Monte Carlo methods in molecular dynamics.

Tony Lelièvre

CERMICS, Ecole Nationale des Ponts et Chaussées, France,

MicMac team, INRIA, France.

`http://cermics.enpc.fr/∼lelievre`

Outline

1 Free energy and metastability,

2 Constrained dynamics,

- 2.1 Thermodynamic integration,
- 2.2 Constrained SDEs,
- 2.3 Non-equilibrium dynamics,

3 Adaptive methods,

- 3.2 Adaptive methods: algorithms,
- 3.2 Adaptive methods: convergence,
- 3.3 Multiple replicas implementations,
- 3.4 Application to Bayesian statistics.

1 Free energy and metastability

We consider a molecular system with N particles with position $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N) = \boldsymbol{x} \in \mathbb{R}^{3N}$ interacting through the potential $V(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$.
In the NVT ensemble, one wants to sample the Boltzmann-Gibbs probability measure:

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

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

Aim: compute “macroscopic quantities” like the likelihood of molecular conformations, reaction rates, ...

1 Free energy and metastability

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

$$V = \sum_{i<j} V_1(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{i<j<k} V_2(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k) + \sum_{i<j<k<l} V_3(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, \boldsymbol{x}_l)$$

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

1 Free energy and metastability

Examples of quantities of interest:

- specific heat

$$C \propto \langle V^2 \rangle_\mu - \langle V \rangle_\mu^2$$

- pressure

$$P \propto -\langle q \cdot \nabla V(q) \rangle_\mu$$

1 Free energy and metastability

Since this is a high-dimensional problem ($N \gg 1$) Monte Carlo methods are used, typically based on Markov chains.
For example, to sample μ, one can use $\boldsymbol{X}_t$ solution to the Stochastic Differential Equation (SDE):

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

(*gradient or over-damped Langevin dynamics*).
Under suitable assumption, we have the ergodic property: for μ-a.e. $\boldsymbol{X}_0$,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \phi(\boldsymbol{X}_t) dt = \int \phi(\boldsymbol{x}) d\mu(\boldsymbol{x}).$$

1 Free energy and metastability

Probabilistic insert (1): discretization of SDEs.
The discretization of (GD) by the Euler scheme is (for a fixed timestep Δt):

$$\boldsymbol{X}_{n+1} = \boldsymbol{X}_n - \nabla V(\boldsymbol{X}_n)\,\Delta t + \sqrt{2\beta^{-1}\Delta t}\boldsymbol{G}_n$$

where $(G_n^i)_{1\le i\le 3N, n\ge 0}$ are i.i.d. random variables with law $\mathcal{N}(0,1)$. Indeed,

$$(\boldsymbol{W}_{(n+1)\Delta t} - \boldsymbol{W}_{n\Delta t})_{n\ge 0} \stackrel{\mathcal{L}}{=} \sqrt{\Delta t}(\boldsymbol{G}_n)_{n\ge 0}.$$

In practice, a sequence of i.i.d. random variables with law $\mathcal{N}(0,1)$ may be obtained from a sequence of i.i.d. random variables with law $\mathcal{U}((0,1))$ (given by the rand() function on computers).

Proof (invariant measure): One needs to show that if the law of $\boldsymbol{X}_0$ is μ, then the law of $\boldsymbol{X}_t$ is also μ. Let us denote $\boldsymbol{X}_t^{\boldsymbol{x}}$ the solution to (GD) such that $\boldsymbol{X}_0 = \boldsymbol{x}$. Let us consider the function $u(t, \boldsymbol{x})$ solution to:

$$\begin{cases} \partial_t u(t,\boldsymbol{x}) = -\nabla V(\boldsymbol{x}) \cdot \nabla u(t,\boldsymbol{x}) + \beta^{-1} \Delta u(t,\boldsymbol{x}), \\ u(0,\boldsymbol{x}) = \phi(\boldsymbol{x}) (\text{+ assumptions on decay at infinity}), \end{cases}$$

then, $u(t,\boldsymbol{x}) = \mathbb{E}(\phi(\boldsymbol{X}_t^{\boldsymbol{x}}))$. Thus, the measure μ is invariant:

$$\frac{d}{dt} \int \mathbb{E}(\phi(\boldsymbol{X}_t^{\boldsymbol{x}})) d\mu(\boldsymbol{x}) = Z^{-1} \int \partial_t u(t,\boldsymbol{x}) \exp(-\beta V(\boldsymbol{x})) d\boldsymbol{x}$$

$$= Z^{-1} \int \left(-\nabla V \cdot \nabla u + \beta^{-1} \Delta u\right) \exp(-\beta V) = 0.$$

Therefore, $\int \mathbb{E}(\phi(\boldsymbol{X}_t^{\boldsymbol{x}})) d\mu(\boldsymbol{x}) = \int \phi(\boldsymbol{x}) d\mu(\boldsymbol{x})$.

1 Free energy and metastability

Probabilistic insert (2): Feynman-Kac formula.

Why $u(t,\boldsymbol{x}) = \mathbb{E}(\phi(\boldsymbol{X}_t^{\boldsymbol{x}}))$? For $0 < s < t$, we have (*characteristic method*):

$$du(t-s,\boldsymbol{X}_s^{\boldsymbol{x}}) = -\partial_t u(t-s,\boldsymbol{X}_s^{\boldsymbol{x}})\,ds + \nabla u(t-s,\boldsymbol{X}_s^{\boldsymbol{x}}) \cdot d\boldsymbol{X}_s^{\boldsymbol{x}}$$

$$+\beta^{-1}\Delta u(t-s,\boldsymbol{X}_s^{\boldsymbol{x}})\,ds,$$

$$= \Big(- \partial_t u(t-s,\boldsymbol{X}_s^{\boldsymbol{x}}) - \nabla V(\boldsymbol{X}_s^{\boldsymbol{x}}) \cdot \nabla u(t-s,\boldsymbol{X}_s^{\boldsymbol{x}})$$

$$+ \beta^{-1}\Delta u(t-s,\boldsymbol{X}_s^{\boldsymbol{x}}) \Big)\,ds + \sqrt{2\beta^{-1}}\nabla u(t-s,\boldsymbol{X}_s^{\boldsymbol{x}}) \cdot d\boldsymbol{W}_s.$$

Thus, integrating over $s \in (0,t)$ and taking the expectation:

$$\mathbb{E}(u(0,\boldsymbol{X}_t^{\boldsymbol{x}})) - \mathbb{E}(u(t,\boldsymbol{X}_0^{\boldsymbol{x}})) = \sqrt{2\beta^{-1}}\mathbb{E}\left(\int_0^t \nabla u(t-s,\boldsymbol{X}_s^{\boldsymbol{x}}) \cdot d\boldsymbol{W}_s\right)$$

$$= 0.$$

Probabilistic insert (3): Itô's calculus. (in 1d.)

Where does the term Δu come from ? Starting from the discretization:

$$X_{n+1} = X_n - V'(X_n)\,\Delta t + \sqrt{2\beta^{-1}\Delta t} G_n,$$

we have (for a time-independent function u):

$$u(X_{n+1}) = u\left(X_n - V'(X_n)\,\Delta t + \sqrt{2\beta^{-1}\Delta t} G_n\right),$$

$$= u(X_n) - u'(X_n)V'(X_n)\,\Delta t + \sqrt{2\beta^{-1}\Delta t} u'(X_n) G_n$$

$$+\beta^{-1}(G_n)^2 u''(X_n)\Delta t + o(\Delta t).$$

Thus, summing over $n \in [0...t/\Delta t]$ and taking the limit $\Delta t \to 0$,

$$u(X_t) = u(X_0) - \int_0^t V'(X_s)u'(X_s)\,ds + \sqrt{2\beta^{-1}} \int_0^t u'(X_s) dW_s$$

$$+\beta^{-1} \int_0^t u''(X_s)\,ds.$$

1 Free energy and metastability

In practice, (GD) is discretized in time, and Cesaro means are computed: $\lim_{N_T\to\infty}\frac{1}{N_T}\sum_{n=1}^{N_T}\phi(\boldsymbol{X}_n)$.

Remark: Practitioners do not use over-damped Langevin dynamics but rather *Langevin dynamics*:

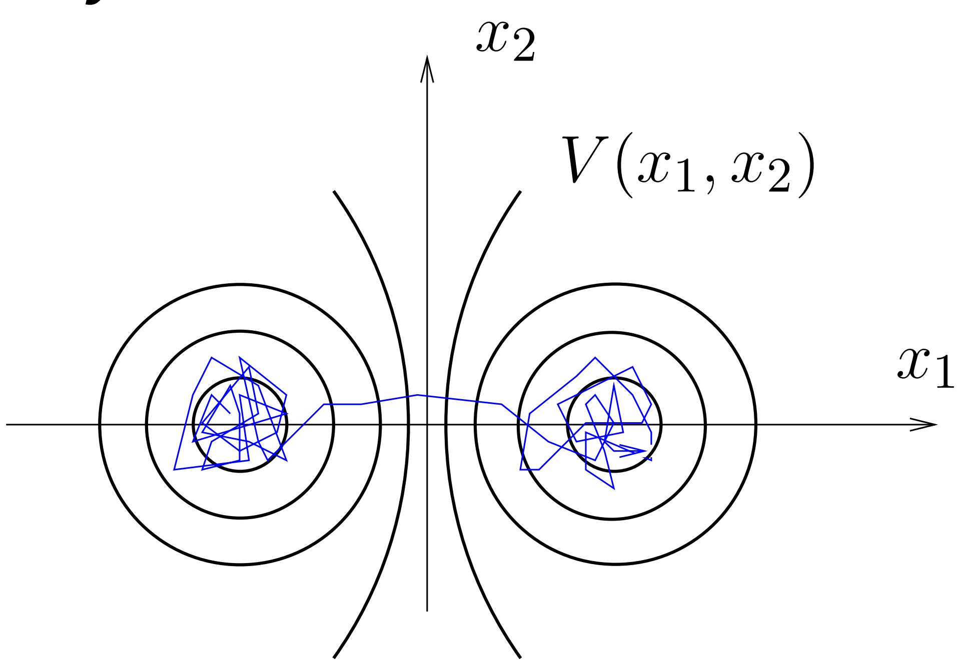
$$\begin{cases} d\boldsymbol{X}_t = M^{-1}\boldsymbol{P}_t\,dt, \\ d\boldsymbol{P}_t = -\nabla V(\boldsymbol{X}_t)\,dt - \gamma M^{-1}\boldsymbol{P}_t\,dt + \sqrt{2\gamma\beta^{-1}}d\boldsymbol{W}_t, \end{cases}$$

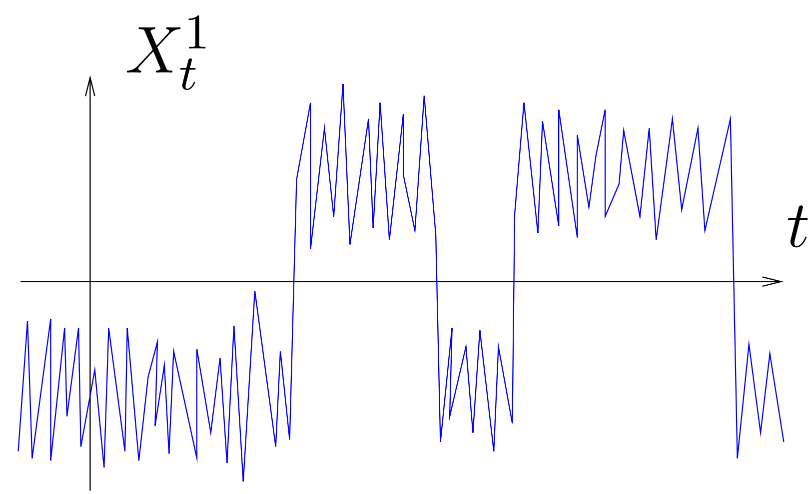
where M is the mass tensor and γ is the friction coefficient. In the following, we only consider over-damped Langevin dynamics.

1 Free energy and metastability

We therefore have a method to compute (an approximation of) $\int \phi(\boldsymbol{x}) d\mu(\boldsymbol{x})$, using $\boldsymbol{X}_t$. But, in practice, $\boldsymbol{X}_t$ is a metastable process, so that the convergence of the ergodic limit is very slow.

A bi-dimensional example: X_t^1 *is a slow variable of the system.*





1 Free energy and metastability

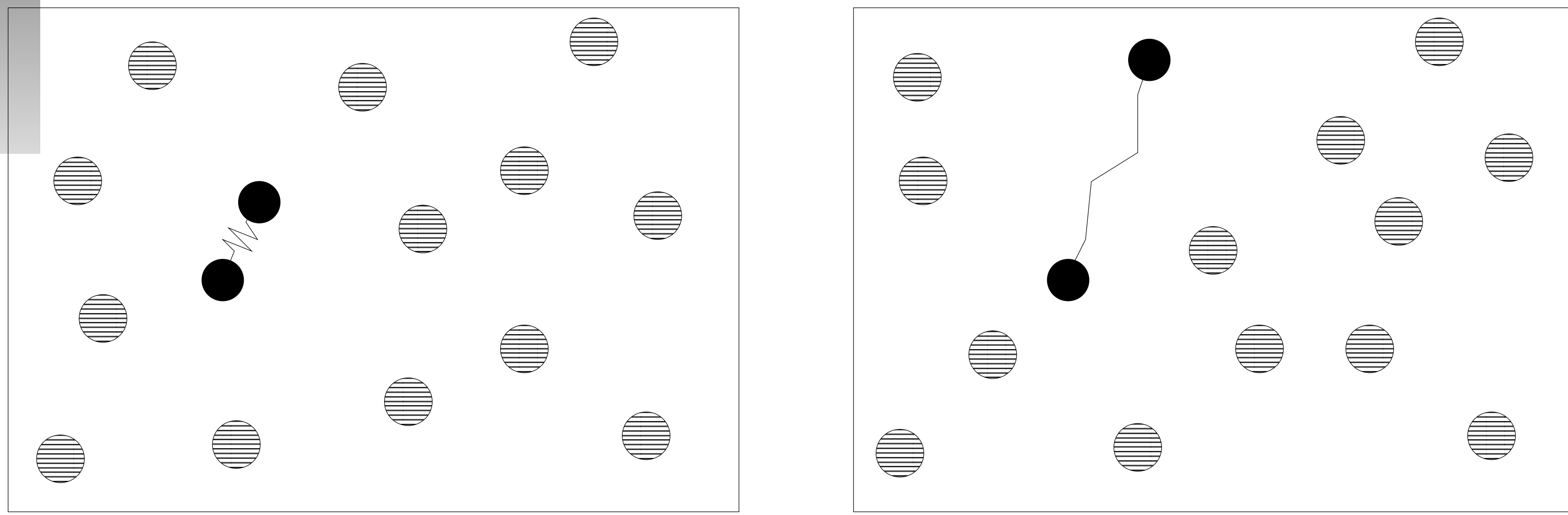
A more realistic example : (Dellago, Geissler): Influence of the solvation on a dimer conformation. The interaction potentials are *(J.D. Weeks, D. Chandler et H.C. Andersen)*:

$$V_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon & \text{if } r \leq r_0, \\ 0 & \text{if } r > r_0, \end{cases}$$

$$V_{\mathrm{S}}(r) = h \left[1 - \frac{(r - r_0 - w)^2}{w^2} \right]^2,$$

where ϵ, σ and w are positive constants and $r_0 = 2^{1/6}\sigma$. V_{S} is a double-well potential.

