Advanced sampling techniques for numerical simulations: lecture notes

Alessandro Laio

April 22, 2009

1 Introduction

Consider a system of coordinates $x \in \Omega$. The system evolves according to a dynamics allowing an equilibrium distribution $P(x) = \frac{1}{Z}\exp(-\beta V(x))$ where β is the inverse temperature and $Z = \int_\Omega dx \exp(-\beta V(x))$. Qualitatively, the system will display metastability if the probability is large in a set of disconnected regions A_i separated by regions in which the probability is low. Hence, we must have

$$\sum_i \int_{A_i} dx P(x) \approx 1$$

In addition one must also require that the regions A_i are well separated, namely that transitions between these regions are rare. We will come back to this second condition in detail later.

1.1 Reduced probability distributions and the free energy

In order to define more quantitatively these regions and their properties, it is useful to consider *reduced* probability distributions. Namely, instead of monitoring the full trajectory $x(t)$ we only monitor the evolution of a set of collective variables or reaction coordinates $s(x)$, and look at their trajectory $s(t) = s(x(t))$. Their probability distribution $P(s)$, for an infinitely long trajectory is given by the histogram of s :

$$P(s) = \lim_{t\to\infty} \frac{1}{t}\int_0^t dt \delta(s - s(t))$$

In real applications, $P(s)$ can be estimated as

$$P(s) \sim \frac{1}{n\Delta s}\sum_{t=1}^{n} \chi_s(s(t))$$

where $\chi_s(x) = 1$ if $x \in [s, s+\Delta s]$ and zero otherwise.

If the system is ergodic and the dynamics allows an equilibrium distribution at an inverse temperature β, the knowledge of $P(s)$ allows defining the free energy$F(s)$:

$$F(s) = -\frac{1}{\beta}\log\left(P(s)\right)$$

$P(s)$ (or $F(s)$) provide a first indication of the metastability properties: in metastable systems s resides for the big majority of time in disconnected regions and $F(s)$ has a characteristic shape with wells and barriers. Unfortunately, as it will be clear in the following, a free energy profile gives useful indications on the metastability properties of the system only if the variable s is properly chosen.

1.2 Reactive flux model and the rate constant

A very simple example of Markov process is the reactive flux model that describes the evolution of the probability to observe the system in a set of states i:

$$\frac{dp_i}{dt} = \sum_{j=1}^{n}\left(-k_{i\to j}p_i + k_{j\to i}p_j\right), i = 1,\ldots,n \tag{1}$$

where n is the number of states. $k_{i\to j}$ is the probability to perform a transition between state i and state j in a unit of time and is called *rate constant*. This model is commonly used for studying chemical reaction, etc.

For two wells (A and B), we have

$$\frac{dp_A}{dt} = -k_{A\to B}p_A + k_{B\to A}p_B \tag{2}$$

$$\frac{dp_B}{dt} = k_{A\to B}p_A - k_{B\to A}p_B \tag{3}$$

The stationary ptobability is obtained imposing $\frac{dp_A}{dt} = 0$ and is given by$p_A \doteq p_A^{\mathrm{eq}} = \frac{k_{B\to A}}{k_{A\to B}+k_{B\to A}}$ and $p_B \doteq p_B^{\mathrm{eq}} = \frac{k_{A\to B}}{k_{A\to B}+k_{B\to A}}$. Thus, Eq. 2 can be written as

$$\frac{dp_A}{dt} = \left(p_A^{eq} - p_A\right)\left(k_{A\to B} + k_{B\to A}\right) \tag{4}$$

With initial conditions $p_A(0) = 1$ and $p_B(0) = 0$ we have

$$p_A(t) = 1 + \frac{k_{A\to B}}{k_{A\to B}+k_{B\to A}}\left(e^{-(k_{A\to B}+k_{B\to A})t} - 1\right) \tag{5}$$

$$p_B(t) = 1 - p_A(t) \tag{6}$$

2 Methods for computing the free energy

We have seen that the equilibrium behavior of a set of CVs is completely defined by the probability distribution:

$$P(s) = \frac{\exp(-\beta F(s))}{\int ds \exp(-\beta F(s))} \tag{7}$$

where by s we denote the d dimensional vector $(s_1, \ldots, s_d)$ and the free energy $F(s)$ is given by

$$F(s) = -\frac{1}{\beta} \ln \left(\frac{1}{t} \int dt \delta (s - s(x(t))) \right) = \tag{8}$$

$$= -\frac{1}{\beta} \ln \left(\int dx \exp(-\beta V(x)) \delta (s - s(x)) \right). \tag{9}$$

The free energy as a function of a *relevant* and *smartly chosen* variable provides a very precious insight in the equilibrium and metastability properties of the system. For instance, the minima in a free energy surface correspond approximately to the metastable sets of a system: the system spends by definition a lot of time in the minima and only rarely it visits the barrier regions in between and has a chance to perform a transition.

The free energy profiles can also be also used to estimate the transition time between two metastable sets. Transition state theory (TST) states that the average transition time from a states A to a state B is proportional to the free energy difference between the minimum and the barrier:

$$t_{A \to B} \propto \exp(\beta (F(s_{\max}) - F(s_A))),$$

where $F(s_{\max})$ is the maximum of $F(s)$ between s_A and s_B. . Beside other important limitations of TST that we will discuss later, we here notice that this equation requires evaluating $F(s)$ at the extrema (maximum and minimum). Even if $s(x)$ is a good reaction coordinate for the system, the free energy differences are pretty ill-defined. Consider in fact another variable $\sigma = \sigma(s)$ which is a invertable function of s. These change of variables are of course allowed and commonly used: for instance, one can describe a bond breaking in a chemical reaction using as CV the distance between two atoms or their coordination number defined as a fermi function of the same distance. The free energy profile as a function of σ and s will NOT look the same. In fact, by the invariance properties of a probability distribution, we must have $P(\sigma) d\sigma = P(s) ds$ and thus

$$F(\sigma) = F(s) + \frac{1}{\beta} \log \left(\frac{d\sigma}{ds} \right)$$

Exercise: deduce this property directly from the definition of $F(s)$8.

2.1 Umbrella sampling

Consider a system of coordinates x and potential $V(x)$, whose dynamics admits a canonical distribution at an inverse temperature β. For this system we want to compute the free energy 8 as a function of a collective variable $s(x)$. In umbrella sampling the normal dynamics of the system is biased by a suitably chosen bias potential $V^B(s(x))$ that depends on x only via $s(x)$.

Consider for example a langevin evolution of the form

$$dx = -D\beta\partial_x \left(V(x) + V^B(s(x))\right) dt + \sqrt{2D}dW$$

The probability distribution as a function of s for the system evolved with $V(x) + V^B(s(x))$ will be

$$\begin{aligned} P^B(s) &= \frac{1}{Z^B} \int dx \exp\left(-\beta \left(V(x) + V^B(s(x))\right)\right) \delta(s - s(x)) = \\ &= \frac{Z}{Z^B} \exp\left(-\beta V^B(s)\right) \frac{1}{Z} \int dx \exp\left(-\beta V(x)\right) \delta(s - s(x)) = \\ &= \frac{Z}{Z^B} \exp\left(-\beta \left(V^B(s) + F(s)\right)\right) \end{aligned} \tag{10}$$

where Z^B is the canonical partition function for the potential $V(x)+V^B(s(x))$. This equation tells that measuring a probability distribution in the presence of a bias $V^B(s(x))$ provides a measure for the unbiased free energy (and for the unbiased probability distribution). In fact we have

$$F(s) = -\frac{1}{\beta} \log\left(P^B(s)\right) - V^B(s) - f^B \tag{11}$$

where

$$f^B = \frac{1}{\beta} \log \frac{Z}{Z^B}$$

is a constant that does not depend on s. Moreover,

$$P(s) = P^B(s) \exp\left(\beta \left(V^B(s) - f^B\right)\right) \tag{12}$$

This method for estimating $F(s)$ is called *umbrella sampling.*

A question that immediatly arises in this approach is how to choose the exact form for $V^B(s)$. It is clear that the statistical uncertainty in Eq. 11 derives from errors in the evaluation of $P^B(s)$. Assume we estimate $P^B(s)$ from a time series $s_1, s_s, \ldots, s_n$ of n values that the collective variable takes during the biased simulation. Imagine we observe i entries in the interval $[s, s + \Delta s]$. $P^B(s)$ can be estimated as

$$P^B(s)\,\Delta s \sim \frac{i}{n} = \frac{1}{n} \sum_{t=1}^{n} \chi_s(s_t) \tag{13}$$

where $\chi_s(x) = 1$ if $x \in [s, s + \Delta s]$ and zero otherwise. This equation can be easily implemented in a computed program: in practice estimating $P^B(s)$

requires computing the normalized histogram of s. At finite simulation time, the value of the probability $P^B(s)\Delta s$ is estimated by i/n. The statistical error on $\frac{i}{n}$ can be easiliy estimated if the entries are assumed to be totally uncorrelated. In fact, the the probability to observe i out of n entries in an interval whose probability is $p \doteq P^B(s)\Delta s$ is given by the Poisson distribution:

$$\mathcal{P}_{p,n}(i) = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

For this distribution, the average value of $\frac{i}{n}$ is given by p: $\left\langle \frac{i}{n} \right\rangle = \sum_i \mathcal{P}_{p,n}(i) \frac{i}{n} = p$. The expected error in $\frac{i}{n}$ is

$$\sigma^2\left(\frac{i}{n}\right) = \left\langle \left(\frac{i}{n} - \left\langle \frac{i}{n} \right\rangle\right)^2 \right\rangle = \sum_i \mathcal{P}_{p,n}(i) \left(\frac{i}{n} - p\right)^2 = \frac{p}{n}.$$

Hence,

$$\sigma^2\left(P^B(s)\right)\Delta s^2 = \sigma^2\left(\frac{i}{n}\right) = \frac{p}{n} = \frac{\Delta s P^B(s)}{n}. \tag{14}$$

This formula requires a few comments.

