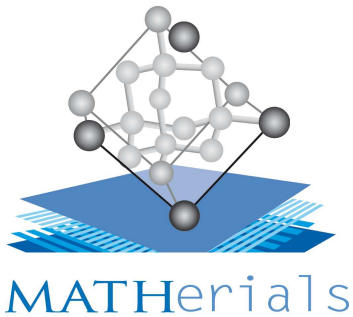
# Sampling problems in computational statistical physics

## 1- Free energy adaptive biasing methods

T. Lelièvre

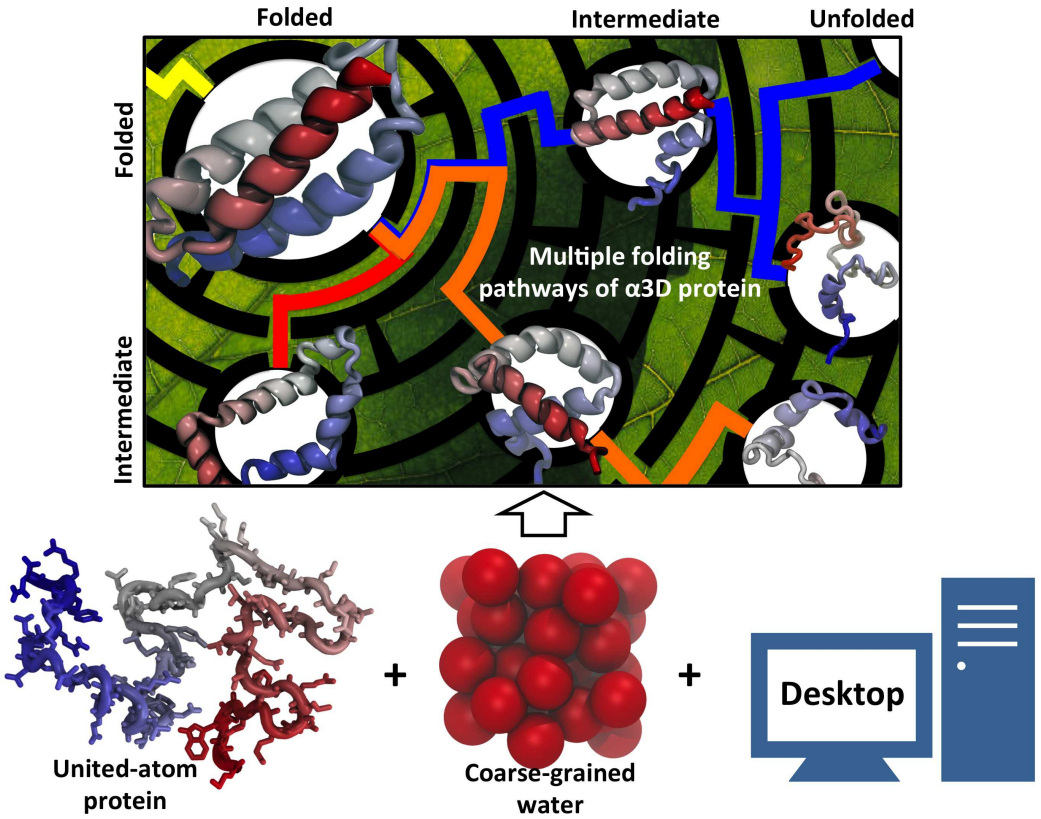
CERMICS - Ecole des Ponts ParisTech & Equipe Matherials - INRIA



*Brummer & Partners MathDataLab, KTH*, 18/01/2021

# Molecular dynamics




Simulation of protein folding (Courtesy of K. Schulten's group)

# Molecular dynamics

Molecular dynamics consists in simulating on the computer the evolution of atomistic systems, as a numerical microscope:

- Understand the link bewteen macroscopic properties and microscopic ingredients
- Explore matter at the atomistic scale
- Simulate new materials, new molecules
- Interprete experimental results

Applications: biology, chemistry, materials science

Molecular dynamics comes of age:

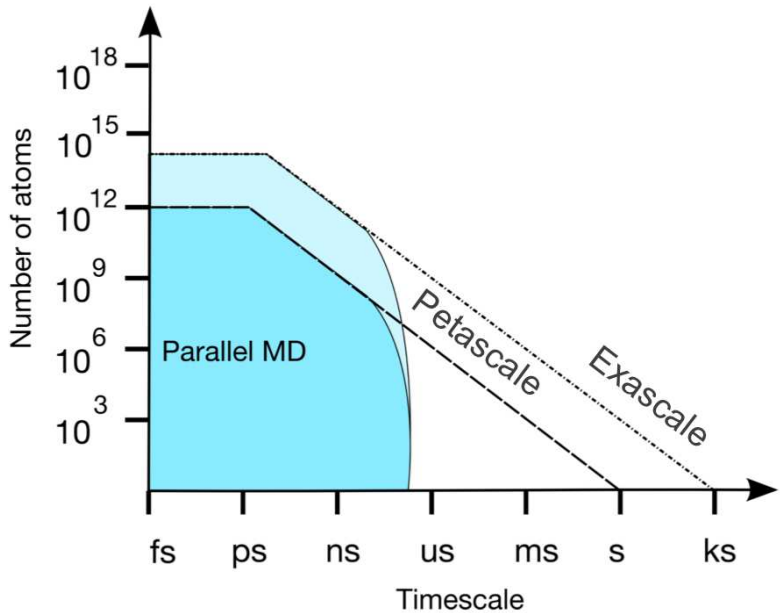
- 1/4 of CPU time worldwide is devoted to computations at the molecular scale
- 2013 Chemistry Nobel prize: Arieh Warshel, Martin Karplus and Michael Levitt. "Today the computer is just as important a tool for chemists as the test tube. Simulations are so realistic that they predict the outcome of traditional experiments."

# Challenges

Main challenges:

- Improved models (force fields, coarse-grained force fields): polarisability, water, chemical reactions
- Improved sampling methods (access long time scales): thermodynamic quantities and dynamical properties
- Incorporate data: Bayesian approaches, data sciences

Spatial parallelism is very effective, but temporal reach of heroic brute force MD is limited to $1\mu s$ or less.




Courtesy of Danny Perez (LANL)

# Challenges

Why is mathematics useful?

- Rigorous links between models at different scales (space and time): coarse-graining, coupling algorithms, certification
- Improve and analyze algorithms (efficiency, robustness, error analysis)
- Develop new algorithms, in particular on parallel architectures
- Modern data assimilation techniques

... but MathSciNet hints: fluid 95543, Navier Stokes 24026, Molecular dynamics 3166, ... ???

# Challenges

Examples of hot topics in mathematics for MD:

- Sampling of probability measures on manifolds, constrained MD (P. Breiding, P. Diaconis, J. Goodman, TL, ...)
- Sampling of reactive trajectories, rare event sampling (A. Guyader, C. Hartmann, TL, E. Vanden Eijnden, J. Weare, ...) Lecture 2
- Sampling of non equilibrium stationary state, non-reversible dynamics (J. Bierkens, G. Stoltz, ...)
- Towards better force fields (G. Csanyi, C. Ortner, A.V. Shapeev, ...)
- Effective dynamics, Mori-Zwanzig (T. Hudson, F. Legoll, TL, W. Zhang, ...)
- Sampling of metastable dynamics, accelerated dynamics methods (D. Aristoff, TL, D. Perez, A. Voter) Lecture 3

Today: Free energy and adaptive biasing methods

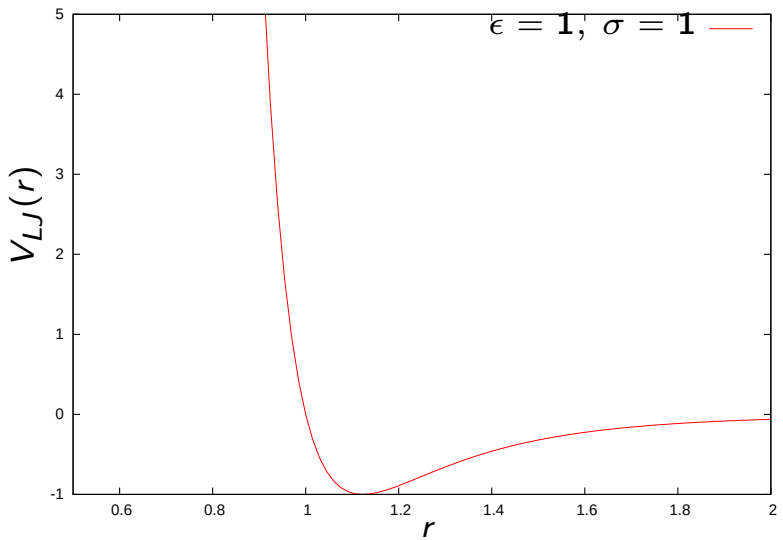
# The force field

The basic modelling ingredient: a potential $V$ which associates to a configuration $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N) = \boldsymbol{x} \in \mathbb{R}^{3N}$ an energy $V(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$.

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.



# Dynamics

Newton equations of motion:

$$\begin{cases} d\boldsymbol{X}_t = M^{-1}\boldsymbol{P}_t\, dt, \\ d\boldsymbol{P}_t = -\nabla V(\boldsymbol{X}_t)\, dt, \end{cases}$$

# Dynamics

Newton equations of motion + thermostat: Langevin dynamics:

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

where $\gamma > 0$. Langevin dynamics is ergodic wrt $\mu(d\boldsymbol{x}) \otimes Z_p^{-1} \exp\left(-\beta \frac{\boldsymbol{p}^t M^{-1}\boldsymbol{p}}{2}\right) d\boldsymbol{p}$ with

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

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.

