TRANSITION PATH THEORY FOR OVERDAMPED LANGEVIN DYNAMICS

MARIA CAMERON

Contents

1. Settings

We consider a system evolving according to the overdamped Langevin dynamics

$$dx = -\nabla V(x)dx + \sqrt{2\beta^{-1}}dw, \quad x \in \mathbb{R}^d, \tag{1}$$

where $V(x)$ is a smooth function, β^{-1} is the temperature times Boltzmann's constant, and w is the standard d-dimensional Brownian motion.

We assume that $V(x)$ is such that the invariant probability density exists and is given by the Gibbs density

$$\mu(x) = \frac{1}{Z}e^{-\beta V(x)}, \quad \text{where} \quad Z = \int_{\mathbb{R}^d} e^{-\beta V(x)}dx. \tag{2}$$

Suppose the invariant density is mostly split between two distinct regions that we denote by A and B. We are interested in studying the transitions of the system evolving according to SDE (1) from A to B. Therefore, we are interested in the so-called *reactive trajectories*, i.e., the trajectories starting at the boundary of A, ∂A, and going next to the boundary of B, ∂B, without returning to A in-between. If we consider a very long trajectory of (1), we can cut out reactive trajectories out of it. A few samples of reactive trajectories for SDE (1) with V being Mueller's potential are shown in Fig. 1.
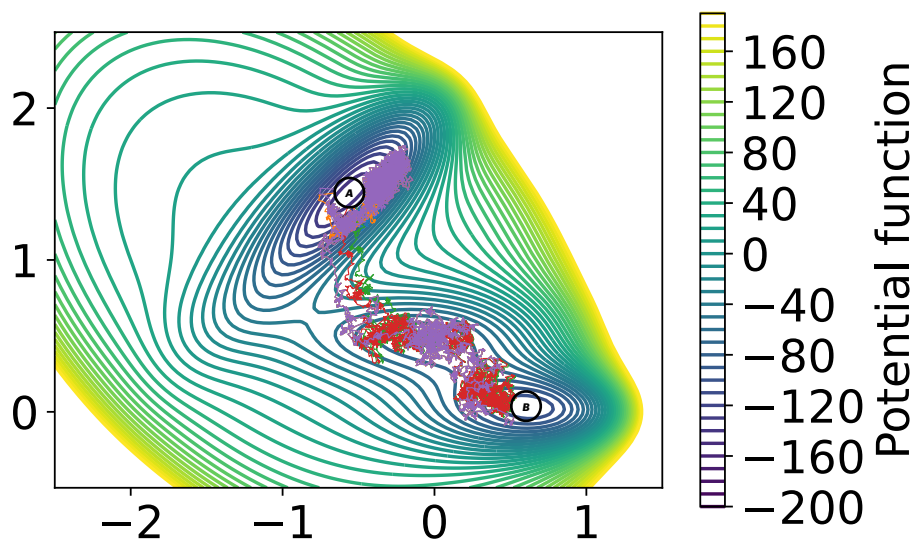
FIGURE 1. A few samples of reactive trajectories of a system evolving according to SDE (1) with V being Mueller's potential.

2. THE COMMITTOR

The transition path theory (TPT) [1, 2] offers a mathematical framework for describing the statistical properties of reactive trajectories and allows us to identify reactive channels and compute the transition rate from A to B. The most important function of TPT is the committor function $q(x)$ which is the probability that a process starting at x and evolving according to SDE (1) will hit first B rather than A. The committor function satisfies the following boundary-value problem (BVP):

$$
(3) \qquad \begin{cases} -\nabla V \cdot \nabla q + \beta^{-1}\Delta q = 0, & x \in \Omega_{AB} := (A \cup B)^c, \\ q = 0, & x \in \partial A, \\ q = 1, & x \in \partial B, \end{cases}
$$

The notation Ω_{AB} is the space $\mathbb{R}^d$ (or the manifold) where the system is living with removed sets A and B. The left-hand side of the PDE in (3) is the backward Kolmogorov operator, or the generator of SDE (1). Note that it can be rewritten in a self-adjoint form which will be helpful for us further:

$$
(4) \qquad \mathcal{L}q := -\nabla V \cdot \nabla q + \beta^{-1}\Delta q = \beta^{-1}e^{\beta V}\nabla \cdot \left(e^{-\beta V}\nabla q \right).
$$

3. TIME-REVERSIBILITY

It is important to note that the overdamped Langevin dynamics (1) are time-reversible. Let us check it.