1. In a real case, $P^B(s)$ is not known, but is estimated using Eq, 13. Hence, the error is actually estimated as

   $$\sigma^2\left(P^B(s)\right)\Delta s^2 \sim \frac{i}{n^2}$$

   For long simulations, i becomes large, but also n grows. As $\frac{i}{n}$ is approximately constant, the quadratic error decays as $1/n$

2. In real applications, if the time series $s_1, s_2, \ldots, s_n$ on which the histogram is computed is not uncorrelated, the error will be larger than the one predicted by Eq. 14 by a factor $\left(1 + 2\frac{\tau(s)}{dt}\right)$ where dt is the time interval between two different s and $\tau(s)$ is the correlation time of $\chi_s(s_t)$ [J. Stat. Phys. 8, 1, (1973)]. This correlation time can depend non trivially on s and is in general hard to compute. In practice, instead of computing $\tau(s)$ one can decimate the time series, i.e. one makes dt large discarding the intermediate s_t. Of course, this procedure reduces the total number of entries n and makes the error larger, but at least avoids systematic errors.

Let's now compute the error on the free energy. From Eq. 11

$$\sigma^2(F(s)) = \frac{1}{\beta^2} \frac{\sigma^2\left(P^B(s)\right)}{\left(P^B(s)\right)^2}.$$

Using 14

$$\sigma^2(F(s)) = \frac{1}{\beta^2} \frac{1}{\Delta s P^B(s)}.$$

If one requires the error on the free energy to be constant one has to choose the bias potential $V^B(s)$ in such a way that $P^B(s)$ is constant. This implies (see equation 10)

$$V^B(s) = -F(s) \tag{15}$$

Hence, the optimal choice for the biasing potential is minus the free energy. It should be noticed that in real applications $F(s)$ is not known (indeed, this is what we are trying to estimate). Hence, condition 15 can be imposed only approximately or by a suitable iterative procedure.

2.2 Weigthted histogram methods

The main problem with umbrella sampling is that it is very difficult to construct $V^B(s)$ without a detailed knowledge of the system. In order to solve this problem, an efficient strategy is the *weigthed histogram method*, initially introduced by Ferremberg and Swendsen[7], then extended by Kapral and Roux[9, 11]. The idea is to combine several histograms constructed with different umbrellas $V^{B_i}(s)$, $i = 1, 2, \ldots$ in order to reconstruct a single estimate of $F(s)$. A typical choice for the bias potentials is

$$V^{B_i}(s) = \frac{1}{2}k(s - s_i)^2$$

in which the s_i-s are usually disposed in a regular grid in order to cover all the interesting values of s. During run i the statistics is collected in the neighborhood of s_i.

From equation 12, we have that, for all i-s,

$$P^i(s) = P^{B_i}(s) \exp\left(\beta\left(V^{B_i}(s) - f^i\right)\right) \tag{16}$$

is an estimate of the unbiased probability distribution. The best estimate of the probability distribution $P(s)$ is assumed to be a linear combination of the P_i-s, with weigths π_i that depend on the values of s :

$$P(s) = C\sum_i \pi_i(s) P^i(s) \tag{17}$$

The weigths are assumed to satisfy the normalization condition

$$\sum_i \pi_i(s) = 1 \tag{18}$$

and C is a constant. The weigths $\pi_i(s)$ are determined minimizing the expected error on $P(s)$

$$\sigma^2(P(s)) = C^2\sum_i \pi_i^2(s)\,\sigma^2\left(P^i(s)\right)$$

under the constraint 18. Thus we require

$$\frac{\delta}{\delta\pi_i(s)}\left(\sigma^2(P(s)) - \lambda\left(\sum_i \pi_i(s) - 1\right)\right) = 2C^2\pi_i(s)\,\sigma^2\left(P^i(s)\right) - \lambda = 0$$

where λ is a Lagrange multiplier. Imposing the constraint 18 on the solution of this equation, we have

$$\pi_i(s) = \frac{\left(\sigma^2\left(P^i(s)\right)\right)^{-1}}{\sum_j \left(\sigma^2\left(P^j(s)\right)\right)^{-1}}$$

From Eq. 16 and 14, we have

$$\begin{aligned}\sigma^2\left(P^i(s)\right) &= \sigma^2\left(P^{B_i}(s)\right)\exp\left(2\beta\left(V^{B_i}(s) - f^i\right)\right) = \\ &= \frac{P^{B_i}(s)}{n_i \Delta s}\exp\left(2\beta\left(V^{B_i}(s) - f^i\right)\right) = \\ &= \frac{P(s)}{n_i \Delta s}\exp\left(\beta\left(V^{B_i}(s) - f^i\right)\right) \end{aligned} \tag{19}$$

and thus

$$\pi_i(s) = \frac{n_i \exp\left(-\beta\left(V^{B_i}(s) - f^i\right)\right)}{\sum_j n_j \exp\left(-\beta\left(V^{B_j}(s) - f^j\right)\right)} \tag{20}$$

The final WHAM estimate of the probability is obtained substituting this expression of the weights into Eq. 17, using Eq. 16:

$$P(s) = C\frac{\sum_k n_k P^{B_k}(s)}{\sum_j n_j \exp\left(-\beta\left(V^{B_j}(s) - f^j\right)\right)}$$

The constants f^i entering in this solution are determined self consistently from their definition:

$$\begin{aligned} e^{-\beta f^i} = \frac{Z^i}{Z} &= \frac{\int ds \exp\left(-\beta\left(V + V^{B_i}\right)\right)}{\int ds \exp\left(-\beta\left(V\right)\right)} = \\ &= \int ds \exp\left(-\beta V^{B_i}(s)\right) P(s) = \\ &= C\int ds \exp\left(-\beta V^{B_i}(s)\right)\frac{\sum_k n_k P^{B_k}(s)}{\sum_j n_j \exp\left(-\beta\left(V^{B_j}(s) - f^j\right)\right)} \end{aligned}$$

Remarks:

1. The procedure is based on the minimization of the estimated error on $P(s)$. If the quantity of interest is $F(s)$, it migth be a better idea to control directly the error on it. If the WHAM procedure is used, the error of F is given by

$$\begin{aligned}\sigma^2\left(F(s)\right) = \frac{\sigma^2\left(P(s)\right)}{P^2(s)} &= \frac{\sum_i \pi_i^2(s)\,\sigma^2\left(P^i(s)\right)}{\left(\sum_i \pi_i(s) P^i(s)\right)^2} = \\ &= \frac{1}{\sum_i n_i P^{B_i}(s)}\end{aligned}$$

where we used equations 14 and 20 to estimate π_i and $\sigma^2\left(P^i(s)\right)$. This shows that the error on $F(s)$ is uniform if the cumulative histogram computed using all the biased trajectory is uniform. Remarkably, the error is complitely independent on the actual form of the biasing potentials and only depends on the biased histogram. This property allows optimizing the choice of the distribution and number of biasing potentials: if a region has not been explored frequently enough, one should add windows in that region.

2. See point 3 of Umbrella Sampling: if one does not know the correlation time of the datas, it is always much safer to decimate them.

3. The number of biasing potentials that one has to use scales exponentially with the dimensionality. The computational price becomes unberable in $d > 2$.

2.3 Thermodynamic integration

Theermodynamic integration is probably the most commonly used methodology for estimating $F(s)$ in complex systems. It was introduced in refs [3, 12] and is based on the fact that the derivative of $F(s)$ with respect to s can be expressed as an equilibrium average over an equilibrium ensamble constrained on the hypersurface $s(x) = s$. In fact, we have

$$\begin{aligned}
F'(s) &= -\frac{1}{\beta}\frac{\int dx \exp\left(-\beta V(x)\right)\delta'\left(s-s(x)\right)}{\int dx \exp\left(-\beta V(x)\right)\delta\left(s-s(x)\right)} = \\
&= \frac{1}{\beta}\frac{\int dx \exp\left(-\beta V(x)\right)\frac{\nabla s(x)}{\|\nabla s(x)\|^2}\cdot\nabla\delta\left(s-s(x)\right)}{\int dx \exp\left(-\beta V(x)\right)\delta\left(s-s(x)\right)} = \\
&= -\frac{1}{\beta}\frac{\int dx\delta\left(s-s(x)\right)\exp\left(-\beta V(x)\right)\left(-\beta\frac{\nabla s(x)\cdot\nabla V(x)}{\|\nabla s(x)\|^2}+\nabla\cdot\left(\frac{\nabla s(x)}{\|\nabla s(x)\|^2}\right)\right)}{\int dx \exp\left(-\beta V(x)\right)\delta\left(s-s(x)\right)} = \\
&= \left\langle \frac{\nabla s(x)\cdot\nabla V(x)}{\|\nabla s(x)\|^2} - \frac{1}{\beta}\nabla\cdot\left(\frac{\nabla s(x)}{\|\nabla s(x)\|^2}\right)\right\rangle_{s=s(x)} \qquad (21)
\end{aligned}$$

$F'(s)$ is usually referred to as the *mean force*. Its value can be used to evaluate $F(s)$ as an integral: for instance,

$$F(s) - F(0) = \int_0^s dx F'(x). \qquad (22)$$

This integral is in practice computed numerically evaluating $F'(s)$ on a sufficiently narrow partition of s.

