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 Non-equilibrium dynamics,

3 Adaptive methods,

  3.2 Adaptive methods: algorithms,

  3.2 Adaptive methods: convergence,

  3.3 Multiple replicas implementations,

  3.4 Extension and another application.

1 Free energy and metastability

The aim of molecular dynamics computations is to evaluate numerically macroscopic quantities from models at the microscopic scale.

Some examples of macroscopic quantities:

- thermodynamics quantities: stress, heat capacity, free energy;
- dynamical quantities: diffusion coefficients, viscosity, transition rates.

Many applications in various fields: biology, physics, chemistry, materials science. Molecular dynamics computations consume today a lot of CPU time.

1 Free energy and metastability

A molecular dynamics model amounts essentially in choosing a potential V which associates to a configuration $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N) = \boldsymbol{x} \in \mathbb{R}^{3N}$ an energy $V(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$.
In the NVT ensemble, configurations are distributed according to 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 averages with respect to μ.

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 $V_{LJ}(r) = 4\epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right)$ is the Lennard-Jones potential.

Difficulties: (i) high-dimensional problem ($N \gg 1$) ; (ii) μ is a multimodal measure.

1 Free energy and metastability

To sample μ, Markov Chain Monte Carlo methods are used.

A typical example is the *over-damped Langevin* (or gradient) dynamics:

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

Under suitable assumption, we have the ergodic property: for μ-a.e. $\boldsymbol{X}_0$,

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

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*:

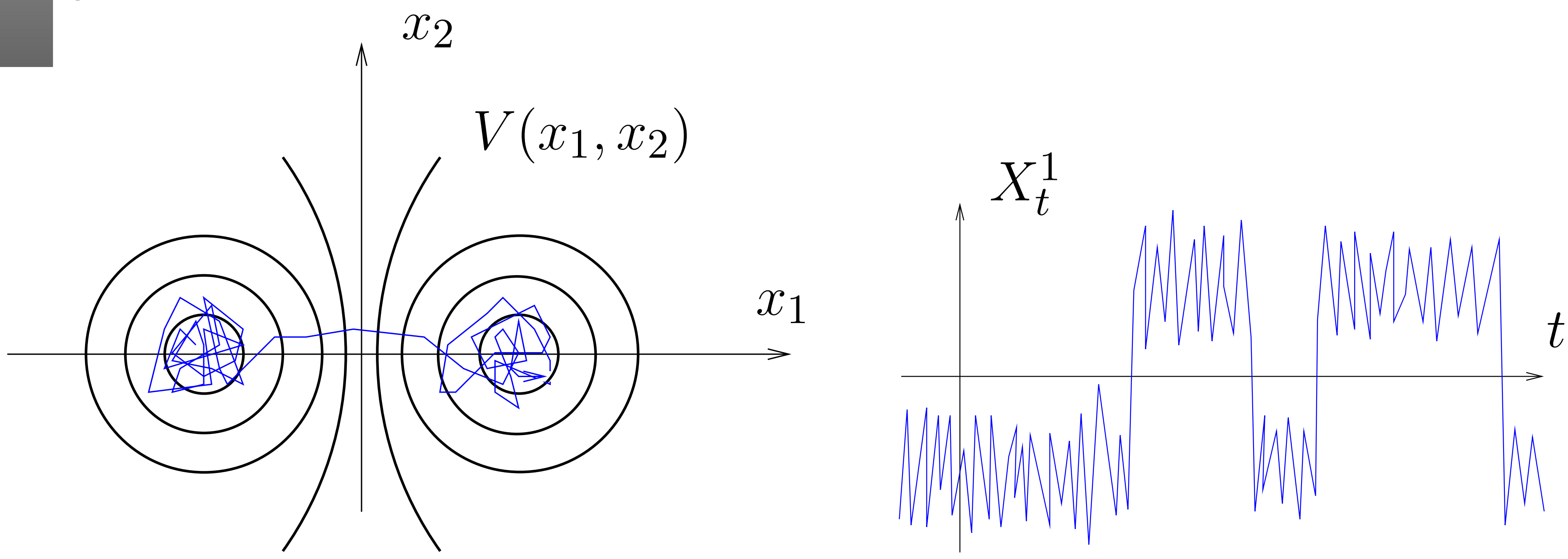
$$\begin{cases} d\boldsymbol{X}_t = M^{-1}\boldsymbol{P}_t\,dt, \\ d\boldsymbol{P}_t = -\nabla V(\boldsymbol{X}_t)\,dt - \gamma M^{-1}\boldsymbol{P}_t\,dt + \sqrt{2\gamma\beta^{-1}}d\boldsymbol{W}_t, \end{cases}$$

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

1 Free energy and metastability

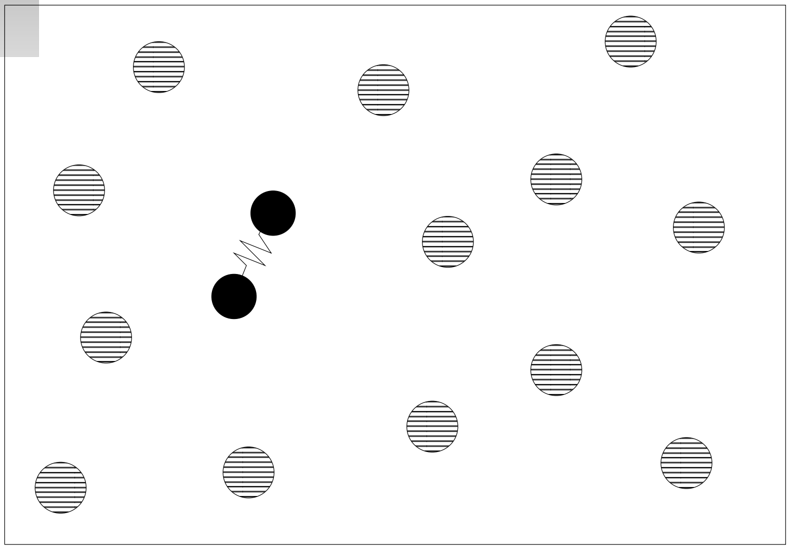
Problem: In practice, $\boldsymbol{X}_t$ is a metastable process, so that the convergence of the ergodic limit is very slow.

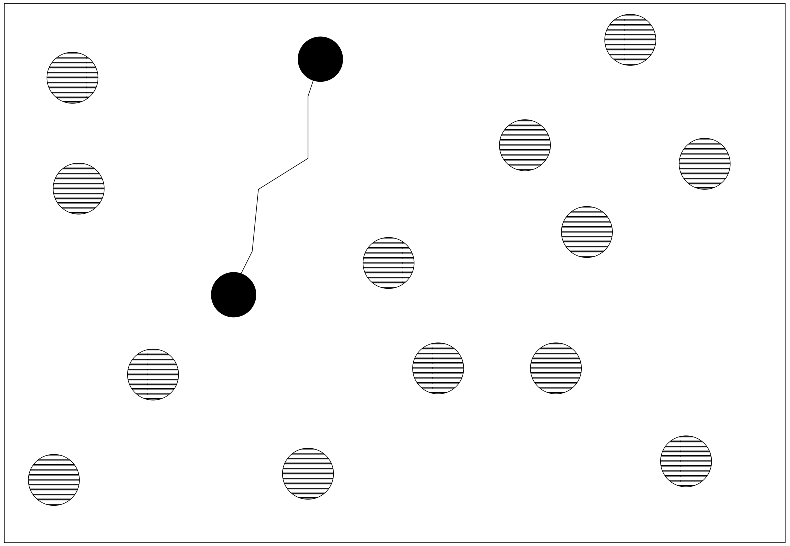
A bi-dimensional example: X_t^1 *is a slow variable of the system.*



1 Free energy and metastability

A more realistic example (Dellago, Geissler): Influence of the solvation on a dimer conformation.



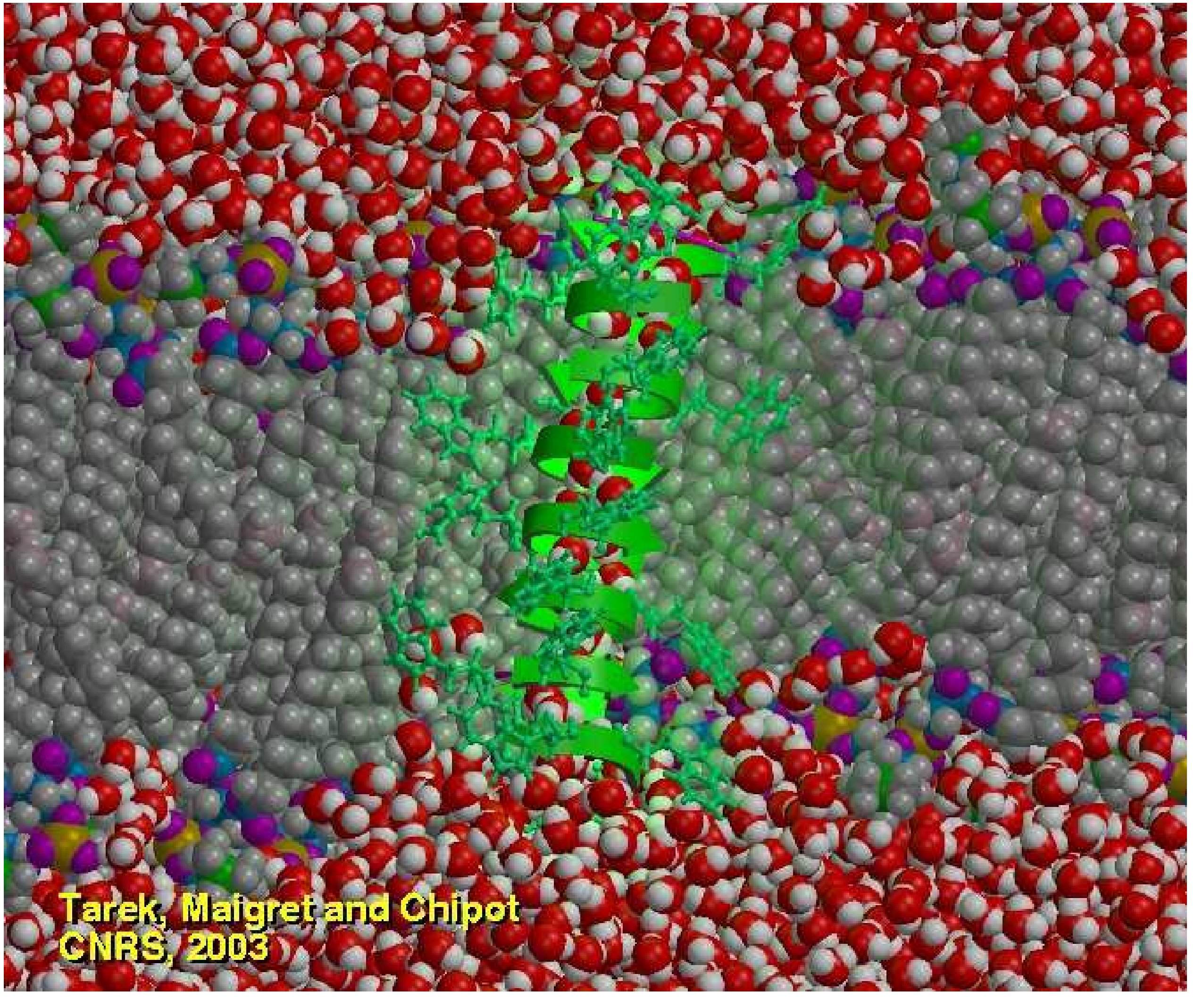


Left: compact state $(\xi = d_0)$. Right: stretched state $(\xi = d_1)$.