For a general SDE of the form

$$(5) \qquad\qquad dx = b(x)dt + \sigma(x)dw,$$

the generator for the forward process is given by

$$(6) \qquad\qquad \mathcal{L}f = b \cdot \nabla + \frac{1}{2}\mathsf{tr}\left(\Sigma^\top \nabla\nabla f\right), \quad \text{where} \quad \Sigma = \sigma\sigma^\top.$$

Let $\mu(x)$ be the invariant probability density. Then the generator for the time-reversed process is given by

$$(7) \qquad\qquad \hat{\mathcal{L}}f = -b \cdot \nabla f + \frac{1}{\mu}\mathsf{div}\left(\Sigma\mu\right) \cdot \nabla f + \frac{1}{2}\mathsf{tr}\left(\Sigma^\top \nabla\nabla f\right).$$

The term $\mathsf{div}\left(\Sigma\mu\right)$ in the coordinate form is:

$$(8) \qquad\qquad \left[\mathsf{div}\left(\Sigma\mu\right)\right]_i = \sum_{j=1}^d \frac{\partial}{\partial x_j}\left(\Sigma_{ij}\mu\right) = \mu \sum_{j=1}^d \frac{\partial \Sigma_{ij}}{\partial x_j} + \sum_{j=1}^d \Sigma_{ij}\frac{\partial\mu}{\partial x_j}.$$

In the case of SDE (1), $b = -\nabla V$, $\mu = Z^{-1}e^{-\beta V}$, and $\Sigma = 2\beta^{-1}I$. Hence, the generator for the time-reversed process is

$$\hat{\mathcal{L}}f = \nabla V \cdot \nabla f - Ze^{\beta V}2\beta^{-1}Z^{-1}\beta\nabla V e^{-\beta V} + \beta^{-1}\Delta f$$
$$= -\nabla V \cdot \nabla f + \beta^{-1}\Delta f = \mathcal{L}f.$$

We see that $\hat{\mathcal{L}} = \mathcal{L}$, i.e., the overdamped Langevin dynamics are time-reversible. In particular, this means that the backward committor is $1 - q(x)$ where $q(x)$ is the forward committor, i.e., the solution to BVP (3).

4. The committor BVP and a minimization problem

Let us show that the committor, i.e., the solution to the BVP (3) is also the solution to the following minimization problem:

$$(9) \qquad q(x) = \arg\min\left\{\int_{\Omega_{AB}} \|\nabla f\|^2 e^{-\beta V}\,dx \,\middle|\, f \in C^1(\Omega_{AB}),\ f(\partial A) = 0,\ f(\partial B) = 1\right\}.$$

In word, we need to find a continuously differentiable function $f(x)$, $x \in \Omega_{AB}$, that minimizes the integral in (9) and satisfies the boundary conditions $f(\partial A) = 0$ and $f(\partial B) = 1$. The integral in (9) is called a *Dirichlet form*.

Let f^* be a minimizer of

$$\mathcal{D}(f) := \int_{\Omega_{AB}} \|\nabla f\|^2 e^{-\beta V}\,dx$$

among all twice continuously differentiable functions f satisfying $f(\partial A) = 0$ and $f(\partial B) = 1$. Hence, if we add a small perturbation $\delta f(x) \in C^1(\Omega_{AB})$ to $f(x)$ such that $\delta f(\partial A) = 0$ and $\delta f(\partial B) = 0$, the perturbed function $f^* + \Delta f$ will satisfy the boundary conditions in (9), and the value of the integral should increase:

$$(10) \qquad\qquad \mathcal{D}(f^* + \Delta f) \geq \mathcal{D}(f^*).$$

Let us play with $\mathcal{D}(f^* + \delta f)$:

$$\mathcal{D}(f^* + \delta f) = \int_{\Omega_{AB}} \|\nabla f^* + \delta f\|^2 e^{-\beta V} dx$$

$$= \int_{\Omega_{AB}} e^{-\beta V} \|\nabla f^*\|^2 + 2 e^{-\beta V} \nabla f^* \cdot \nabla \delta f + e^{-\beta V} \|\delta f\|^2 dx$$

$$= \mathcal{D}(f^*) + 2 \int_{\Omega_{AB}} e^{-\beta V} \nabla f^* \cdot \nabla \delta f dx + \int_{\Omega_{AB}} e^{-\beta V} \|\delta f\|^2 dx$$

$$=: \mathcal{D}(f^*) + 2 I_2 + I_3.$$

Integral I_3 in the last expression is nonnegative as it has a nonnegative integrand. Integral I_2 can be of either sign, in principle. Our goal is to prove that, if f^* is a minimizer of $\mathcal{D}(f)$ satisfying