Remarks:

1. The procedure is subject to error propagation, since $F(s)$ is estimated as a numerical integral. The choice of the optimal partition for evaluating

the integral depends in general on the smoothness properties if $F(s)$. The limiting factor is the second derivative of $F(s)$, namely how fast $F'(s)$ is varying.

2. Like in umbrella sampling, the error of each mean force average Eq. 21 is determined by the number of independent estimate that are actually used, i.e. by the correlation time of the argument of the average. This number cannot expected to be independent on s, except in very special cases. Hence, investing the same computational time for all the mean forces $F'(s)$ is NOT in general the optimal choice. Exactly like in umbrella sampling, one can always choose to decimate the data in order to keep the error under control. But also this choice is usually far from optimal. A very common procedure is to renounce to an explicit error control and evaluate the averages using all the available data. The integral is evaluated in the forward direction as in Eq. 22, and, independently, in the bacward direction. The two free energy profiles obtained in this way are compared and their difference provides a rough estimate of the accuracy.

3. Evaluating the average 21 requires running a molecular dynamics or a montecarlo under the constraint $s(x) = s$. This can be non-trivial for complex functions $s(x)$. For molecular dynamics and Langevin dynamics the problem has been solved in its full generality (see ref. [12]).

Exercise: Consider a 2-dimensional system of potential

$$V(x_1, x_2) = -3 \log \left(3e^{-2x_1^2 - \left(\frac{x_2}{2}\right)^2} + 2e^{-\left(x_1 - \frac{3}{2}\right)^2 - \left(x_2 - \frac{5}{2}\right)^2} \right)$$

evolving according to a Langevin dynamics at a temperature T=1 and with a diffusion coefficient $D = 0.0001$.

1. For $s = x_1$ compute analytically the free energy $F(x_1)$. Repeat the calculation for $s = x_2$ and comare the height of the barriers for the original potential, $F(x_1)$and $F(x_2)$

2. For $s = x_1$write explicitly the expression of the mean force Eq.21.

3. Write a code for simulating the evolution of the system at fixed x_1.

4. Using this code, compute the free energy $F(x_1)$ numerically using thermodynamic integration.

2.4 Metadynamics

The metadynamics method[10, 8, 2], like umbrella sampling, WHAM, and thermodynamic integration, is based on a dimensional reduction, and requires the preliminary identification of a set of collective variables (CVs) s, which are function of the system coordinates, x, and are able to describe the activated process of interest.

In metadynamics the free energy is reconstructed sequentially, beginning from the bottom of the well by a history-dependent random walk that explores a larger and larger portion of configuration space. In the simplest molecular dynamics implementation of the algorithm a small repulsive Gaussian potential is added every τ_G MD steps. Thus, the external ("metadynamics") potential acting on the system at time t is given by

$$V_G(x,t) = w \sum_{\substack{t' = \tau_G, 2\tau_G, \cdots \\ t' < t}} \exp\left(-\frac{\left(s\left(x\right) - s\left(t'\right)\right)^2}{2\delta s^2} \right). \tag{23}$$

where $s\left(t\right) = s\left(x\left(t\right)\right)$ is the value taken by the collective variable at time t. Three parameters enter the definition of the V_G:

1. The gaussian height w
2. The gaussian width δs
3. the frequency τ_G at which the gaussians are deposed.

These parameters influence the accuracy and efficiency of the free energy reconstruction. Qualitatively, they define the size of the "barrows of sand" the walker is throwing. If the Gaussians are large, the free energy surface will be covered fastly, but the reconstructed profile will be affected by large errors. Instead, if the Gaussians are small or are placed infrequently the reconstruction will be very accurate, but it will take a much longer time.

For example, consider the system depicted in Figure 1. This system is evolved with a Langevin overdamped dynamics at a temperature T=1 with a diffusion coefficient D=0.0001 on a one-dimensional potential with three minima. If the dynamics starts from the central local minimum, this is filled by the Gaussians in ~ 20 steps. After that the dynamics escapes over the well from the lowest saddle point, filling the second well in ~ 80 steps. The second highest saddle point is reached in ~ 160 steps, and the full free energy surface is filled in a total of ~ 320 steps. This example also provides a demonstration of the two different manners in which metadynamics can be exploited :

- If the method is used for "escaping free energy minima", i.e. for finding the lowest free energy saddle point out of a local minimum, the metadynamics should be stopped as soon as the walker starts exploring a completely new region of space. For the example of Figure 1, this happens after ~ 20 steps. .
- If the aim is to estimate the free energy in a predefined region in the CV space, metadynamics should be stopped when the walker has explored this whole region and its motion becomes diffusive. For the example of Figure 1, the full free energy profile is filled after ~ 320 steps.

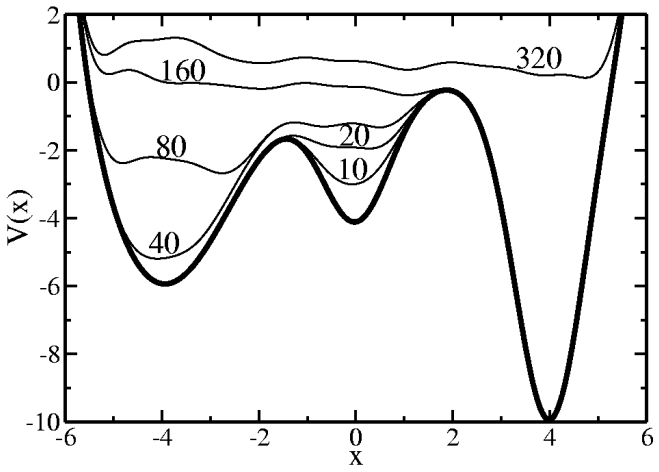

Figure 1: Time evolution of the sum of a one-dimensional model potential V(σ) and the accumulating Gaussian terms of Eq. (**??**). The dynamic evolution (thin lines) is labelled by the number of dynamical iterations (37). The starting potential (thick line) has three minima and the dynamics is initiated in the second minimum. After Ref.[10].

The basic assumption of metadynamics is that $V_G(s,t)$ defined in Eq. 23 provides an estimate of the underlying free energy:

$$\lim_{t\to\infty} V_G(s,t) \sim -F(s). \tag{24}$$

This equation states that an equilibrium quantity, namely the free energy, can be estimated by a non-equilibrium dynamics, in which the underlying potential is changed every time a new Gaussian is added. In Ref. [10], Eq. **??** is postulated in a heuristic manner, observing the behavior of the dynamics 37 on free energy surfaces of known functional form. For instance, in the example of Fig. , it is clear that the sum of F and V_G after ˜320 Gaussians is approximately a constant, with deviations that would be in different positions for a statistically independent runs.

For an atomistic system in which the potential depends on the position of several atoms and the free energy is the result of a complex dimensional reduction, eq. **??** can be qualitatively understood in the limit of slow "deposition" (i.e. $w \to 0$). In this limit, $V_G(s,t)$ varies very slowly and the probability to observe s is always approximately proportional to $\exp\left[-\beta\left(F(s)+V_G(s,t)\right)\right]$. If the function $F(s)+V_G(s,t)$ has some local minimum, s will preferentially be localized in the neighborhood of this minimum and increasing numbers of Gaussians will be deposed there until this minimum is flattened. Let us consider

instead the case in which $F(s) \sim -V_G(s,t)$ in a region $\Omega(s)$. The probability distribution will be approximately flat in this region, and the location of the new Gaussians will not be affected by the bias deriving from the difference $F(s) + V_G(s,t)$. Hence, if $w \to 0$, the only corrugations in the free energy that are not flattened by the dynamics will be of the order of the size of the new Gaussians that are deposed.

Problems:

- The time required to escape from a local minimum in the free energy surface is determined by the number of Gaussians that are needed to fill the well. This number is proportional to $(1/\delta s)^d$, where d is the number of collective variables used in the system. Hence, the efficiency of the method scales exponentially with the number of dimensions involved. If d is large, the only way to obtain a reasonable efficiency is to use Gaussians with a size comparable to that of the well. On the other hand, a sum of Gaussians can only reproduce features of the FES on a scale larger than $\sim \delta s$. Already from these simple considerations it is clear that the metadynamics parameters w and δs strongly influence the quality of the reconstructed free energy and that, for a given problem, have to be carefully chosen in order to ensure the right compromise between accuracy and sampling efficiency. Large values for w and δs will allow a fast exploration of the CV space but at the price of a low accuracy.

- If a relevant variable is forgotten the algorithm is inaccurate: if the system performs a transition in the "hidden" degrees of freedom, the thermodynamic forces become inconsistent with the Gaussian potential. In particular, Eq. 24 holds only if the collective variables on which the history dependent potential acts are chosen in a smart manner. This might be a daunting task, since there is no *a priori* recipe for finding the correct set of collective variables.

