A MATHEMATICAL THEORY OF OPTIMAL MILESTONING (WITH A DETOUR VIA EXACT MILESTONING)

LING LIN, JIANFENG LU, AND ERIC VANDEN-EIJNDEN

ABSTRACT. Milestoning is a computational procedure that reduces the dynamics of complex systems to memoryless jumps between intermediates, or milestones, and only retains some information about the probability of these jumps and the time lags between them. Here we analyze a variant of this procedure, termed optimal milestoning, which relies on a specific choice of milestones to capture exactly some kinetic features of the original dynamical system. In particular, we prove that optimal milestoning permits the exact calculation of the mean first passage times (MFPT) between any two milestones. In so doing, we also analyze another variant of the method, called exact milestoning, which also permits the exact calculation of certain MFPTs, but at the price of retaining more information about the original system's dynamics. Finally, we discuss importance sampling strategies based on optimal and exact milestoning that can be used to bypass the simulation of the original system when estimating the statistical quantities used in these methods.

1. INTRODUCTION

Relying on the enormous power of modern computing technologies, with advances such as special purpose high-performance computers, high-performance graphical processing units (GPUs), massively parallel simulations, etc. scientific computing has been playing an ever growing role as a tool to study complex systems and analyze their dynamics at an unprecedented level of details. Molecular dynamics (MD) simulations, for example, can nowadays be used to probe the function of large biomolecules and other complex molecular systems at spatio-temporal scales that are beyond experimental reach, thereby opening the door to a first-principle understanding of these systems. Similarly, general circulation and global climate models (GCMs) are used to simulate the dynamics of the coupled atmosphere/ocean system at ever higher resolutions and are responsible for the increasing accuracy of weather forecasting and climate change predictions. These advances do not come without challenges, however. The dynamics of complex systems often involve complicated activated processes, such as reactive events arising in kinetic phase transitions, conformational change of macromolecules, or regime changes in climate. These processes require the system to cross over (free) energy barriers or make long diffusive transitions, and they occur on very long time scales that even today are difficult to reach by brute-force numerical simulations. On top of this, bare simulation data in these systems are typically very large and intricate, and therefore hard to analyze. These difficulties call for the development of analytical and computational techniques to (i) identify quantities that characterize the essential features of system's kinetics at a coarser level, and (ii) accelerate the calculation of these quantities via techniques that bypass the brute-force simulation of the original system.

Date: March 14, 2017.
The work of J.L. is partially supported by the National Science Foundation under grant DMS-1454939.

The milestoning method, originally introduced by Elber in [17] and further developed e.g. in [28, 38, 16, 36, 33, 3] is an approach that aims at achieving both these objectives. The main idea behind milestoning is to reduce the overall system's dynamics to transition events between intermediates, or milestones, in its phase-space. Milestoning assumes that such transitions are memoryless, and retains only the information about the probabilities that a given milestone will be reached first after another, and the mean (more generally, the distribution) of the lag-times between these transitions. These statistical quantities may for instance be estimated from large amounts of simulation data, so that milestoning can be viewed as a data-processing tool and used to analyze long time series from numerical simulations. Alternatively, these quantities can be sampled efficiently by running several local simulations independently. In this way the method is akin to an importance sampling technique and can be used to accelerate the numerical simulations by generating sub-trajectories directly in regions of low probability rather than having to wait a long time until an unbiased trajectory visits these regions.

Even though milestoning has become quite a popular method by now, work remains to be done to give it a rigorous mathematical foundation. The main objective of this paper is to contribute to this effort. One of the main issues is to determine under which conditions the coarse-grained description of milestoning still retains useful and accurate kinetic information about the original system. For example, one is often interested in the mean time the dynamics takes to go from one region of its phase space to another. In the context of milestoning, this amounts to asking what is the mean first passage time (MFPT) from one milestone to another one far away, after many transitions via other milestones in between. Does milestoning permit the accurate estimation of such MFPTs? Clearly, one cannot expect it to be the case unless certain conditions about the system dynamics and/or the milestones are met, and this has led to two main routes of justification of the method.

The first is to restrict oneself to systems whose dynamics is metastable, i.e. such that we can identify 'hubs' in its phase-space that the system visits often but between which it seldom transitions. Under appropriate assumption, the transition between these hubs can then be approximately described by a Markov jump process, which justifies the milestoning description if the hubs are used as milestones [29, 18]. The mathematical justification of this picture relies on tools from spectral theory [8, 9, 30] and potential theory [4, 5, 6] that have been used to analyze metastability, and we will consider it in a forthcoming publication. In the present paper, we will instead focus on another route that has been proposed to justify milestoning. This route is based on the observation, originally made in [36], that *there exists a particular way to pick the milestones such that the method permits the exact calculation of MFPTs, regardless on whether its dynamics is metastable or not.* The version of milestoning that involves this particular choice of milestones was termed *optimal milestoning* in [36], and our purpose here is to justify it rigorously. In the process of doing so, we will also discuss another variant of milestoning, the so-called *exact milestoning* [3, 2], which also permits the exact calculation of certain MFPTs but at the price of retaining more information about the original system's dynamics, namely the exact location at which the process reaches a milestone first after hitting another – in this sense exact milestoning is somewhat closer in spirit to methods such as forward flux sampling (FFS) [1, 31] or transition interface sampling (TIS) [26], and even more so to non-equilibrium umbrella sampling methods [37, 7] such as trajectory parallelization and tilting [35], than to the original milestoning method. Finally, we will also discuss how to accelerate the sampling of the statistical quantities needed in optimal milestoning (and in

exact milestoning too). Let us remark that our analysis of optimal milestoning is connected to the study of coarse-graining without timescale separation performed in [21].

The remainder of this paper is organized as follows. In Sec. 2 we start by formulating the set-up of milestoning that we will study (Sec. 2.1) and then list our main results regarding the exact calculation of MFPTs within optimal milestoning (Sec. 2.2). In Sec. 3 we make a detour by exact milestoning, as this discussion will allow us to better understand what the realizability of optimal milestoning entails, in particular in terms of the existence of an invariant family of distributions on the milestones. In Sec. 4, the existence of these distributions is discussed in detail, and these results are then used in Sec. 5 to explain why and how MFPTs can be calculated exactly within optimal milestoning. In Sec. 6 we go on discussing how to accelerate the sampling of the key statistical quantities optimal milestoning relies upon. Sec. 7 gives a few concluding remarks, and several appendices contain the proofs of our more technical results.

2. Set-up and main results

2.1. **Set-up.** We shall focus on situations where the original process is a diffusion on $\mathbb{R}^d$ with infinitesimal generator $\mathcal{L}$ whose action on a test function $f : \mathbb{R}^d \to \mathbb{R}$ is given by

$$(\mathcal{L}f)(x) = \nabla \cdot (a(x)\nabla f(x)) + b(x) \cdot \nabla f(x), \tag{1}$$

where $b(x) \in \mathbb{R}^d$ and $a(x) \in \mathbb{R}^{d\times d}$ is symmetric and positive definite for every x, such that the operator $\mathcal{L}$ is uniformly elliptic. We assume that this diffusion is positive recurrent and possesses a unique invariant distribution with density $\rho(x) > 0$ satisfying $\mathcal{L}^*\rho = 0$, where

$$0 = (\mathcal{L}^*\rho)(x) = \nabla \cdot (a(x)\nabla\rho(x) - b(x)\rho(x)) \,. \tag{2}$$

($\mathcal{L}^*$ is the formal adjoint operator of $\mathcal{L}$.) Note that we do not assume microscopic reversibility (a.k.a. detailed-balance), i.e. $a(x)\nabla\rho(x) - b(x)\rho(x) \neq 0$ in general.

The operator $\mathcal{L}$ is the generator of the Itô stochastic differential equation (SDE)

$$dX(t) = \big(b(X(t)) + \operatorname{div} a(X(t))\big)dt + \sqrt{2}\sigma(X(t))dW(t), \tag{3}$$

where $\sigma(x)$ satisfies $(\sigma\sigma^T)(x) = a(x)$ and $W(t)$ denotes the standard Brownian motion in $\mathbb{R}^d$. We will denote by $X \equiv \{X(t)\}_{t\in\mathbb{R}_+}$ a sample path of this process on $\mathbb{R}_+ = [0, \infty)$, obtained by solving (3) with some initial condition at $t = 0$.

Remark 2.1. *An important example of SDE of the from* (3) *is the overdamped Langevin equation for a particle with position $Q(t) = X(t)$ moving in a potential $V : \mathbb{R}^d \to \mathbb{R}$ and subject to thermal effect at inverse temperature β:*

$$\gamma M dQ(t) = -\nabla V(Q(t))dt + \sqrt{2\beta^{-1}\gamma M}\, dW(t),$$

where γ denotes the friction tensor and M the mass matrix. On the other hand, the inertial Langevin equation for $X(t) = (Q(t), P(t))$,

$$\begin{aligned} \dot{Q}(t) &= M^{-1}P(t) \\ dP(t) &= -\nabla V(X(t))dt - \gamma P(t)dt + \sqrt{2\beta^{-1}\gamma M}\, dW(t), \end{aligned}$$

is not of the form (3) *because its generator is hypoelliptic. This equation is important in view e.g. of its applications to molecular dynamics. We believe that most of the results listed below apply to it after minor modifications, but the proofs would have to be modified.*

To introduce the coarse-grained description used in milestoning, we begin by defining two key quantities:

Definition 2.2 (Milestones)**.** *Let $A_0 \subset A_1 \subset \cdots \subset A_N \subset \mathbb{R}^d$ be a finite collection of nested connected sets with smooth boundaries. We call the boundaries of these sets the* milestones*, $M_i = \partial A_i$, $i \in I \equiv \{0, 1, \ldots, N\}$. We will refer to I as the* index set *of the milestones, and denote $\mathcal{M} = \{M_i : i \in I\}$.*

Definition 2.3 (First and last hitting times)**.** *For any subset $E \subset \mathbb{R}^d$, we define the* first hitting time *of E after time t as*

$$H_E^+(t) = \inf\{s \geq t : X(s) \in E\}$$

and the last hitting time *of E before time t as*

$$H_E^-(t) = \sup\{s \leq t : X(s) \in E\}.$$

The coarse-grained description of the trajectory used in milestoning can now be specified in terms of these objects as follows:

Definition 2.4 (Milestoning index process)**.** *Let $\mathcal{M} = \{M_i : i \in I\}$ be a set of milestones. Define $\tau_0 = H^+_{\cup_i M_i}(0)$. For each $t \geq \tau_0$, define $\Xi(t)$ to be the index of the last milestone hit by $X(t)$, i.e.,*

$$\Xi(t) = \text{ index } i \in I \text{ such that } X\left(H^-_{\cup_i M_i}(t)\right) \in M_i.$$

The process $\{\Xi(t) : t \geq \tau_0\}$ is called the milestoning index process *associated with $\mathcal{M}$.*

Remark 2.5. *Note that the trajectory of the milestoning index process $\{\Xi(t) : t \geq \tau_0\}$ is a piecewise constant function taking values in the index set I which jumps from one value to another whenever the original trajectory X reaches a new milestone. Due to the way we defined the milestones (see Definition 2.2), these jumps can only be ± 1. Note also that consecutive hits of the same milestone without hitting another one in between will not change the value of $\Xi(t)$.*

It will also be useful to decompose the milestoning index process into its temporal and spatial components:

Definition 2.6. *Let $\{\Xi(t) : t \geq \tau_0\}$ be the milestoning index process associated with a set of milestones $\mathcal{M} = \{M_i : i \in I\}$. Set $\xi_0 = \Xi(\tau_0)$ and define recursively for $n \geq 1$,*

$$\tau_n = \inf\{t \geq \tau_{n-1} : \Xi(t) \neq \xi_{n-1}\},$$

and

$$\xi_n = \Xi(\tau_n).$$

The sequence $\{(\xi_n, \tau_n) : n \in \mathbb{N}_0\}$ is called the coarse-grained milestoning chain associated with $\mathcal{M}$*. In addition, the sequence $\{\xi_n : n \in \mathbb{N}_0\}$ is called the* skeleton *of the milestoning index process.*

Remark 2.7. *Thus the skeleton $\{\xi_n : n \in \mathbb{N}_0\}$ of the milestoning index process gives the indices of successive milestones that the original trajectory X hits and the sequence of jump times $\{\tau_n : n \in \mathbb{N}_0\}$ records the first times at which these successive milestones are hit.*

In the sequel we will denote the lags between the jump times as

$$\alpha_n = \tau_n - \tau_{n-1}, \qquad n \in \mathbb{N}. \tag{4}$$

2.2. **Main results.** Having introduced the main objects used in milestoning, we now ask what kind of kinetic information about the original process we can extract from them. We will focus here on the mean first passage time (MFPT) $T_{i,j}$ from M_i to M_j for any $i, j \in I$ with $i \neq j$ – other quantities of interest include the probability that, starting from milestone M_i, milestone M_j will be hit before milestone M_k, or the invariant distribution π_i giving the stationary probability that the last milestone hit was M_i, etc. Our analysis below will indicate how to calculate these quantities as well.

The MFPT $T_{i,j}$ can be defined directly from a sampling view point as follows: First, employ a subset $\mathcal{M}^{(i,j)} \subset \mathcal{M}$ consisting of only the two milestones M_i and M_j. Second, introduce as in Definition 2.6 a coarse-grained sequence $\{(\xi_n^{(i,j)}, \tau_n^{(i,j)}) : n \in \mathbb{N}_0\}$ associated with $\mathcal{M}^{(i,j)} = \{M_i, M_j\}$. Then set

$$T_{i,j} = \lim_{n\to\infty} \frac{\sum_{p=1}^{n} \alpha_p^{(i,j)} \delta_{i,\xi_{p-1}^{(i,j)}}}{\sum_{p=1}^{n} \delta_{i,\xi_{p-1}^{(i,j)}}}, \tag{5}$$

where $\alpha_p^{(i,j)} = \tau_p^{(i,j)} - \tau_{p-1}^{(i,j)}$. (5) estimates from a long ergodic trajectory the average time it takes to go from the set M_i to M_j after each return to M_i from M_j. We will prove below that this limit exists and give an expression for it in terms of the quantities used in milestoning that is exact under certain conditions. Specifically, we will show that under these conditions (to be specified in a moment) $T_{i,j}$ satisfies the linear system

$$T_{i,j} = t_i + \sum_{k\in I} p_{i,k} T_{k,j}, \qquad i \in I \quad i \neq j \tag{6}$$

with the boundary condition $T_{j,j} = 0$. Here t_i is the average time the trajectory is associated with the ith milestone, which can be defined empirically as

$$t_i = \lim_{n\to\infty} \frac{1}{n} \int_{\tau_0}^{\tau_n} \delta_{i,\Xi(t)} dt, \tag{7}$$

and p_{ij} is the probability that the trajectory hits the jth milestone after leaving the ith milestone, which can be defined as

$$p_{i,j} = \frac{\sum_{p=1}^{n} \delta_{i,\Xi(\tau_{p-1})} \delta_{j,\Xi(\tau_p)}}{\sum_{p=1}^{n} \delta_{i,\Xi(\tau_{p-1})}} \tag{8}$$

We prove that the system (6) gives the exact MFPT *iff* the milestones are chosen to be level sets of the backward committor function, defined as follows:

Definition 2.8 (Backward committor functions)**.** *Let A and B be two non-overlapping bounded closed sets of $\mathbb{R}^d$, each of which is the closure of a nonempty, simply connected, open set. The* backward committor function *is the classical solution to the Dirichlet problem:*

$$\begin{cases} \mathcal{L}^\dagger q^- = 0 & \text{in } \mathbb{R}^d \setminus (A \cup B), \\ q^-|_A = 1, \quad q^-|_B = 0, \end{cases} \tag{9}$$

where

$$(\mathcal{L}^\dagger q^-)(x) = \nabla \cdot \big(a(x)\nabla q^-(x)\big) - b(x) \cdot \nabla q^-(x) + \frac{2}{\rho(x)} \langle \nabla \rho(x), a(x) \nabla q^-(x) \rangle$$

is the generator of the time-reversed diffusion process, and $\langle , \rangle$ denotes the standard inner product on $\mathbb{R}^d$. (Note that $\mathcal{L}^\dagger \neq \mathcal{L}$ in general, since we do not assume microscopic reversibility.)

Specifically, we prove that (6) is exact if we pick any A and B consistent with Definition 2.8 and set $M_i = \{x : q^-(x) = z_i\}$ for $i \in I$ with any $1 \geq z_0 > z_1 > \cdots > z_N \geq 0$ – that is, use isocommittor surfaces as milestones. We call *optimal milestoning* the method used with such a set of milestones.

Observe the backward committor function $q^-(x)$ is the probability that the trajectory associated with the time-reversed diffusion process will hit A first rather than B. Roughly speaking, it is the probability that the trajectory $X(t)$ located at x at time 0 came from the set A rather than B. The backward committor function (along with the forward one) plays a central role in Transition Path Theory (TPT) [13, 24, 32, 14, 25, 15, 19], and this framework will also prove essential in our analysis.

The remainder of this paper is devoted to make the statements above rigorous. As pointed out before, our analysis of optimal milestoning is connected to the study of coarse-graining without timescale separation in [21]. In fact, the latter can be viewed as using all the isocommittor surfaces as a continuous family of milestones, while the optimal milestoning, as shown in this work, is chosen as a (discrete) collection of isocommittor surfaces. In some sense, the coarse-graining proposed in [21] amounts to a continuous limit of optimal milestoning.

3. Warm-up: exact milestoning

To understand better what the validity of (6) entails, it is useful to consider first a variant of milestoning, termed *exact milestoning* [2, 3], in which more information about the process is kept than in optimal milestoning. Specifically, exact milestoning uses the *first hitting chain* defined as follows:

Definition 3.1 (First hitting chain)**.** *Given a set of milestones $\mathcal{M} = \{M_i : i \in I\}$, let $\{(\xi_n, \tau_n) : n \in \mathbb{N}_0\}$ be the coarse-grained milestoning chain associated with $\mathcal{M}$ introduced in Definition 2.6 and set $Y_n = X(\tau_n)$. The* first hitting chain *associated with $\mathcal{M}$ is the process $\{Y_n : n \in \mathbb{N}_0\}$.*

Thus, the index chain $\{\xi_n : n \in \mathbb{N}_0\}$ can be viewed as a coarse-grained sequence of the first hitting chain $\{Y_n : n \in \mathbb{N}_0\}$ in which one reduces the exact positions on the milestones to the indices of these milestones. The key observation, which immediately follows from strong Markovianity, is that:

Proposition 3.2. *Let $\mathcal{M} = \{M_i : i \in I\}$ be a set of milestones as in Definition 2.2 and $\{Y_n : n \in \mathbb{N}_0\}$ the first hitting chain associated with these milestones. Then $\{Y_n : n \in \mathbb{N}_0\}$ is a Markov chain with transition probability kernel*

$$\nu(x, B) = \mathbb{P}^x(X(\tau_1) \in B) = \mathbb{P}^x(Y_1 \in B), \tag{10}$$

where $\mathbb{P}^x$ denotes the probability conditional on $Y_0 = x$.

Note that, by construction, $\nu(x, B) = 0$ if $x \in M_i$ and $B \subset M_i$ since $Y_0 \in M_i$ implies that $Y_1 \notin M_i$. We will discuss in Sec. 6 how to sample $\nu(x, \cdot)$ in accelerated ways. Note also that, unlike $\{Y_n : n \in \mathbb{N}_0\}$, the index chain $\{\xi_n : n \in \mathbb{N}_0\}$ is not Markov, in general. What we show next is that we can compute the MFPT exactly if we allow ourselves to use $\{Y_n : n \in \mathbb{N}_0\}$ and the sequence of jump times $\{\tau_n : n \in \mathbb{N}_0\}$. This will also help us understand which property we need to require from the milestones in order that this exact computation be possible with $\{(\xi_n, \tau_n) : n \in \mathbb{N}_0\}$ instead.

