Importance splitting for rare event simulation

F. Cerou
Frederic.Cerou@inria.fr

Inria Rennes Bretagne Atlantique

Simulation of hybrid dynamical systems and applications to molecular dynamics
September 27-30 2010, IHP

Rare event

Low but non-zero probability, say $0 < \mathbb{P}(X \in A) \leq 10^{-8}$.
It is very important to know this probability, or to generate typical realizations.
We assume that we know how to draw (pseudo) realizations of X.
Different from extreme values: here we have a model that we can simulate, extreme values refer to some dataset used with purely statistical approaches.

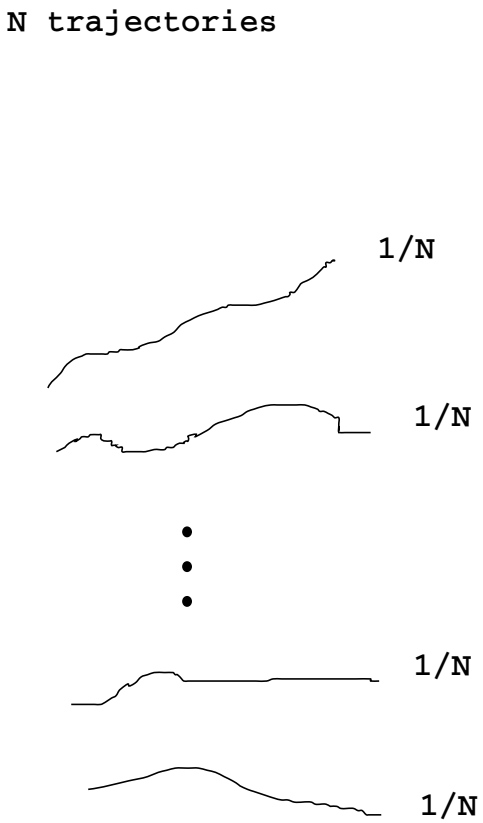
Why naive Monte-Carlo does not work ? Let $p = \mathbb{P}(X \in A)$.
$\hat{p} = \sum_{i=1}^{N} \mathbb{1}_A(X_i)$, and the relative mean square error is $\frac{\text{var}(\hat{p})}{p^2} = \frac{1-p}{Np}$. So we need $N \simeq 10/p$ at least...

Two main frameworks: importance sampling vs. importance splitting.
Splitting can work with a simulation blackbox, and gives rare trajectories drawn with the original dynamics.

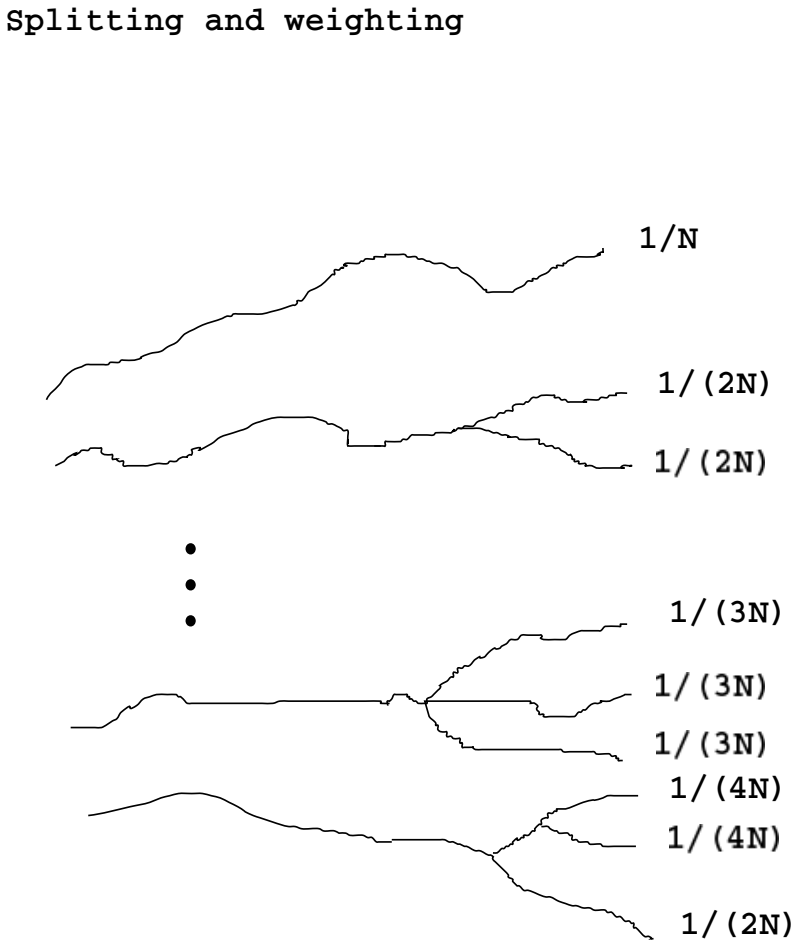
Some applications

- Particle transmission [KH51]
- Queueing networks
- Air traffic management
- Satellite versus debris collision
- Finance
- Food contaminant exposure
- Molecular dynamics

Naive Monte-Carlo



Trajectory splitting



Trajectory splitting

Splitting (time and number of offsprings) may depend on the past of the current trajectory.
With the right weighting, the splitting MC estimates are unbiased.
But the goal is to lower the variance, by using a "clever" way of splitting.
Consider splitting those trajectories that goes "towards" the rare event.

Precise setting

X continuous time stochastic process in $\mathbb{R}^d$, with initial distribution μ_0. X has almost surely right continuous, left limited trajectories (RCLL).

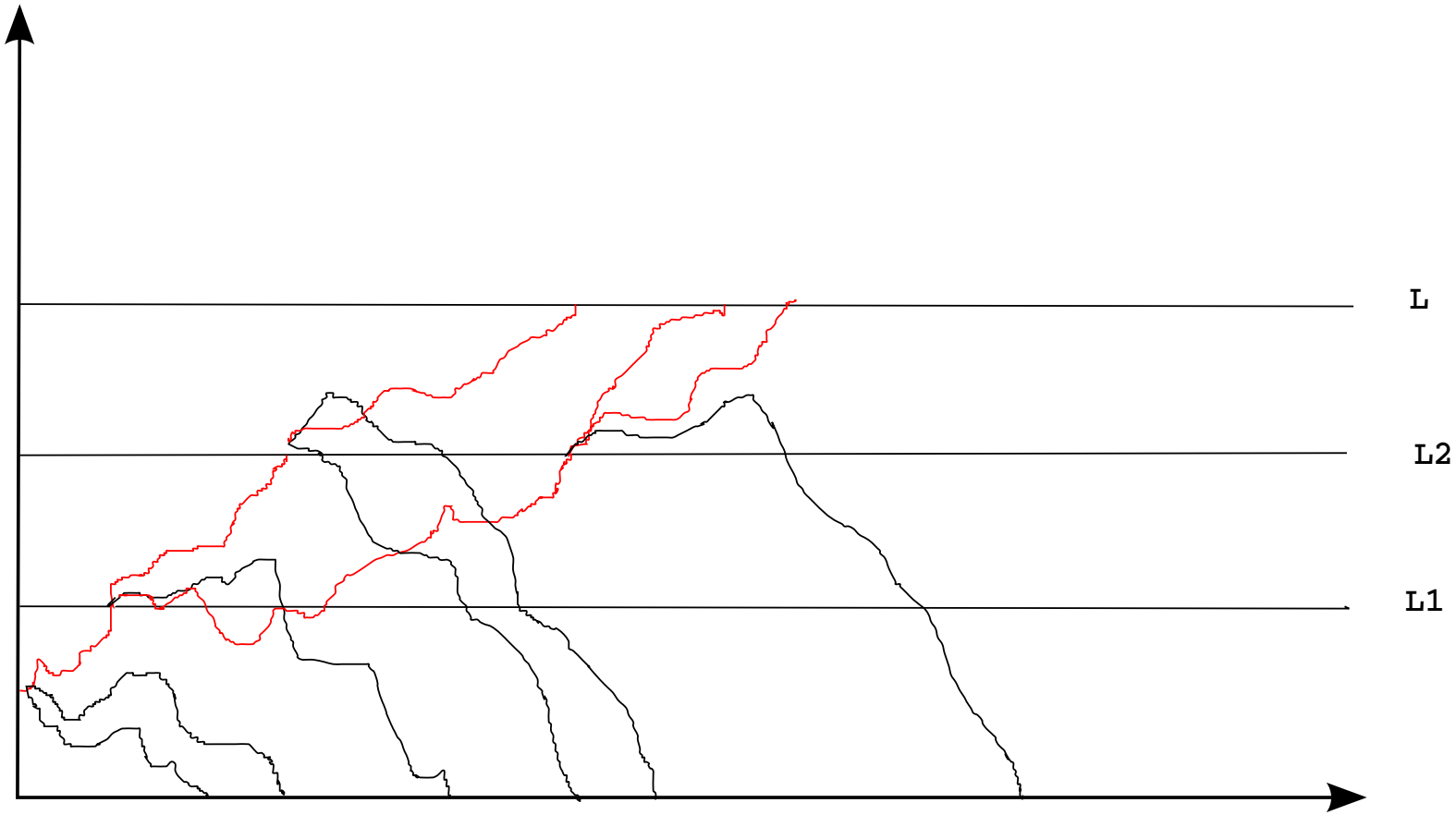
A_0 is a recurrent event for X, i.e. $\mathbb{P}(X \text{ returns to } A_0 \text{ i.o. }) = 1$.