1 Free energy and metastability

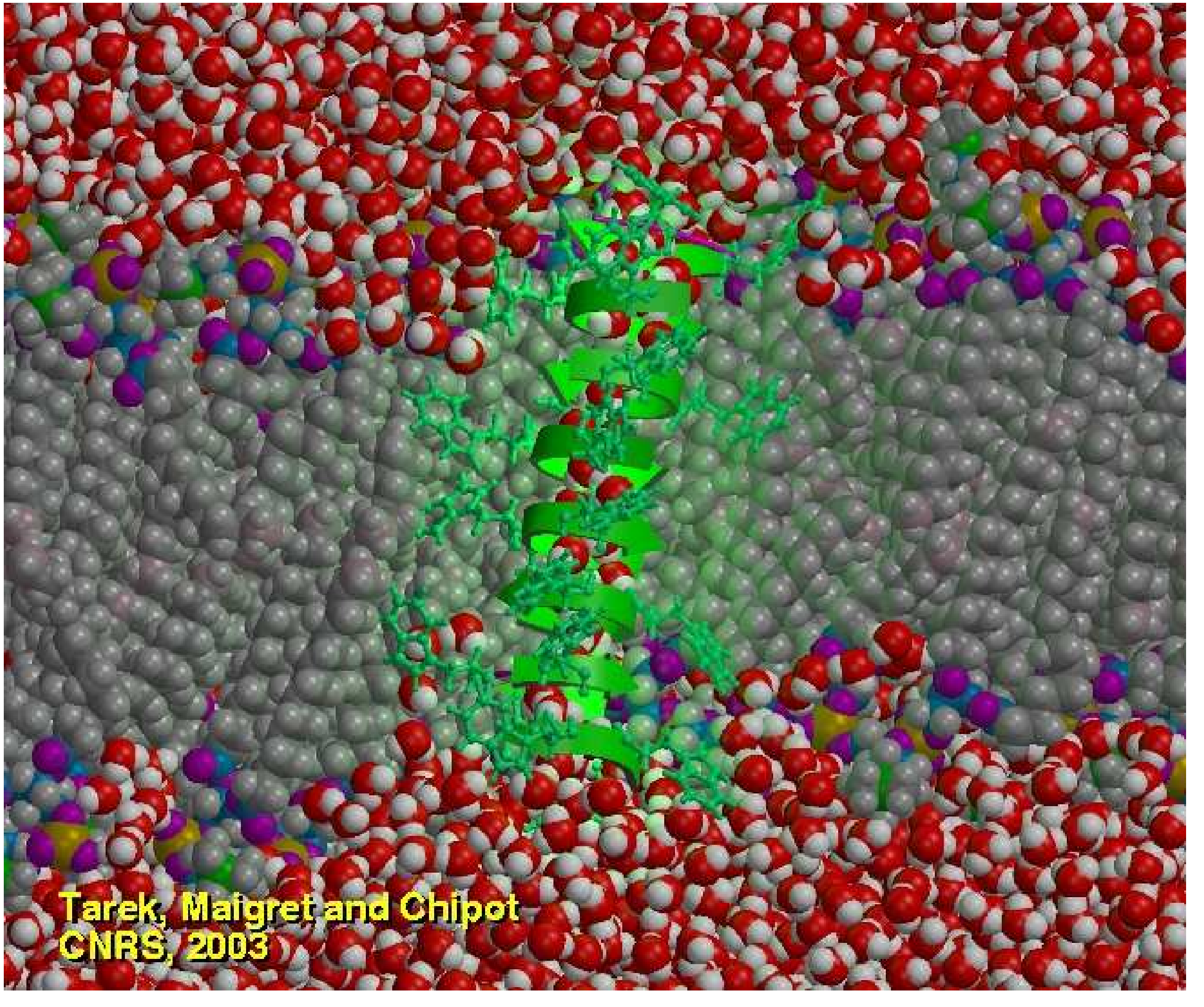


Left: compact state ($\xi = 0$).   Right: stretched state ($\xi = 1$).

A slow variable is $\xi(\boldsymbol{X}_t)$ where $\xi(\boldsymbol{x}) = \frac{|\boldsymbol{x}_1 - \boldsymbol{x}_2| - r_0}{2w}$ is a so-called reaction coordinate.

1 Free energy and metastability

A "real" example: ions canal in a cell membrane.
(C. Chipot).



1 Free energy and metastability

We suppose in the following that the slow variable is of dimension 1 and known: $\xi(\boldsymbol{x})$, where $\xi : \mathbb{R}^n \to \mathbb{R}$.

This slow variable contains most of the information needed in practice so that it would be enough to compute the law of $\xi(\boldsymbol{X})$, for $\boldsymbol{X}$ with law μ.

Lemme 1 *The image of the measure μ by ξ is $Z^{-1}\exp(-\beta A(z))\,dz$, where*

$$A(z) = -\beta^{-1}\ln\left(\int_{\Sigma_z} e^{-\beta V}|\nabla\xi|^{-1} d\sigma_{\Sigma_z}\right) = -\beta^{-1}\ln Z_{\Sigma_z},$$

where $\Sigma_z = \{\boldsymbol{x},\, \xi(\boldsymbol{x}) = z\}$ is a (smooth) submanifold of $\mathbb{R}^n$, and σ_{Σ_z} is the Lebesgue measure on Σ_z.

1 Free energy and metastability

Co-area formula: Let X be a random variable with law $\psi(x)\,dx$ in $\mathbb{R}^n$. Then $\xi(X)$ has law $\int_{\Sigma_z} \psi \,|\nabla \xi|^{-1}\, d\sigma_{\Sigma_z}\, dz$, and the law of X conditioned to a fixed value z of $\xi(X)$ is $d\mu_z = \frac{\psi\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}{\int_{\Sigma_z}\psi\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}$.

Indeed, for any bounded functions f and g,

$$
\begin{aligned}
\mathbb{E}(f(\xi(X))g(X)) &= \int_{\mathbb{R}^n} f(\xi(x))g(x)\psi(x)\,dx,\\
&= \int_{\mathbb{R}^p}\int_{\Sigma_z} f\circ\xi\, g\,\psi\,|\nabla\xi|^{-1} d\sigma_{\Sigma_z}\,dz,\\
&= \int_{\mathbb{R}^p} f(z)\frac{\int_{\Sigma_z} g\,\psi\,|\nabla\xi|^{-1}d\sigma_{\Sigma_z}}{\int_{\Sigma_z}\psi\,|\nabla\xi|^{-1}d\sigma_{\Sigma_z}}\int_{\Sigma_z}\psi\,|\nabla\xi|^{-1}d\sigma_{\Sigma_z}\,dz.
\end{aligned}
$$

Remarks:

- The measure $|\nabla \xi|^{-1} d\sigma_{\Sigma_z}$ is sometimes denoted $\delta_{\xi(\boldsymbol{x})-z}$ in the literature.

- A is the free energy associated with the reaction coordinate or collective variable ξ (angle, length, ...). A is defined up to an additive constant, so that it is enough to compute free energy differences, or the derivative of A (the mean force).

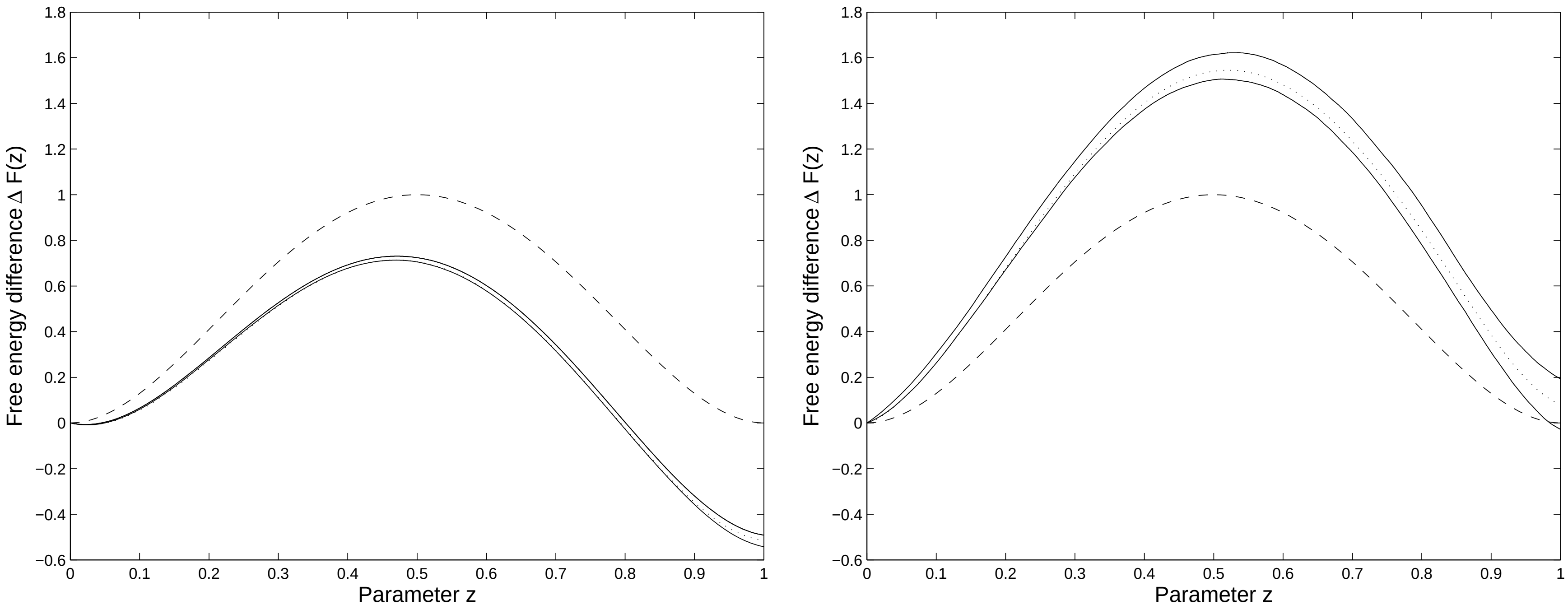
- $A(z) = -\beta^{-1} \ln Z_{\Sigma_z}$ and Z_{Σ_z} is the partition function associated with the conditioned probability measures:

$$\mu_{\Sigma_z} = Z_{\Sigma_z}^{-1} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}.$$

1 Free energy and metastability

Example of a free energy profile (solvation of a dimer)

(Profiles computed using TI)

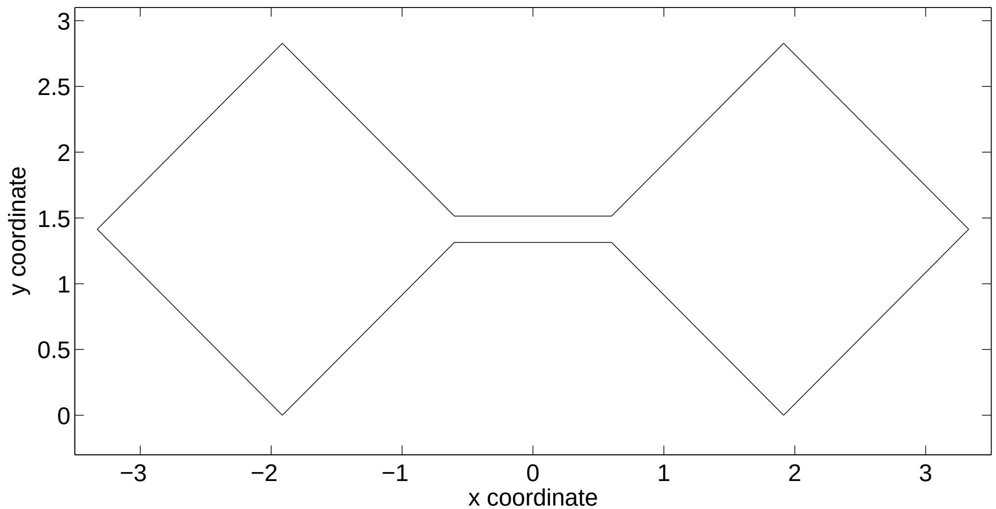


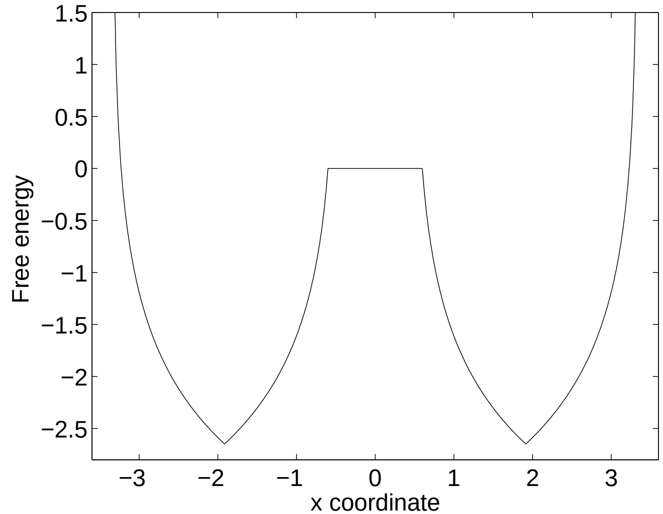
The density of the solvent molecules is lower on the left than on the right. At high density, the compact state is more likely but (claim of physicists) spontaneous transitions are less frequent (free energy barrier) ... to be better understood.

1 Free energy and metastability

Some direct numerical simulations...

Remark: Free energy is not energy !





Left: The potential is 0 in the region enclosed by the curve, and $+\infty$ outside.