A slow variable is $\xi(\boldsymbol{X}_t)$ where $\xi(\boldsymbol{x}) = |\boldsymbol{x}_1 - \boldsymbol{x}_2|$ 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

Metastability: How to quantify this bad behaviour ?

1. Escape time from a potential well.
2. Asymptotice variance of the estimator.
3. "Decorrelation time".
4. Rate of convergence of the law of $\boldsymbol{X}_t$ to μ.

In the following we use the fourth criterium.

1 Free energy and metastability

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

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

which can be rewritten as $\partial_t \psi = \beta^{-1} \mathrm{div}\left(\psi_\infty \nabla \left(\frac{\psi}{\psi_\infty}\right)\right)$.

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):

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

1 Free energy and metastability

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

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 $E(t) \leq C\exp(-2\beta^{-1}Rt)$ and thus ψ converges to ψ_∞ exponentially fast with rate $\beta^{-1}R$.

Metastability $\iff$ small R

1 Free energy and metastability

Metastability: How to attack this problem ?
We suppose in the following that the slow variable is of dimension 1 and known: $\xi(\boldsymbol{X}_t)$, where $\xi : \mathbb{R}^n \to \mathbb{T}$.

Functionals to be averaged are typically functions of this slow variable.
Let us introduce the free energy A which is such that the image of the measure μ by ξ is $Z^{-1} \exp(-\beta A(z))\, dz$.
From the co-area formula, one gets:

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

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.

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 $d\mu_{\Sigma_z} = \dfrac{\psi\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}{\int_{\Sigma_z}\psi\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma_z}}$.

Indeed, for any bounded functions f and g,

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

1 Free energy and metastability

Remarks:

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

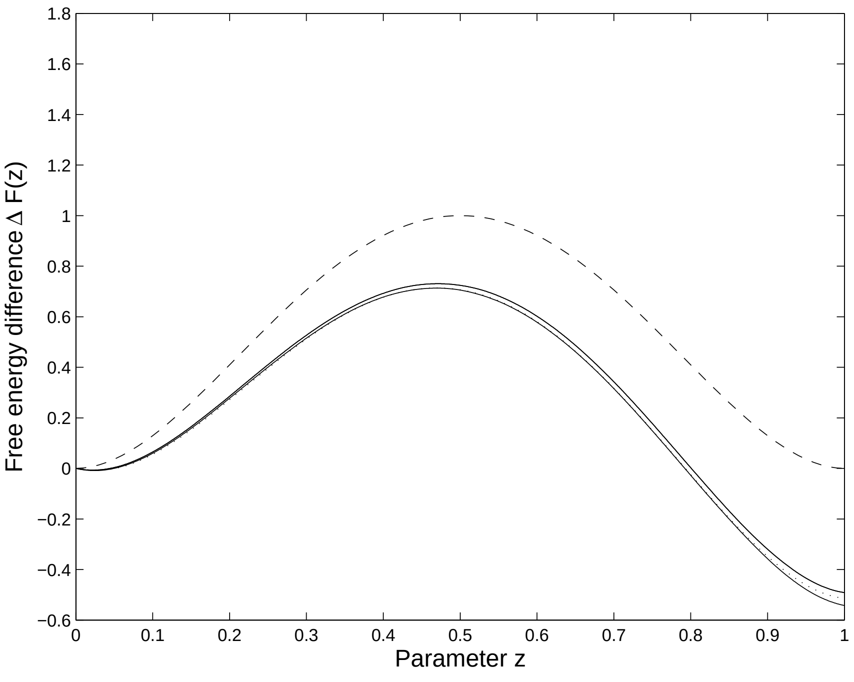
- A is the free energy associated with the reaction coordinate or collective variable ξ (angle, length, ...). A is defined up to an additive constant, so that it is enough to compute free energy differences, or the derivative of A (the mean force).

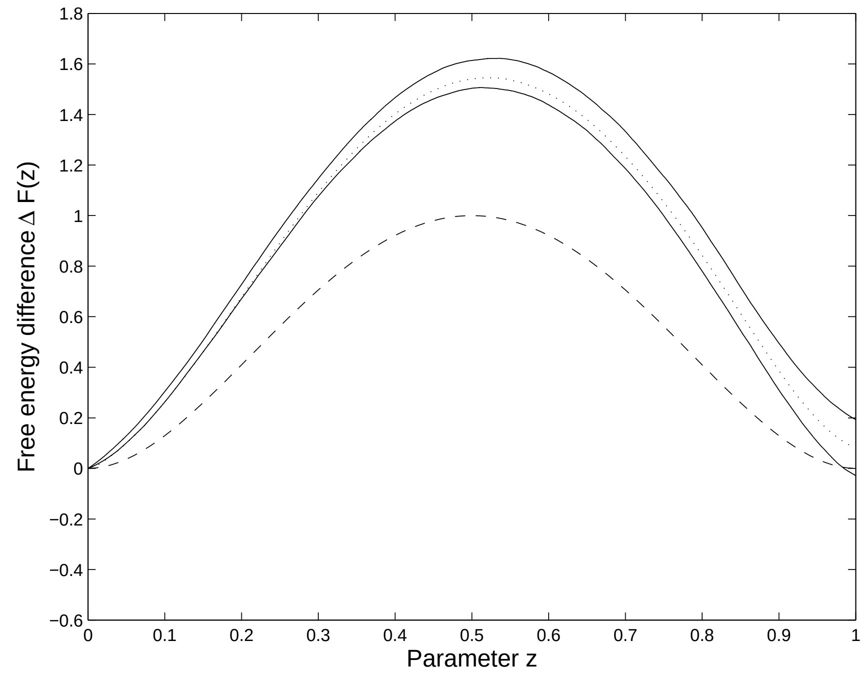
- $A(z) = -\beta^{-1} \ln Z_{\Sigma_z}$ and Z_{Σ_z} is the partition function associated with the conditional probability measures:

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

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

(Profiles computed using TI)



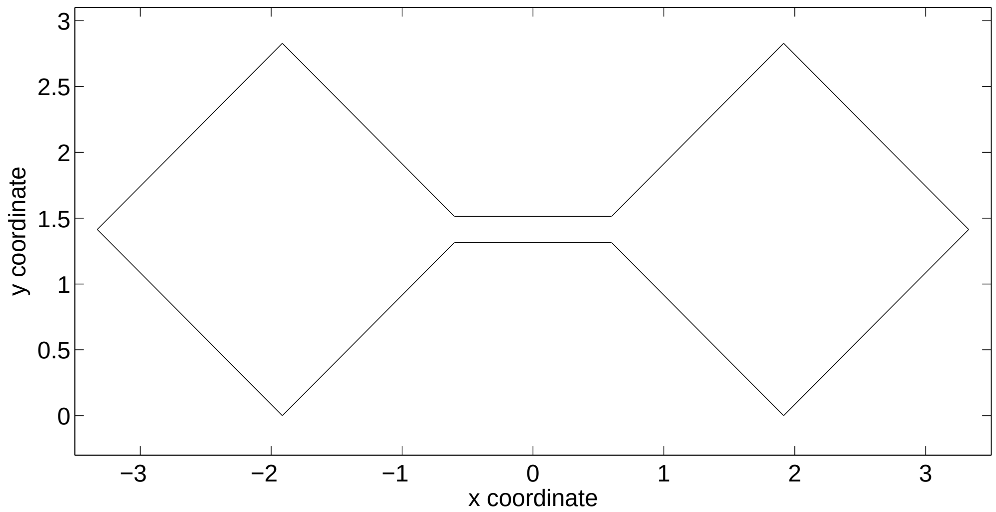


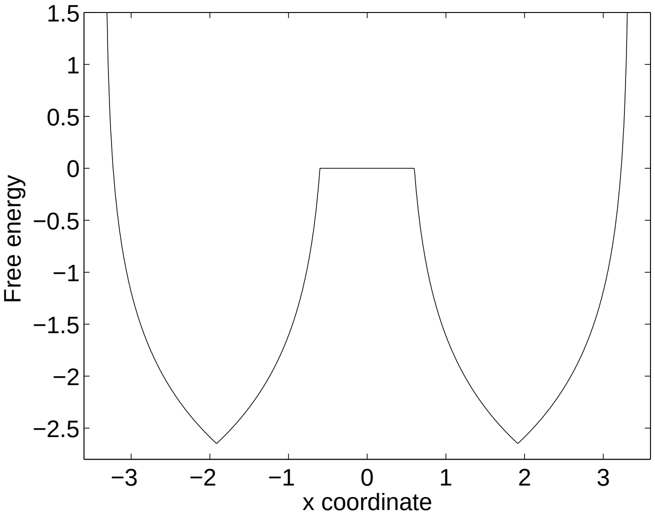
The density of the solvent molecules is lower on the left than on the right. At high (resp. low) density, the compact state is more (resp. less) likely. The “free energy barrier” is higher at high density than at low density. Question: interpretation of the free energy barriers in terms of dynamics ?

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.

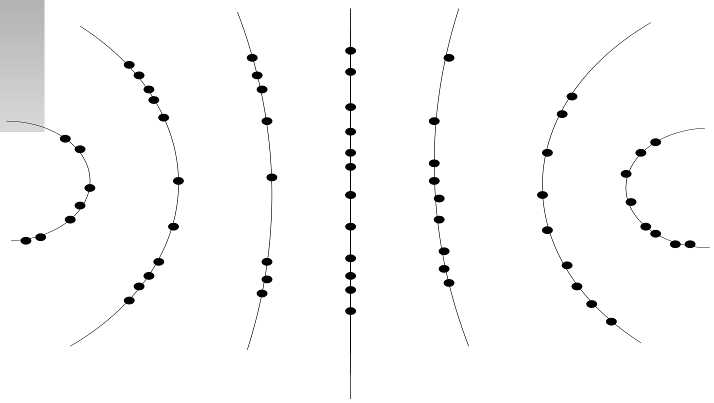
Right: Associated free energy profile when the x coordinate is the reaction coordinate ($\beta = 1$).

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

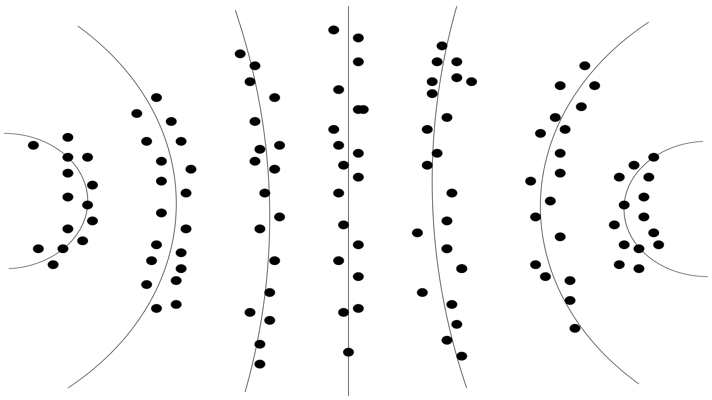
- Thermodynamic integration *(Kirkwood)* (homogeneous Markov process),
- Perturbation methods *(Zwanzig)* and histogram methods,
- Out of equilibrium dynamics *(Jarzynski)* (non-homogeneous Markov process),
- Adaptive methods *(ABF, metadynamics)* (non-homogeneous and non-linear Markov process).

Numerically, this amounts to: (i) sampling efficiently a multi-modal measure in high dimension, (ii) computing the marginal law of such a measure along a given low-dimensional function.

