Computing free energy differences using non-equilibrium dynamics

Régis SANTET

(CERMICS, École des Ponts & MATHERIALS Team, Inria Paris)

d^2 reading group, Dept. of Statistics, Oxford

Free energy

• **State function** of a thermodynamic system (internal energy, enthalpy, entropy, etc.)

• **Change in free energy** = maximum amount of **work** the system can perform in a **process** at **constant temperature**

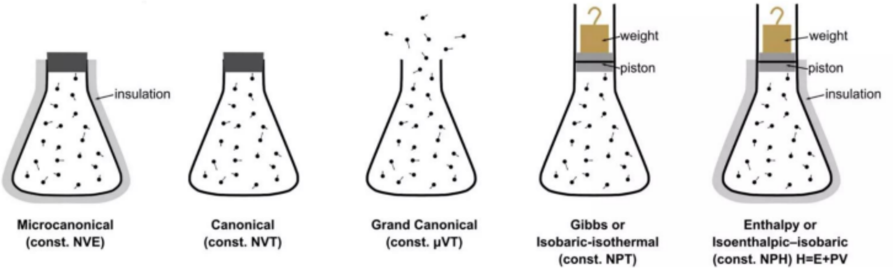
• Helmoltz free energy

$$\boxed{F = U - T_b S}$$

• Applications: gas-phase reactions, energetics of a process: a drug binding a protein or its partioning across cell membranes, ...

NVT ensemble

• F has a minimum at equilibrium as long as certain variables are held constant: **NVT** thermodynamic ensemble



Courtesy of P. Gkeka

• **Canonical measure**

$$\mu(dq\,dp) = Z^{-1}\mathrm{e}^{-\beta H(q,p)}, \quad H(q,p) = V(q) + \frac{1}{2}p^{\mathsf{T}}M^{-1}p, \quad \beta^{-1} = kT_b$$

• Z is the **partition function** [normalizing constant]

Langevin dynamics

- Admits μ as its invariant measure

$$\begin{cases} dq_t = M^{-1} p_t \, dt \\ dp_t = -\nabla V(q_t) \, dt - \gamma M^{-1} p_t \, dt + \sqrt{2\gamma\beta^{-1}} \, dW_t \end{cases}$$

- Overdamped Langevin: reversible w.r.t. $\pi \propto \mathrm{e}^{-\beta V}$

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

- Ergodic averages:

$$\langle \varphi \rangle = \int \varphi \, d\pi = \lim_{T \to +\infty} \frac{1}{T} \int_0^T \varphi(X_t) dt$$

$\rightarrow$ It all comes down to the ability to perform an efficient **sampling** of the configurational space

- In practice: **Markov Chain Monte Carlo** methods (e.g. GHMC, ULA)

Free energy and partition function

• Denote by $\lambda \in [0,1]$ an **external parameter** such that the system is in state A for $\lambda = 0$, and state B for $\lambda = 1$

Example: insertion of a particle

$$V_\lambda(q_1, q_2, q_3) = V\left(\|q_2 - q_1\|\right) + \lambda\left[V\left(\|q_3 - q_1\|\right) + V\left(\|q_3 - q_2\|\right)\right]$$

• One can show that

$$\boxed{F(\lambda) = -\beta^{-1} \ln Z_\lambda = -\beta^{-1} \ln \int \mathrm{e}^{-\beta V_\lambda}}$$

→ The change in free energy is

$$\Delta F = F(1) - F(0) = -\beta^{-1} \ln\left(Z_1/Z_0\right) \iff \mathrm{e}^{-\beta \Delta F} = \frac{Z_1}{Z_0}$$

One needs to compute **a ratio of normalizing constants**

Available methods

• Many methods[1] have been constructed to compute ΔF:
$\rightarrow$ thermodynamic integration, non-equilibrium methods, adaptive methods (ABF, metadynamics), selection mechanisms and parallel replicas, etc.