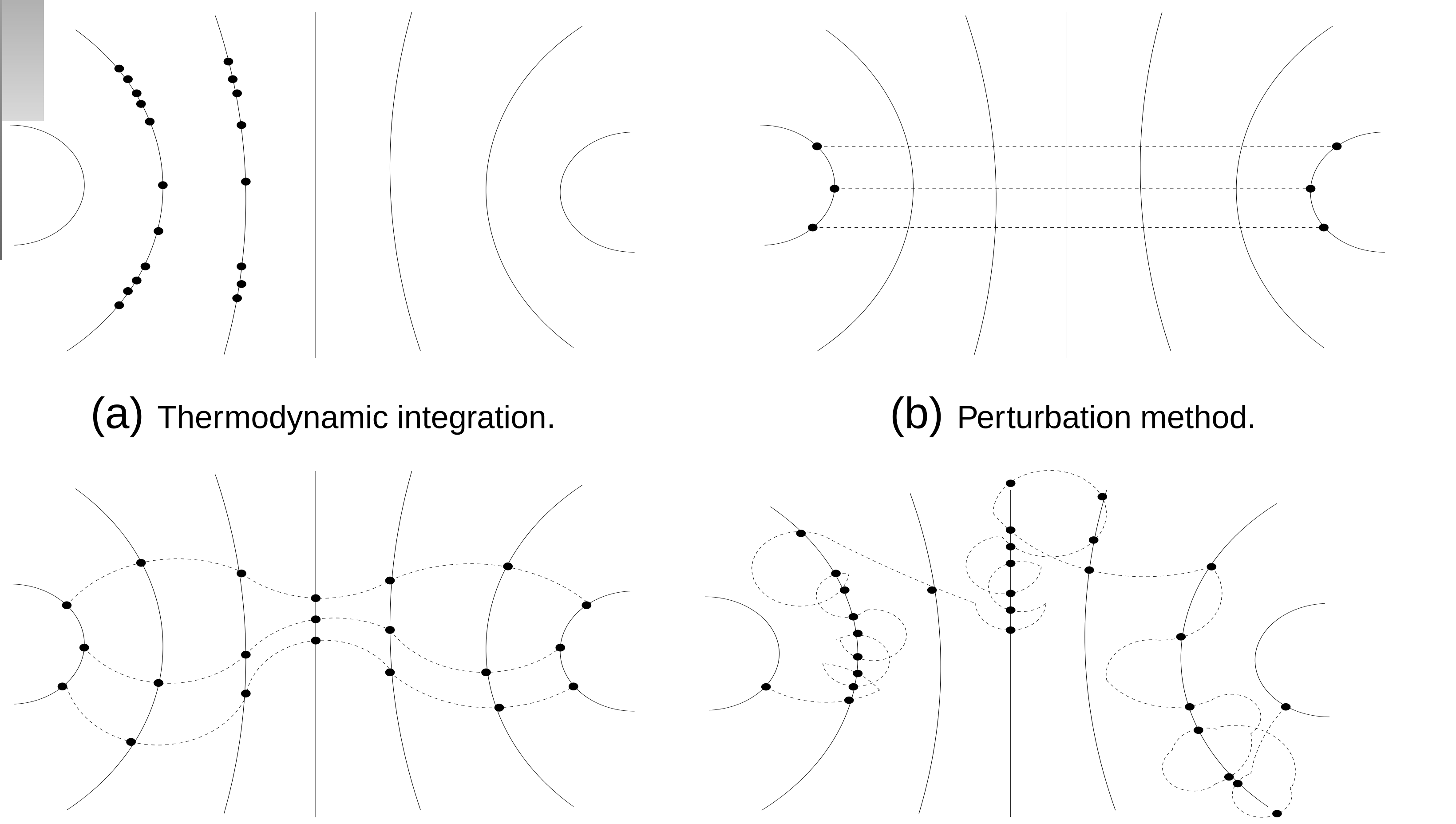
Right: Associated free energy profile when the x coordinate is the reaction coordinate ($\beta = 1$).

1 Free energy and metastability

Examples of methods to compute free energy differences $A(z_2) - A(z_1)$:

- Thermodynamic integration *(Kirkwood)* (homogeneous Markov process),
- Perturbation methods *(Zwanzig)* (importance sampling),
- Out of equilibrium dynamics *(Jarzynski)* (non-homogeneous Markov process),
- Adaptive methods *(ABF, metadynamics)* (non-homogeneous and non-linear Markov process).

1 Free energy and metastability



(a) Thermodynamic integration. (b) Perturbation method.

(c) Out of equilibrium dynamics. (d) Adaptive dynamics.

2 Constrained dynamics

Examples of methods to compute free energy differences $A(z_2) - A(z_1)$:

- Thermodynamic integration *(Kirkwood)* (homogeneous Markov process),
- Perturbation methods *(Zwanzig)* (importance sampling),
- Out of equilibrium dynamics *(Jarzynski)* (non-homogeneous Markov process),
- Adaptive methods *(ABF, metadynamics)* (non-homogeneous and non-linear Markov process).

2.1 Thermodynamic integration

Thermodynamic integration is based on two remarks:

(1) The derivative $A'(z)$ can be obtained by sampling the conditioned probability measure μ_{Σ_z} *(Sprik, Ciccotti, Kapral, Vanden-Eijnden, E, den Otter, ...)*

$$\begin{aligned}
A'(z) &= Z_{\Sigma_z}^{-1} \int \left(\frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right) \exp(-\beta V) |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \\
&= Z_{\Sigma_z}^{-1} \int \frac{\nabla \xi}{|\nabla \xi|^2} \cdot \left(\nabla \tilde{V} + \beta^{-1} \boldsymbol{H} \right) \exp(-\beta \tilde{V}) d\sigma_{\Sigma_z}, \\
&= \int f d\mu_{\Sigma_z},
\end{aligned}$$

where $\tilde{V} = V + \beta^{-1} \ln |\nabla \xi|$, $f = \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right)$

and $\boldsymbol{H} = -\nabla \cdot \left(\frac{\nabla \xi}{|\nabla \xi|} \right) \frac{\nabla \xi}{|\nabla \xi|}$ is the mean curvature vector.

2.1 Thermodynamic integration

Proof: (based on the co-area formula)

$$\int \left(\int \exp(-\beta \tilde{V}) d\sigma_{\Sigma_z} \right)' \phi(z)\, dz = -\int \int \exp(-\beta \tilde{V}) d\sigma_{\Sigma_z} \phi' \, dz,$$

$$\begin{aligned}
&= -\int \int \exp(-\beta \tilde{V}) \phi' \circ \xi \, d\sigma_{\Sigma_z} \, dz, \\
&= -\int \exp(-\beta \tilde{V}) \phi' \circ \xi |\nabla \xi| d\boldsymbol{x}, \\
&= -\int \exp(-\beta \tilde{V}) \nabla(\phi \circ \xi) \cdot \frac{\nabla \xi}{|\nabla \xi|^2} |\nabla \xi| d\boldsymbol{x}, \\
&= \int \nabla \cdot \left(\exp(-\beta \tilde{V}) \frac{\nabla \xi}{|\nabla \xi|} \right) \phi \circ \xi \, d\boldsymbol{x},
\end{aligned}$$

$$= \int \int \left(-\beta \frac{\nabla \tilde{V} \cdot \nabla \xi}{|\nabla \xi|^2} + |\nabla \xi|^{-1} \nabla \cdot \left(\frac{\nabla \xi}{|\nabla \xi|} \right) \right) \exp(-\beta \tilde{V}) d\sigma_{\Sigma_z} \phi(z)\, dz.$$

2.1 Thermodynamic integration

(2) It is possible to sample the conditioned probability measure $\mu_{\Sigma_z} = Z_{\Sigma_z}^{-1} \exp(-\beta \tilde{V}) d\sigma_{\Sigma_z}$ by considering the following constrained dynamics:

$$\text{(RCD)} \begin{cases} d\boldsymbol{X}_t = -\nabla \tilde{V}(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t + \nabla \xi(\boldsymbol{X}_t) d\Lambda_t, \\ d\Lambda_t \text{ such that } \xi(\boldsymbol{X}_t) = z. \end{cases}$$

Moreover, we have $d\Lambda_t = d\Lambda_t^{\mathrm{m}} + d\Lambda_t^{\mathrm{f}}$, with $d\Lambda_t^{\mathrm{m}} = -\sqrt{2\beta^{-1}} \frac{\nabla \xi}{|\nabla \xi|^2}(\boldsymbol{X}_t) \cdot d\boldsymbol{W}_t$ and $d\Lambda_t^{\mathrm{f}} = \frac{\nabla \xi}{|\nabla \xi|^2} \cdot \left(\nabla \tilde{V} + \beta^{-1} \boldsymbol{H} \right) (\boldsymbol{X}_t)\, dt = f(\boldsymbol{X}_t)\, dt$ so that

$$A'(z) = \lim_{T \to \infty} \frac{1}{T} \int_0^T d\Lambda_t = \lim_{T \to \infty} \frac{1}{T} \int_0^T d\Lambda_t^{\mathrm{f}}.$$

2.1 Thermodynamic integration

The free energy profile is then obtained by thermodynamic integration:

$$A(z) - A(0) = \int_0^z A'(z)\,dz \simeq \sum_{i=0}^{K} \omega_i A'(z_i).$$

2.1 Thermodynamic integration

The rigidly constrained dynamics can also be written:

$$(RCD) \quad d\boldsymbol{X}_t = P(\boldsymbol{X}_t)\left(-\nabla\tilde{V}(\boldsymbol{X}_t)\,dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}_t\right) + \beta^{-1}\boldsymbol{H}(\boldsymbol{X}_t)\,dt,$$

where $P(\boldsymbol{x})$ is the orthogonal projection operator:

$$P(\boldsymbol{x}) = \mathrm{Id} - \boldsymbol{n}(\boldsymbol{x}) \otimes \boldsymbol{n}(\boldsymbol{x}),$$

with $\boldsymbol{n}$ the unit normal vector: $\boldsymbol{n}(\boldsymbol{x}) = \dfrac{\nabla\xi}{|\nabla\xi|}(\boldsymbol{x})$.

(RCD) can also be written using the Stratonovitch product: $d\boldsymbol{X}_t = -P(\boldsymbol{X}_t)\nabla\tilde{V}(\boldsymbol{X}_t)\,dt + \sqrt{2\beta^{-1}}P(\boldsymbol{X}_t) \circ d\boldsymbol{W}_t$.

It is easy to check that $\xi(\boldsymbol{X}_t) = \xi(\boldsymbol{X}_0) = z$ for $\boldsymbol{X}_t$ solution to (RCD).

Assume $z = 0$.

μ_{Σ_0} is the unique invariant measure with support in Σ_0 for (RCD).

Proposition 1 *Let $\boldsymbol{X}_t$ be the solution to (RCD) such that the law of $\boldsymbol{X}_0$ is μ_{Σ_0}. Then, for all smooth function ϕ and for all time $t > 0$,*

$$\mathbb{E}(\phi(\boldsymbol{X}_t)) = \int \phi(\boldsymbol{x}) d\mu_{\Sigma_0}(\boldsymbol{x}).$$

Proof: Introduce the infinitesimal generator and apply the divergence theorem on submanifolds :
$\forall \boldsymbol{\phi} \in \mathcal{C}^1(\mathbb{R}^{3N}, \mathbb{R}^{3N})$,

$$\int \mathrm{div}_{\Sigma_0}(\boldsymbol{\phi}) \, d\sigma_{\Sigma_0} = -\int \boldsymbol{H} \cdot \boldsymbol{\phi} \, d\sigma_{\Sigma_0},$$

2.1 Thermodynamic integration

Discretization: These two schemes are consistent with (RCD):

$$(S1)\begin{cases} \boldsymbol{X}_{n+1} = \boldsymbol{X}_n - \nabla\tilde{V}(\boldsymbol{X}_n)\Delta t + \sqrt{2\beta^{-1}}\Delta\boldsymbol{W}_n + \lambda_n \nabla\xi(\boldsymbol{X}_{n+1}), \\ \text{with } \lambda_n \in \mathbb{R} \text{ such that } \xi(\boldsymbol{X}_{n+1}) = 0, \end{cases}$$

$$(S2)\begin{cases} \boldsymbol{X}_{n+1} = \boldsymbol{X}_n - \nabla\tilde{V}(\boldsymbol{X}_n)\Delta t + \sqrt{2\beta^{-1}}\Delta\boldsymbol{W}_n + \lambda_n \nabla\xi(\boldsymbol{X}_{n}), \\ \text{with } \lambda_n \in \mathbb{R} \text{ such that } \xi(\boldsymbol{X}_{n+1}) = 0, \end{cases}$$

where $\Delta\boldsymbol{W}_n = \boldsymbol{W}_{(n+1)\Delta t} - \boldsymbol{W}_{n\Delta t}$. The constraint is exactly satisfied (important for longtime computations). The discretization of $A'(0) = \lim_{T\to\infty} \frac{1}{T}\int_0^T d\Lambda_t$ is:

$$\lim_{T\to\infty} \lim_{\Delta t\to 0} \frac{1}{T} \sum_{n=1}^{T/\Delta t} \lambda_n = A'(0).$$

In practice, the following variance reduction scheme may be used:

$$\begin{cases} \boldsymbol{X}_{n+1} = \boldsymbol{X}_n - \nabla \tilde{V}(\boldsymbol{X}_n)\Delta t + \sqrt{2\beta^{-1}}\Delta \boldsymbol{W}_n + \lambda \nabla \xi(\boldsymbol{X}_{n+1}), \\ \text{with } \lambda \in \mathbb{R} \text{ such that } \xi(\boldsymbol{X}_{n+1}) = 0, \end{cases}$$

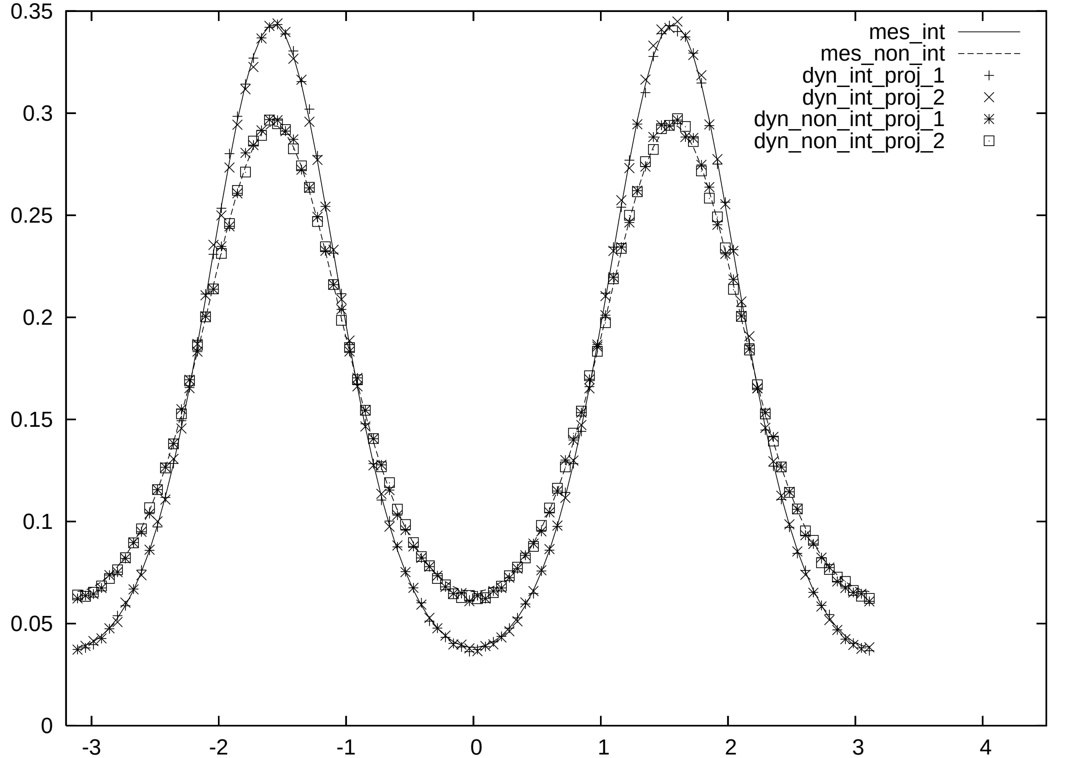
$$\begin{cases} \boldsymbol{X}_* = \boldsymbol{X}_n - \nabla \tilde{V}(\boldsymbol{X}_n)\Delta t - \sqrt{2\beta^{-1}}\Delta \boldsymbol{W}_n + \lambda_* \nabla \xi(\boldsymbol{X}_*), \\ \text{with } \lambda_* \in \mathbb{R} \text{ such that } \xi(\boldsymbol{X}_*) = 0, \end{cases}$$

and $\lambda_n = (\lambda + \lambda_*)/2$.

The martingale part $d\Lambda_t^{\mathrm{m}}$ (*i.e.* the most fluctuating part) of the Lagrange multiplier is removed.

2.1 Thermodynamic integration

An over-simplified illustration: in dimension 2,
$V(\boldsymbol{x}) = \frac{\beta^{-1}}{2}|\boldsymbol{x}|^2$ and $\xi(\boldsymbol{x}) = \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} - 1$.



Measures samples theoretically and numerically (as a function of the angle θ), with $\beta = 1$, $a = 2$, $b = 1$, $\Delta t = 0.01$, and 50 000 000 timesteps.