2.4.1 Derivation of Eq. 24 for an overdumped Langevin dynamics

In order to demonstrate equation 24 we model the CVs evolution as a Langevin type dynamics whose equilibrium distribution as a function of s is proportional to $\exp(-\beta F(s))$:

$$ds(t) = -D \left.\frac{\partial F(s)}{\partial s}\right|_{s=s(t)} dt + \sqrt{2D} dW$$

where $ds(t) = s(t+dt) - s(t)$, dW is a Wigner noise, namely a Gaussian process of variance 0 and covariance dt , D is a diffusion coefficient and we measure the energies in unit of temperature. We also assume that the system is confined in a region Ω, and that the dynamics satisfies reflecting boundary conditions on $\partial\Omega$.In metadynamics, the normal equilibrium dynamics of the CVs is biased by an history-dependent term that discourages the system from revisiting positions

in the CV space that have already been explored. The evolution of the system can be modelled by the SDE

$$ds\left(t\right)=-D\,\frac{\partial}{\partial s}\left[F(s)+\int_{0}^{t}dt'g(s,s\left(t'\right))\right]\Big|_{s=s(t)}dt+\sqrt{2D}dW\ . \tag{25}$$

The time integral term in Eq. 25 is an history dependent potential, generated through the kernel $g(s,s')$. If one takes

$$g(s,s')=\frac{w}{\tau_G}\exp\left(-\frac{|s-s'|^2}{2\delta s^2}\right)$$

one recovers Eq. 23. We want to prove that for large enough t, $\int_0^t dt'g(s,s(t'))$ is an unbiased estimator of $F\left(s\right)$, namely

$$\overline{\int_{0}^{t}dt'g(s,s(t'))}=F\left(s\right)$$

where the overbar denotes an average over several realizations of the dynamic process (all of time duration t). The kernel $g(s,s')$ is required to be such that it exists a function $\varphi_0(s)$ such that the equation

$$\int ds'g(s,s')\varphi_0(s')+F(s)=0 \tag{26}$$

has a solution (namely, we require the free energy to be representable as a convolution with the kernel g).

Eq. 25 contains a history dependent term (the bias potential) and is clearly non-Markovian. In order to circumvent this problem we define a time dependent field $\varphi(s;t)$

$$\varphi(s;t)=\int_{0}^{t}dt\delta(s-s(t))+\varphi_0(s) \tag{27}$$

that is made of two terms: the histogram of the positions already visited by the system and a time independent term $\varphi_0(s)$. With this choice of the gauge it is implicitly assumed that the initial conditions are $\varphi(s;0)=\varphi_0(s)$, so that $\int ds'g(s,s')\varphi(s;0)=F\left(s\right)$. In terms of the variables $s(t)$ and $\varphi(s;t)$ the stochastic process in Eq. 25 can be rewritten in the simple form

$$ds(t)=-D\int ds'\frac{\partial g(s,s')}{\partial s}\varphi(s';t)\Big|_{s=s(t)}dt+\sqrt{2D}dW \tag{28}$$

$$d\varphi(s;t)=\delta(s-s(t))dt \tag{29}$$

as can be verified by direct substitution. This is the crucial step which allows replacing the non-Markovian evolution of a single dynamic variable $s(t)$ in Eq. 25 with a Markovian evolution for the extended set of variables which includes $s(t)$ and the field $\varphi(s;t)$. In fact, the state of the system at time $t+dt$, $\left[s\left(t+dt\right),\varphi\left(s;t+dt\right)\right]$ depends only on the state of the system at time

$t, [s(t), \varphi(s;t)]$. The information related to the underlying free energy $F(s)$ has disappeared from the equation of motion but is still present through the initial condition for $\varphi(s;t)$, see Eq. 27.

Using the Markovian property it is possible to analyze in a rigorous manner the behavior of Eq. 29. In order to study the average properties of an ensemble of independent metadynamics calculations we have to transform the stochastic description of Eq. 25 in a probabilistic description. When the stochastic evolution is Markovian, this is done using the Fokker-Planck equation. Recall that a SDE of the form

$$dx = a(x)\,dt + bdW$$

corresponds to a Fokker-Planck equation of the form

$$\frac{\partial P(x,t)}{\partial t} = \frac{\partial a(x) P(x,t)}{\partial x} + \frac{b^2}{2}\frac{\partial^2 P}{\partial x^2}$$

We write a generalized Fokker-Plank equation for the probability distribution associated with Eq. 29. Afterwards, we study its asymptotic behaviour for large t. We consider an ensemble of independent metadynamics runs, and define an ensemble density. Since our dynamic variables are the position of the walker s and the field $\varphi(s)$, the probability density will be a function of s and a functional of φ. We denote this probability as $P(\{\varphi\}, s;t)$. The Fokker-Planck equation for $P(\{\varphi\}, s;t)$ is

$$\begin{aligned}\frac{\partial P(\{\varphi\},s;t)}{\partial t} = & -\frac{\delta P(\{\varphi\},s;t)}{\delta\varphi(s)} + DP(\{\varphi\},s;t)\int ds' \frac{\partial^2 g(s,s')}{\partial s^2}\varphi(s') + \\ & + D\frac{\partial P(\{\varphi\},s;t)}{\partial s}\int ds' \frac{\partial g(s,s')}{\partial s}\varphi(s') + D\frac{\partial^2 P(\{\varphi\},s;t)}{\partial s^2}\end{aligned} \tag{30a}$$

Here, if the dimensionality of the system is higher than 1, a trace is implied and the second derivative is in fact a Laplacian.

We now look for the limiting distribution of Eq. 30a namely for the probability density $\bar{P}$ which satisfies $\frac{\partial \bar{P}(\{\varphi\},s;t)}{\partial t} = 0$. We assume that the equilibrium probability is independent on the walker position, i.e. $\bar{P}(\{\varphi\}, s) = \bar{P}(\{\varphi\})$. Inserting this *ansatz* in Eq. 30a we find that if $\frac{\partial^2 g(s,s')}{\partial s^2}$ is symmetric and negative definite, the equilibrium distribution is

$$\bar{P}(\{\varphi\}) = C\exp\left(\frac{D}{2}\int ds ds' \varphi(s)\frac{\partial^2 g(s,s')}{\partial s^2}\varphi(s')\right) \tag{31}$$

where C is a normalization constant.

Equation 31 expresses the probability of obtaining a given field φ at the end of a metadynamics simulation. Since the negative of the biasing potential is used to estimate the free energy, we define the error $\epsilon(s)$ as the sum of the exact underlying free energy and the biasing potential. Using Equations 26 and 27 we find that the error is linearly related to the field φ through

$$\epsilon(s) = \int ds' g(s,s')\varphi(s';t)\ . \tag{32}$$

Equation 31 states that any possible field (or error) is observable at the end of a simulation, but small fields (or errors) will be more likely to be obtained. Using Eq. 31 we can explicitly calculate the expected average error of a series of runs. Since the distribution is even with respect to the field $\varphi(s)$, namely $\bar{P}(\{\varphi\}) = \bar{P}(\{-\varphi\})$, the expectation value of this field is vanishing. The error $\epsilon(s)$ is linear in the field $\varphi(s)$, and consequently also its expectation value is vanishing:

$$\langle \epsilon(s) \rangle = 0 \tag{33}$$

Thus, we proved that the average of the biasing potential over a series of metadynamics runs provides an unbiased estimate for the underlying free energy.

Using Eq. 31 we can also address the problem of the accuracy, determining the expected quadratic deviation $\langle \epsilon^2(s) \rangle$ of a single metadynamics run from the average. Since the distribution in Eq. 31 is a Gaussian with respect to φ, this expectation value can be easily calculated on a suitable basis, for instance, the eigenvectors of the Laplacian operator on Ω. Assuming a cubic d dimensional domain with side S and reflecting boundaries, the eigenvectors of the Laplacian are $a_k(s) = \sqrt{\frac{2}{S^d}} \cos(\frac{\pi k \cdot s}{S})$, where k is a d dimensional array of integers. Thus we have

$$g(s, s') = \sum_{k \neq 0} g_k a_k(s) a_k(s'), \text{ with } g_k > 0 \ , \tag{34}$$

and

$$\langle \epsilon^2(s) \rangle = \sum_k \frac{g_k a_k^2(s) S^2}{D \pi^2 k^2} \tag{35}$$

The average value of the error on the domain Ω is

$$\langle \epsilon^2 \rangle = \frac{1}{S^d} \int ds \langle \epsilon^2(s) \rangle = S^{2-d} \sum_k \frac{g_k}{D \pi^2 k^2} \tag{36}$$

This expression for the error has been shown to reproduce the standard deviation of many metadynamics realizations in real system.

3 Metastability from the spectral properties of the transfer operator

In this section we provide a general and quantitative definition of metastability, which is valid if the dynamics of the system is described by a Markov process.

The time-dependent probability of the system is assumed to satisfy the equation

$$P(x, t+\tau) = \int dx' \pi_\tau (x' \to x) P(x', t) \tag{37}$$

The transfer operator π_τ is a *transition probability*, namely it satisfies

$$\pi_\tau (x' \to x) \geq 0 \tag{38}$$

$$\int dx \pi_\tau (x' \to x) = 1 \tag{39}$$

The first consequence of properties 39 is that the dynamics 37 preserves the normalization of P. In fact, integrating both members of eq. 37 and using 39, we have

$$\int dx P(x, t+\tau) = \int dx \int dx' \pi_\tau (x' \to x) P(x', t) = \int dx' P(x', t)$$

In MC and MD, π_t satisfies a supplementary condition called *detailed balance*. Namely, it exists a positive defined function $P_{eq}(x)$ such that

$$\pi_\tau (x' \to x) P_{eq}(x') = \pi_\tau (x \to x') P_{eq}(x)$$

It is easy to see that $P_{eq}(x)$ is the equilibrium distribution for the dynamics 37. In fact we have

$$\begin{aligned} P(x, t+\tau) &= \int dx' \pi_\tau (x' \to x) P_{eq}(x') = \\ &= \int dx' \pi_\tau (x \to x') P_{eq}(x) = P_{eq}(x) \end{aligned}$$

The latter equation also says that $P_{eq}(x)$ is an eigenfunction of the operator $\pi_\tau (x \to x')$ associated with an eigenvalue 1.

