starting from two different initial conditions.

Alanine di-peptide (5/6)



Transition time obtained for two values of z_{min}: D=DNS, 1=ξ_1, 2=ξ_2. Reference value obtained by DNS over a 97 DNS simulations of $2\mu s$.

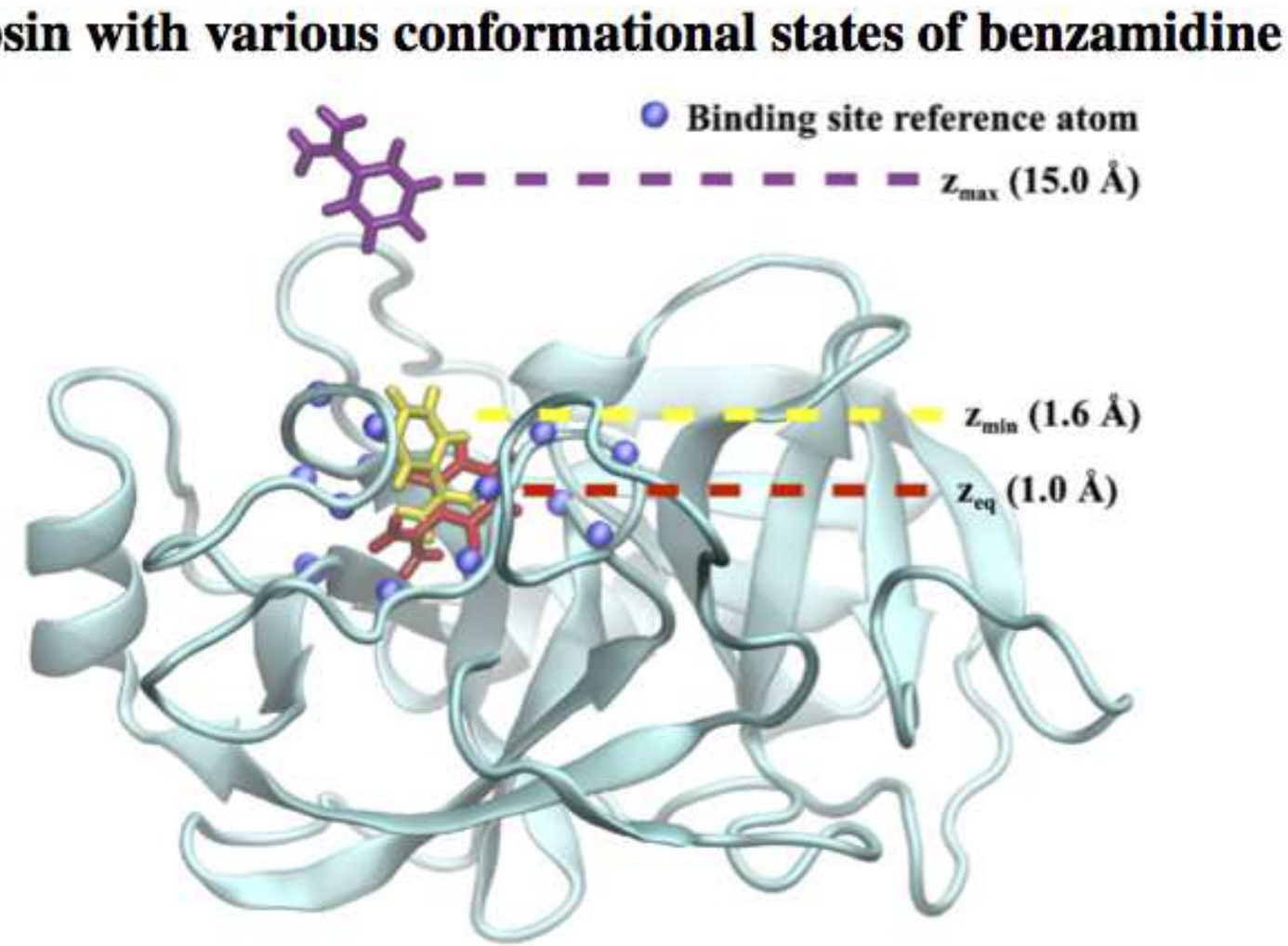
Alanine di-peptide (6/6)



Estimate of the committor function using AMS.

Benzamidine-trypsin (1/2)

We recently used AMS to estimate the off rate of benzamidine from trypsin [I. Teo, C. Mayne, K. Schulten and TL, 2016].



Benzamidine-trypsin (2/2)

We obtain a dissociation rate $k_{\text{off}} = (260 \pm 240) s^{-1}$ within the same order of magnitude as the experimentally measured rate $(600 \pm 300) s^{-1}$.