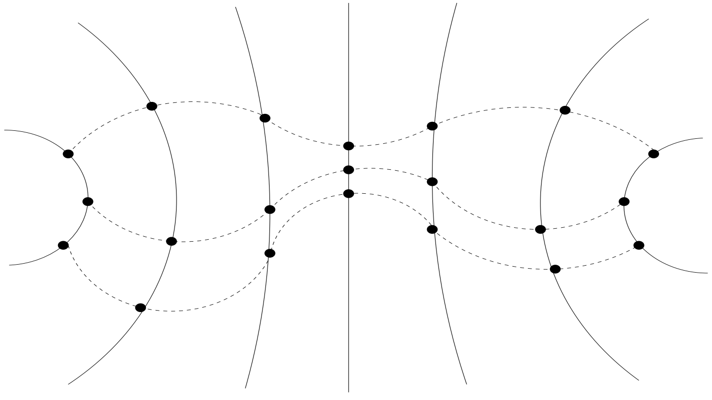
1 Free energy and metastability



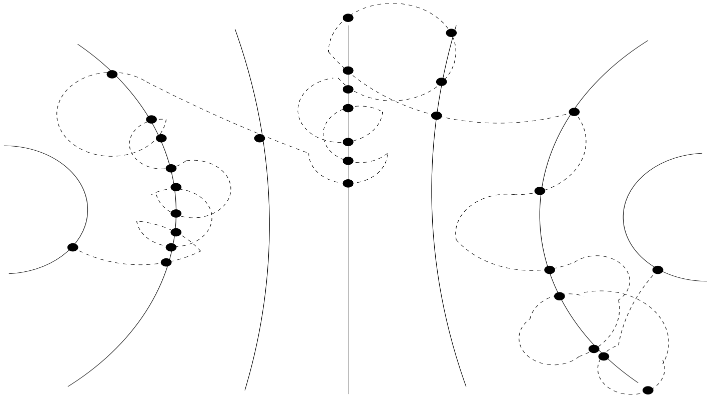
(a) Thermodynamic integration.



(b) Histogram method.



(c) Out of equilibrium dynamics.



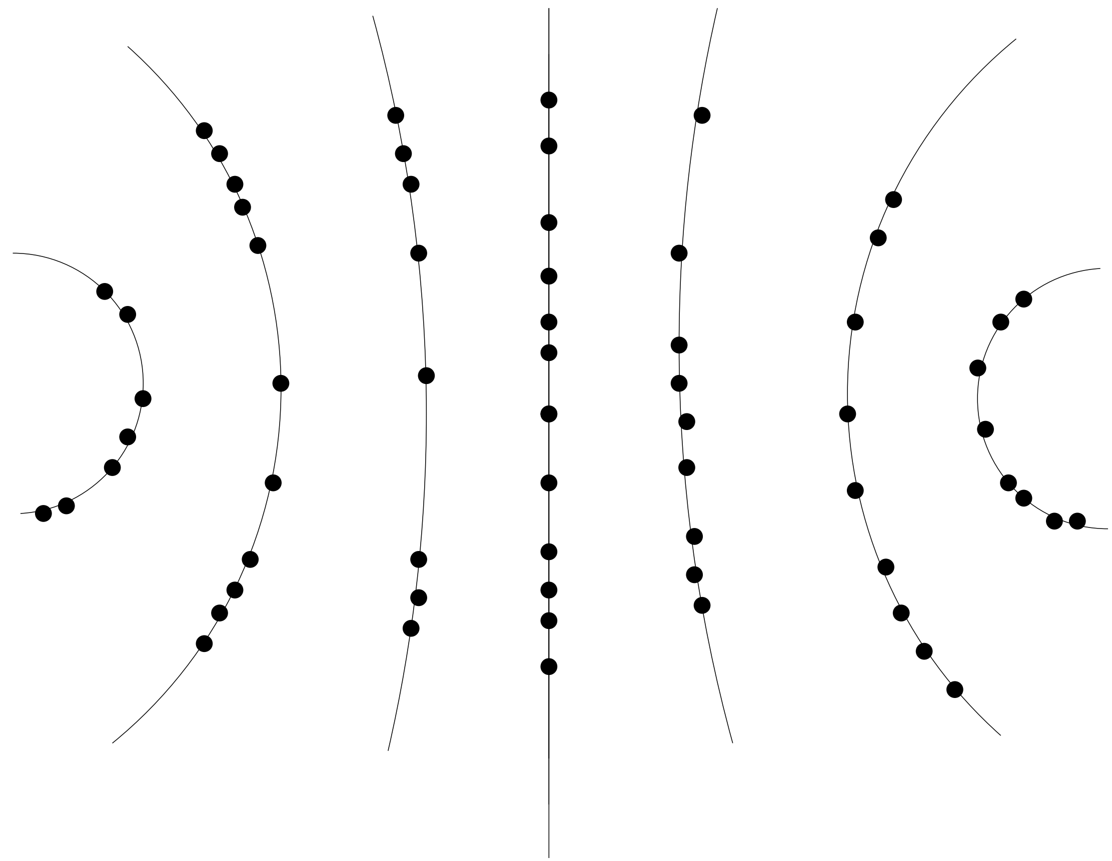
(d) Adaptive dynamics.

2 Constrained dynamics

- Thermodynamic integration *(Kirkwood)*
- Perturbation methods *(Zwanzig)* and histogram methods,
- Out of equilibrium dynamics *(Jarzynski)*,
- Adaptive methods *(ABF, metadynamics)*.

2.1 Thermodynamic integration

Thermodynamic integration



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:

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

Thus, $A'(z) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(\boldsymbol{X}_t)\, dt$.

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

Notice that there is actually no need to compute f in practice since the mean force may be obtained by averaging the Lagrange multipliers.

Indeed, 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

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

Of course, this comes at a price: essentially, we are using the fact that

$$\lim_{M\to\infty}\lim_{\Delta t\to 0}\frac{1}{M\Delta t}\sum_{m=1}^{M}\left[\xi\left(q+\sqrt{\Delta t}\,G^m\right) - 2\xi(q) + \xi\left(q-\sqrt{\Delta t}\,G^m\right)\right] = \Delta\xi(q),$$

and this estimator has a non zero variance.

2.1 Thermodynamic integration

More explicitly, the rigidly constrained dynamics writes:

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

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).

2.1 Thermodynamic integration

[G. Ciccotti, TL, E. Vanden-Einjden, 2008] Assume $z = 0$. μ_{Σ_0} is the unique invariant measure with support in Σ_0 for (RCD).

In particular, if $\boldsymbol{X}_t$ is 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 : $\forall \boldsymbol{\phi} \in \mathcal{C}^1(\mathbb{R}^{3N}, \mathbb{R}^{3N})$,

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

where $\mathrm{div}_{\Sigma_0}(\boldsymbol{\phi}) = \mathrm{tr}(P \nabla \boldsymbol{\phi})$.

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:

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

2.1 Thermodynamic integration

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

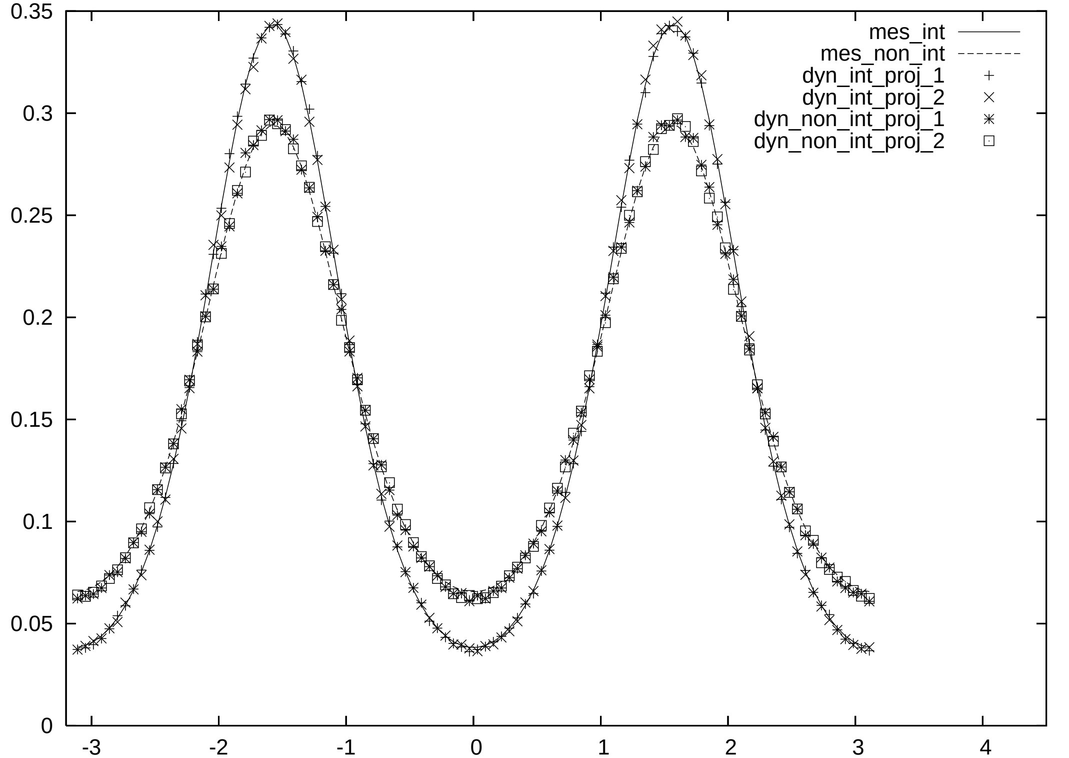
$$\begin{cases} \boldsymbol{X}_{n+1} = \boldsymbol{X}_n - \nabla \tilde{V}(\boldsymbol{X}_n)\Delta t + \sqrt{2\beta^{-1}}\Delta \boldsymbol{W}_n + \lambda \nabla \xi(\boldsymbol{X}_{n+1}), \\ \text{with } \lambda \in \mathbb{R} \text{ such that } \xi(\boldsymbol{X}_{n+1}) = 0, \end{cases}$$

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

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.

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.

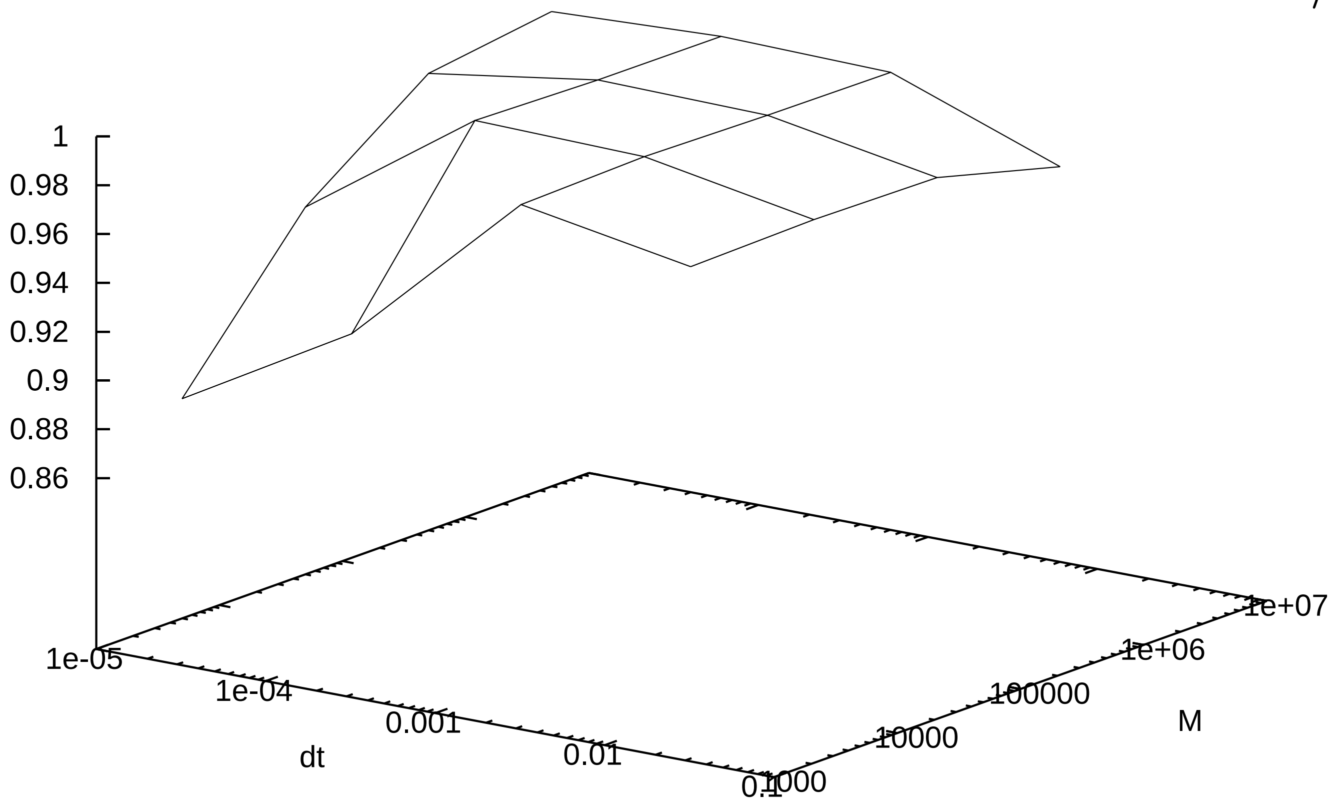
2.1 Thermodynamic integration

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

[E. Faou, TL, 2009] Using classical techniques (Talay-Tubaro like proof), one can check that the ergodic measure $\mu^{\Delta t}_{\Sigma_0}$ sampled by the Markov chain $(\boldsymbol{X}_n)$ is an approximation of order one of μ_{Σ_0}: for all smooth functions $g : \Sigma_0 \to \mathbb{R}$,