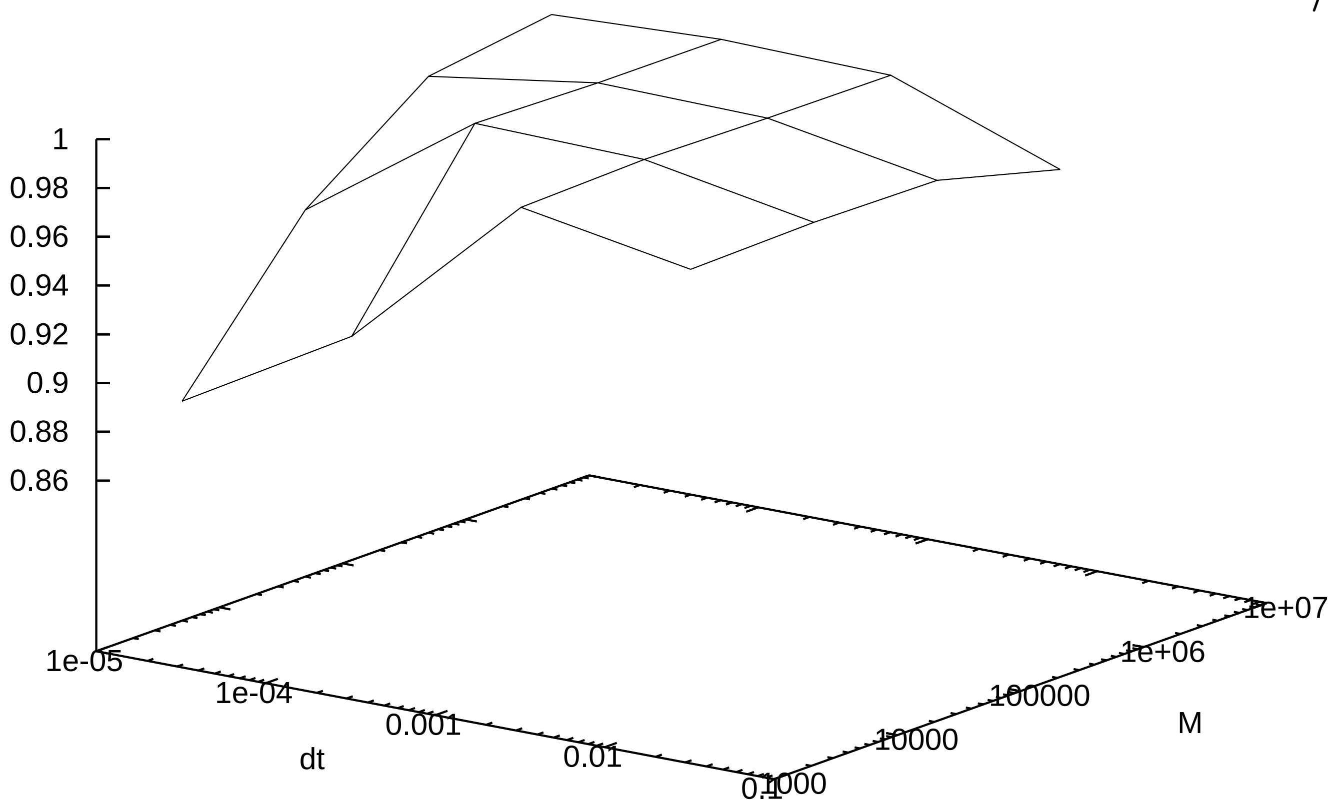
2.1 Thermodynamic integration

Computation of the mean force: $\beta = 1$, $a = 2$, $b = 1$. The exact value is: 0.9868348150. The numerical result (with $\Delta t = 0.001$, $M = 50000$) is: $[0.940613\,;\,1.03204]$.

The variance reduction method reduces the variance by a factor 100. The result (with $\Delta t = 0.001$, $M = 50000$) is: $[0.984019\,;\,0.993421]$.

2.1 Thermodynamic integration

App. mean force as a function of Δt and $M = T/\Delta t$:



A balance needs to be find between the discretization error ($\Delta t \to 0$) and the convergence in the ergodic limit ($T \to \infty$).

2.1 Thermodynamic integration

Using classical technics (Talay-Tubaro like proof), one can check that the ergodic measure $\mu_{\Sigma_0}^{\Delta t}$ sampled by the Markov chain $(\boldsymbol{X}_n)$ is an approximation of order one of μ_{Σ_0}: for all smooth functions $g : \Sigma_0 \to \mathbb{R}$,

$$\left| \int_{\Sigma_0} g d\mu_{\Sigma_0}^{\Delta t} - \int_{\Sigma_0} g d\mu_{\Sigma_0} \right| \leq C \Delta t.$$

Remarks:
- There are many ways to constrain the dynamics (GD). We chose one which is simple to discretize. We may also have used, for example (for $z = 0$)

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

where the constraint is penalized. One can show that $\lim_{\eta\to 0} X_t^\eta = X_t$ (in $L^\infty_{t\in[0,T]}(L^2_\omega)$-norm) where X_t satisfies (RCD). Notice that we used V and not $\tilde{V}$ in the penalized dynamics.

2.1 Thermodynamic integration

The statistics associated with the dynamics where the constraints are rigidly imposed and the dynamics where the constraints are softly imposed through penalization are different: “a stiff spring $\neq$ a rigid rod” *(van Kampen, Hinch,...)*.

2.1 Thermodynamic integration

- TI yields a way to compute $\int \phi(\boldsymbol{x}) d\mu(\boldsymbol{x})$:

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

$$= Z^{-1} \int_z \int_{\Sigma_z} \phi e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \, dz, \qquad \text{(co-area formula)}$$

$$= Z^{-1} \int_z \frac{\int_{\Sigma_z} \phi e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int_{\Sigma_z} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} \int_{\Sigma_z} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \, dz,$$

$$= \left(\int_z e^{-\beta A(z)} \, dz \right)^{-1} \int_z \left(\int_{\Sigma_z} \phi d\mu_{\Sigma_z} \right) e^{-\beta A(z)} \, dz.$$

ith $\Sigma_z = \{\boldsymbol{x}, \, \xi(\boldsymbol{x}) = z\}$, $A(z) = -\beta^{-1} \ln \left(\int_{\Sigma_z} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \right)$ and $\mu_{\Sigma_z} = e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z} / \int_{\Sigma_z} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}$.

- For a general SDE (with a non isotropic diffusion), the following diagram does not commute:

$$
\begin{tikzcd}
& \text{Projected continuous process} \arrow[r, "\Delta t"] & \text{Discretized projected continuous process} \arrow[dd, leftrightarrow, "?"] \\
\text{Continuous process} \arrow[ur, "\mathrm{P}_{\text{cont}}"] \arrow[dr, "\Delta t"'] & & \\
& \text{Discretized process} \arrow[r, "\mathrm{P}_{\text{disc}}"'] & \text{Projected discretized process}
\end{tikzcd}
$$

2.2 Constrained SDEs

We are interested in simulating a SDE:

$$d\boldsymbol{X}_t = b(\boldsymbol{X}_t)\,dt + \sigma(\boldsymbol{X}_t)d\boldsymbol{W}_t$$

subject to the constraint

$$q(\boldsymbol{X}_t) = 0.$$

$\boldsymbol{X}_t \in \mathbb{R}^n$, $b : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ (with $\sigma\sigma^T > 0$), $\boldsymbol{W}_t$ is a m-dimensional standard Brownian process with filtration $\mathcal{F}_t$. The functions b, σ and q are supposed to be smooth.

In this section, $q : \mathbb{R}^n \to \mathbb{R}$ and we suppose that $\forall \boldsymbol{x} \in \Sigma = \{\boldsymbol{x},\ q(\boldsymbol{x}) = 0\}$, $|\nabla q|(\boldsymbol{x}) \neq 0$.
(In the MD framework, q may be the reaction coordinate or some molecular constraints.)

2.2 Constrained SDEs: continuous level

As such, the problem is ill-posed. We want to find a $\mathcal{F}_t$-adapted process $\boldsymbol{Y}_t$ such that:

$$\begin{cases} d\boldsymbol{X}_t = b(\boldsymbol{X}_t)\,dt + \sigma(\boldsymbol{X}_t) d\boldsymbol{W}_t + d\boldsymbol{Y}_t, \\ q(\boldsymbol{X}_t) = 0, \end{cases}$$

where $d\boldsymbol{Y}_t = dA_t + S_t\,d\boldsymbol{W}_t$. Additional assumption:

dA_t and $S_t\,d\boldsymbol{W}_t$ are colinear to $D(\boldsymbol{X}_t)\nabla q(\boldsymbol{X}_t)\,dt$,

where $D(\boldsymbol{x})$ is a $n \times n$ symmetric positive matrix.

$D(\boldsymbol{x})\nabla q(\boldsymbol{x})$ is the normal to Σ at point $\boldsymbol{x}$ and can be given by some additional assumptions on the constraining term $\boldsymbol{Y}_t$ (D'alembert's principle for example).

2.2 Constrained SDEs: continuous level

By Itô's calculus on the constraint $q(\boldsymbol{X}_t) = 0$, one then obtains:

$$(SDE) \left\{ \begin{array}{ll} d\boldsymbol{X}_t = & P(\boldsymbol{X}_t)\left(b(\boldsymbol{X}_t)\,dt + \sigma(\boldsymbol{X}_t)d\boldsymbol{W}_t\right) \\ & -\frac{1}{2}\left(\nabla^2 q : (P\sigma\sigma^T P^T)\frac{D\nabla q}{||\nabla q||_D^2}\right)(\boldsymbol{X}_t)\,dt, \end{array} \right.$$

where the projection operator $P(\boldsymbol{x})$ is:

$$P(\boldsymbol{x}) = \mathrm{Id} - \frac{D(\boldsymbol{x})\nabla q(\boldsymbol{x}) \otimes \nabla q(\boldsymbol{x})}{||\nabla q(\boldsymbol{x})||^2_{D(\boldsymbol{x})}},$$

and, for any $\boldsymbol{Y} \in \mathbb{R}^n$ and any SDP matrix S,

$$||\boldsymbol{Y}||_S^2 = (\boldsymbol{Y} \cdot S\boldsymbol{Y}).$$

2.2 Constrained SDEs: discrete level

Two “natural” schemes:

$$\left\{ \begin{aligned} \widetilde{\boldsymbol{X}}_{n+1} &= \boldsymbol{X}_n + b(\boldsymbol{X}_n)\Delta t + \sigma(\boldsymbol{X}_n)\Delta \boldsymbol{W}_n, \\ \boldsymbol{X}_{n+1} &= \arg\min \left\{ ||\widetilde{\boldsymbol{X}}_{n+1} - \boldsymbol{Y}||^2_{S_n}, \boldsymbol{Y} \in \mathbb{R}^n, \; q(\boldsymbol{Y}) = 0 \right\}, \end{aligned} \right.$$

where Δt is the time step, $\Delta \boldsymbol{W}_n = \boldsymbol{W}_{(n+1)\Delta t} - \boldsymbol{W}_{n\Delta t}$ and S_n is a SDP matrix, $\mathcal{F}_{n\Delta t}$-measurable. For sufficiently small time step, this is well posed and equivalent to:

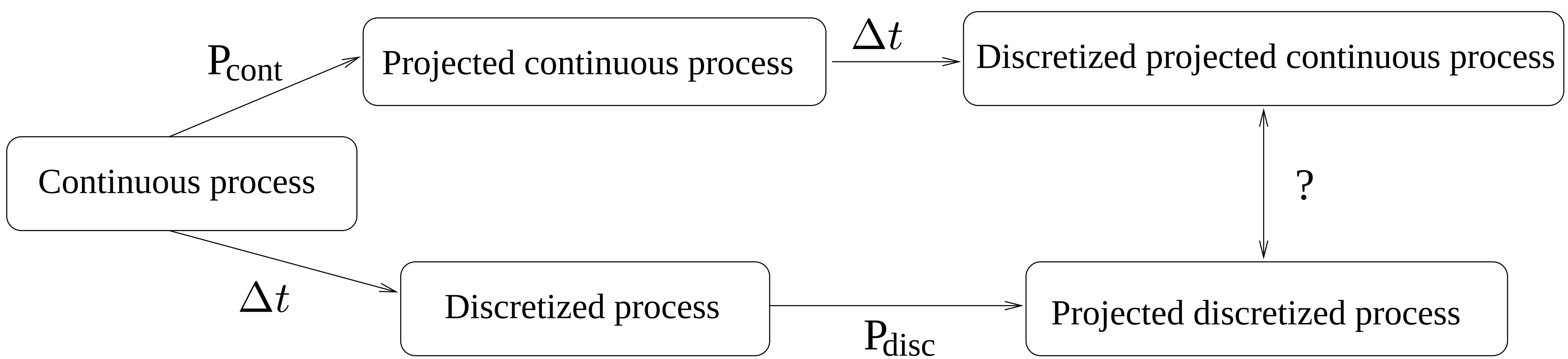
$$(S1) \left\{ \begin{array}{l} \boldsymbol{X}_{n+1} = \boldsymbol{X}_n + b(\boldsymbol{X}_n)\Delta t + \sigma(\boldsymbol{X}_n)\Delta \boldsymbol{W}_n + \lambda_n S_n^{-1} \nabla q(\boldsymbol{X}_{n+1}), \\ \text{where } \lambda_n \in \mathbb{R} \text{ is such that } q(\boldsymbol{X}_{n+1}) = 0. \end{array} \right.$$

2.2 Constrained SDEs: discrete level

A “more explicit” scheme is then:

$$(S2)\begin{cases} \boldsymbol{X}_{n+1} = \boldsymbol{X}_n + b(\boldsymbol{X}_n)\Delta t + \sigma(\boldsymbol{X}_n)\Delta \boldsymbol{W}_n + \lambda_n S_n^{-1}\nabla q(\boldsymbol{X}_n), \\ \text{where } \lambda_n \in \mathbb{R} \text{ is such that } q(\boldsymbol{X}_{n+1}) = 0. \end{cases}$$

Question: Are (S1) and (S2) consistent with (SDE) for any q and b ?



2.2 Constrained SDEs: consistency

Théorème 1

- *In the ODE case ($\sigma = 0$), (S1) and (S2) are consistent with (SDE) for any q and b iff $D \propto S^{-1}$.*
- *In the SDE case ($\sigma\sigma^T > 0$),*
  - *(S2) is consistent with (SDE) for any q and b iff $D \propto S^{-1}$.*
  - *(S1) is consistent with (SDE) for any q and b iff $D \propto S^{-1} \propto \sigma\sigma^T$.*

Proof: Expansions w.r.t. Δt and $\Delta \boldsymbol{W}_n$.

