Adaptive Importance Sampling Strategies

TONY LELIEVRE, MATHIAS ROUSSET, GABRIEL STOLTZ

MICMAC project team, INRIA & CERMICS, Ecole des Ponts et Chaussées

http://cermics.enpc.fr/∼stoltz/

The metastability problem

- Applications in computational physics and Bayesian statisitics, when some high dimensional probability measure has to be sampled
- Measure to be sampled $\mu(dq) = Z^{-1}\,\mathrm{e}^{-\beta V(q)}\,dq$ with $Z = \int_{\mathcal{D}} \mathrm{e}^{-\beta V(q)}\,dq$
- For an ergodic dynamics such as $dq_t = -\nabla V(q_t)\,dt + \sqrt{\frac{2}{\beta}}\,dW_t$, ensemble averages can be approximated by trajectorial averages:

$$\langle A\rangle = \int_{\mathcal{D}} A(q)\,\mu(dq) = \lim_{T\to+\infty}\frac{1}{T}\int_0^T A(q_t)\,dt \qquad (1)$$

- Although the convergence (1) is theoretically ensured, it can be very slow from a numerical viewpoint
- Metastability arises from free-energy barriers, which can have either energetic or entropic origins

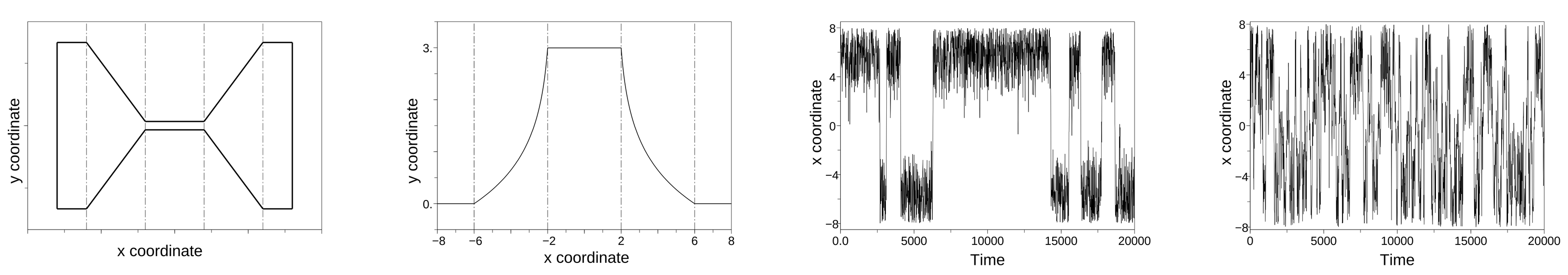
Free-energy biased sampling

- Consider a function $\xi : \mathcal{D}\to\mathbb{R}^m$ ($m \ll \dim(\mathcal{D})$) such that $\xi(q_t)$ is some slowly evolving degree of freedom (a notion to be precised...)
- Marginal equilibrium distribution ($m = 1$ to simplify)

$$\mu^\xi(dz) = Z^{-1}\int_{\Sigma(z)} \mathrm{e}^{-\beta V(q)}\,\delta_{\xi(q)-z}(dq)\,dz = Z^{-1}\int_{\Sigma(z)} \mathrm{e}^{-\beta V(q)}\,\frac{d\sigma_{\Sigma(z)}(dq)}{|\nabla\xi(q)|}\,dz = \mathrm{e}^{-\beta F(z)}\,dz,$$

where $\Sigma(z) = \left\{q\in\mathcal{D}\,\middle|\,\xi(q) = z\right\}$ is a submanifold of $\mathcal{D}$

- Conditional equilibrium distribution $\quad \nu^\xi(dq\,|\,z) = \dfrac{\mathrm{e}^{-\beta V(q)}\,|\nabla\xi(q)|^{-1}\,\sigma_{\Sigma(z)}(dq)}{\displaystyle\int_{\Sigma(z)} \mathrm{e}^{-\beta V}\,|\nabla\xi|^{-1}\,d\sigma_{\Sigma(z)}}$
- Mean force $\quad \nabla F(z) = \displaystyle\int_{\Sigma(z)} f(q)\,\nu^\xi(dq\,|\,z) \quad$ with $\quad f = \dfrac{\nabla\xi\cdot\nabla V}{|\nabla\xi|^2} - \beta^{-1}\mathrm{div}\left(\dfrac{\nabla\xi}{|\nabla\xi|^2}\right)$
- When the potential is biased by the free-energy, $\mathcal{V}(q) = V(q) - F(\xi(q))$, the new marginal distribution is constant = uniform sampling in ξ and the metastability is removed in this direction!
- Application: entropic barrier



(1) Potential for which entropic barriers have to be overcome (0 in the region enclosed by the curve, $+\infty$ outside) and (2) associated free energy profile when $\xi(x, y) = x$. Typical trajectories for a simple Metropolis random walk (3) and a dynamics biased by the free energy (4).