If the transition probability is known only for short times, eq. 37 is more conveniently written in the form of a master equation. Using 39 we have

$$\begin{aligned} P(x, t+\tau) - P(x, t) &= \int dx' \pi_\tau (x' \to x) P(x', t) - P(x, t) = \\ &= \int dx' \left(\pi_\tau (x' \to x) - \delta (x' - x) \right) P(x', t) \end{aligned}$$

and thus

$$\frac{\partial P(x,t)}{\partial t} = \int dx' m(x' \to x) P(x',t) \tag{40}$$

where

$$m(x' \to x) = \lim_{\tau \to 0} \frac{\pi_\tau(x' \to x) - \delta(x' - x)}{\tau}$$

3.1 En example: the transfer operator for Langevin dynamics

Consider an ovrerdumped Langevin process in x:

$$x(t+dt) = x(t) - dt\beta D\nabla V + \sqrt{2D}W(t) \tag{41}$$

where $V(x)$ is the potential, β is the inverse temperature, D is the diffusion coefficient and $W(t)$ is a Wiener process, namely a stochastic process satisfying the following properies:

1. Zero mean: $\langle W \rangle = 0$
2. Gaussian probability distribution, with covariance equal to dt: $P(W) = \frac{1}{\sqrt{2\pi dt}} \exp\left(-\frac{W^2}{2dt}\right)$
3. Markovian: $\langle W(t) W(t') \rangle = 0$ for $t \neq t'$

We now write the evolution defined by this equation in the tranfer operator form introduced in the previous Section. Denoting $x(t+dt) = x'$ and $x(t) = x$, the explicit form of the transition probability $\pi_{dt}(x \to x')$ can be deduced noticing that the probability distribution of W is Gaussian of covariance dt. From equation 41 we have $W = \frac{x' - x + D\beta\nabla V dt}{\sqrt{2D}}$. Hence

$$\begin{aligned}
\pi_{dt}(x \to x') &= \frac{1}{\sqrt{2D}} P\left(W^2\right) = \\
&= \frac{1}{\sqrt{2D}} \frac{1}{\sqrt{2\pi dt}} \exp\left(-\frac{W^2}{2dt}\right) = \\
&= \frac{1}{\sqrt{4\pi D dt}} \exp\left(-\frac{(x' - x + D\beta\nabla V dt)^2}{4dtD}\right)
\end{aligned}$$

It is easy to verify that $\pi_{dt}(x \to x') > 0$ and $\int dx' \pi_{dt}(x \to x') = 1$. Moreover,

$$\lim_{dt \to 0} \left(\pi_{dt}(x \to x') \exp(-\beta V(x)) - \pi_{dt}(x' \to x) \exp(-\beta V(x'))\right) = 0$$

namely, for dt small enough, π_{dt} satisfies detailed balance for $P_{eq}(x) = \frac{1}{Z} \exp(-\beta V(x))$. This conditions are sufficient to ensure that the equilibrium distribution of a dynamics generated with eq. 41 is canonical.

3.2 Properties of the transfer operator

We now discuss the properties of Markov processes using the more general form eq. 37.

1. *The eigenvalues of π_τ satisfying detailed balance are all real.* The eigenvalue equation for π_τ is

$$\int dx' \pi_\tau (x' \to x) \varphi_i (x') = \lambda_i \varphi_i (x) \tag{42}$$

Consider the operator

$$h_\tau (x' \to x) = \sqrt{\frac{P_{eq}(x')}{P_{eq}(x)}} \pi_\tau (x' \to x)$$

h_τ is Hermitian (its elements are real and it satisfies $h_\tau (x' \to x) = h_\tau (x \to x')$) and has the same eigenvalues of π_τ : from 42

$$\int dx' \pi_\tau (x' \to x) \sqrt{\frac{P_{eq}(x')}{P_{eq}(x)}} \frac{\varphi_i (x')}{\sqrt{P_{eq}(x')}} = \lambda_i \frac{\varphi_i (x)}{\sqrt{P_{eq}(x)}}$$

and thus

$$\int dx' h_\tau (x' \to x) \psi_i (x') = \lambda_i \psi_i (x)$$

where

$$\psi_i (x) = \frac{\varphi_i (x)}{\sqrt{P_{eq}(x)}}$$

Since the ψ_i-s are the eigenfunctions of an hermitian operator, they form an orthonormal set:

$$\int dx \psi_i (x) \psi_j (x) = \delta_{ij} \tag{43}$$

2. *The eigenvalue of π_τ satisfying 39 are all smaller than one in absolute value (Perron-Flobenius theorem).* In fact, consider the absolute value of both members of the left eigenvalue equation:

$$|\lambda_i| |\varphi_i (x)| = \left| \int dx' \pi_\tau (x \to x') \varphi_i (x') \right|$$

Next, using the triangular inequality and $\pi_\tau (x' \to x) \geq 0$ for the second member

$$|\lambda_i| |\varphi_i (x)| \leq \int dx' \pi_\tau (x \to x') |\varphi_i (x')|$$

Denoting by C the maximum of $|\varphi_i (x')|$, and using $\int dx' \pi_\tau (x \to x') = 1$

$$\int dx' \pi_\tau (x \to x') |\varphi_i (x')| \leq \int dx' \pi_\tau (x \to x') C = C$$

Thus,

$$|\lambda_i|\,|\varphi_i(x)| \leq C$$

This inequality must be valid for all x, including the $\widehat{x}$ one for which $|\varphi_i(\widehat{x})| = C$. This gives $|\lambda_i|\,|\varphi_i(\widehat{x})| = |\lambda_i|\,C \leq C$, and thus $\lambda_i \leq 1$, c.v.d.

Using the eigenvector expansion, it is easy to construct the formal solution for the dynamics 37 for an initial condition $P(x,0)$. We look for a solution of the form

$$P(x,t) = \sum_i c_i(t)\,\varphi_i(x)$$

where $c_i(t)$ are coefficients to be determined. Substituting in the dynamics equation 37 we have $\sum_i c_i(t+\tau)\,\varphi_i(x) = \sum_i c_i(t)\,\lambda_i\varphi_i(x)$, which is solved by $c_i(t+\tau) = \lambda_i c_i(t)$, and thus

$$c_i(n\tau) = \lambda_i^n c_i(0)$$

From the initial condition and using Eq. 43 we have

$$c_i(0) = \int dx \frac{\varphi_i(x)}{P_{eq}(x)} P(x,0)$$

Notice that, since $\varphi_1(x) = P_{eq}(x)$, $c_1(0) = \int dx P(x,0) = 1$.The probability after n steps will be given by

$$P(x,n\tau) = \varphi_1(x)\,c_1(0) + \sum_{i>1} \lambda_i^n \varphi_i(x)\,c_i(0) = P_{eq}(x) + \sum_{i>1} \lambda_i^n \varphi_i(x)\,c_i(0)$$

Since $|\lambda_i| < 1$ for all $i > 1$, $\lim_{n\to\infty} P(x,n\tau) = P_{eq}(x)$ regardless on the initial conditions $P(x,0)$.

Imagine now that we sort the eigenvalues according to their absolute value and that $|\lambda_2| \gg |\lambda_3|$ (namely a spectral gap exists). In these conditions, after a small number of steps only the contributions from $\lambda_1 = 1$ and λ_2 will survive:

$$P(x,n\tau) \sim P_{eq}(x) + \lambda_2^n \varphi_2(x)\,c_2(0) \tag{44}$$

This observation leads to a very popular definition of metastability: a system is metastable if the spectrum of its evolution operator has a gap. Of course it is necessary to specify what a gap is...

Eq. 44 allows also the regions A_i-s which have been introduced in Section 1 for giving a first definition of metastability. We notice in fact that $\int dx\varphi_2(x) = 0$ (to show this integrate both members of Eq. 44). Thus, $\varphi_2(x)$ is somewhere negative and somewhere positive. Eq. 44 states that at small time (small n) it is more likely to observe the system in values of x in which $\varphi_2(x) > 0$. For large n this will not be the case anymore. Thus Eq. 44 describes transitions from the region

$$A = \{x \in \Omega : \varphi_2(x) > 0\} \tag{45}$$

to its complement. If λ_2 is very close to one, these transitions will be unlikely, rare, and a large number of steps will be required before the system reaches equilibrium.

More in general, the number of metastable subsets (the number of regions A_i) depends on the number of eigenvalues of the propagator close to its maximum eigenvalue $\lambda = 1$ and before the gap.