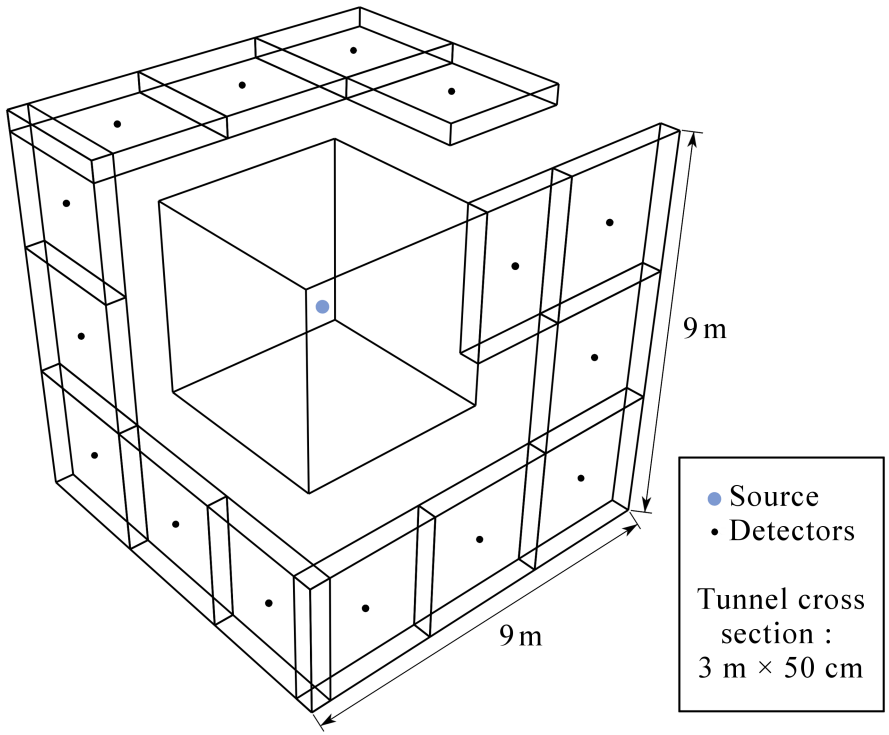
The overall simulation time taken, summed over all 1000 replicas, was $2.1 \mu s$ ($2.3 \mu s$ after including direct MD and steered MD simulations), which is four orders of magnitude shorter than the estimated dissociation time of one event.

The main practical difficulty seems to be the determination of a 'good' domain A.

Computational setting: 68 789 atoms, with 21 800 water molecules, 62 sodium ions, and 68 chloride ions. Water: TIP3P model. CHARMM36 force field, with parameters for benzamidine obtained from the CGenFF force field. NPT conditions, at 298 K and 1 atm Langevin thermostat and barostat settings, using 2 fs time steps. AMS with $n = 1000$ replicas and $k = 1$.

Another example: Radiation protection (1/2)

Monte Carlo particle transport



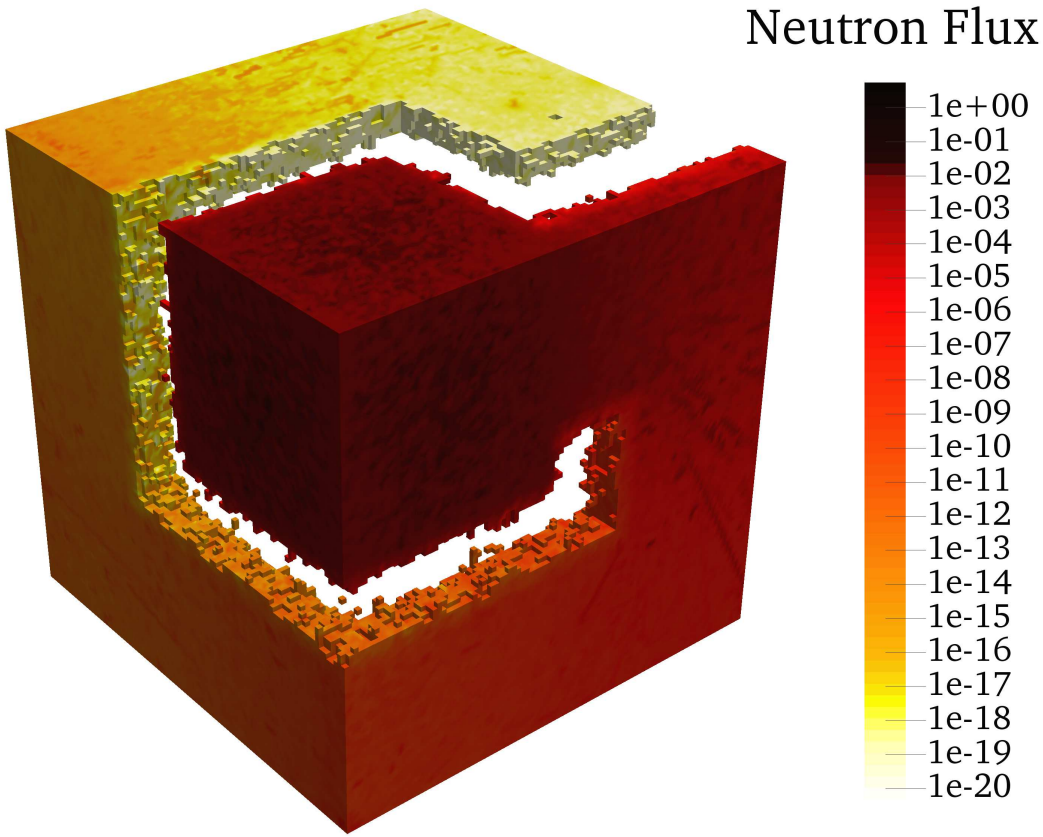
Concrete tunnel with a neutron source

How to compute the neutron flux at the detector ?

Challenge: the flux is very small

Another example: Radiation protection (2/2)

Example 2: In collaboration with CEA (Eric Dumonteil, Cheikh Diop and Henri Louvin), AMS is now implemented in the Tripoli code.



Concluding remarks on AMS (1/2)

Practical recommendations:

- A careful implementation of the splitting step leads to unbiased estimators for non-normalized quantities.
- Perform many independent realizations of AMS.
- Use ξ as a numerical parameter.

The algorithm is very versatile:

- Non-intrusivity: the MD integrator is a black box.
- Can be adapted to generate trajectories of any stopped Markov process.
- Can be applied to both entropic and energetic barriers, to non-equilibrium systems, non-homogeneous Markov process, random fields, ...
- Algorithmic variants: other resampling procedure, additional selection, ...