2.3 Non-equilibrium dynamics

Let us consider a stochastic process such that $\boldsymbol{X}_0 \sim \mu_{\Sigma_{z(0)}}$ and

$$\left\{ \begin{array}{rl} d\boldsymbol{X}_t = & -P(\boldsymbol{X}_t)\nabla\tilde{V}(\boldsymbol{X}_t)\,dt + \sqrt{2\beta^{-1}}P(\boldsymbol{X}_t)\circ d\boldsymbol{W}_t \\ & +\nabla\xi(\boldsymbol{X}_t)d\Lambda_t^{\text{ext}}, \\ d\Lambda_t^{\text{ext}} = & \frac{z'(t)}{|\nabla\xi(\boldsymbol{X}_t)|^2}\,dt, \end{array} \right.$$

where $z : [0,T] \to [0,1]$ is a fixed deterministic evolution of the reaction coordinate ξ, such that $z(0) = 0$ and $z(T) = 1$. The idea is to associate to each trajectory $\boldsymbol{X}_t$ a weight $\mathcal{W}(t)$ and to compute free energy differences by a Feynman-Kac formula:

$$A(1) - A(0) = -\beta^{-1}\ln\left(\mathbb{E}\left(\exp(-\beta\,\mathcal{W}(T))\right)\right).$$

2.3 Non-equilibrium dynamics

The dynamics can also be written using a Lagrange multiplier:

$$\begin{cases} d\boldsymbol{X}_t = -\nabla \tilde{V}(\boldsymbol{X}_t)\,dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}_t + \nabla\xi(\boldsymbol{X}_t)d\Lambda_t, \\ \xi(\boldsymbol{X}_t) = z(t). \end{cases}$$

And we have

$$d\Lambda_t = d\Lambda_t^{\mathrm{m}} + d\Lambda_t^{\mathrm{f}} + d\Lambda_t^{\mathrm{ext}},$$

where $d\Lambda_t^{\mathrm{m}} = -\sqrt{2\beta^{-1}}\frac{\nabla\xi}{|\nabla\xi|^2}(\boldsymbol{X}_t)\cdot d\boldsymbol{W}_t$, $d\Lambda_t^{\mathrm{f}} = f(\boldsymbol{X}_t)\,dt$ and $d\Lambda_t^{\mathrm{ext}} = \frac{z'(t)}{|\nabla\xi(\boldsymbol{X}_t)|^2}\,dt$.

Let us introduce the weight

$$\mathcal{W}(t) = \int_0^t f(\boldsymbol{X}_s) z'(s)\, ds = \int_0^t z'(s) d\Lambda^{\mathrm{f}}_s$$

. One can show that:

Théorème 2

$$A(z(t)) - A(z(0)) = -\beta^{-1} \ln \left(\mathbb{E}\left(\exp(-\beta \mathcal{W}(t))\right)\right).$$

The proof consists in introducing the semi-group associated with the dynamics

$$u(s, \boldsymbol{x}) = \mathbb{E}\left(\exp\left(-\beta \int_s^t f(\boldsymbol{X}_r^{s,\boldsymbol{x}}) z'(r)\, dr\right)\right)$$

and to show that $\frac{d}{ds} \int u(s,.) \exp(-\beta \tilde{V}) d\sigma_{\Sigma_{z(s)}} = 0$ using the divergence theorem on submanifolds.

2.3 Non-equilibrium dynamics

The discretization is (as before):

$$(S1)\begin{cases} \boldsymbol{X}_{n+1} = \boldsymbol{X}_n - \nabla\tilde{V}(\boldsymbol{X}_n)\Delta t + \sqrt{2\beta^{-1}}\Delta \boldsymbol{W}_n + \lambda_n \nabla\xi(\boldsymbol{X}_{n+1}), \\ \text{with } \lambda_n \text{ such that } \xi(\boldsymbol{X}_{n+1}) = z(t_{n+1}), \end{cases}$$

$$(S2)\begin{cases} \boldsymbol{X}_{n+1} = \boldsymbol{X}_n - \nabla\tilde{V}(\boldsymbol{X}_n)\Delta t + \sqrt{2\beta^{-1}}\Delta \boldsymbol{W}_n + \lambda_n \nabla\xi(\boldsymbol{X}_{n}), \\ \text{with } \lambda_n \text{ such that } \xi(\boldsymbol{X}_{n+1}) = z(t_{n+1}). \end{cases}$$

To extract λ_n^{f} from λ_n, one can *e.g.* compute:

$$\lambda_n^{\mathrm{f}} = \lambda_n - \frac{z(t_{n+1}) - z(t_n)}{|\nabla\xi(\boldsymbol{X}_n)|^2} + \sqrt{2\beta^{-1}}\frac{\nabla\xi}{|\nabla\xi|^2}(\boldsymbol{X}_n)\cdot\Delta\boldsymbol{W}_n.$$

Another method to compute λ_n^{f} consists in:

$$\begin{cases} \boldsymbol{X}_{n+1}^R = \boldsymbol{X}_n - \nabla \tilde{V}(\boldsymbol{X}_n)\Delta t - \sqrt{2\beta^{-1}}\Delta \boldsymbol{W}_n + \lambda_n^R \nabla \xi(\boldsymbol{X}_{n+1}^R), \\ \text{with } \lambda_n^R \text{ such that } \frac{1}{2}\Big(\xi(\boldsymbol{X}_{n+1}^R) + \xi(\boldsymbol{X}_{n+1})\Big) = \xi(\boldsymbol{X}_n). \end{cases}$$

We then have $\lambda_n^{\mathrm{f}} = \frac{1}{2}\left(\lambda_n + \lambda_n^R\right)$.
The weight is then approximated by

$$\begin{cases} \mathcal{W}_0 = 0, \\ \mathcal{W}_{n+1} = \mathcal{W}_n + \frac{z(t_{n+1}) - z(t_n)}{t_{n+1} - t_n} \lambda_n^{\mathrm{f}}, \end{cases}$$

and a (biased) estimator of the free energy difference $A(z(T)) - A(z(0))$ is $-\beta^{-1} \ln\left(\frac{1}{M}\sum_{m=1}^M \exp\left(-\beta \mathcal{W}_{T/\Delta t}^m\right)\right)$.

2.3 Non-equilibrium dynamics

In practice, the efficieny of this numerical method is not clearly demonstrated. If the transition is too fast, the variance of the estimator is very large. If the transition is slow, we are back to thermodynamic integration...

3 Adaptive methods

Examples of methods to compute free energy differences $A(z_2) - A(z_1)$:

- Thermodynamic integration *(Kirkwood)* (homogeneous Markov process),
- Perturbation methods *(Zwanzig)* (importance sampling),
- Out of equilibrium dynamics *(Jarzynski)* (non-homogeneous Markov process),
- Adaptive methods *(ABF, metadynamics)* (non-homogeneous and non-linear Markov process).

3 Adaptive methods

The principle of adaptive methods is to modify the potential seen by the particles in function of there history in order to:

- efficiently explore the free energy surface,
- compute free energy profiles.

The time dependent potential is of the form

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

where A_t is an approximation of A computed by using the history of the configurations of the systems conditioned at a given value of the reaction coordinate.

References: *Darve, Pohorille, Hénin, Chipot, Laio, Parrinello, Wang, Landau,...*

3.1 Adaptive methods: algorithms

How to update A_t ? Assume for the moment that the process is instantaneously at equilibrium $\psi = \psi^{\rm eq} \propto \exp(-\beta \mathcal{V}_t)(\boldsymbol{x})\, d\boldsymbol{x} = \exp(-\beta(V - A_t \circ \xi))(\boldsymbol{x})\, d\boldsymbol{x}$.

Recall the definition of free energy:

$$A(z) = -\beta^{-1} \ln \left(\int_{\Sigma_z} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \right) = -\beta^{-1} \ln Z_{\Sigma_z},$$

and associated mean force:

$$A'(z) = \frac{\displaystyle\int \left(\frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right) e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\displaystyle\int e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} = \int f \, d\mu_{\Sigma_z}$$

For adaptive dynamics, we replace V by $\mathcal{V}_t$ in these formulas, to get observed free energy or mean force.

3.1 Adaptive methods: algorithms

Two basic methods : estimate the free energy or the mean force at time t.

Observed free energy :

$$-\beta^{-1} \ln \left(\int_{\Sigma_z} e^{-\beta \mathcal{V}_t} |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \right) = (A - A_t),$$

Observed mean force :

$$\frac{\int \left(\frac{\nabla \mathcal{V}_t \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right) e^{-\beta \mathcal{V}_t} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int e^{-\beta \mathcal{V}_t} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} = (A' - A'_t).$$

Idea: use these expressions to update A_t (resp. A'_t) in such a way that $\lim_{t \to \infty} A'_t = A'$.

3.1 Adaptive methods: algorithms

Two basic methods : estimate the free energy or the mean force at time t.

Observed free energy :

$$\tau \frac{\partial A_t}{\partial t} = -\beta^{-1} \ln \left(\int_{\Sigma_z} e^{-\beta \mathcal{V}_t} |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \right) = (A - A_t),$$

Observed mean force :

$$\tau \frac{\partial A'_t}{\partial t} = \frac{\int \left(\frac{\nabla \mathcal{V}_t \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right) e^{-\beta \mathcal{V}_t} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int e^{-\beta \mathcal{V}_t} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} = (A' - A'_t)$$

Idea: use these expressions to update A_t (resp. A'_t) in such a way that $\lim_{t \to \infty} A'_t = A'$.

3.1 Adaptive methods: algorithms

Now, $\boldsymbol{X}_t$ is not instantaneously at equilibrium... We use the previous argument as a guideline to build updating methods (ψ^{eq} is replaced by ψ):

(ABP) $$\frac{\partial A_t}{\partial t}(z) = -\frac{1}{\tau}\beta^{-1}\ln\int \psi|\nabla\xi|^{-1}d\sigma_{\Sigma_z},$$

(ABF) $$\frac{\partial A_t'}{\partial t}(z) = \frac{1}{\tau}\left(\frac{\int f\,\psi|\nabla\xi|^{-1}d\sigma_{\Sigma_z}}{\int \psi|\nabla\xi|^{-1}d\sigma_{\Sigma_z}} - A_t'(z)\right),$$

(where $'$ denotes a derivative with respect to z).

Remark: Since $\psi \neq \psi^{\mathrm{eq}}$ (no equilibrium), ABP $\neq$ ABF.

3.1 Adaptive methods: algorithms

Consistency of the method : the stationary state yields the mean force. Indeed, if the system reaches a stationary state

$$(\psi_t(\boldsymbol{x}), A_t(z)), \longrightarrow (\psi_\infty(\boldsymbol{x}), A_\infty(z)),$$

then

$$\psi_\infty = Z^{-1} \exp(-\beta(V - A_\infty \circ \xi))$$

and we have:

- for (ABP), $0 = -\beta^{-1} \ln \int \psi_\infty |\nabla \xi|^{-1} d\sigma_{\Sigma_z}$,
- for (ABF), $0 = \frac{\int f \, \psi_\infty |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int \psi_\infty |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} - A'_\infty(z)$,

and thus, in both cases, (up to an additive constant),

$$A_\infty = A.$$

3.1 Adaptive methods: algorithms

More generally, one can consider for F_t and G_t (such that $G_t(0)=0$) two increasing functions :

(ABP) $$\frac{\partial A_t}{\partial t}(z) = F_t\left(-\beta^{-1}\ln\int \psi_t |\nabla \xi|^{-1} d\sigma_{\Sigma_z}\right),$$

The biasing potential is increased (resp. decreased) where the observed free energy is high (resp. low).

(ABF) $$\frac{\partial A'_t}{\partial t}(z) = G_t\left(\frac{\int f\,\psi_t |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int \psi_t |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} - A'_t(z)\right),$$

The biasing force is increased (resp. decreased) where the observed mean force is positive (resp. negative).

3.1 Adaptive methods: algorithms

A typical adaptive dynamics is thus (ABF):

$$\begin{cases} d\boldsymbol{X}_t = -\nabla(V - A_t \circ \xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t, \\ \dfrac{\partial A_t'}{\partial t}(z) = \dfrac{1}{\tau} \left(\mathbb{E}\left(f(\boldsymbol{X}_t) | \xi(\boldsymbol{X}_t) = z \right) - A_t'(z) \right). \end{cases}$$

In terms of the pdf ψ, we have:

$$\begin{cases} \partial_t \psi = \mathrm{div}\, \left(\nabla(V - A_t \circ \xi)\psi + \beta^{-1} \nabla \psi \right), \\ \dfrac{\partial A_t'}{\partial t}(z) = \dfrac{1}{\tau} \left(\dfrac{\int f\, \psi |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int \psi |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} - A_t'(z) \right). \end{cases}$$

3.1 Adaptive methods: algorithms

The principle of metadynamics is to extend the configuration space to $(\boldsymbol{x}, z) \in \mathbb{R}^{n+1}$ and to consider the meta-potential

$$V^k(\boldsymbol{x}, z) = V(\boldsymbol{x}) + k(z - \xi(\boldsymbol{x}))^2.$$

Then, one chooses $(\boldsymbol{x}, z) \mapsto z$ as a reaction coordinate.

In this case, $A^k(z) = -\beta^{-1} \ln \left(\frac{\int \exp(-\beta V^k(\boldsymbol{x},z))\, d\boldsymbol{x}}{\int \exp(-\beta V^k(\boldsymbol{x},z))\, d\boldsymbol{x}\, dz} \right)$.

Notice that

$$\frac{\int \exp(-\beta V^k(\boldsymbol{x}, z))\, d\boldsymbol{x}}{\int \exp(-\beta V^k(\boldsymbol{x}, z))\, d\boldsymbol{x}\, dz} = \frac{\int \exp(-\beta V(\boldsymbol{x})) \frac{\exp(-\beta k (z-\xi(\boldsymbol{x}))^2)}{\sqrt{\pi/(k\beta)}}\, d\boldsymbol{x}}{\int \exp(-\beta V(\boldsymbol{x}))\, d\boldsymbol{x}},$$

$$\xrightarrow[k\to\infty]{} \frac{\int \exp(-\beta V) |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int \exp(-\beta V(\boldsymbol{x}))\, d\boldsymbol{x}},$$

and thus $A^k \xrightarrow[k\to\infty]{} A$ (up to an additive constant).

3.1 Adaptive methods: algorithms

Four possible combinations:

	$\frac{dA'_t}{dt}$	$\frac{dA_t}{dt}$
V	ABF	ABP
V^k	M-ABF	M-ABP

In practice, to compute $\int \psi_t |\nabla \xi|^{-1} d\sigma_{\Sigma_z}$, or $\frac{\int f(x) \psi_t |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int \psi_t |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}$, one can use empirical means or longtime averaging, and various regularizations:

- in time, for example (ABF): $A'_t(z) = \frac{\kappa_\tau * \int f \, \psi_. |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\kappa_\tau * \int \psi_. |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}$, where $\kappa_\tau = 1_{t \geq 0} \exp(-t/\tau)$.
- in space: replace $|\nabla \xi|^{-1} d\sigma_{\Sigma_z}$ by $\delta_\epsilon(\xi(\boldsymbol{x}) - z)$.

3.2 Adaptive methods: convergence

Recall: the gradient dynamics

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

is metastable, and thus the ergodic limit is difficult to reach.
Is the adaptive method (ABF and $\tau = 0$)

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

better ?

3.2 Adaptive methods: convergence

How to quantify the bad behaviour of (GD) ?

1. Escape time from a potential well.
2. Asymptotice variance of the estimator.
3. “Decorrelation time”.
4. Rate of convergence of the law of X_t to μ.

In the following we use the fourth criterium.

3.2 Adaptive methods: convergence

The PDE point of view: convergence of the pdf $\psi(t,\boldsymbol{x})$ of $\boldsymbol{X}_t$ to $\psi_\infty(\boldsymbol{x}) = Z^{-1}e^{-\beta V(\boldsymbol{x})}$. ψ satisfies the Fokker-Planck equation

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

which can be rewritten as $\partial_t \psi = \operatorname{div}\left(\psi_\infty \nabla \left(\frac{\psi}{\psi_\infty}\right)\right)$.