Adaptive methods: A general framework and consistency results [5]

- Bottom line of adaptive methods: add a biasing term, depending on ξ only, and adapt it on-the-fly in order to reach a uniform distribution of $\xi(q_t)$.
- The resulting potential is $\mathcal{V}_t = V - F_t\circ\xi$, and rules to update the bias are needed
- Denote by $\psi(t, q)$ the law of the process $dq_t = -\nabla(V - F_t\circ\xi)(q_t)\,dt + \sqrt{\frac{2}{\beta}}\,dW_t$
- General update formula for Adaptive Biasing Potential method [4, 8] using the *observed free energy*

$$\frac{dF_t(z)}{dt} = \mathcal{F}_t\Big(F_{\mathrm{obs}}(t,z)\Big), \qquad F_{\mathrm{obs}}(t,z) = -\beta^{-1}\ln\left(\int_{\Sigma(z)}\psi(t,\cdot)\,\frac{d\sigma_{\Sigma(z)}}{|\nabla\xi|}\right)$$

- General update formula for Adaptive Biasing Force method [1, 2] using the *observed mean force*

$$\frac{d\Gamma_t(z)}{dt} = \mathcal{G}_t\Big(\Gamma_{\mathrm{obs}}(t,z) - \Gamma_t(z)\Big), \qquad \Gamma_{\mathrm{obs}}(t,z) = \int_{\Sigma(z)} f\,d\psi^\xi(t,\cdot|z)$$

Possibly, set $\Gamma_t(z) = \Gamma_{\mathrm{obs}}(t,z)$

- If some equilibrium is reached and the updating functions $\mathcal{F}_t$ and $\mathcal{G}_t$ are strictly increasing (with $\mathcal{G}_t(0) = 0$), then $F_\infty = F + c$ and $\Gamma_\infty = \nabla F_\infty$
- Issues with the case $m > 1$: ABF is not a gradient dynamics

References

[1] E. DARVE AND A. POHORILLE, Calculating free energies using average force, *J. Chem. Phys.* **115**(20) (2001) 9169–9183

[2] J. HÉNIN AND C. CHIPOT, Overcoming free energy barriers using unconstrained molecular dynamics simulations, *J. Chem. Phys.* **121**(7) (2004) 2904–2914

[3] B. JOURDAIN, T. LELIÈVRE AND R. ROUX, Existence, uniqueness and convergence of a particle approximation for the adaptive biasing force process, In preparation.

[4] A. LAIO AND M. PARINELLO, Escaping free-energy minima, *Proc. Natl. Acad. Sci. U.S.A* **99** (2002) 12562-12566.

[5] T. LELIÈVRE, M. ROUSSET AND G. STOLTZ, Computation of free energy profiles with parallel adaptive dynamics, *J. Chem. Phys.* **126** (2007) 134111

[6] T. LELIÈVRE, M. ROUSSET AND G. STOLTZ, Long-time convergence of an adaptive biasing force method, *Nonlinearity* **21** (2008) 1155-1181

[7] P. RAITERI, A. LAIO, F. L. GERVASIO, C. MICHELETTI, AND M. PARRINELLO, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, *J. Phys. Chem. B* **110**(8) (2006) 3533–3539

[8] F. G. WANG AND D. P. LANDAU, Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram, *Phys. Rev. E* **64**(5) (2001) 056101

Convergence of the Adaptive Biasing Force method [6]

- Dynamics $\quad \begin{cases} dq_t = -\nabla\Big(V - F_t\circ\xi - \beta^{-1}\ln(|\nabla\xi|^{-2})\Big)(q_t)\,|\nabla\xi|^{-2}(q_t)\,dt + \sqrt{2\beta^{-1}}|\nabla\xi|^{-1}(q_t)\,dW_t,\\ F_t'(z) = \mathbb{E}\Big(f(X_t)\,\Big|\,\xi(X_t) = z\Big) = \displaystyle\int_{\Sigma(z)} f(q)\,d\psi^\xi(t,z)\end{cases}$
- Existence and uniqueness of the solution [3]
- Nonlinear PDE on the law of q_t: $\quad \begin{cases} \partial_t\psi = \mathrm{div}\left(\dfrac{\nabla\big(V - F_t\circ\xi\big)\psi + \beta^{-1}\nabla\psi}{|\nabla\xi|^{-2}}\right),\\ F_t'(z) = \dfrac{\displaystyle\int_{\Sigma(z)} f\,|\nabla\xi|^{-1}\psi(t,\cdot)d\sigma_{\Sigma(z)}}{\displaystyle\int_{\Sigma(z)} |\nabla\xi|^{-1}\psi(t,\cdot)d\sigma_{\Sigma(z)}}.\end{cases}$
- Expected longtime limits: $F_t(z)\to F(z)$, $\psi_t(q)\to\psi_\infty(q) = \mathrm{e}^{-\beta(V(q)-F(\xi(q)))}$
- Proof using entropy estimates. The relative entropy of μ with respect to ν is $H(\mu|\nu) = \displaystyle\int \ln\left(\frac{d\mu}{d\nu}\right)d\mu$. Then the total entropy can be decomposed as