Concluding remarks on AMS (2/2)

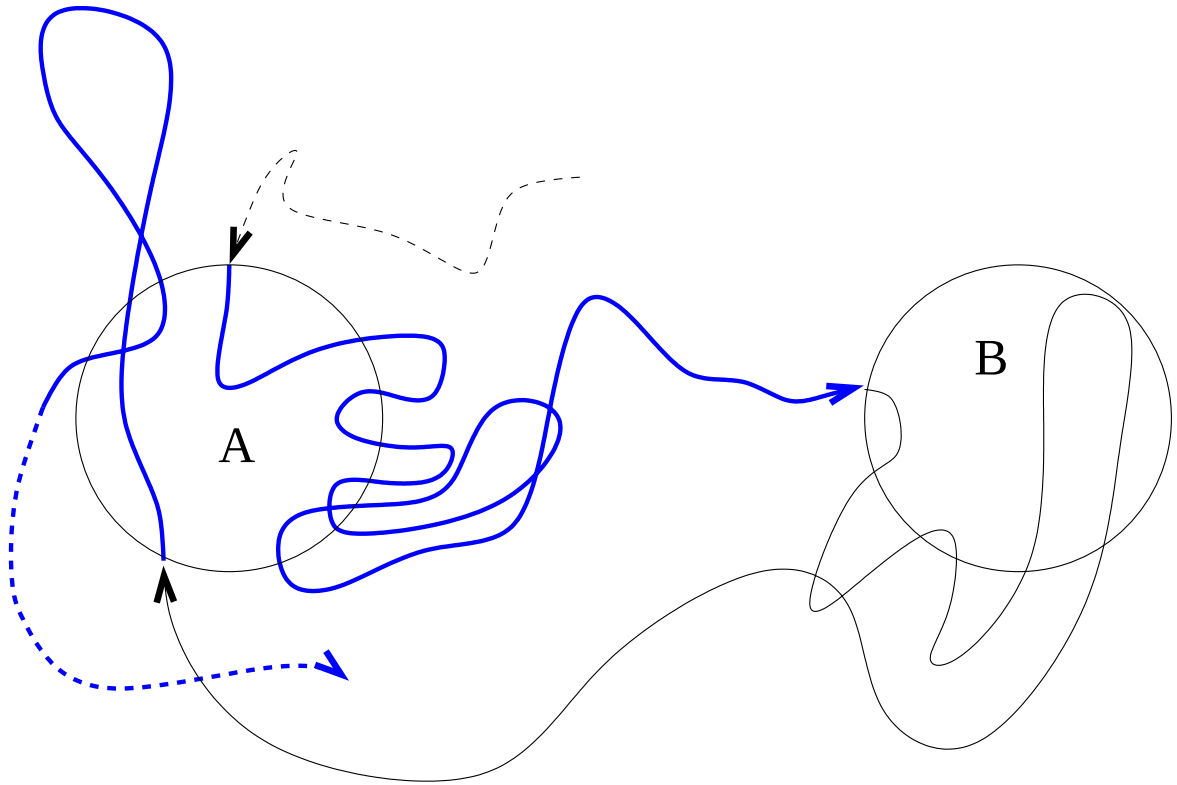
Works in progress:

- Tests on complicated biological systems (collab. with J. Hénin and L. Silva Lopes)
- Adaptive computation of ξ.
- Analysis of the efficiency as a function of ξ. For optimal choice of ξ, the cost of AMS is (for n large)

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

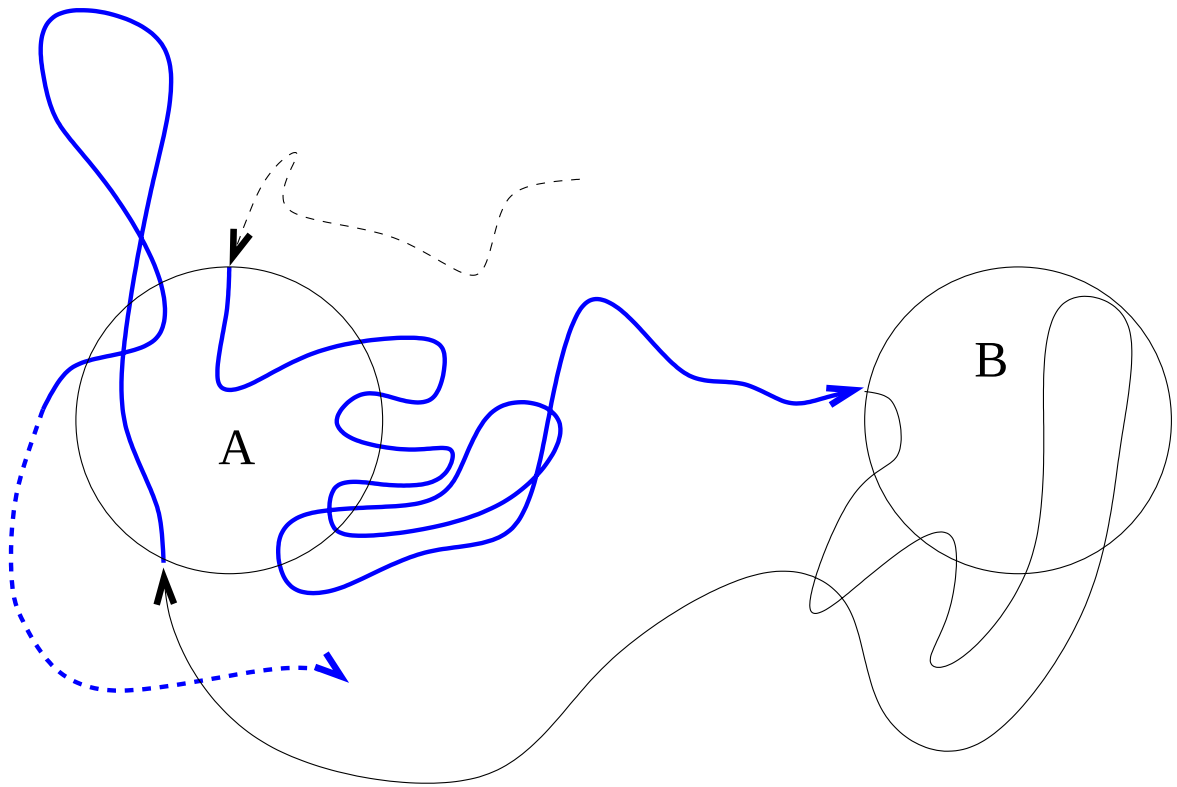
much better than the cost of naive Monte Carlo: $\frac{1-p}{p}$. How does this degrade when ξ departs from the optimal case ?