# Dynamics

Newton equations of motion + thermostat: Langevin dynamics:

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

where $\gamma > 0$. Langevin dynamics is ergodic wrt $\mu(d\boldsymbol{x}) \otimes Z_p^{-1} \exp\left(-\beta \frac{\boldsymbol{p}^t M^{-1}\boldsymbol{p}}{2}\right) d\boldsymbol{p}$ with

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

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.

In the following, we focus on the *over-damped Langevin* (or gradient) dynamics

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

which is also ergodic wrt $\mu$.

# From micro to macro

These dynamics are used to compute macroscopic quantities:

(i) Thermodynamic quantities: averages wrt $\mu$ of some observables. Examples: stress, heat capacity, free energy,... These are approximated by trajectorial averages over dynamics ergodic wrt $\mu$:

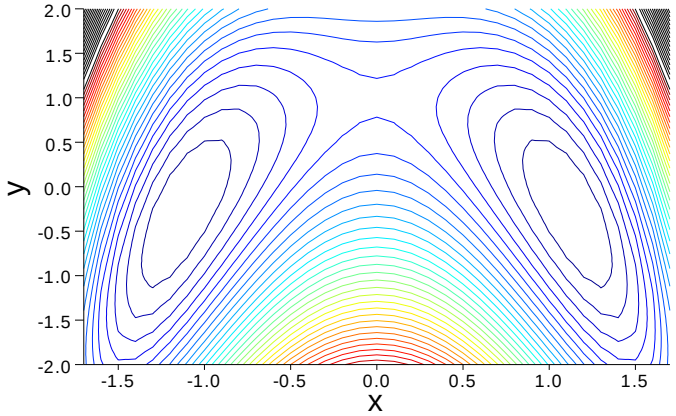
$$\int_{\mathbb{R}^d} \varphi(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \simeq \frac{1}{T} \int_0^T \varphi(\boldsymbol{X}_t)\, dt.$$

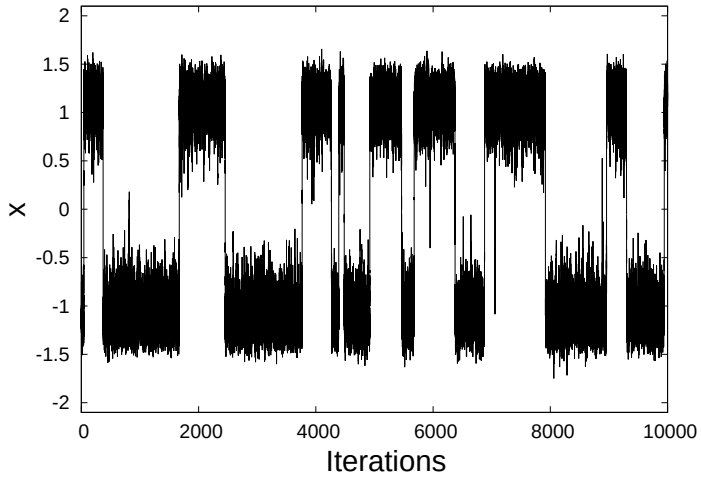
(ii) Dynamical quantities: statistical quantities depending on the trajectories. Examples: transition rates, transition paths, diffusion coefficients, viscosity,... These require to sample trajectories.

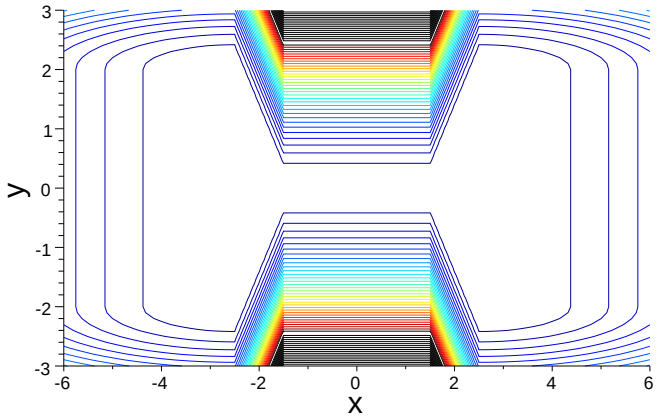
Difficulties: (i) high-dimensional problem ($N \gg 1$); (ii) $\boldsymbol{X}_t$ is a metastable process and $\mu$ is a multimodal measure.

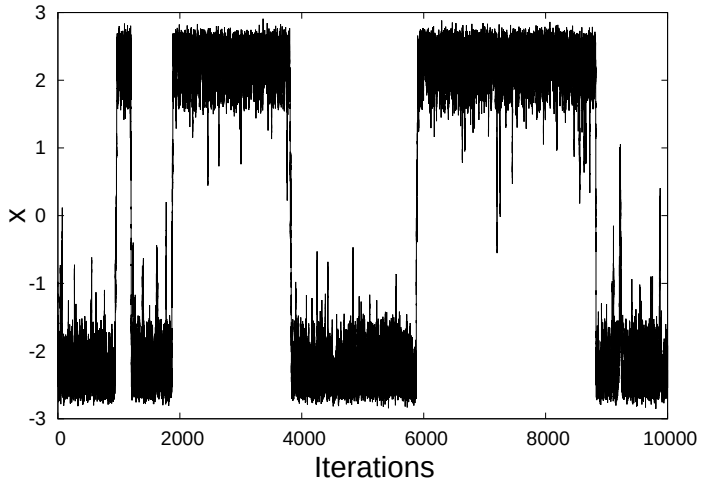
# Metastability: energetic and entropic barriers

## A two-dimensional schematic picture



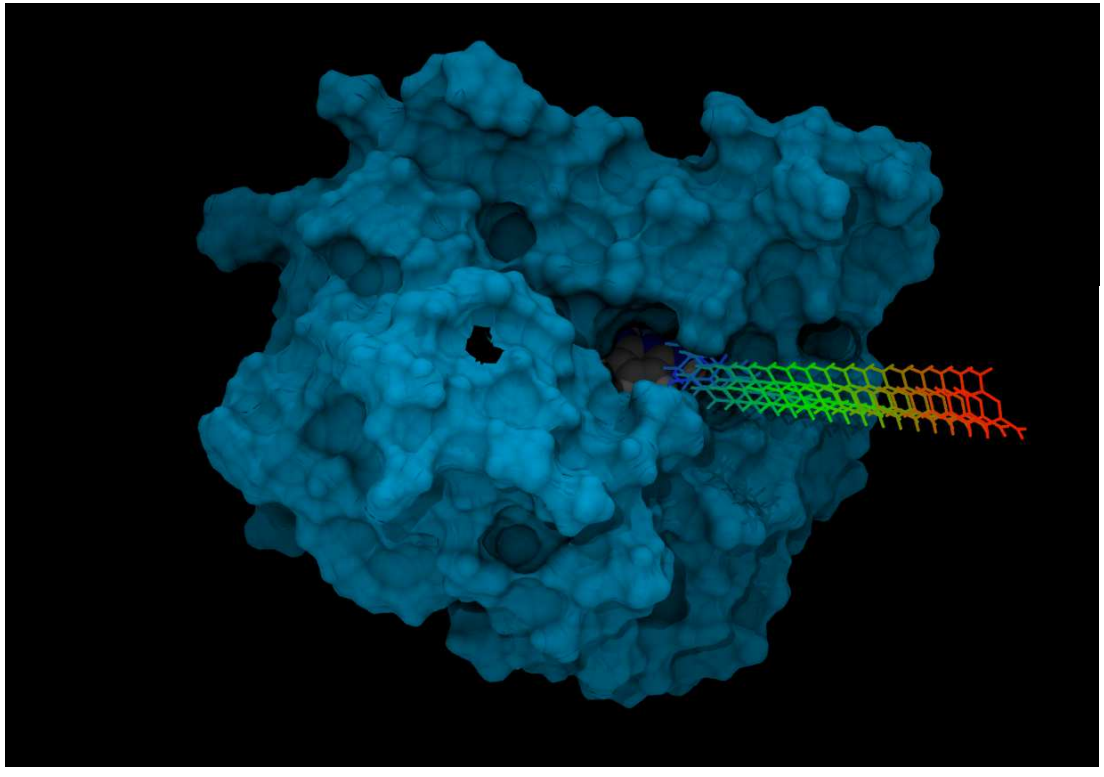

$\longrightarrow$ • Slow convergence of trajectorial averages

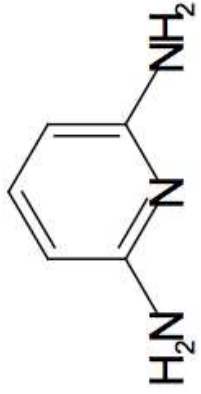
• Transitions between metastable states are rare events

# Simulations of biological systems

## Unbinding of a ligand from a protein

(Diaminopyridine-HSP90, Courtesy of SANOFI)






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

Dissociation time $= 0.5\,\mathrm{s}$

# Limitation of direct molecular dynamics

Direct molecular dynamics is a very powerful technique to generate atomistic trajectories. These trajectories can be useful in themselves (dynamical quantities) or to get ensemble averages (thermodynamic quantities).

Orders of magnitude: LJ potential costs $\sim 2\mu s$/atom/timestep; EAM potential costs $\sim 5\mu s$/atom/timestep; AIMD costs (at least) 1 min/atom/timestep.

Thus, molecular dynamics' reach is limited in terms of time and length scales. $\longrightarrow$ Depending on the quantity of interest, MD is combined with other algorithms to get better sampling.

