Transition Path Theory

E. Vanden-Eijnden

Courant Institute of Mathematical Sciences, New York University New York, NY 10012
eve2cims.nyu.edu

Eric Vanden-Eijnden

E. Vanden-Eijnden: *Transition Path Theory*, Lect. Notes Phys. **703**, 439–478 (2006)
DOI 10.1007/3-540-35273-2_13

1 Introduction

The dynamical behavior of many systems arising in physics, chemistry, biology, etc. is dominated by rare but important transition events between long lived states. For over 70 years, transition state theory (TST) has provided the main theoretical framework for the description of these events [17,33,34]. Yet, while TST and evolutions thereof based on the reactive flux formalism [1, 5] (see also [30,31]) give an accurate estimate of the transition rate of a reaction, at least in principle, the theory tells very little in terms of the *mechanism* of this reaction. Recent advances, such as transition path sampling (TPS) of Bolhuis, Chandler, Dellago, and Geissler [3, 7] or the action method of Elber [15, 16], may seem to go beyond TST in that respect: these techniques allow indeed to sample the ensemble of reactive trajectories, i.e. the trajectories by which the reaction occurs. And yet, the reactive trajectories may again be rather uninformative about the mechanism of the reaction. This may sound paradoxical at first: what more than actual reactive trajectories could one need to understand a reaction? The problem, however, is that the reactive trajectories by themselves give only a very indirect information about the *statistical properties* of these trajectories. This is similar to why statistical mechanics is not simply a footnote in books about classical mechanics. What is the probability density that a trajectory be at a given location in state-space conditional on it being reactive? What is the probability current of these reactive trajectories? What is their rate of appearance? These are the questions of interest and they are not easy to answer directly from the ensemble of reactive trajectories. The right framework to tackle these questions also goes beyond standard equilibrium statistical mechanics because of the nontrivial bias that the very definition of the reactive trajectories imply – they must be involved in a reaction. The aim of this chapter is to introduce the reader to the probabilistic framework one can use to characterize the mechanism of a reaction and obtain the probability density, current, rate, etc. of the reactive trajectories.

Since our results are rather general, it is useful to set the stage somewhat abstractly. We shall consider a system whose state-space is $\Omega \subseteq \mathbb{R}^n$ and denote by $x(t)$ the current state of the system in Ω at time t. For instance, $x(t)$ may be the set of instantaneous positions and momenta of the atoms of a molecular system. We assume that the system is ergodic (in the sense of (3) below) and that we have observed an infinitely long equilibrium trajectory, $\{x(t) : -\infty < t < \infty\}$. Out of this observation, we wish to understand the mechanism of reaction between a *reactant state* A and a *product state* B, here thought of as two disjoint regions in Ω, $A \subset \Omega$ and $B \subset \Omega$ with $A \cap B = \emptyset$. (Of course our final aim is to bypass the need of actually observing $\{x(t) : -\infty < t < \infty\}$ and find more efficient ways to understand the reaction, but it is useful to formulate the questions this way.) Since the system is ergodic, the trajectory $\{x(t) : -\infty < t < \infty\}$ will go infinitely many times from A to B, and each time the reaction happens. This reaction involves *reactive trajectories* defined

as follows: given the trajectory $\{x(t) : -\infty < t < \infty\}$, we say that the reactive pieces of it are those pieces during which $x(t)$ is out of both A and B and such that it came out of A last and will go to B next. To formalize things, given a trajectory $\{x(t) : -\infty < t < \infty\}$, let

$$\begin{aligned} t^+_{AB}(t) &= \text{smallest } t' \geq t \text{ such that } x(t') \in A \cup B \\ t^-_{AB}(t) &= \text{largest } t' \leq t \text{ such that } x(t') \in A \cup B \end{aligned} \tag{1}$$

Then

$$\begin{gathered} \text{ensemble of reactive trajectories} = \{x(t) : t \in R\} \\ \text{where } t \in R \text{ if and only if } x(t) \notin A \cup B,\ x(t^+_{AB}(t)) \in B \text{ and } x(t^-_{AB}(t)) \in A \end{gathered} \tag{2}$$

and each continuous piece of trajectory going from A to B in the ensemble (2) is a specific reactive trajectory (see Fig. 1).

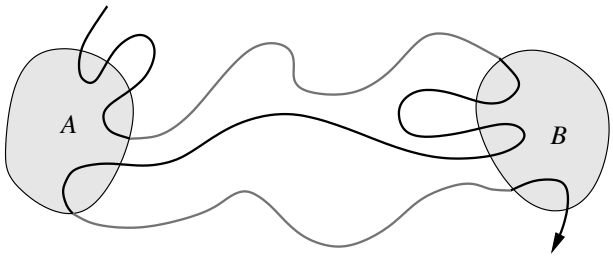

Fig. 1. Schematic representation of the reactant state A, the product state B, a piece of an equilibrium trajectory (shown in black) and the two reactive trajectories along it (shown in light gray)

Our objective is to understand the statistical properties of the reactive trajectories in the ensemble (2). We will try to do so under minimum assumptions about the dynamics of $x(t)$, but we have to require the following from the start. First we require that the dynamics be *Markov*, i.e. given $x(t)$, its future $\{x(t') : t' > t\}$ and its past $\{x(t') : t' < t\}$ are statistically determined. Second, that it be *ergodic* with respect to some equilibrium probability density $m(x)$, i.e. given a suitable observable $F(x)$ and a generic trajectory $\{x(t) : -\infty < t < \infty\}$, we have

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} F(x(t))dt = \int_{\Omega} F(x)m(x)dx \tag{3}$$

Ergodicity guarantees that the trajectory oscillates infinitely often between A and B, so that the number of reactive trajectories in the ensemble (2) is infinite and this ensemble is statistically independent of which particular trajectory $\{x(t) : -\infty < t < \infty\}$ one chooses to generate it. In other words,

we try to understand the generic way the reaction happens at equilibrium, by opposition to the way it may happen once if one prepares the system in some specific initial state.

We then ask the following:

1. What is the *probability density function* of reactive trajectories, i.e. the probability density $m_{AB}(x)$ that the trajectory be at point $x \in \Omega \setminus (A \cup B)$ at time t conditional on this trajectory being reactive at time t? We stress that $m_{AB}(x)$ is not simply $m(x)$ properly renormalized on $\Omega \setminus (A \cup B)$ since a trajectory may be in $\Omega \setminus (A \cup B)$ by leaving B, or by leaving A and re-entering this set afterwards without visiting B in between (in both cases it is not considered as reactive). From (2) $x(t)$ belongs to the set of reactive trajectories if and only if $\chi_{\Omega\setminus(A\cup B)}(x(t))\chi_B(x(t^+_{AB}(t)))\chi_A(x(t^-_{AB}(t))) = 1$ (here and below, given any set C, $\chi_C(x) = 1$ if $x \in C$ and $\chi_C(x) = 0$ otherwise). Therefore, by ergodicity, $m_{AB}(x)$ can be defined as the probability density such that for any suitable function $F(x)$, we have (compare (3))

$$\begin{aligned}\lim_{T\to\infty} \frac{\int_{-T}^{T} F(x(t))\chi_{\Omega\setminus(A\cup B)}(x(t))\chi_B(x(t^+_{AB}(t)))\chi_A(x(t^-_{AB}(t)))dt}{\int_{-T}^{T} \chi_{\Omega\setminus(A\cup B)}(x(t))\chi_B(x(t^+_{AB}(t)))\chi_A(x(t^-_{AB}(t)))dt} \\ = \int_{\Omega\setminus(A\cup B)} F(x)m_{AB}(x)dx\end{aligned} \tag{4}$$

2. What is the *probability current* of reactive trajectories? This probability current is the vector field $J_{AB}(x)$ defined in $\Omega \setminus (A \cup B)$ which is such that, given any surface $S \subset \Omega \setminus (A \cup B)$ which is the boundary of a region Ω_S, the surface integral of $J_{AB}(x)$ over S gives the probability flux of reactive trajectories across S. More precisely,

$$\begin{aligned}\lim_{s\to 0^+} \frac{1}{s} \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \big(\chi_{\Omega_S}(x(t))\chi_{\Omega\setminus\Omega_S}(x(t+s)) \\ - \chi_{\Omega\setminus\Omega_S}(x(t))\chi_{\Omega_S}(x(t+s))\big) \\ \times \chi_A(x(t^-_{AB}(t)))\chi_B(x(t^+_{AB}(t+s)))dt \\ = \int_S \hat{n}_S(x) \cdot J_{AB}(x)d\sigma_S(x)\end{aligned} \tag{5}$$

where $\hat{n}_S(x)$ is the unit normal on S pointing outward Ω_S and $d\sigma_S(x)$ is the surface element on S. The integral at the left hand-side of (5) counts the balance of how many reactive trajectories go in and out of Ω_S in the infinitesimal interval $[t, t+s]$ and averages this count over time. Therefore the current $J_{AB}(x)$ is the true indicator of the mechanism of the reaction since, roughly, it measures the average flow of the reactive trajectories at a given point $x \in \Omega \setminus (A \cup B)$: in particular, the streamlines of $J_{AB}(x)$ indicate the average pathway of the reaction as they allow to define regions (or tubes) joining A to B which contribute to a specific percentage of the reaction.

3. What is the *rate* of the reaction, i.e. what is the mean frequency k_{AB} of transitions from A to B? If N_T^R is the number of reactive trajectories observed during the time interval $[-T, T]$ in the ensemble (2), the rate is the limit

$$k_{AB} = \lim_{T\to\infty} \frac{N_T^R}{2T}. \tag{6}$$

The answers to these questions and some extra ones are given below. This chapter is a summary of the material presented originally in [9–14, 26–28].

2 Probability Density of Reactive Trajectories

What is the probability density $m_{AB}(x)$ that a trajectory be at point $x \in \Omega \setminus (A \cup B)$ at time t conditional on it being reactive at time t? To answer this question, let us first derive the probability density $m_R(x)$ that there be a reactive trajectory at point x at time t; in terms of the trajectory $\{x(t) : -\infty < t < \infty\}$, $m_R(x)$ can be defined as the density such that for any suitable function $F(x)$ we have

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} F(x(t))\chi_{\Omega\setminus(A\cup B)}(x(t))\chi_B(x(t_{AB}^+(t)))\chi_A(x(t_{AB}^-(t)))dt \\ = \int_{\Omega\setminus(A\cup B)} F(x)m_R(x)dx\end{aligned} \tag{7}$$

Observe that this expression involves the integral at numerator in (4) divided by $2T$ in the limit as $T \to \infty$. Using Markovianity, $m_R(x)$ is the probability density that any trajectory (reactive or not) be at x at time t, times the probability $P_R(x)$ that the trajectory be reactive (i.e. that it came from A last and will go to B next) conditional on it being at x at time t. The probability $P_R(x)$ can be expressed in terms of the forward and backward *committor functions*, defined as follows. Given $x \in \Omega \setminus (A \cup B)$, the forward committor function is

$$\begin{aligned}q_+(x) = &\text{ probability} \\ &\text{that } x(t) \text{ goes next to } B \text{ rather than } A \text{ given that } x(0) = x\end{aligned} \tag{8}$$

whereas the backward commitor function is

$$\begin{aligned}q_-(x) = &\text{ probability} \\ &\text{that } x(t) \text{ came last from } A \text{ rather than } B \text{ given that } x(0) = x\end{aligned} \tag{9}$$

Note that since the dynamics need not be time-reversible, $q_-(x) \neq 1 - q_+(x)$ in general.[1] In terms of $q_+(x)$ and $q_-(x)$, we simply have $P_R(x) = q_+(x)q_-(x)$

[1] $q_+(x) = 1 - q_-(x)$ if and only if the dynamics is time-reversible, which is here understood in the following sense. By definition, time-reversible Markov processes

and therefore we deduce that the probability density that there be a reactive trajectory at point x at time t is

$$m_R(x) = q_+(x)q_-(x)m(x) \tag{10}$$

Note that the integral of $m_R(x)$ over $\Omega \setminus (A \cup B)$ is in general less than one since it gives the probability the trajectory be reactive at time t:

$$\begin{aligned} Z_{AB} &= \int_{\Omega\setminus(A\cup B)} q_+(x)q_-(x)m(x)dx \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \chi_{\Omega\setminus(A\cup B)}(x(t))\chi_B(x(t_{AB}^+(t)))\chi_A(x(t_{AB}^-(t)))dt \end{aligned} \tag{11}$$

which is also the limit as $T \to \infty$ of the integral at denominator in (4) divided by $2T$. Therefore Z_{AB} is the factor by which $m_R(x)$ needs to be divided to account for the extra condition that the trajectory be reactive at time t and answers our original question about $m_{AB}(x)$:

$$m_{AB}(x) = Z_{AB}^{-1} m_R(x) = Z_{AB}^{-1} q_+(x)q_-(x)m(x) \ , \tag{12}$$

This relation between $m_{AB}(x)$, $Z_{AB}(x)$ and $m_R(x)$ as well as the expression (10) for $m_R(x)$ in a simpler setting were first derived in [20].

Since the dynamics is Markov by assumption, the committor functions $q_+(x)$, $q_-(x)$ and hence $m_R(x)$ and $m_{AB}(x)$ from (10) and (12) are well-defined. However, these functions could be quite nasty. For instance, if the dynamics is deterministic (say, Hamiltonian), $q_+(x)$ and $q_-(x)$ are either 1 or 0, depending on whether the trajectory passing through x goes next to B or not, and came last from A or not. So, to avoid pathologies, we will assume that the dynamics is not deterministic and $q_+(x)$ and $q_-(x)$ are smooth functions taking value in $[0, 1]$. Later, we will give conditions under which this assumption holds true and give closed equations for $q_+(x)$ and $q_-(x)$ (see (23) and (24) below).

The density $m_{AB}(x)$ gives a first interesting indicator of the mechanism of the reaction since it allows e.g. to determine the proportion of time that the trajectory spends in any region $C \subseteq \Omega \setminus (A \cup B)$ while it is reactive. This proportion of time is simply given by (4) with $F(x) = \chi_C(x)$:

$$\int_C m_{AB}(x)dx \tag{13}$$

---

are such that, at equilibrium, the joint probability density that the system be at x at time t and at y at time $t + s$ with $s > 0$ is the same as the joint probability density that the system be at y at time t and at x at time $t + s$. Note that this definition is the standard one in stochastic processes theory, but it is different from what time-reversibility means in the context of Hamiltonian systems. In particular, an Hamiltonian system is *not* time-reversible with our definition, since one is not allowed to revert the directions of the momenta.