• I'll present a non-equilibrium method based on Jarzynski's equality[2] and introduce one diffusion models framework[3] based on sequential Monte Carlo samplers[4]

[1] Lelièvre/Rousset/Stoltz (2010)
[2] Jarzynski (1997)
[3] Doucet *et al* (2022)
[4] Del Moral *et al* (2006)

Jarzynski's equality

- Choose an **annealing schedule** $\Lambda : [0,T] \to [0,1]$ that transports π_0 to π_1 using *interpolant* distributions $\pi_{\Lambda(t)} \equiv \pi_{\lambda_t} \propto \mathrm{e}^{-V_{\lambda_t}}$

Example: $\pi_\lambda \propto \pi_0^{(1-\lambda)} \pi_1^\lambda$

- Define the SDE

$$dX_t = -\nabla V_{\lambda_t}(X_t) + \sqrt{2}\, dW_t, \qquad X_0 \sim \pi_0$$

and the path functional

$$\mathcal{W}(\{X_t\}_{0 \leqslant t \leqslant T}) = \int_0^T \dot{\lambda}_t \, \partial_\lambda V_{\lambda_t}(X_t)\, dt$$

- Then it holds

$$\boxed{\mathrm{e}^{-\beta \Delta F} = \left\langle \mathrm{e}^{-\beta \mathcal{W}} \right\rangle}$$

where the average is with respect to $X_0 \sim \pi_0$ and the realizations of the Brownian motion

Numerical example - I

- Setting: transporting one Gaussian to another

$$\begin{cases} \pi_0 = \mathcal{N}(\mu_0, \sigma_0^2), \mu_0 = -2, \sigma_0 = 1 \\ \pi_1 = \mathcal{N}(\mu_1, \sigma_1^2), \mu_1 = 2, \quad \sigma_1 = 0.5 \end{cases}$$

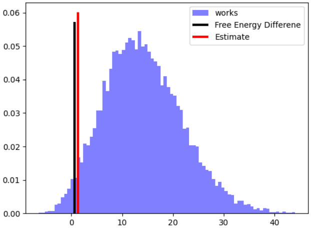
- Time step $\Delta t \sim 10^{-4}$, linear schedule $\Lambda(t) = t/T$, $N_{\text{samples}} \sim 10^5$

- Integrating the dynamics using the Euler–Maruyama scheme:

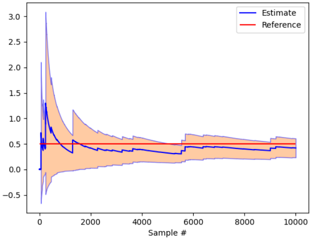
$$x_{k+1} = x_k - \Delta t \nabla V_{\lambda_{k+1}}(x_k) + \sqrt{2\Delta t}\, G_{k+1}, \qquad G_{k+1} \sim \mathcal{N}(0, 1)$$

- Remark: one could use a 'backward' scheduling, depending on which one is more favorable thermodynamically (insertion vs. deletion)

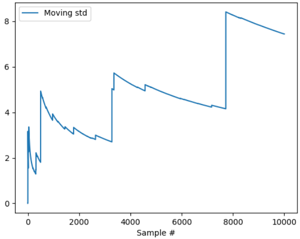
Numerical example - II



Work distribution



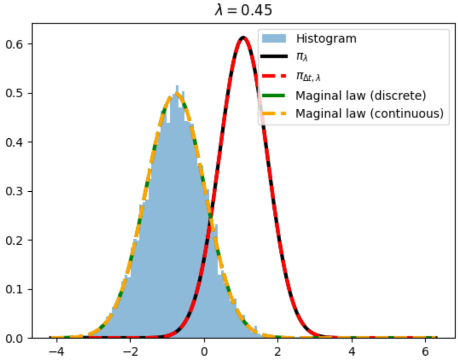
Moving estimate



Moving std of the estimate

Numerical example - III

• The variance of the estimator is intuitively linked to the **variance of the work distribution**



→ The law of the process q_t **'lags behind'**: variance would be minimal if the switching was **infinitely slow**

Escorted Jarzynski

- Construct an **'escorting drift'** $u(x, \lambda)$

$$\boxed{dX_t = -\nabla V_{\lambda_t}(X_t)\, dt + \dot{\Lambda}_t u(X_t, \lambda_t)\, dt + \sqrt{2}\, dW_t}$$

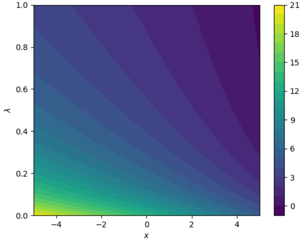
$\rightarrow$ If $q_t^u = \pi_{\lambda_t}$, only 1 sample is needed and you only get discretization errors

- An optimal escorting drift $u^\star$ solves

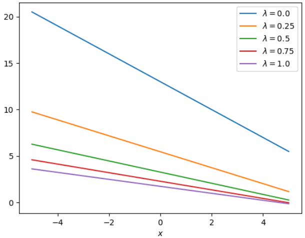
$$\partial_\lambda \pi_\lambda + \nabla \cdot (u^\star \pi_\lambda) = 0$$