Thermodynamic quantities: variance reduction methods (importance sampling, stratification, control variate, ...)
Dynamic quantities: accelerated dynamics (using Markov State models), rare event sampling methods (splitting), ...

# Sampling problems in molecular dynamics

In molecular dynamics, we would like to sample:

- multimodal measures in high-dimension,
- and metastable dynamics in high-dimension.

Outline of the lectures:

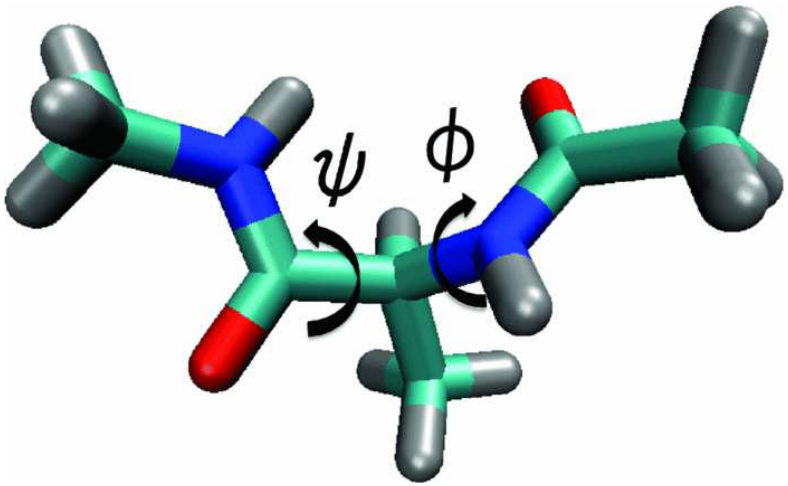
- Lecture 1: Free energy and adaptive biasing methods (importance sampling)
- Lecture 2: Sampling of reactive trajectories, rare event sampling (interacting particle systems, splitting methods)
- Lecture 3: Accelerated dynamics methods (the QSD approach to study the exit event from a metastable state)

# Outline

We will present adaptive biasing techniques which are based on the computation of the free energy:

1. Adaptive biasing force techniques *Mathematical tool: Entropy techniques and Logarithmic Sobolev Inequalities.*
2. Wang Landau algorithm *Mathematical tool: Convergence of Stochastic Approximation Algorithms.*

# Free energy and adaptive biasing techniques




[Esque, Cecchini, J. Phys. Chem. B, 2015.]

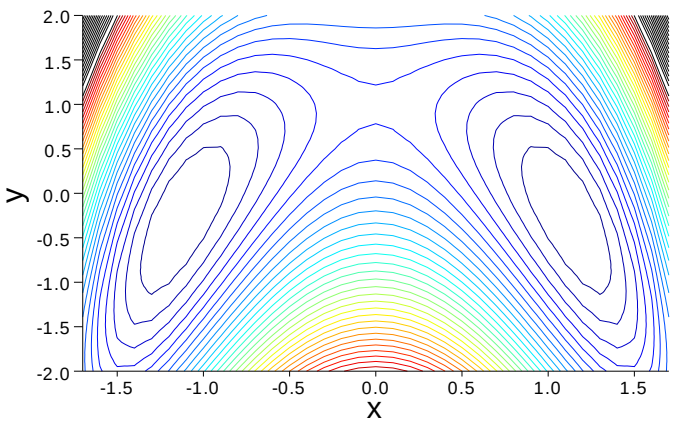
# Reaction coordinate

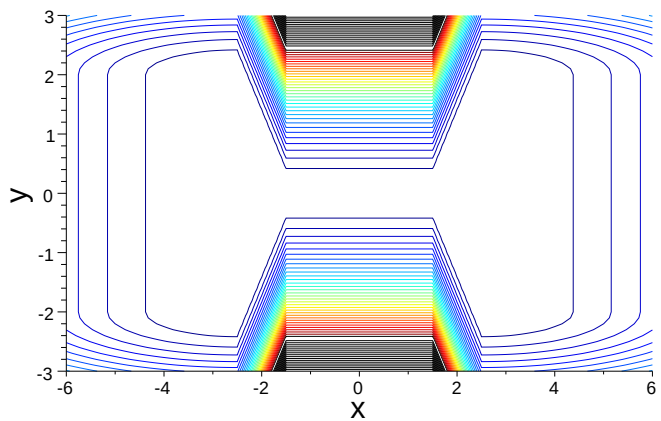
Let us be given a slow variable of interest $\xi(\boldsymbol{X}_t)$ (say of dimension 1 and periodic $\xi : \mathbb{R}^d \to \mathbb{T}$). The function $\xi$ is called reaction coordinate, collective variable, order parameter,...

This reaction coordinate can be used to efficiently sample the canonical measure: (i) constrained dynamics (thermodynamic integration) and (ii) biased dynamics (adaptive free energy importance sampling technique).

Free energy will play a central role.

For example, in the 2D simple examples: $\xi(x, y) = x$.



# Free energy

Let us introduce the image of the measure $\mu$ by $\xi$:

$$\xi_* \mu\,(dz).$$

If $\boldsymbol{X} \sim \mu$, then $\xi(\boldsymbol{X}) \sim \xi_* \mu$.

The free energy $A$ is defined by:

$$\exp(-\beta A(z))\, dz = \xi_* \mu\,(dz)$$

namely

$$A(z) = -\beta^{-1} \ln \left( \int_{\Sigma(z)} e^{-\beta V} \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x}) \right),$$

where $\Sigma(z) = \{\boldsymbol{x},\, \xi(\boldsymbol{x}) = z\}$ is a submanifold of $\mathbb{R}^d$, and $\delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})\, dz = d\boldsymbol{x}$.

# Free energy (2d case)

In the simple case $\xi(x, y) = x$, the image of the measure $\mu$ by $\xi$ is

$$\xi_* \mu\,(dx) = \left( \int_{\Sigma(x)} e^{-\beta V(x,y)} dy \right) dx$$

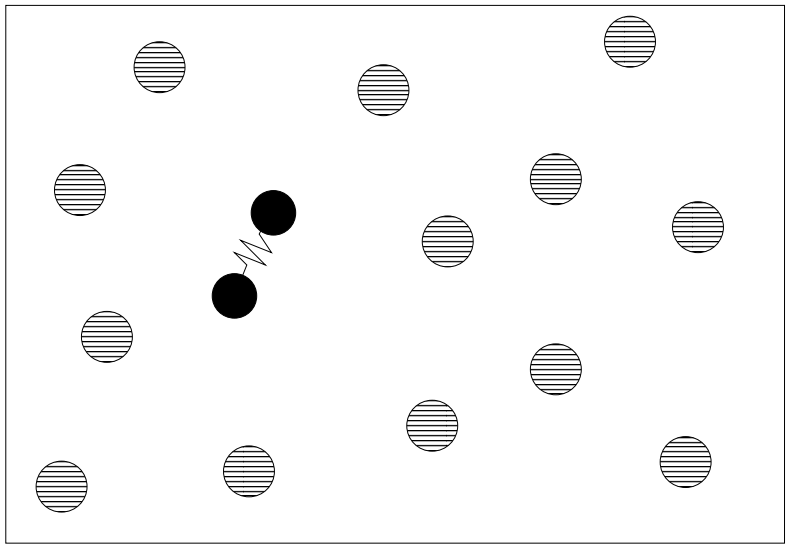
where $\Sigma(x) = \{(x, y),\ y \in \mathbb{R}\}$.

And thus the free energy $A$ is defined by:

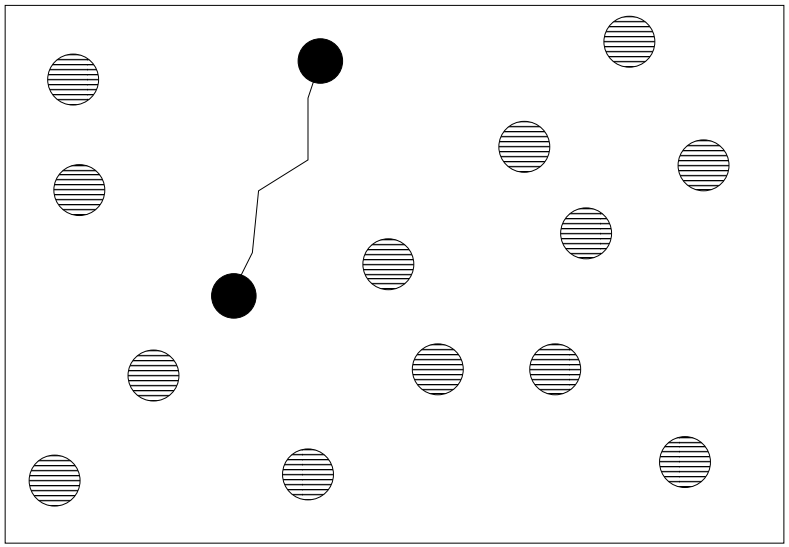
$$A(x) = -\beta^{-1} \ln \left( \int_{\Sigma(x)} e^{-\beta V(x,y)} dy \right).$$

# Free energy on a simple example (1/2)

What is free energy? A toy model for the solvation of a dimer [Dellago, Geissler].



Compact state.