The first hitting chain inherits ergodicity properties of the original process X. We state this result as:

Lemma 3.3. *Assume that the process X is positive recurrent. Then the first hitting chain $\{Y_n : n \in \mathbb{N}_0\}$ is positive recurrent as well, and its invariant measure μ satisfies*

$$\mu(\cdot) = \int_{\cup_i M_i} \mu(\mathrm{d}x)\nu(x,\cdot). \tag{11}$$

Note that the invariant measure μ is supported on the union of the milestones $\cup_i M_i$, and so it can be decomposed as

$$\mu(\cdot) = \sum_{i\in I} \pi_i \mu_i(\cdot), \tag{12}$$

where $\mu_i(\cdot)$ is supported on M_i and normalized so that $\mu_i(M_i) = 1$ by introducing

$$\pi_i = \mu(M_i), \qquad \sum_{i\in I} \pi_i = 1, \tag{13}$$

The distribution π_i gives the invariant probability distribution of the index chain $\{\xi_n : n \in \mathbb{N}_0\}$, and it is easy to derive an exact equation for it. To see how, start by decomposing

$$\nu(x, dy) = \sum_{j\sim i} P_{i,j}(x)\nu_{i,j}(x, \mathrm{d}y), \qquad x \in M_i, \quad i \in I \tag{14}$$

where $j \sim i$ denote the indices of the milestones adjacent to M_i (that is, $j = i + 1$ and $j = i - 1$ if $i = 1, \ldots, N - 1$, $j = 1$ if $i = 0$, and $j = N - 1$ if $i = N$), $P_{i,j}(x)$ is the conditional probability that, if $x \in M_i$, the next milestone to be reached will be M_j, i.e.

$$P_{i,j}(x) = \mathbb{P}^x(Y_1 \in M_j) = \nu(x, M_j), \qquad x \in M_i, \quad i \in I, \tag{15}$$

and $\nu_{i,j}(x,\cdot)$ is the transition probability kernel from $x \in M_i$ to M_j, conditional on hitting M_j next, i.e., $\nu_{i,j}(x, M_j) = 1$ for all $x \in M_i$. Note that $P_{i,j}(x)$ does depend on x in general (rather than only on $M_i \ni x$).

Using the decompositions (12) and (14) in (11) we obtain

$$\sum_{k\in I} \pi_k \mu_k(\cdot) = \sum_{j\in I} \pi_j \int_{M_j} \mu_j(\mathrm{d}x) \sum_{k\sim j} P_{j,k}(x)\nu_{j,k}(x,\cdot) \tag{16}$$

If we evaluate this equation on M_i, we arrive at the desired equation for π_i, a result we summarize as:

Proposition 3.4. *The invariant distribution π_i of the index chain $\{\xi_n : n \in \mathbb{N}_0\}$ is the solution to*

$$\pi_i = \sum_{j\sim i} \pi_j p_{j,i}, \tag{17}$$

where

$$p_{j,i} = \int_{M_j} \mu_j(\mathrm{d}x) P_{j,i}(x) \tag{18}$$

As we will see in Sec. 6 we can sample $p_{i,j}$ directly (i.e. without having to evaluate $\nu(x,\cdot)$ or even $P_{i,j}(x)$ beforehand) in accelerated ways.

Next we use these relations to calculate the MFPT from any $x \in \cup_{i\neq j} M_i$ to M_j. Denoting this MFPT by $T_j(x)$, it is defined as

Definition 3.5. *Given a set of milestones $\mathcal{M} = \{M_i : i \in I\}$ as Definition 2.2, the* mean first passage time (MFPT) *from $x \in \cup_{i\in I} M_i$ to M_j is given by:*

$$T_j(x) = \mathbb{E}\, t_j(x), \qquad t_j(x) = \inf\{t : X(t) \in M_j, X(0) = x \in \cup_{i\in I} M_i\}.$$

We have

Proposition 3.6. *The MFPT $T_j(x)$ satisfies the equation*

$$\tau(x) = T_j(x) - \int_{\cup_{i\in I} M_i} \nu(x, \mathrm{d}y) T_j(y), \qquad x \in \cup_{i\neq j} M_j, \tag{19}$$

with the boundary condition $T_j(x) = 0$ if $x \in M_j$. Here $\tau(x)$ denotes the average time the first hitting chain remains assigned to a milestone after hitting this milestone at location x; in the notation of Definition 2.6, it is

$$\tau(x) = \mathbb{E}(\alpha_1 | Y_0 = x). \tag{20}$$

We will skip the proof of this proposition, as it is similar to the proof of Lemma 4.6 (a rigorous version of (23)) below, using strong Markovianity and time homogeneity of the process.

Equation (19) is exact, but it obviously requires more information than (6), which is a closed equation for

$$T_{i,j} = \int_{M_i} \mu_i(dx) T_j(x). \tag{21}$$

It is clear that (6) cannot be derived from (19) without additional assumptions. Suppose, however, that the following property holds (this will be made more precise below in Definition 4.1):

$$\begin{aligned} & \frac{\int_{M_i} \mu_i(dx)\nu(x, B_j)}{\int_{M_i} \mu_i(dx)\nu(x, M_j)} = \mu_j(B_j), \qquad \forall B_j \subset M_j, \ \ \forall i \sim j \\ \Leftrightarrow \quad & \frac{\int_{M_i} \mu_i(dx)\nu(x, B_j)}{\int_{M_i} \mu_i(dx) P_{i,j}(x)} = \mu_j(B_j), \qquad \forall B_j \subset M_j, \ \ \forall i \sim j \\ \Leftrightarrow \quad & \int_{M_i} \mu_i(dx) P_{i,j}(x) \nu_{i,j}(x, \cdot) = p_{i,j}\mu_j(\cdot), \qquad \forall i \sim j. \end{aligned} \tag{22}$$

Intuitively, the above property means that, conditioned on hitting M_j, the push forward of the distribution μ_i by the transition kernel ν of the first hitting chain is given by μ_j. Assuming that (22) holds, we can average (19) with respect to $\mu_i(\cdot)$. Since $\int_{M_i} \mu_i(\mathrm{d}x)\tau(x) = t_i$ (defined in (7)) due to ergodicity, this gives

$$\begin{aligned} t_i &= T_{i,j} - \int_{M_i} \mu_i(\mathrm{d}x) \sum_{k\sim i} P_{i,k}(x) \int_{M_k} \nu_{i,k}(x, \mathrm{d}y) T_j(y) \\ &= T_{i,j} - \sum_{k\sim i} p_{i,k} \int_{M_k} \mu_k(\mathrm{d}y) T_j(y) \qquad \text{(by (22))} \\ &= T_{i,j} - \sum_{k\sim i} p_{i,k} T_{k,j}, \end{aligned} \tag{23}$$

which is precisely (6). In order words, (6) is exact if it is associated with milestones such that (22) holds. Theorem 4.2 below establishes that such milestones do indeed exists.

We remark that the exact milestoning [3] permits via solution of (19) to obtain the MFPTs $T_j(x)$ for any set of milestones consistent with Definition 2.2, but it requires to sample both the kernel $\nu(x, \cdot)$ (or at least expectations such as $\int_{M_k} \nu(x, \mathrm{d}y) T_j(y)$) and $\tau(x)$. Clearly, it is computationally more expensive to gather this information than that entering (19) – this is why optimal milestoning is more efficient. Still, it is posible to compute $\nu(x, \cdot)$ and $\tau(x)$, as explained in Sec. 6.1).

4. INVARIANT FAMILY OF DISTRIBUTIONS

The exact calculation of the MFPTs in optimal milestoning is based on the existence of invariant families of distributions, for which (22) hold. To introduce them precisely, let us define a shift operator $\mathcal{P}^*$ of a probability measures μ, associated with the first hitting chain $\{Y_n : n \in \mathbb{N}\}$ by

$$(\mathcal{P}^*\mu)(B) = \mathbb{P}_\mu\big[Y_1 \in B\big], \qquad B \subset \cup_{i\in I} M_i, \tag{24}$$

where we have used $\mathbb{P}_\mu$ to denote the law of the diffusion with the initial distribution μ. Note that the kernel associated with $\mathcal{P}^*$ is exactly $\nu(x,\cdot)$, the transition probability kernel of the first hitting chain used in (11), as

$$(\mathcal{P}^*\mu)(B) = \int_{\mathcal{M}} \mu(\mathrm{d}x)\nu(x,B). \tag{25}$$

We then have:

Definition 4.1 (Invariant family of distributions)**.** *A set of milestones $\mathcal{M} = \{M_i : i \in I\}$ is said to have an invariant family of distributions if there exists a family of probability measures $\{\mu_i : i \in I\}$ with each μ_i concentrated on M_i such that the conditional distribution of $\mathcal{P}^*\mu_i$ given M_j, if it makes sense, is μ_j. Such family $\{\mu_i : i \in I\}$ is called an invariant family of distributions associated with $\mathcal{M} = \{M_i : i \in I\}$.*

The next theorem proves that invariant families of distributions exist. Recall that q^- is the backward committor function defined in (9). Let us define a set of milestones $\mathcal{M} = \{M_i : 0 \le i \le N\}$ as

$$M_i = \{x \in \mathbb{R}^d \setminus (A \cup B)^\circ : q^-(x) = z_i\}, \quad \text{for } i \in I, \tag{26}$$

where $1 \ge z_0 > z_1 > \cdots > z_N \ge 0$. We assume that all the z_i's are regular values of q^-, i.e., the regularity condition $|\nabla q^-(x)| > 0$ holds for every $x \in \mathcal{M}$. We also assume that all the surface integrals

$$Z_i = \int_{M_i} \frac{\rho(x)}{|\nabla q^-(x)|} \langle a(x)\nabla q^-(x), \nabla q^-(x)\rangle \, \mathrm{d}\sigma_{M_i}(x), \quad i \in I \tag{27}$$

are finite. Then we are able to define on each M_i a probability measure μ_i with the density function

$$\rho_i(x) = Z_i^{-1} \frac{\rho(x)}{|\nabla q^-(x)|} \langle a(x)\nabla q^-(x), \nabla q^-(x)\rangle, \qquad i \in I \tag{28}$$

and we have:

Theorem 4.2. *Let $\mathcal{M} = \{M_i : i \in I\}$ be a set of milestones made of the backward isocommittor surfaces as in* (26) *satisfying the regular condition $|\nabla q^-(x)| > 0$ for every $x \in M$. Then the family of probability measures $\{\mu_i : i \in I\}$ with density ρ_i defined in (28) is an invariant family of distributions associated with $\mathcal{M}$. Actually we have*

$$\mathcal{P}^*\mu_i = \sum_{j\in I} q_{i,j}\mu_j, \quad \textit{for } i \in I \tag{29}$$

where $q_{i,j}$ is given by

$$q_{i,j} = \mathbb{P}_{\mu_i}\big[\xi_1 = j\big] = \begin{cases} \dfrac{z_i - z_{i+1}}{z_{i-1} - z_{i+1}} & \textit{if } j = i-1, \\ \dfrac{z_{i-1} - z_i}{z_{i-1} - z_{i+1}} & \textit{if } j = i+1, \\ 0 & \textit{otherwise}. \end{cases} \tag{30}$$

where we set $z_{-1} = -\infty$ *and* $z_{N+1} = +\infty$ *so that* $q_{0,1} = q_{N,N-1} = 1$.