$\rightarrow$ Non-uniqueness, impossible to construct exactly in general

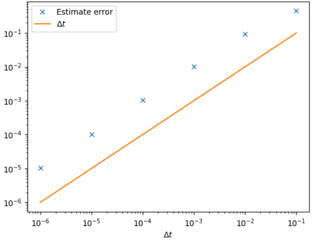
Numerical example - IV



$u^\star$



$u(\cdot, \lambda)$



Error as a function of Δt

Using diffusion models

• Computing normalization constants have been investigated a lot in the ML community recently, in particular with the rise of diffusion models and its link with optimal transport

• Setting based on a paper by Doucet *et al* (2022) & Geffner/Domke (2023)

Constructing the estimate

- Set $\pi_i = Z_i^{-1}\gamma_i$ with $\gamma_i = \mathrm{e}^{-V_i}$, $i \in \{0,1\}$

- Annealing schedule $\pi_\gamma \propto \mathrm{e}^{-V_\lambda}$, forward transition kernel

$$F_{k+1}(x_{k+1}|x_k) = \mathcal{N}\left(x_{k+1}; x_k - \Delta t \nabla V_{\lambda_{k+1}}(x_k), 2\Delta t\right)$$

- Choose **any** backward transition kernel B_k, i.e. it only has to satisfy $\int B_k(x|x')dx = 1$ for any x'

- Then $\mathrm{e}^{-\Delta F} = \left\langle \mathrm{e}^{-\mathcal{W}} \right\rangle$ with

$$\boxed{\mathrm{e}^{-\mathcal{W}} = \frac{\gamma_1(x_N)}{\gamma_0(x_0)} \prod_{k=0}^{N-1} \frac{B_k(x_k|x_{k+1})}{F_{k+1}(x_{k+1}|x_k)}}$$

Two links with Jarzynski's methods - I

- Annealed Importance Sampling: choosing

$$B_k(x_k|x_{k+1}) = \pi_{\lambda_{k+1}}(x_k)\frac{F_{k+1}(x_{k+1}|x_k)}{\pi_{\lambda_{k+1}}(x_{k+1})}$$

leads to

$$\mathcal{W} = \sum_{k=0}^{N-1}(V_{\lambda_{k+1}} - V_{\lambda_k})(x_k) \approx \int_0^T \dot{\lambda}_t \partial_\lambda V_{\lambda_t}(x_t)dt$$

$\rightarrow$ We recover the usual work in **Jarzynski's equality**

Link with Jarzynski's methods - II

- The optimal backward kernel (minimizing the variance of the estimator) is

$$B_k^{\text{opt}}(x_k|x_{k+1}) = \frac{q_{\lambda_k}(x_k)F_{k+1}(x_{k+1}|x_k)}{q_{\lambda_{k+1}}(x_{k+1})}$$

yielding the estimator $\frac{\gamma_1(x_N)}{Z_0 q_N(x_N)}$

- If instead

$$F_{k+1}^{u^\star}(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}; x_k - \Delta t \nabla V_{\lambda_{k+1}}(x_k) + \Delta t \dot{\lambda}_{k+1} u^\star(x_k, \lambda_{k+1}), 2\Delta t),$$

then $q_{\lambda_k} = \pi_{\lambda_k}$ for any k so that

$$B_k^{\text{opt},u^\star}(x_k|x_{k+1}) = \frac{\pi_{\lambda_k}(x_k)F_{k+1}^{u^\star}(x_{k+1}|x_k)}{\pi_{\lambda_{k+1}}(x_{k+1})},$$

and $\mathrm{e}^{-\mathcal{W}}$ **does not depend on the trajectory**
$\rightarrow$ we recover the 0 variance estimator for the optimal escorting drift

Learning the score

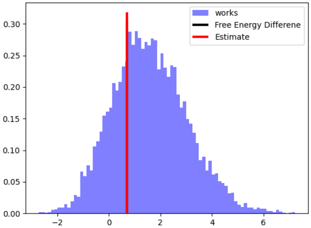
• In practice, one approximates the optimal backward kernel, which lead to a transition kernel related to the structure of the **reversed SDE**

$$dY_t = \nabla V_{\lambda_{T-t}}(Y_t)\,dt + 2\nabla \log q_{T-t}(Y_t)\,dt + \sqrt{2}\,dW_t, \qquad Y_T \sim q_T$$