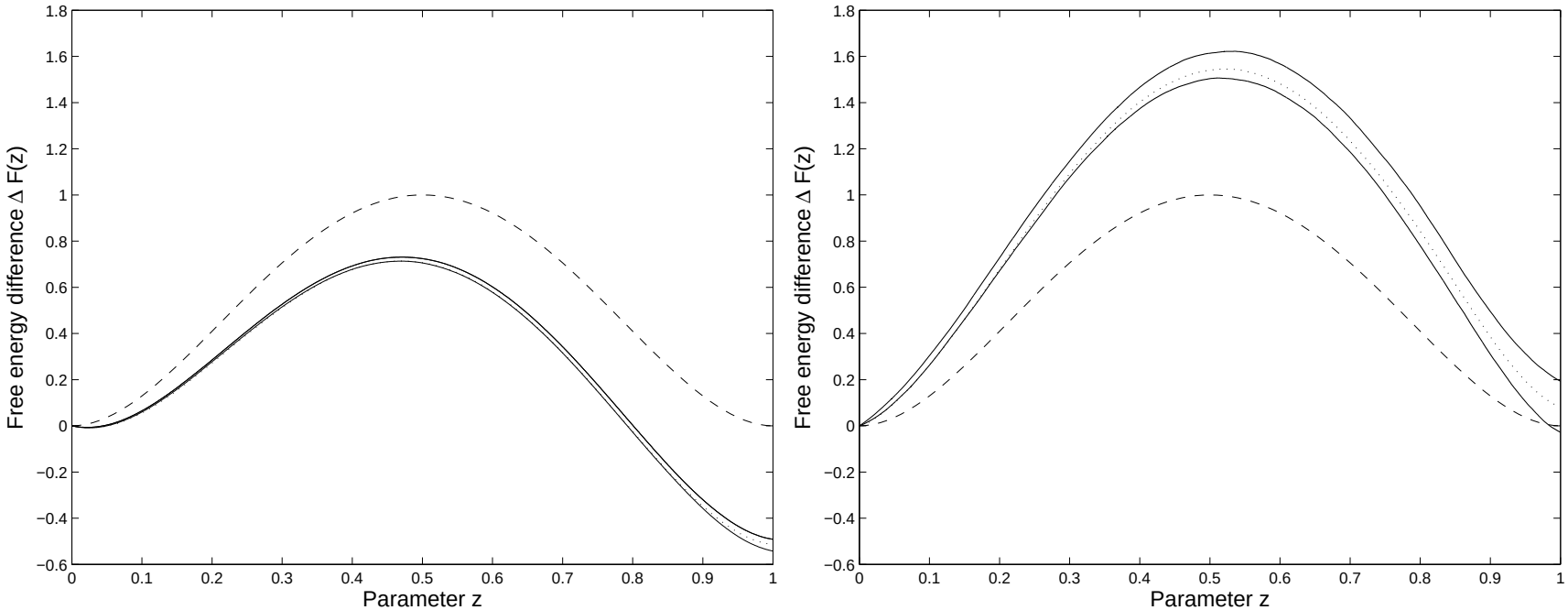


Stretched state.

The particles interact through a pair potential: truncated LJ for all particles except the two monomers (black particles) which interact through a double-well potential. A slow variable is the distance between the two monomers.

# Free energy on a simple example (2/2)

Profiles computed using thermodynamic integration.




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.

# Free energy

The free energy $A$ associated with the reaction coordinate $\xi$ (angle, length, ...) can be seen as an effective potential along $\xi$.

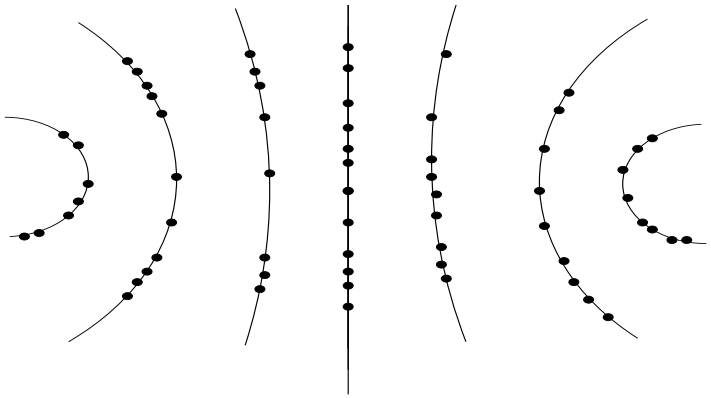
Free energy is nowadays one of the major thermodynamic quantity that the practitioners would like to compute.

The associated Boltzmann-Gibbs measure $\exp(-\beta A(z))\, dz$ is exact in terms of thermodynamic quantities.

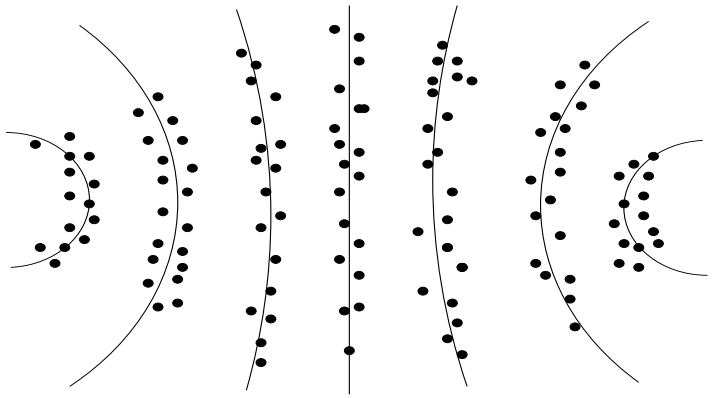
*Remark*: Interesting question: what is the dynamical content of $A$?

# Free energy calculation techniques

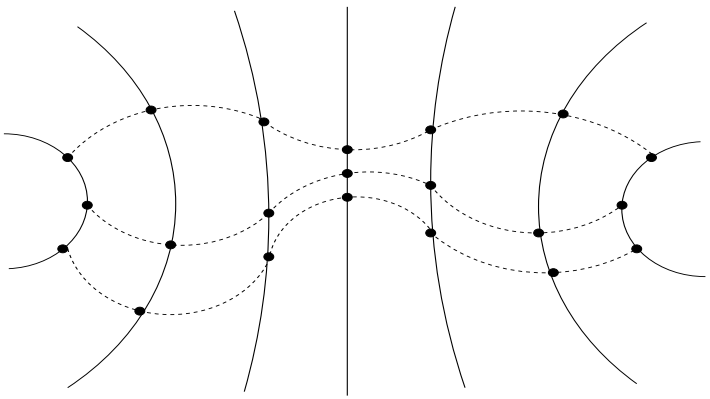
There are many free energy calculation techniques:



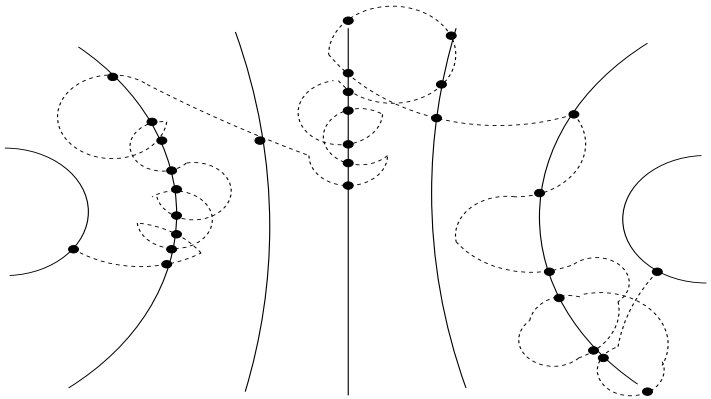
(a) Thermodynamic integration.



(b) Histogram method.



(c) Non equilibrium dynamics.



(d) Adaptive dynamics.

# Adaptive biasing techniques

The bottom line of adaptive methods is the following: for "well chosen" $\xi$ the potential $V - A \circ \xi$ is less rugged than $V$. Indeed, by construction $\xi_* \exp(-\beta(V - A \circ \xi)) = 1_{\mathbb{T}}$.

Problem: $A$ is unknown ! Idea: use a time dependent potential of the form

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

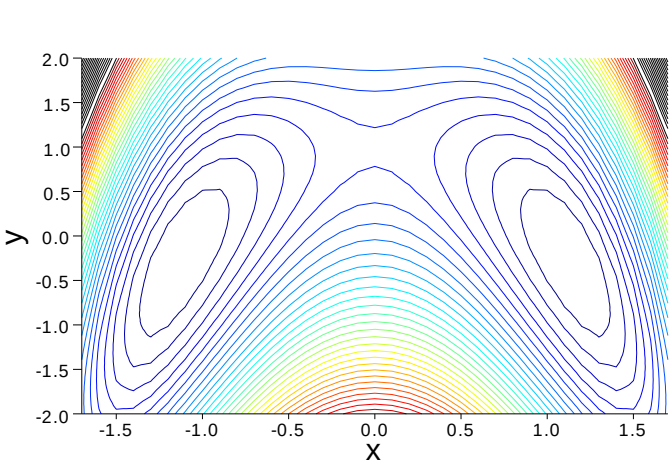
where $A_t$ is an approximation at time $t$ of $A$, given the configurations visited so far.