Let $\tau_A = \inf\{t > 0,\ X_t \in A\}$ and $\tau_{A_0} = \inf\{t > 0,\ X_t \in A_0\}$. Consider the rare event $A = \mathbb{P}(\tau_A < \tau_{A_0})$.

In many cases, A has the following form: $A = \{x \in \mathbb{R}^d,\ \phi(x) \geq L\}$ for some scalar continuous function ϕ, and some real L.

Algorithm: fixed splitting

Parameters: Splitting levels $L_1 < L_2 < \cdots < L_k < \cdots < L$ for a chosen importance function ϕ, and mean offspring numbers R. We estimate the total probability by $\hat{p}_{n,\mathrm{FS}} = \frac{\#\text{particles reaching } L}{NR^{n-1}}$.



Efficiency

How to measure efficiency ?
Let $\tau_k = \inf\{t > 0,\ \phi(X_t) \geq L_k\}$ and $p_k = \mathbb{P}(\tau_k < \tau_{A_0})$. Assume $\lim_{k\to+\infty} p_k = 0$.
Work normalized variance $\mathrm{var}(\hat{p}_k)w_k$.
Naive Monte-Carlo $\mathrm{var}(\hat{p}_{k,\mathrm{MC}})w_k = \mathrm{var}(\hat{p}_{k,\mathrm{MC}})N = p_k(1-p_k)$

Definition

We define the asymptotic efficiency [GW92] of an estimate $\hat{p}_k$ by

$$\lim_{k\to+\infty} \frac{\log(\mathrm{var}(\hat{p}_k)w_k)}{\log p_k} = 2$$

By Jensen inequality, we always have ≤ 2 for unbiased estimators.
For Naive Monte-Carlo, we get
$\frac{\log(\mathrm{var}(\hat{p}_{k,\mathrm{MC}})w_k)}{\log p_k} = \frac{\log(p_k(1-p_k))}{\log(p_k)} = 1 + \frac{\log(1-p_k)}{\log(p_k)} \to 1.$

Efficiency: fixed splitting

Assume $\lim_{k\to+\infty} \frac{\log p_k}{k} = \log \rho$ exists.
From [GHSZ99], we have

Theorem

Under some mixing conditions,

$$\lim_{k\to+\infty} \frac{\log(\mathrm{var}(\hat{p}_{k,\mathrm{FS}})w_{k,\mathrm{FS}})}{\log p_k} = 2 \text{ for } R = 1/\rho,$$

and

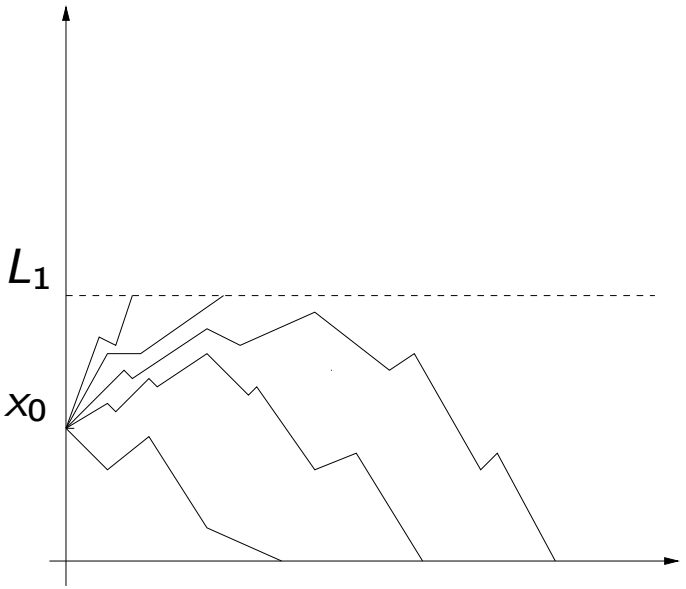
$$\lim_{k\to+\infty} \frac{\log(\mathrm{var}(\hat{p}_{k,\mathrm{FS}})w_{k,\mathrm{FS}})}{\log p_k} < 2 \text{ for } R \neq 1/\rho,$$

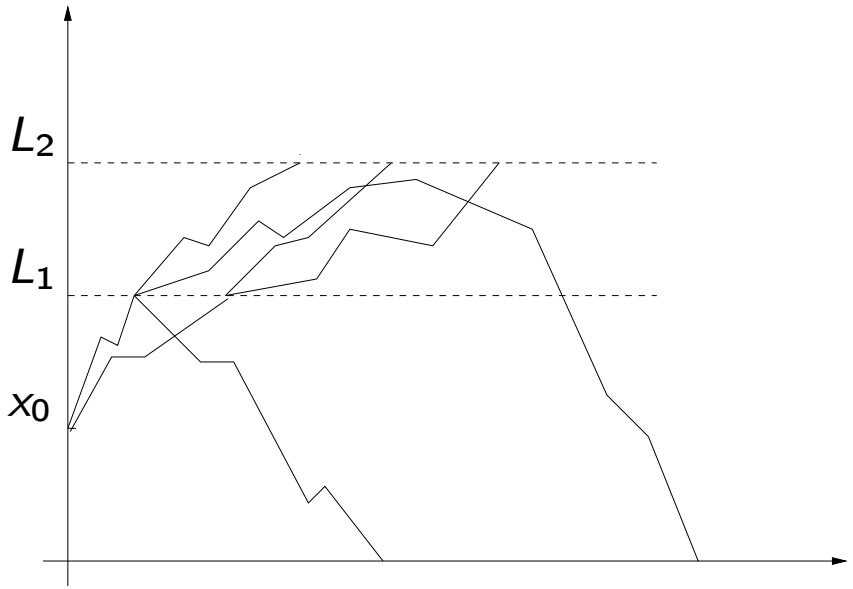
Fixed splitting efficient only in critical regime. From simulations, it proved to be very sensitive to the branching rate R.
But we can choose R adaptively...

Fixed effort splitting

At each level, we randomly branch N offsprings on the remaining paths.

Let $\hat{p}_{k|k-1} = \frac{\#\text{particles reaching } L_k}{N}$, we estimate the total probability by $\hat{p}_{n,\mathrm{FE}} = \prod_{k=1}^{n} \hat{p}_{k|k-1}$.





Feynman-Kac representation

Following [CDLL06] let $\tau_k = \inf\{t > 0, \phi(X_t) \geq L_k\}$, and $T_k = \tau_k \wedge \tau_{A_0}$.

$$\mathbb{E}[f(X_t, t \leq T_n) | \tau_A \leq \tau_{A_0}] = \frac{\mathbb{E}[f(X_t, t \leq T_n) \prod_{k=0}^{n} \mathbb{1}_{\phi(X_{T_k}) \geq L_k}]}{\mathbb{E}[\prod_{k=0}^{n} \mathbb{1}_{\phi(X_{T_k}) \geq L_k}]},$$

and

$$\mathbb{P}(T_A \leq T_{A_0}) = \mathbb{E}[\prod_{k=0}^{n} \mathbb{1}_{\phi(X_{T_k}) \geq L_k}].$$

Special case of general Feynman-Kac formula

$$\eta_n(f) = \frac{\mathbb{E}[f(Z_1, \ldots, Z_n) \prod_{k=0}^{n} G_k(Z_k)]}{\mathbb{E}[\prod_{k=0}^{n} G_k(Z_k)]}$$

where G_k are measurable non-negative functions called potentials.
The corresponding Interacting Particle System approximation is exactly the fixed effort splitting algorithm.

Convergence results

Once we have our problem in the form of a Feynman-Kac formula, associated with an IPS, we can use directly results in the monograph [DM04]. This includes unbiasedness of $\hat{p}_{n,\mathrm{FE}}$, SLLN, CLT...
Central Limit Theorem:

Theorem

$$\sqrt{N}\,\frac{\hat{p}_{n,\mathrm{FE}}-p}{p} \xrightarrow[N\to+\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2),$$

with

$$\begin{aligned}\sigma^2 &= \sum_{k=1}^{n}\frac{1-p_{k|k-1}}{p_{k|k-1}} \\ &\quad+\sum_{k=0}^{n-1}\frac{\mathrm{var}(\mathbb{P}(\phi(X_{\tau_n})\geq L|X_{\tau_k},\phi(X_{\tau_k})\geq L_k))}{\mathbb{P}^2(\phi(X_{\tau_n})\geq L|\phi(X_{\tau_k})\geq L_k)}\frac{1-p^2_{k|k-1}}{p_{k|k-1}}.\end{aligned}$$

Some remarks

$$\sigma^2 = \sum_{k=1}^{n} \frac{1-p_{k|k-1}}{p_{k|k-1}} + \sum_{k=0}^{n-1} \frac{\mathrm{var}(\mathbb{P}(\phi(X_{\tau_n}) \geq L | X_{\tau_k}, \phi(X_{\tau_k}) \geq L_k))}{\mathbb{P}^2(\phi(X_{\tau_n}) \geq L | \phi(X_{\tau_k}) \geq L_k)} \frac{1-p_{k|k-1}^2}{p_{k|k-1}}.$$

- The vaviance in the second term is zero if the importance function ϕ is the committor function, i.e. the levels L_k define iso-committor surfaces.
- First term is minimal when all the $p_{k|k-1}$ are equal (optimization with the constraint that their product is p).