We defer the proof of this theorem to the appendix. We remark that the optimal milestones are iso-surfaces of the backward committor function is due to the assignment of the milestoning index process to the last milestone the trajectory hit as in Definition 2.4.

Next we give four lemmas that list some properties of the invariant family of distributions associated with a set of milestones.

The first lemma states that an invariant family of distributions always forms an invariant distribution for the first hitting chain. The lemma justifies the construction in (12) where we obtain the normalized marginal distribution on M_i through the decomposition of the invariant measure.

Lemma 4.3. *Let* $\{\mu_i : i \in I\}$ *be an invariant family of distributions associated with a set of milestones* $\mathcal{M} = \{M_i : i \in I\}$. *Then there exists an invariant distribution* μ *for the first hitting chain* $\{Y_n : n \in \mathbb{N}_0\}$ *such that each* μ_i *is the conditional distribution of* μ *given* M_i.

Proof. By assumption and the formula of total probability, we obtain the following decomposition of $\mathcal{P}^*\mu_i$ for each $i \in I$,

$$\mathcal{P}^*\mu_i = \sum_{j\in I} p_{i,j}\mu_j,$$

where $p_{i,j} = (\mathcal{P}^*\mu_i)(M_j) = \mathbb{P}_{\mu_i}\big[\xi_1 = j\big]$. It is easy to check that this agrees with $p_{i,j}$ defined in (17). Let $(\pi_i)_{i\in I}$ be an invariant distribution for the transition matrix $(p_{i,j})_{i,j\in I}$. Then it is straightforward to check that $\mu = \sum_{i\in I}\pi_i\mu_i$ is an invariant distribution for $\{Y_n : n \in \mathbb{N}_0\}$. By definition, we have $\mu(M_i) = \pi_i$ and hence it is also consistent with (13). □

Remark 4.4. *While* $p_{i,j}$ *and* $q_{i,j}$ *defined in* (30) *are the same quantity, we reserve the notation* $q_{i,j}$ *for the transition probability when the milestones are chosen to be backward isocommittor surfaces, and hence the values are explicitly known as in* (30). *For general choice of milestones, we use* $p_{i,j}$ *instead.*

The second lemma justifies a key assumption made in milestoning for the set of milestones possessing an invariant family of distributions.

Lemma 4.5. *Suppose that* $\{\mu_i : i \in I\}$ *is an invariant family of distributions associated with a set of milestones* $\mathcal{M} = \{M_i : i \in I\}$. *Let* μ *be the corresponding invariant distribution for the first hitting chain, introduced in Lemma 4.3. Then under the law* $\mathbb{P}_\mu$, *the following properties hold.*

(i) *The index chain* $\{\xi_n : n \in \mathbb{N}_0\}$ *is a Markov chain;*

(ii) *For any* $n \geq 1$ *and* $i_k \in I$, $0 \leq k \leq n$,

$$\mathbb{E}_\mu\big[\alpha_n \mid \xi_k = i_k, 0 \leq k \leq n\big] = \mathbb{E}_\mu\big[\alpha_1 \mid \xi_0 = i_{n-1}, \xi_1 = i_n\big].$$

The third lemma is related to the exact calculation of the MFPTs and will be used to justify (23):

Lemma 4.6. *Assume that properties* (i) *and* (ii) *in Lemma 4.5 hold. Let* D *be the first time step* n *such that* $\xi_n = j$. *Define* $h_{i,j} = \mathbb{E}_\mu\big[\tau_D \big|\, \xi_0 = i\big]$. *Then* $(h_{i,j})_{i,j\in I}$ *is the unique solution to the following discrete Poisson problem*

$$\begin{cases} h_{i,j} = \displaystyle\sum_{k\in I} p_{i,k}t_{i,k} + \sum_{k\in I} p_{i,k}h_{k,j}, & i \in I, \quad i \neq j, \\ h_{j,j} = 0, \end{cases} \tag{31}$$

where $p_{i,j} = \mathbb{P}_\mu\big[\xi_1 = j \big|\, \xi_0 = i\big]$ *and* $t_{i,j} = \mathbb{E}_\mu\big[\alpha_1 \big|\, \xi_0 = i, \xi_1 = j\big]$.

The fourth lemma gives a restriction property of the invariant family of distributions:

Lemma 4.7. *Suppose that* $\{\mu_i : i \in I\}$ *is an invariant family of distributions associated with a set of milestones* $\mathcal{M} = \{M_i : i \in I\}$. *Then for any subset* $I' \subset I$, $\{\mu_i : i \in I'\}$ *is likewise an invariant family of distributions associated with the set of milestones* $\mathcal{M}' = \{M_i : i \in I'\}$.

We defer the proofs of these three lemmas to the appendix.

5. Exact calculation of mean first passage times in optimal milestoning

We prove in this section that optimal milestoning permits the exact calculation of mean first passage times $T_{i,j}$. Let us first define $T_{i,j}$ more rigorously by evaluating the limit in (5), which yields an equivalent probabilistic definition by ergodic theorem.

Proposition 5.1. *Assume that the first hitting chain* $\{Y_n^{(i,j)} : n \in \mathbb{N}_0\}$ *associated with* $\mathcal{M}^{(i,j)} = \{M_i, M_j\}$ *is uniquely ergodic and denote its unique invariant distribution by* $\mu^{(i,j)}$. *Then almost surely with respect to the law* $\mathbb{P}_{\mu^{(i,j)}}$, *the limit in the definition* (5) *of MFPT* $T_{i,j}$ *exists and can be expressed as*

$$T_{i,j} = \mathbb{E}_{\mu^{(i,j)}}\left[\alpha_1^{(i,j)} \middle| Y_0^{(i,j)} \in M_i\right] = \mathbb{E}_{\mu_i^{(i,j)}}\left[\alpha_1^{(i,j)}\right], \tag{32}$$

where $\mu_i^{(i,j)}$ *is the conditional distribution of* $\mu^{(i,j)}$ *given* M_i.

Proof. It is easily seen that $\{(\alpha_n^{(i,j)}, Y_n^{(i,j)}) : n \in \mathbb{N}_0\}$ and $\{Y_n^{(i,j)} : n \in \mathbb{N}_0\}$ are both Markov chains. Also note that there is a one-to-one correspondence between the invariant distributions of these two Markov chains and they are both induced from the law $\mathbb{P}_{\mu^{(i,j)}}$. In view of the ergodic theorem, we obtain almost surely with respect to $\mathbb{P}_{\mu^{(i,j)}}$,

$$T_{i,j} = \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{k=1}^n \alpha_{k+1}^{(i,j)} \mathbf{1}_{M_i}\big(Y_k^{(i,j)}\big)}{\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{M_i}\big(Y_k^{(i,j)}\big)} = \frac{\mathbb{E}_{\mu^{(i,j)}}\left[\alpha_1^{(i,j)} \mathbf{1}_{M_i}\big(Y_0^{(i,j)}\big)\right]}{\mathbb{P}_{\mu^{(i,j)}}\left[Y_0^{(i,j)} \in M_i\right]},$$

as asserted in (32). □

We are now ready to justify (6) rigorously.

Theorem 5.2. *Assume that* $\mathcal{M} = \{M_i : 0 \le i \le N\}$ *is a set of milestones made of the backward isocommittor surfaces as in (26) satisfying the regularity condition* $|\nabla q^-(x)| > 0$ *for every* $x \in \mathcal{M}$. *For each* $0 \le i \le N$, *let* μ_i *be the probability measure concentrated on* M_i *with the density* ρ_i *given in* (28). *Also assume that* $\{Y_n^{(i,j)} : n \in \mathbb{N}_0\}$ *is uniquely ergodic. Then the mean first passage times* $(T_{i,j})_{i,j\in I}$ *is the unique solution to the following discrete Poisson problem:*

$$\begin{cases} T_{i,j} = t_i + \displaystyle\sum_{k\in I} q_{i,k} T_{k,j}, & i \in I, \quad i \ne j, \\ T_{j,j} = 0. \end{cases} \tag{33}$$

where $q_{i,j}$ *is given by (30) and*

$$t_i = \sum_{j\in I} q_{i,j} \mathbb{E}_{\mu_i}\big[\alpha_1 \big| \xi_1 = j\big] = \mathbb{E}_{\mu_i}\big[\alpha_1\big].$$

Proof. Much of the work needed to prove this theorem has been done in proving the previous lemmas. By Theorem 4.2, Lemma 4.7, and the proof of Lemma 4.3, $\frac{1}{2}\mu_i + \frac{1}{2}\mu_j$ is an invariant distribution for the Markov chain $\{Y_n^{(i,j)} : n \in \mathbb{N}_0\}$. Note that $H^+_{M_j}(0) = D$ (recall D from Lemma 4.6) due to $\lim_{n\to\infty} \tau_n = \infty$ (see the proof of Lemma 4.7). So by Proposition 5.1,

$$T_{i,j} = \mathbb{E}_{\mu_i}\left[H^+_{M_j}(0)\right] = \mathbb{E}_{\mu}\left[D\middle|\, \xi_0 = i\right] = h_{i,j},$$

and the assertion follows immediately from Lemma 4.6. □

6. Accelerated sampling methods based on milestoning

As shown in Section 5, for diffusion processes we can calculate the MFPT exactly by using backward isocommittor surfaces as milestones. The only required quantities in this calculation are $p_{i,j} = \mathbb{P}_{\mu_i}\big[\xi_1 = j\big]$ and $t_i = \mathbb{E}_{\mu_i}\big[\alpha_1\big]$, where μ_i is the invariant family of distributions on the milestone M_i. In words, t_i is the average time a trajectory initiated on the ith milestone randomly from μ_i takes before it hits another milestone, and $p_{i,j}$ is the probability that the next milestone hit by this trajectory (other than the ith milestone) is the jth one.