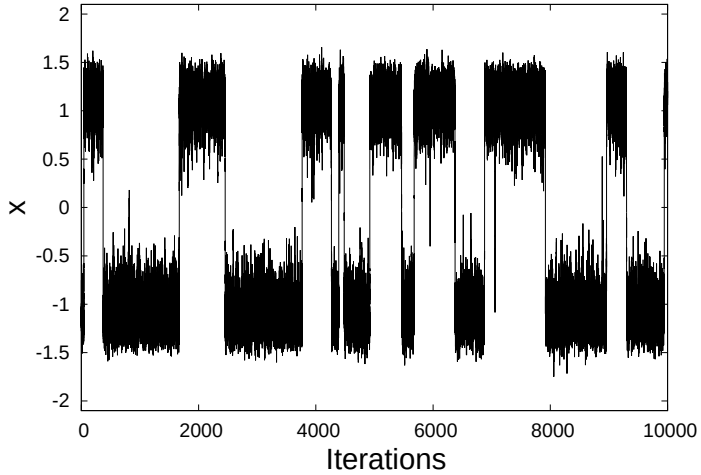
Hopes:

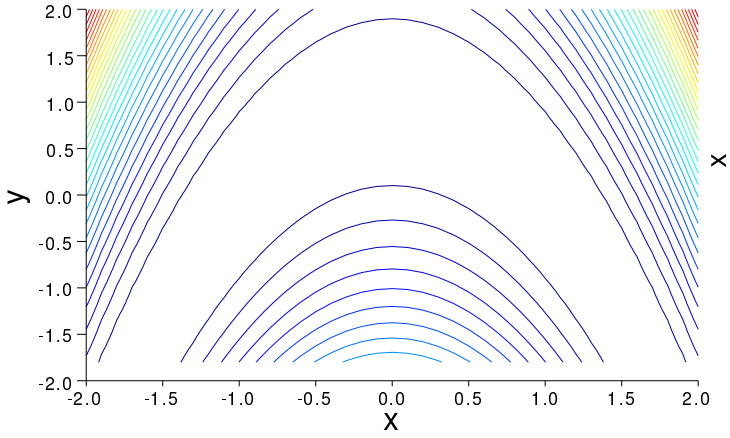
- build a dynamics which goes quickly to equilibrium,
- compute free energy profiles.

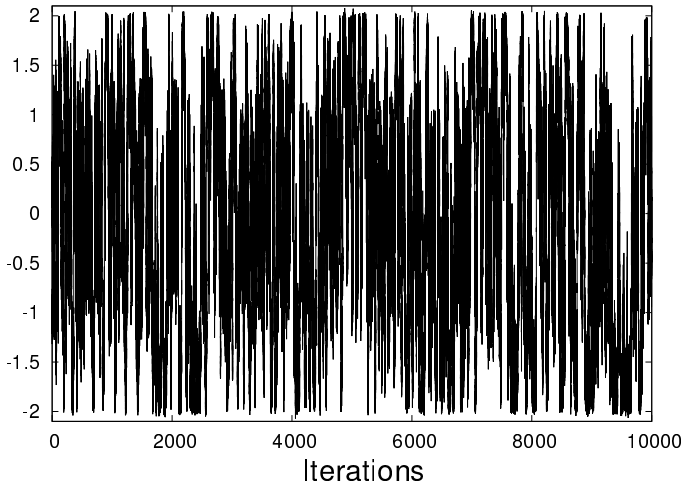
Wang-Landau, ABF, metadynamics: *Darve, Pohorille, Hénin, Chipot, Laio, Parrinello, Wang, Landau,...*

# Free energy biased dynamics (1/2)



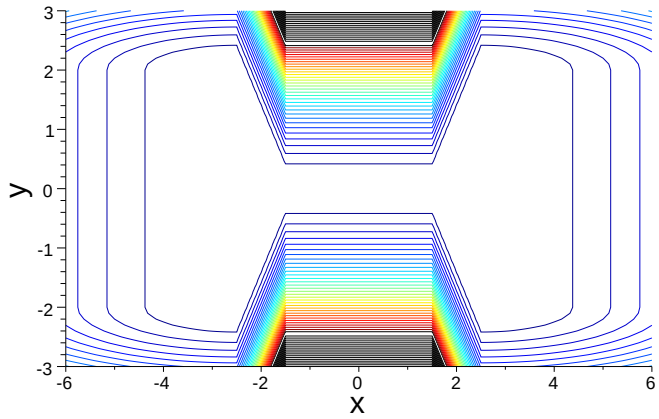



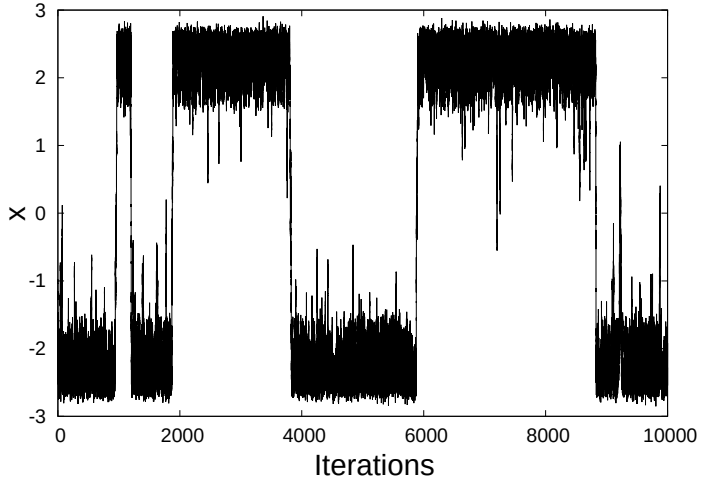



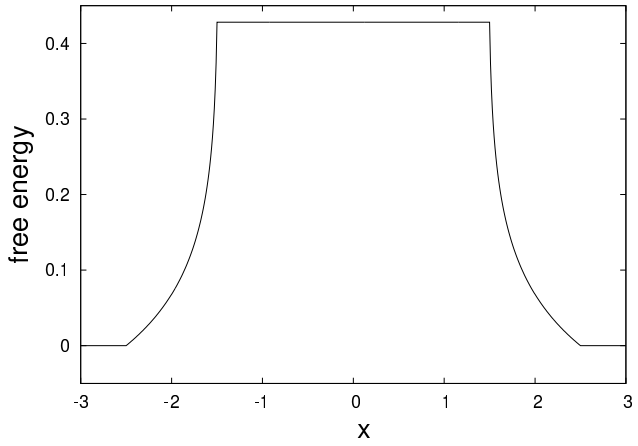



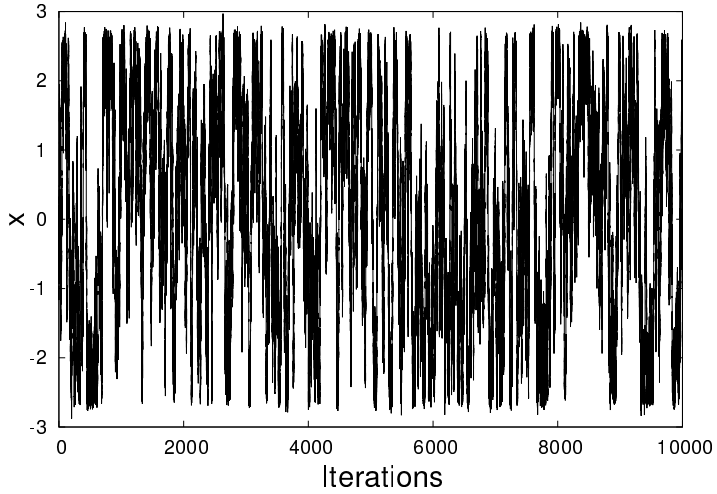

A 2D example of a free energy biased trajectory: energetic barrier.

# Free energy biased dynamics (2/2)




A 2D example of a free energy biased trajectory: entropic barrier.

# Updating strategies

How to update $A_t$? Two methods depending on wether $A'_t$ (Adaptive Biasing Force) or $A_t$ (Adaptive Biasing Potential) is approximated.
To avoid geometry problem, an extended configurational space $(\boldsymbol{x}, z) \in \mathbb{R}^{n+1}$ may be considered, together with the meta-potential:

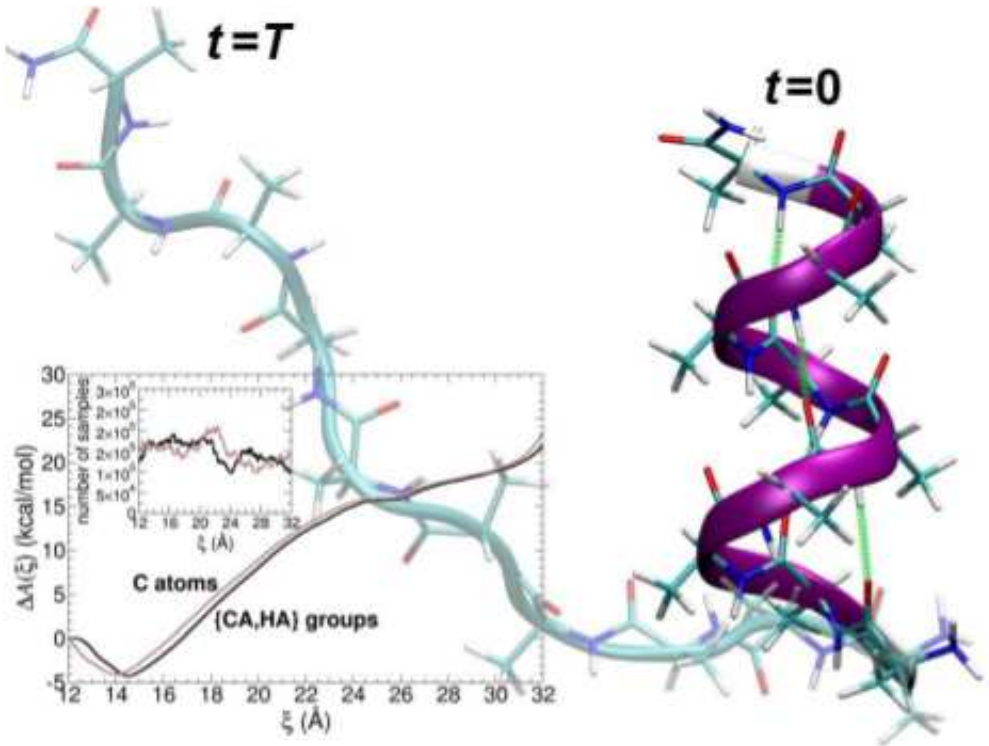
$$V^k(\boldsymbol{x}, z) = V(\boldsymbol{x}) + k(z - \xi(\boldsymbol{x}))^2.$$

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).