Computing transition times with AMS



Transition time

Let us consider an ergodic stochastic continuous in time process $(X_t)_{t\geq 0}$ in $\mathbb{R}^d$, and two disjoint subsets $A \subset \mathbb{R}^d$ and $B \subset \mathbb{R}^d$. The objective is to compute the mean transition time at equilibrium from A to B, denoted by $\Delta_{A\to B}$.



Remark: we are also interested in any statistical property of the equilibrium reactive paths from A to B.

Metastability

Examples: Molecular dynamics (A and B are defined in positions space)

- Langevin dynamics (M mass matrix, $\gamma > 0$, $\beta = (k_B T)^{-1}$)

$$\begin{cases} dQ_t = M^{-1} P_t \, dt, \\ dP_t = -\nabla V(Q_t) \, dt - \gamma M^{-1} P_t \, dt + \sqrt{2\gamma\beta^{-1}} dW_t, \end{cases}$$

ergodic wrt $\mu(dq) \otimes Z_p^{-1} \exp\left(-\beta \frac{p^t M^{-1} p}{2}\right) dp$ with

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

where $Z = \int \exp(-\beta V)$.

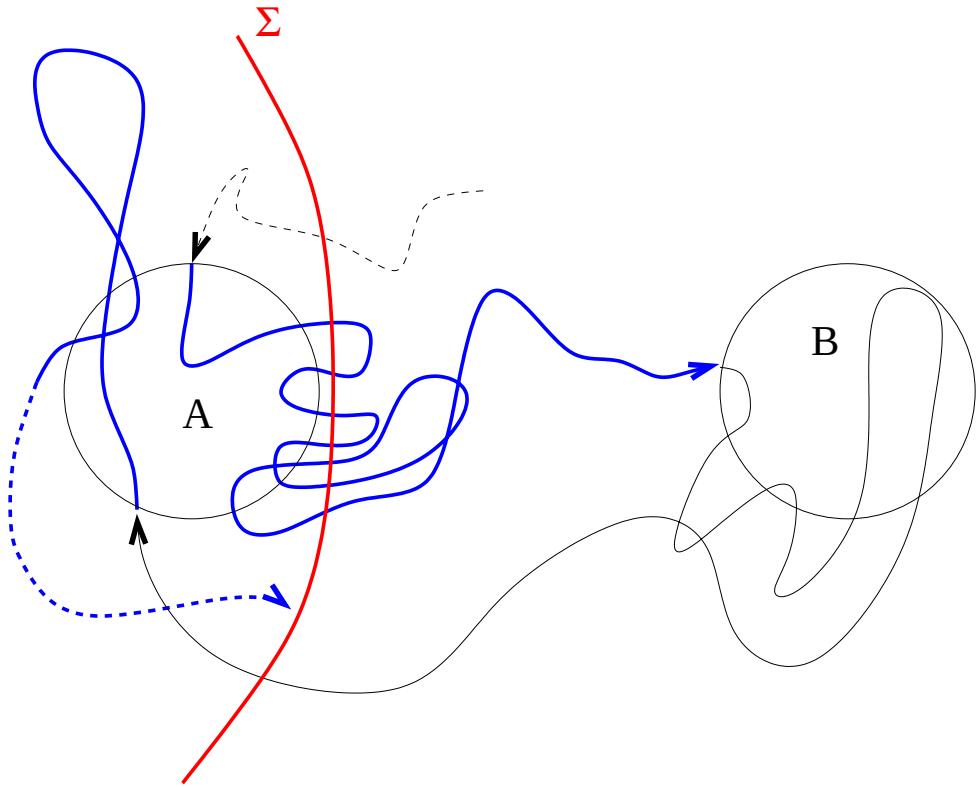
- over-damped Langevin dynamics

$$dX_t = -\nabla V(X_t) \, dt + \sqrt{2\beta^{-1}} dW_t,$$

which is also ergodic wrt μ.

Challenge: A and B are typically metastable states, so that observing transitions from A to B is a rare event.

From continous time to discrete time



Let Σ be a co-dimension 1 submanifold in-between A and B. Then, $(Y_n)_{n\geq 0}$ is the sequence of successive intersections of $(X_t)_{t\geq 0}$ with $\mathcal{A} = \partial A$ or $\mathcal{B} = \partial B$, while hitting Σ in-between.