Efficiency of fixed effort

Theorem ([CDG10])

Under mixing conditions, and provided that the levels are well chosen (in the sense that $\forall k, 0 < \rho_1 \leq p_{k|k-1} \leq \rho_2 < 1$*), we have*

$$\mathrm{var}(\hat{p}_{k,\mathrm{FE}}) \leq p_k^2[(1 + \frac{C}{(N-1)\rho_1})^k - 1]$$

for some $C > 0$.

and thus, taking $w_{k,\mathrm{FE}} = kN$,

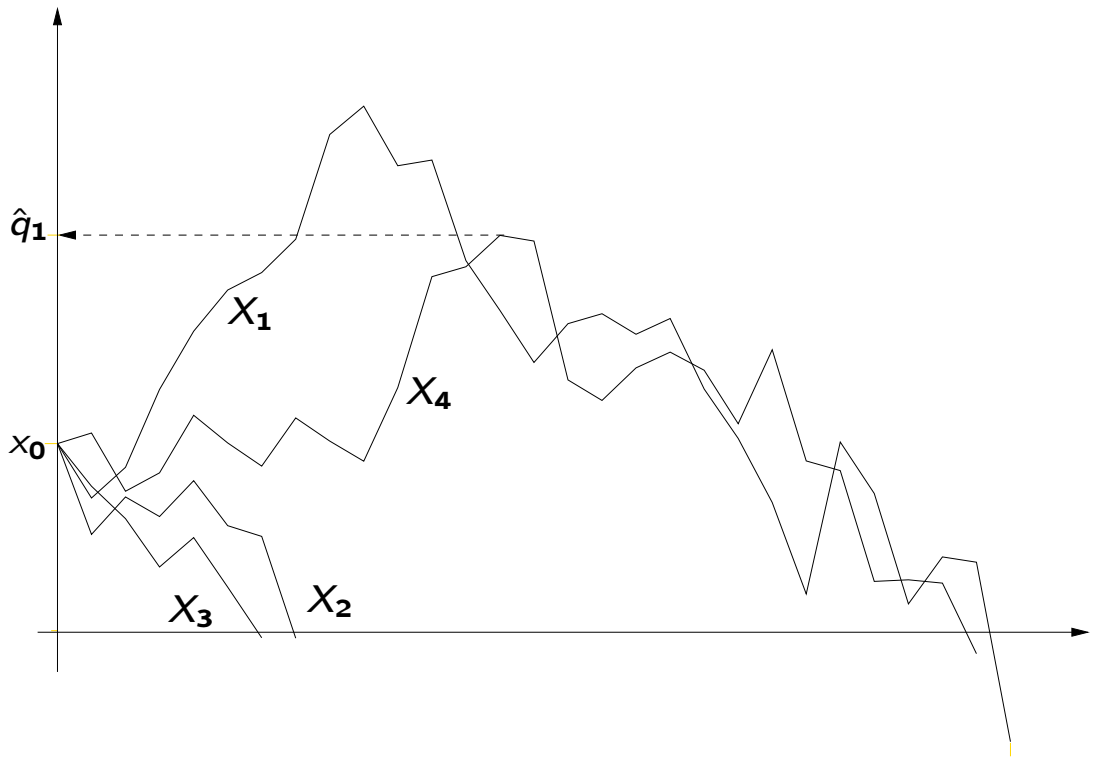
Corollary

$$\liminf_{k\to\infty} \frac{\log(w_{k,\mathrm{FE}}\,\mathrm{var}(\hat{p}_{k,\mathrm{FE}}))}{\log p_k} \geq 2 + \frac{1}{2\log\rho_2}\log(1 + \frac{C}{\rho_1(N-1)}).$$

Almost get efficiency, closer as N gets larger.
Paul Dupuis and Yi Cai recently obtained a similar result.

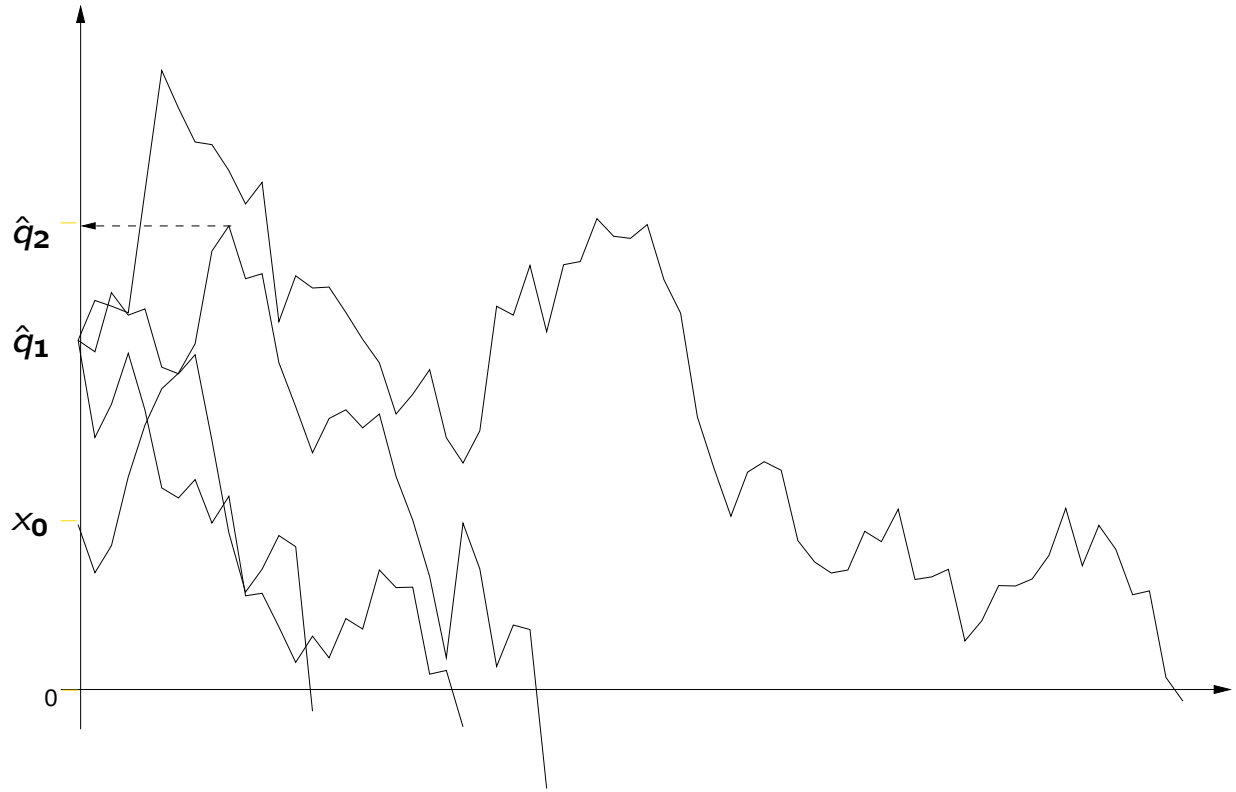
Adaptive splitting

To each trajectory X we associate $S = \sup_{t \leq \tau_A \wedge \tau_{A_0}} \phi(X_t)$.
Choose $K < N$, and compute the K/N quantile $\hat{q}_1$ of $S_1, \ldots, S_N$. This quantile will be the first level of splitting.



Adaptive splitting

Keep the trajectories above $\hat{q}_1$, and split (branch) them when they first corss this level. From the whole N again trajectories, all above $\hat{q}_1$, compute $\hat{q}_2$, and iterate.



Convergence of adaptive splitting

Only in dimension 1: a.s. convergence and CLT.

Theorem ([CG07])

$$\sqrt{N}(p_n - \hat{p}_{n,AS}) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

with

$$\sigma^2 = p_n^2((n-1)\frac{1-p_0}{p_0} + \frac{1-\rho}{\rho}),$$

with $p_0 = \frac{K}{N}$.

Minimal variance in the CLT, but no asyptotic efficiency results.
Complexity in $nN \log N$.
New version with $K = N - 1$: discard only the lower trajectory at each step.

Application to reactive trajectories

Let $V : \mathbb{R}^d \to \mathbb{R}$ the potential function, consider the overdamped Langevin dynamics:

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

where $\beta = 1/(k_B T)$
Equilibrium canonical measure:

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

where $Z = \displaystyle\int_{\mathbb{R}^d} \exp(-\beta V(x))\, dx$

Reactive trajectories

A and B two metastable (recurrent) regions in $\mathbb{R}^d$.