Adaptive algorithms used in molecular dynamics fall into one of these four possible combinations [TL, M. Rousset, G. Stoltz, J Chem Phys, 2007]:

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

# The Adaptive Biasing Force method



# The ABF method

For the Adaptive Biasing Force (ABF) method [Darve, Pohorille, Chipot, Hénin], the idea is to use the formula

$$A'(z) = \frac{\displaystyle\int \left( \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left( \frac{\nabla \xi}{|\nabla \xi|^2} \right) \right) e^{-\beta V} \, \delta_{\xi(x)-z}(dx)}{\displaystyle\int e^{-\beta V} \, \delta_{\xi(x)-z}(dx)}$$

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

The mean force $A'(z)$ is the mean of $f$ with respect to $\mu_{\Sigma(z)}$.

# The ABF method

In the simple case $\xi(x,y) = x$, remember that

$$A(x) = -\beta^{-1} \ln \left( \int e^{-\beta V(x,y)} dy \right),$$

so that

$$A'(x) = \frac{\int \partial_x V \, e^{-\beta V(x,y)} \, dy}{\int e^{-\beta V(x,y)} \, dy}$$

$$= \int \partial_x V \, d\mu_{\Sigma(x)}.$$

Notice that actually, whatever $A_t$ is,

$$A'(z) = \frac{\int f e^{-\beta(V - A_t \circ \xi)} \, \delta_{\xi(x)-z}(dx)}{\int e^{-\beta(V - A_t \circ \xi)} \, \delta_{\xi(x)-z}(dx)}.$$

# The ABF method

Thus, we would like to simulate:

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

but $A$ is unknown...

# The ABF method

The ABF dynamics is then:

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

# The ABF method

The ABF dynamics is then:

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

The associated (nonlinear) Fokker-Planck equation writes:

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

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

# The ABF method

The ABF dynamics is then:

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

The associated (nonlinear) Fokker-Planck equation writes:

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

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

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

# Longtime convergence and entropy (1/3)

Recall the original gradient dynamics:

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

The associated (linear) Fokker-Planck equation writes:

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

where $\boldsymbol{Q}_t \sim \phi(t, \boldsymbol{q})\, d\boldsymbol{q}$.

The metastable behaviour of $\boldsymbol{Q}_t$ is related to the multimodality of $\mu$, which can be quantified through the rate of convergence of $\phi$ to $\phi_\infty = Z^{-1} \exp(-\beta V)$.

A classical approach for partial differential equations (PDEs): entropy techniques.

# Longtime convergence and entropy (2/3)

Notice that the Fokker-Planck equation rewrites

$$\partial_t \phi = \beta^{-1} \mathrm{div} \left( \phi_\infty \nabla \left( \frac{\phi}{\phi_\infty} \right) \right).$$

Let us introduce the entropy:

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

We have (Csiszár-Kullback inequality):

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

# Longtime convergence and entropy (3/3)

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

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

$$H(\phi|\phi_\infty) \leq \frac{1}{2R}I(\phi|\phi_\infty)$$

then $E(t) \leq E(0)\exp(-2\beta^{-1}Rt)$ and thus $\phi$ converges to $\phi_\infty$ exponentially fast with rate $\beta^{-1}R$.

Metastability $\iff$ small $R$

# Convergence of ABF (1/2)

A convergence result [TL, M. Rousset, G. Stoltz, Nonlinearity 2008]: Recall the ABF Fokker-Planck equation:

$$\begin{cases} \partial_t \psi = \mathrm{div}\, \left( \nabla (V - A_t \circ \xi) \psi + \beta^{-1} \nabla \psi \right), \\ A_t'(z) = \dfrac{\int f\, \psi\, \delta_{\xi(x)-z}(dx)}{\int \psi\, \delta_{\xi(x)-z}(dx)}. \end{cases}$$

Suppose:

(H1) "Strong ergodicity" of the microscopic variables: the conditional probability measures $\mu_{\Sigma(z)}$ satisfy a LSI($\rho$),

(H2) Bounded coupling: $\left\| \nabla_{\Sigma(z)} f \right\|_{L^\infty} < \infty$,
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 $\overline{\psi} = \int \psi\, \delta_{\xi(x)-z}(dx)$ to $\overline{\psi_\infty}$,
- the LSI constant $\rho$ (the real limitation).

# Convergence of ABF (2/2)

In summary:

- Original gradient dynamics: $\exp(-\beta^{-1}Rt)$ where $R$ is the LSI constant for $\mu$;
- ABF dynamics: $\exp(-\beta^{-1}\rho t)$ where $\rho$ is the LSI constant for the conditioned probability measures $\mu_{\Sigma(z)}$.

If $\xi$ is well chosen, $\rho \gg R$: the free energy can be computed very efficiently.

Two ingredients of the proof:

(1) The marginal $\overline{\psi}(t,z) = \int \psi(t,\boldsymbol{x})\,\delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})$ satisfies a closed PDE:

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

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

# Convergence of ABF (3)

(2) The total entropy can be decomposed as [N. Grunewald, F. Otto, C. Villani, M. Westdickenberg, Ann. IHP, 2009]:

$$E = E_M + E_m$$

where

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

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

The microscopic entropy is

$$E_m = \int H\Big(\psi(\cdot|\xi(\boldsymbol{x}) = z)\Big|\psi_\infty(\cdot|\xi(\boldsymbol{x}) = z)\Big)\,\overline{\psi}(z)\,dz.$$

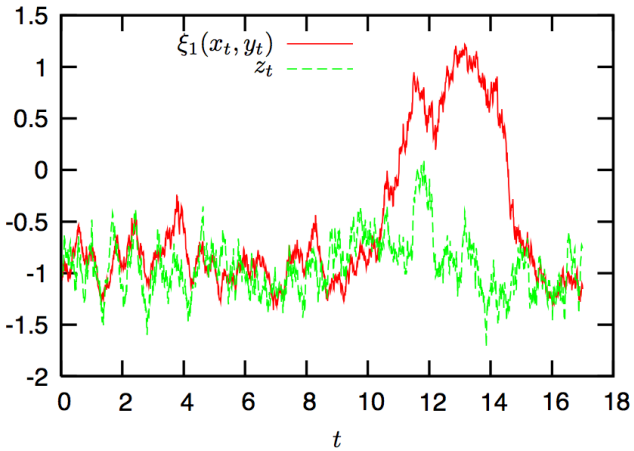
We already know that $E_M$ goes to zero: it remains only to consider $E_m$...

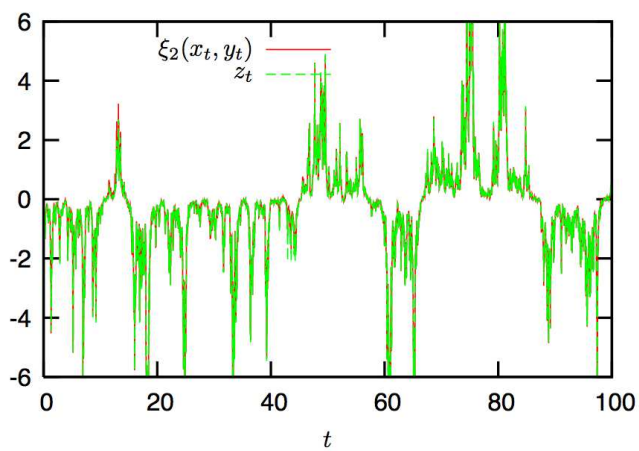
# Other results based on this set of assumptions:

- [TL, JFA 2008] LSI for the cond. meas. $\mu_{\Sigma(z)}$
  + LSI for the marginal $\overline{\mu}(dz) = \xi_* \mu(dz)$
  + bdd coupling ($\|\nabla_{\Sigma(z)} f\|_{L^\infty} < \infty$) $\Longrightarrow$ LSI for $\mu$.

- [F. Legoll, TL, U. Sharma, W. Zhang, 2010-2019] Effective dynamics for $\xi(\boldsymbol{Q}_t)$. Uniform control in time:

$$H(\mathcal{L}(\xi(\boldsymbol{Q}_t))|\mathcal{L}(z_t)) \leq C \left( \frac{\|\nabla_{\Sigma(z)} f\|_{L^\infty}}{\rho} \right)^2 H(\mathcal{L}(\boldsymbol{Q}_0)|\mu).$$



# Discretization of ABF

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

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

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

  This approach is easy to parallelize, flexible (selection mechanisms) and efficient in cases with multiple reactive paths. [TL, M. Rousset, G. Stoltz, 2007; C. Chipot, TL, K. Minoukadeh, 2010 ; TL, K. Minoukadeh, 2010]

- Use trajectorial averages along a single path:

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

  The longtime behavior is much more difficult to analyze

# Back to the original problem

How to use free energy to compute canonical averages $\int \varphi d\mu = Z^{-1} \int \varphi e^{-\beta V}$?