Since the reactive trajectories will in general slow down near the dynamical bottlenecks of the reaction, this allows one to identify the *transition state regions* roughly as the regions where $m_{AB}(x)$ is peaked. Observe however that these regions can be multiple (i.e. there may be more than one dynamical bottleneck for a reaction) and quite wide (i.e. the dynamical bottleneck may be a rather extended region in state-space).

Quite interestingly, the above argument gives also the rate of the reaction k_{AB} in terms of Z_{AB} and the *mean reaction time* t_{AB} defined as

$$t_{AB} = \lim_{T\to\infty} \frac{1}{N_T^R} \int_{-T}^{T} \chi_{\Omega\setminus(A\cup B)}(x(t))\chi_B(x(t_{AB}^+(t)))\chi_A(x(t_{AB}^-(t)))dt \qquad (14)$$

Indeed, letting

$$T_T^R = \int_{-T}^{T} \chi_{\Omega\setminus(A\cup B)}(x(t))\chi_B(x(t_{AB}^+(t)))\chi_A(x(t_{AB}^-(t)))dt \qquad (15)$$

be the total time the trajectory is reactive in $[-T, T]$ and combining (6), (11) and (14), one sees that

$$\begin{aligned} k_{AB} &= \lim_{T\to\infty} \frac{N_T^R}{2T} = \lim_{T\to\infty} \frac{T_T^R}{2T}\frac{N_T^R}{T_T^R} \\ &= \frac{\lim_{T\to\infty} T_T^R/2T}{\lim_{T\to\infty} T_T^R/N_T^R} = Z_{AB}/t_{AB} \end{aligned} \qquad (16)$$

An alternative to that expression will be given below (see (43)) in terms of $m(x)$, $q_+(x)$ and $q_-(x)$. Note that (16) is not fully explicit since we have not expressed t_{AB} in terms of $m(x)$, $q_+(x)$ and $q_-(x)$ (this will also be done below, see (44)). A formula equivalent to (16) was first derived in [20].

Finally, notice that $m_{AB}(x)$ restricted to any dividing surface S and properly normalized to this surface gives the probability density that a reactive trajectory crosses S at point $x \in S$. In other words, if t_j^S with $j \in \mathbb{Z}$ is the set of all times at which the reactive trajectories in the ensemble (2) cross S, i.e.

$$\{t_j^S : j \in \mathbb{Z}\} \text{ is the set of all times such that } x(t_j^S) \in S \text{ and } t_j^S \in R \qquad (17)$$

then for any suitable function $F(x)$

$$\lim_{N\to\infty} \frac{1}{2N} \sum_{j=-N}^{N} F(x(t_j^S)) = Z_S^{-1} \int_S F(x)q_+(x)q_-(x)m(x)d\sigma_S(x) \qquad (18)$$

where $Z_S = \int_S q_+(x)q_-(x)m(x)d\sigma_S(x)$ and $d\sigma_S(x)$ is the surface element on S.[2]

[2] Formula (17) and (18) assume that we can count the crossing times t_j^S which is not possible in general if the dynamics of $x(t)$ is governed by a stochastic differential equation such as (19). However, (17) and (18) can be generalized to this case, and the claim remains valid: $m_{AB}(x)$ restricted to any dividing surface S and properly normalized to this surface gives the probability density that the reactive trajectories cross S at point $x \in S$.

3 Probability Current of Reactive Trajectories

To proceed further, we must make some additional assumptions about the equation of motion governing the evolution of $x(t)$. We shall suppose that $x(t)$ satisfies the following *stochastic differential equation*

$$\dot{x}(t) = b(x(t)) + \sqrt{2}\,\sigma(x(t))\,\eta(t) \tag{19}$$

where $b(x) = (b_1(x), \ldots, b_n(x))^T \in \mathbb{R}^n$ is the drift vector,

$$\sigma(x) = \begin{pmatrix} \sigma_{11}(x) & \cdots & \sigma_{1n}(x) \\ \vdots & \ddots & \vdots \\ \sigma_{n1}(x) & \cdots & \sigma_{nn}(x) \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n \tag{20}$$

is the square root of the diffusion tensor, and $\eta(t) = (\eta_1(t), \ldots, \eta_n(t))^T \in \mathbb{R}^n$ is an n dimensional white-noise, i.e. a Gaussian process with mean 0 and covariance $\langle \eta_i(t)\eta_j(s)\rangle = \delta_{ij}\delta(t-s)$. An important example of (19) which arises as a special case of this equation is the *Langevin equation*

$$\begin{cases} \dot{r}_i = v_i \\ \mu_i \dot{v}_i = -\dfrac{\partial V(r)}{\partial r_i} - \gamma v_i + \sqrt{2k_B T\gamma}\,\eta_i(t) \end{cases} \tag{21}$$

where $(r, v) = (r_1, \ldots, r_m, v_1, \ldots, v_m)^T \in \mathbb{R}^m \times \mathbb{R}^m$ are the position and velocities of the particles, $V(r)$ is the potential, $-\nabla V = -(\partial V/\partial r_1, \ldots, \partial V/\partial r_m)^T$ is the force, γ is the friction coefficient and μ_i is the mass of the particle. (21) can be put in the form of (19) by writing $x = (r, v)^T$, $b(x) = (v, -\mu^{-1}\nabla V(r) - \gamma\mu^{-1}v)^T$, etc.

We also need some background material about (19). If $m(x)$ denotes the *equilibrium probability density function* of $x(t)$, i.e. the probability density to find a trajectory (reactive or not) at position x at time t, $m(x)$ satisfies the (steady) *forward Kolmogorov equation* (also known as Fokker-Planck equation)

$$0 = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(b_i(x)m(x)) + \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}(x)m(x)) \tag{22}$$

where $a(x) = a^T(x) = \sigma(x)\sigma^T(x) \in \mathbb{R}^n \times \mathbb{R}^n$ is the nonnegative-definite diffusion tensor (to avoid the problems with the deterministic dynamics discussed in Sect. 2, we assume that at least some of the entries $a_{ij}(x)$ are nonzero). If (22) has a unique solution such that $\int_\Omega m(x)dx = 1$, then the process defined by (19) is ergodic, i.e. it satisfies (3). In addition, it can be shown that the forward committor function $q_+(x)$ satisfies the *backward Kolmogorov equation* associated with (19):

$$\begin{cases} 0 = \displaystyle\sum_{i=1}^{n} b_i(x)\frac{\partial q_+(x)}{\partial x_i} + \sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2 q_+(x)}{\partial x_i \partial x_j} \\ q_+(x)|_{x\in\partial A} = 0, \qquad q_+(x)|_{x\in\partial B} = 1, \end{cases} \tag{23}$$

whereas the backward committor function $q_-(x)$ satisfies the backward Kolmogorov equation associated with the time-reversed process:

$$\begin{cases} 0 = \sum_{i=1}^n b_i^R(x) \dfrac{\partial q_-(x)}{\partial x_i} + \sum_{i,j=1}^n a_{ij}(x) \dfrac{\partial^2 q_-(x)}{\partial x_i \partial x_j} \\ q_-(x)|_{x\in\partial A} = 1, \qquad q_-(x)|_{x\in\partial B} = 0 \end{cases} \tag{24}$$

where

$$b_i^R(x) = -b_i(x) + \frac{2}{m(x)} \sum_{j=1}^n \frac{\partial}{\partial x_i}(a_{ij}(x) m(x)) \tag{25}$$

The operator acting on $q_+(x)$ at the right hand-side of (23) is called the *generator* of the process, and it plays a very important role. In particular, for any suitable $F(x)$ we have

$$\lim_{t\to 0^+} \frac{1}{t}\big(\mathbf{E}_x F(x(t)) - F(x)\big) = \sum_{i=1}^n b_i(x) \frac{\partial F(x)}{\partial x_i} + \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 F(x)}{\partial x_i \partial x_j} \tag{26}$$

where $x(t)$ denotes the solution of (19) in a given realization of the white-noise $\eta(t)$ and $\mathbf{E}_x$ denotes the expectation conditional on $x(0) = x$ over the ensemble of solutions generated with different realizations of the white-noise The first term at the right hand side of (26) is similar to the one one would get for the ordinary differential equation $\dot{x}(t) = b(x(t))$ by chain rule: $\dot{F}(x(t)) = \sum_{i=1}^n \dot{x}_i(t) \partial F(x(t))/\partial x_i = \sum_{i=1}^n b_i(x(t)) \partial F(x(t))/\partial x_i$. The second term at the right hand-side of (26) arises because of the presence of the white-noise in (19). The derivation of (23), (24) and (26) from the definition of $q_+(x)$, $q_-(x)$ and $x(t)$ is beyond the scope of the present chapter but it can be found in any elementary textbook on probability and stochastic processes theory like e.g. chapters in 4 and 5 in [8] (another accessible reference is [18]).

Going back to the problem of the probability current of reactive trajectories, notice first that (22) can be written in the form of a continuity equation,

$$0 = -\sum_{i=1}^n \frac{\partial J_i(x)}{\partial x_i} \tag{27}$$

where $J(x) = (J_1(x), \ldots, J_n(x))^T \in \mathbb{R}^n$ is the equilibrium *probability current*

$$J_i(x) = b_i(x) m(x) - \sum_{j=1}^n \frac{\partial}{\partial x_j}(a_{ij}(x) m(x)) \tag{28}$$

Of course, this current accounts for what the trajectory does, irrespective of whether it is reactive or not. In particular, if Ω_S is the region enclosed by the surface S, then the equilibrium current $J(x)$ is such that (compare (5))

$$\lim_{s\to 0^+}\frac{1}{s}\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\big(\chi_{\Omega_S}(x(t))\chi_{\Omega\setminus\Omega_S}(x(t+s)) - \chi_{\Omega\setminus\Omega_S}(x(t))\chi_{\Omega_S}(x(t+s))\big)dt = \int_S \hat{n}_S(x)\cdot J(x)d\sigma_S(x) = 0 \tag{29}$$

where $\hat{n}_S(x)$ is the unit normal on S pointing outward Ω_S and $d\sigma_S(x)$ is the surface element on S (below it is actually shown how to derive (28) from (29)). This is not what we want. For instance, the equilibrium current (28) is identically zero for processes that are time-reversible in the sense of the footnote on page 444.[3] But even for these processes, there must be a probability current of reactive trajectories since, by construction, these trajectories flow from A to B.

To determine what is the probability current of reactive trajectories let us go back to the definition (5). Taking the limit as $T\to\infty$ in this expression using ergodicity gives

$$\lim_{s\to 0^+}\frac{1}{s}\Big(\int_{\Omega_S} m(x)q_-(x)\mathbf{E}_x\big(q_+(x(s))\chi_{\Omega\setminus\Omega_S}(x(s))\big)dx - \int_{\Omega\setminus\Omega_S} m(x)q_-(x)\mathbf{E}_x\big(q_+(x(s))\chi_{\Omega_S}(x(s))\big)dx\Big) = \int_S \hat{n}_S(x)\cdot J_{AB}(x)d\sigma_S(x) \tag{30}$$

where $\mathbf{E}_x$ denotes expectation conditional on $x(0)=x$ as in (26). Taking the limit as $s\to 0^+$ can now be done using (26) but this operation is somewhat tricky because of the presence of the discontinuous functions $\chi_{\Omega_S}(x)$ and $\chi_{\Omega\setminus\Omega_S}(x)$ over which the generator in (26) arising in the limit must act. The spatial derivatives on $\chi_{\Omega_S}(x)$ and $\chi_{\Omega\setminus\Omega_S}(x)$ bring delta distributions concentrated on the boundary S of Ω_S, and so intuitively in the limit as $s\to 0^+$, the left hand-side of (30) should reduce to a surface integral over S. This is indeed the case, but since the argument leading to this conclusion is somewhat technical we defer it till the end of this section. The result is that (30) gives

$$\int_S \sum_{i=1}^n \hat{n}_{S,i}(x)\Big(q_-(x)q_+(x)J_i(x) + q_-(x)m(x)\sum_{j=1}^n a_{ij}(x)\frac{\partial q_+(x)}{\partial x_j} - q_+(x)m(x)\sum_{j=1}^n a_{ij}(x)\frac{\partial q_-(x)}{\partial x_j}\Big)d\sigma_S(x) = \int_S \hat{n}_S(x)\cdot J_{AB}(x)d\sigma_S(x) \tag{31}$$

[3] For processes satisfying a stochastic differential equation such as (19), it can be shown that the time-reversibility is precisely equivalent to the *detailed balance condition* $J(x)=0$ which is, in fact, a constraint on $b(x)$ and $a(x)$ since $m(x)$ is determined by these coefficients through (22).

where $J(x)$ is the probability current (28). Since this relation must hold for any surface S, we deduce that the *probability current of reactive trajectories* is

$$J_{AB,i}(x) = q_-(x)q_+(x)J_i(x) + q_-(x)m(x)\sum_{j=1}^n a_{ij}(x)\frac{\partial q_+(x)}{\partial x_j} - q_+(x)m(x)\sum_{j=1}^n a_{ij}(x)\frac{\partial q_-(x)}{\partial x_j} \tag{32}$$

Derivation of (31).