Reactive trajectory: piece of equilibrium trajectory that leaves A and goes to B without going back to A in the meantime (excursion from A to B)
Problem: one may wait a long time before the trajectory eventually reaches B

Reaction coordinate

Also called "importance function" in rare event literature
Smooth one-dimensional function:

$$\xi : \mathbb{R}^d \to \mathbb{R}$$

such that:

$$|\nabla \xi| \neq 0, \ A \subset \{x \in \mathbb{R}^d, \xi(x) < z_{\min}\} \text{ and } B \subset \{x \in \mathbb{R}^d, \xi(x) > z_{\max}\},$$

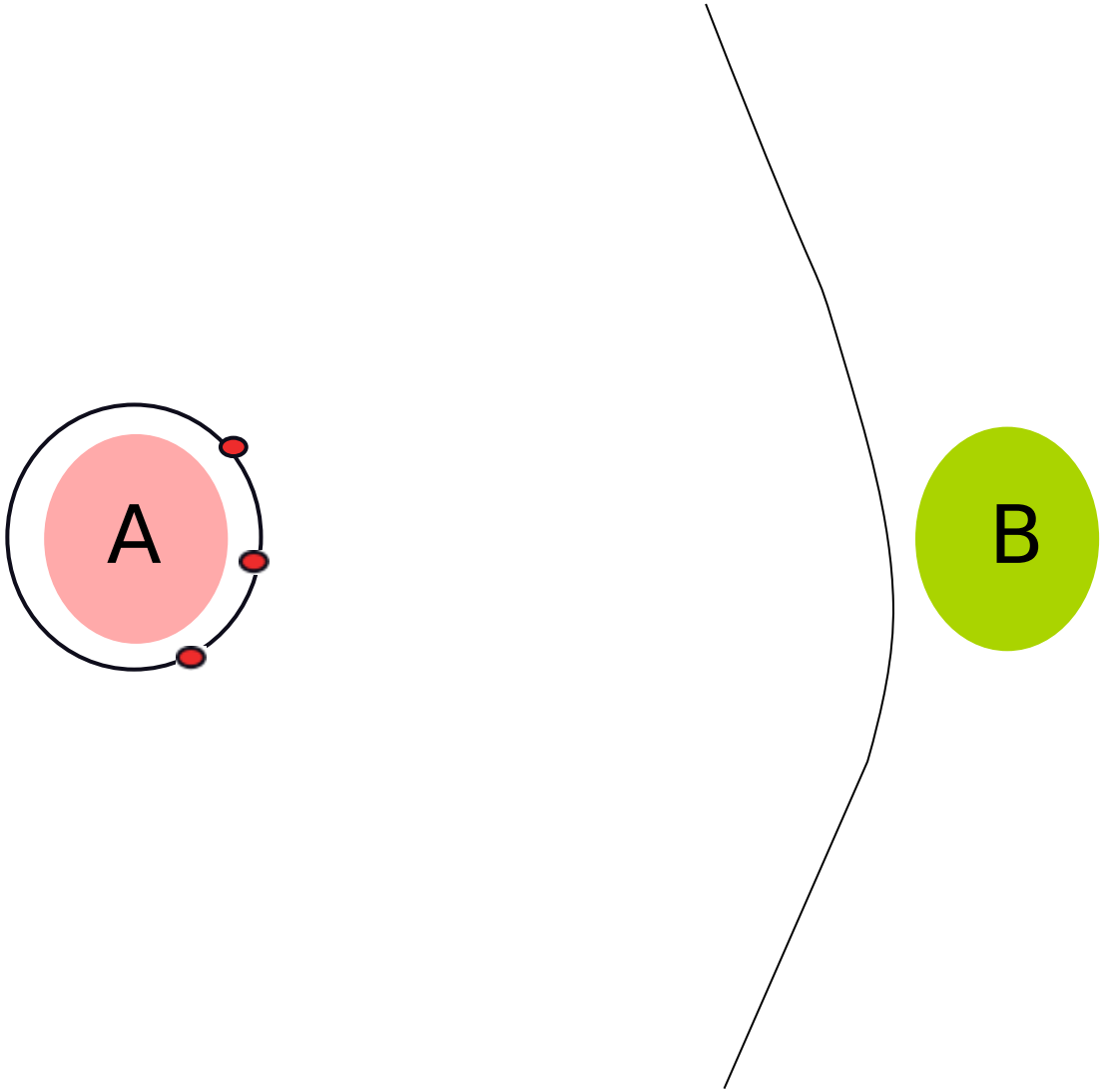
where $z_{\min} < z_{\max}$ are two given real numbers
Example: $\xi(x) = \|x - x_A\|$ with $x_A \in A$ denotes a reference configuration in A

Committor function

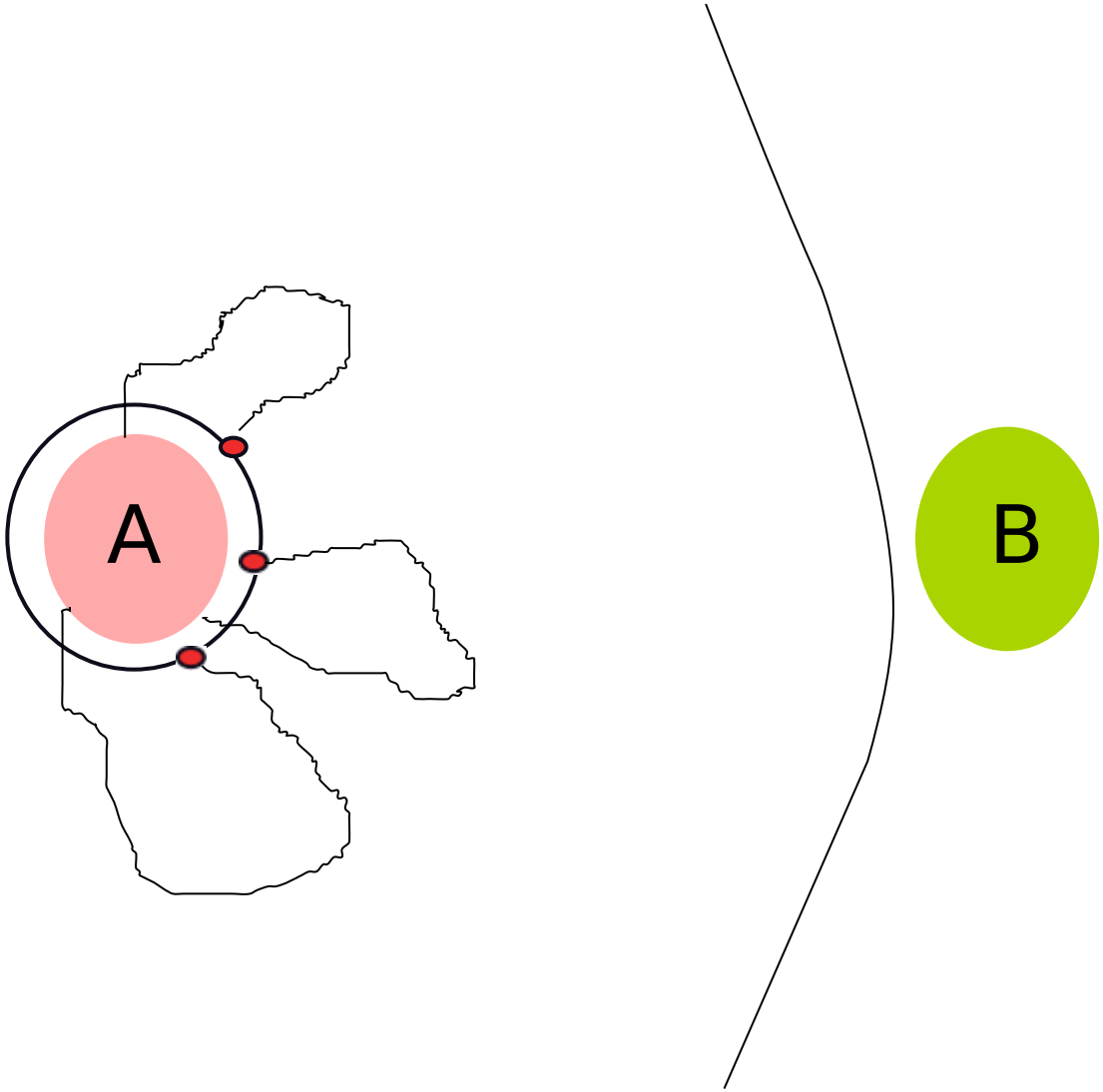
It is well known (e.g. Metzner, Schütte and Vanden-Eijnden (2006), Vanden-Eijnden, Venturoli, Ciccotti and Elber (2008)) that the optimal reaction coordinate/importance function is the committor function $q(x) = \mathbb{P}(\tau_A(x) > \tau_B(x))$.
q is solution of

$$\begin{cases} -\nabla V \cdot \nabla q + \beta^{-1} \Delta q = 0 \text{ in } \mathbb{R}^d \setminus (\overline{A} \cup \overline{B}), \\ q = 0 \text{ on } \partial A \text{ and } q = 1 \text{ on } \partial B. \end{cases}$$