In this section we address the question of how to sample these quantities. Note that this involves two practical issues: (i) how to sample $p_{i,j}$ and t_i based on short trajectories given a set of milestones, and (ii) assuming that we want to do optimal milestoning, how to pick milestones that approximate level sets of q^-. Below we will discuss these two issues separately, without necessarily assuming in (i) that we use an optimal set of milestones.

6.1. Sampling $p_{i,j}$ and t_i.

We start with the issue of how to compute $p_{i,j} = \mathbb{P}_{\mu_i}\big[\xi_1 = j\big]$ and $t_i = \mathbb{E}_{\mu_i}\big[\alpha_1\big]$ given an arbitrary set of milestones consistent with Definition 2.2. In principle, these quantities can be sampled by reinitializing short trajectories on each milestone M_i according to the distributions μ_i defined in (12) and running until each trajectory hits another milestone. This procedure is not easy to implement in practice, however, since it requires one to sample from μ_i, which we do not know *a priori* (recall that in this section we do not assume that the milestones are optimal, i.e. the density of μ_i is not given by (28) in general). In the original milestoning procedure, it was assumed that each μ_i can be approximated by the invariant distribution conditional on M_i, but the accuracy of this approximation is difficult to assess.

One way to get around this difficulty and estimate $p_{i,j} = \mathbb{P}_{\mu_i}\big[\xi_1 = j\big]$ and $t_i = \mathbb{E}_{\mu_i}\big[\alpha_1\big]$ directly in an unbiased way is to use a sampling strategy that bypasses the need of the reinitialization and thereby avoids the problem of having to know μ_i beforehand. In fact, as we will see below, this procedure permits to sample μ_i, as well as $\tau(x)$ and $\nu(x,\cdot)$, which is useful in the context of exact milestoning [3]. The key result behind this strategy is summarized in the following lemma that uses ergodicity:

Lemma 6.1. *We have*

$$\begin{aligned}
p_{i,j} &= \mathbb{P}_{\mu_i}\big[\xi_1 = j\big] = \mathbb{P}_{\mu}\big[\xi_1 = j \,\big|\, \xi_0 = i\big] \\
&= \lim_{n\to\infty} \frac{\sum_{k=1}^n \delta_{i,\xi_{k-1}}\delta_{j,\xi_k}}{\sum_{k=1}^n \delta_{i,\xi_{k-1}}} = \lim_{T\to\infty} \frac{N_{ij}^T}{N_i^T} = \lim_{T\to\infty} \frac{N_{ij}^T}{\sum_j N_{ij}^T}, \\
t_i &= \mathbb{E}_{\mu_i}\big[\alpha_1\big] = \mathbb{E}_{\mu}\big[\alpha_1\big|\, \xi_0 = i\big] \\
&= \lim_{n\to\infty} \frac{\sum_{k=1}^n \alpha_k\delta_{i,\xi_{k-1}}}{\sum_{k=1}^n \delta_{i,\xi_{k-1}}} = \lim_{T\to\infty} \frac{R_i^T}{N_i^T} = \lim_{T\to\infty} \frac{R_i^T}{\sum_j N_{ij}^T},
\end{aligned}$$

where N_{ij}^T is the number of times observed in $[0,T]$ that the trajectory visits M_j after hitting M_i last, $N_i^T = \sum_j N_{ij}^T$ is the number of times in $[0,T]$ that the trajectory hits M_i after hitting another milestone last, and R_i^T is the total time in $[0,T]$ during which M_i is the milestone that the trajectory hits last.

For a proof of a similar result with two milestones, see [19]. The above lemma gives estimators for $p_{i,j}$ and t_i in terms of an unbiased long trajectory of the milestoning index process. In practice, it is more efficient to use short parallel trajectories. Such a sampling method was proposed in [33] based on Voronoi tessellation. A variant can be formulated as follows: Consistent with Definition 2.2, suppose that we define the milestones as level sets of a function $f : \mathbb{R}^n \to [0,1]$, assuming that these level sets form a nested family of smooth surfaces (like e.g. the level sets of the backward committor function q^- in the context of optimal milestoning, but with f not necessarily equal to q^-). Specifically, given $0 < z_0 < z_1 \cdots < z_N < 1$, define

$$M_i = \{x : f(x) = z_i\}, \qquad i \in I. \tag{34}$$

Then set

$$\Omega_i = \{x : z_{i-1} \le f(x) \le z_{i+1}\}, \qquad i \in I \tag{35}$$

where we interpret $z_{-1} = -\infty$ and $z_{N+1} = +\infty$. Thus, for $i = 1, \ldots, N-1$, Ω_i is the region comprised between M_{i-1} and M_{i+1}, Ω_0 is the region on the side of M_1 that contains M_0 and Ω_N is the region on the side of M_{N-1} that contains M_N. It is then easy to see that if one considers the solution to the SDE in (3) in each Ω_i, with reflective (no-flux, Neumann) boundary conditions at $\partial\Omega_i = M_{i-1} \cup M_{i+1}$, then these solutions can be used to sample $p_{i,j}$ and t_i. Specifically, if we consider the set of milestones $\mathcal{M}^{(i)} \equiv \{M_{i-1}, M_i, M_{i+1}\}$ for $i = 1, \ldots, N-1$, $\mathcal{M}^{(0)} = \{M_0, M_1\}$ and $\mathcal{M}^{(N)} = \{M_{N-1}, M_N\}$, and use the solutions of the SDE in Ω_i with reflective boundary condition to construct the sequence $\{(\xi_n^{(i)}, \tau_n^{(i)}) : n \in \mathbb{N}_0\}$ associated with $\mathcal{M}^{(i)}$, then Lemma 6.1 still holds if we replace ξ_n by $\xi_n^{(i)}$ and α_n by $\alpha_n^{(i)} = \tau_n^{(i)} - \tau_{n-1}^{(i)}$ in the ergodic averages. We state this result as:

Proposition 6.2. *We have*

$$p_{i,j} = \lim_{n\to\infty} \frac{\sum_{k=1}^n \delta_{i,\xi_{k-1}^{(i)}} \delta_{j,\xi_k^{(i)}}}{\sum_{k=1}^n \delta_{i,\xi_{k-1}^{(i)}}},$$

$$t_i = \lim_{n\to\infty} \frac{\sum_{k=1}^n \alpha_k^{(i)} \delta_{i,\xi_{k-1}^{(i)}}}{\sum_{k=1}^n \delta_{i,\xi_{k-1}^{(i)}}}$$

Proof. The proposition is a consequence of the fact that the law of the solutions to the SDE in Ω_i with reflective boundary conditions on $\partial\Omega_i$ is identical to that of the solutions to the original SDE pruned to Ω_i. □

The computational advantage of this result is clear, as it permits to replace the sampling of one long trajectory across the whole domain by that of $N+1$ trajectories in the domains Ω_i, $i \in I$. This calculation can be done in parallel, and it guarantees that we can sample the process in regions that a long unbiased trajectory may visit only very infrequently.

We also note that if we consider the set of milestones $\mathcal{M}^{(i)} \equiv \{M_{i-1}, M_i, M_{i+1}\}$ for $i = 1, \ldots, N-1$, $\mathcal{M}^{(0)} = \{M_0, M_1\}$ and $\mathcal{M}^{(N)} = \{M_{N-1}, M_N\}$, and use the solutions of the SDE in Ω_i with reflective boundary condition to construct the sequence

$\{(Y_n^{(i)}, \tau_n^{(i)}) : n \in \mathbb{N}_0\}$, we can also sample $\mu_i(\cdot)$, $\tau(x)$, and $\nu(x,\cdot)$, at least in principle. Indeed, this amounts to recording the first time and locations the solution of the SDE in Ω_i with reflective boundary condition on $\partial\Omega_i$ crosses M_i first after hitting $\partial\Omega_i$, as well as the time and locations the trajectory hit this boundary the first time after that. This calculation is clearly more computationally expensive than that of t_i and $p_{i,j}$, as it requires much more data to converge, but it offers a route towards exact calculations within exact milestoning. We refer the reader to [35] for details in the context of a similar procedure termed trajectory parallelization and tilting. We also remark that the full information about $\nu(x,\cdot)$ etc. is not necessarily needed to compute certain quantities. For example, $T_j(x)$ as in Definition 3.5 can be solved using (19) iteratively, by noticing that the integral $\int_{\cap_{i\in I} M_i} \nu(x,\partial y) f(y)$ can be sampled by trajectories starting from $x \in M_i$ that hits one of the neighboring milestones.