Let us introduce the entropy

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

We have (Csiszár-Kullback inequality):

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

3.2 Adaptive methods: convergence

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

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

$$H(\psi|\psi_\infty) \leq \frac{1}{2R} I(\psi|\psi_\infty)$$

then $E(t) \leq C \exp(-2Rt)$ and thus ψ converges to ψ_∞ exponentially fast with rate R.

Metastability $\iff$ small R

3.2 Adaptive methods: convergence

We use the same technics on the adaptive dynamics:

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

Or, in terms of the pdf $\psi(t,\boldsymbol{x})$ of $\boldsymbol{X}_t$:

$$\left\{\begin{array}{l} \partial_t\psi = \mathrm{div}\ \left(\nabla(V - A_t \circ \xi)\psi + \beta^{-1}\nabla\psi\right), \\ A_t'(z) = \dfrac{\int f\psi|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}{\int \psi|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}. \end{array}\right.$$

Recall $f = \frac{\nabla V\cdot\nabla\xi}{|\nabla\xi|^2} - \beta^{-1}\mathrm{div}\ \left(\frac{\nabla\xi}{|\nabla\xi|^2}\right)$.

3.2 Adaptive methods: convergence

Theorem: Suppose

(H1) ergodicity of the microscopic variables: the conditioned probability measures μ_{Σ_z} satisfy a logarithmic Sobolev inequality LSI(ρ),

(H2) bounded coupling: $\|\nabla_{\Sigma_z} f\|_{L^\infty} < \infty$,

then

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

The rate of convergence is limited by:

- the rate r of convergence of $\overline{\psi} = \int \psi |\nabla \xi|^{-1} \, d\sigma_{\Sigma_z}$ to $\overline{\psi_\infty}$, at the macroscopic level,
- the constant ρ of LSI at the microscopic level.

3.2 Adaptive methods: convergence

To simplify the problem, let us consider the case $n = 2$, the configuration space is $\mathbb{T} \times \mathbb{R}$, and $\xi(x, y) = x$. In this case, the dynamics writes:

$$\begin{cases} d\boldsymbol{X}_t = -\nabla(V - A_t \circ \xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}}\, dW_t, \\ A_t'(x) := \frac{dA_t}{dx} = \mathbb{E}\left(\partial_x V(\boldsymbol{X}_t) \middle| \xi(\boldsymbol{X}_t) = x\right). \end{cases}$$

Or, in terms of the pdf $\psi(t, \boldsymbol{x})$ of $\boldsymbol{X}_t$:

$$\begin{cases} \partial_t \psi = \mathrm{div}\ \left(\nabla V \psi + \beta^{-1} \nabla \psi\right) - \partial_x (A_t' \psi), \\ A_t'(x) = \dfrac{\int \partial_x V(x, y) \psi(t, x, y)\, dy}{\int \psi(t, x, y)\, dy}. \end{cases}$$

3.2 Adaptive methods: convergence

Our aims are:

- to show that the metastable features of X_t^x have been eliminated,
- to show that $A_t'(x) = \frac{\int \partial_x V(x,y) \psi(t,x,y)\, dy}{\int \psi(t,x,y)\, dy}$ converges to $A'(x) = \frac{\int \partial_x V(x,y) \exp(-\beta V)(x,y)\, dy}{\int \exp(-\beta V)(x,y)\, dy}$.

Ingredient 1: It is easy to check that $\overline{\psi}(t,x) = \int \psi(t,x,y)\, dy$ satisfies a closed PDE

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

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

3.2 Adaptive methods: convergence

For the proof of convergence, we use relative entropies $H(\mu|\nu) = \int \ln\left(\frac{d\mu}{d\nu}\right) \, d\mu$ to measure the distance to equilibrium $\psi_\infty = Z^{-1} \exp(-\beta(V - A \circ \xi))$.

the total entropy $E(t) = H(\psi(t,.)|\psi_\infty)$,

the macroscopic entropy $E_M(t) = H(\overline{\psi}(t,.)|\overline{\psi_\infty})$,

and the microscopic entropy

$$E_m(t) = \int H\left(\frac{\psi(t,x,.)}{\overline{\psi}(t,x)} \Big| \frac{\psi_\infty(x,.)}{\overline{\psi_\infty}(x)}\right) \overline{\psi}(t,x) \, dx.$$

Ingredient 2: Notice that $E = E_M + E_m$. We know that E_M goes to zero: it remains to consider E_m.

3.2 Adaptive methods: convergence

Ingredient 3: We have (algebraïc miracle)

$$\partial_t E_m = \partial_t E - \partial_t E_M$$

$$\leq -\beta^{-1} \iint \left| \partial_y \ln \left(\frac{\psi}{\psi_\infty} \right) \right|^2 \psi - \int \partial_x \ln \left(\frac{\overline{\psi}}{\overline{\psi_\infty}} \right) \overline{\psi} (A'_t - A').$$

Using

(H1) the conditioned probability measures $\frac{\psi_\infty(x,y)}{\overline{\psi_\infty}(x)} \, dy$ satisfy a logarithmic Sobolev inequality LSI(ρ), then

$$-\beta^{-1} \iint \left| \partial_y \ln \left(\frac{\psi}{\psi_\infty} \right) \right|^2 \psi \leq -2\rho\beta^{-1} E_m.$$

3.2 Adaptive methods: convergence

(H1) also implies a Talagrand inequality (Ingredient 4):

$$
\begin{aligned}
&\left|A_t'(x) - A'(x)\right| \\
&= \left| \int \partial_x V(x,y) \frac{\psi(t,x,y)}{\int \psi(t,x,y)\,dy}\,dy - \int \partial_x V(x,y) \frac{\psi_\infty(x,y)}{\int \psi_\infty(x,y)\,dy}\,dy \right|, \\
&\leq \|\partial_{x,y} V\|_{L^\infty} \int \left|y - y'\right| \pi_{t,x}(dy, dy'), \\
&\leq \|\partial_{x,y} V\|_{L^\infty} \sqrt{\frac{2}{\rho} H\left(\frac{\psi(t,x,.)}{\overline{\psi}(t,x)} \middle| \frac{\psi_\infty(x,.)}{\overline{\psi_\infty}(x)} \right)},
\end{aligned}
$$

where $\pi_{t,x}$ is any coupling measure:

$$
\int (f(y) + g(y'))\pi_{t,x}(dy, dy') = \int f(y) \frac{\psi(t,x,y)}{\int \psi(t,x,y)\,dy}\,dy + \int g(y') \frac{\psi_\infty(x,y')}{\int \psi_\infty(x,y)\,dy}\,dy'.
$$

This requires (H2) $\partial_{x,y} V \in L^\infty$.

3.2 Adaptive methods: convergence

Thus, we have

$$-\int \partial_x \ln\left(\frac{\overline{\psi}}{\overline{\psi_\infty}}\right)\overline{\psi}(A'_t - A') \le \sqrt{\int |A'_t - A'|^2\,\overline{\psi}}\sqrt{\int \left|\partial_x \ln\left(\frac{\overline{\psi}}{\overline{\psi_\infty}}\right)\right|^2 \overline{\psi}}$$

$$\le \|\partial_{x,y}V\|_{L^\infty}\sqrt{\frac{2}{\rho}E_m}C\exp(-4\pi^2\beta^{-1}t).$$

We have proved that

$$\partial_t E_m \le -2\rho\beta^{-1}E_m + \|\partial_{x,y}V\|_{L^\infty}\sqrt{\frac{2}{\rho}E_m}C\exp(-4\pi^2\beta^{-1}t),$$

and this yields $\sqrt{E_m}(t) \le C\exp(-\beta^{-1}\min(\rho, 4\pi^2)t)$.

3.2 Adaptive methods: convergence

Then, the mean force A'_t observed at time t converges to the mean force A' in the following sense:

$$\int |A'_t - A'|^2(z)\overline{\psi}(t,z)\,dz \le C\exp(-2\beta^{-1}\min(\rho,4\pi^2)t),$$

and thus, $\exists t^*, C^* > 0$, $\forall t \ge t^*$,

$$\|A'_t - A'\|_{L^2} \le C^*\exp(-\beta^{-1}\min(\rho,4\pi^2)t).$$

3.2 Adaptive methods: convergence

These arguments can be generalized to prove the theorem in the following frameworks:

- $\xi : \mathbb{R}^n \to \mathbb{T}$ (with a slight modification of the dynamics),
- $\xi : \mathbb{R}^n \to \mathbb{R}$ (with a slight modification of the dynamics and a constraining potential on $\xi(\boldsymbol{x})$),
- $\xi : \mathbb{R}^n \to \mathbb{T}^m$ or $\xi : \mathbb{R}^n \to \mathbb{R}^m$ under an additional orthogonality assumption: $\nabla \xi_i \cdot \nabla \xi_j = 0$ for $i \neq j$,
- $\xi : \mathbb{R}^n \to \mathbb{T}^m$ or $\xi : \mathbb{R}^n \to \mathbb{R}^m$ with the original ABF dynamics, without orthogonality condition, if the coupling is small enough.

3.2 Adaptive methods: convergence

The case $\xi : \mathbb{R}^n \to \mathbb{R}$: the convergence result holds for the following adaptive dynamics:

$$d\boldsymbol{X}_t = -\nabla\Big(V - \beta^{-1}\ln(|\nabla\xi|^{-2}) - A_t \circ \xi + \Pi \circ \xi\Big)(\boldsymbol{X}_t)|\nabla\xi|^{-2}(\boldsymbol{X}_t)\,dt + \sqrt{2\beta^{-1}}|\nabla\xi|^{-1}(\boldsymbol{X}_t)\,dW_t,$$

$$A_t'(z) = \mathbb{E}\left(\left(\frac{\nabla V\cdot\nabla\xi}{|\nabla\xi|^2} - \beta^{-1}\mathrm{div}\left(\frac{\nabla\xi}{|\nabla\xi|^2}\right)\right)(\boldsymbol{X}_t)\middle|\xi(\boldsymbol{X}_t) = z\right).$$

The blue terms are naturally required to obtain a closed parabolic PDE on $\overline{\psi}(t,z) = \int_{\Sigma_z}|\nabla\xi|^{-1}\psi(t,.)d\sigma_{\Sigma_z}$:

$$\partial_t\overline{\psi} = \partial_z(\Pi'\,\overline{\psi} + \beta^{-1}\partial_z\overline{\psi}).$$

The green term is required for $\overline{\psi}$ to converge to a stationary state.

3.2 Adaptive methods: convergence

Side result: The techniques of proof can be used to prove the following result (generalization of a result by F. Otto and M. Reznikoff):

For a measure μ and a function ξ, assume

- LSI for the conditioned measures $\mu(\cdot|\xi(x) = z)$,
- LSI for the marginal $\overline{\mu}(dz)$,
- bounded coupling ($\|\nabla_{\Sigma_z} f\|_{L^\infty} < \infty$),

then the measure μ satisfies a LSI.

3.3 Multiple replicas implementations

In these adaptive methods, an implementation using many replicas (a system of interacting particles) is natural to compute the conditional expectations by empirical means.

The numerical method is thus very easy to parallelize, with a small amount of information to pass from one node to the other.

One additional interest: A selection mechanism may be added to duplicate “innovative particles” and kill “redundant particles”.

3.3 Multiple replicas implementations

Numerical analysis of the particle system

Theorem: We suppose that the configuration space is $\mathbb{T}^d$, V is smooth, and $\xi(\boldsymbol{x}) = x^1$. We consider the following particle approximation:

$$d\boldsymbol{X}_{t,n,N} = \left(-\nabla V(\boldsymbol{X}_{t,n,N}) + \frac{\sum_{m=1}^{N} \phi_\epsilon^\alpha (X_{t,n,N}^1 - X_{t,m,N}^1) \partial_1 V(\boldsymbol{X}_{t,m,N})}{\sum_{m=1}^{N} \phi_\epsilon^\alpha (X_{t,n,N}^1 - X_{t,m,N}^1)} \mathbf{e}_1 \right) dt + \sqrt{2} d\boldsymbol{W}_t^n$$

where $\phi_\epsilon^\alpha = \alpha + \epsilon^{-1} \phi(\epsilon^{-1} \cdot)$. Then we have

$$\int_0^T \left\| \frac{\sum_{m=1}^{N} \phi_\epsilon^\alpha (\cdot - X_{t,m,N}^1) \partial_1 V(\boldsymbol{X}_{t,m,N})}{\sum_{m=1}^{N} \phi_\epsilon^\alpha (\cdot - X_{t,m,N}^1)} - A_t' \right\|_{L^\infty_{\mathbb{T}}} dt$$

$$= O\left(\sqrt{\alpha} + \epsilon^{1/4} + \frac{\exp\left(\frac{K}{\alpha \epsilon^2}\right)}{\alpha^2 \epsilon^3} \frac{1}{\sqrt{N}} \right).$$

3.3 Multiple replicas implementations

The selection mechanism

On the ABF dynamics, a selection mechanism can enhance the diffusion at the “macroscopic“ level.

$$\begin{cases} \partial_t \psi = \mathrm{div}\ \left(|\nabla \xi|^{-2} \left(\nabla(V - A_t \circ \xi)\psi + \beta^{-1} \nabla \psi\right)\right) + W_{\overline{\psi}} \circ \xi\, \psi, \\ A_t'(z) = \int_{\Sigma_z} \left(\frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div}\ \left(\frac{\nabla \xi}{|\nabla \xi|^2}\right)\right) |\nabla \xi|^{-1} \psi(t,.) d\sigma_{\Sigma_z} \\ \qquad \times \left(\int_{\Sigma_z} |\nabla \xi|^{-1} \psi(t,.) d\sigma_{\Sigma_z} \right)^{-1}. \end{cases}$$

Then, we have: $\partial_t \overline{\psi} = \beta^{-1} \partial_{z,z} \overline{\psi} + W_{\overline{\psi}}\, \overline{\psi}$.

3.3 Multiple replicas implementations

How to choose W? A typical choice :

$$W_{\overline{\psi}} = c \frac{\partial_{z,z}\overline{\psi}}{\overline{\psi}}$$

so that

$$\partial_t \overline{\psi} = (\beta^{-1} + c)\partial_{z,z}\overline{\psi}.$$

The rate of convergence of $\overline{\psi}$ to $\overline{\psi_\infty}$, at the “macroscopic“ level, is thus enhanced.

Numerically, it amounts to associate a weight

$$w_{n,N}(t) = \exp\left(\int_0^t W_{\overline{\psi}}(\xi(\boldsymbol{X}_{s,n,N}))\, ds \right)$$

to the n-th replica trajectory, and to make weighted means to compute A'_t.