(11) $$f \in C^2(\Omega_{AB}), \quad f(\partial A) = 0, \quad f(\partial B) = 1,$$

then $I_2 = 0$ for any feasible δf, i.e. satisfying $\delta f(x) \in C^1(\Omega_{AB})$, $\delta f(\partial A) = 0$, and $\delta f(\partial B) = 0$. Indeed, if $I_2 \neq 0$ for some feasible δf, we can multiply δf by an arbitrary positive real number α. Then I_3 will be proportional to α^2 while I_2 will be proportional to α. Therefore, if $I_2 \neq 0$, picking α small enough we will be always able to make $2|\alpha I_2| > I_3$. Furthermore, if $I_2 > 0$, we replace δf with $-\delta f$ and hence make I_2 negative. Thus, unless I_2 is zero for all feasible δf, we always take δf for which I_2 is nonzero and pick a real number α such that $\mathcal{D}(f^* + \alpha \delta f)$ will be less than $\mathcal{D}(f^*)$ which contradicts to the fact that f^* is the minimizer.

Let us show that $I_2 = 0$. We would like to switch the gradient operator from δf to $e^{-\beta V} \nabla f^*$. This can be accomplished using *the divergence theorem*:

(12) $$\int_U (\psi \nabla \cdot \Gamma + \nabla \psi \cdot \Gamma) \, dx = \oint_{\partial U} \psi \, (\Gamma \cdot \hat{n}) \, dS,$$

where $\Gamma \in C^1(U)$, $\psi \in C^1(U)$, $\hat{n}$ is the unit outward normal vector for the surface ∂U, and dS is the surface element. We set $\Gamma := e^{-\beta V} \nabla f$, $\psi := \delta f$, $U := \Omega_{AB}$, and apply (12):

$$I_2 = \int_{\Omega_{AB}} \nabla \delta f \cdot e^{-\beta V} \nabla f^* dx$$

(13) $$= - \int_{\Omega_{AB}} \delta f \nabla \cdot \left(e^{-\beta V} \nabla f^* \right) dx + \int_{\partial A \cup \partial B} \delta f \left(e^{-\beta V} \nabla f^* \cdot \hat{n} \right) dS.$$

The integral over $\partial A \cup \partial B$ is equal to zero because δf is zero on ∂A and ∂B. Therefore, in order to make I_2 be equal to zero for any δf, we must have Now, let us scrutinize the integrand of the remaining integral:

$$\nabla \cdot \left(e^{-\beta V} \nabla f^* \right) = \beta e^{-\beta V} \mathcal{L} f^* \equiv 0.$$

Since $\beta e^{-\beta V}$ is never zero, we must have $\mathcal{L} f^* \equiv 0$. Since f^* satisfies the boundary conditions (11) we conclude that f^* is a solution to the BVP (3).

5. The probability density of reactive trajectories

The probability density of reactive trajectories is given by

$$(14) \qquad m_R(x) := \mu(q) q^-(x) q^+(x)$$

The integral of m_R over any region gives the probability to find a reactive trajectory in this region at an arbitrary moment of time. Since $q^- = 0$ in B and $q^+ = 0$ in A, the integral of m_R over the whole space

$$(15) \qquad \int_{\mathbb{R}^d} m_R(x) dx = \int_{\Omega_{AB}} m_R(x) dx = \mathbb{P}(t \in T_R),$$

is equal to the probability that a trajectory is reactive at an arbitrary time t, i.e., is on its way from A to B.

In the case of the overdamped Langevin dynamics (1), the probability density of reactive trajectories becomes:

$$(16) \qquad m_R(x) := Z^{-1} e^{-\beta V} q(x)(1 - q(x)).$$

6. The probability current

In general, *the probability current* is such a vector field J that its flux through any small surface segment dS with a unit outer normal $\hat{n}$ is the difference of the probability to cross dS from inside to outside per unit time and the probability to cross dS from outside to inside per unit time.

The probability current can be found from the forward Kolmogorov equation which governs the time evolution of the probability density. Recall that *the forward Kolmogorov operator* is *the adjoint generator* that can be found by considering the inner product

$$(17) \qquad (\mathcal{L}f, g) = (f, \mathcal{L}^*g), \quad \text{where} \quad (f, g) = \int_{\mathbb{R}^d} f(x) g(x) dx.$$

Let us find the adjoint generator for SDE (1):

$$\begin{aligned} (\mathcal{L}f, g) &= \int_{\mathbb{R}^d} \left[-\nabla V \cdot \nabla f + \beta^{-1} \Delta f \right] g dx \\ &= \int_{\mathbb{R}^d} \left[f \nabla \cdot [g \nabla V] - \beta^{-1} \nabla g \cdot \nabla f \right] dx \\ &= \int_{\mathbb{R}^d} \left[\nabla \cdot [g \nabla V] + \beta^{-1} \Delta g \right] f dx = (f, \mathcal{L}^*g). \end{aligned}$$