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

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 $z = 0$)

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

where the constraint is penalized. One can show that $\lim_{\eta\to 0}\boldsymbol{X}_t^\eta = \boldsymbol{X}_t$ (in $L^\infty_{t\in[0,T]}(L^2_\omega)$-norm) where $\boldsymbol{X}_t$ 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 $\int \phi(\boldsymbol{x}) d\mu(\boldsymbol{x})$:

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

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

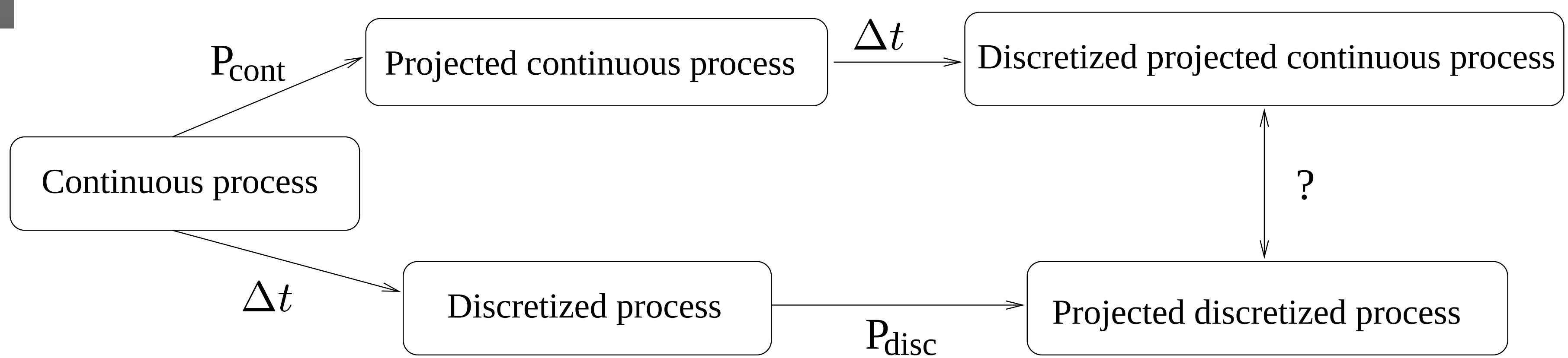
$$= Z^{-1} \int_z \frac{\int_{\Sigma_z} \phi e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\int_{\Sigma_z} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} \int_{\Sigma_z} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z} \, dz,$$

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

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

2.1 Thermodynamic integration

- [C. Le Bris, TL, E. Vanden-Einjden] For a general SDE (with a non isotropic diffusion), the following diagram does not commute:



2.1 Thermodynamic integration

How to correct the error due to the time discretization? A natural answer is to use a Metropolis-Hastings algorithm by adding an acceptance rejection step, with the acceptance probability:

$$r(x,y) = \frac{T(y,dx)\mu_{\Sigma_0}(dy)}{T(x,dy)\mu_{\Sigma_0}(dx)}.$$

Problem: the transition kernel $T(x,dy)$ has no simple analytical expression.

A solution: go to phase space, and use a Metropolis adjusted algorithm based on a proposal kernel which is symplectic (RATTLE scheme).

2.1 Thermodynamic integration

1. From an initial position $q^n \in \Sigma_0$, sample a random vector in tangent space:

$$\begin{cases} p^n = \beta^{-1/2}\, G^n + \nabla\xi(q^n)\widetilde{\lambda}^n, \\ \nabla\xi^T(q^n)\, p^n = 0. \end{cases}$$

2. Perform one integration step of the RATTLE scheme:

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\delta t}{2}\,\nabla V(q^n) + \nabla\xi(q^n)\,\lambda^{n+1/2}, \\ q^{n+1} = q^n + \delta t\, p^{n+1/2}, \\ \xi(q^{n+1}) = 0, \\ p^{n+1} = p^{n+1/2} - \dfrac{\delta t}{2}\,\nabla V(q^{n+1}) + \nabla\xi(q^{n+1})\,\lambda^{n+1}, \\ \nabla\xi^T(q^{n+1})\, p^{n+1} = 0. \end{cases}$$

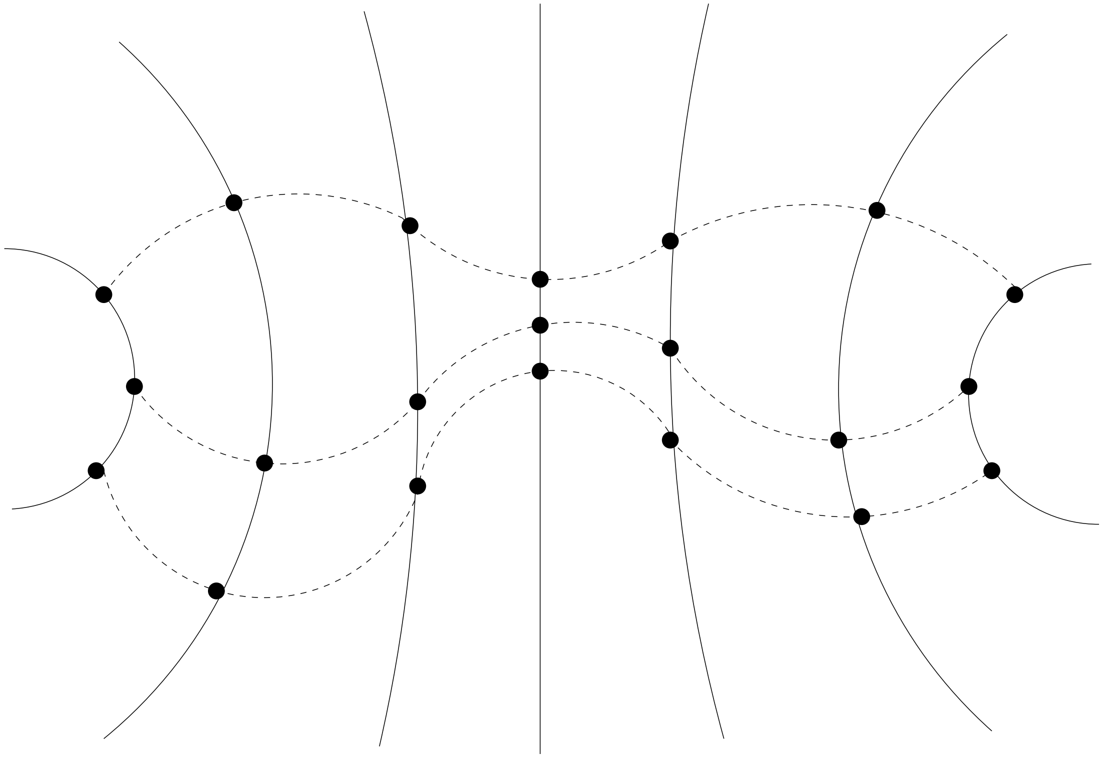
3. Accept the proposal with probability

$$\min\Big(\exp\big(-\beta\left(H(q^{n+1},p^{n+1}) - H(q^n,p^n)\right)\big), 1\Big),$$

where $H(q,p) = \frac{1}{2}\,|p|^2 + V(q)$. Otherwise, reject and set $q^{n+1} = q^n$.

2.2 Non-equilibrium dynamics

Non-equilibrium dynamics



2.2 Non-equilibrium dynamics

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

$$\begin{cases} d\boldsymbol{X}_t = & -P(\boldsymbol{X}_t)\nabla\tilde{V}(\boldsymbol{X}_t)\,dt + \sqrt{2\beta^{-1}}P(\boldsymbol{X}_t)\circ d\boldsymbol{W}_t \\ & +\nabla\xi(\boldsymbol{X}_t)d\Lambda_t^{\mathrm{ext}}, \\ d\Lambda_t^{\mathrm{ext}} = & \dfrac{z'(t)}{|\nabla\xi(\boldsymbol{X}_t)|^2}\,dt, \end{cases}$$

where $z : [0,T] \to [0,1]$ is a fixed deterministic evolution of the reaction coordinate ξ, such that $z(0) = 0$ and $z(T) = 1$.

2.2 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 $d\Lambda_t^{\mathrm{ext}} = \frac{z'(t)}{|\nabla\xi(\boldsymbol{X}_t)|^2}\,dt$.

2.2 Non-equilibrium dynamics

How to get equilibrium quantities (like the free energy) through non-equilibrium simulations ?

The idea is to associate to each trajectory $\boldsymbol{X}_t$ a weight

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

and to compute free energy differences by a Feynman-Kac formula (Jarzynski identity):

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

2.2 Non-equilibrium dynamics

[TL, M. Rousset, G. Stoltz, 2007] The proof consists in introducing the semi-group associated with the dynamics

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

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. Then

$$\int u(t,.) \exp(-\beta \tilde{V}) d\sigma_{\Sigma_{z(t)}} = \int u(0,.) \exp(-\beta \tilde{V}) d\sigma_{\Sigma_{z(0)}}$$

is equivalent to

$$\int \varphi \exp(-\beta \tilde{V}) d\sigma_{\Sigma_{z(t)}} = \exp(-\beta A(z(0))) \mathbb{E}\left(\int \varphi(\boldsymbol{X}_t) \exp\left(-\beta \int_0^t f(\boldsymbol{X}_r) z'(r)\, dr\right)\right).$$

This can be generalized to dynamics in phase-space (Langevin dynamics).