3.3 Multiple replicas implementations

We use an histogram to discretize $\overline{\psi}$ and thus

$$W_{\overline{\psi}}(z) \simeq c \frac{\overline{\psi}(z+\delta z) - 2\overline{\psi}(z) + \overline{\psi}(z-\delta z)}{\overline{\psi}(z)\delta z^2}$$

$$\simeq \frac{3c}{\overline{\psi}(z)\delta z^2} \left(\frac{\overline{\psi}(z+\delta z) + \overline{\psi}(z) + \overline{\psi}(z-\delta z)}{3} - \overline{\psi}(z) \right)$$

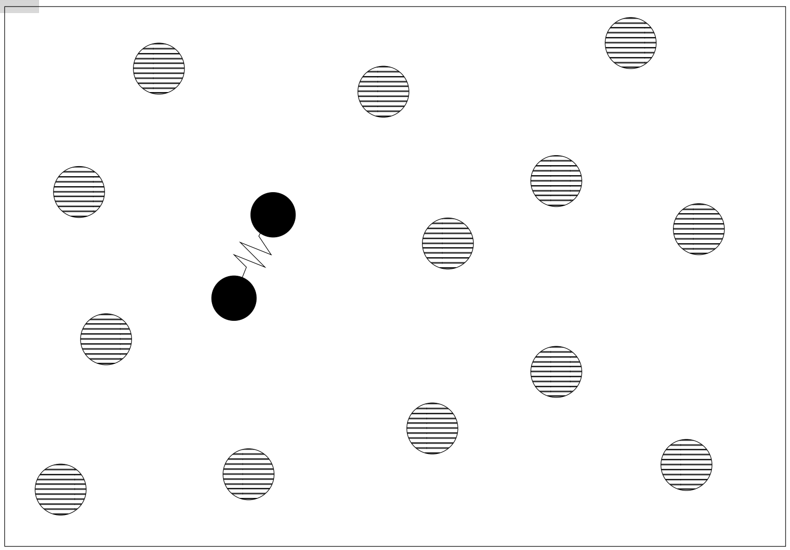
Weights of particles in locally under-explored regions are increased.

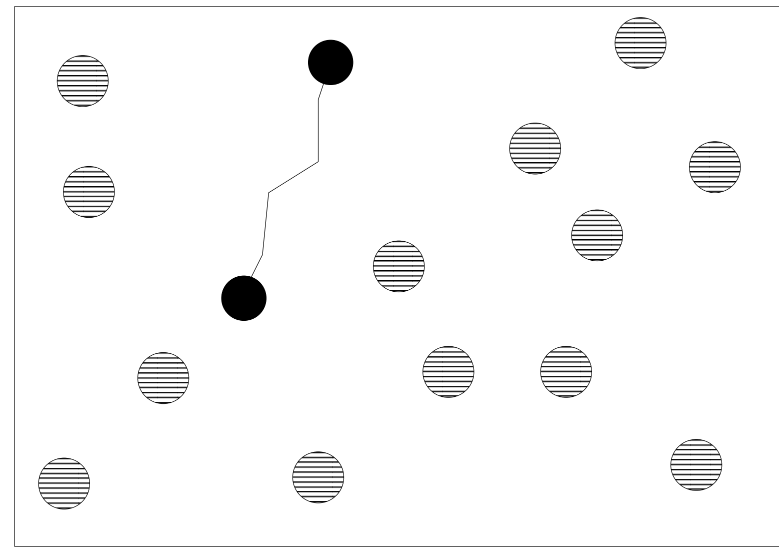
An adequate selection process can then be implemented, using these weights (like in genetic algorithm).

This should help to efficiently detect and take advantage of rare events.

3.3 Multiple replicas implementations

Tests on the numerical example *(Dellago, Geissler)*:
Influence of the solvation on a dimer conformation.

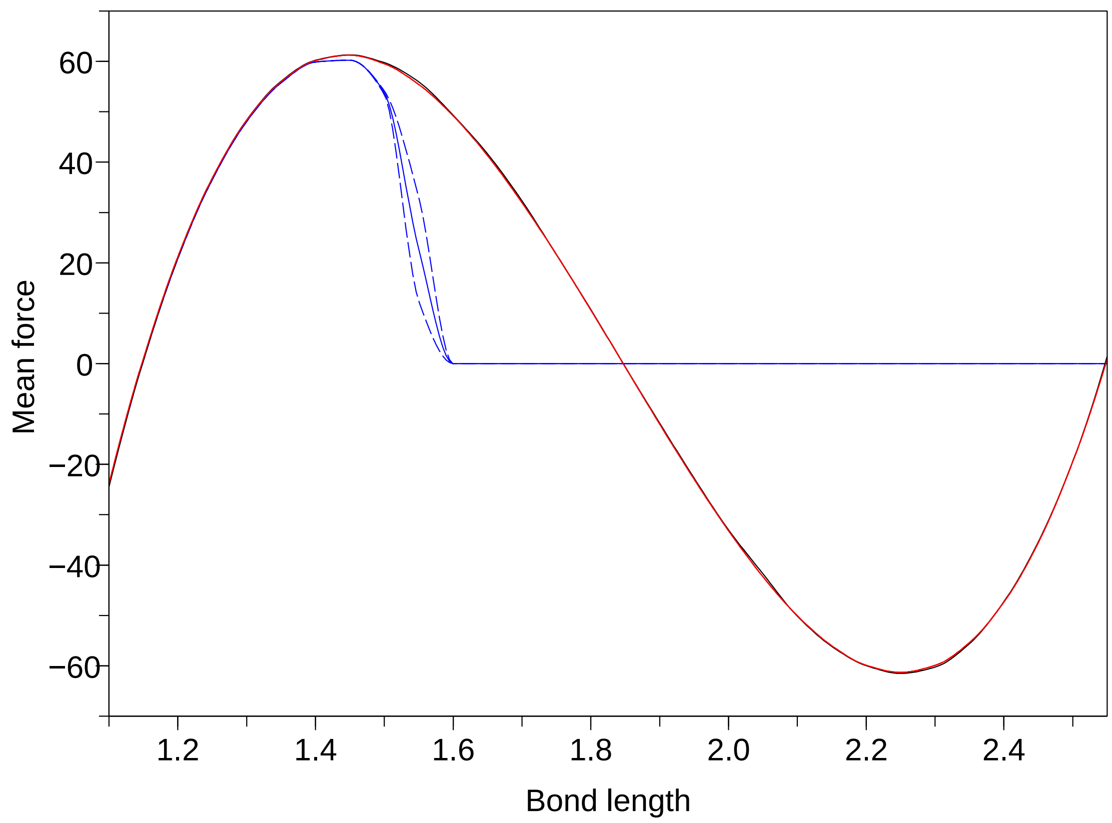




Left: compact state ($\xi = 0$). Right: stretched state ($\xi = 1$).

3.3 Multiple replicas implementations

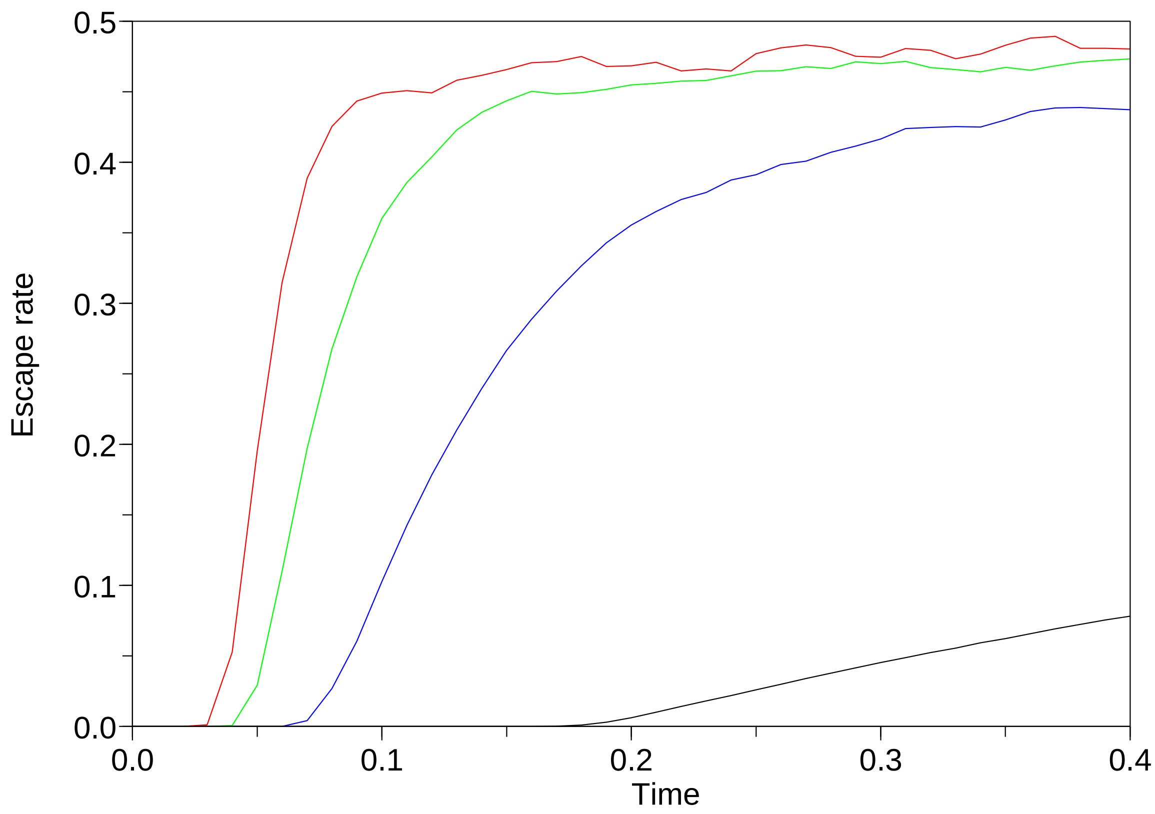
Free energy profile with parallel ABF obtained at t = 0.1, with 2000 replicas.



Red: with selection (c = 10); Blue: without selection
Dashed lines: 95 % confidence interval.

3.3 Multiple replicas implementations

Proportion of replicas which have crossed the free energy barrier.



Black: without selection; Blue: c=2; Green: c=5; Red: c=10.

3.3 Multiple replicas implementations

Another numerical experiment: ergodic averaging or empirical means ?

Compare:

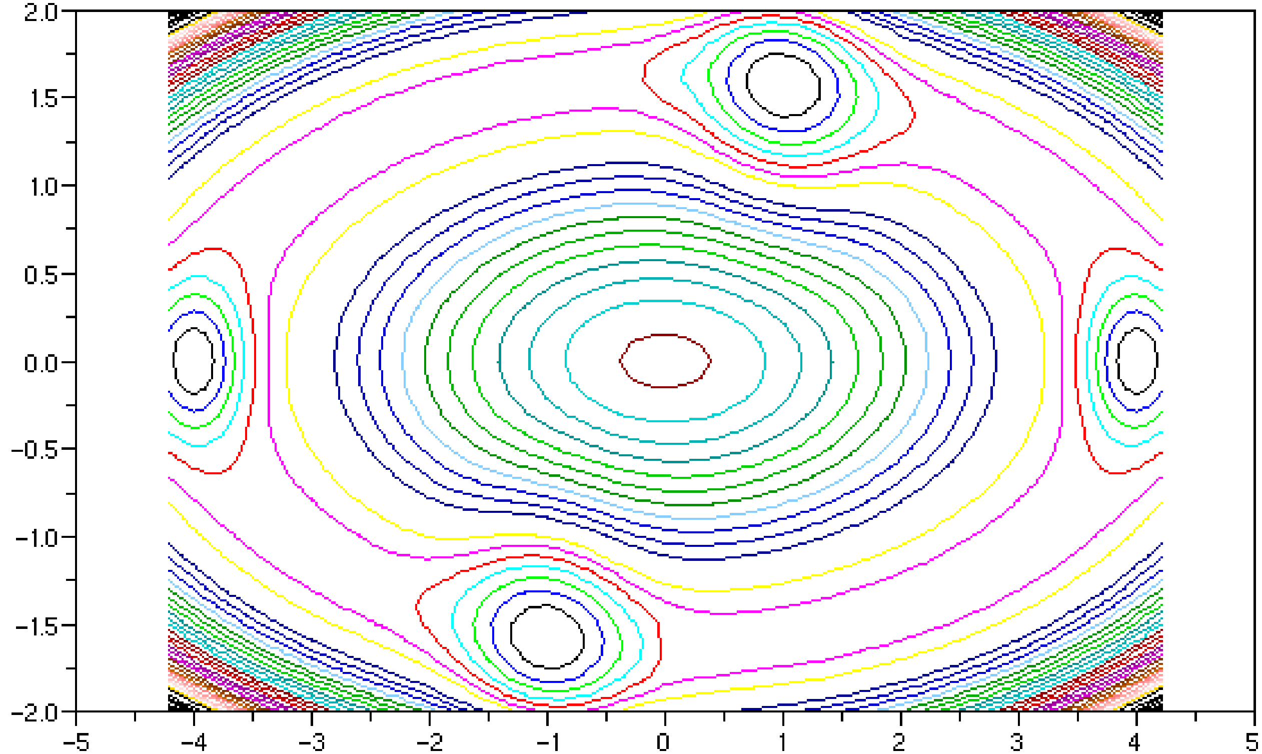
- one replica and ergodic averaging with exponential memory kernel:

$$A_t'(z) = \frac{\int_0^t \exp(-(t-s)/\tau) f(X_s) \delta_{\xi(X_s)-z} \, ds}{\int_0^t \exp(-(t-s)/\tau) \delta_{\xi(X_s)-z} \, ds}$$

- with many (10 000) replicas and empirical means to compute $A_t'(z)$.

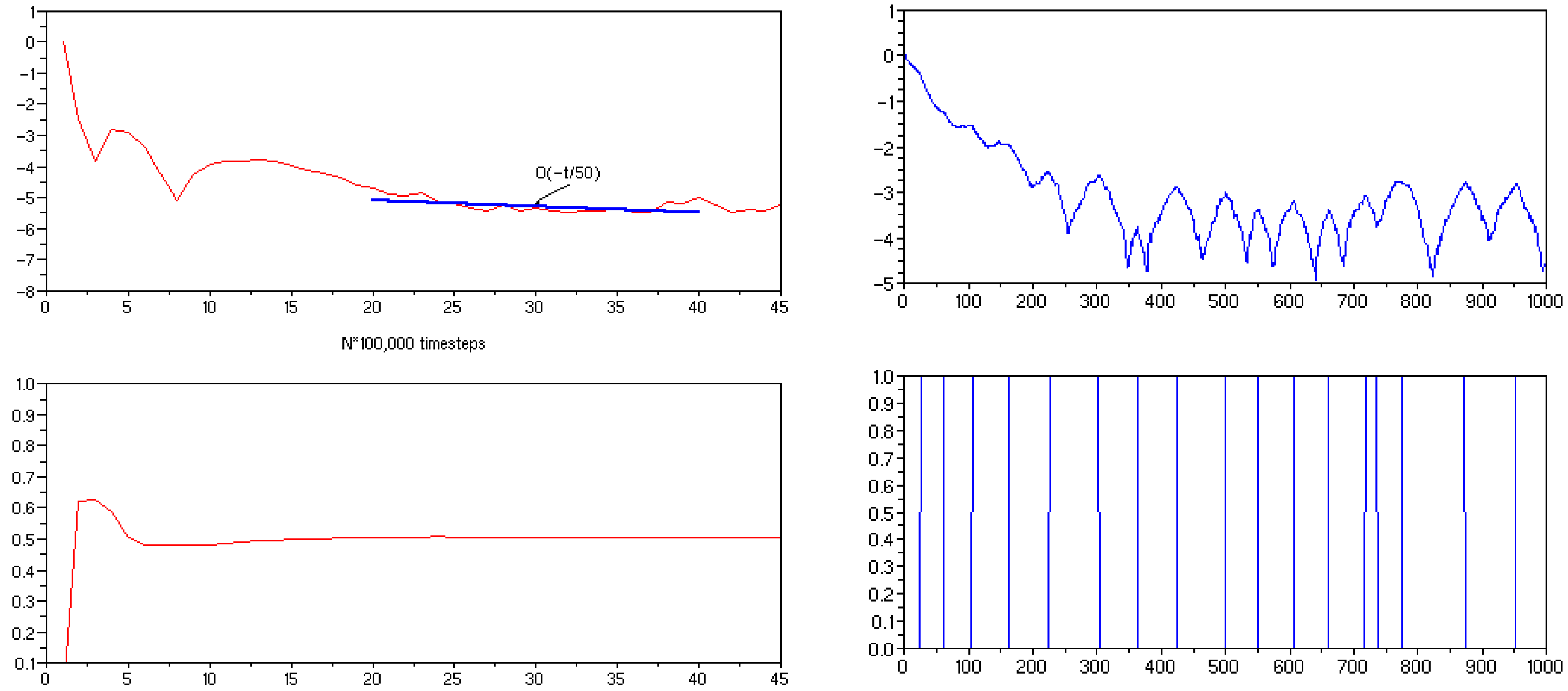
3.3 Multiple replicas implementations

Numerical experiments on a toy example (2d): the bi-channel case (hopefully representative of the case when a “bad” reaction coordinate has been chosen).