The proper way to take the limit as $s \to 0^+$ and avoid ambiguities on how to interpret the derivatives of $\chi_{\Omega_S}(x)$ and $\chi_{\Omega\setminus\Omega_S}(x)$ is to *mollify* these functions, that is, replace them by functions varying rapidly on S but smooth, then let $s \to 0^+$ and finally remove the mollification. Let then $f_\delta(x)$ be a smooth function which is 1 in Ω_S at a distance δ from S, 0 out of Ω_S at a distance δ from S and varies rapidly but smoothly from 0 to 1 in the strip of size 2δ around S. Thus (30) is the limit as $\delta \to 0$ of

$$I_\delta = \lim_{s\to 0^+}\frac{1}{s}\int_\Omega \Big(m(x)q_-(x)f_\delta(x)\mathbf{E}_x\big(q_+(x(s))(1-f_\delta(x(s)))\big) - m(x)q_-(x)(1-f_\delta(x))\mathbf{E}_x\big(q_+(x(s))f_\delta(x(s))\big)\Big)dx \tag{33}$$

Inserting

$$0 = -m(x)q_-(x)f_\delta(x)\big(q_+(x)(1-f_\delta(x))\big) + m(x)q_-(x)(1-f_\delta(x))\big(q_+(x)f_\delta(x)\big)$$

under the integral then letting $s \to 0^+$ using (26), (33) gives

$$I_\delta = \int_\Omega \Big(m(x)q_-(x)f_\delta(x)L\big(q_+(x)(1-f_\delta(x))\big) - m(x)q_-(x)(1-f_\delta(x))L\big(q_+(x)f_\delta(x)\big)\Big)dx \tag{34}$$

where L is a short hand notation for the generator defined in (26): for any suitable $F(x)$

$$LF(x) = \sum_{i=1}^n b_i(x)\frac{\partial F(x)}{\partial x_i} + \sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 F(x)}{\partial x_i \partial x_j} \tag{35}$$

Expanding the integrand in (34), several terms cancel and we are simply left with

$$I_\delta = -\int_\Omega m(x)q_-(x)L\big(q_+(x)f_\delta(x)\big)dx \tag{36}$$

Using the explicit form (35) for L and expanding, this is

$$I_\delta = -\int_\Omega m(x)q_-(x)\Bigg(f_\delta(x)Lq_+(x) + \sum_{i,j=1}^n a_{ij}(x)\frac{\partial}{\partial x_i}\Big(q_+(x)\frac{\partial f_\delta(x)}{\partial x_j}\Big) + \sum_{i=1}^n \frac{\partial f_\delta(x)}{\partial x_i}\Big(b_i(x)q_+(x) + \sum_{j=1}^n a_{ij}(x)\frac{\partial q_+(x)}{\partial x_j}\Big)\Bigg)dx \tag{37}$$

By (23), $Lq_+(x) = 0$ and integrating by parts the second term in the parenthesis under the integral, we arrive at

$$I_\delta = -\int_\Omega \sum_{i=1}^n \frac{\partial f_\delta(x)}{\partial x_i}\Big(q_+(x)q_-(x)J_i(x) + q_-(x)m(x)\sum_{j=1}^n a_{ij}(x)\frac{\partial q_+(x)}{\partial x_j} - q_+(x)m(x)\sum_{j=1}^n a_{ij}(x)\frac{\partial q_-(x)}{\partial x_j}\Big)dx \tag{38}$$

where $J(x)$ is the probability current (28). Now let $\delta \to 0$ and recall that for any suitable $F(x) = (F_i(x), \ldots, F_n(x))^T$

$$\begin{aligned}\lim_{\delta\to 0}\int_\Omega \sum_{i=1}^n \frac{\partial f_\delta(x)}{\partial x_i}F_i(x)dx &= -\lim_{\delta\to 0}\int_\Omega f_\delta(x)\sum_{i=1}^n \frac{\partial F_i(x)}{\partial x_i}dx \\ &= -\int_{\Omega_S}\sum_{i=1}^n \frac{\partial F_i(x)}{\partial x_i}dx \\ &= -\int_S \sum_{i=1}^n \hat{n}_{S,i}(x)F_i(x)d\sigma_S(x)\end{aligned} \tag{39}$$

where the first equality follows by integration by parts, the second by definition of $f_\delta(x)$, and the third by the divergence theorem. Using (39), we conclude that the limit of (38) as $\delta \to 0$ is the surface integral at the left hand-side of (31), and we are done.

4 Reaction Rate, Streamlines of the Current and Transition Tubes

Using (22), (23) and (24), it is easy to see that $J_{AB}(x)$ is a divergence free vector field in $\Omega \setminus (A \cup B)$. Indeed, after some straightforward algebraic manipulations, one arrives at

$$\begin{aligned}\sum_{i=1}^n \frac{\partial J_{AB,i}(x)}{\partial x_i} &= -q_+(x)q_-(x) \times \text{RHS(22)} - m(x)q_-(x) \times \text{RHS(23)} \\ &\quad - m(x)q_+(x) \times \text{RHS(24)} = 0\end{aligned} \tag{40}$$

where RHS stands for right hand-side. As a result, by the divergence theorem, the probability flux across any closed surface $S \subset \Omega \setminus (A \cup B)$ is zero,

$$\int_S \hat{n}_S(x) \cdot J_{AB}(x) d\sigma_S(x) = \int_{\Omega_S} \mathrm{div} J_{AB}(x) dx = 0 \quad \text{(closed } S\text{)} \tag{41}$$

where Ω_S is the volume whose boundary is S, $\hat{n}_S(x)$ is the unit normal to S pointing outward Ω_S, $d\sigma_S(x)$ is the surface element on S and $\mathrm{div} J_{AB}(x) = 0$ denotes the divergence of the current that we calculated in (40). For the same reason, the probability flux across any dividing surface S, i.e. any surface which separates Ω into two pieces, one which contains A and the other B, is constant. This is consistent with the fact that there are no source nor sink of reactive trajectories in $\Omega \setminus (A \cup B)$: every reactive trajectory that leaves A must reach B. By construction, the constant probability flux across any dividing surface is simply the *reaction rate*,

$$k_{AB} = \int_S \hat{n}_S(x) \cdot J_{AB}(x) d\sigma_S(x) \qquad \text{(dividing } S\text{)}, \tag{42}$$

where $\hat{n}_S(x)$ is the unit normal to S pointing toward B. This quantity is the limit defined in (6), i.e. it gives the *exact* mean frequency at which the reactive trajectories are observed within a given trajectory.

The expression (42) for the rate can be simplified and transformed into a volume integral over $\Omega \setminus (A \cup B)$:

$$k_{AB} = \int_{\Omega \setminus (A \cup B)} m(x) \sum_{i,j=1}^n a_{ij}(x) \frac{\partial q_+(x)}{\partial x_i} \frac{\partial q_+(x)}{\partial x_j} dx \tag{43}$$

(Equivalently, we could use $q_-(x)$ instead of $q_+(x)$ in this expression.) Since the derivation of (43) from (42) involves a few technical steps we postpone it till the end of this section. To the best of our knowledge, this expression was first given in [14]. It provides an exact alternative to the expression for the reaction rate given by TST and the reactive flux formalism (see Sect. 5). Finally, notice that (11) and (43) can be combined with (16) to give the following expression for the *mean reaction time* t_{AB} defined in (14):

$$t_{AB} = \frac{\int_{\Omega \setminus (A \cup B)} q_+(x) q_-(x) m(x) dx}{\int_{\Omega \setminus (A \cup B)} m(x) \sum_{i,j=1}^n a_{ij}(x) (\partial q_+ / \partial x_i)(\partial q_+(x) / \partial x_j) dx} \tag{44}$$

There is more than the reaction rate that can be extracted from the probability current of reactive trajectories. First, notice that on ∂A where $q_+(x) = 0$ and $q_-(x) = 1$, we have

$$\begin{aligned} \hat{n}_{\partial A}(x) \cdot J_{AB}(x) &= m(x) \sum_{i,j=1}^n \hat{n}_{\partial A,i}(x) a_{ij}(x) \frac{\partial q_+(x)}{\partial x_j} \\ &= m(x) |\nabla q_+(x)|^{-1} \sum_{i,j=1}^n a_{ij}(x) \frac{\partial q_+(x)}{\partial x_i} \frac{\partial q_+(x)}{\partial x_j} \geq 0. \end{aligned} \tag{45}$$

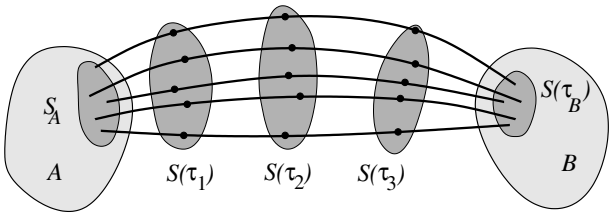

Fig. 2. Schematic of the streamlines of the probability current of reactive trajectories out of $S_A \subset \partial A$, and the pushed forward region $S(s)$ at "times" $0 < \tau_1 < \tau_2 < \tau_3 < s_A$. The collection of these regions, $\{S(\tau) : \tau \geq 0\} = \cup_{\tau \geq 0} S(\tau)$ forms a tube carrying a given percentage of the probability flux of reactive trajectories

where $\hat{n}_{\partial A}(x)$ is the unit normal on ∂A pointing outward A which can be expressed as $\hat{n}_{\partial A}(x) = |\nabla q_+(x)|^{-1} \nabla q_+(x)$. The last inequality in (45) follows from the fact that $a(x)$ is a nonnegative definite tensor. Similarly on ∂B where $q_+(x) = 1$ and $q_-(x) = 0$, we have

$$\begin{aligned} \hat{n}_{\partial B}(x) \cdot J_{AB}(x) &= m(x) \sum_{i,j=1}^{n} \hat{n}_{\partial B,i}(x) a_{ij}(x) \frac{\partial q_-(x)}{\partial x_j} \\ &= m(x) |\nabla q_-(x)|^{-1} \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial q_-(x)}{\partial x_i} \frac{\partial q_-(x)}{\partial x_j} \geq 0. \end{aligned} \tag{46}$$

where $\hat{n}_{\partial B}(x)$ is the unit normal on ∂B pointing inward B which can be expressed as $\hat{n}_{\partial B}(x) = -|\nabla q_-(x)|^{-1} \nabla q_-(x)$. Equations (45) and (46) shows that the probability current of reactive trajectories point outward A everywhere on ∂A and inward B everywhere on ∂B, as it should since every reactive trajectory connects A to B by construction. Now suppose that we identify a region on ∂A, say $S_A \subset \partial A$, which is such that $X\%$ of the probability current goes out of A through S_A, i.e.

$$\int_{S_A} \hat{n}_{\partial A}(x) \cdot J_{AB}(x) d\sigma_{\partial A}(x) = \frac{X}{100} \int_{\partial A} \hat{n}_{\partial A}(x) \cdot J_{AB}(x) d\sigma_{\partial A}(x) \equiv \frac{X}{100} k_{AB} \tag{47}$$

We can then push forward the surface S_A, using the artificial dynamics

$$\frac{dx(\tau)}{d\tau} = J_{AB}(x(\tau)), \qquad x(0) \in S_A \subset \partial A. \tag{48}$$

This equation defines the *streamlines* of $J_{AB}(x)$ (see Fig. 2). Eventually, every streamline $x(\tau)$ must reach B for some $\tau_B > 0$ and we will terminate the streamlines on ∂B by assuming that $x(\tau) = x(\tau_B) \in \partial B$ for $\tau \geq \tau_B$ (notice that τ_B depends on $x(0) \in \partial A$ and may be different for every streamline). Using the divergence theorem again, the probability flux through the surface $S(\tau) = \cup_{x(0) \in S_A} x(\tau)$ (that is, the push-forward of the surface S_A along the

streamlines), is equal to the probability flux through the surface $S_A \equiv S(0)$, i.e.

$$\int_{S_A} \hat{n}_{\partial A}(x) \cdot J_{AB}(x) d\sigma_{\partial A}(x) = \int_{S(\tau)} \hat{n}_{S(\tau)}(x) \cdot J_{AB}(x) d\sigma_{S(\tau)}(x). \tag{49}$$

As a result, the region

$$\{S(\tau) : \tau \geq 0\} = \bigcup_{\tau \geq 0} S(\tau) \tag{50}$$

defines one or more *transition tubes* in $\Omega \setminus (A \cup B)$ which connect A and B and which carries $X\%$ of the probability flux of reactive trajectories (see Fig. 2).

Derivation of (43).

To check that (43) gives the rate, let $S_+(z) = \{q_+(x) = z\}$ be the (forward) isocommittor surface with committor value $z \in [0, 1]$, and consider the integral

$$A(z) = \int_{S_+(z)} m(x) \sum_{i,j=1}^{n} \hat{n}_{S_+(z),i}(x) a_{ij}(x) \frac{\partial q_+(x)}{\partial x_j} d\sigma_{S_+(z)}(x) \tag{51}$$

where $\hat{n}_{S_+(z)}(x)$ is the unit normal to $S_+(x)$ pointing toward B and $d\sigma_{S_+(z)}(x)$ is the surface element on $S_+(z)$. Since $S_+(0) \equiv \partial A$, is easy to see from (42) and (45) that:

$$A(0) = \int_{\partial A} m(x) \sum_{i,j=1}^{n} \hat{n}_{\partial A,i}(x) a_{ij}(x) \frac{\partial q_+(x)}{\partial x_j} d\sigma_{\partial A}(x) \equiv k_{AB} \tag{52}$$

Next, we show that $A(z) = A(0) = k_{AB}$ for all $z \in [0, 1]$. Using the Dirac delta function we can express $A(z)$ as

$$A(z) = \int_{\Omega} m(x) \sum_{i,j=1}^{n} \frac{\partial q_+(x)}{\partial x_i} a_{ij}(x) \frac{\partial q_+(x)}{\partial x_j} \delta(q_+(x) - z) dx \tag{53}$$

and hence

$$\begin{aligned} \frac{dA(z)}{dz} &= -\int_{\Omega} m(x) \sum_{i,j=1}^{n} \frac{\partial q_+(x)}{\partial x_i} a_{ij}(x) \frac{\partial q_+(x)}{\partial x_j} \delta'(q_+(x) - z) dx \\ &= -\int_{\Omega} m(x) \sum_{i,j=1}^{n} \frac{\partial q_+(x)}{\partial x_i} a_{ij}(x) \frac{\partial}{\partial x_j} \delta(q_+(x) - z) dx \end{aligned} \tag{54}$$

Integrating by parts, this gives

$$\begin{aligned}\frac{dA(z)}{dz} &= \int_\Omega m(x) \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 q_+(x)}{\partial x_i \partial x_j} \delta(q_+(x) - z) dx \\ &\quad + \int_\Omega \sum_{i,j=1}^n \frac{\partial q_+(x)}{\partial x_i} \frac{\partial}{\partial x_j} (a_{ij}(x) m(x)) \delta(q_+(x) - z) dx \\ &= - \int_\Omega m(x) \sum_{i=1}^n b_i(x) \frac{\partial q_+(x)}{\partial x_i} \delta(q_+(x) - z) dx \\ &\quad + \int_\Omega \sum_{i,j=1}^n \frac{\partial q_+(x)}{\partial x_i} \frac{\partial}{\partial x_j} (a_{ij}(x) m(x)) \delta(q_+(x) - z) dx \end{aligned} \tag{55}$$

where in the second step we used (23). Using the definition (28) for the equilibrium current $J(x)$, the two integrals in the last equality can be recombined into

$$\begin{aligned}\frac{dA(z)}{dz} &= - \int_\Omega \sum_{i=1}^n \frac{\partial q_+(x)}{\partial x_i} J_i(x) \delta(q_+(x) - z) dx \\ &= - \int_{S_+(z)} \sum_{i=1}^n n_{S_+(z),i}(x) J_i(x) d\sigma_{S_+(z)}(x) = 0 \end{aligned} \tag{56}$$

where in the last equality we use the fact that the probability flux of the regular (by opposition to reactive) trajectories through any surface is zero at equilibrium. Equation (56) implies that $A(z) = A(0) = k_{AB}$ for all $z \in [0, 1]$ as claimed. Hence, $\int_0^1 A(z) dz = k_{AB}$ which from (53) gives

$$\begin{aligned} k_{AB} &= \int_0^1 \int_\Omega m(x) \sum_{i,j=1}^n \frac{\partial q_+(x)}{\partial x_j} a_{ij}(x) \frac{\partial q_+(x)}{\partial x_j} \delta(q_+(x) - z) dx dz \\ &= \int_{\Omega \setminus (A \cup B)} m(x) \sum_{i,j=1}^n \frac{\partial q_+(x)}{\partial x_j} a_{ij}(x) \frac{\partial q_+(x)}{\partial x_j} dx \end{aligned} \tag{57}$$

This is (43).