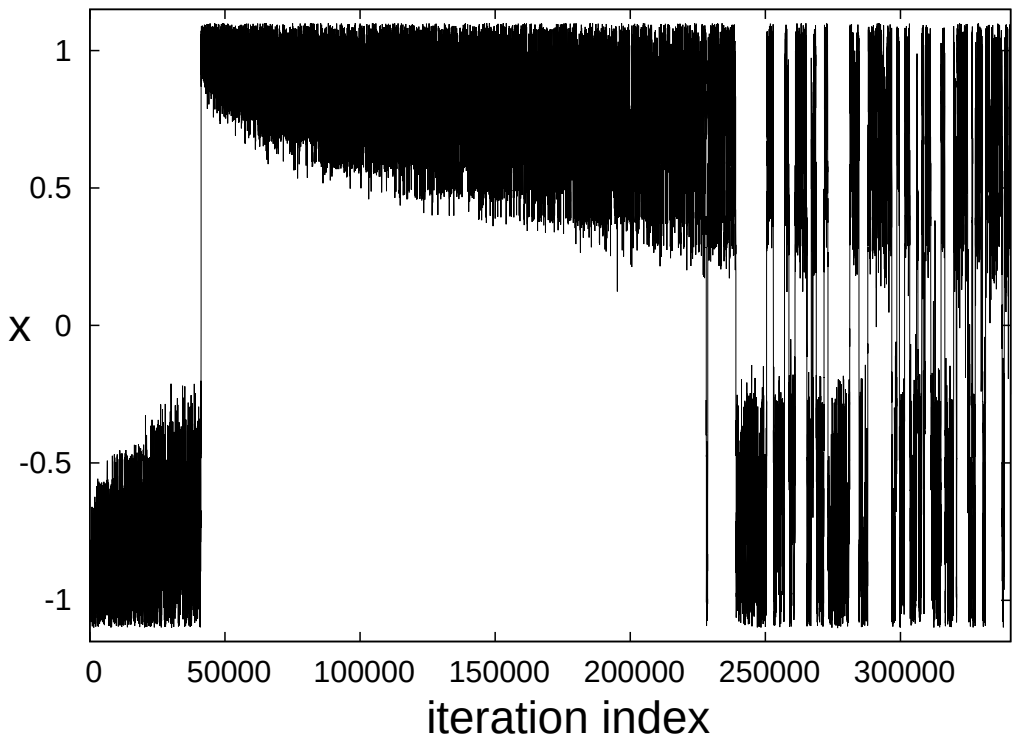
- Importance sampling:

$$\int \varphi \, d\mu = \frac{\int \varphi e^{-\beta A \circ \xi} \, Z_A^{-1} e^{-\beta(V - A \circ \xi)}}{\int e^{-\beta A \circ \xi} \, Z_A^{-1} e^{-\beta(V - A \circ \xi)}}.$$

- Conditioning:

$$\int \varphi \, d\mu = \frac{\int_z \left( \int_{\Sigma(z)} \varphi \, d\mu_{\Sigma(z)} \right) e^{-\beta A(z)} \, dz}{\int_z e^{-\beta A(z)} \, dz}.$$

# The Wang Landau algorithm



# The Wang Landau algorithm

The Wang Landau algorithm is one example of an Adaptive Biasing Potential strategy.

Let us present the algorithm in the following setting:

- The original (non-adaptive) dynamics is the Metropolis-Hastings algorithm with target measure $\exp(-V(x))\,dx$ ($\beta = 1$ for simplicity)
- The reaction coordinate is discrete: $I : \mathbb{R}^d \to \{1, \ldots, m\}$

The free energy $A$ is thus defined by: for $i \in \{1, \ldots, m\}$,

$$\exp(-A(i)) = Z^{-1} \int_{\mathbb{R}^d} 1_{\{I(x)=i\}} \exp(-V(x))\,dx$$

In addition, we will consider an algorithm where

- Trajectorial averages are used to approximate the free energy

# The original algorithm: Metropolis-Hastings

Iterate the following on $n \geq 1$:

1. Propose a move from $X_n$ to $Y_{n+1}$ according to a proposal kernel $q(x, y)\, dy$ (assumed to be symmetric for simplicity $q(x, y) = q(y, x)$)
2. With probability $\min(1, \exp[-V(Y_{n+1}) + V(X_n)])$ accept the move $(X_{n+1} = Y_{n+1})$ ; otherwise reject $(X_{n+1} = Y_n)$

For a multimodal target $\exp(-V(x))\, dx$, the dynamics is metastable, and convergence is very slow.

# The free energy biased algorithm

If the free energy $A$ was known, one would use instead the following algorithm:

Iterate the following on $n \geq 1$:

1. Propose a move from $X_n$ to $Y_{n+1}$ according to a proposal kernel $q(x, y)\, dy$ (assumed to be symmetric for simplicity $q(x, y) = q(y, x)$)
2. With probability $\min\left(1, \exp[-V(Y_{n+1}) + A(I(Y_{n+1})) + V(X_n) - A(I(X_n))]\right)$ accept the move $(X_{n+1} = Y_{n+1})$ ; otherwise reject $(X_{n+1} = Y_n)$

# The Wang Landau algorithm

Iterate the following on $n \geq 1$:

1. Propose a move from $X_n$ to $Y_{n+1}$ according to a proposal kernel $q(x, y)\, dy$ (assumed to be symmetric for simplicity $q(x, y) = q(y, x)$)
2. With probability $\min\left(1, \exp[-V(Y_{n+1}) + A_n(I(Y_{n+1})) + V(X_n) - A_n(I(X_n))]\right)$ accept the move $(X_{n+1} = Y_{n+1})$ ; otherwise reject $(X_{n+1} = Y_n)$
3. Update $A_n$: for $i \in \{1, \ldots, m\}$,

$$\exp(-A_{n+1}(i)) = \exp(-A_n(i))(1 + \gamma_{n+1} 1_{\{I(X_{n+1})=i\}}).$$

Here, $(\gamma_n)_{n\geq 1}$ is a given sequence of penalizing weights $(\gamma_n > 0)$. The sequence $(\gamma_n)_{n\geq 1}$ should converge to zero (to observe convergence) but not too fast (premature convergence).

The idea is to penalize already visited states, the states being indexed by $I$.

# Heuristics (1/2)

We expect $\theta_n(i) := \frac{\exp(-A_n(i))}{\sum_{j=1}^m \exp(-A_n(j))}$ to converge to $\exp(-A(i))$.

Why?

$$\begin{aligned}\sum_{j=1}^m \exp(-A_{n+1}(j)) &= \sum_{j=1}^m \exp(-A_n(j))[1+\gamma_{n+1}1_{\{I(X_{n+1})=j\}}] \\ &= \sum_{j=1}^m \exp(-A_n(j)) + \gamma_{n+1}\exp(-A_n(I(X_{n+1}))) \\ &= \sum_{j=1}^m \exp(-A_n(j))\left[1+\gamma_{n+1}\theta_n(I(X_{n+1}))\right]\end{aligned}$$

and thus

$$\begin{aligned}\theta_{n+1}(i) &= \frac{\exp(-A_n(i))(1+\gamma_{n+1}1_{\{I(X_{n+1})=i\}})}{\sum_{j=1}^m \exp(-A_n(j))\left[1+\gamma_{n+1}\theta_n(I(X_{n+1}))\right]} \\ &= \theta_n(i)\left[1+\gamma_{n+1}\left(1_{\{I(X_{n+1})=i\}}-\theta_n(I(X_{n+1}))\right)\right]+\mathcal{O}(\gamma_{n+1}^2)\end{aligned}$$

# Heuristics (2/2)

$$\theta_{n+1}(i) = \theta_n(i) + \gamma_{n+1}\theta_n(i)\left(1_{\{I(X_{n+1})=i\}} - \theta_n(I(X_{n+1}))\right) + \mathcal{O}(\gamma_{n+1}^2)$$

If $X_{n+1}$ was at equilibrium wrt the target measure

$$Z_n^{-1}\exp(-(V(x) - A_n(I(x)))\,dx = \tilde{Z}_n^{-1}\exp(-V(x))/\theta_n(I(x)),$$

where $\tilde{Z}_n = \sum_{i=1}^m \exp(-A(i))/\theta_n(i)$, the average drift would be:

$$\begin{aligned}&\tilde{Z}_n^{-1}\int \theta_n(i)\left(1_{\{I(x)=i\}} - \theta_n(I(x))\right)\exp(-V(x))/\theta_n(I(x))\,dx\\ &= \tilde{Z}_n^{-1}\left[\exp(-A(i)) - \theta_n(i)\right].\end{aligned}$$

Thus, up to fluctuations,

$$\theta_{n+1}(i) \simeq \theta_n(i) + \gamma_{n+1}\tilde{Z}_n^{-1}\left[\exp(-A(i)) - \theta_n(i)\right].$$

If $\theta_n$ converges when $n \to \infty$, it is thus expected that $\theta_n \to \exp(-A)$.

# Convergence result (1/3)

The convergence of $\theta_n$ to $\exp(-A)$ can be obtained using results for the convergence of SAA [Andrieu, Moulines, Priouret], assuming that

$$\sum_{n\geq 1} \gamma_n = \infty \text{ and } \sum_{n\geq 1} \gamma_n^2 < \infty.$$

This requires in particular to prove the recurrence property:

$$\text{almost surely, } \limsup_{n\to\infty} \min_{j\in\{1,\ldots,d\}} \theta_n(j) > 0.$$

The sequence $(\theta_n)_{n\geq 0}$ returns infinitely often to a compact set of $\Theta = \left\{(\theta(1),\ldots,\theta(d)),\, \theta(i) > 0, \sum_{i=1}^d \theta(i) = 1\right\}$.

Under additional assumptions on $(\gamma_n)_{n\geq 1}$, it is possible to prove a central limit theorem on $\sqrt{\gamma_n}(\theta_n - \exp(-A))$, see [Fort].