2.2 Non-equilibrium dynamics

A more general relation is the so-called Crooks identity:

$$\exp\Big(-\beta\big(A(z(T))-A(z(0))\big)\Big)\,\mathbb{E}\left(\varphi(\{\boldsymbol{X}^b_{T-s}\}_{0\le s\le T})\right)$$
$$=\mathbb{E}\left(\varphi(\{\boldsymbol{X}^f_s\}_{0\le s\le T})\exp(-\beta\mathcal{W}^f(T))\right)$$

where $\boldsymbol{X}^f_0\sim\mu_{\Sigma_{z(0)}}$, $\boldsymbol{X}^b_0\sim\mu_{\Sigma_{z(T)}}$, $\mathcal{W}^f(T)=\int_0^T f(\boldsymbol{X}^f_s)z'(s)\,ds$

$$\left\{\begin{array}{l} d\boldsymbol{X}^f_t=-\nabla\tilde{V}(\boldsymbol{X}^f_t)\,dt+\sqrt{2\beta^{-1}}d\boldsymbol{W}^f_t+\nabla\xi(\boldsymbol{X}^f_t)d\Lambda^f_t,\\ \xi(\boldsymbol{X}^f_t)=z(t),\end{array}\right.$$

$$\left\{\begin{array}{l} d\boldsymbol{X}^b_t=-\nabla\tilde{V}(\boldsymbol{X}^b_t)\,dt+\sqrt{2\beta^{-1}}d\boldsymbol{W}^b_t+\nabla\xi(\boldsymbol{X}^b_t)d\Lambda^b_t,\\ \xi(\boldsymbol{X}^b_t)=z(T-t).\end{array}\right.$$

The discretization of the constrained process is (as before):

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

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

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:

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

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 $-\beta^{-1} \ln\left(\frac{1}{M}\sum_{m=1}^M \exp\left(-\beta \mathcal{W}_{T/\Delta t}^m\right)\right)$.

2.2 Non-equilibrium dynamics

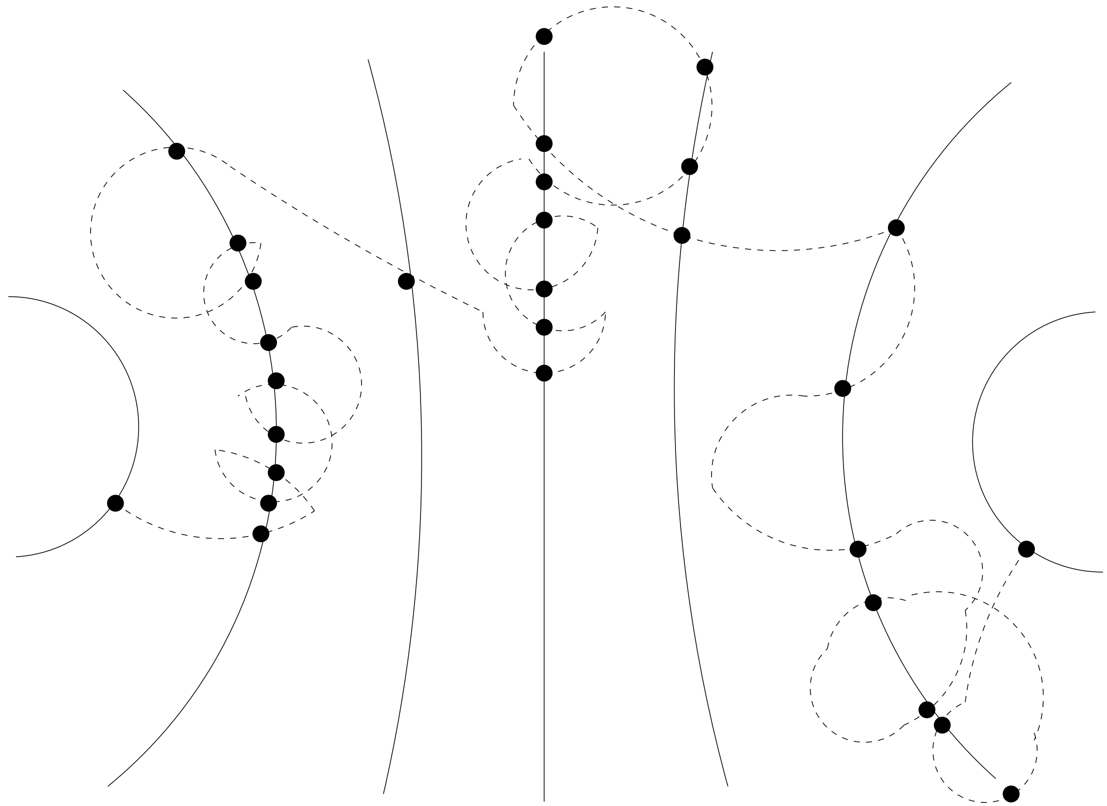
In practice, the efficiency of the Jarzynski identity as a numerical method to evaluate free energy differences is not clearly demonstrated. If the transition is too fast, the variance of the estimator is very large. If the transition is slow, we are back to thermodynamic integration...

Ideas: (i) combine forward and backward trajectories or (ii) add selection mechanisms [M. Rousset, G. Stoltz, 2006].

3 Adaptive methods

- Thermodynamic integration *(Kirkwood)*,
- Perturbation methods *(Zwanzig)* and histogram methods,
- Out of equilibrium dynamics *(Jarzynski)*,
- Adaptive methods *(ABF, metadynamics)*.

Adaptive methods



3.1 Adaptive methods: algorithms

The bottom line of adaptive methods is the following: for “good” ξ the potential $V - A \circ \xi$ is less metastable than V. But A is unknown !

Principle: use a time dependent potential of the form

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

where A_t is an approximation at time t of A, given the configurations visited so far.

Hopes:

- build a dynamics which goes quickly to equilibrium,
- compute free energy profiles.

Wang-Landau, ABF, metadynamics: *Darve, Pohorille, Hénin, Chipot, Laio, Parrinello, Wang, Landau,...*

3.1 Adaptive methods: algorithms

How to update A_t ? Two methods depending on wether A' (Adaptive Biasing Force) or A_t (Adaptive Biasing Potential) is approximated.

For the Adaptive Biasing Force method, the idea is to use the formula

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

$$= \int f \, d\mu_{\Sigma_z} = \mathbb{E}_\mu(f(\boldsymbol{X}) | \xi(\boldsymbol{X}) = z).$$

The mean force $A'(z)$ is the mean of f with respect to $\mu_{\Sigma_z} = Z_{\Sigma_z}^{-1} e^{-\beta V} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}$.

3.1 Adaptive methods: algorithms

Important remark: whatever A_t, the mean force associated with the Gibbs distribution

$$\psi^{\mathrm{eq}} \propto \exp(-\beta \mathcal{V}_t)(\boldsymbol{x})\, d\boldsymbol{x} = \exp(-\beta(V - A_t \circ \xi))(\boldsymbol{x})\, d\boldsymbol{x}$$

is the original mean force A':

$$\frac{\displaystyle\int f \psi^{\mathrm{eq}} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}}{\displaystyle\int \psi^{\mathrm{eq}} |\nabla \xi|^{-1} d\sigma_{\Sigma_z}} = A'(z).$$

Thus, use as an approximation of $A'(z)$:

$$A_t'(z) = \mathbb{E}(f(\boldsymbol{X}_t) | \xi(\boldsymbol{X}_t) = z).$$

3.1 Adaptive methods: algorithms

A typical ABF dynamics is thus:

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

The associated (nonlinear) Fokker-Planck equation writes:

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

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

3.1 Adaptive methods: algorithms

Two variants:

- A may be approximated instead of A', using the formula

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

This leads to Adaptive Biasing Potential (ABP) methods. A typical example is:

$$\begin{cases} d\boldsymbol{X}_t = -\nabla(V - A_t \circ \xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t, \\ \dfrac{\partial A_t}{\partial t}(z) = -\dfrac{1}{\tau}\beta^{-1} \ln \Big(\mathbb{E}\left(\delta(\xi(\boldsymbol{X}_t) - z) \right) \Big). \end{cases}$$

3.1 Adaptive methods: algorithms

- To avoid geometry problem, an extended configurational space $(\boldsymbol{x}, z) \in \mathbb{R}^{n+1}$ is considered, together with the meta-potential:

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

  Choosing $(\boldsymbol{x}, z) \mapsto z$ as a reaction coordinate, the associated free energy A^k is close to A (in the limit $k \to \infty$, up to an additive constant).

3.1 Adaptive methods: algorithms

Adaptive algorithms used in molecular dynamics fall into one of these four possible combinations:

	A'_t	A_t
V	ABF	Wang-Landau
V^k	...	metadynamics

3.2 Adaptive methods: convergence

Let us now study the rate of convergence of the ABF methods:

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

Questions: Does A_t' converge to A' ? What did we gain compared to the original gradient dynamics ?

Using entropy techniques, the convergence of ψ (and thus A_t') to its equilibrium value can be shown to be exponential.

3.2 Adaptive methods: convergence

A fundamental remark. Let us consider the problem in a simple situation: $n = 2$, the configuration space is $\mathbb{T} \times \mathbb{R}$, and $\xi(x, y) = x$.

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

Let $\overline{\psi}(t, x) = \int \psi(t, x, y)\, dy$. Then

$$\begin{aligned} \partial_t \overline{\psi} &= \beta^{-1} \partial_{x,x} \overline{\psi} + \partial_x \int \partial_x V \psi\, dy - \partial_x (A_t' \overline{\psi}) \\ &= \beta^{-1} \partial_{x,x} \overline{\psi}. \end{aligned}$$

The metastability along the reaction coordinate direction has been eliminated.

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. $\longrightarrow$ The real limitation.

3.2 Adaptive methods: convergence

Main ingredients of the proof in the simple setting ($n = 2$ on $\mathbb{T} \times \mathbb{R}$, with $\xi(x, y) = x$).

Ingredient 1: $\overline{\psi}(t, x) = \int \psi(t, x, y)\, dy$ 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 $C \exp(-4\pi^2 \beta^{-1} t)$.

Ingredient 2: Decomposition of entropy: $E = E_M + E_m$. "Total entropy = macroscopic entropy + microscopic entropy."

Cf. works by F. Otto *et al.*

3.2 Adaptive methods: convergence

Equilibrium is $\psi_\infty = Z^{-1}\exp(-\beta(V - A\circ\xi))$.

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

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

The microscopic entropy is

$$E_m(t) = \int H\left(\psi(\cdot|\xi(x)=z)\middle|\psi_\infty(\cdot|\xi(x)=z)\right)\overline{\psi}(z)\,dz$$

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

We already know that E_M goes to zero: it remains to consider E_m.

3.2 Adaptive methods: convergence

Notice that

$$\partial_t \psi = \beta^{-1} \mathrm{div} \left(\psi_\infty \nabla \left(\frac{\psi}{\psi_\infty} \right) \right) + \partial_x ((A' - A'_t) \psi).$$

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 prob. 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):

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

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

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

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

We have proved that

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

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

3.2 Adaptive methods: convergence

In summary [TL, G. Stoltz, M. Rousset, Nonlinearity 2008] :

- Original gradient dynamics: $\exp(-\beta^{-1}Rt)$ where R is the ISL constant for μ ;
- ABF dynamics: $\exp(-\beta^{-1}\rho t)$ where ρ is the ISL constant for the conditioned probability measures $\mu(\cdot|\xi(x) = z)$.

If ξ is well chosen, $\rho \gg R$.

Remarks:

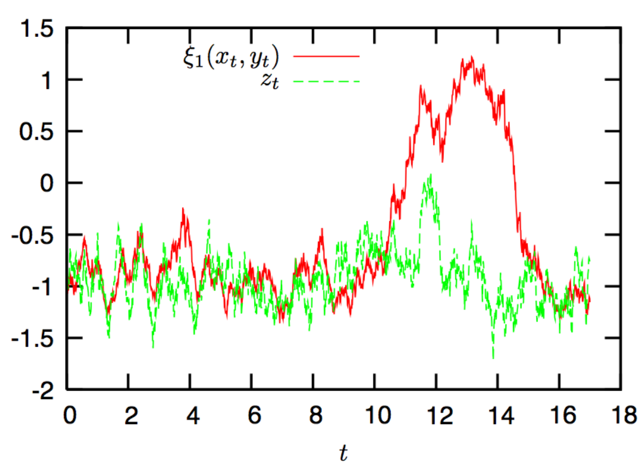
- if there are metastabilities in μ_{Σ_z}, only "local LSI" is needed (work in progress with K. Minoukadeh)
- the ABP case is not understood so far...