5 Comparison with Transition State Theory (TST) and Transition Path Sampling (TPS)

A recent account of transition state theory (TST) has been given in [31] (see also [30]). Here we shall content ourselves with briefly summarizing TST (by which we mean the modern version of the theory which accounts for dynamical corrections) and contrasting it with the present approach. The main result of TST is an expression for the reaction rate; this expression is equivalent to (42) but different from it and less general (it only applies to specific cases of (19), such as the Langevin equation (21), and only with specific dividing surfaces).

These restrictions arise because TST is based on a two-step procedure [1, 5]: first one measures the frequency of crossing of a dividing surface S by the trajectory (this gives the bare TST rate of S), then one corrects this frequency to account only for the last crossing by each reactive trajectories (this gives the transmission coefficient of the surface S, which is the coefficient by which the bare TST rate of S must be multiplied to get the actual reaction rate).[4] This two-step procedure is in general not possible with (19) because every reactive trajectory cross any dividing surface infinitely many times, hence the bare TST rate is infinite, the transmission coefficient 0, and their product undetermined. In addition, while the transmission coefficient can be expressed in terms of the functions $q_+(x)$ and $q_-(x)$ (see [31]), within TST these functions are not defined through (23) and (24) (though, of course, they could) but rather estimated via running trajectories from the dividing surface S itself and following them till they reach A or B (which is a step which may prove rather inefficient numerically). Finally, and most importantly, since TST is based on a single dividing surface, this precludes TST from giving any information about the probability density and current of the reactive trajectories away from this surface. Summarizing: TST is oblivious to the probabilistic framework discussed here and, as a result, it gives no information about the mechanism of the reaction.

Transition path sampling (TPS) is discussed in detail in another chapter of this book, and here too we will content ourselves with a few remarks. TPS is based on the observation that the probability density to observe a piece of trajectory $\{x(t) : 0 \le t \le T\}$ satisfying (19) is proportional to

$$\exp\Big(-\frac{1}{4}\int_0^T \sum_{i=1}^n \Big(\sum_{j=1}^n \sigma_{ij}^{-1}(x(t))(\dot{x}_j(t) - b_j(x(t))\Big)^2 dt\Big). \tag{58}$$

where $\sigma^{-1}(x)$ is the inverse of the tensor $\sigma(x)$ defined in (20) (assuming it to be invertible). Therefore (58) allows one to design Metropolis Monte-Carlo schemes in trajectory space, as was suggested earlier by Pratt [25]. If the constraints that $x(0) = x_0 \in A$ with x_0 distributed according to the equilibrium density $m(x)$ restricted and normalized to A and $x(T) \in B$ are added, then these Monte-Carlo schemes permit to sample trajectories which, by construction, start in A and end in B without making any a priori assumption about their behavior in between. The big advantage of TPS is that, if A and B are long-lived (as is the case in situations of interest), by definition these trajectories would be much more difficult to observe by direct simulation of (19). In addition, these trajectories belong to an ensemble (the TPS ensemble) which is close to the ensemble of reactive trajectories defined in (2), except for the additional constraint that only the reactive trajectories which make the transition from A to B in less than T are accounted for. In principle, this bias can

[4] The fact that TST can be viewed this way may not be readily apparent from the traditional exposition of TST, but it is clear from the viewpoint on TST taken in [30, 31] (see also [5, 32]).

be minimized by taking T large enough, though this may leads to additional difficulties both conceptual (when T is large, a piece of trajectory may go back and forth between A and B more than once which requires to re-weight each of these carefully in the ensemble) and practical (the longer the pieces of trajectories, the costlier they are to handle). On the other hand, TPS does not use nor rely on the probabilistic framework discussed in this chapter and as a result it offers no direct way to estimate the probability density of the reactive trajectories, their probability current or their rate. As we know, these quantities depends on the committor functions $q_-(x)$ and $q_+(x)$. In TPS, these functions must be estimated *a posteriori*, via the processing of the trajectories in the TPS ensemble, which is a nontrivial operation (see e.g. [22] where this was attempted on the example of alanine dipeptide). Nevertheless, using the probabilistic framework discussed in this chapter in conjunction with TPS may prove useful and is certainly worth further considerations (for some results in this direction see [2, 20]).

6 The Situations with Localized Transition Tubes

The results presented so far indicates that the isocommittor functions $q_-(x)$ and $q_+(x)$ are essential to understand the mechanism of a reaction. However these results do not say how to compute $q_-(x)$ and $q_+(x)$, except via the solution of (24) and (23) which is a formidable task, even numerically, when the dimensionality of the system is large (that is, in any situation of interest). In this section, we show that the transition tube carrying most of the probability flux of reactive trajectories can be identified under the assumption that this tube is localized, in a sense made precise below. As show in Sect. 7, this is a way to make practical the probabilistic framework presented so far: while standard numerical methods based on finite difference or finite element are inappropriate to determine $q_-(x)$ and $q_+(x)$, under the assumptions of this section one can develop algorithms to estimate these functions locally inside the tubes carrying most of the probability flux of reactive trajectories, i.e. where they matter most.

For simplicity, we will consider first the time-reversible situation, when

$$0 = J_i(x) = b_i(x)m(x) - \sum_{j=1}^{n} \frac{\partial}{\partial x_j}(a_{ij}(x)m(x)) \tag{59}$$

holds (see the footnote on p. 449). An important example which is *not* time-reversible is that of the Langevin equation (21) (recall that reverting the momenta is not allowed in our definition of time-reversibility, see the footnote on p. 444): how to deal with this example will be briefly explained in Sect. 6.5. (59) used in (25) implies that $b^R(x) = b(x)$ and as a result $q_-(x) = 1 - q_+(x)$ since (24) and (23) are identical when $b^R(x) = b(x)$ except for the boundary

conditions. Denoting $q_+(x)$ by $q(x)$ to stress that (59) holds, it is then easy to see that (23) can be written as

$$\begin{cases} 0 = \sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}\Big(a_{ij}(x)m(x)\frac{\partial q(x)}{\partial x_j}\Big) \\ q(x)|_{x\in\partial A} = 0, \qquad q(x)|_{x\in\partial B} = 1. \end{cases} \tag{60}$$

In addition the probability density of reactive trajectories (12) reduces to

$$m_{AB}(x) = Z_{AB}^{-1} q(x)(1-q(x))m(x) \tag{61}$$

where $Z_{AB} = \int_{\Omega\setminus(A\cup B)} q(x)(1-q(x))m(x)dx$, and the probability current of reactive trajectories (32) to

$$J_{AB,i}(x) = \sum_{j=1}^{n} a_{ij}(x)m(x)\frac{\partial q(x)}{\partial x_j} \tag{62}$$

6.1 The Localized Tube Assumption

Let $S(z) = \{x : q(x) = z\}$, $z \in [0,1]$, be the family of *isocommittor surfaces* which foliate $\Omega\setminus(A\cup B)$. The results in this section are based on the following *local transition tube assumption* (see Fig. 3):[5]

> *There exist a family of regions $C(z) \subset S(z)$, $z \in [0,1]$, such that $\cup_{z\in[0,1]} C(z) \equiv \{C(z) : z \in [0,1]\}$ forms a tube in $\Omega \setminus (A \cup B)$ joining A and B with the following properties: $C(z)$ contributes most to the integral of $m(x)$ on $S(z)$, i.e.*
>
> $$\int_{S(z)} m(x)d\sigma_{S(z)}(x) \approx \int_{C(z)} m(x)d\sigma_{S(z)}(x), \tag{63}$$
>
> *and inside each $C(z)$ we have*
>
> $$a_{ij}(x) \approx cst_{ij}, \quad \frac{\partial q(x)}{\partial x_i} \approx cst_i \qquad x \in C(z), \quad i,j = 1,\ldots,n \tag{64}$$

(63) says that the equilibrium probability density restricted to $S(z)$ is actually concentrated on $C(z)$; (64) says that the diffusion tensor $a_{ij}(x)$ is approximately constant in $C(z)$ (the reason for this assumption will become clear shortly) and that the isocommittor surfaces $S(z)$ are locally planar inside the tube. We will denotes these planes by $P(z)$, $z \in [0,1]$, so that $S(z) \approx P(z)$

[5] Observe that it may be necessary to introduce more than one tube $\cup_{z\in[0,1]} C(z)$ if the reaction can proceed by several localized channels. The arguments below can be generalized to these situations, but we will not consider them here for simplicity.

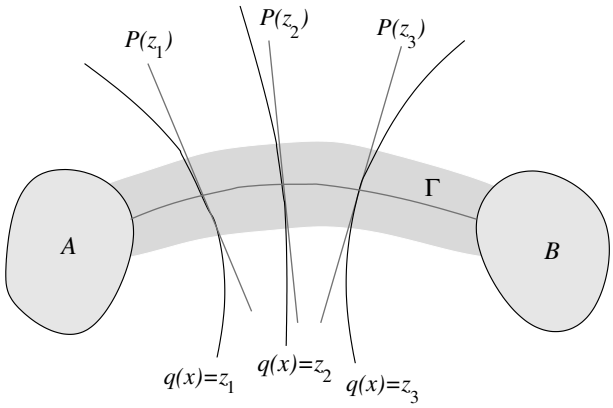

Fig. 3. Schematic representation of tube $\cup_{z\in[0,1]}C(z)$ connecting A and B (shown in *dark grey*), the isocommittor surfaces $S(z)$ where $q(x) = z$, the planes $P(z)$ which approximate $S(z)$ inside the tube, and the curve $\Gamma = \{\varphi(z) : z \in [0,1]\}$ whose geometric location is that of a streamline of the probability current of reactive trajectories inside the tube

inside $\cup_{z\in[0,1]}C(z)$. Equation (63) implies that inside each $P(z)$ there must be a region $D(z) \subset P(z)$ which is such that $D(z) \approx C(z)$ and contributes most to the integral of $m(x)$ on $P(z)$, i.e.

$$\int_{D(z)} m(x)d\sigma_{P(z)}(x) \approx \int_{P(z)} m(x)d\sigma_{P(z)}(x) \; . \tag{65}$$

Observe that the localized assumption does not necessarily mean that the region $C(z)$ in $S(z)$ (or equivalently $D(z)$ in $P(z)$) is small: beside depending on how $a_{ij}(x)$ varies, its extension in various directions actually depends on how much the isocommittor surface $q(x) = z$ is curved in these directions (the more planar it is in one direction, the wider the region can be in this direction).

The main claim of this section is that, if the localized tube assumption holds, then the tube $\cup_{z\in[0,1]}C(z) \approx \cup_{z\in[0,1]}D(z)$ is the tube carrying most of the probability flux of reactive trajectories and it can be determined by an algorithmic procedure which is much simpler than solving (23) and (24) for $q_+(x)$ and $q_-(x)$.

To see this, recall that $m_{AB}(x)$ restricted and properly re-normalized to any dividing surface S is the probability density that the reactive trajectories cross S at $x \in S$ (see (18) in Sect. 2). By (61) $m_{AB}(x)$ restricted and re-normalized to the isocommittor surface $S(z) = \{x : q(x) = z\}$ is the same as $m(x)$ restricted and re-normalized to $S(z)$ since $m_{AB}(x) = Z_{AB}^{-1}z(1-z)m(x)$ in $S(z)$ by construction and the constant factor $Z_{AB}^{-1}z(1-z)$ is absorbed by the re-normalization. Therefore, $m(x)$ restricted and normalized to the isocommittor surface $S(z) = \{x : q(x) = z\}$ is also the probability density

that the reactive trajectories cross $S(z)$ at $x \in S(z)$. As a result (63) implies that the reactive trajectories cross $S(z)$ mostly in the region $C(z)$. Since this is true for all $z \in [0,1]$, this implies that the reactive trajectories remain preferably in the tube $\cup_{z\in[0,1]}C(z)$. As a result, most of the probability flux of reactive trajectories must go through $\cup_{z\in[0,1]}C(z)$ since there cannot be any significant probability current of the reactive trajectories in regions that they do not visit to begin with.

It follows that the streamlines of the the probability current of reactive trajectories must stay inside the tube $\cup_{z\in[0,1]}C(z)$. In fact, (64) implies that these streamlines form a bundle of quasi-parallel curves inside $\cup_{z\in[0,1]}C(z)$, the collection of which actually represent the tube $\cup_{z\in[0,1]}C(z)$. This is a consequence of the equation (48) for the streamlines, which in the present time-reversible context is explicitly

$$\frac{dx_i(\tau)}{d\tau} = m(x(\tau))\sum_{j=1}^{n} a_{ij}(x(\tau))\frac{\partial q(x(\tau))}{\partial x_j}. \tag{66}$$

Together with assumption (64) which says that the vector $(\partial q/\partial x_1, \ldots, \partial q/\partial x_n)^T$ evaluated at any $x \in C(z) \approx D(z) \subset P(z)$ is approximately parallel to the unit normal $\hat{n}(z)$ to the plane $P(z)$, (66) implies that the streamlines solution of (66) cross every $C(z) \approx D(z)$ at approximately the same angle (for instance, if $a(x)$ was the identity matrix, (66) would imply that each streamline is perpendicular to each plane).