From continous time to discrete time

More precisely:

$$Y_n = X_{\tau_n}$$

where

$$\tau_n^{\Sigma} = \inf\{t > \tau_{n-1}, X_t \in \Sigma\}$$

$$\tau_n = \inf\{t > \tau_n^{\Sigma},\, X_t \in \mathcal{A} \cup \mathcal{B}\}.$$

The Markov chain $(Y_n)_{n\geq 0}$ is with values in $\mathcal{A} \cup \mathcal{B}$, with kernel:

$\forall x \in \mathcal{A} \cup \mathcal{B}$, $\forall C \subset \mathcal{A} \cup \mathcal{B}$,

$$K(x, C) = \int_{z\in\Sigma} \mathbb{P}^x(X_{\tau_1^{\Sigma}} \in dz)\mathbb{P}^z(X_{\tau_1} \in C)\, dz.$$

Reactive entrance distribution

Let us define the successive entrance times in $\mathcal{A}$ and $\mathcal{B}$ [Lu, Nolen, 2013] [E, Vanden Eijnden, 2006]:

$$T_{k+1}^{\mathcal{A}} = \inf\{n > T_k^{\mathcal{B}},\ Y_n \in \mathcal{A}\}$$

$$T_{k+1}^{\mathcal{B}} = \inf\{n > T_{k+1}^{\mathcal{A}},\ Y_n \in \mathcal{B}\}.$$

The reactive entrance distribution in $\mathcal{A}$ at equilibrium is defined by:

$$\nu_E = \lim_{K\to\infty} \hat{\nu}_{E,K}$$

where

$$\hat{\nu}_{E,K} = \frac{1}{K}\sum_{k=1}^{K} \delta_{Y_{T_k^{\mathcal{A}}}}.$$

Remark: ν_E is independant on the choice of Σ and is also the reactive entrance distribution for the original continuous time process.

Back to the mean transition time

The mean transition time at equilibrium is (strong Markov property):

$$\Delta_{A\to B} = \mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} \Delta(Y_n)\right)$$

where

$$T_{\mathcal{B}} = \inf\{n \geq 0,\ Y_n \in \mathcal{B}\}$$

and for all $x \in \mathcal{A}$,

$$\Delta(x) = \mathbb{E}^x(\tau_1).$$

Remark: Notice that

$$\Delta(x) = \mathbb{E}^x(\tau_1 1_{Y_1\in\mathcal{A}}) + \mathbb{E}^x(\tau_1 1_{Y_1\in\mathcal{B}})$$

is the average time of loop from x back to $\mathcal{A}$ when $Y_1 \in \mathcal{A}$ and the average time of a reactive trajectory from x to $\mathcal{B}$ when $Y_1 \in \mathcal{B}$.

Summary

Objective: Given a discrete-time Markov chain $(Y_n)_{n\geq 0}$ with values in $\mathcal{A} \cup \mathcal{B}$ and a bounded measurable function $f : \mathcal{A} \to \mathbb{R}$, estimate:

$$\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right).$$

Two challenges: The sets $\mathcal{A}$ and $\mathcal{B}$ are metastable, so that (i) $T_{\mathcal{B}}$ is very large, and (ii) ν_E is difficult to sample.

Ideas: For (i), use rare event sampling method (forward flux sampling -FFS- or adaptive multilevel splitting -AMS-). For (ii), use the fact that $\mathcal{A}$ is metastable: the process $(Y_n)_{n\geq 0}$ reaches "equilibrium within $\mathcal{A}$" (quasi stationary distribution) before transitioning to $\mathcal{B}$.

Assumptions and notation

Assumptions: In the following, we assume that the Markov chain $(Y_n)_{n\geq 0}$ satisfies the following hypothesis:

[A1] $(Y_n)_{n\geq 0}$ is weak-Feller meaning that $(Kf) \in \mathcal{C}(\mathcal{A}\cup\mathcal{B},\mathbb{R})$ whenever $f \in \mathcal{C}(\mathcal{A}\cup\mathcal{B},\mathbb{R})$.

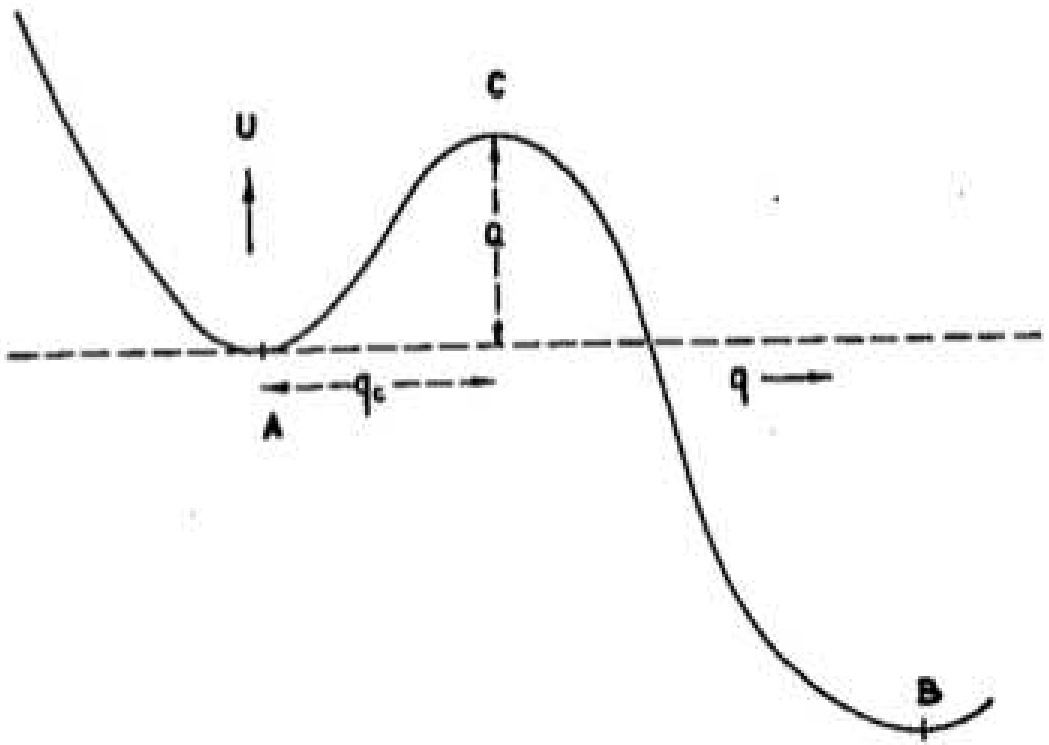
[A2] $(Y_n)_{n\geq 0}$ is positive Harris recurrent, and π_0 denotes its unique stationary probability measure.