Therefore, the adjoint generator for the overdamped Langevin dynamics is given by

$$(18) \qquad \mathcal{L}^*g = \nabla \cdot [g \nabla V] + \beta^{-1} \Delta g.$$

The time evolution of a pdf $\rho(x)$ for SDE (1) is governed by

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \mathcal{L}^*g = \nabla \cdot [\rho \nabla V] + \beta^{-1} \Delta \rho \\ (19) \qquad &\equiv -\nabla \cdot \left[-\rho \nabla V - \beta^{-1} \nabla \rho \right] =: -\nabla \cdot J. \end{aligned}$$

The vector field J defined in (19) is the probability current. Indeed, according to the *divergence theorem*, the integral over a region U of divergence of a vector field Γ is equal to the flux of that vector field through the boundary of that region:

$$(20) \qquad \int_U \nabla \cdot \Gamma dx = \oint_{\partial U} (\Gamma \cdot \hat{n}) \, dS,$$

where $\hat{n}$ is the outer unit normal to ∂U. Returning to (19), we observe that the time derivative of the probability to find a system in a region U is negative of the total outflux of the probability current from U which matches its physical sense declared at the beginning of this subsection:

$$(21) \qquad \frac{\partial}{\partial t} \int_U \rho(x) dx = -\int_U \nabla \cdot J dx = -\oint_{\partial U} (J \cdot \hat{n}) \, dS.$$

It is easy to check that the Gibbs density $\mu(x) = Z^{-1} e^{-\beta V}$ is the invariant pdf for (1). Indeed,

$$\mathcal{L}^* \mu = \nabla \cdot \left[Z^{-1} e^{-\beta V} \nabla V + \beta^{-1} \nabla Z^{-1} \nabla e^{-\beta V} \right]$$

$$= Z^{-1} \nabla \cdot \left[e^{-\beta V} \nabla V - e^{-\beta V} \nabla V \right] = \nabla \cdot 0 = 0.$$

This calculation shows that not only the divergence of the probability current associated with the Gibbs density is zero, but also *the stationary probability current is zero*. The last fact is a manifestation of time-reversibility.

7. The reactive current

The reactive current is the probability current in which only the crossings of a surface segment dS by those trajectories that last visited A rather than B are counted. We do not need to demand that these trajectories go to B next because if they return to A, they will cross any dividing surface separating A and B in both directions an equal number of times and hence will not contribute to the transition rate from A to B.

Let us consider the following process. The system evolves according to SDE (1) in Ω_{AB}. As soon as its trajectory hits ∂B, it disappears and reappears at the boundary of ∂A according to the invariant pdf $Z^{-1} e^{-\beta V}$ restricted to ∂A. The dynamics of these trajectories are still governed by the original SDE (1). Therefore, the time evolution of the pdf for this process satisfies the following PDE with boundary conditions:

$$(22) \qquad \begin{cases} \frac{\partial \hat{\mu}}{\partial t} = \nabla \cdot \left[\hat{\mu} \nabla V + \beta^{-1} \nabla \hat{\mu} \right], & x \in \Omega_{AB} \\ \hat{\mu} = \mu, & x \in \partial A, \\ \hat{\mu} = 0, & x \in \partial B. \end{cases}$$

Let us show that $\hat{\mu} := \mu(x) q^-(x) \equiv Z^{-1} e^{-\beta V} (1 - q)$, where $(1 - q)$ is the backward committor, is the invariant pdf for this process. Indeed, $\hat{\mu}$ satisfies the boundary conditions

in (22). Plugging it into the right-hand side of the PDE in (22) we get:

$$\mathcal{L}^*\hat{\mu} = \nabla \cdot \left[\hat{\mu}\nabla V + \beta^{-1}\nabla\hat{\mu}\right]$$

$$(23) \qquad = \nabla \cdot \left[q^-\mu\nabla V + \beta^{-1}\mu\nabla q^- - q^-\mu\nabla V\right] = -\nabla \cdot \left[-\beta^{-1}\mu\nabla q^-\right]$$

The last expression gives us the probability current associated with the transition path process:

$$(24) \qquad J_{AB} := -\beta^{-1}\mu\nabla q^-$$

Let us show that its divergence is indeed zero:

$$(25) \qquad \mathcal{L}^*\hat{\mu} = -\nabla \cdot \left[-\beta^{-1}\mu\nabla q^-\right] = \mu\nabla V \cdot \nabla q^- + \beta^{-1}\mu\Delta q^-$$

$$(26) \qquad = \mu\left[-\nabla V \cdot \nabla q + \beta^{-1}\Delta q\right] = \mu\mathcal{L}q = 0.$$

Here we have used the facts that $q^- = 1 - q$ and $\mathcal{L}q = 0$ in Ω_{AB}.