Remarkably, the constraint that the streamlines must form a bundle of quasi-parallel curves inside the tube gives us an extra constraint on this tube which permits to determine its geometric location. To see this, it is convenient to re-express (66) as a geometric constraint on the streamlines inside the tube by viewing these streamlines as a collection of curves $\Gamma_x = \{\varphi^x(z) : z \in [0,1]\}$, where $x \in C(0) \subset \partial A$ indexes the curves and

$$\varphi^x(z) = \{x(\tau) : x(0) = x \in C(0) \subset \partial A, \tau \in [0, \tau_B]\} \cap C(z) \tag{67}$$

is the location in $C(z)$ where the streamline starting from $x \in C(0) \subset \partial A$ intersects $C(z)$ (recall that $\tau_B = \tau_B(x)$ is the time at which the streamline starting from $x(0) = x \in \partial A$ reaches ∂B, i.e. $x(\tau_B) \in \partial B$). By construction of Γ_x, we can now represent the tube as

$$\bigcup_{z\in[0,1]} C(z) \approx \bigcup_{z\in[0,1]} D(z) \approx \bigcup_{x\in C(0)\subset\partial A} \Gamma_x \tag{68}$$

and from (66) it follows that each $\varphi^x(z)$ must satisfy

$$\frac{d\varphi_i^x(z)}{dz} \text{ parallel to } \sum_{j=1}^{n} a_{ij}(\varphi^x(z))\frac{\partial q(\varphi^x(z))}{\partial x_j} \tag{69}$$

Since the unit normal $\hat{n}(z)$ to $P(z)$ is approximately parallel to $(\partial q/\partial x_1, \ldots, \partial q/\partial x_n)^T$ evaluated at $\varphi^x(z)$, (69) can be written as

$$\frac{d\varphi_i^x(z)}{dz} \text{ parallel to } \sum_{j=1}^{n} a_{ij}(\varphi^x(z))\hat{n}_j(z) \tag{70}$$

Similar to (66), this equation is a constraint on the angle at which each of the curves $\Gamma_x = \{\varphi^x(z) : z \in [0,1]\}$ intersects each of the planes $P(z)$. Taken together with (63) and (65), (70) is the additional constraint which determines the geometric location tube $\cup_{z\in[0,1]}C(z) \approx \cup_{z\in[0,1]}D(z)$. Let us see why.

Identification of the Tube

Assume that the localized assumption holds and suppose that we give ourself one specific curve $\Gamma = \{\varphi(z) : z \in [0,1]\}$ joining some point in ∂A to some point on ∂B; view Γ as the tentative location of a streamline. Knowing this curve, we can then construct the collection of planes $P(z)$, $z \in [0,1]$, along the curve which are such that $\varphi(z) \in P(z)$ and the unit normal to $P(z)$ satisfies (this is a re-writing of (70) with $\varphi^x(z) \equiv \varphi(z)$):

$$\hat{n}_i(z) \text{ parallel to } \sum_{j=1}^{n} a_{ij}^{-1}(\varphi(z))\frac{d\varphi_j(z)}{dz} \tag{71}$$

Once we have the planes $P(z)$, $z \in [0,1]$, we can also construct the bundle of all the curves Γ_x satisfying (70) (one of the curves Γ_x is the initial Γ.) Observe that this bundle may be bigger than the one in (67) and (68) since we have not specified the regions $D(z)$ yet. In fact, the only thing that limits the extension of this bundle at this point is that it must be defined in the region where the planes $P(z)$ do not intersect, that is in the union of the regions $Q(z) \subset P(z)$, $z \in [0,1]$, around Γ which are such that any point x in $Q(z)$ is closest to $\varphi(z)$ than to any other point along Γ:

$$Q(z) = \{x : x \in P(z) \text{ and } |x - \varphi(z)| = \min_{z'\in[0,1]} |x - \varphi(z')|\} \tag{72}$$

Now look at the regions $D(z)$ in the planes $P(z)$ which concentrate the probability, i.e. the smallest regions in $P(z)$ where (65) holds to some prescribed accuracy. If the initial curve Γ that we choose is a streamline, then we will have

$$\bigcup_{z\in[0,1]} D(z) \subseteq \bigcup_{z\in[0,1]} Q(z), \qquad \text{i.e. } D(z) \subseteq Q(z) \text{ for all } z \in [0,1] \tag{73}$$

In addition, consistent with (68), any curve in the bundle of curves in $\cup_{z\in[0,1]}Q(z)$ will either be in the tube $\cup_{z\in[0,1]}D(z)$ or out of it:

$$\begin{aligned} \text{either: } & \Gamma_x = \{\varphi^x(z) : z \in [0,1]\} \in \cup_{z\in[0,1]}D(z), \\ & \qquad \text{i.e. if } \varphi^x(0) \in D(0) \text{ then } \varphi^x(z) \in D(z) \text{ for all } z \in [0,1], \\ \text{or: } & \Gamma_x = \{\varphi^x(z) : z \in [0,1]\} \notin \cup_{z\in[0,1]}D(z), \\ & \qquad \text{i.e. if } \varphi^x(0) \notin D(0) \text{ then } \varphi^x(z) \notin D(z) \text{ for all } z \in [0,1] \end{aligned} \tag{74}$$

(Since the second equality in (68) is only supposed to be satisfied approximately to some prescribed accuracy, it is allowed to fiddle things a little bit at the frontier of $\cup_{z\in[0,1]}D(z)$ in function of what the curves do at this frontier.) If (73) and (74) hold, then $\cup_{z\in[0,1]}D(z) \approx \cup_{x\in D(0)}\Gamma_x \approx \cup_{z\in[0,1]}C(z)$ is the transition tube.

In contrast, if the initial curve Γ that we choose is not an actual streamline, then the regions $D(z)$, $z\in[0,1]$ obtained via (65) will not be consistent with the bundle of curves Γ_x obtained via (70): either some of the regions $D(z)$ will be at least partially out of the possible regions $Q(z)$, i.e. we will have $D(z) \not\subseteq Q(z)$ for some $z\in[0,1]$ and (73) will fail, or some of the curves in the bundle will enter or leave the tube $\cup_{z\in[0,1]}D(z)$ and (74) will fail, or both.

Summarizing: we need to adjust the initial curve Γ in order to get collection of planes $P(z)$ $z\in[0,1]$ via (71), a collection of regions $D(z)\subset P(z)$, $z\in[0,1]$ via (65) and $Q(z)\subset P(z)$, $z\in[0,1]$ via (72), and an associated bundle of curves via (70) which are such that both (73) and (74) hold. On the other hand, once this collection of $D(z)\subset P(z)$, $z\in[0,1]$ and the associated bundle have been identified, then the tube $\cup_{z\in[0,1]}D(z) \approx \cup_{x\in D(0)}\Gamma_x \approx \cup_{z\in[0,1]}C(z)$ is the transition tube. In fact, such an adjustment procedure is algorithmic in nature, i.e. it can be turned into an numerical scheme to determine the tube, as explained in Sect. 7.

Finally, observe that the adjustment procedure described above determines the geometric location of the tube and the bundle of curves Γ_x, but not their parametrization, i.e. even if we have localized the tube and the bundle of curves Γ_x, we still do not know where are $D(z)$ and $\varphi^x(z)$ along Γ_x: this is important since $D(z)$ is the local approximation of the isocommittor surface $S(z)=\{q(x)=z\}$. We handle this question in the next section.

6.2 Tube Parametrization and Committor Function

In this section, we assume that the geometric location of the tube $\cup_{z\in[0,1]}D(z)$ $\approx \cup_{z\in[0,1]}C(z)$ has been determined by the adjustment procedure described at the end of Sect. 6.1 and that so has the bundle of curves Γ_x which are the geometric location of the streamlines inside the tube. For simplicity we will denote by Γ a specific curve within the bundle (it is not important which one in particular), and we ask how to parametrize it as $\Gamma=\{\varphi(z), z\in[0,1]\}$ where, as in (67), $\varphi(z)$ is the location along Γ where Γ intersects $D(z)$. Recall that this is important, because $D(z)\approx C(z)\subset S(z)$, i.e. it is a local approximation of the isocommittor surface $S(z)=\{q(x)=z\}$: in other words, $q(x)\approx q(\varphi(z))\equiv z$ if $x\in D(z)$.

We will proceed as follows: let $\phi(s)$ be the parametrization of Γ by arclength, i.e. $\Gamma=\{\phi(s): s\in[0,L_\Gamma]\}$ where L_Γ is the length of Γ, and let $\bar{P}(s)$ be the plane which intersects Γ at point $\phi(s)$ in such a way that the unit normal to $\bar{P}(s)$, which we shall denote by $\hat{\nu}(s)$, satisfies the equivalent of (70): $d\phi_i/ds$ is parallel to $\sum_{j=1}^n a_{ij}(\phi(s))\hat{\nu}_j(s)$. Since by construction $|d\phi/ds|=1$,

where $|d\phi/ds|$ is the Euclidean norm of the vector $(d\phi_1/ds, \ldots, d\phi_n/ds)^T$, we have

$$\frac{d\phi_i(s)}{ds} = \frac{\sum_{j=1}^n a_{ij}(\phi(s))\hat{\nu}_j(s)}{|a(\phi(s))\hat{\nu}(s)|} \tag{75}$$

where $|a(\phi(s))\hat{\nu}(s)|$ is the Euclidean norm of

$$(\textstyle\sum_{j=1}^n a_{1,j}(\phi(s))\hat{\nu}_j(s), \ldots, \sum_{j=1}^n a_{n,j}(\phi(s))\hat{\nu}_j(s))^T.$$

Since Γ is assumed to be known, all these quantities are known and what remains to be done is relate $\varphi(z)$ to $\phi(s)$, $\hat{n}(z)$ to $\hat{\nu}(s)$ and $P(z)$ to $\bar{P}(s)$, i.e. find the function $f(z)$ such that $\varphi(z) = \phi(f(z))$, $\hat{n}(z) = \hat{\nu}(f(z))$ and $P(z) = \bar{P}(f(z))$.

To do so, recall that $z = q(\varphi(z)) = q(\phi(f(z)))$. Differentiating with respect to z, we deduce that

$$\begin{aligned} 1 &= \frac{df(z)}{dz} \sum_{i=1}^n \frac{d\phi_i(f(z))}{ds} \frac{\partial q(\phi(f(z)))}{\partial x_i} \\ &= \frac{df(z)}{dz} \frac{\sum_{i,j=1}^n a_{ij}(\phi(f(z)))\hat{\nu}_j(f(z))\partial q(\phi(f(z)))/\partial x_i}{|a(\phi(f(z)))\hat{\nu}(f(z))|} \end{aligned} \tag{76}$$

where we used (75) in the second step. This equation is not closed because we do not know $\partial q/\partial x_i$. To get the latter, let $\Omega(z)$ be the region enclosed between ∂A and $S(z)$ and integrate (60) over $\Omega(z)$. Using the divergence theorem, this gives

$$\begin{aligned} 0 &= \int_{\Omega(z)} \sum_{i,j=1}^n \frac{\partial}{\partial x_i}\Big(a_{ij}(x)m(x)\frac{\partial q(x)}{\partial x_j}\Big) \\ &= -\int_{\partial A} \sum_{i,j=1}^n \hat{n}_{\partial A,i}(x)a_{ij}(x)\frac{\partial q(x)}{\partial x_j}m(x)d\sigma_{\partial A}(x) \\ &\quad + \int_{S(z)} \sum_{i,j=1}^n \hat{n}_{S(z),i}(x)a_{ij}(x)\frac{\partial q(x)}{\partial x_j}m(x)d\sigma_{S(z)}(x) \end{aligned} \tag{77}$$

where $\hat{n}_{\partial A}(x)$ is the unit normal to ∂A pointing outward A and $d\sigma_{\partial A}(x)$ is the surface element on ∂A. From (42) and (62), both integrals at the right-hand side of (77) are the reaction rate k_{AB}, so this equation can be written as

$$k_{AB} = \int_{S(z)} \sum_{i,j=1}^n \hat{n}_{S(z),i}(x)a_{ij}(x)\frac{\partial q(x)}{\partial x_j}m(x)d\sigma_{S(z)}(x) \tag{78}$$

So far, we have not made any approximation, i.e. (78) is exact. Let us now use the localized tube assumption, i.e. (63), (64) and (65): it implies that

$$
\begin{aligned}
k_{AB} &\approx \sum_{i,j=1}^{n} \hat{n}_i(z) a_{ij}(\varphi(z)) \frac{\partial q(\varphi(z))}{\partial x_j} Z_{P(z)} \\
&= \sum_{i,j=1}^{n} \hat{\nu}_i(f(z)) a_{ij}(\phi(f(z))) \frac{\partial q(\phi(f(z)))}{\partial x_j} Z_{\bar{P}(f(z))}
\end{aligned}
\tag{79}
$$

where given any plane P, Z_P is defined as

$$
Z_P = \int_P m(x) d\sigma_P(x) \tag{80}
$$

(79) gives the information on $\partial q/\partial x_i$ that we were missing in (76): combining these two equations we arrive at the following equation for $f(z)$:

$$
\frac{df(z)}{dz} = \frac{1}{k_{AB}} Z_{\bar{P}(f(z))} |a(\phi(f(z)))\hat{\nu}(f(z))| \tag{81}
$$

This differential equation actually determines both $f(z)$ and k_{AB}. Indeed it is a first order differential equation, but with two boundary conditions for $f(z)$: $f(0) = 0$ (since $\phi(0) = \varphi(0)$) and $f(1) = L_\Gamma$ (since $\phi(L_\Gamma) = \varphi(1)$). Due to the special structure of (81) (namely because k_{AB} enters as an overall multiplicative factor at the right hand-side), we can actually imposes these two boundary conditions, which gives $f(z)$ and fixes k_{AB} as well.

Once $f(z)$ is known, using $\varphi(z) = \phi(f(z))$ and (75), we have

$$
\frac{d\varphi_i(z)}{dz} = \frac{df(z)}{dz} \frac{d\phi_i(f(z))}{ds} = \frac{df(z)}{dz} \frac{\sum_{j=1}^{n} a_{ij}(\varphi(z))\hat{n}_j(z)}{|a(\varphi(z))\hat{n}(z)|} \tag{82}
$$

This equation gives the proportionality factor that was missing in (70).

6.3 Reaction Rate and Probability to be Reactive

Once the locations of $\varphi(z)$ and the associated planes $P(z)$ along Γ are known, we can give simple expressions for the various quantities of interest in terms of these objects.