3.3 The rate constants from the microscopic transition probability

If the spectrum of the transfer operator of the system has a gap, it is possible to write its dynamics in the form of a rate equation3. Consider for example the case in which a single eigenvalue λ is close to one. As we have shown, the probability at time ndt has the form $P(x, ndt) \sim P_{eq}(x) + \lambda^n \varphi_2(x) c_2(0)$ and the probability p_A to observe the system in the region A defined in Eq. 45 is

$$p_A(ndt) = p_A^{eq} + \lambda^n \alpha \tag{46}$$

with $p_A = \int_A dx P(x, ndt)$and $\alpha = c_2(0) \int_A dx \varphi_2(x)$. We derive from this a differential equation for p_A:

$$\begin{aligned} \frac{dp_A}{dt} &= \frac{p_A((n+1)dt) - p_A(ndt)}{dt} = \\ &= \lambda^n \alpha \frac{(\lambda - 1)}{dt} = \\ &= \frac{(1-\lambda)}{dt}(p_A^{eq} - p_A) \end{aligned}$$

where, in the second passage, we have solved Eq. 46 with respect to $\lambda^n \alpha$. Remarkably, this has exactly the form of Eq. 4. *This allows identifying explicitly the rates from the eigenvalues and the eigenvectors of the microscopic transition probability.* Indeed, we have

$$\begin{aligned} \frac{(1-\lambda)}{dt} &= k_{A\to B} + k_{B\to A} \\ \int_A dx P_{eq}(x) &= \frac{k_{B\to A}}{k_{A\to B} + k_{B\to A}} \end{aligned}$$

These two equations, if $P_{eq}(x)$, $\varphi_2(x)$ and λ are known, can be easily solved with respect to the rate constants.

It is also easy to verify that the macroscopic rate equation inherits the properties of the microscopic dynamics 37 from which it can be derived. Indeed, Eq. 3 can be written in the form $p(t + dt) = w_{dt} p(t)$with

$$w_{dt} = \begin{pmatrix} 1 - dtk_{A\to B} & dtk_{B\to A} \\ dtk_{A\to B} & 1 - dtk_{B\to A} \end{pmatrix}$$

It can be easily verified that:

1. w_{dt} satisfies the sum rule 39
2. Its eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 1 - dt\,(k_{A\to B} + k_{B\to A})$.
3. w_{dt} satisfies detailed balance with respect to $p^{eq} = \begin{pmatrix} \frac{k_{B\to A}}{k_{A\to B}+k_{B\to A}} \\ \frac{k_{B\to A}}{k_{A\to B}+k_{B\to A}} \end{pmatrix}$, the eigenvactor corresponding to $\lambda = 1$
4. Finally,

   $$p\,(t) - p^{equilibrium} \quad \propto \quad \lambda_2^{t/dt} = (1 - dt\,(k_{A\to B} + k_{B\to A}))^{t/dt} = \exp\left(-t\,(k_{A\to B} + k_{B\to A})\right)$$

   This is consistent with the explicit solution Eq. 5.

3.4 Relaxation dynamics on a network

Consider a network of sites labeled by i. The connettivity of the network is defined by a matrix χ_{ij}, whose elements are equal to 1 if a link between nodes i and j exists and 0 otherwise. The transition probabilities in a time dt are given by

$$\begin{aligned} \pi_{i\to j} &= \frac{dt}{\tau_{ij}}\chi_{ij}\exp\left(\frac{\beta}{2}\,(F_i - F_j)\right) \\ \pi_{i\to i} &= 1 - \sum_{j\neq i}\pi_{i\to j} \end{aligned}$$

where β is the inverse temperature and $\tau_{ij}$is a constant that fixes the time scale of the transitions and that, in general, depends on i and j. Clearly, if dt is sufficiently small, this dynamics satisfies Eqs. 39. Moreover, it satisfies detailed balance with respect to a probability

$$p_i^{eq} \propto \exp\left(-\beta F_i\right)$$

Exercise: consider a one dimensional network with 100 sites, $\chi_{ij} = \delta_{i,i+1}$. Take $dt/\tau_{ij} = 0.001$, $F_1 = F_{100} = \infty$ and

$$F_i = \left(\frac{i-50}{25}\right)^4 - 3\left(\frac{i-50}{25}\right)^2$$

Compute the spectrum of the transfer operator as a function of β and discuss the result in terms of the metastability properties of the system. For $\beta = 1$ also plot the eigenfunctions corresponding to the three largest eigenvalues.

4 Microscopic definition of the rate constants: transition state theory

In Section 1 we described a general procedure for computing the rate constant from the eigenvalues of the transfer operator. Unfortunately, the approach requires evaluating explicitly the spectrum of the transfer operator. If the dimensionality of the system is too large, this can be very difficult or impossible. In this section, we introduce transition state theory, a more traditional manner of estimating the rate constant, that is less general, but that can be applied also for systems with many degrees of freedom. In this Section we partially follow the derivation in Ref. [6].

4.1 The rate constant as the ratio between flux and population

Consider once again a system whose phase space Ω can be divided in a region A and a region B. We have seen that if the system displays metastability, the population of the two regions satisfies a rate equation of the form:

$$\frac{dp_A}{dt} = -k_{A\to B}p_A + k_{B\to A}p_B \tag{47}$$

$$\frac{dp_B}{dt} = k_{A\to B}p_A - k_{B\to A}p_B \tag{48}$$

We want to estimate $k_{A\to B}$ and $k_{B\to A}$ appearing in this equation from the microscopic properties of the system, If the two regions are "well-defined" in a sense that we will specify in the following, the rate constant $k_{A\to B}$ can be estimated as the ratio between the flux $\Phi_{A\to B}$ between A and B and the equilibrium probability $p_A^{\text{equilibrium}}$ to observe the system in A . In fact, in the rate model 48 we have

$$\Phi_{A\to B} = \begin{array}{c}\text{probability per unit time}\\ \text{to observe a transition}\\ \text{between } A \text{ and } B\end{array} = \frac{1}{\tau_{A\to B}+\tau_{B\to A}} = \frac{k_{A\to B}k_{B\to A}}{k_{A\to B}+k_{B\to A}}$$

where we used $\tau_{A\to B} = 1/k_{A\to B}$ and $\tau_{B\to A} = 1/k_{B\to A}$. Moreover

$$p_A^{\text{equilibrium}} = \frac{k_{B\to A}}{k_{A\to B}+k_{B\to A}}$$

Thus, as we anticipated, $k_{A\to B} = \Phi_{A\to B}/p_A^{\text{equilibrium}}$.

4.2 The rate constant in one-dimensional Langevin Dynamics

4.3 Transition state theory

In transition state theory both $\Phi_{A\to B}$ and $p_A^{\text{equilibrium}}$ are expressed as ensemble averages over the probability $P(x,v)$ to observe the system in (x,v). The

equilibrium population to observe the system in A is given by

$$p_A^{\text{equilibrium}} = \int d\mathbf{x} \int d\mathbf{v} P(\mathbf{x}, \mathbf{v}) \chi_A(\mathbf{x})$$

where $\chi_A(x)$ is the characteristic function of the region A. The flux $\Phi_{A\to B}$ between A and B is approximated by the flux through the boundary of A (we will comment later on the validity of this assumption):

$$\Phi_{\partial A} = \frac{1}{2} \int_{\partial A} d\sigma(\mathbf{x}) \int d\mathbf{v} P(\mathbf{x}, \mathbf{v}) |\mathbf{n}(\mathbf{x}) \cdot \mathbf{v}|$$

where $\mathbf{n}(\mathbf{x})$ is a unit vector orthogonal to ∂A in $\mathbf{x}$. If we assume that ∂A is parametrized by an equation $s(\mathbf{x}) = 0$ we have $\mathbf{n}(\mathbf{x}) = \frac{\nabla s(\mathbf{x})}{\|\nabla s(\mathbf{x})\|}$ and $\int_{\partial A} d\sigma(\mathbf{x}) = \int d\mathbf{x} \delta(s(\mathbf{x})) \|\nabla s(\mathbf{x})\|$. Thus,

$$\Phi_{\partial A} = \frac{1}{2} \int d\mathbf{x} \delta(s(\mathbf{x})) \int d\mathbf{v} P(\mathbf{x}, \mathbf{v}) |\nabla s(\mathbf{x}) \cdot \mathbf{v}|$$

Putting all together, we have

$$k_{A\to B}^{TST} = \frac{\Phi_{\partial A}}{p_A^{\text{equilibrium}}} = \frac{\int d\mathbf{x} \int d\mathbf{v} P(\mathbf{x}, \mathbf{v}) \delta(s(\mathbf{x})) |\nabla s(\mathbf{x}) \cdot \mathbf{v}|}{2 \int d\mathbf{x} \int d\mathbf{v} P(\mathbf{x}, \mathbf{v}) \chi_A(\mathbf{x})}$$

We now take $P(\mathbf{x}, \mathbf{v}) \propto e^{-\beta\left(\frac{m}{2}\|\mathbf{v}\|^2 + V(\mathbf{x})\right)}$ (to simplify the notation we have assumed that the system is made of identical particles of mass m). The integral on the velocities can be computed explicitly. This gives

$$k_{A\to B}^{TST} = \sqrt{\frac{4}{\pi m \beta}} \frac{\int d\mathbf{x} \delta(s(\mathbf{x})) |\nabla s(\mathbf{x})| \exp(-\beta V(\mathbf{x}))}{\int d\mathbf{x} \chi_A(\mathbf{x}) \exp(-\beta V(\mathbf{x}))} \tag{49}$$

The rate constant evaluated by Eq. 49 is well-defined only if the regions A and B are properly chosen. In fact, while the integral at the denominator is well behaved and largely insensitive to the exact definition of A, the surface integral in the numerator depends strongly on the specific choice of ∂A. Indeed, a trajectory that has just crossed the boundary of A might have a chance to recross it back right away, giving no contribution to the true (macroscopic) flux between A and B. In other words,

$$\Phi_{A\to B} \leq \Phi_{\partial A}$$

The ratio

$$\frac{k_{A\to B}}{k_{A\to B}^{TST}} = \frac{\Phi_{A\to B}}{\Phi_{\partial A}} \doteq \kappa$$

is by construction smaller than one and is called transmission coefficient of ∂A.