Therefore, the reactive current for SDE (1) is given by

$$(27) \qquad \boxed{J_{AB} := \beta^{-1}Z^{-1}e^{-\beta V}\nabla q.}$$

8. THE TRANSITION RATE

The transition rate ν_{AB} from A to B is the average number of transitions from A to B per unit time:

$$(28) \qquad \nu_{AB} = \lim_{t\to\infty}\frac{1}{t}N_{AB}(t),$$

where N_{AB} is the number of transitions from A to B accomplished within the time interval $[0, t]$. It follows from the definition of the reactive current that the transition rate ν_{AB} is the total flux of the probability current through any surface S that separates the sets A and B:

$$(29) \qquad \nu_{AB} = \int_S J_{AB} \cdot \hat{n}dS,$$

where $\hat{n}$ is the unit normal to the surface S pointing to the side that contains set B. The surface S can be arbitrary as the divergence of the reactive current is zero. Indeed, let us consider a union of two different dividing surfaces S_1 and S_2. Assume $S_1 \cap S_2 = \emptyset$. We pick the unit normals $\hat{n}_1$ and $\hat{n}_2$ for S_1 and S_2 respectively so that $\hat{n}_1$ points to the side containing A while $\hat{n}_2$ points to the side containing B. Then we consider the integral of $\nabla \cdot J_{AB}$ over the region $D(S_1, S_2)$ bounded by these surfaces. Clearly, this integral is zero because $\nabla \cdot J_{AB} \equiv 0$. On the other hand, by the divergence theorem,

$$0 = \int_{D(S_1,S_2)} \nabla \cdot J_{AB}dx = \oint_{S_1 \cup S_2} J_{AB} \cdot \hat{n}dS = \int_{S_1} J_{AB} \cdot \hat{n}_1 dS + \int_{S_2} J_{AB} \cdot \hat{n}_2 dS.$$

Hence the last two integrals are equal in their absolute values. Flipping $\hat{n}_2$ to point to the side containing B we find that these two integrals are equal and conclude that the flux of the reactive current is the same for any dividing surface.

Since the flux of the reactive current is the same through any dividing surface, we can pick an isocommittor surface as a dividing surface, i.e.

$$(30) \qquad S_{q^*} = \left\{ x \in \mathbb{R}^d \mid q(x) = q^* \right\}, \quad \text{where} \quad 0 < q^* < 1.$$

The isocommittor surfaces foliate the domain Ω_{AB}. Noting the trivial fact that

$$\int_0^1 dq^* = 1,$$

and that the unit normal to an isocommittor surface is given by

$$\hat{n}(S_{q^*}) = \frac{\nabla q}{\|\nabla q\|},$$

we can express the transition rate as the integral over the whole region Ω_{AB}:

$$\nu_{AB} = \int_0^1 dq^* \int_{S_{q^*}} J_{AB} \cdot \hat{n} dS$$

$$= \beta^{-1} Z^{-1} \int_0^1 dq^* \int_{S_{q^*}} e^{-\beta V} \nabla q \cdot \frac{\nabla q}{\|\nabla q\|} dS$$

$$(31) \qquad = \beta^{-1} Z^{-1} \int_0^1 dq^* \int_{S_{q^*}} e^{-\beta V} \|\nabla q\| dS$$

Next, we observe that the volume element

$$(32) \qquad dq^* dS = \|\nabla q\| dx.$$

Indeed, let z be a coordinate normal to the isocommittor surface with a positive direction along the gradient of the committor. Then $dq = \|\nabla q\| dz$ which implies (32). Using (32), we obtain the following expession for the transition rate:

$$(33) \qquad \boxed{\nu_{AB} = Z^{-1} \beta^{-1} \int_{\Omega_{AB}} e^{-\beta V} \|\nabla q\|^2 dx.}$$

REFERENCES

[1] W. E and E. Vanden-Eijnden, "Towards a theory of transition paths," *Journal of statistical physics*, vol. 123, no. 3, p. 503, 2006.

[2] W. E and E. Vanden-Eijnden, "Transition-path theory and path-finding algorithms for the study of rare events.," *Annual review of physical chemistry*, vol. 61, pp. 391–420, 2010.