First, we can re-express the rate k_{AB} in terms of $\varphi(z)$ and $\hat{n}(z)$ alone as follows. Using $\phi(f(z)) = \varphi(z)$ and $\hat{\nu}(f(z)) = \hat{n}(z)$ in (76), this equation can be written as

$$
1 = \frac{df(z)}{dz} \frac{\sum_{i,j=1}^{n} a_{ij}(\varphi(z))\hat{n}_j(z)\partial q(\varphi(z))/\partial x_i}{|a(\varphi(z))\hat{n}(z)|} \tag{83}
$$

where $|a(\varphi(z))\hat{n}(z)|$ denotes the Euclidean norm of

$$
(\textstyle\sum_{j=1}^{n} a_{1,j}(\varphi(z))\hat{n}_j(z), \ldots, \sum_{j=1}^{n} a_{n,j}(\varphi(z))\hat{n}_j(z))^T.
$$

Solving (83) in $\sum_{i,j=1}^{n} a_{ij}(\varphi(z))\hat{n}_j(z)\partial q(\varphi(z))/\partial x_i$ and using the result in the first equality in (79) gives

$$k_{AB} \approx \frac{|a(\varphi(z))\hat{n}(z)|Z_{P(z)}}{df(z)/dz} \tag{84}$$

Multiplying both members by $df(z)/dz$ and integrating on $z \in [0,1]$ using $f(0) = 0$ and $f(1) = L_\Gamma$, leads to

$$L_\Gamma k_{AB} \approx \int_0^1 |a(\varphi(z))\hat{n}(z)|Z_{P(z)}dz \tag{85}$$

Since $L_\Gamma = \int_0^1 |d\varphi/dz|dz$, where $|d\varphi/dz|$ denotes the Euclidean norm of the vector $(d\varphi_1/dz, \ldots, d\varphi_n/dz)^T$, we arrive at the following expression for k_{AB}:

$$k_{AB} \approx \frac{\int_0^1 |a(\varphi(z))\hat{n}(z)|Z_{P(z)}dz}{\int_0^1 |d\varphi/dz|dz} \tag{86}$$

Proceeding similarly, we can express Z_{AB} in terms of $\varphi(z)$ and $\hat{n}(z)$. Observe that

$$\begin{aligned} Z_{AB} &= \int_{\Omega\setminus(A\cup B)} q(x)(1-q(x))m(x)dx \\ &= \int_0^1 z(1-z)\Big(\int_{S(z)} \frac{m(x)}{|\nabla q(x)|} d\sigma_{S(z)}(x)\Big)dz \\ &\approx \int_0^1 z(1-z)\frac{Z_{P(z)}}{|\nabla q(\varphi(z))|}dz \end{aligned} \tag{87}$$

where we used (63) and (64) in the last step. Since $z = q(\varphi(z))$, we have

$$1 = \sum_{i=1}^n \frac{d\varphi_i(z)}{dz}\frac{\partial q(\varphi(z))}{\partial x_i} = |\nabla q(\varphi(z))| \sum_{i=1}^n \frac{d\varphi_i(z)}{dz}\hat{n}_i(z) \tag{88}$$

Solving this equation for $|\nabla q(\varphi(z))|$ and using the result in (87) gives

$$Z_{AB} \approx \int_0^1 z(1-z)\sum_{i=1}^n \frac{d\varphi_i(z)}{dz}\hat{n}_i(z)Z_{P(z)}dz \tag{89}$$

6.4 Working in Collective Variables

In Sect. 6.1 we assumed that the local transition tube assumption was valid in the original variables x. This is a strong assumption which may fail in many situations of interest. On the other hand, it is reasonable to assume that the range of validity of the local transition tube assumption may be extended in some suitable set of collective variables which are sufficient to describe the reaction. This is the situation that we describe in this section.

Suppose that $q(x)$ depends of x only through a set of *collective variables*, $(\theta_1(x), \ldots, \theta_N(x))$, i.e.

$$q(x) \approx Q(\theta_1(x), \ldots, \theta_N(x)), \tag{90}$$

for some unknown function $Q(\theta_1, \ldots, \theta_N)$. The collective variables can go all the way from being a set of variables introduced to account for some symmetries (like e.g. translation or rotation), in which case $N < n$ but $N \approx n$ and there is no approximation in (90), to being a much smaller set of variables (like e.g torsion angles, bond distances, etc.), chosen because one has reason to believe a priori that this set of variables is large enough to describe the reaction. In this second case, $N \ll n$, but the there is no reason to assume that N is small, i.e. there could still be many collective variables $(\theta_1(x), \ldots, \theta_N(x))$, so the approximation made in (90) is not as restrictive as it may look at first sight.

What is the equation for $Q(\theta_1, \ldots, \theta_n)$ if one assumes that (90) holds? Let us go back to (60): to incorporate (90) in (60), observe that (60) is the Euler-Lagrange equation associated with the minimization of [6]

$$I(q) = \int_\Omega \sum_{i,j=1}^n a_{ij}(x) \frac{\partial q(x)}{\partial x_i} \frac{\partial q(x)}{\partial x_j} m(x) dx \tag{91}$$

over all $q(x)$ satisfying $q(x)|_{x\in A} = 0$ and $q(x)|_{x\in B} = 0$. Consistent with (90), we can minimize (91) over all functions $q(x)$ in the form of (90) and get a good approximation of the committor function. Using (90) in (91), this object function can be reduced to the following object function for Q:

$$\begin{aligned} \bar{I}(Q) &= \int_\Omega \sum_{k,l=1}^N \sum_{i,j=1}^n a_{ij}(x) \frac{\partial \theta_k(x)}{\partial x_i} \frac{\partial \theta_l(x)}{\partial x_j} \Big(\frac{\partial Q(\theta^*)}{\partial \theta_k} \frac{\partial Q(\theta^*)}{\partial \theta_l} \Big)_{\theta(x)=\theta^*} m(x) dx \\ &= \int_{\Omega_\theta} \sum_{k,l=1}^N A_{kl}(\theta^*) \frac{\partial Q(\theta^*)}{\partial \theta_k} \frac{\partial Q(\theta^*)}{\partial \theta_l} M(\theta^*) d\theta^* \end{aligned} \tag{92}$$

where Ω_θ,is the image of Ω by the map $\theta : \mathbb{R}^n \mapsto \mathbb{R}^N$,

$$M(\theta^*) = \int_\Omega m(x) \delta(\theta_1(x) - \theta_1^*) \cdots \delta(\theta_N(x) - \theta_N^*) dx \tag{93}$$

is the equilibrium probability density of the collective variable $\theta(x)$ and

$$\begin{aligned} A_{kl}(\theta^*) &= M^{-1}(\theta^*) \int_\Omega \sum_{i,j=1}^n a_{ij}(x) \frac{\partial \theta_k(x)}{\partial x_i} \frac{\partial \theta_l(x)}{\partial x_j} \\ &\qquad \times m(x) \delta(\theta_1(x) - \theta_1^*) \cdots \delta(\theta_N(x) - \theta_N^*) dx \\ &\equiv \Big\langle \sum_{i,j=1}^n a_{ij}(x) \frac{\partial \theta_k(x)}{\partial x_i} \frac{\partial \theta_l(x)}{\partial x_j} \Big\rangle_{\theta(x)=\theta^*}. \end{aligned} \tag{94}$$

[6] Observe that the object function (91) may be used as a starting point to implement other approximations than (90). Observe also that as a direct consequence of (43)), $k_{AB} = \min I(q)$ over all $q(x)$ satisfying $q(x)|_{x\in A} = 0$ and $q(x)|_{x\in B} = 1$, i.e. the actual $q(x)$ solution of (60) is the one that minimizes the reaction rate.

is the conditional average of $\sum_{i,j=1}^{n} a_{ij}(x)(\partial\theta_k/\partial x_i)(\partial\theta_l/\partial x_j)$ in $\theta(x) = \theta^*$. (92) must be minimized over all functions $Q(\theta)$ satisfying $Q(\theta)|_{\theta\in A_\theta} = 0$ and $Q(\theta)|_{\theta\in B_\theta} = 1$ where A_θ and B_θ are the images of A and B, respectively, by the map $\theta : \mathbb{R}^n \mapsto \mathbb{R}^N$

Equation (92) is simpler than (60) but otherwise structurally identical to it. In particular, the Euler-Lagrange equation associated with minimizing (92) has the form of a backward Kolmogorov equation:

$$\begin{cases} 0 = \displaystyle\sum_{k,l=1}^{N} \frac{\partial}{\partial\theta_k}\Big(A_{kl}(\theta)M(\theta)\frac{\partial Q(\theta)}{\partial\theta_l}\Big) \\ Q(\theta)|_{\theta\in\partial A_\theta} = 0, \qquad Q(\theta)|_{\theta\in\partial B_\theta} = 1. \end{cases} \tag{95}$$

Therefore (95) instead of (60) can be taken as a starting point for what we did in Sect. 6.1. Observe that (95) is the backward Kolmogorov equation associated with the stochastic differential equation (compare (19))

$$\dot\theta(\tau) = B(\theta(\tau)) + \sqrt{2}\bar\sigma(\theta(\tau))\eta(\tau) \tag{96}$$

where $\eta_k(\tau)$ is a white-noise satisfying $\langle\eta_k(\tau)\eta_l(\tau')\rangle = \delta_{kl}\delta(\tau - \tau')$, $\bar\sigma(\theta)$ is such that $\sum_{l=1}^{N} \bar\sigma_{kl}(\theta)\bar\sigma_{k'l}(\theta) = A_{kk'}(\theta)$ and

$$B_k(\theta) = \frac{1}{M(\theta)}\sum_{l=1}^{N}\frac{\partial}{\partial\theta_l}(M(\theta)A_{kl}(\theta)) \tag{97}$$

The time τ in (96) is an artificial time whose relation with the physical time t in (19) is unspecified by our argument. In other words, (96) allows one to understand the mechanism of the reaction but not its timing or rate. To obtain the latter, we must go back to the object function in (92) and observe that $k_{AB} = \min_{q(x)} I(q) \approx \min_{Q(\theta)} \bar I(Q)$, i.e. within approximation (90) we have

$$k_{AB} = \int_{\Omega_\theta}\sum_{k,l=1}^{N} A_{kl}(\theta)\frac{\partial Q(\theta)}{\partial\theta_k}\frac{\partial Q(\theta)}{\partial\theta_l}M(\theta)d\theta \tag{98}$$

where $Q(\theta)$ is the solution of (95).

6.5 The Case of the Langevin Dynamics

Consider now the Langevin equation in (21). This is an important example which is not time-reversible, i.e. $q_+(r,v) \neq 1 - q_-(r,v)$ (in fact it is easy to see that $q_+(r,v) = 1 - q_-(r,-v)$ since, in order to know what happens when we revert time, we need to revert the momenta). The backward Kolmogorov equation (23) in the present case is

$$
\begin{aligned}
0 = &\sum_{i=1}^m \Big(v_i \frac{\partial q_+(r,v)}{\partial r_i} - \mu_i^{-1} \frac{\partial V(r)}{\partial r_i} \frac{\partial q_+(r,v)}{\partial v_i} \Big) \\
&+ \gamma \sum_{i=1}^m \Big(- \mu_i^{-1} v_i \frac{\partial q_+(r,v)}{\partial v_i} + \mu_i^{-2} k_B T \frac{\partial^2 q_+(r,v)}{\partial v_i^2} \Big)
\end{aligned}
\tag{99}
$$

plus boundary conditions. To deal with this case, let us assume that $q_+(x,v)$ can be approximated by a function independent of the momenta v and depending on the positions only through the collective variables $(\theta_1(r), \ldots, \theta_N(r))$, i.e.

$$
q_+(r,v) \approx Q(\theta_1(r), \ldots, \theta_N(r)) \tag{100}
$$

for some unknown function $Q(\theta_1, \ldots, \theta_N)$. (Observe that since $q_-(r,v) = 1 - q_+(r,-v)$, we would then also have $q_-(r,v) \approx 1 - Q(\theta_1(r), \ldots, \theta_N(r))$ for the same function Q as in the time-reversible case.) To incorporate the approximation (100) into (99), observe that the solution of this equation minimizes the object function

$$
\begin{aligned}
I(q_+) = \int_{\Omega \setminus (A \cup B)} &\Big| \sum_{i=1}^m \Big(v_i \frac{\partial q_+(r,v)}{\partial r_i} - \mu_i^{-1} \frac{\partial V(r)}{\partial r_i} \frac{\partial q_+(r,v)}{\partial v_i} \Big) \\
&+ \gamma \sum_{i=1}^m \Big(- \mu_i^{-1} v_i \frac{\partial q_+(r,v)}{\partial v_i} + \mu_i^{-2} k_B T \frac{\partial^2 q_+(r,v)}{\partial v_i^2} \Big) \Big|^2 e^{-\beta H(r,v)} dr dv
\end{aligned}
\tag{101}
$$

where $H(r,v) = \frac{1}{2} \sum_{i=1}^m \mu_i v_i^2 + V(r)$ is the Hamiltonian. Now insert the ansatz (99) into (101). It can be checked that (101) then reduces precisely to (92) with $m(x) = Z^{-1} e^{-\beta V(r)}$ and $a_{ij}(x) = \mu_i^{-1} \delta_{ij}$. In other words, we can again apply all what we did before in Sects. 6.1–6.4 to the present case.

It should be stressed, however, that the approximation in (100) may be adequate to describe the mechanism of the transition but it does not allow us to compute the rate of the reaction. Indeed, from (43), the reaction rate in the present case is

$$
k_{AB} = Z_H^{-1} k_B T \gamma \int_{\Omega \setminus (A \cup B)} \sum_{i=1}^m \mu_i^{-2} \Big(\frac{\partial q_+(r,v)}{\partial v_i} \Big)^2 e^{-\beta H(r,v)} dr dv \tag{102}
$$

where $Z_H = \int_{\mathbb{R}^m \times \mathbb{R}^m} e^{-\beta H(r,v)} dr dv$ ($Z_H^{-1} e^{-\beta H(r,v)}$ is the equilibrium probability density function of the Langevin equation (21)). Thus (102) depends on the corrections to (100) which are not accounted for by this approximation. Observe that this does not make (100) inconsistent: the rate depends on the derivatives of $q(r,v)$, which may be more difficult to approximate that $q(r,v)$ itself. How is this possible is illustrated by the following example.

Remark: the high friction limit of (99).