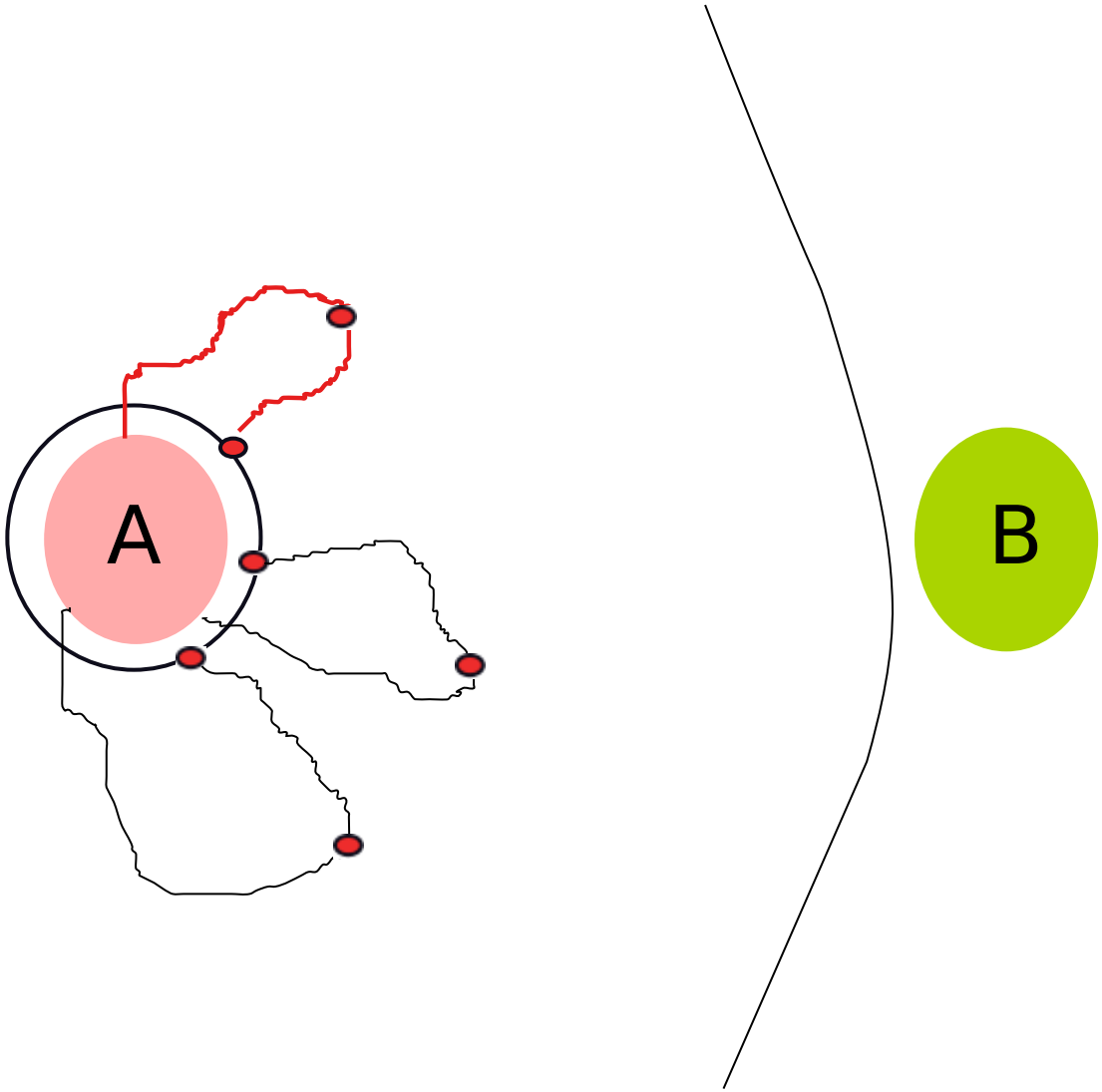
Algorithm



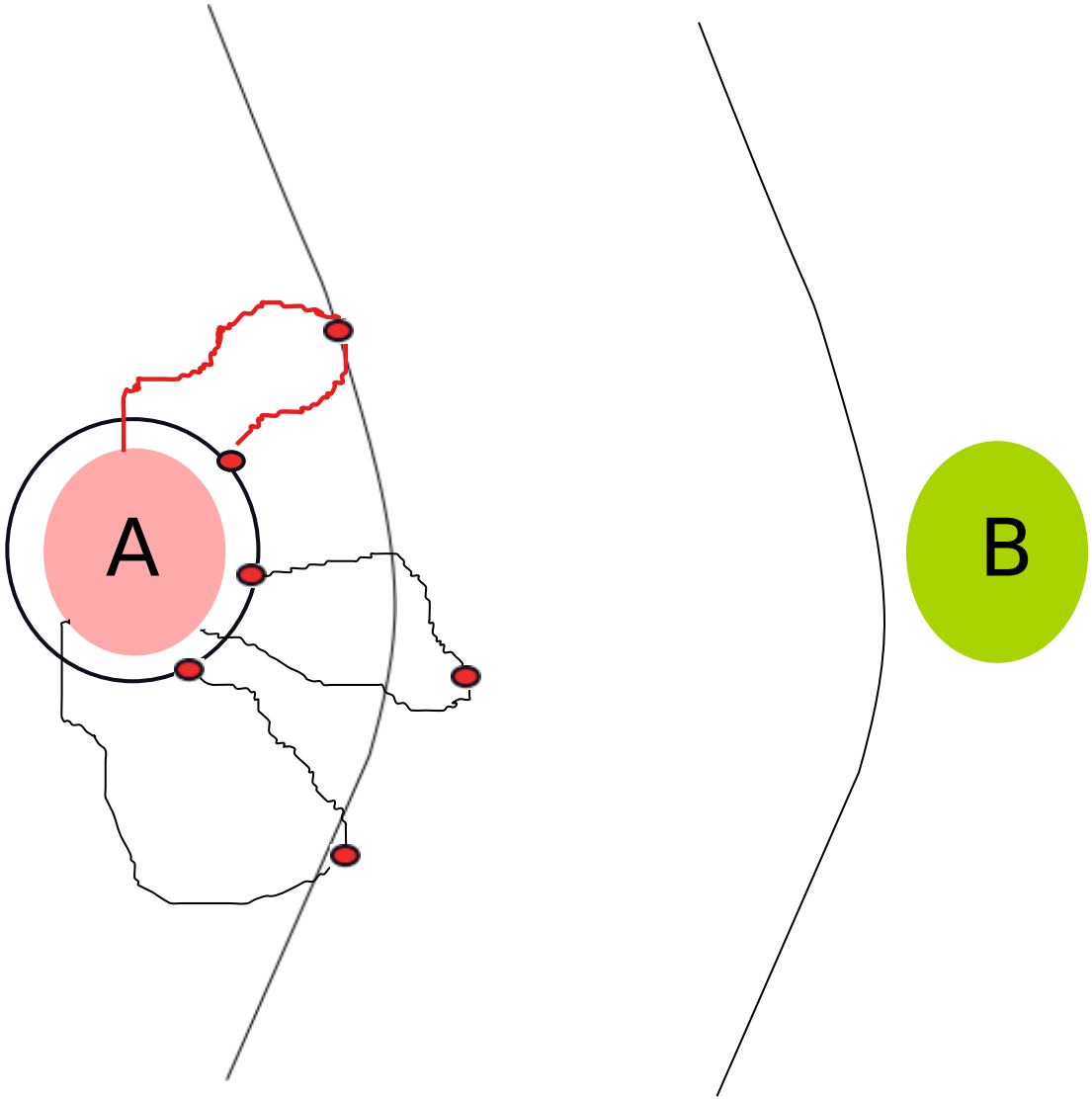
Algorithm



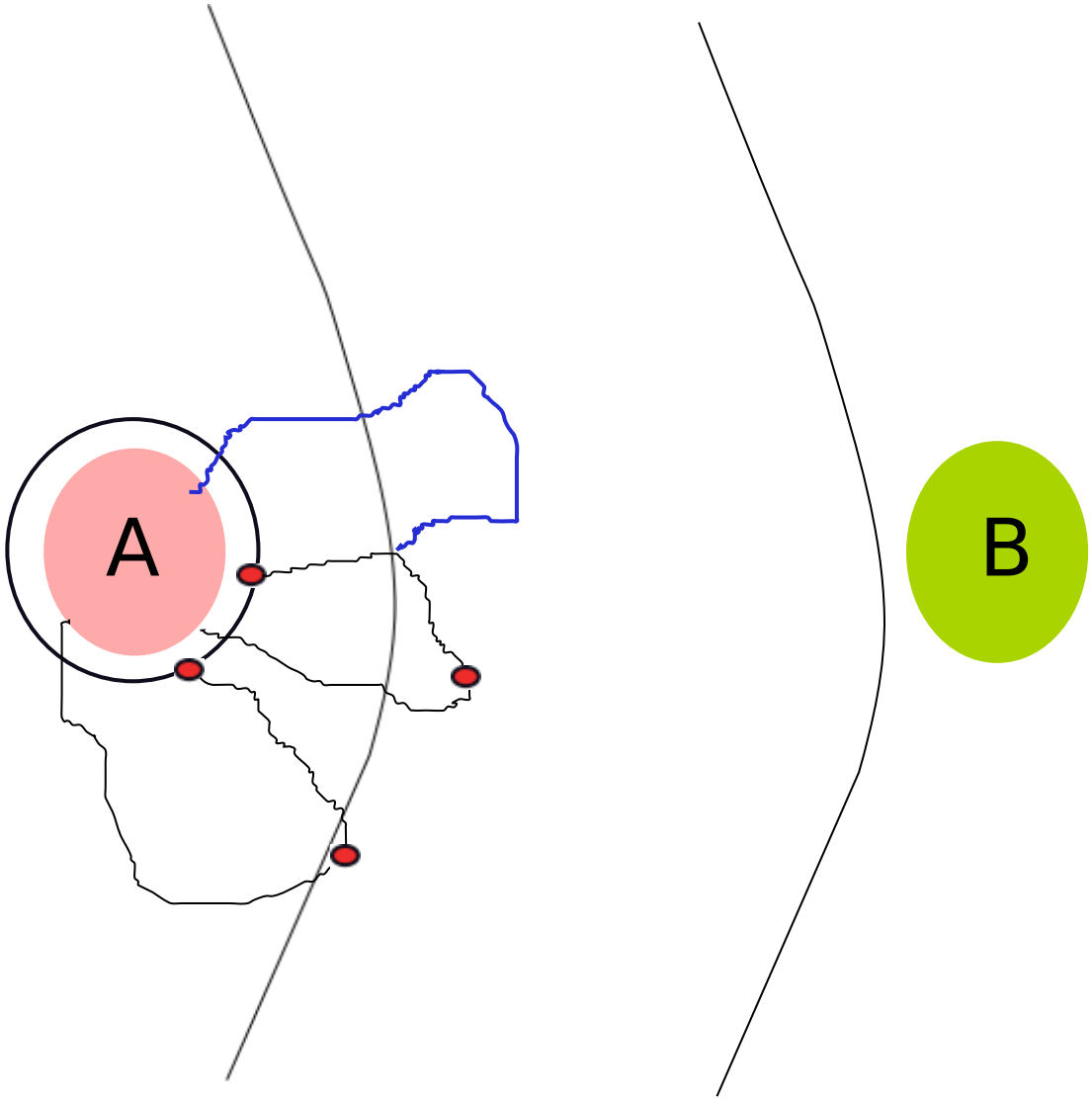
Algorithm



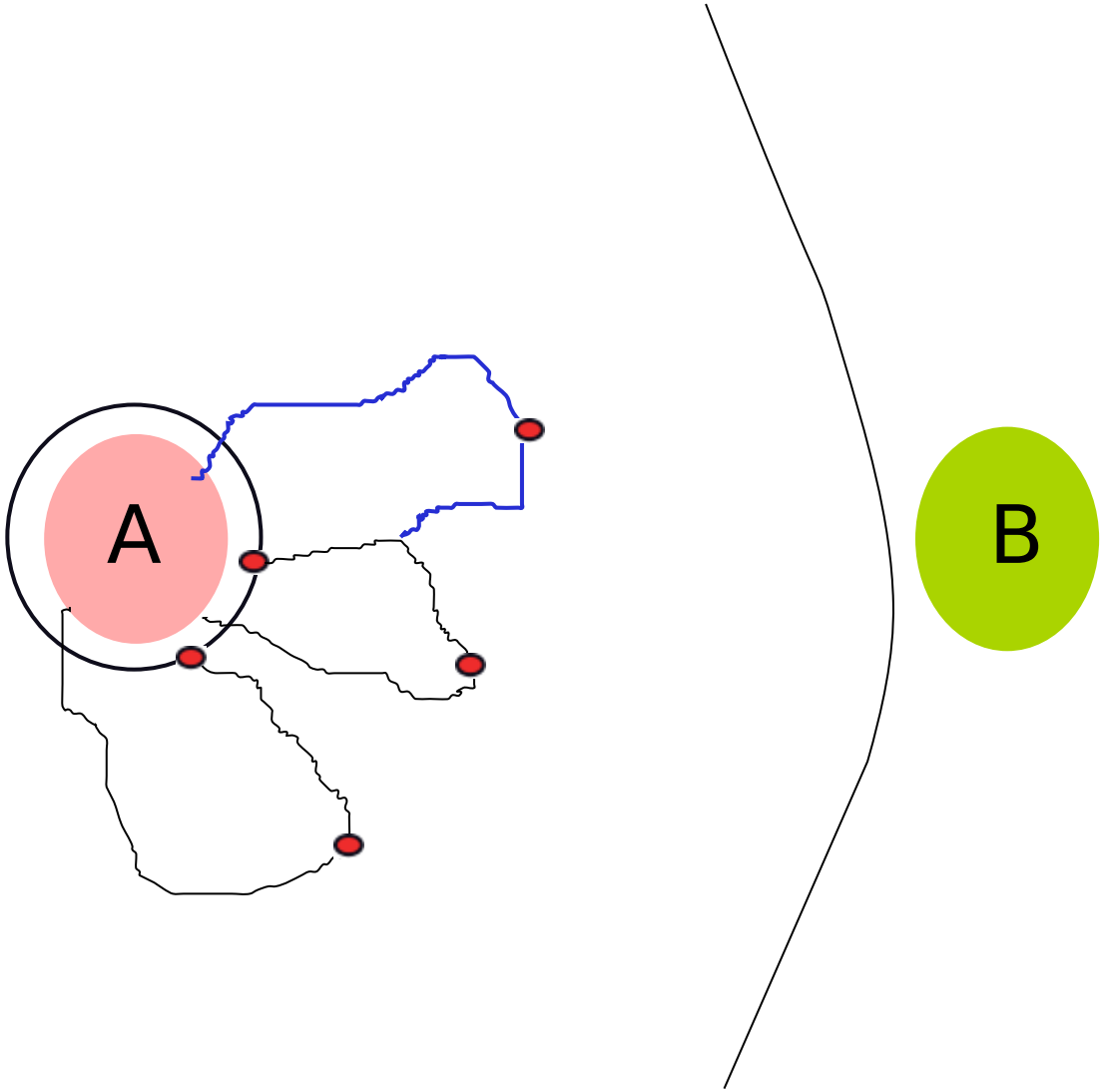
Algorithm



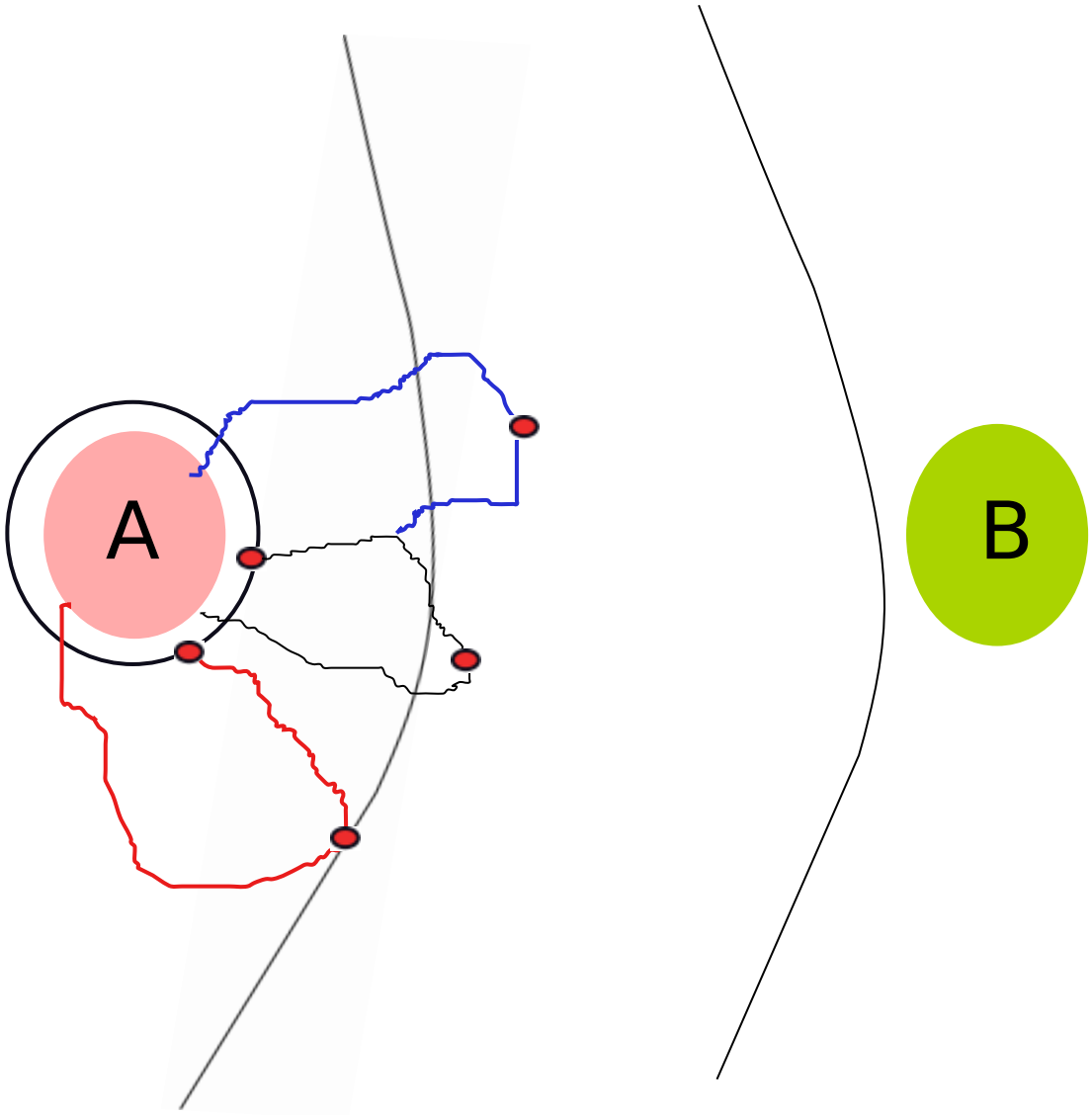
Algorithm



Algorithm



Algorithm



Algorithm

Iterate until all the paths reach $\Sigma_{\max}$, and keep only those which actually reach B.

Initialization: generate a long trajectory "around" A (wait until equilibrium), and keep (a subsample of) the upward crossing points of $\Sigma_{\min}$, and the corresponding piece of trajectory coming from A.

Remark: the algorithm is a kind of adaptive Forward Flux Sampling (e.g. Allen, Valeriani and ten Wolde (2009)), or more generally Importance Splitting

Diffusion process approximated by Euler scheme.

Convergence and efficiency

Unbiased estimates $\hat{p}_{k,\mathrm{IR}}$.
In dimension 1, almost have efficiency. In this case
$\mathrm{var}(\hat{p}_{k,\mathrm{IR}}) = p_k^2(p_k^{-\frac{1}{N}} - 1)$ and $w_{k,\mathrm{IR}} = -N \log N \log p_k$.

$$\lim_{k \to +\infty} \frac{\log(\mathrm{var}(\hat{p}_{k,\mathrm{IR}}) w_{k,\mathrm{IR}})}{\log p_k} = 2 - \frac{1}{N}.$$

(Guyader,Hengertner,Matzner-Løber, not yet published).