6.2. **Approximation of the isocommittor surfaces.** The other issue is how to identify the isocommittor surfaces to be used as milestones in system of high dimensions. This is probably the most complicated problem to address. There are several algorithms to effectively approximate the isocommittor surfaces under suitable assumptions, e.g., [11, 12, 23, 22, 27, 34]. In particular, when the reaction channels are localized either in the original variables or in suitable collective variables, the string method [10, 11, 12, 23, 34] can be used to calculate the isocommittor surfaces. The output of the string method is one or more curves, each corresponding to a reaction tube. Assume that the curve γ is parametrized by $\varphi(s)$ with $s \in [0,1]$. We may take the parameter s to be the normalized arclength along γ and so $|\varphi'(s)| \equiv$ the total length of the curve γ. Associated with γ, we define a function $s_\gamma(x)$ as

$$s_\gamma(x) = \arg\min_{s\in[0,1]} |x - \varphi(s)|.$$

In words, $s_\gamma(x)$ is the value of the parameter s identifying the point $\varphi(s_\gamma(x))$ the closest to x along the curve to the point x. If the minimum is achieved by more than one value of s, we conventionally assign to $s_\gamma(x)$ the smallest value. It is easily seen that $s_\gamma(x)$ satisfies

$$\varphi'(s_\gamma(x)) \cdot (x - \varphi(s_\gamma(x))) = 0.$$

The key approximation in the string method is that the isocommittor surfaces are approximated locally in the reaction tube by the level sets of $s_\gamma(x)$. In other words, we assume that we can approximate the committor function q^- locally in the vicinity of the curve γ by

$$q^-(x) \approx f(x) = \int_{\mathbb{R}^n} K_\delta(x-y) Q(s_\gamma(y)) \mathrm{d}y$$

where $Q : [0,1] \to [0,1]$ with $Q'(x) > 0$ is some rescaling function and $K_\delta(x)$ is some smoothing kernel introduced to guarantee that the approximation $f(x)$ for $q^-(x)$ is smooth: for example, one could take

$$K_\delta(x) = (2\pi\delta^2)^{-n/2} e^{-\frac{1}{2}|x|^2/\delta^2} \qquad \text{for some } \delta > 0.$$

The accuracy is unfortunately hard to assess, except in very special circumstances. We will not dwell on this issue further here, even though it should be stressed that the results of optimal milestoning will crucially depend on the accuracy of the isocommittor surfaces we choose as optimal milestones.

7. CONCLUDING REMARKS

The main objective of this paper was to prove that a specific version of milestoning, termed optimal milestoning, retains information about the kinetics of the original process and permits e.g. the exact calculation of mean first passage times (MFPTs). As we saw, this property requires one to use specific sets of milestones, namely level sets of the backward committor function associated with the reaction from any set A to any B. We also explained why such milestones must by used by considering exact milestoning, which is akin to other non-equilibrium umbrella sampling methods, and also permits the exact calculation of certain MFPTs, but at the price of also storing the locations at which the process transitions from milestone to milestone. These results set standards to meet in order to use milestoning as an accelerated sampling scheme. What now remains to be developed are more computational tools to efficiently compute the isocommittor surfaces needed in optimal milestoning, along with theoretical tools to assess the error introduced, say, in the MFPTs if one uses isocommittor surfaces that are only approximated, like e.g. those given by the string method. Alternatively, one could use exact milestoning, in which case the main issue becomes the efficient computation and storage of the quantities needed in that approach.

APPENDIX A. PROOF OF THEOREM 4.2

In the proof of Theorem 4.2, the following lemma will be needed.

Lemma A.1. *All the surface integrals Z_i's in (27) are identical.*

Proof. Let Ω_j^i denote the open region enclosed by M_i and M_j, i.e.,

$$\Omega_j^i = \{x : z_j < q^-(x) < z_i\} \quad \text{for } i < j.$$

Also for convenience and consistent with our former convention, we allow $-1 \le i < j \le N+1$, and Ω_j^{-1} is understood as the region at one side of M_j that contains A, and Ω_{N+1}^i as that at one side of M_i that contains B. Notice that $\mathcal{L}^\dagger$ can be expressed as

$$\mathcal{L}^\dagger = \frac{1}{\rho}\big[\nabla \cdot (\rho a \nabla) - J \cdot \nabla\big], \tag{36}$$

where $J = \rho b - a\nabla\rho$ is the stationary probability current and is divergence free. Thus from $\mathcal{L}^\dagger q^- = 0$, we deduce that

$$\nabla \cdot \left(\rho a \nabla q^-\right) = J \cdot \nabla q^- = \nabla \cdot \left(q^- J\right). \tag{37}$$

Observe that

$$n(x) = \frac{\nabla q^-(x)}{|\nabla q^-(x)|} \tag{38}$$

is the unit normal vector of the surface M_i. Then for any $0 \le i < j \le N$, by the divergence theorem,

$$\begin{aligned} Z_j - Z_i &= \int_{M_j} \langle \rho a \nabla q^-, n \rangle \, \mathrm{d}\sigma_{M_j} - \int_{M_i} \langle \rho a \nabla q^-, n \rangle \, \mathrm{d}\sigma_{M_i} = - \int_{\Omega_j^i} \nabla \cdot \left(\rho a \nabla q^- \right) \, \mathrm{d}x \\ &= - \int_{\Omega_j^i} \nabla \cdot \left(q^- J \right) \, \mathrm{d}x = \int_{M_j} q^- J \cdot n \, \mathrm{d}\sigma_{M_j} - \int_{M_i} q^- J \cdot n \, \mathrm{d}\sigma_{M_i} \\ &= z_j \int_{M_j} J \cdot n \, \mathrm{d}\sigma_{M_j} - z_i \int_{M_i} J \cdot n \, \mathrm{d}\sigma_{M_i} \\ &= z_j \int_{\Omega_{N+1}^j} \nabla \cdot J \, \mathrm{d}x - z_i \int_{\Omega_{N+1}^i} \nabla \cdot J \, \mathrm{d}x = 0, \end{aligned}$$

where the last equality follows since J is divergence free. □

Proof of Theorem 4.2. By duality, it suffices to show

$$\int_{M_i} (\mathcal{P} f)(x) \, \mu_i(\mathrm{d}x) = \sum_{j=0}^{N} q_{i,j} \int_{M_j} f(x) \, \mu_j(\mathrm{d}x), \quad 0 \le i \le N,$$

for any nice test function f, where $(\mathcal{P} f)(x) = \mathbb{E}_x \big[f(Y_1) \big]$. For each $0 \le i \le N$, let u_i be the solution to the Dirichlet problem:

$$\begin{cases} \mathcal{L} u_i = 0 \quad \text{in } \Omega_{i+1}^{i-1}, \\ u_i|_{M_{i-1} \cup M_{i+1}} = f|_{M_{i-1} \cup M_{i+1}}. \end{cases}$$

Then $\mathcal{P} f$ and u_i coincide on M_i. Thus we need to verify

$$\int_{M_i} u_i \rho_i \, \mathrm{d}\sigma_{M_i} = q_{i,i-1} \int_{M_{i-1}} u_i \rho_{i-1} \, \mathrm{d}\sigma_{M_{i-1}} + q_{i,i+1} \int_{M_{i+1}} u_i \rho_{i+1} \, \mathrm{d}\sigma_{M_{i+1}} \tag{39}$$

for each $0 \le i \le N$ and every bounded smooth function u_i defined on $\Omega_{i+1}^{i-1} \cup M_{i-1} \cup M_{i+1}$ satisfying $\mathcal{L} u_i = 0$ in Ω_{i+1}^{i-1}. By Lemma A.1 and definition of ρ_i in (28), (39) reduces to

$$\begin{aligned} \int_{M_i} \langle u_i \rho a \nabla q^-, n \rangle \, \mathrm{d}\sigma_{M_i} &= q_{i,i-1} \int_{M_{i-1}} \langle u_i \rho a \nabla q^-, n \rangle \, \mathrm{d}\sigma_{M_{i-1}} \\ &\quad + q_{i,i+1} \int_{M_{i+1}} \langle u_i \rho a \nabla q^-, n \rangle \, \mathrm{d}\sigma_{M_{i+1}}. \end{aligned}$$

Let us insert the values of $q_{i,i-1}$ and $q_{i,i+1}$ given by (30) into the last equation, and multiply both sides by $z_{i-1} - z_{i+1}$, then we are left with checking

$$\begin{aligned} (z_{i-1} - z_{i+1}) \int_{M_i} \langle u_i \rho a \nabla q^-, n \rangle \, \mathrm{d}\sigma_{M_i} &= (z_i - z_{i+1}) \int_{M_{i-1}} \langle u_i \rho a \nabla q^-, n \rangle \, \mathrm{d}\sigma_{M_{i-1}} \\ &\quad + (z_{i-1} - z_i) \int_{M_{i+1}} \langle u_i \rho a \nabla q^-, n \rangle \, \mathrm{d}\sigma_{M_{i+1}}. \end{aligned}$$

Moving the right hand side to the left, and regrouping these terms properly into three surface integrals, we obtain

$$\begin{aligned}
&z_{i+1}\left(\int_{M_{i-1}}\langle u_i\rho a\nabla q^-,n\rangle\,\mathrm{d}\sigma_{M_{i-1}}-\int_{M_i}\langle u_i\rho a\nabla q^-,n\rangle\,\mathrm{d}\sigma_{M_i}\right)\\
+&z_{i-1}\left(\int_{M_i}\langle u_i\rho a\nabla q^-,n\rangle\,\mathrm{d}\sigma_{M_i}-\int_{M_{i+1}}\langle u_i\rho a\nabla q^-,n\rangle\,\mathrm{d}\sigma_{M_{i+1}}\right)\\
-&z_i\left(\int_{M_{i-1}}\langle u_i\rho a\nabla q^-,n\rangle\,\mathrm{d}\sigma_{M_{i-1}}-\int_{M_{i+1}}\langle u_i\rho a\nabla q^-,n\rangle\,\mathrm{d}\sigma_{M_{i+1}}\right)=0.
\end{aligned}$$

Applying divergence theorem to each surface integral yields

$$\begin{aligned}
&z_{i+1}\int_{\Omega_i^{i-1}}\nabla\cdot\left(u_i\rho a\nabla q^-\right)\,\mathrm{d}x+z_{i-1}\int_{\Omega_{i+1}^{i}}\nabla\cdot\left(u_i\rho a\nabla q^-\right)\,\mathrm{d}x\\
-&z_i\int_{\Omega_{i+1}^{i-1}}\nabla\cdot\left(u_i\rho a\nabla q^-\right)\,\mathrm{d}x=0.
\end{aligned}\tag{40}$$

Now we calculate

$$\nabla\cdot(u_i\rho a\nabla q^-)=\langle\nabla u_i,\rho a\nabla q^-\rangle+u_i\nabla\cdot(\rho a\nabla q^-),$$

$$\nabla\cdot(q^-\rho a\nabla u_i)=\langle\nabla q^-,\rho a\nabla u_i\rangle+q^-\nabla\cdot(\rho a\nabla u_i).$$