Suppose that one is interested in a situation where the friction coefficient in (99) is big, $\gamma \gg 1$. In this case (99) can be solved by singular perturbation

techniques: here we used the techniques developed in [24] where the proofs of the claims that we make below can be found. Look for a solution of (99) in the form of

$$q_+(r,v) = q_0(r,v) + \gamma^{-1} q_1(r,v) + \gamma^{-2} q_2(r,v) + \cdots \tag{103}$$

Inserting this ansatz into (99) and equating equal powers of γ leads to the hierarchy:

$$\begin{cases} \displaystyle\sum_{i=1}^{m} \Big(-\mu_i^{-1} v_i \frac{\partial q_0(r,v)}{\partial v_i} + \mu_i^{-2} k_B T \sum_{i=1}^{n} \frac{\partial^2 q_0(r,v)}{\partial v_i^2} \Big) = 0, \\ \displaystyle\sum_{i=1}^{m} \Big(-\mu_i^{-1} v_i \frac{\partial q_1(r,v)}{\partial v_i} + \mu_i^{-2} k_B T \sum_{i=1}^{n} \frac{\partial^2 q_1(r,v)}{\partial v_i^2} \Big) \\ \qquad\qquad = -\displaystyle\sum_{i=1}^{m} \Big(v_i \frac{\partial q_0(r,v)}{\partial r_i} - \mu_i^{-1} \frac{\partial V(r)}{\partial r_i} \frac{\partial q_0(r,v)}{\partial v_i} \Big), \\ \displaystyle\sum_{i=1}^{m} \Big(-\mu_i^{-1} v_i \frac{\partial q_2(r,v)}{\partial v_i} + \mu_i^{-2} k_B T \sum_{i=1}^{m} \frac{\partial^2 q_2(r,v)}{\partial v_i^2} \Big) \\ \qquad\qquad = -\displaystyle\sum_{i=1}^{m} \Big(v_i \frac{\partial q_1(r,v)}{\partial r_i} - \mu_i^{-1} \frac{\partial V(r)}{\partial r_i} \frac{\partial q_1(r,v)}{\partial v_i} \Big), \\ \quad \cdots \end{cases} \tag{104}$$

The first equation implies that $q_0(r,v)$ belongs to the null-space of the operator at the left hand-side, $L_v \equiv \sum_{i=1}^{m} (-\mu_i^{-1} v_i \partial/\partial v_i + \mu_i^{-2} k_B T \partial^2/\partial v_i^2)$. It can be shown [24] that this requires that $q_0(r,v)$ be a function of r only, i.e. $q_0(r,v) = Q(r)$. Since the null-space of the operator L_v is non-trivial, this operator is not invertible and the second and the third equations in (104) require solvability conditions. To understand how these solvability conditions come about, multiply both sides of the second equation in (104) by $(2\pi k_B T)^{-m/2} e^{-\frac{1}{2}\beta |v|_\mu^2}$ where $|v|_\mu^2 = \sum_{i=1}^{m} \mu_i v_i^2$, and integrate on v. Integrating by parts at the left hand-side using

$$0 = \sum_{i=1}^{m} \left(\mu_i^{-1} \frac{\partial}{\partial v_i} \Big(v_i e^{-\frac{1}{2}\beta |v|_\mu^2} \Big) + \mu_i^{-2} k_B T \frac{\partial^2}{\partial v_i^2} e^{-\frac{1}{2}\beta |v|_\mu^2} \right) \tag{105}$$

we arrive at the condition (using $\partial q_0/\partial v_i = \partial Q/\partial v_i = 0$)

$$0 = -\int_{\mathbb{R}^n} (2\pi k_B T)^{-m/2} e^{-\frac{1}{2}\beta |v|_\mu^2} \sum_{i=1}^{m} v_i \frac{\partial Q(r)}{\partial r_i} dv \tag{106}$$

This condition is the solvability condition for the second equation in (104); our argument above shows that it is a necessary condition in order that this equation be solvable and it can be proven [24] that this condition is also sufficient to that end. Equation (106) is automatically satisfied since

$\partial Q/\partial r_i$ can be pulled out of the integral (it is independent of v) and $\int_{\mathbb{R}^m}(2\pi k_BT)^{-m/2}e^{-\frac{1}{2}\beta|v|_\mu^2}v_i dv = 0$. This means that one can solve the equation for q_1: it is easy to see that

$$q_1(r,v) = \sum_{i=1}^n \mu_i v_i \frac{\partial Q(r)}{\partial r_i}. \tag{107}$$

is a solution to this equation and it can be shown [24] that this solution is in fact unique. Inserting this expression into the equation for q_2 in (104) gives:

$$\begin{aligned}\sum_{i=1}^m \Big(-\mu_i^{-1} v_i \frac{\partial q_2(r,v)}{\partial v_i} + \mu_i^{-2} k_BT \frac{\partial^2 q_2(r,v)}{\partial v_i^2}\Big) \\ = -\sum_{i,j=1}^n \mu_i v_i v_j \frac{\partial^2 Q(r)}{\partial r_i \partial r_j} + \sum_{i=1}^n \frac{\partial V(r)}{\partial r_i}\frac{\partial Q(r)}{\partial r_i}.\end{aligned} \tag{108}$$

The solvability condition for this equation is also obtained by multiplying both sides by $(2\pi k_BT)^{-m/2}e^{-\frac{1}{2}\beta|v|_\mu^2}$ and integrating over v. Integrating by parts at the left hand-side and using (105), we arrive at

$$\begin{aligned}0 &= \int_{\mathbb{R}^n}(2\pi k_BT)^{-n/2}e^{-\frac{1}{2}\beta|v|_\mu^2}\left(-\sum_{i,j=1}^n \mu_i v_i v_j \frac{\partial^2 Q(r)}{\partial r_i \partial r_j} + \sum_{i=1}^n \frac{\partial V(r)}{\partial r_i}\frac{\partial Q(r)}{\partial r_i}\right) dv \\ &= \sum_{i=1}^n \Big(-k_BT \frac{\partial^2 Q(r)}{\partial r_i^2} + \frac{\partial V(r)}{\partial r_i}\frac{\partial Q(r)}{\partial r_i}\Big),\end{aligned} \tag{109}$$

This equation is the limiting equation for $Q(r)$: it is the backward Kolmogorov equation associated with the overdamped equation obtained from (21) in the limit as $\gamma \to \infty$ (the overdamped dynamics is considered in Sect. 6.6). On the other hand, inserting (107) into (103) gives

$$q_+(r,v) = Q(r) + \gamma^{-1}\sum_{i=1}^n \mu_i v_i \frac{\partial Q(r)}{\partial r_i} + O(\gamma^{-2}) \tag{110}$$

This expression is consistent with (100) to leading order in γ, but it is the next order correction which gives the rate since $Q(r)$ is independent of v and (102) depends on the gradient of $q_+(r,v)$ with respect to v. In fact, using (110) in (102) and integrating out the momenta, we see that

$$k_{AB} = Z^{-1}k_BT\gamma^{-1}\int_{\mathbb{R}^m\setminus(A_r\cup B_r)}\left(\sum_{i=1}^n \frac{\partial Q(r)}{\partial r_i}\right)^2 e^{-\beta V(r)}dr \tag{111}$$

where A_r and B_r are the projections of A and B in configuration space.

Of course, in general one is not interested in the high friction limit in the original Cartesian space. However, assuming that $q(r,v) \approx Q(\theta_1(r),\ldots,\theta_N(r))$,

it may be reasonable to suppose that the dynamics is overdamped at the level of the θ_j's. In this case a generalization of (111) where one would take $Q(r) \approx Q(\theta_1(r), \ldots, \theta_N(r))$ may be appropriate provided that one can decide what γ is. On the other hand, it may be safer to estimate the rate via other techniques, for instance by estimating t_{AB} by some other means (e.g. using reactive trajectories generated by TPS or obtained by initiating trajectories from the isocommittor $\frac{1}{2}$ surface) and using (16) with Z_{AB} given by (89) (the latter expression being dominated by the leading order term $q(r, v) \approx Q(\theta_1(r), \ldots, \theta_N(r))$ since (11) depends on q itself and not on its gradient).

6.6 Remark: the Role of the Minimum Energy Path

Consider the *overdamped dynamics* when (19) takes the form

$$\gamma\, \dot{r}_i(t) = -\frac{\partial V(r(t))}{\partial r_i} + \sqrt{2k_B T \gamma}\, \eta_i(t) \tag{112}$$

As explained in Sect. 6.5, this equation arises from (21) in the high friction limit as $\gamma \gg 1$. The equilibrium probability density associated with (112) is

$$m(r) = Z^{-1} e^{-\beta V(r)} \quad \text{where} \quad Z = \int_{\mathbb{R}^m} e^{-\beta V(r)} dr \tag{113}$$

In addition, $q(r)$ satisfies (60), which in the present context is (this is nothing but (109))

$$\begin{cases} 0 = -\nabla V(r) \cdot \nabla q(r) + k_B T\, \Delta q(r) \\ q(r)|_{r \in \partial A} = 0, \qquad q(r)|_{r \in \partial B} = 1\ , \end{cases} \tag{114}$$

where $\nabla q = (\partial q/\partial r_1, \ldots, \partial q/\partial r_m)^T$, the dot denotes the standard inner product and $\Delta q = \sum_{i=1}^m \partial^2 q/\partial r_i^2$. In the present case the calculations in Sect. 6.1 indicates that the curves $\Gamma_x = \{\varphi^x(z) : z \in [0,1]\}$ must be perpendicular to each plane $P(z)$ (this is (70)). Quite interestingly, if one makes the additional assumption that the temperature is so small that the density $e^{-\beta V(r)}$ in (113) is strongly peaked at the minimum of $V(r)$ in $P(z)$, then each $D(z) \approx C(z)$ shrinks to a single point located at the minimum of $V(r)$ in $P(z)$; hence the tube $\cup_{z\in[0,1]} C(z) \approx \cup_{z\in[0,1]} D(z)$ shrinks to a single curve, $\Gamma = \{\varphi(z) : z \in [0,1]\}$, which is such that

$$\varphi(z) = \text{minimum of } V(r) \text{ in } P(z)\ . \tag{115}$$

This relation requires that

$$\hat{n}_i(z) \ \text{ parallel to } \ \frac{\partial V(\varphi(z))}{\partial r_i} \tag{116}$$

Combining this equation with (70), which in the present context reads

$$\frac{d\varphi(z)}{dz} \quad \text{parallel to} \quad \hat{n}_i(z) \tag{117}$$

we therefore deduce that

$$\{\varphi(z) : z \in [0,1]\} = \text{curve } \Gamma \text{ such that } \nabla V(r) \text{ along } \Gamma \text{ is parallel to } \Gamma \tag{118}$$

This equation is sometimes written as $0 = [\nabla V(r)]^\perp$, which is supposed to hold everywhere along Γ and where $[\cdot]^\perp$ denotes the projection perpendicular to Γ.

The curve defined by (118) is a very well-known object: it is the *minimum energy path* (MEP) which connects two minima of $V(r)$ via a saddle point. How, why and by whom the MEP was first introduced in the context of molecular dynamics is not clear (to the author at least). The argument above indicates that it is the relevant object that concentrates most of the probability current of the reactive trajectories in its vicinity in the case of the overdamped dynamics when the temperature is small and the potential is sufficiently smooth (otherwise, if $V(r)$ has many critical points, (118) has many solutions, none of which taken alone is relevant). It is however not clear when such situations arise, and the MEP may often prove irrelevant.

7 Some Computational Aspects

In this section, we briefly discuss the computational aspects of the theory discussed so far. As explained in Sect. 6, the committors functions $q_+(x)$ and $q_-(x)$ are the key objects to determine. Whereas this cannot be done by solving (23) and (24) directly (in most applications of practical interest these equations are simply too large for standard numerical techniques such as finite differences or finite elements), $q_+(x)$ and $q_-(x)$ can be determined locally in the situations where the localized tube assumption discussed in Sect. 6.1 (or its equivalent in collective variables, see Sects. 6.4 and 6.5) is valid. This is the essence of the *string method* which we briefly discuss now. Since the focus of the present chapter is mainly the theory, we will be rather sketchy here and content ourselves with indicating how the ideas presented in Sect. 6 naturally lead to the string method: for the details on how to implement this algorithm, we refer the reader to the original references [9–12, 23, 26–28].

As explained in Sect. 6.1, if the localized tube assumption is valid, it is possible to identify this tube by moving a single curve, Γ, viewed as a candidate for the location of a streamline in the tube, to get a collection of planes $P(z)$ $z \in [0,1]$ via (71), a collection of regions $D(z) \subset P(z)$, $z \in [0,1]$ via (65) and $Q(z) \subset P(z)$, $z \in [0,1]$ via (72), and an associated bundle of curves via (70) which are such that both (73) and (74) hold. Now, since the location of the tube depends at least in part on the equilibrium probability density $m(x)$, it is natural to simplify the procedure described at the end of Sect. 6.1 and try to estimate a priori (that is, guess) the location of a specific streamline

inside the tube based on the value of $m(x)$ inside the tube. There are two natural a priori estimates for the location of the streamline: the first leads to the zero-temperature string method and is explained in Sect. 7.1; the second leads to the finite temperature string method and is explained in Sect. 7.2. Both methods can be generalized to work in collective variables, as explained in Sect. 7.3. As shown below, making an a priori estimate for the location of the streamline simplifies the actual identification of the tube, though it introduces an additional assumption beyond the localized tube assumption of Sect. 6.1 and hence is somewhat more restrictive.

7.1 The Zero-Temperature String Method

The first natural a priori estimate for the location of a specific streamline $\Gamma = \{\varphi(z) : z \in [0,1]\}$ inside the tube is to take for $\varphi(z)$ the location in $C(z)$ where $m(x)$ is maximum, i.e. take

$$\varphi(z) = \text{location of the maximum of } m(x) \text{ in } C(z). \tag{119}$$

Since $C(z) \approx D(z) \subset P(z)$, this implies that $\varphi(z)$ satisfies

$$\frac{\partial m(\varphi(z))}{\partial x_i} \text{ parallel to } \hat{n}_i(z) \tag{120}$$

This equation and (70) can then be combined into the following single equation for the curve:

$$\{\varphi(z) : z \in [0,1]\} = \text{curve } \Gamma \text{ such that } a(x)\nabla m(x) \text{ along } \Gamma \text{ is parallel to } \Gamma \tag{121}$$

where $a(x)\nabla m(x)$ is the vector

$$(\textstyle\sum_{j=1}^{n} a_{1,j}(x)\partial m/\partial x_j, \ldots, \sum_{j=1}^{n} a_{n,j}(x)\partial m/\partial x_j)^T$$

Using (59), (121) can also be written as

$$\begin{aligned}\{\varphi(z) : z \in [0,1]\} = \text{curve } \Gamma \text{ such that } & b(x) \\ & - \text{ div } (a(x)) \text{ along } \Gamma \text{ is parallel to } \Gamma\end{aligned} \tag{122}$$

where $\text{div}(a(x))$ is the vector with components

$$(\text{div}(a(x)))_i = \sum_{j=1}^{n} \frac{\partial a_{ij}(x)}{\partial x_j} \tag{123}$$

Of course, even if the localized tube assumption is valid (and hence the tube can be identified by the adjustment procedure described at the end of Sect. 6.1), since (119) may not be consistent with being the location of a streamline in the tube, there is no guarantee that this curve will lead to

regions $D(z)$ defined in the planes $P(z)$ via (65) and an associated bundle of curves Γ_x via (70) such that both (73) and (74) are valid. In particular, if $m(x)$ has many critical points (local minima, maxima, etc.) (119) may lead to a curve which is too wiggly, in which case the potential regions $Q(z) \subset P(z)$ where the bundle lives (see (72)) may be too small to contain the regions $D(z) \subset P(z)$. But again, this does not mean that the localized tube assumption is invalid, simply that the a priori estimate (119) is to restrictive. How to improve upon (119) is explained in Sect. 7.2.