$$E(t) = H(\psi(t,\cdot)\,|\,\psi_\infty) = E_M(t) + E_m(t),$$

where the macroscopic and microscopic entropies are respectively

$$E_M(t) = H\big(\psi^\xi(t,\cdot)\,|\,\psi^\xi_\infty\big), \qquad E_m(t) = \int_{\mathcal{M}} e_m(t,z)\,\psi^\xi(t,z)\,dz \qquad e_m(t,z) = H\big(\nu^\xi(t,\cdot\,|\,z)\,|\,\nu^\xi(\infty,\cdot\,|\,z)\big)$$

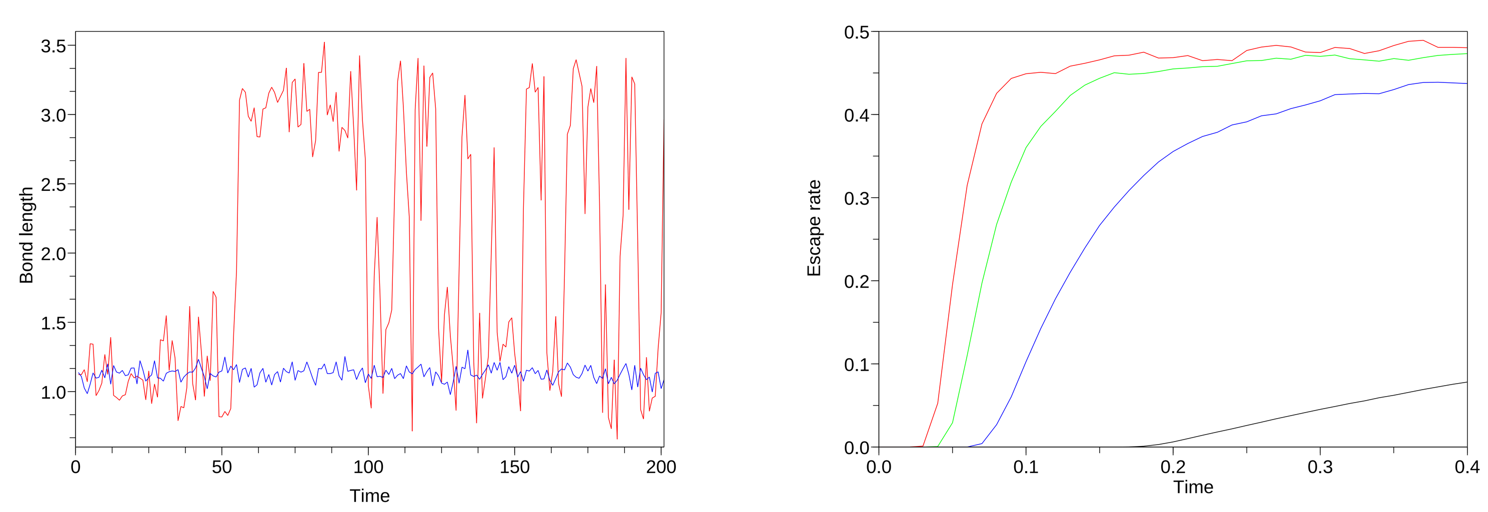
- The marginal density satisfies a simple diffusion equation $\partial_t\psi^\xi = \beta^{-1}\partial_{zz}\psi^\xi$, therefore $E_M\to 0$
- Control of the microscopic entropy when assuming some uniform ergodicity for the dynamics at fixed $\xi(q) = z$ (logarithmic Sobolev inequality)
- The overall rate of convergence of the method is limited by the rate of convergence of the projected dynamics, so that the metastability in the ξ direction is removed

Application to Statistical Physics: Mulitple replica & Selection

- Model dimer in a solvent (double-well potential), solvent particles interacting through the purely repulsive potential (truncated Lennard-Jones):

$$V_{\mathrm{WCA}}(r) = \begin{cases} 4\varepsilon\left[\left(\dfrac{\sigma}{r}\right)^{12} - \left(\dfrac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{if } r\le\sigma,\\ 0 & \text{if } r > \sigma,\end{cases} \qquad V_{\mathrm{dimer}}(r) = h\left[1 - \frac{(r-\sigma-w)^2}{w^2}\right]^2$$