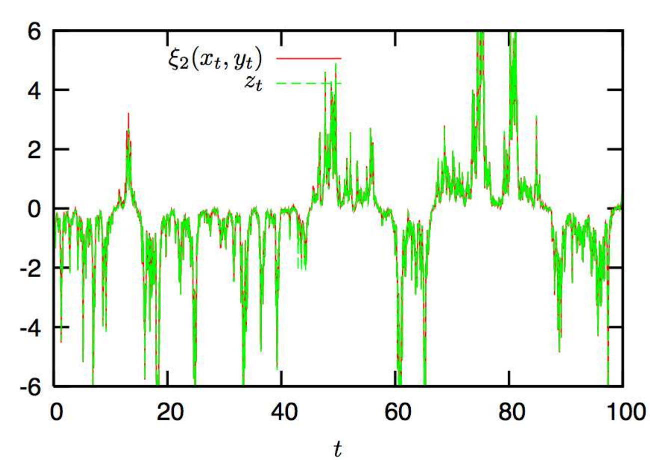
3.2 Adaptive methods: convergence

Other results based on this set of assumptions:

- [TL, JFA 2008] LSI for the cond. meas. $\mu(\cdot|\xi(x) = z)$
  + LSI for the marginal $\overline{\mu}(dz) = \xi * \mu(dz)$
  + bdd coupling ($\|\nabla_{\Sigma_z} f\|_{L^\infty} < \infty$) $\implies$ LSI for μ.
- [F. Legoll, TL, 2009] Effective dynamics for $\xi(\boldsymbol{X}_t)$. Uniform control in time:

$$H(\mathcal{L}(\xi(X_t))|\mathcal{L}(z_t)) \leq C \left(\frac{\|\nabla_{\Sigma_z} f\|_{L^\infty}}{\rho} \right)^2 H(\mathcal{L}(X_0)|\mu).$$





3.3 Multiple replicas implementations

Discretization of adaptive methods can be done using two (complementary) approaches:

- Use trajectorial averages along a single path:

$$\mathbb{E}(f(\boldsymbol{X}_t)|\xi(\boldsymbol{X}_t)=z) \simeq \frac{\int_0^t f(\boldsymbol{X}_s)\delta^\alpha(\xi(\boldsymbol{X}_s)-z)\,ds}{\int_0^t \delta^\alpha(\xi(\boldsymbol{X}_s)-z)\,ds}.$$

- Use empirical means over many replicas (intreacting particle system):

$$\mathbb{E}(f(\boldsymbol{X}_t)|\xi(\boldsymbol{X}_t)=z) \simeq \frac{\sum_{m=1}^N f(\boldsymbol{X}_t^{m,N})\delta^\alpha(\xi(\boldsymbol{X}_t^{m,N})-z)}{\sum_{m=1}^N \delta^\alpha(\xi(\boldsymbol{X}_t^{m,N})-z)}.$$

3.3 Multiple replicas implementations

Interest of a discretization using an interacting particle system:

- Very efficient parallelization.
- Better sampling of all reactive paths.
- A selection mechanism may be added to duplicate “innovative particles” and kill “redundant particles”.

$\longrightarrow$ We propose a selection mechanism which accelerates the convergence “at the macroscopic level” (increase r). [TL, G. Stoltz, M. Rousset, J Chem Phys 2007].

3.3 Multiple replicas implementations

Numerical analysis of the particle system [B. Jourdain, TL, R. Roux, 2009]

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:

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

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

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

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

3.3 Multiple replicas implementations

The selection mechanism

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

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

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

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

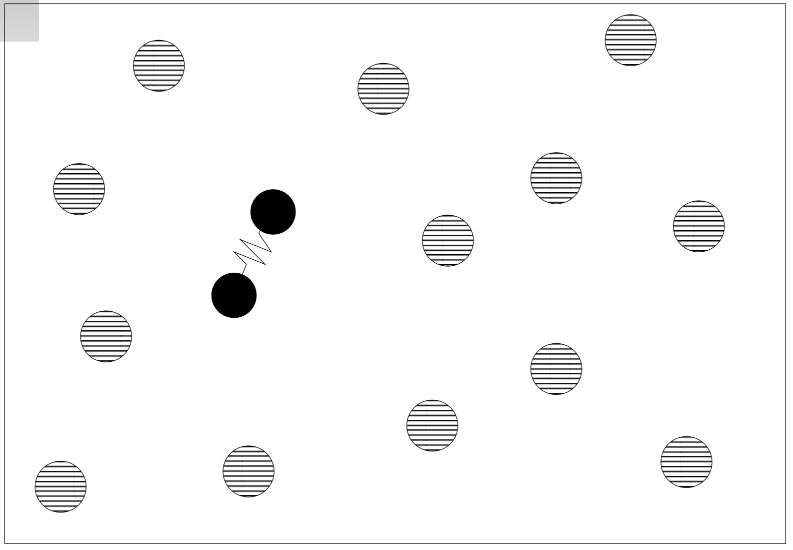
Weights of particles in locally under-explored regions are increased.

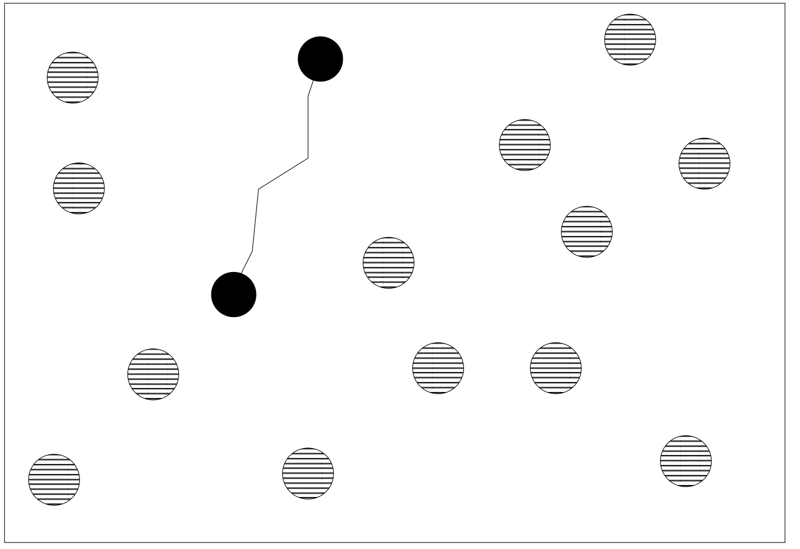
An adequate selection process can then be implemented, using these weights (like in genetic algorithm).

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

3.3 Multiple replicas implementations

Numerical illustration on the example of the solvation of a dimer.



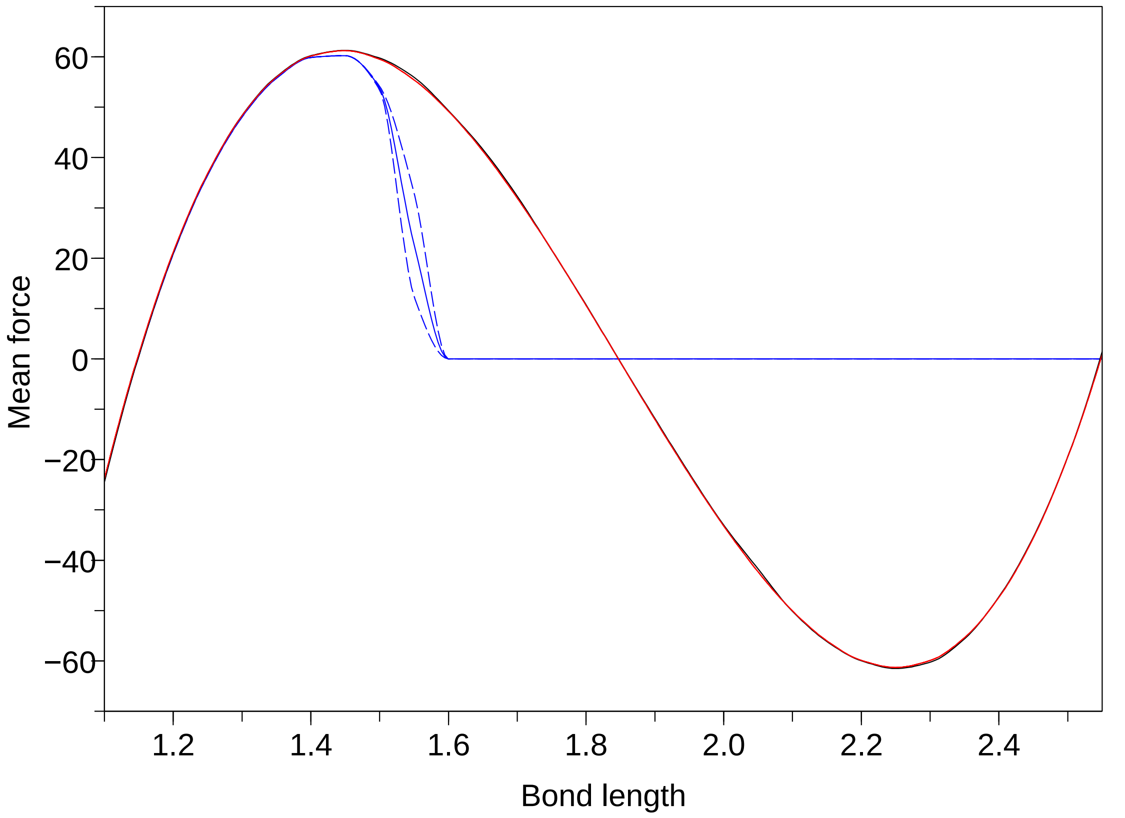


Left: compact state Right: stretched state.

Recall the reaction coordinate is $\xi(x) = |x_1 - x_2|$.

3.3 Multiple replicas implementations

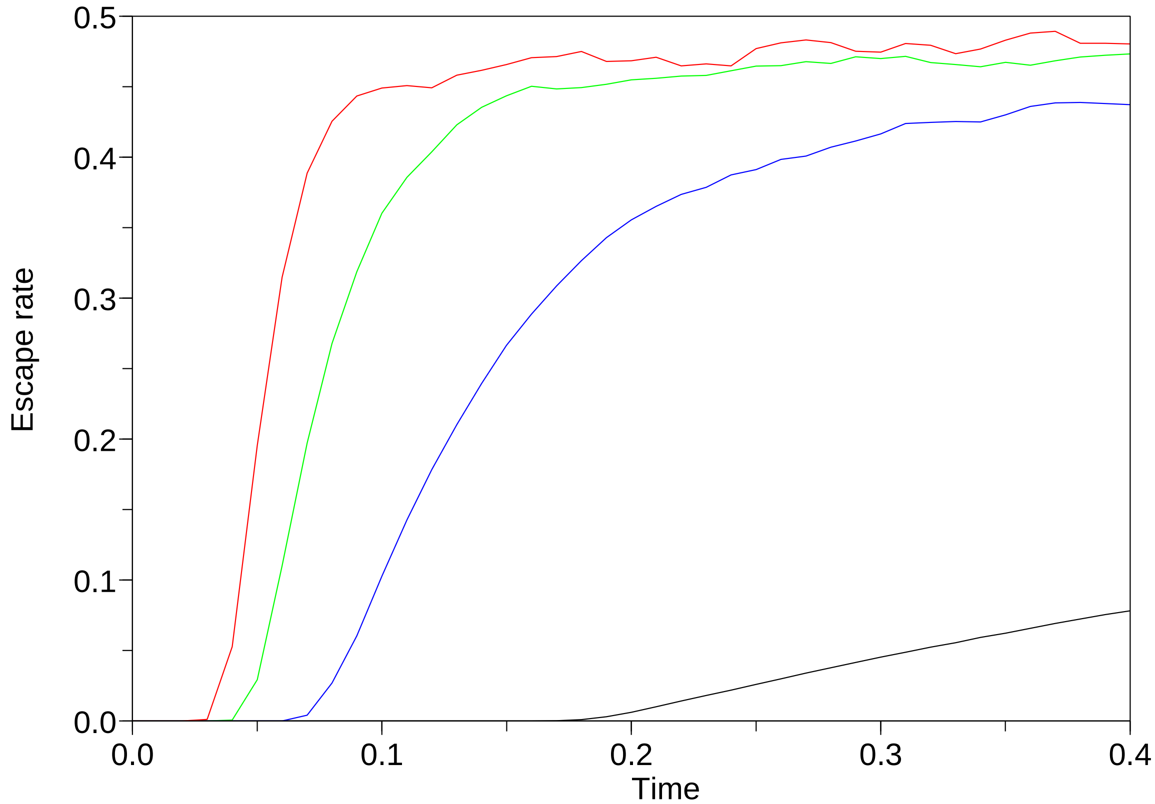
Free energy profile with parallel ABF obtained at t = 0.1, with 2000 replicas.