# Convergence result (2/3)

Once $\theta_n$ is known to converge, one can apply results on the convergence of Markov chains with non-constant transition kernels [Fort, Moulines, Priouret] to obtain:

$$\begin{cases} \lim_{n\to\infty} \mathbb{E}(\varphi(X_n)) = Z_A^{-1} \int \varphi(x) \exp(-V(x) + A(I(x)))\, dx \\ \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \varphi(X_k) = Z_A^{-1} \int \varphi(x) \exp(-V(x) + A(I(x)))\, dx \text{ a.s.} \end{cases}$$

where $Z_A = \int \exp(-V(x) + A(I(x)))\, dx = m$.

# Convergence result (3/3)

Back to the original problem (sampling of $\mu$):

- Importance sampling:

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n}\sum_{j=1}^{m}\theta_{k-1}(j)\varphi(X_k)1_{\{I(X_k)=j\}}}{\sum_{k=1}^{n}\sum_{j=1}^{m}\theta_{k-1}(j)1_{\{I(X_k)=j\}}} = Z^{-1}\int \varphi(x)\exp(-V(x))\,dx \text{ a.s.}$$

$$\lim_{n\to\infty} \frac{m}{n}\sum_{k=1}^{n}\sum_{j=1}^{m}\theta_{k-1}(j)\varphi(X_k)1_{\{I(X_k)=j\}} = Z^{-1}\int \varphi(x)\exp(-V(x))\,dx \text{ a.s.}$$

- Conditioning:

$$\lim_{n\to\infty} \sum_{j=1}^{m}\theta_n(j)\frac{\sum_{k=1}^{n}\varphi(X_k)1_{\{I(X_k)=j\}}}{\sum_{k=1}^{n}1_{\{I(X_k)=j\}}} = Z^{-1}\int \varphi(x)\exp(-V(x))\,dx \text{ a.s.}$$

$$\lim_{n\to\infty} \sum_{j=1}^{m}\theta_n(j)\frac{m}{n}\sum_{k=1}^{n}\varphi(X_k)1_{\{I(X_k)=j\}} = Z^{-1}\int \varphi(x)\exp(-V(x))\,dx \text{ a.s.}$$

# Efficiency

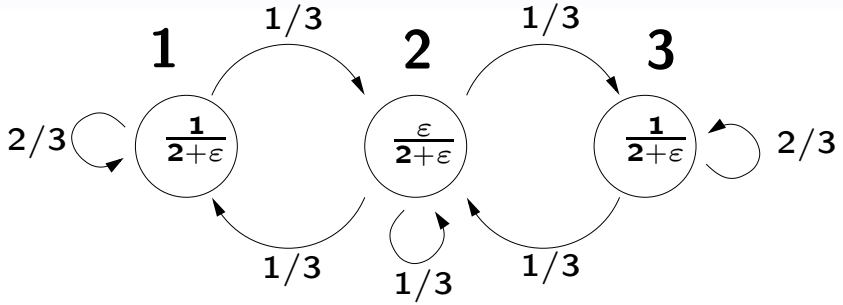
Related works: [Jacob, Ryder]
Convergence is nice, but what about the efficiency of the whole procedure?

Possible approaches:

- Compare the asymptotic variances of the original dynamics with the adaptive dynamics
- Compare the exit times from metastable states of the original dynamics with the adaptive dynamics
- ... ???

Let us look at the average exit time from a metastable state on a toy problem.

# A three state model




Proposal kernel:

$$Q = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Target probability:

$$Z^{-1}(\exp -(V(i)))_{i\in\{1,2,3\}} = \frac{1}{2+\varepsilon}(1,\varepsilon,1).$$

Reaction coordinate: for $i \in \{1,2,3\}$, $I(i) = i$

We look at the average time to go from 1 to 3, in the limit $\varepsilon \to 0$.

# A three state model: the MH dynamics

For the original (non adaptive) Metropolis Hastings dynamics, one can compute

$$\mathbb{E}(T_{1\to 3}^{MH}) \simeq \frac{6}{\varepsilon}$$

More precisely, $\lim_{\varepsilon\to 0} \varepsilon T_{1\to 3}^{MH} = \mathcal{E}(1/6)$ in distribution.

# A three state model: the WL dynamics

For the Wang-Landau dynamics with stepsize sequence $\gamma_n = \gamma_\star n^{-\alpha}$, one has:

- for $\alpha \in (1/2, 1)$,

$$\mathbb{E}(T_{1\to 3}^{WL}) \simeq C_{\alpha,\gamma_\star} \, |\ln(\varepsilon)|^{1/(1-\alpha)}$$

- for $\alpha = 1$,

$$\mathbb{E}(T_{1\to 3}^{WL}) \simeq C\varepsilon^{-1/(1+\gamma_\star)}$$

In any case, in the limit $\varepsilon \to 0$,

$$\mathbb{E}(T_{1\to 3}^{WL}) \ll \mathbb{E}(T_{1\to 3}^{MH}).$$

# Recent developments

We recently analyzed [Fort, Jourdain, TL, Stoltz, 2017-2018] advanced WL-like dynamics (SHUS, well-tempered metadynamics) with:

- Self-tuned stepsize sequence ($\gamma_{n+1}$ is a function of the past);
- Partial biasing (the target is $\exp(-V(x) + aA(I(x)))$ with $a \in (0,1)$).

Practical interests:

- Adapt the stepsize sequence (large $\gamma$ in the exploration phase, small $\gamma$ in the asymptotic regime);
- Control the efficiency factor in the importance sampling step.

# Conclusion

## ABF versus ABP

- ABP can treat discrete reaction coordinates
- ABP can be used for purely entropic barriers
- ABF computes directly the force (which is generally needed in MD)
- ABF acts locally (bias is changed only at the RC value)
- ABF has less numerical parameters to tune

## Trajectorial averages versus Interacting Particle Systems

- IPS $\longrightarrow$ Nonlinear PDE / Trajectorial average $\longrightarrow$ convergence of SAA
- IPS allows for selection mechanisms and easy parallelization
- IPS lead to a better behaviour in a multiple channel case

Adaptive biasing techniques can be used whenever the sampling of a multimodal measure is involved, for example for statistical inference in Bayesian statistics [N. Chopin, TL, G. Stoltz, 2011].

# Recent developments and open problems

Numerical aspects:

- Multiple walker ABF [C. Chipot, TL, K. Minoukadeh]
- Projection on a gradient of the mean force (Helmholtz decomposition) [H. Alrachid, J. Hénin, TL, 2016-2017]
- Reaction coordinates in larger dimension: exchange bias, separated representations [V. Ehrlacher, TL, P. Monmarché, 2019], learning techniques
- What happens for non-gradient force fields?

Theoretical aspects:

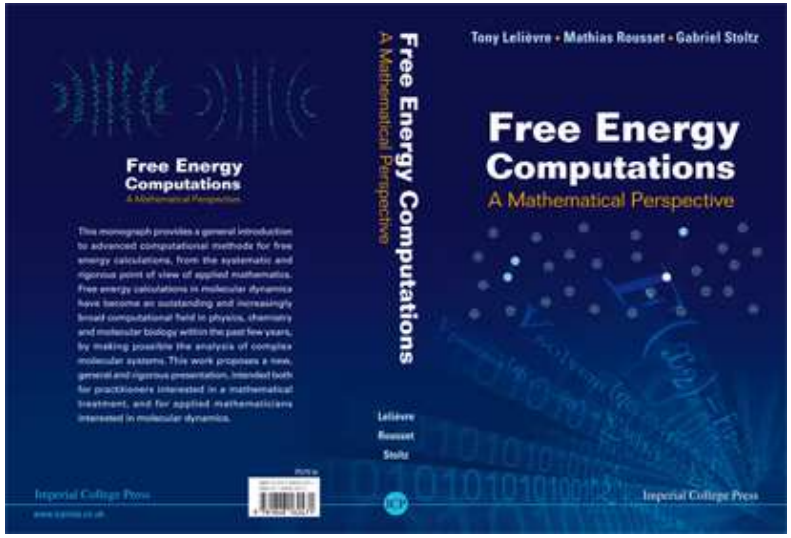
- Analysis when the mean force (or the free energy) is approximated using time averages [G. Fort, B. Jourdain, E. Kuhn, TL, G. Stoltz, P.A. Zitt, 2014-2019]
- Extension of the analysis to the Langevin dynamics?

# References


- TL, M. Rousset and G. Stoltz, *Long-time convergence of an Adaptive Biasing Force method*, Nonlinearity, 21, 2008.
- TL and K. Minoukadeh, *Long-time convergence of an Adaptive Biasing Force method: the bi-channel case*, Archive for Rational Mechanics and Analysis, 202, 2011.
- G. Fort, B. Jourdain, E. Kuhn, TL and G. Stoltz, *Efficiency of the Wang-Landau algorithm: a simple test case*, AMRX, 2, 2014.
- G. Fort, B. Jourdain, E. Kuhn, TL and G. Stoltz, *Convergence of the Wang-Landau algorithm*, Mathematics of Computation, 84, 2015.
- V. Ehrlacher, TL and P. Monmarché, *Adaptive force biasing algorithms: new convergence results and tensor approximations of the bias*, `hal-02314426`, 2019.
- B. Jourdain, TL and P.A. Zitt, *Convergence of metadynamics: discussion of the adiabatic hypothesis*, Annals of Applied Probability, 2021.

# References


- TL, M. Rousset and G. Stoltz, *Free energy computations: A mathematical perspective*, Imperial College Press, 2010.




- TL and Gabriel Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*. Acta Numerica, 25, 2016.