Since a is symmetric, we have

$$\nabla\cdot(u_i\rho a\nabla q^-)=\nabla\cdot(q^-\rho a\nabla u_i)-q^-\nabla\cdot(\rho a\nabla u_i)+u_i\nabla\cdot(\rho a\nabla q^-).$$

By (37), the last term on the right side is just $u_i\nabla\cdot(q^-J)$. On the other hand, since $\mathcal{L}$ can be expressed as

$$\mathcal{L}=\frac{1}{\rho}\big[\nabla\cdot(\rho a\nabla)+J\cdot\nabla\big],$$

a calculation similar to the derivation of (37) leads to

$$\nabla\cdot(\rho a\nabla u_i)=-\nabla\cdot(u_iJ).\tag{41}$$

Combining the above identities, we obtain

$$\nabla\cdot(u_i\rho a\nabla q^-)=\nabla\cdot(q^-\rho a\nabla u_i+q^-u_iJ).\tag{42}$$

Substituting this into (40) and applying the divergence theorem and $q^-=z_i$ on M_i, we get

$$\begin{aligned}
&z_{i+1}\left(z_{i-1}\int_{M_{i-1}}\langle\rho a\nabla u_i+u_iJ,n\rangle\,\mathrm{d}\sigma_{M_{i-1}}-z_i\int_{M_i}\langle\rho a\nabla u_i+u_iJ,n\rangle\,\mathrm{d}\sigma_{M_i}\right)\\
+&z_{i-1}\left(z_i\int_{M_i}\langle\rho a\nabla u_i+u_iJ,n\rangle\,\mathrm{d}\sigma_{M_i}-z_{i+1}\int_{M_{i+1}}\langle\rho a\nabla u_i+u_iJ,n\rangle\,\mathrm{d}\sigma_{M_{i+1}}\right)\\
-&z_i\left(z_{i-1}\int_{M_{i-1}}\langle\rho a\nabla u_i+u_iJ,n\rangle\,\mathrm{d}\sigma_{M_{i-1}}-z_{i+1}\int_{M_{i+1}}\langle\rho a\nabla u_i+u_iJ,n\rangle\,\mathrm{d}\sigma_{M_{i+1}}\right)\\
=&\,0.
\end{aligned}$$

Regrouping these terms again into three new surface integrals and applying the divergence theorem again, we may convert the last equation into

$$z_{i-1}z_{i+1}\int_{\Omega_{i+1}^{i-1}}\nabla\cdot(\rho a\nabla u_i+u_iJ)\,\mathrm{d}x-z_iz_{i+1}\int_{\Omega_{i+1}^{i}}\nabla\cdot(\rho a\nabla u_i+u_iJ)\,\mathrm{d}x$$
$$-z_{i-1}z_i\int_{\Omega_{i}^{i-1}}\nabla\cdot(\rho a\nabla u_i+u_iJ)\,\mathrm{d}x=0.$$

Finally this follows from (41) and the proof is complete. □

APPENDIX B. PROOF OF LEMMA 4.5

Proof. We will prove (i) by showing that for all $n\geq 1$ and $i_k\in I$, $0\leq k\leq n$,

$$\mathbb{P}_\mu\big[\xi_k=i_k,1\leq k\leq n\big|\,\xi_0=i_0\big]=p_{i_0,i_1}p_{i_1,i_2}\cdots p_{i_{n-1},i_n}.$$

The proof goes by induction on n. It is trivial for $n=1$. Suppose that this is true for some $n\geq 1$. Then for the case $n+1$, the strong Markov property of X gives

$$\begin{aligned}&\mathbb{P}_\mu\big[\xi_k=i_k,1\leq k\leq n+1\big|\,\xi_0=i_0\big]=\mathbb{P}_{\mu_{i_0}}\big[Y_k\in M_{i_k},1\leq k\leq n+1\big]\\ =&\mathbb{E}_{\mu_{i_0}}\big[\mathbf{1}_{\{Y_1\in M_{i_1}\}}\mathbb{P}_{\mu_i}\big[Y_{k-1}\in M_{i_k},2\leq k\leq n+1\,\big|\mathscr{F}_{\tau_1}\big]\big]\\ =&\mathbb{E}_{\mu_{i_0}}\big[\mathbf{1}_{M_{i_1}}(Y_1)\mathbb{P}_{X(\tau_1)}\big[Y_{k-1}\in M_{i_k},2\leq k\leq n+1\big]\big],\end{aligned}$$

where the second line follows from time homogeneity. Note that the distribution of $Y_1=X(\tau_1)$ relative to the probability law $\mathbb{P}_{\mu_{i_0}}$ is $\mathcal{P}^*\mu_{i_0}$, which by assumption, is given by $\sum_{j\in I}p_{i_0,j}\mu_j$. So the last display equals to

$$\begin{aligned}&=\int_{\mathcal{M}}\mathbf{1}_{M_{i_1}}(x)\mathbb{P}_x\big[Y_k\in M_{i_{k+1}},1\leq k\leq n\big]\,(\mathcal{P}^*\mu_{i_0})(\mathrm{d}x)\\ &=\sum_{j\in I}p_{i_0,j}\int_{M_{i_1}}\mathbb{P}_x\big[\xi_k=i_{k+1},1\leq k\leq n\big]\,\mu_j(\mathrm{d}x)\\ &=p_{i_0,i_1}\mathbb{P}_{\mu_{i_1}}\big[\xi_k=i_{k+1},1\leq k\leq n\big]\\ &=p_{i_0,i_1}\mathbb{P}_\mu\big[\xi_k=i_{k+1},1\leq k\leq n\big|\,\xi_0=i_1\big].\end{aligned}$$

Using the induction hypothesis yields the desired equality for $n+1$, which completes the inductive step and the assertion on the Markovianity of $\{\xi_n:n\in\mathbb{N}\}$ is proved.

The proof of (ii) is similar by induction on n. We just outline the inductive step below. By the strong Markov property, we compute

$$\begin{aligned}&\mathbb{E}_\mu\big[\alpha_n\mathbf{1}_{\{\xi_k=i_k,1\leq k\leq n\}}\big|\,\xi_0=i_0\big]=\mathbb{E}_{\mu_{i_0}}\Big[\alpha_n\prod_{1\leq k\leq n}\mathbf{1}_{M_{i_k}}(Y_k)\Big]\\ =&\mathbb{E}_{\mu_{i_0}}\Big[\mathbf{1}_{M_{i_1}}(Y_1)\mathbb{E}_{X(\tau_1)}\Big[\alpha_{n-1}\prod_{2\leq k\leq n}\mathbf{1}_{M_{i_k}}(Y_{k-1})\Big]\Big]\\ =&p_{i_0,i_1}\mathbb{E}_{\mu_{i_1}}\Big[\alpha_{n-1}\prod_{1\leq k\leq n-1}\mathbf{1}_{M_{i_{k+1}}}(Y_k)\Big]\\ =&p_{i_0,i_1}\mathbb{E}_\mu\big[\alpha_{n-1}\mathbf{1}_{\{\xi_k=i_{k+1},1\leq k\leq n-1\}}\big|\,\xi_0=i_1\big],\end{aligned}$$

where in the third equality we have used the assumption that under the law $\mathbb{P}_{\mu_{i_0}}$, the probability of the event $\{Y_1\in M_{i_1}\}$ is p_{i_0,i_1}, and given this event, the conditional distribution of Y_1 is μ_{i_1}. On the other hand, by (i), we have

$$\mathbb{P}_\mu\big[\xi_k=i_k,1\leq k\leq n\big|\,\xi_0=i_0\big]=p_{i_0,i_1}\mathbb{P}_\mu\big[\xi_k=i_{k+1},1\leq k\leq n-1\big|\,\xi_0=i_1\big].$$

So we obtain

$$\mathbb{E}_\mu\big[\alpha_n\big|\xi_k = i_k, 0 \le k \le n\big] = \mathbb{E}_\mu\big[\alpha_{n-1}\big|\xi_k = i_{k+1}, 0 \le k \le n-1\big].$$

This completes the inductive step. □

APPENDIX C. PROOF OF LEMMA 4.6

Proof. Let us focus on the case $j = N$, the proof for other j is the same. For simplicity of notation, we will write $h_i = h_{i,N}$.

We first show that $(h_i)_{i\in I}$ satisfies (31). Note that the event $\{D \ge n\}$ belongs to the σ-field generated by $\xi_0, \cdots, \xi_{n-1}$. Then by Fubini's Theorem and Lemma 4.5, we obtain

$$\begin{aligned} h_i = \mathbb{E}_{\mu_i}[\tau_D] = \mathbb{E}_{\mu_i}\left[\sum_{n=1}^{D}\alpha_n\right] = \mathbb{E}_{\mu_i}\left[\sum_{n=1}^{\infty}\alpha_n\mathbf{1}_{\{D\ge n\}}\right] = \sum_{n=1}^{\infty}\mathbb{E}_{\mu_i}\big[\alpha_n\mathbf{1}_{\{D\ge n\}}\big] \\ = \sum_{n=1}^{\infty}\mathbb{E}_{\mu_i}\big[\mathbf{1}_{\{D\ge n\}}\mathbb{E}_{\mu_i}\big[\alpha_n\big|\xi_1,\cdots,\xi_n\big]\big] = \sum_{n=1}^{\infty}\mathbb{E}_{\mu_i}\big[\mathbf{1}_{\{D\ge n\}}t_{\xi_{n-1},\xi_n}\big]. \end{aligned} \tag{43}$$

Let us assume $i \ne N$. Then $D \ge 1$ and the last-written sum can be split into

$$\mathbb{E}_{\mu_i}\big[t_{i,\xi_1}\big] + \sum_{n=2}^{\infty}\mathbb{E}_{\mu_i}\big[\mathbf{1}_{\{D\ge n\}}t_{\xi_{n-1},\xi_n}\big],$$

where the first term amounts to $\sum_{j\in I} t_{i,j}p_{i,j}$, while the second summation term is equal to

$$\begin{aligned} &\sum_{n=2}^{\infty}\sum_{j\in I}\mathbb{E}_\mu\big[\mathbf{1}_{\{D\ge n\}}t_{\xi_{n-1},\xi_n}\big|\ \xi_0 = i, \xi_1 = j\big]\mathbb{P}_\mu\big[\xi_1 = j\ \big|\ \xi_0 = i\big] \\ =&\sum_{j\in I}p_{i,j}\sum_{n=2}^{\infty}\mathbb{E}_\mu\big[\mathbf{1}_{\{\xi_m\ne N, 1\le m\le n-1\}}t_{\xi_{n-1},\xi_n}\big|\ \xi_0 = i, \xi_1 = j\big]. \end{aligned}$$

By Lemma 4.5, under the law $\mathbb{P}_\mu$, $\{\xi_n : n \in \mathbb{N}\}$ has the time-homogeneous Markov property, therefore the last display equals to

$$\begin{aligned} &\sum_{j\in I}p_{i,j}\sum_{n=2}^{\infty}\mathbb{E}_j\big[\mathbf{1}_{\{\xi_m\ne N, 0\le m\le n-2\}}t_{\xi_{n-2},\xi_{n-1}}\big] \\ =&\sum_{j\in I}p_{i,j}\sum_{n=1}^{\infty}\mathbb{E}_j\big[\mathbf{1}_{\{D\ge n\}}t_{\xi_{n-1},\xi_n}\big] = \sum_{j\in I}p_{i,j}h_j, \end{aligned}$$

where the last step follows from (43).