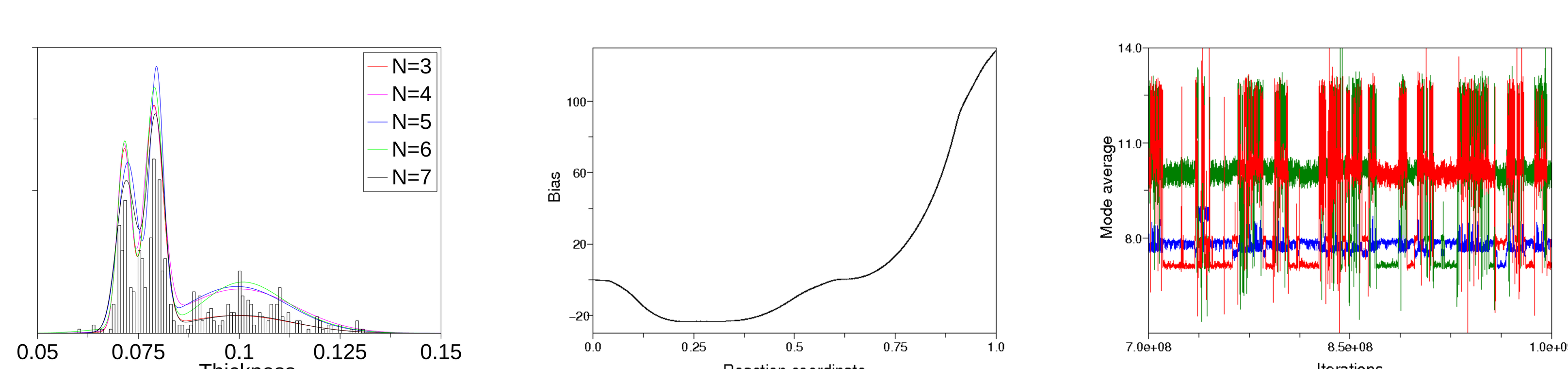
- Two energy minima (compact state $r = r_0 = 2^{1/6}\sigma$, stretched state $r = r_0 + 2w$), energy barrier h
- Reaction coordinate = dimer bond length: $\xi(q) = \dfrac{|q_1 - q_2| - r_0}{2w}$
- Implementation with multiple replicas [7] and selection procedure with the fitness function $S = c\dfrac{\partial_{zz}\psi^\xi_t}{\psi^\xi_t}$



Left: Dynamics with and without bias. Right: Selection procedure with increasing selection strength c.

Application to Bayesian statistics (Monte-Carlo ABF)

- Hidalgo stamp problem: the thickness of $N_{\mathrm{data}} = 485$ stamps are measured, and the corresponding histogram is approximated by a mixture of N Gaussians.
- Parameters $x = (q_1,\dots,q_{N-1},\mu_1,\dots,\mu_N,v_1,\dots,v_N) \in \mathcal{S}_{N-1}\times[\mu_{\min},\mu_{\max}]^N\times[v_{\min},+\infty)\subset\mathbb{R}^{3N-1}$, with $\mathcal{S}_{N-1} = \left\{(q_1,\dots,q_{N-1})\,\middle|\,0\le q_i\le 1,\ q_1+\dots q_{N-1}\le 1\right\}$
- The corresponding mixture is $f(y\,|\,x) = \displaystyle\sum_{i=1}^N q_i\sqrt{\frac{v_i}{2\pi}}\exp\left(-\frac{v_i}{2}(y-\mu_i)^2\right)$, where $q_N = 1 - \sum_{i=1}^{N-1} q_i$
- The likelihood of observing the data $\{y_i, 1\le i\le N_{\mathrm{data}}\}$ is $\quad \Pi(y\,|\,x) = \displaystyle\prod_{d=1}^{N_{\mathrm{data}}} f(y_d\,|\,x) \propto \mathrm{e}^{-\beta V_{\mathrm{likelihood}}}$
- Potential $V = V_{\mathrm{prior}} + V_{\mathrm{likelihood}}$ such that the probability of a given configuration is proportional to $\exp(-V)$
- A simple Metropolis random-walk is metastable
- Use a Monte-Carlo ABF dynamics where $\xi(x) = q_1$ is the reaction coordinate.
- Principle of the method = update the average force experienced in the q_1 direction and obtain the free-energy bias by integrating the approximated mean force



Left: Histogram of the data, and fit with several gaussian modes. Middle: Biasing potential obtained from a Monte-Carlo ABF dynamics. Right: Evolution of the averages μ_1, μ_2 and μ_3 for a biased dynamics