[A3] $\pi_0(\mathcal{A}) > 0$ and $\pi_0(\mathcal{B}) > 0$.

All these assumptions are satisfied for the discrete processes built from the Langevin or overdamped Langevin dynamics.

Notation: In the following we use the block-decomposition of the kernel K of the chain $(Y_n)_{n\geq 0}$ over $\mathcal{A}\cup\mathcal{B}$: $K = \begin{bmatrix} K_{\mathcal{A}} & K_{\mathcal{A}\mathcal{B}} \\ K_{\mathcal{B}\mathcal{A}} & K_{\mathcal{B}} \end{bmatrix}$.

The Hill relation



[Kramers, 1940]

The π-return process and the Hill relation

Let π be a probability measure on $\mathcal{A}$. The π-return process $(Y_n^\pi)_{n\geq 0}$ is the Markov chain with values in $\mathcal{A}$ and transition kernel: $\forall x \in \mathcal{A}$, $\forall C \subset \mathcal{A}$,

$$K^\pi(x, C) = \mathbb{P}^x(Y_1 \in C, T_{\mathcal{B}} > 1) + \mathbb{P}^x(Y_1 \in \mathcal{B})\pi(C).$$

In words, $(Y_n^\pi)_{n\geq 0}$ is the chain $(Y_n)_{n\geq 0}$ "reset to π" each time Y_n enters $\mathcal{B}$.

Lemma. $(Y_n^\pi)_{n\geq 0}$ admits a unique stationary distribution, denoted by $R(\pi)$, where

$$R(\pi) = \frac{\pi(\mathrm{Id}_{\mathcal{A}} - K_{\mathcal{A}})^{-1}}{\mathbb{E}^\pi(T_{\mathcal{B}})}.$$

Remark: Such processes are typically used in MD when people introduce a sink in B and a source in A to create a non-equilibrium flux from A to B [Farkas, 1927] [Kramers, 1940], Weighted Ensemble [Zuckerman, Aristoff], Milestoning [Elber, Vanden Eijnden], TIS [Bolhuis, Van Erp].

The π-return process and the Hill relation

We are now in position to state the Hill relation [Hill, 1977] [Aristoff, 2018].

Proposition. For any bounded measurable function $f : \mathcal{A} \to \mathbb{R}$,

$$\mathbb{E}^{\pi}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right) = \frac{R(\pi)f}{\mathbb{P}^{R(\pi)}(Y_1 \in \mathcal{B})}.$$

Remark: If $R(\pi)$ is easy to sample, the RHS is typically easier to compute, since it only involves one step of (Y_n).

Application of the Hill relation to $\pi = \nu_E$

Lemma. The probability measure $R(\nu_E)$ is the stationary distribution π_0 restricted to $\mathcal{A}$:

$$R(\nu_E) = \frac{\pi_0 1_{\mathcal{A}}}{\pi_0(\mathcal{A})} =: \pi_{0|\mathcal{A}}.$$

As a consequence,

$$\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right) = \frac{\pi_{0|\mathcal{A}}(f)}{\mathbb{P}^{\pi_{0|\mathcal{A}}}(Y_1 \in \mathcal{B})}.$$

The Hill relation to compute $\Delta_{A\to B}$

Back to the mean transition time:

$$\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1}\Delta(Y_n)\right) = \Delta_{Loop}(\pi_{0|\mathcal{A}})\left(\frac{1}{P_{React}(\pi_{0|\mathcal{A}})}-1\right)+\Delta_{React}(\pi_{0|\mathcal{A}})$$

where

- $\Delta_{Loop}(\pi_{0|\mathcal{A}}) = \mathbb{E}^{\pi_{0|\mathcal{A}}}(\tau_1|Y_1\in\mathcal{A})$ is the mean time for a loop from $\pi_{0|\mathcal{A}}$ back to $\mathcal{A}$ (computed by brute force Monte Carlo)
- $\Delta_{React}(\pi_{0|\mathcal{A}}) = \mathbb{E}^{\pi_{0|\mathcal{A}}}(\tau_1|Y_1\in\mathcal{B})$ is the mean time of a reactive trajectory from $\pi_{0|\mathcal{A}}$ to $\mathcal{B}$ (computed by FFS/AMS)
- $P_{React}(\pi_{0|\mathcal{A}}) = \mathbb{P}^{\pi_{0|\mathcal{A}}}(Y_1\in\mathcal{B})$ is the probability to get a reactive traj. starting from $\pi_{0|\mathcal{A}}$ (computed by FFS/AMS)

The difficulty is that π_0 and, a fortiori, $\pi_{0|\mathcal{A}}$ are in general unknown and difficult to sample.