For now, however, let us consider a situation where (119) is valid and the problem reduces to determining the curve Γ satisfying (122). To do so, let Γ_0 be an initial guess for Γ. We wish to find a sequence of curves, Γ_0, Γ_1, etc. such that Γ_n converges towards the solution of (122) as $n \to \infty$. To write the iterative procedure which gives Γ_{n+1} from Γ_n, it is convenient to represent this curve parametrically as $\Gamma_n = \{\phi_n(\alpha) : \alpha \in [0,1]\}$ where α is some parameter not necessarily related to the parameter z used in $\varphi(z)$ in (122). How to choose and fix the parametrization of Γ_n will be explained in a moment, but for time being it suffices to recall that, once $\Gamma = \lim_{n\to\infty} \Gamma_n$ is found, no matter how it is parametrized, the actual $\varphi(z)$ entering (122) can be obtained from $\phi(\alpha) = \lim_{n\to\infty} \phi_n(\alpha)$ by the procedure explained in Sect. 6.2 and k_{AB} and Z_{AB} by the procedure explained in Sect. 6.3.

Now recall that (122) is the combination of (70) and (120). This suggests to use the following iterative procedure. Given $\Gamma_n = \{\phi_n(\alpha) : \alpha \in [0,1]\}$, construct the planes $P_n(\alpha)$ with unit normal $\hat{n}_n(\alpha)$ satisfying (this is the equivalent of (70))

$$\hat{n}_{n,i}(\alpha) \;\; \text{parallel to} \;\; \sum_{j=1}^{n} a_{ij}^{-1}(\phi_n(\alpha)) \frac{\partial \phi_{n,j}(\alpha)}{\partial \alpha} \tag{124}$$

where $a_{ij}^{-1}(x)$ are the entries of the inverse of the diffusion matrix $a(x)$. Next find the maximum of $m(x)$ in each $P_n(\alpha)$, which can be done e.g. by running the dynamical equation

$$\dot{x}_i^{\alpha}(t) = \frac{\partial m(x^{\alpha}(t))}{\partial x_i} - \hat{n}_{n,i}(\alpha) \sum_{j=1}^{n} \hat{n}_{n,j}(\alpha) \frac{\partial m(x^{\alpha}(t))}{\partial x_j}, \qquad x^{\alpha}(0) = \varphi_n(\alpha) \tag{125}$$

The right hand side of this equation is the component in $P_n(\alpha)$ of the vector $\nabla m(x) = (\partial m/\partial x_1, \ldots, \partial m/\partial x_n)^T$, and the steady state solution of (125) is where this component is zero, i.e. $\nabla m(x)$ is parallel to $\hat{n}_n(\alpha)$, consistent with (120). Now construct the curve Γ_{n+1} out of the solutions of (125), i.e. take $\phi_{n+1}(\alpha) = x^{\alpha}(\Delta t)$ for some $\Delta t > 0$. If $\Gamma_{n+1} = \Gamma_n$ within required accuracy stop; if $\Gamma_{n+1} \neq \Gamma_n$ repeat. This iteration can be shown to converge, at least if the initial curve Γ_0 is close enough to the actual one satisfying (122) and Δt is small enough. Finally, once this curve has been identified, construct the associated regions $D(z)$ via (65), $Q(z)$ via (72) and the associated bundle of curves via (70). Then check whether (73) and (74) are satisfied. In practice,

the regions $D(z)$ and $Q(z)$ (or part of the latter at least) can be identified by running some constrained version of (19) such that $x(t)$ stays in the plane and samples the density $m(x)$ restricted to this plane (which can be done by generalizing the idea of the blue moon sampling [6, 29] to (19), see [4]): $D(z)$ is then the region in which the constrained trajectory is most likely to stay, and each region $D(z)$ is contained in the region $Q(z)$ if the region $D(z)$ do not intersect with any other region $D(z')$, $z' \neq z$. If (73) and (74) are not satisfied, then either the localized tube assumption is invalid or the curve defined by (122) is not a streamline in the tube (the latter reason of failure can be excluded or eliminated by using (70) and (127) instead of (122), see below).

The above iterative procedure is straightforward to implement except maybe for the following catch. In practice the curves Γ_n must be discretized, i.e. one does not work with $\{\phi_n(\alpha) : \alpha \in [0,1]\}$ but rather with a collection of representative points or *images* along this curve, $\{\phi_n(\alpha_j) : j = 1, \ldots, J\}$ for some $J \in \mathbb{N}$. If we naively apply the iterative procedure outlined above to these discretized curves, then the representative points $\{\phi_n(\alpha_j) : j = 1, \ldots, J\}$ may end up clustering in certain regions along the continuous curve while leaving other regions along this curve under-resolved. This may lead to numerical instabilities. This difficulty is easy to solve: at each iteration, we can interpolate a continuous curve along the images $\{\phi_n(\alpha_j) : j = 1, \ldots, J\}$, then redistribute new images evenly along this interpolated curve and restart with the planes associated with the redistributed images rather than the original ones. This way, it is guaranteed that the images always remain evenly spaced and no numerical instability arises. The iterative procedure just described, including the reparametrization step is, in essence, the *zero-temperature string method*, introduced in [9,26] and further developed and used in [10,27]. [7] The zero-temperature string method is a generalization of the nudged elastic band method [19,21].

7.2 The Finite Temperature String Method

Suppose now that (122) is a curve leading to regions $D(z)$ defined in the planes $P(z)$ via (65) and an associated bundle of curves Γ_x via (70) such that (73) and (74) are not satisfied. This means that the a priori estimate (119) is not consistent with being the location of a streamline and must be improved upon. A second natural a priori estimate for a streamline is to take the mean of $m(z)$ restricted to $C(z)$, i.e

$$\varphi(z) = Z_{C(z)}^{-1} \int_{C(z)} x m(x) d\sigma_{S(z)}(x) \tag{126}$$

[7] The zero-temperature string method was primarily formulated as a time-continuous evolution of a curve $\Gamma(t)$ rather than the time-discrete evolution as explained here, but this difference is not essential.

where $Z_{C(z)} = \int_{C(z)} m(x) d\sigma_{C(z)}(x)$. Since $P(z)$ approximates $S(z)$ inside the tube $\cup_{z\in[0,1]} C(z)$, from (63) it follows that the equation above can also be written as

$$\varphi(z) \approx Z_{P(z)}^{-1} \int_{P(z)} x m(x) d\sigma_{P(z)}(x) \tag{127}$$

Again, there is no guarantee that (127) will lead to regions $D(z)$ defined in the planes $P(z)$ via (65) and an associated bundle of curves Γ_x via (70) such that (73) and (74) are satisfied. If (127) fails to do the job because this curve is too wiggly again, then other guesses should be used that lead to smoother curves, or one should go back to the full procedure based on moving a whole bundle of curves as outlined in Sect. 6.1 (which, by the way, has not been done yet).

Assuming that (126) is valid, it is possible to determine the curve Γ satisfying (70) and (127) instead of (122) via a straightforward generalization of the iterative procedure described in Sect. 7.1. Instead of using (125) to find the maximum of $m(x)$ in each $P_n(\alpha)$, one uses a constrained version of (19) to sample $m(x)$ and get the mean of this density in each plane. The new curve $\Gamma_{n+1} = \{\phi_{n+1}(\alpha) : \alpha \in [0,1]\}$ is then constructed by taking for $\phi_{n+1}(\alpha)$ a weighted average between $\phi_n(\alpha)$ and the mean position in $P_n(\alpha)$. Except for this modification, the iterative procedure is conducted as above, including the reparametrization step at each iteration, and in essence it is the *finite-temperature string method* introduced in [11] and further developed in [12,28]. At the end of the iteration, when the curve satisfying both (70) and (127) has been identified, one has again to check whether it leads to regions $D(z)$ defined in the planes $P(z)$ via (65) and an associated bundle of curves Γ_x via (70) such that (73) and (74) are satisfied. If not, one has to go into other procedures to estimate a priori the location of Γ inside the tube, e.g. by smoothing the curve defined by (127). Eventually, if none of these work, we may have to conclude that the localized tube assumption is invalid, at least in the original variables, and look for collective variables in which it may be satisfied.

7.3 The String Method in Collective Variables

To generalize the zero- or the finite-temperature string methods to work in collective variables (Sects. 6.4 and 6.5), it suffices to do the following. At each iteration, it is now required to estimate the tensor $A_{kl}(\theta)$ defined in (94) and determine the gradient of the density $M(\theta)$ defined in (93) to be used in the equivalent of (125) (in the zero-temperature version) or sample with respect to this density to identify the equivalent of (127) (in the finite-temperature version). This requires to introduce an additional loop of sampling of blue moon type, now in the original ensemble with density $m(x)$ under the additional constraint that $\theta(x) = \theta^*$. For details, see [23].

8 Outlook

In this chapter, we have shown why the recent transition path theory (TPT) offers the correct probabilistic framework to understand the mechanism by which rare events occur by analyzing the statistical properties of the reactive trajectories involved in these events. The main results of TPT are the probability density of reactive trajectories and the probability current (and associated streamlines) of reactive trajectories, which also allows one to compute the probability flux of these trajectories and the rate of the reaction. It was also shown that TPT is a constructive theory: under the assumption that the reaction channels are local, TPT naturally leads to algorithms that allow to identify these channels in practice and compute the various quantities that TPT offers.

Acknowledgments

I am grateful to Weinan E and Weiqing Ren: the results reported here are part of a joint research project with them. I also thank very warmly Giovanni Ciccotti who helped me to clarify things when I was confused (and often unaware of it) and to whom the presentation owns a lot (the good parts anyway; I am sole responsible for the rest). I also thank David Chandler, Ron Elber, Christof Schuette, Alexander Fischer, Luca Maragliano, Phillip Metzner and Paul Maragakis for many interesting discussions. This work was partially supported by NSF grants DMS02-09959 and DMS02-39625, and by ONR grant N00014-04-1-0565.

References

Au:In Ref 1, please provide page No. if any.

Au: Please provide titles for these Refs. 2–12, 14–17, 19, 20, 22, 24–34

1. C. H. Bennett (1977) In *Algorithms for Chemical Computation*, eds. A. S. Nowick and J. J. Burton ACS Symposium Series No. 46, **63**
2. R. B. Best, and G. Hummer (2005) *Proc. Natl. Acad. Sci. USA* **102**, p. 6732
3. P. G. Bolhuis, D. Chandler, C. Dellago, and P. Geissler (2002) *Ann. Rev. Phys. Chem.* **59**, p. 291
4. G. Ciccotti, R. Kapral, and E. Vanden-Eijnden (2005) *Chem. Phys. Chem.* **6**, p. 1809
5. D. Chandler (1978) *J. Chem. Phys.* **68**, p. 2959
6. E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral (1989) *Chem. Phys. Lett.* **156**, p. 472
7. C. Dellago, P. G. Bolhuis, and P. L. Geissler (2002) *Advances in Chemical Physics* **123**, p. 1
8. R. Durrett (1996) *Stochastic Calculus.* CRC Press
9. W. E, W. Ren and E. Vanden-Eijnden (2002) *Phys. Rev. B* **66**, p. 052301
10. W. E, W. Ren and E. Vanden-Eijnden (2003) *J. App. Phys.* **93**, p. 2275
11. W. E, W. Ren and E. Vanden-Eijnden (2005) *J. Phys. Chem. B* **109**, p. 6688

12. W. E, W. Ren and E. Vanden-Eijnden (2005) *Chem. Phys. Lett.* **413**, p. 242
13. W. E and E. Vanden-Eijnden (2004) in: *Multiscale Modelling and Simulation*, eds. S. Attinger and P. Koumoutsakos (LNCSE **39**, Springer Berlin Heidelberg
14. W. E and E. Vanden-Eijnden, to appear in *J. Stat. Phys.*
15. R. Elber, A. Ghosh, and A. Cárdenas (2002) *Account of Chemical Research* **35**, p. 396
16. R. Elber, A. Ghosh, A. Cárdenas, and H. Stern (2003) *Advances in Chemical Physics* **126**, p. 93
17. H. Eyring (1935) *J. Chem. Phys.* **3**, p. 107
18. C. W. Gardiner (1997) Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences, Springer Berlin Heidelberg
19. G. Henkelman and H. Jónsson (2000) *J. Chem. Phys.* **113**, p. 9978
20. G. Hummer (2004) *J. Chem. Phys.* **120**, p. 516
21. H. Jónsson, G. Mills, and K. W. Jacobsen (1998) in: *Classical and Quantum Dynamics in Condensed Phase Simulations*, ed. by: B. J. Berne, G. Ciccoti, and D. F., Coker, World Scientific
22. A. Ma, A. R. Dinner (2005) *J. Phys. Chem. B* **109**, p. 6769
23. L. Maragliano, A. Fischer, E. Vanden-Eijnden and G. Ciccotti, submitted to: *J. Chem. Phys.*
24. G. Papanicolaou (1976) *Rocky Mountain Math. J* **6**, p. 653
25. L. R. Pratt (1986) *J. Chem. Phys.* **9**, p. 5045
26. W. Ren (2002) *Numerical Methods for the Study of Energy Landscapes and Rare Events.* PhD thesis, New York University
27. W. Ren (2003) *Comm. Math. Sci.* **1**, p. 377
28. W. Ren, E. Vanden-Eijnden, P. Maragakis, and W. E. (2005) *J. Chem. Phys.* **123**, p. 134109
29. M. Sprik and G. Ciccotti (1998) *J. Chem Phys.* **109**, p. 7737
30. F. Tal and E. Vanden-Eijnden (2006) *Nonlinearity* **19**, p. 501
31. E. Vanden-Eijnden and F. Tal (2005) *J. Chem. Phys.* **123**, p. 184103
32. T. S. van Erp and P. G. Bolhuis (2005) *J. Comp. Phys.* **205**, p. 157
33. E. Wigner (1938) *Trans. Faraday Soc.* **34**, p. 29
34. T. Yamamoto (1960) *J. Chem Phys.* **33**, p. 281