1D example

Simple example to validate the method against DNS

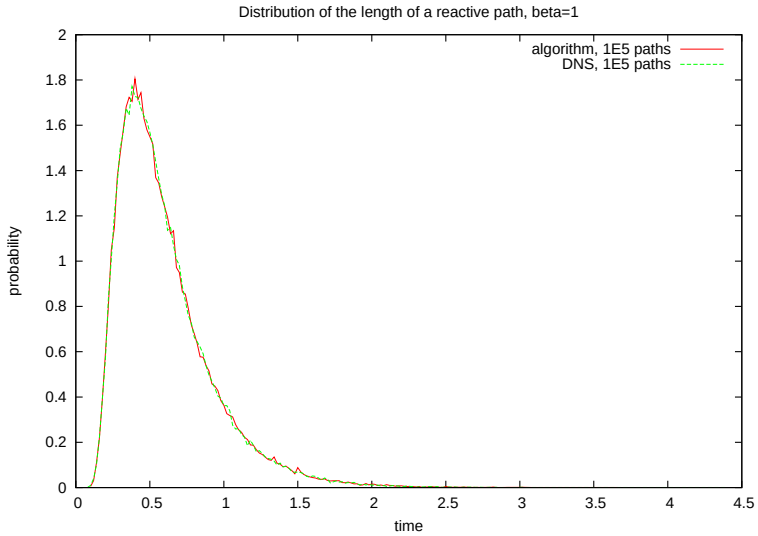
$$V(x) = x^4 - 2x^2.$$

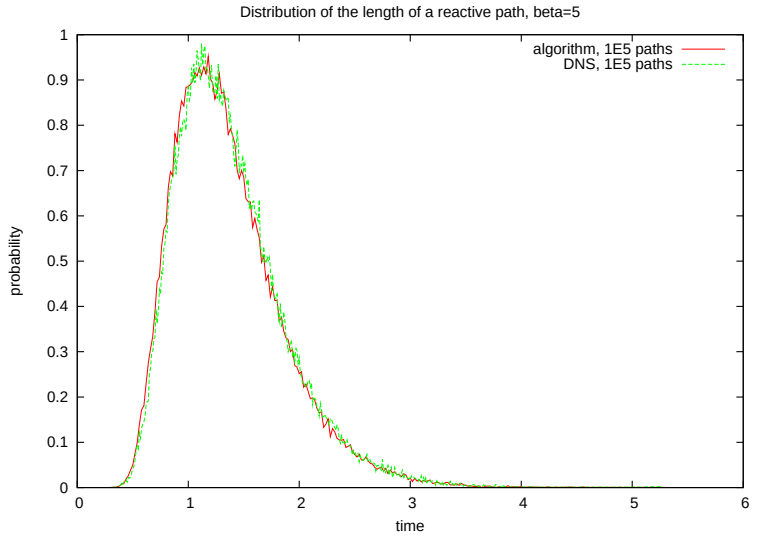
This potential has two minima at ± 1 and one saddle point at 0. In this simple one dimensional setting, we set as metastable states $A = \{-1\}$ and $B = \{+1\}$, and the reaction coordinate is taken to be simply

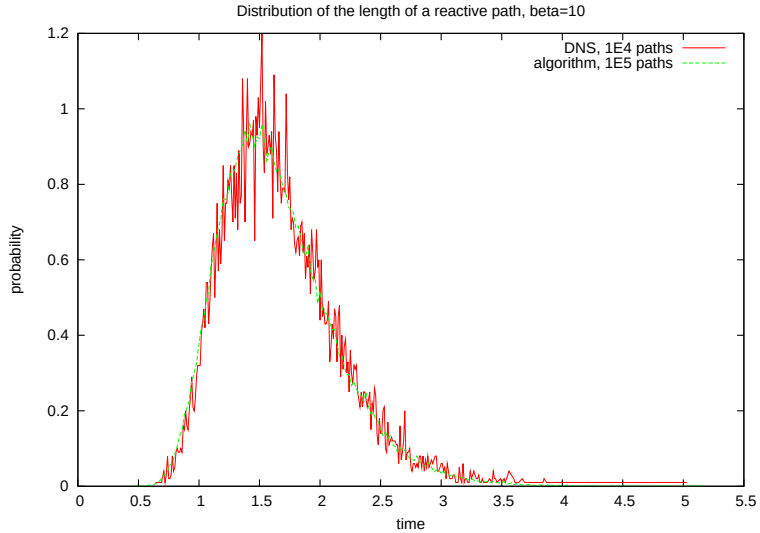
$$\xi(x) = x.$$

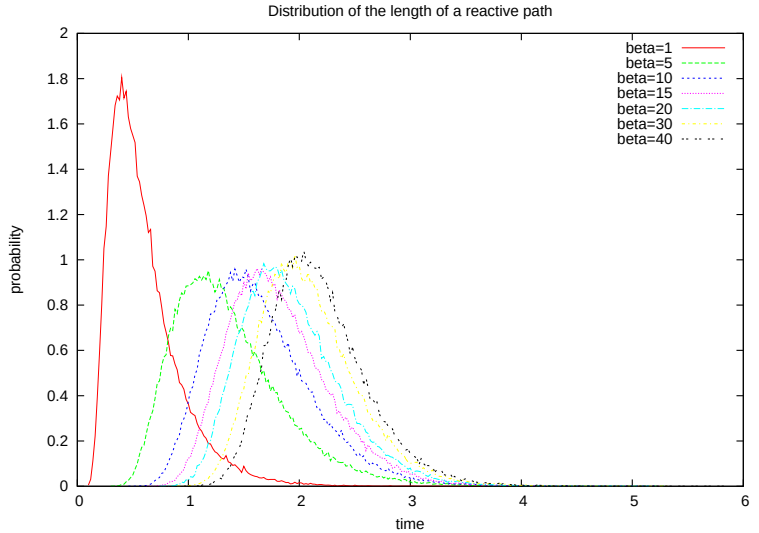
We computed the distribution of the duration of the reactive paths.

1D example


Distribution of the length of a reactive path, beta=1
algorithm, 1E5 paths
DNS, 1E5 paths
probability
time


Distribution of the length of a reactive path, beta=5
algorithm, 1E5 paths
DNS, 1E5 paths
probability
time


Distribution of the length of a reactive path, beta=10
DNS, 1E4 paths
algorithm, 1E5 paths
probability
time


Distribution of the length of a reactive path
beta=1
beta=5
beta=10
beta=15
beta=20
beta=30
beta=40
probability
time

2D example

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

as considered by Metzner, Schütte and Vanden-Eijnden (2006)

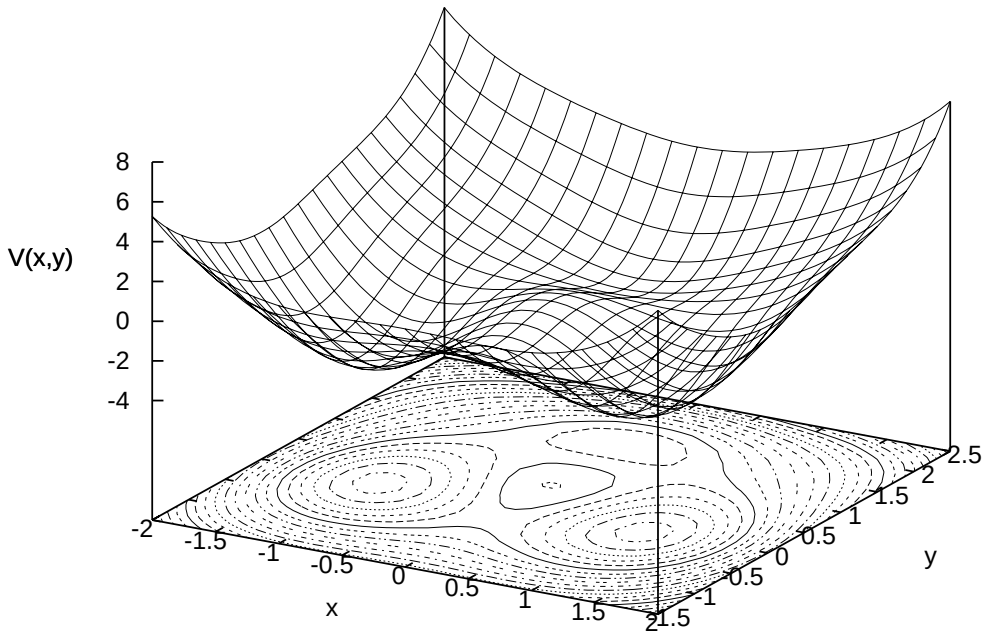
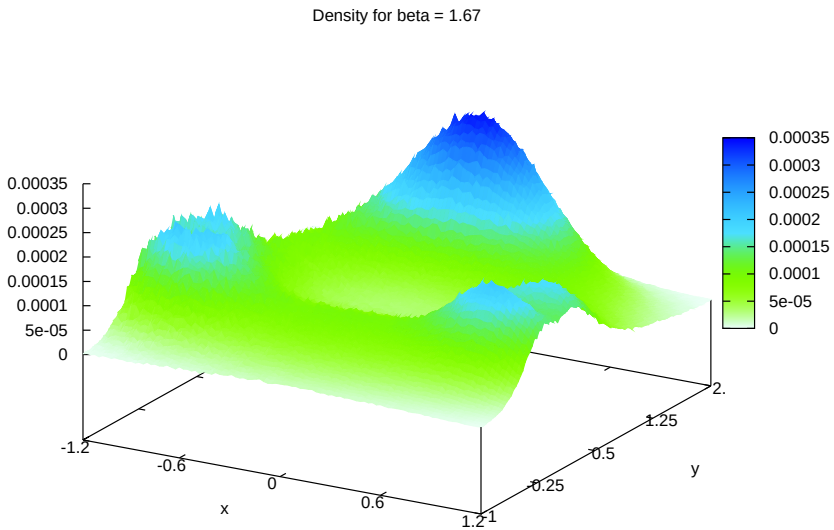
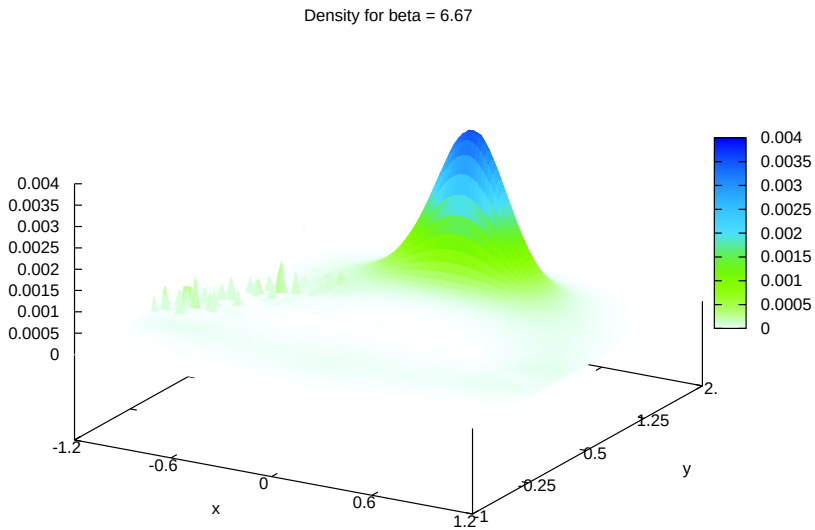