Reaction coodinate: $\xi(x,y) = x$.

3.3 Multiple replicas implementations



Left: many replicas. Right: one replica. Top: error (log scale). Bottom: (mean) position of the particles.
Works in progress:

- the case of ergodic averaging with $\tau = \infty$,
- Rate of convergence for many particles.

3.4 Application to Bayesian statistics

- Hidalgo stamp problem: the thickness of $N_{\rm data} = 485$ stamps are measured, and the corresponding histogram is approximated by a mixture of N Gaussians:

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

- parameters describing the mixture ($q_N = 1 - \sum_{i=1}^{N-1} q_i$): $x = (q_1, \ldots, q_{N-1}, \mu_1, \ldots, \mu_N, v_1, \ldots, v_N) \in \mathcal{S}_{N-1} \times [\mu_{\rm min}, \mu_{\rm max}]^N \times [v_{\rm min}, +\infty) \subset \mathbb{R}^{3N-1}$, where $\mathcal{S}_{N-1} = \left\{ (q_1, \ldots, q_{N-1}) \ \middle|\ 0 \le q_i \le 1,\ \sum_{i=1}^{N-1} q_i \le 1 \right\}$.

3.4 Application to Bayesian statistics

- the likelihood of observing the data $\{y_i, 1 \leq i \leq N_{\mathrm{data}}\}$ is

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

- potential $V = V_{\text{prior}} + V_{\text{likelihood}}$ such that the probability of a given configuration is proportional to $\exp(-V)$ where $V_{\text{prior}}(x) = \sum_{i=1}^{N} \frac{1}{2}\mu_i^2 + a \ln v_i + \frac{b}{v_i}$ with $a = 0.9$ and $b = 10^{-5}$, while the likelihood part is

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

3.4 Application to Bayesian statistics

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

This is a metastable measure.

Idea: use ABF together with a Metropolis Hasting algorithm, using $\xi(x) = q_1$ as a reaction coordinate. We use fixed gaussian proposals $T(x, x')$.

3.4 Application to Bayesian statistics

Algorithm: Metropolis Hasting-ABF.

Iterate on $n \geq 0$

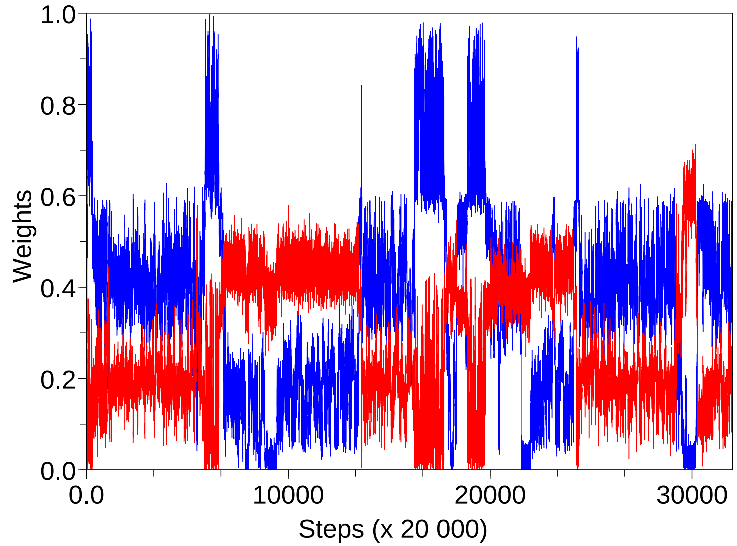
1. Update the biasing potential by computing and then integrating $(A^{n+1})'$.
2. Propose a move from x^n to y^{n+1} according to $T(x^n, y^{n+1})$;
3. Acceptance rate

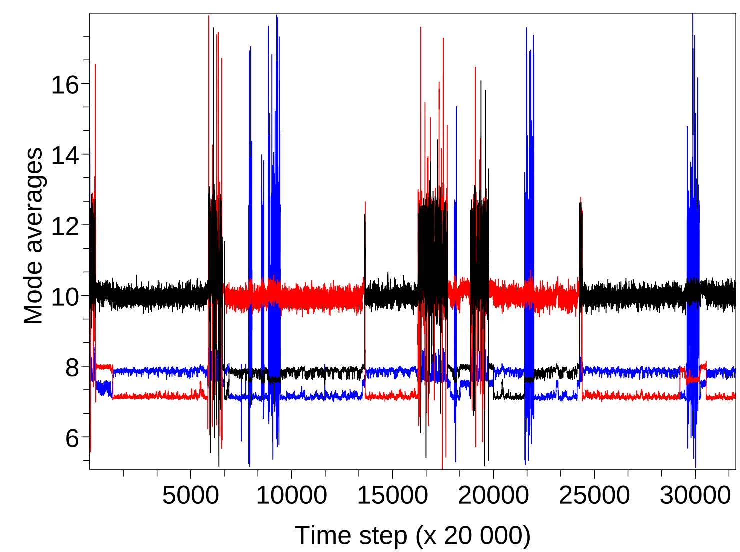
$$\alpha^n = \min\left(\frac{\pi_{A^{n+1}}(y^{n+1})\, T(y^{n+1}, x^n)}{\pi_{A^{n+1}}(x^n)\, T(x^n, y^{n+1})}, 1\right),$$

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

3.4 Application to Bayesian statistics

Some results for $N = 3$.



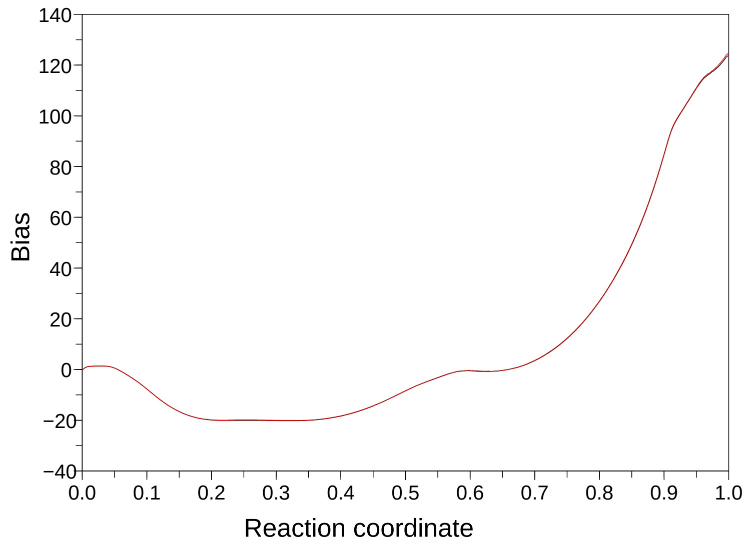


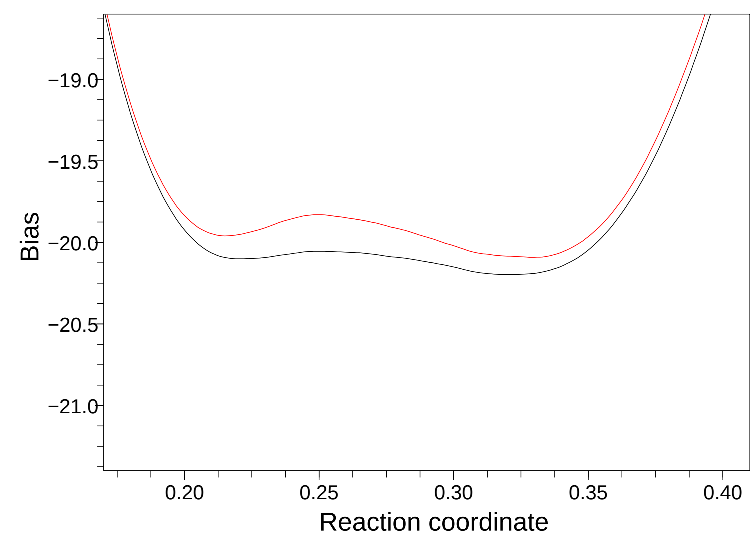
Left: Evolution of the weights q_1 (reaction coordinate, blue) and q_2.

Right: Evolution of the averages μ_1, μ_2 and μ_3.

3.4 Application to Bayesian statistics

Convergence for the bias.

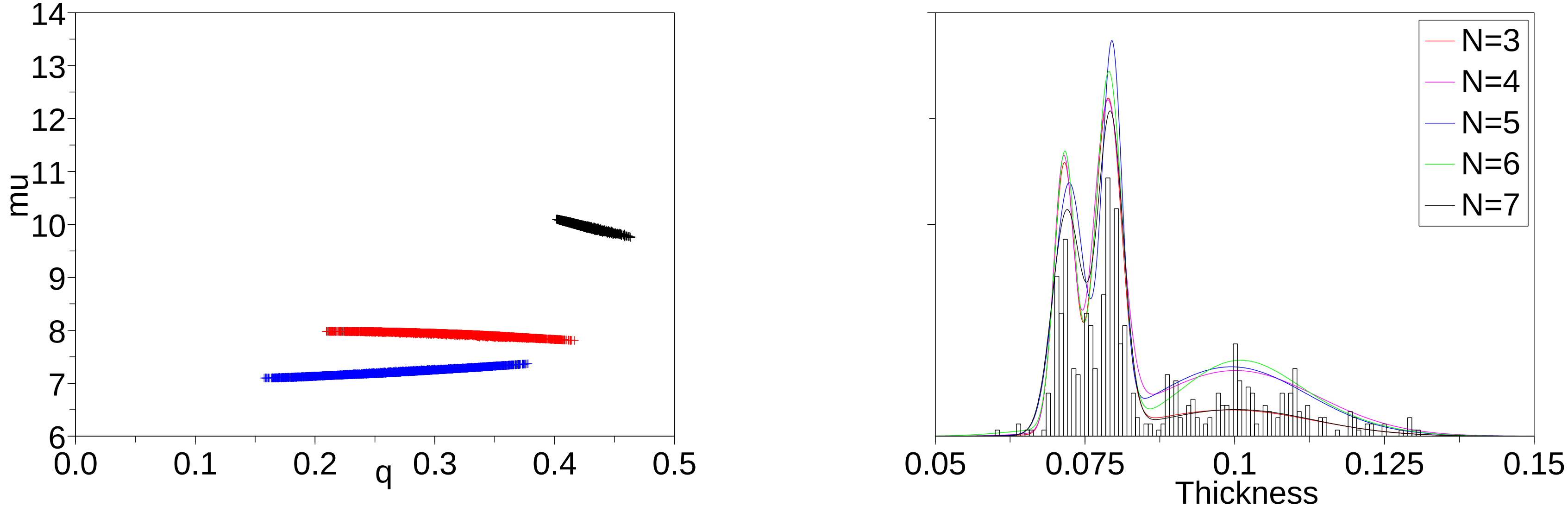




Bias after $N = 5 \times 10^8$ steps (black) and after $N = 2.5 \times 10^8$ steps (red).

3.4 Application to Bayesian statistics

Comparison of the mixture with the datas, after minimization of the configurations with lowest energies.



Left: Distribution of couples (q_i, μ_i) ($N = 3$). Right: Mixture densities obtained for $N = 3, \dots, 7$.

Conclusion

SDEs with constraints:

- The discretization of the projected dynamics may be different from the projection of the discretized dynamics,
- Constraining the dynamics with “rigid bonds” is different from constraining the dynamics with “springs”,
- The mean force can be computed by averaging the Lagrange multipliers associated with the constraints,
- The free energy differences can be obtained by non-equilibrium stochastic dynamics.

Conclusion

We proposed a unified formulation of adaptive methods using conditional distributions.

Theoretically, this allows a proof of convergence in the longtime limit for a certain class of algorithm (ABF-like algorithms). The rate of convergence is related to the logarithmic Sobolev inequality constant of the conditioned Boltzmann-Gibbs probability measures at fixed values of the reaction coordinate.

Numerically, the conditional distributions are naturally approximated by empirical means on many replicas. We have shown how a selection mechanism on the replicas can speed up the computation.

Conclusion

These techniques can be seen as adaptive importance sampling methods. They may be applied more generally to the sampling of metastable potentials, as soon as some knowledge of the directions of metastability is assumed.

Work in progress: Metropolis Hasting for measures on submanifolds, constrained Langevin equations, generalized adaptive importance sampling methods, effective dynamics and free energy...

Acknowledgements

These are joint (on-going) works with :

- C. Chipot (CNRS Nancy)
- B. Jourdain, C. Le Bris, K. Minoukadeh, R. Roux, G. Stoltz (CERMICS)
- F. Otto (Bonn)
- M. Rousset (INRIA Lilles)
- E. Vanden Eijnden (NYU)

References

Monographs on numerical methods in molecular dynamics:

- M.P. Allen and D.J. Tildesley, *Computer simulation of liquids*, Oxford Science Publications, 1987.
- C. Chipot and A. Pohorille, *Free energy calculations*, Springer, 2007.
- D. Frankel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Academic Press, 2002.
- B. Leimkuhler and S. Reich, *Simulating Hmiltonian dynamics*, Cambridge University Press, 2004.

References

- TI and constrained dynamics:

- G. Ciccotti, TL et E. Vanden-Eijnden, *Sampling Boltzmann-Gibbs distributions restricted on a manifold with diffusions*, CPAM, **61**(3), 371-408, (2008).
- C. Le Bris, TL et E. Vanden-Eijnden, *Analysis of some discretization schemes for constrained Stochastic Differential Equations*, C. R. Acad. Sci. Paris, Ser. I, **346**(7-8), 471-476, (2008).
- E. Faou et TL, *Conservative stochastic differential equations: Mathematical and numerical analysis*, to appear in Mathematics of Computation.

- Out of equilibrium methods:

- M. Rousset et G. Stoltz, *Equilibrium sampling from nonequilibrium dynamics*, J. Stat. Phys., **123** (6), 1251-1272, (2006).
- TL, M. Rousset et G. Stoltz, *Computation of free energy differences through nonequilibrium stochastic dynamics: the reaction coordinate case*, J. Comp. Phys., **222**(2), 624-643, (2007).

References

- Adaptive methods:

- TL, M. Rousset et G. Stoltz, *Computation of free energy profiles with adaptive parallel dynamics*, J. Chem. Phys. **126**, 134111 (2007).
- TL, M. Rousset et G. Stoltz, *Long-time convergence of the Adaptive Biasing Force method*, Nonlinearity, **21**, 1155-1181 (2008).
- *TL, A general two-scale criteria for logarithmic Sobolev inequalities*, to appear in Journal of Funtional Analysis.