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.

3.4 Extensions and another application

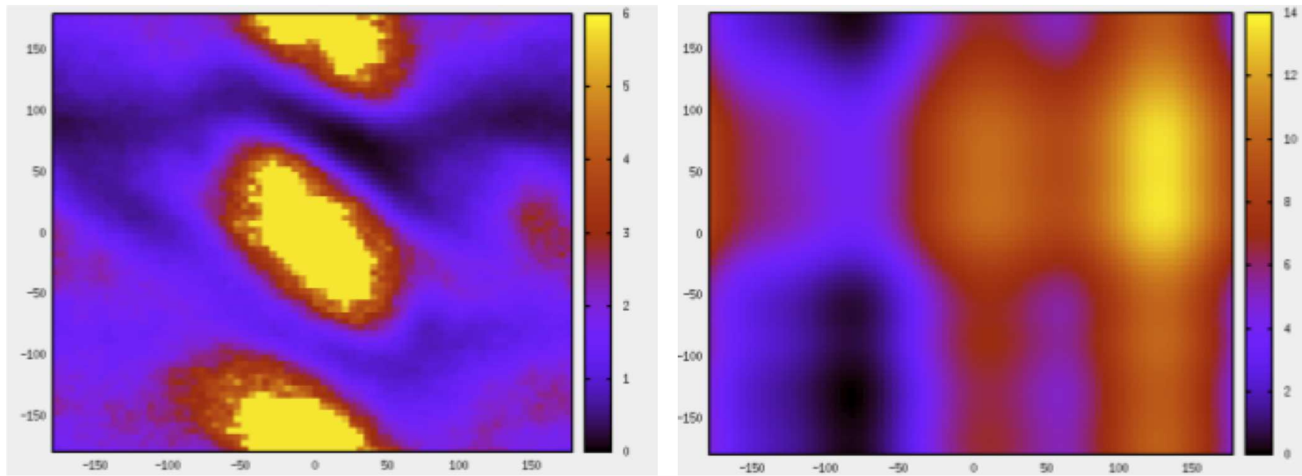
Adaptive methods can be seen as adaptive importance sampling methods rather than free energy calculation methods. $\longrightarrow$ compute a bias adaptively, and then unbias.
This gives many freedom in the way to use them. For example:

- Instead of computing the complicated local mean force $f = \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right)$, use simpler expressions, like $\frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2}$.
- Use ABF for high dimensional reaction coordinates by postulating a separated representations of the mean force:

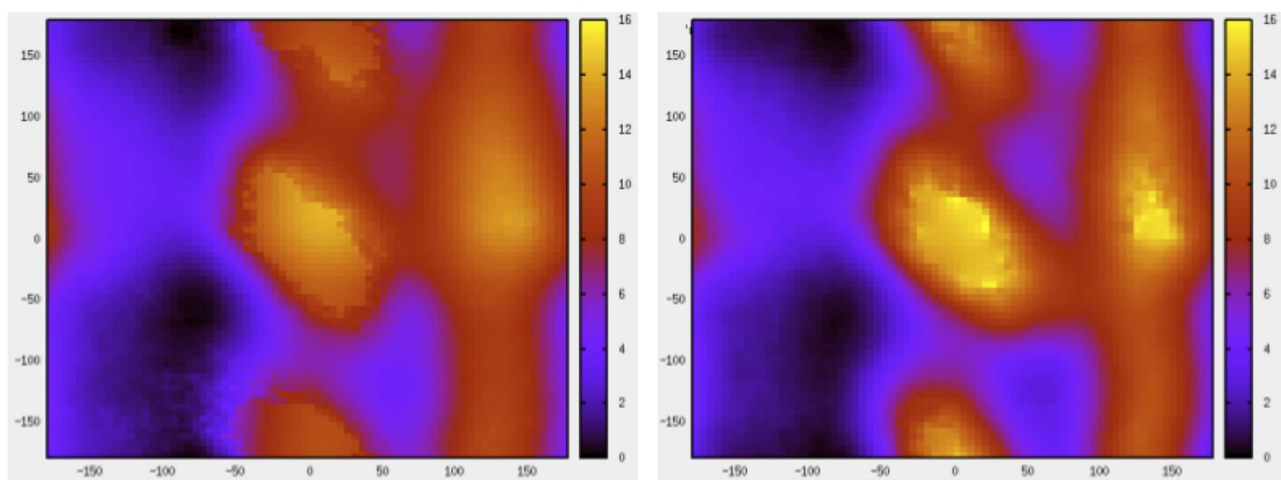
$$A(z_1, \ldots, z_N) = A_1(z_1) + A_{2,3}(z_2, z_3) + A_4(z_4) + \ldots$$

3.4 Extensions and another application

Preliminary results on the alanine dipeptide: $A_1(\phi) + A_2(\psi)$.



$-kT\ln(\mathbb{P}(\delta_{(\phi,\psi)(\boldsymbol{X}_t)-(\phi_0,\psi_0)}))$ Tensor product of the bias



Reconstructed PMF Reference PMF

Work in progress with C. Chipot and J. Hénin.

3.4 Extensions and another application

Application to Bayesian statistics: Sampling of posterior distributions using a MCMC ABF algorithm.

- The fishery problem: the size of $N_{\rm data} = 256$ fishes 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 ($q_N = 1 - \sum_{i=1}^{N-1} q_i$): $x = \big(q_1, \ldots, q_{N-1}, \mu_1, \ldots, \mu_N, v_1, \ldots, v_N\big) \in \mathcal{S}_{N-1} \times [\mu_{\min}, \mu_{\max}]^N \times [v_{\min}, +\infty) \subset \mathbb{R}^{3N-1}$, 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 Extensions and another application

- given the parameters, the likelihood of observing the data $\{y_i, 1 \leq i \leq N_{\text{data}}\}$ is
$$\Pi(y \,|\, x) = \prod_{d=1}^{N_{\text{data}}} f(y_d \,|\, x).$$
- the prior on the parameters is $\mu_i \sim \mathcal{N}(M, R^2/4)$, $v_i \sim \mathrm{Gamma}(a, \beta)$ with $\beta \sim \mathrm{Gamma}(g, h)$ and $(q_1, \ldots, q_N) \sim \mathrm{Dirichlet}_N(1, \ldots, 1)$ for fixed values (M, R, a, g, h).

So actually $x = (q_1, \ldots, q_{N-1}, \mu_1, \ldots, \mu_N, v_1, \ldots, v_N, \beta)$.

Objective: sample the posterior distribution (distribution of the parameters given the observations):
$$\Pi(x|y) = \frac{\Pi(y|x)\,\mathrm{Prior}(x)}{\int \Pi(y|x)\,\mathrm{Prior}(x)\,dx}.$$

3.4 Extensions and another application

The potential associated with the posterior (posterior is proportional to $\exp(-V)$) is

$$V = V_{\text{prior}} + V_{\text{likelihood}}$$

with $V_{\text{prior}} = \frac{2}{R^2}\sum_{i=1}^N (\mu_i - M)^2 - N\alpha \ln\beta + \beta\sum_{i=1}^N v_i - (a-1)\sum_{i=1}^N \ln v_i + h\beta - (g-1)\ln\beta$ and $V_{\text{likelihood}} = \sum_{d=1}^{N_{\text{data}}} \ln\left[\sum_{i=1}^N q_i\sqrt{v_i}\exp\left(-\frac{v_i}{2}(y_d-\mu_i)^2\right)\right].$

The posterior distribution is a metastable (multimodal) measure.

Idea: use ABF together with a Metropolis Hasting algorithm (with a fixed proposal). What is a good "reaction coordinate" $\xi(x)$?

3.4 Extensions and another application

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 $\bar{x}^{n+1}$ according to $T(x^n, \bar{x}^{n+1})$.
3. Acceptance ratio

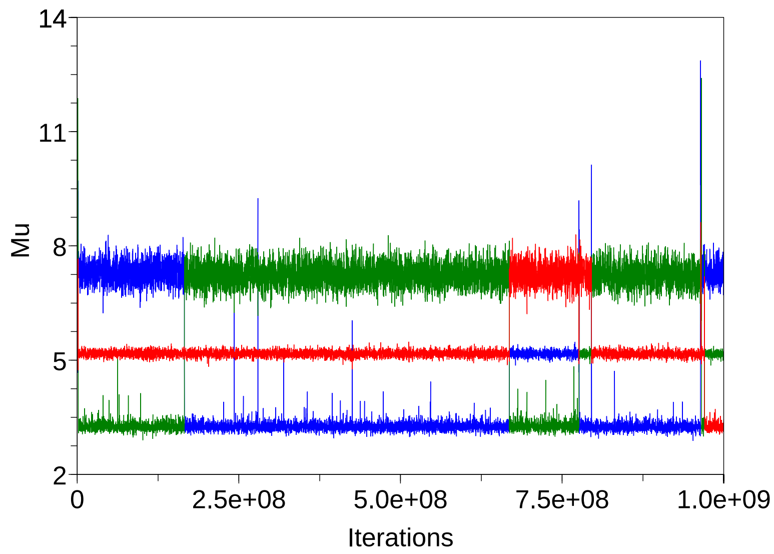
$$r^n = \min\left(\frac{\pi_{A^{n+1}}(\bar{x}^{n+1})\,T(\bar{x}^{n+1}, x^n)}{\pi_{A^{n+1}}(x^n)\,T(x^n, \bar{x}^{n+1})}, 1\right),$$

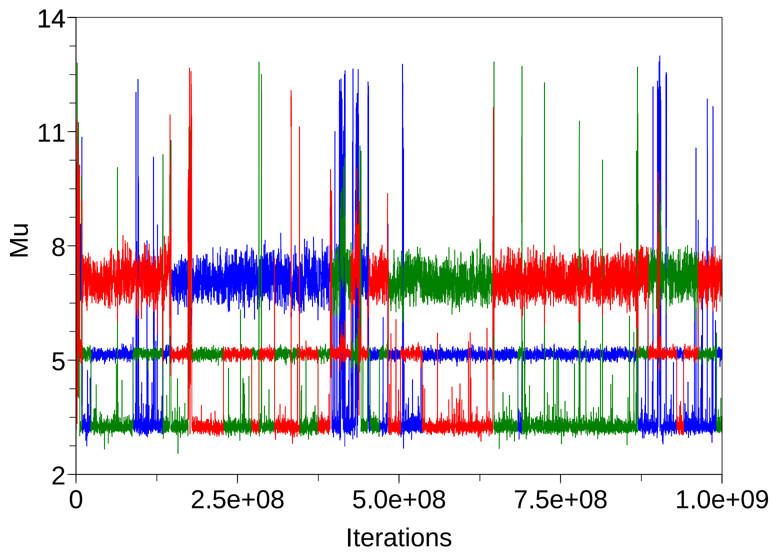
   where the biased probability is $\pi_{A^{n+1}}(x) \propto \pi(x)\exp(A^{n+1}(\xi(x)))$.
4. Draw a random variable U^n uniformly distributed in $[0, 1]$ ($U^n \sim \mathcal{U}[0, 1]$).
   (a) if $U^n \leq r^n$, accept the move and set $x^{n+1} = \bar{x}^{n+1}$;
   (b) if $U^n > r^n$, reject the move and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

3.4 Extensions and another application

[Work in progress with N. Chopin and G. Stoltz.]