4.4 Bennett-Chandler procedure for computing κ

In order to define more rigorously the rate constant between two regions it is necessary to assume that these are "well-separated" by a buffer region in which the equilibrium probability to observe the system is almost zero. In order to estimate the rate, we need to count the reactive trajectories, namely the trajectory actually leading from A to B, and not the flux across a surface in between them. Let's consider in detail the case of Hamiltonian dynamics, in which the evolution of the system depends only on the initial conditions (x, v). We choose a dividing surface Σ belonging to the buffer region between A and B. For all (x, v) with $x \in \Sigma$, we can evolve the system backward and forward in time until it reaches A or B. For this trajectory, we compute the number of times $N_\Sigma(x, v)$ it crosses Σ. If $N_\Sigma(x, v)$ is even, the trajectory connects A with A or B with B. Thus, it gives no contribution to the macroscopic flux between A and B. If instead $N_\Sigma(x, v)$ is odd, the trajectory connects A with B, and gives a contribution to the flux. We also notice that looping over all the possible (x, v), we will find this trajectory $N_\Sigma(x, v)$ times. Thus, in order to quantify the macroscopic flux, we introduce a function

$$g_\Sigma(x, v) = \frac{1 - (-1)^{N_\Sigma(x,v)}}{2N_\Sigma(x, v)}$$

For non reactive (x, v) this function is zero. For reactive trajectory it is equal to $1/N_\Sigma(x, v)$. The flux between A and B can be written as an explicit average of this function.

$$\begin{aligned} k_{A\to B} &= \frac{\text{flux between } A \text{ and } B}{\text{equilibrium population of } A} \\ &= \frac{\int_\Sigma d\sigma(x) \int dv P(x, v) g_\Sigma(x, v) |n(x) \cdot v|}{\int dx \int dv P(x, v) \chi_A(x)} = \\ &= k^{TST}_{A\to B} \frac{\int_\Sigma d\sigma(x) \int dv P(x, v) g_\Sigma(x, v) |n(x) \cdot v|}{\int_\Sigma d\sigma(x) \int dv P(x, v) |n(x) \cdot v|} = \\ &= k^{TST}_{A\to B} \langle g_\Sigma(x, v) \rangle_\Sigma \end{aligned}$$

where $k^{TST}_{A\to B}$ is the TST rate constant associated with Σ. This manner of computing $k_{A\to B}$ was introduced by Bennett and Chandler[1], but the derivation presented here is taken from [6]. Its advantage is that the result is insensitive on the choice of Σ. Its practical implementation requires first choosing a reasonable Σ and computing $k^{TST}_{A\to B}$ using Eq. 49. Afterwards, the transmission coefficient $\kappa = \langle g_\Sigma(x, v) \rangle_\Sigma$ is estimated running several short trajectories. The initial points of these trajectories are located on the surface Σ and are distributed with a probability proportional to $P(x, v) |n(x) \cdot v|$. Then, one evolves (x, v) backward and forward in time until the trajectory reaches A or B, computes $g_\Sigma(x, v)$ on this trajectory and evaluates the average $\langle g_\Sigma(x, v) \rangle$.

The problem of this procedure is that if Σ is not properly chosen, κ can be very small and the true value of $k_{A\to B}$ can be obtained only computing

accurately this very small number. This, in practical applications, can be computationally heavy. Moreover, if the dynamics is intrinsically diffusive, whatever Σ is chosen the trajectories will cross it several times. In these cases, N_Σ can be intrinsically large, and computing g_Σ is difficult even if the best possible dividing surface is chosen.

4.5 Optimizing TST: minimize the flux through the dividing surface

From what we discussed in the previous section, it is clear that in order to estimate $k_{A\to B}$ it is necessary to choose Σ in such a way that $k^{TST}_{A\to B}$ is minimum. In these conditions, κ will be as close as possible to 1 and it will be easy to compute. Since the denominator in 49 is practically unchanged by the choice of Σ, minimizing $k^{TST}_{A\to B}$ is equivalent to minimizing the flux

$$\Phi_\Sigma = \int d\mathbf{x} \exp\left(-\beta V\left(\mathbf{x}\right)\right) \left\|\nabla s\left(\mathbf{x}\right)\right\| \delta\left(s\left(\mathbf{x}\right)\right)$$

We now compute the first order variation of Φ_Σ corresponding to a variation $u\left(x\right) = \delta s\left(x\right)$ of $s\left(x\right)$ (the notation $u\left(x\right)$ is used in order to avoid confusion between the variation $\delta s\left(\mathbf{x}\right)$of $s\left(\mathbf{x}\right)$ and the Dirac delta $\delta\left(s\left(\mathbf{x}\right)\right)$):

$$\delta\Phi_\Sigma = \int d\mathbf{x} e^{-\beta V(x)} \left[\delta\left(s\left(\mathbf{x}\right)\right) \mathbf{n}\left(x\right) \cdot \nabla u\left(\mathbf{x}\right) + \left\|\nabla s\left(\mathbf{x}\right)\right\| \delta'\left(s\left(\mathbf{x}\right)\right) u\left(\mathbf{x}\right)\right]$$

where $\mathbf{n}\left(\mathbf{x}\right) = \frac{\nabla s(\mathbf{x})}{\|\nabla s(\mathbf{x})\|}$ is the unit vector orthogonal to Σ in x. Using $\delta'\left(s\left(\mathbf{x}\right)\right) = \frac{\nabla s(\mathbf{x})}{\|\nabla s(\mathbf{x})\|^2} \cdot \nabla\delta\left(s\left(\mathbf{x}\right)\right) = \frac{1}{\|\nabla s(\mathbf{x})\|}\mathbf{n}\left(\mathbf{x}\right) \cdot \nabla\delta\left(s\left(\mathbf{x}\right)\right)$ and integrating by part the term in $\nabla\delta\left(s\left(\mathbf{x}\right)\right)$ we have

$$\begin{aligned}
\delta\Phi_\Sigma &= \int d\mathbf{x} e^{-\beta V(\mathbf{x})} \left[\delta\left(\mathbf{s}\left(\mathbf{x}\right)\right) \mathbf{n}\left(\mathbf{x}\right) \cdot \nabla u\left(\mathbf{x}\right) + \mathbf{n}\left(\mathbf{x}\right) \cdot \nabla\delta\left(s\left(\mathbf{x}\right)\right) u\left(\mathbf{x}\right)\right] = \\
&= \int d\mathbf{x}\delta\left(s\left(\mathbf{x}\right)\right) \left[e^{-\beta V(\mathbf{x})}\mathbf{n}\left(\mathbf{x}\right) \cdot \nabla u\left(\mathbf{x}\right) - \nabla \cdot \left(e^{-\beta V(\mathbf{x})}\mathbf{n}\left(\mathbf{x}\right) u\left(\mathbf{x}\right)\right)\right] = \\
&= \int d\mathbf{x}\delta\left(s\left(\mathbf{x}\right)\right) e^{-\beta V(\mathbf{x})} \\
&\quad \left[\mathbf{n}\left(\mathbf{x}\right) \cdot \nabla u\left(\mathbf{x}\right) + \beta\left(\nabla V\left(\mathbf{x}\right) \cdot \mathbf{n}\left(\mathbf{x}\right)\right) u\left(\mathbf{x}\right) - \nabla \cdot \mathbf{n}\left(\mathbf{x}\right) u\left(\mathbf{x}\right) - \mathbf{n}\left(\mathbf{x}\right) \cdot \nabla u\left(\mathbf{x}\right)\right] = \\
&= \int d\mathbf{x}\delta\left(s\left(\mathbf{x}\right)\right) e^{-\beta V(\mathbf{x})} \left[\beta\left(\nabla V\left(\mathbf{x}\right) \cdot \mathbf{n}\left(\mathbf{x}\right)\right) - \nabla \cdot \mathbf{n}\left(\mathbf{x}\right)\right] u\left(\mathbf{x}\right)
\end{aligned}$$

At the minimum we must have $\delta\Phi_\Sigma = 0$ for all $u\left(\mathbf{x}\right)$. Then, the last equation implies

$$\beta\left(\nabla V\left(\mathbf{x}\right) \cdot \mathbf{n}\left(\mathbf{x}\right)\right) - \nabla \cdot \mathbf{n}\left(\mathbf{x}\right) = 0 \tag{50}$$

This is an equation in $\mathbf{n}\left(\mathbf{x}\right)$ that can be used for finding the optimal dividing surface.