Summary

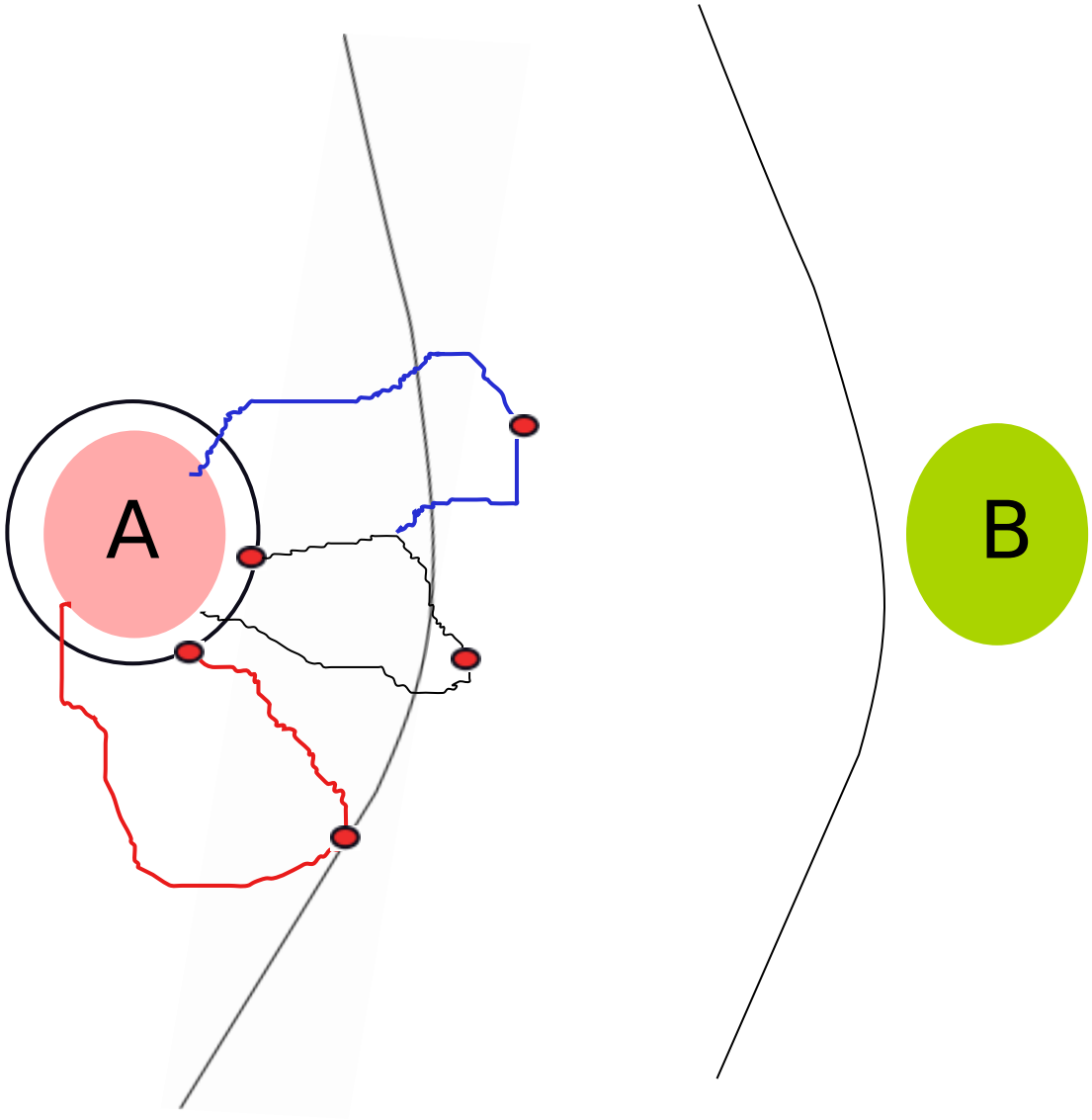
The formula

$$\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right) = \frac{\pi_{0|\mathcal{A}}(f)}{\mathbb{P}^{\pi_{0|\mathcal{A}}}(Y_1 \in \mathcal{B})}$$

is not practical since $\pi_{0|\mathcal{A}}$ is difficult to sample.

Hope: since $\mathcal{A}$ is metastable, maybe it is not needed to sample ν_E or $\pi_{0|\mathcal{A}}$ since, typically, the process will reach a local equilibrium within $\mathcal{A}$ before going to $\mathcal{B}$.

A practical algorithm



The quasi-stationary distribution (QSD)

Lemma. Under the assumptions above, the process $(Y_n)_{n\geq 0}$ admits a quasi-stationary distribution (QSD) ν_Q in $\mathcal{A}$, namely a probability measure ν_Q over $\mathcal{A}$ such that: $\forall C \subset A$,

$$\nu_Q(C) = \mathbb{P}^{\nu_Q}(Y_1 \in C | T_{\mathcal{B}} > 1).$$

In the following, we assume that

[B] $(Y_n)_{n\geq 0}$ admits a unique quasi-stationary distribution ν_Q.

Properties of the QSD:

- For any intial condition $x \in \mathcal{A}$, for any $C \subset A$,

$$\lim_{n\to\infty} \mathbb{P}^x(Y_n \in C | n < T_{\mathcal{B}}) = \nu_Q(C).$$

- The ν_Q-return process admits ν_Q as an invariant distribution:

$$R(\nu_Q) = \nu_Q.$$

The Hill relation applied to $\pi = \nu_Q$

As a consequence

$$\mathbb{E}^{\nu_Q}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right) = \frac{\nu_Q(f)}{\mathbb{P}^{\nu_Q}(Y_1 \in \mathcal{B})}.$$

Remark: Starting from ν_Q, $T_{\mathcal{B}}$ is geometrically distributed, with parameter $\mathbb{P}^{\nu_Q}(Y_1 \in \mathcal{B}) = P_{React}(\nu_Q)$.

Back to the mean transition time [Cérou, Guyader, TL, Pommier, 2011]:

$$\mathbb{E}^{\nu_Q}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} \Delta(Y_n)\right) = \Delta_{Loop}(\nu_Q)\left(\frac{1}{P_{React}(\nu_Q)} - 1\right) + \Delta_{React}(\nu_Q)$$

What did we gain, compared to $\pi = \nu_E$? The probability distribution ν_Q can be sampled by brute force Monte Carlo.