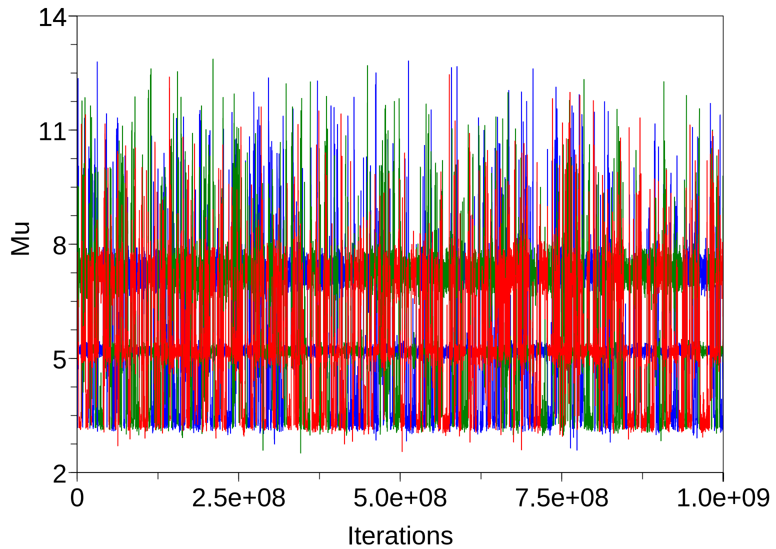
Some results for $N = 3$.

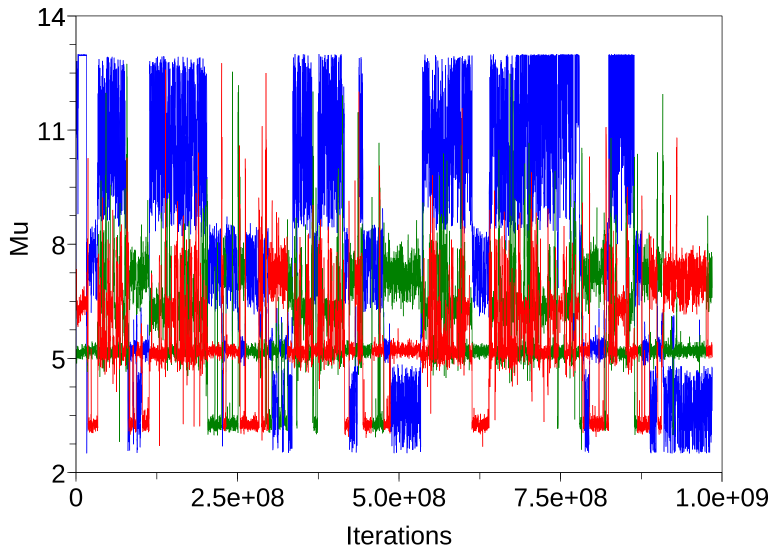




Left: evolution of the averages μ_i without bias.
Right: evolution of the averages μ_i with $\xi = q_1$.

3.4 Extensions and another application

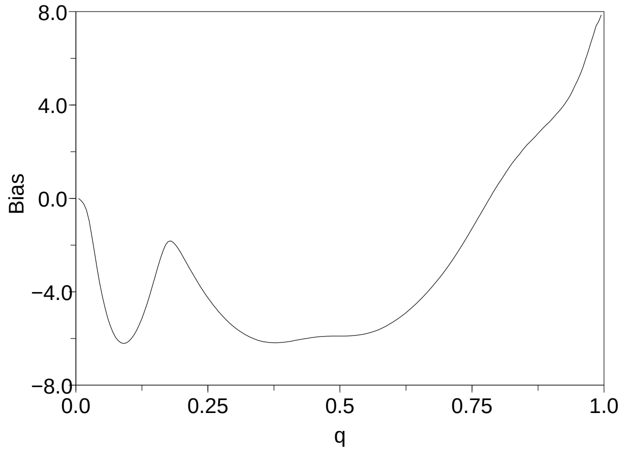


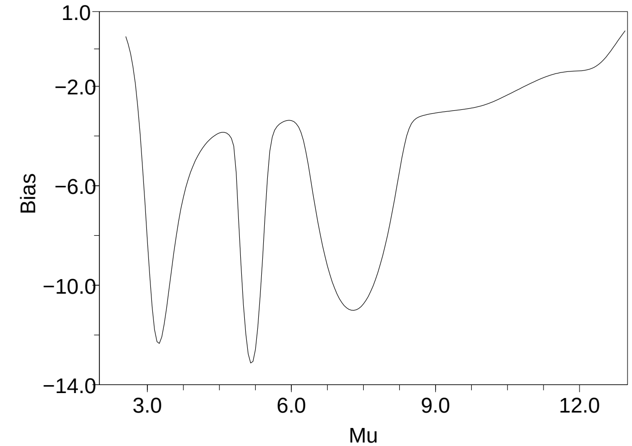


Left: evolution of the averages μ_i with $\xi = \beta$.
Right: evolution of the averages μ_i with $\xi = \mu_1$.

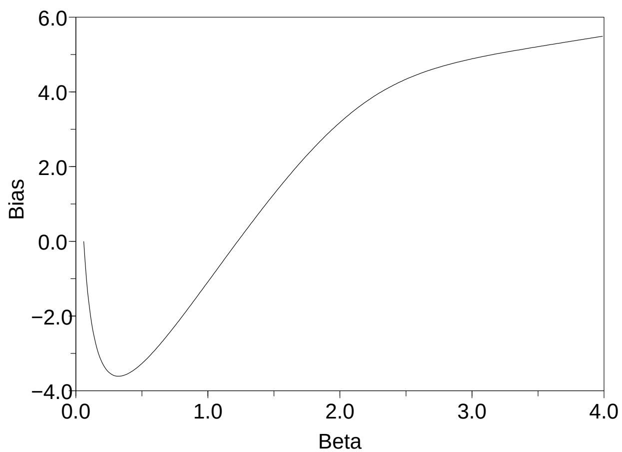
A good reaction coordinate seems to be $\xi = \beta$.

3.4 Extensions and another application





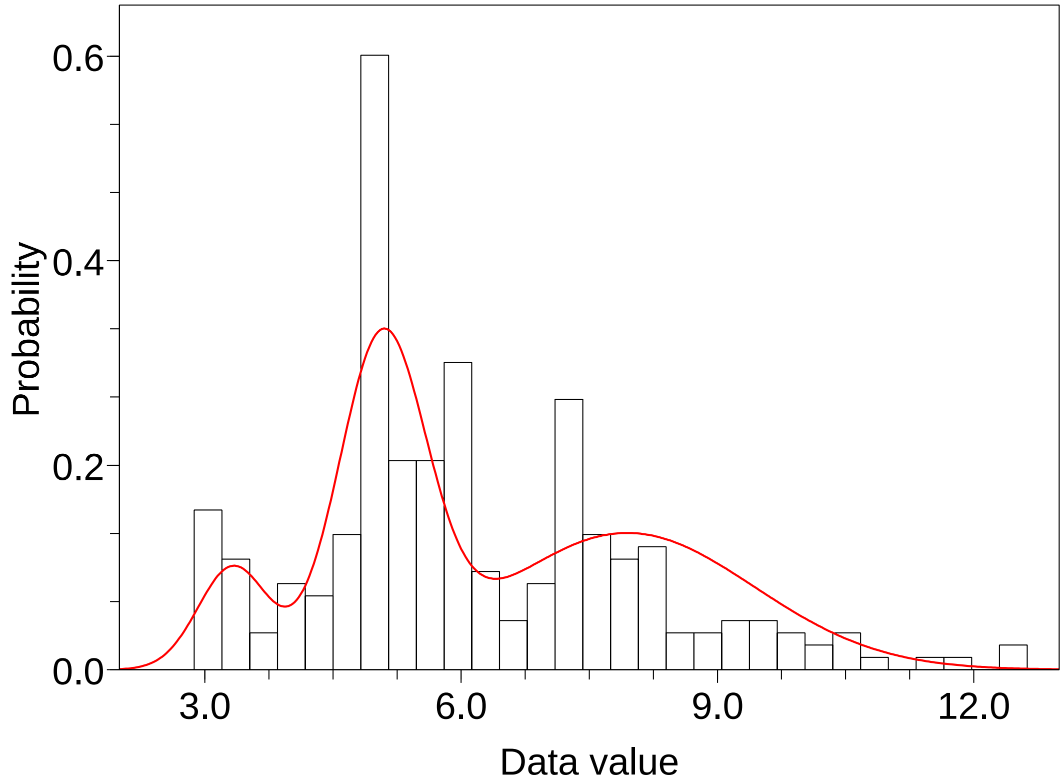
Converged bias.



The effective sample size is observed to be approximately $0.18N$.

3.4 Extensions and another application

Comparison of the mixture with the datas.

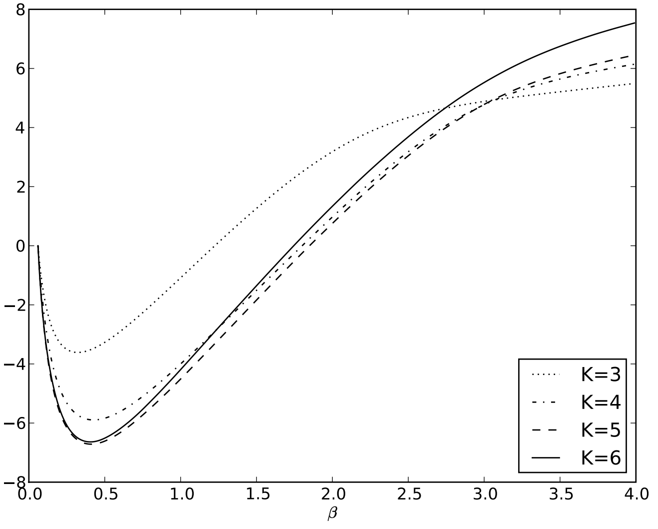


$$q_1 = 0.42227 \quad q_2 = 0.118506$$
$$\mu_1 = 5.1818 \quad \mu_2 = 3.29704 \quad \mu_3 = 7.79154$$

3.4 Extensions and another application

Why does it work with $\xi = \beta$? The bias is relatively small; forcing large values of β is forcing large values of the variances, which allows for a mixing of the components.

Extension: Bayesian model choice. Look for the best number of components. It seems that the bias (for $\xi = \beta$) for $K = 3$ is also a good bias for $K = 4$ and $K = 5$.



Acknowledgements

These are joint (on-going) works with :

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

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 Hamiltonian 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*, Mathematics of Computation, **78**, 2047-2074 (2009).

- 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*, Journal of Funtional Analysis, **256(7)**, 2211-2221 (2009).

- Effective dynamics:

- F. Legoll, TL, *Effective dynamics using conditional expectations*, `http://hal.archives-ouvertes.fr/hal-00399986`.

Forthcoming book: TL, M. Rousset and G. Stoltz, *Free energy computations: a mathematical perpsective*, Imperial College Press.

And also: http://cermics.enpc.fr/˜lelievre/rapports/ECODOQUI_notes.pdf