Let's consider the case in which A and B are divided by a single saddle point of coordinates x^S. In a neighbourhood of x_S we have

$$V(\mathbf{x}) = V^S + \frac{1}{2} H^S_{\alpha\beta} \left(x - x^S\right)_\alpha \left(x - x^S\right)_\beta$$

where $V^S = V\left(\mathbf{x}^S\right)$ and $H^S_{\alpha\beta} = \left.\frac{\partial^2 V(\mathbf{x})}{\partial x_\alpha \partial x_\beta}\right|_{x^S}$. Since $\mathbf{x}^S$ is a saddle point, $H^S_{\alpha\beta}$ has one negative eigenvalue, that we denote by ω_1^S. It is straightforward to see that the surface Σ solving Eq. 50 is an hyperplane passing from $\mathbf{x}^S$and orthogonal to the eigenvector v_1 associated to ω_1^S. In fact, for an hyperplane $\nabla \cdot \mathbf{n}(\mathbf{x}) = 0$ as $\mathbf{n}(\mathbf{x})$does not depend on x. Moreover,

$$\begin{aligned} \nabla V(\mathbf{x}) \cdot \mathbf{n} &= H^S_{\alpha\beta} \left(x - x^S\right)_\alpha n_\beta = \\ &= H^S_{\alpha\beta} \left(x - x^S\right)_\alpha v_{1,\beta} = \\ &= \omega_1^S \left(x - x^S\right)_\alpha n_\alpha = 0 \end{aligned}$$

for all x belonging to the hyperplane.

4.6 Harmonic (or canonical)TST

(Missing: this part should be done using mass-reduced coordinates. This derivation is valid only if m is the same for all the particles)

This choice for the dividing surface leads to naturally a very useful expression for the rate constant that is usually referred to as "canonical TST rate constant" or simply "TST rate constant". We also Taylor expand $V(\mathbf{x})$ around the minimum A :

$$V(\mathbf{x}) = V^A + \frac{1}{2} H^A_{\alpha\beta} \left(x - x^A\right)_\alpha \left(x - x^A\right)_\beta$$

Under these assumptions, the integrals appearing in eq. 49 can be evaluated explicitly:

$$k^{TST}_{A\to B} = \sqrt{\frac{4}{\pi m \beta}} \exp\left(-\beta\left(V^S - V^A\right)\right) \frac{\prod_{\alpha>1} \omega^S_\alpha}{\prod_\alpha \omega^A_\alpha}$$

where ω^A_α and ω^S_α are, respectively, the eigenvalues of $H^A_{\alpha\beta}$ and of $H^S_{\alpha\beta}$.

5 Transition path sampling

In transition path sampling techniques[4, 5], an ensemble of "reactive" trajectories is generated by a montecarlo procedure in trajectory space.The weight of a reactive trajectory $\{x_0, \ldots, x_n\}$ is equal to

$$P(\{x_0, \ldots, x_n\}) = P(\{x_t\}) = C p_A(x_0) \chi_B(x_n) \prod_{t=0}^{n-1} \pi(x_t \to x_{t+1}) \qquad (51)$$

where C is a normalization constant and:

1. $p_A(x_0)$ is the probability distribution of the initial points of the trajectory. We take
$$p_A(x_0) \propto \exp(-\beta V(x_0))\,\chi_A(x_0) \tag{52}$$

2. $\pi(x_t \to x_{t+1})$ is the transition probability from x to x'. For an overdumped langevin process of the form
$$dx = -\beta D \partial_x V(x)\,dt + \sqrt{2D}dW$$
we have
$$\pi(x \to x') \frac{1}{\sqrt{4\pi D dt}} \exp\left(-\frac{(x' - x + D\beta\partial_x V(x)\,dt)^2}{4dtD}\right)$$
In any case, we assume that π satisfies detailed balance:
$$\pi(x \to x')\exp(-\beta V(x)) = \pi(x' \to x)\exp(-\beta V(x')) \tag{53}$$

We generate a new path by a "shooting procedure" that consists in generating a forward and backward trajectory from a point x_l chosen at random. The probability to generate the path $\{x'_t\}$ is given by
$$P_{gen}(\{x_t\} \to \{x'_t\}) = G(x_l \to x'_{l'}) \prod_{t=l'}^{1} \pi(x'_t \to x'_{t-1}) \prod_{t=l'}^{n-1} \pi(x'_t \to x'_{t+1})$$
where $G(x_l \to x'_{l'})$ is the probability to generate a new shooting point $x'_{l'}$ from the old one. We require $G(x \to x') = G(x' \to x)$. We now apply metropolis criterion, namely we accept the new path with a probability $\min(1, r(\{x_t\}, \{x'_t\}))$, with
$$\begin{aligned} r(\{x_t\}, \{x'_t\}) &= \frac{P_{gen}(\{x'_t\} \to \{x_t\})\,P(\{x'_t\})}{P_{gen}(\{x_t\} \to \{x'_t\})\,P(\{x_t\})} = \\ &= \frac{p_A(x'_0)}{p_A(x_0)} \chi_B(x'_n) \frac{w(l', \{x'_t\})}{w(l, \{x_t\})} \end{aligned} \tag{54}$$
where we assumed $\chi_B(x_n) = 1$ (the old trajectory belongs to the TPE) and
$$w(l, \{x_t\}) = \prod_{t=l-1}^{0} \frac{p(x_t \to x_{t+1})}{p(x_{t+1} \to x_t)}$$
Using detailed balance (Eq. 53) we have
$$w(l, \{x_t\}) = \prod_{t=l-1}^{0} \frac{\exp(-\beta V(x_{t+1}))}{\exp(-\beta V(x_t))} = \exp(\beta(V(x_0) - V(x_l)))$$
Using also $p_A(x_0) \propto \exp(-\beta V(x_0))\,\chi_A(x_0)$, we finally have
$$r(\{x_t\}, \{x'_t\}) = \chi_B(x'_n)\,\chi_A(x'_0) \exp(-\beta(V(x'_{l'}) - V(x_l)))$$

The acceptance only depends on the difference between the potential energies of the two shooting points (old and new) a topological requirement (the new trajectory must begin in A and end in B). Of course this latter condition is the most difficult to satisfy. In practice, $\chi_B(x_n)\chi_A(x_0)$ is likely to be different from zero only if the shooting point is in the transition state region. If a successful shooting point is available, the new trial shooting point has to be generated from from a $G(x_l \to x'_{l'})$ that has to be localized, both in time and in space, namely l' must be rather close to l and the distance $\|x_l - x'_{l'}\|$ must be rather small. As it is usual in montecarlo procedures, no big jumps are allowed, but too short jumps imply slow decorrelation, and a compromise has to be seeked.

5.1 The committor

Let us imagine that for every x not belonging to A or B we run several trajectories until they reach the boundary of A or the boundary of B. Then, we compute the fraction of these trajectories that reach the boundary of A. This fraction is called committor function and is denoted as $C_A(x)$. Of course $C_A(x) \in [0,1]$ and $C_A(x) = 1$ for $x \in \partial A$ and $C_A(x) = 0$ for $x \in \partial B$. We define a configuration x to be a transition state (TS) if both states A and B are equally accessible from that configuration, namely if

$$C_A(x) = 0.5 \tag{55}$$

This equation in general defines an hypersurface of dimensionality $n-1$, where n is the dimensionality of the system. This surface is called in the literature "transition path ensamble" (TPE). The "shape" of this surface, can of course be far from trivial in real cases. The committor C_A is, by definition, the best possible reaction coordinate.

TPS provides a very natural manner to find the configurations x satisfying equation 55. In fact, it is clear that from a configuration x it is likely to generate a reactive trajectory, for which $\chi_B(x_n)\chi_A(x_0) = 1$ only if $C_A(x) \sim 0.5$.More rigorously, the expected value of $\chi_B(x_n)\chi_A(x_0)$ is equal to $C_A(x)(1 - C_A(x))$.Hence, the shooting points generated in a TPS procedure are approximately distributed according to $e^{-\beta V(x)} C_A(x)(1 - C_A(x))$.

References

[1] C.H. Bennett. *Algorithms for Chemical Computations*. ACS Symposium Series No. 46, edited R.E. Christofferson (American Chemical Society, Washington D,C.), 1977.

[2] G. Bussi, A. Laio, and M. Parrinello. Equilibrium free energies from non-equilibrium metadynamics. *Phys. Rev. Lett., in press*, 2005.

[3] E.A. Carter, G. Ciccotti, J.T. Hynes, and R. Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.*, 156:472, 1989.

[4] C. Dellago, P. Bolhuis, F. S. Csajka, and D. Chandler. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.*, 108:1964–1977, 1998.

[5] C. Dellago, P.G. Bolhuis, and P.L. Geissler. Transition path sampling. *Adv. Chem. Phys.*, 123:1–78, 2002.

[6] W. E and E. Vanden-Eijnden. *Metastability, conformation dynamics, and transition pathways in complex systems.* Lecture notes in computational science and engineering, Ed. S. Attinger and P. Koumoutsakos, Springer, 2004.

[7] A.M. Ferrenberg and R.W. Swendsen. Optimized monte-carlo data-analysis. *Phys. Rev. Lett.*, 61:2635, 1988.

[8] M. Iannuzzi, A. Laio, and M. Parrinello. Efficient exploration of reactive potential energy surfaces using car-m. parrinelloolecular dynamics. *Phys Rev. Lett.*, 90:238302, 2003.

[9] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, 16:1339–1350, 1995.

[10] A. Laio and M. Parrinello. Escaping free energy minima. *Proc. Natl. Acad. Sci. USA*, 99:12562–12566, 2002.

[11] B. Roux. The calculation of the potential of mean force using computer-simulations. *Comput. Phys. Comm.*, 91:275–282, 1995.

[12] M. Sprik and G. Ciccotti. Free energy from constrained molecular dynamics. *J. Chem. Phys.*, 109:7737, 1998.