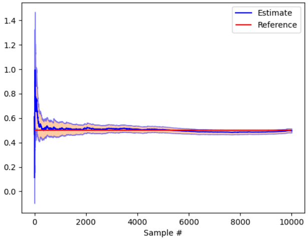
• One then approximates the score term using a neural network $s_\theta(T-t, y)$, minimizing the KL divergence between the forward and (parameterized) backward path distributions

$$\begin{aligned} D_{\mathrm{KL}}(Q||P_\theta) &= \mathbb{E}_Q\left[\int_0^T \|s_\theta(t, x_t) - \nabla \log q_t(x_t)\|^2\,dt\right] + C_1 \\ &\approx \Delta t \sum_{k=1}^{K} \mathbb{E}_Q\left[\|s_\theta(t_k, x_k) - \nabla \log F_k(x_k|x_{k-1})\|^2\right] + C_2 \end{aligned}$$

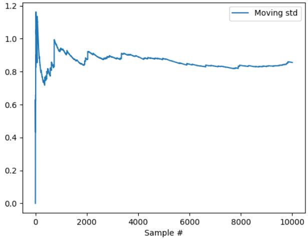
Numerical example - V



Work distribution



Moving estimate



Moving std of the estimate

Investigations

- Learning s_θ to compute ΔF in the overdamped/underdamped setting (which architecture ?)

- Study the connection with **Schrödinger bridge**[5]: sequence of estimators

- Adapt to the **reaction coordinate** framework

[5] Leonard (2014), Vargas/Nusken (2023)

Reaction coordinate framework - I

• **Reaction coordinate**: $\xi : \mathbb{R}^d \to \mathbb{R}$, system constrained to the submanifold

$$\Sigma(\lambda) = \left\{ x \in \mathbb{R}^d \,\middle|\, \xi(x) = \lambda \right\}$$

$\to$ It is assumed that $|\nabla \xi|$ is nonzero at the vicinity of $\Sigma(\lambda)$

Example: dihedral angles, distances between two molecular groups

• Free energy is

$$F(\lambda) = -\beta^{-1} \ln \left(\int_{\Sigma(\lambda)} \pi^{\xi}(dx|\lambda) \right)$$

$\to$ $\pi^{\xi}(\cdot|\lambda)$ is the measure π **conditioned to a fixed value of λ** of the map ξ

• This presentation adapts to the case $\xi : \mathbb{R}^d \to \mathbb{R}^m$ with $m \geqslant 1$

Reaction coordinate framework - II

- **Switched dynamics**: schedule $\Lambda(0) = \lambda_0, \Lambda(T) = \lambda_T$.

$$\begin{cases} dX_t = -\nabla V^\xi(X_t)\,dt + \sqrt{2\beta^{-1}}\,dW_t + \nabla\xi(X_t)\,d\theta_t, & q_0 \sim \pi^\xi(\cdot|0), \\ \xi(X_t) = \lambda_t \end{cases}$$

$\rightarrow V^\xi = V + \beta^{-1}\ln|\nabla\xi|$, $(\theta_t)_{t\in[0,T]}$ are Lagrange multipliers (with available expressions)

- Work is defined as the integral of the local mean force f

$$\mathcal{W}(\{X_t\}_{0\leqslant t\leqslant T}) = \int_0^T \dot{\Lambda}(s) f(X_s) ds$$

with

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

$\rightarrow$ Can we adapt diffusion models estimates to this framework ?

Free energy and partition function

- Internal energy is $U = \langle H \rangle = Z^{-1} \int H \, \mathrm{e}^{-\beta H} = -\partial_\beta \ln Z$
- A change of an external variable λ applies a force equal to

$$F = -\partial_\lambda H, \qquad \langle F \rangle = \beta^{-1} \partial_\lambda \ln Z$$

- If both β and λ vary, then [chain rule]

$$d(\ln Z) = \partial_\beta \ln Z d\beta + \partial_\lambda \ln Z d\lambda = -U d\beta + \beta F d\lambda = -d(\beta U) + \beta dU + \beta F d\lambda$$

The change in internal energy is

$$T dS - F d\lambda = dU = \beta^{-1} d\left(\ln Z + \beta U\right) - F d\lambda$$

Hence

$$S = k \ln Z + U/T$$

so that

$$\boxed{F = -\beta^{-1} \ln Z}$$