The algorithm to compute $\Delta_{A\to B}$

In practice:

- Simulate the process $(X_t)_{t\geq 0}$ (or $(Q_t, P_t)_{t\geq 0}$ in a neighborhood of A, registering the successive loops from A to Σ. This gives samples distributed according to ν_Q, and $\Delta_{Loop}(\nu_Q)$.
- Use AMS to simulate reactive trajectories, starting from the QSD ν_Q. This gives an estimate of $P_{React}(\nu_Q)$.

Remark: Typically, one has $P_{React}(\nu_Q) \ll 1$ and $\Delta_{React}(\nu_Q) \ll \frac{\Delta_{Loop}(\nu_Q)}{P_{React}(\nu_Q)}$ so that

$$\mathbb{E}^{\nu_Q}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} \Delta(Y_n)\right) \simeq \frac{\Delta_{Loop}(\nu_Q)}{P_{React}(\nu_Q)}.$$

This is the formula used in FFS to compute transition times [Allen, Valeriani, ten Wolde, 2009].

Error analysis

$$\left| \frac{\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right) - \mathbb{E}^{\nu_Q}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right)}{\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right)} \right| \ll 1?$$

Error analysis

In practice, we thus compute $\mathbb{E}^{\nu_Q}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right)$ instead of the truth $\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right)$.

Objective: Quantify the relative error

$$ERR = \left| \frac{\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right) - \mathbb{E}^{\nu_Q}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right)}{\mathbb{E}^{\nu_E}\left(\sum_{n=0}^{T_{\mathcal{B}}-1} f(Y_n)\right)} \right|.$$

as a function of how large is the transition time wrt the convergence time to the QSD.

Transition time

The time to observe a transition to $\mathcal{B}$ is measured by

$$\frac{1}{p^+}$$

where $p^+ = \sup_{x \in \mathcal{A}} \mathbb{P}^x(Y_1 \in \mathcal{B})$.

Remark: One obviously has, for any $x \in \mathcal{A}$,

$$\frac{1}{p^+} \leq \mathbb{E}^x(T_{\mathcal{B}}).$$

Convergence time to the QSD

The convergence time to the QSD is measured by:

$$T_Q^E = \|\nu_E H_Q(x,\cdot)\|_{TV}$$

where

$$H_Q(x,\cdot) = \sum_{n=0}^{\infty} \left((K^{\nu_Q})^n(x,\cdot) - \nu_Q\right).$$

Why can T_H be seen as a convergence time to the QSD?
One has

$$T_Q^E \leq \sum_{n=0}^{\infty} \|\mathcal{L}^{\nu_E}(Y_n | T_{\mathcal{B}} > n) - \nu_Q\|.$$

Example: the geometrically ergodic case

In the context of the over-damped Langevin dynamics, one can show that: $\exists \alpha > 0, \exists \rho \in (0,1), \forall n \geq 0$,

$$\|\mathcal{L}^{\nu_E}(Y_n | T_{\mathcal{B}} > n) - \nu_Q\|_{TV} \leq \alpha \rho^n.$$

In this case,

$$T_Q^E \leq \frac{\alpha}{1-\rho}.$$

Error analysis

Proposition. Assume that $p^+ T_Q^E < 1$. Then,

$$ERR \leq \frac{p^+ T_Q^E}{1 - p^+ T_Q^E} \left(1 + \frac{\|f\|_\infty}{|\pi_{0|\mathcal{A}}(f)|} \right).$$

This shows that the error is small if the transition time is large compared to the convergence time to the QSD, i.e.

$$\frac{1}{p^+} \gg T_Q^E.$$

Remark: We have checked on examples that the upper bound is sharp in various ways. In particular, one cannot replace p^+ by $P_{React}(\nu_Q)$ neither by $P_{React}(\nu_E)$ in the RHS.

Conclusion (1/2)

We now have a good understanding of the formula which is used by many algorithms (FFS, AMS and the "source and sink methods": TIS, WE, milestoning) to compute the mean transition time:

- These methods are exact if the process is initialized in the initial state with the correct distribution: the reactive entrance distribution
- The reactive entrance distribution can be replaced by the QSD if A is metastable.

Conclusion (2/2)

Current research directions:

- We analyzed the bias, and not the variance or the efficiency of the whole procedure. This should be possible, at least in simple prototypical cases, and maybe give some hints on good choices for some numerical parameters (position of Σ).
- In practice, it is observed that the initial conditions that indeed yield a transition to B are concentrated on some parts of the boundary ∂A. We are currently working on good sampling methods for these initial conditions [collab. Laura Lopes].

The practical problem is not to replace ν_E by ν_Q, but to sample ν_Q correctly.

References

- M. Baudel, A. Guyader, and TL, *On the Hill relation and the mean reaction time for metastable processes*, `https://arxiv.org/abs/2008.09790` .
- C.-E. Bréhier, M. Gazeau, L. Goudenège , TL, and M. Rousset, *Unbiasedness of some generalized Adaptive Multilevel Splitting algorithms*, Annals of Applied Probability, 2016.
- F. Cérou, A. Guyader, TL, and D. Pommier, *A multiple replica approach to simulate reactive trajectories*, J. Chem. Phys. 2011.
- C.M. Diop, E. Dumonteil, TL, H. Louvin, and M. Rousset, *Adaptive Multilevel Splitting for Monte Carlo particle transport*, Eur. Phys. J. N, 2017.
- TL and L.J.S. Lopes, *Analysis of the Adaptive Multilevel Splitting method with the alanine di-peptide's isomerization*, J. Comput. Chem., 2019.
- I. Teo, C. Mayne, K. Schulten, and TL, *Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time*, J. Chem. Theory Comput., 2016.