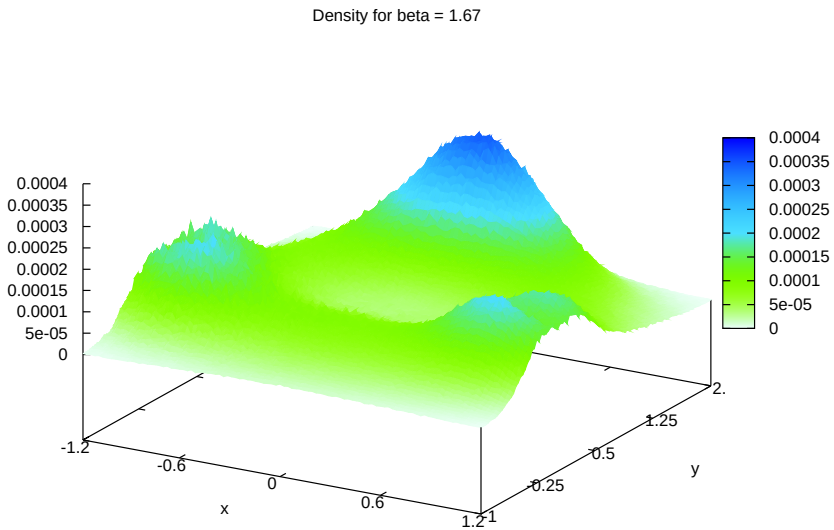
Figure: The potential V.

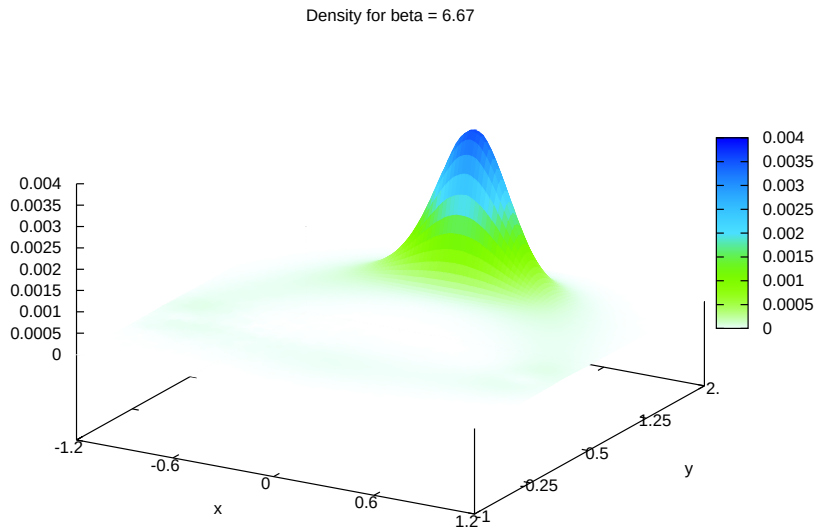
Density of reactive paths





Density of reactive paths





Flux of reactive trajectories

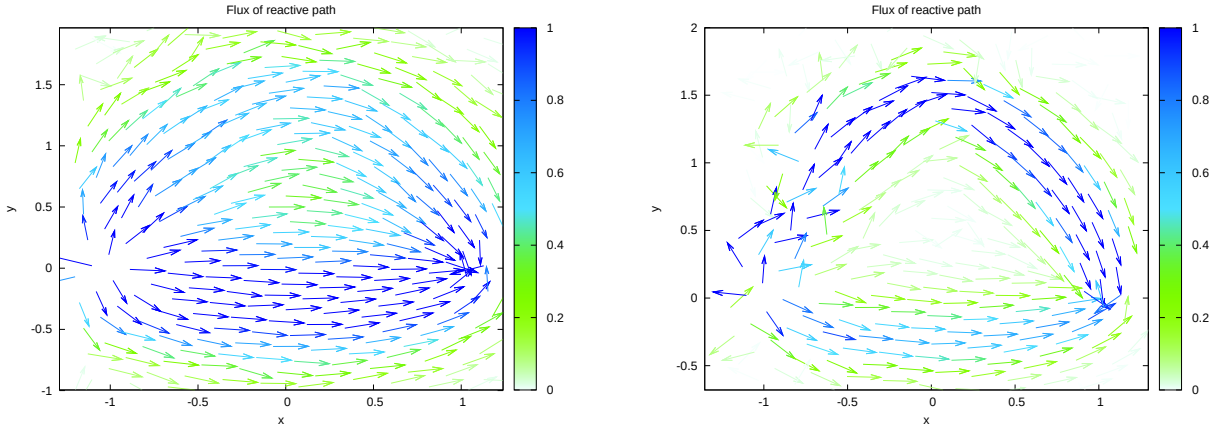


Figure: Flux of reactive trajectories, at inverse temperature $\beta = 1.67$ on the left, and $\beta = 6.67$ on the right. The color indicates the norm of the flux.

A few reactive trajectories

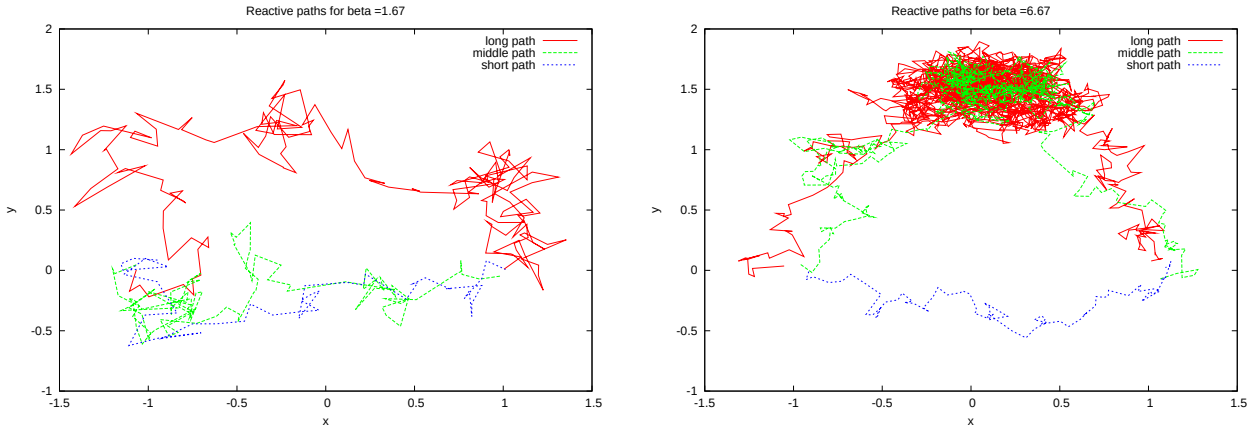


Figure: A few reactive paths for $\beta = 1.67$ (left), $\beta = 6.67$ (right).

F. Cerou, P. Del Moral, et al.
A non asymptotic variance theorem for unnormalized feynman-kac particle models.
Annales de l'IHP, 2010.

F. Cérou, P. Del Moral, et al.
Genetic genealogical models in rare event analysis.
Latin American Journal of Probability and Mathematical Statistics, 1, 2006.

F. Cérou and A. Guyader.
Adaptive multilevel splitting for rare event analysis.
Stoch. Anal. Appl., 25(2):417–443, 2007.

P. Del Moral.
Feynman-Kac formulae, Genealogical and interacting particle systems with applications.
Probability and its Applications. Springer-Verlag, New York, 2004.


P. Glasserman, P. Heidelberger, et al.
Multilevel splitting for estimating rare event probabilities.
Oper. Res., 47(4):585–600, 1999.


P. W. Glynn and W. Whitt.
The asymptotic efficiency of simulation estimators.
Oper. Res., 40(3):505–520, 1992.


H. Kahn and T. Harris.
Estimation of particle transmission by random sampling.
National Bureau of Standards Appl. Math. Series, 12:27–30, 1951.