To prove the uniqueness, we will show that $(h_i)_{i\in I}$ is the minimal nonnegative solution to the discrete Poisson problem (31). To this end, suppose that $(y_i)_{i\in I}$ is another nonnegative solution. Then $y_N = 0$. Consider $i \ne N$, we have

$$y_i = \sum_{j\in I}p_{i,j}t_{i,j} + \sum_{j\in I}p_{i,j}y_j = \sum_{j\in I}p_{i,j}t_{i,j} + \sum_{j\ne N}p_{i,j}y_j.$$

Using this identity to replace y_j on the right hand side and by Lemma 4.5, we deduce

$$\begin{aligned}
y_i &= \sum_{j\in I} p_{i,j}t_{i,j} + \sum_{j\neq N} p_{i,j}\left(\sum_{k\in I} p_{j,k}t_{j,k} + \sum_{k\neq N} p_{j,k}y_k\right)\\
&= \sum_{j\in I} \mathbb{P}_\mu\big[\xi_1 = j \,\big|\, \xi_0 = i\big]\mathbb{E}_\mu\big[\alpha_1\big|\, \xi_0 = i, \xi_1 = j\big]\\
&+ \sum_{j\neq N}\sum_{k\in I} \mathbb{P}_\mu\big[\xi_1 = j, \xi_2 = k \,\big|\, \xi_0 = i\big]\mathbb{E}_\mu\big[\alpha_2\big|\, \xi_0 = i, \xi_1 = j, \xi_2 = k\big]\\
&+ \sum_{j\neq N}\sum_{k\neq N} p_{i,j}p_{j,k}y_k\\
&= \mathbb{E}_{\mu_i}\big[\alpha_1\big] + \mathbb{E}_{\mu_i}\big[\alpha_2\mathbf{1}_{\{D\geq 2\}}\big] + \sum_{j\neq N}\sum_{k\neq N} p_{i,j}p_{j,k}y_k.
\end{aligned}$$

By repeated substitution in the last term, we obtain after n steps

$$y_i = \sum_{m=1}^{n} \mathbb{E}_{\mu_i}\big[\alpha_m\mathbf{1}_{\{D\geq m\}}\big] + \text{ Nonnegative Remainder} \geq \sum_{m=1}^{n} \mathbb{E}_{\mu_i}\big[\alpha_m\mathbf{1}_{\{D\geq m\}}\big].$$

Sending $n\to\infty$, by (43), we find

$$y_i \geq \sum_{m=1}^{\infty} \mathbb{E}_{\mu_i}\big[\alpha_m\mathbf{1}_{\{D\geq m\}}\big] = h_i,$$

which completes the proof. □

APPENDIX D. PROOF OF LEMMA 4.7

Proof. Let $\{(\xi'_n, \tau'_n) : n\in\mathbb{N}\}$ be the coarse-grained milestoning chain associated with $\mathcal{M}'$ and set $Y'_n = X(\tau'_n)$. Let $(\mathcal{P}')^*$ denote the linear shift operator associated with the transition probability of the Markov chain $\{Y'_n : n\in\mathbb{N}\}$.

What we need to prove is that for every $i, j\in I'$, and any measurable set $E\subset M_j$,

$$((\mathcal{P}')^*\mu_i)(E) = p'_{ij}\mu_j(E), \quad \text{with } p'_{ij} = ((\mathcal{P}')^*\mu_i)(M_j).$$

Fix i and j in I'. By definition, the left hand member is

$$\mathbb{P}_{\mu_i}\left[X\left(H^+_{M'\setminus M'_i}\right)\in E\right].$$

Put $\eta = X\left(H^+_{M'\setminus M'_i}\right)$ and $J' = (I\setminus I')\cup\{i\}$. Using the strong Markov property of X and the assumption that

$$\mathcal{P}^*\mu_k = \sum_{\ell\in I} p_{k,\ell}\mu_\ell, \quad \text{with } p_{k,\ell} = (\mathcal{P}^*\mu_k)(M_\ell),$$

we obtain that for $k\in J'$,

$$\begin{aligned}
\mathbb{P}_{\mu_k}\big[\eta\in E\big] &= \mathbb{E}_{\mu_k}\big[\mathbb{P}_{Y_1}\big[\eta\in E\big]\big] = \sum_{\ell\in I} p_{k,\ell}\mathbb{P}_{\mu_\ell}\big[\eta\in E\big]\\
&= \sum_{\ell\in I'\setminus\{i\}} p_{k,\ell}\mathbb{P}_{\mu_\ell}\big[\eta\in E\big] + \sum_{\ell\in J'} p_{k,\ell}\mathbb{P}_{\mu_\ell}\big[\eta\in E\big]\\
&= p_{k,j}\mu_j(E) + \sum_{\ell\in J'} p_{k,\ell}\mathbb{P}_{\mu_\ell}\big[\eta\in E\big].
\end{aligned}$$

In the last step, we use the fact that under the law $\mathbb{P}_{\mu_\ell}$ for $\ell \in I' \setminus \{i\}$, we have $H^+_{M' \setminus M'_i} = 0$ and $\eta = X(0)$ almost surely. Using this identity to replace the term $\mathbb{P}_{\mu_\ell}\big[\eta \in E\big]$ on the right hand side, we see that

$$\begin{aligned}\mathbb{P}_{\mu_k}\big[\eta \in E\big] &= p_{k,j}\mu_j(E) + \sum_{\ell \in J'} p_{k,\ell}p_{\ell,j}\mu_j(E) + \sum_{\ell,m \in J'} p_{k,\ell}p_{\ell,m}\mathbb{P}_{\mu_m}\big[\eta \in E\big] \\ &= \mathbb{P}_{\mu_k}\big[\xi_1 = j\big]\mu_j(E) + \mathbb{P}_{\mu_k}\big[\xi_1 \in J', \xi_2 = j\big]\mu_j(E) \\ &\quad + \sum_{\ell,m \in J'} p_{k,\ell}p_{\ell,m}\mathbb{P}_{\mu_m}\big[\eta \in E\big].\end{aligned}$$

Continuing in the obvious way, we find

$$\mathbb{P}_{\mu_k}\big[\eta \in E\big] = \mu_j(E)\sum_{n=1}^{m}\mathbb{P}_{\mu_k}\big[\xi_\ell \in J', 1 \le \ell < n,\ \xi_n = j\big] + \text{ Nonnegative Remainder.}$$

Sending $m \to \infty$ yields

$$\mathbb{P}_{\mu_k}\big[\eta \in E\big] \ge \mu_j(E)\sum_{n=1}^{\infty}\mathbb{P}_{\mu_k}\big[\xi_\ell \in J', 1 \le \ell < n,\ \xi_n = j\big].$$

Observe that the summation on the right hand side of the last inequality actually gives the probability that the chain $\{\xi_n : n \in \mathbb{N}\}$ starting at $\xi_0 = k$ will visit $I' \setminus \{i\}$ in some step and occupy the state j at the first time of visiting $I' \setminus \{i\}$, which may also be expressed as

$$\mathbb{P}_{\mu_k}\left[H^+_{M' \setminus M'_i} = \tau_n \text{ for some } n \ge 0, \eta = X\left(H^+_{M' \setminus M'_i}\right) \in M_j\right].$$

Since the diffusion process X has continuous trajectories, and $\lim_{n\to\infty} X(\tau_n)$ does not exist by construction, it follows that $\lim_{n\to\infty}\tau_n = \infty$. This implies

$$\mathbb{P}_{\mu_k}\left[H^+_{M' \setminus M'_i} = \tau_n \text{ for some } n \ge 0\right] = 1.$$

Thus we obtain

$$\mathbb{P}_{\mu_k}\big[\eta \in E\big] \ge \mu_j(E)\mathbb{P}_{\mu_k}\big[\eta \in M_j\big].$$

To turn the last inequality into an equality, we observe that the last inequality also holds for $M_j \setminus E$ and therefore

$$\begin{aligned}\mathbb{P}_{\mu_k}\big[\eta \in M_j\big] &= \mathbb{P}_{\mu_k}\big[\eta \in E\big] + \mathbb{P}_{\mu_k}\big[\eta \in M_j \setminus E\big] \\ &\ge \mu_j(E)\mathbb{P}_{\mu_k}\big[\eta \in M_j\big] + \mu_j(M_j \setminus E)\mathbb{P}_{\mu_k}\big[\eta \in M_j\big] \\ &= \mathbb{P}_{\mu_k}\big[\eta \in M_j\big].\end{aligned}$$

Since the left and right hand sides are equal, we must have

$$\mathbb{P}_{\mu_k}\big[\eta \in E\big] = \mu_j(E)\mathbb{P}_{\mu_k}\big[\eta \in M_j\big].$$

In particular, for $k = i$, this is exactly what was to be shown. The proof is complete. □

REFERENCES

[1] R. J. Allen, D. Frenkel, and P. R. ten Wolde. Forward flux sampling-type schemes for simulating rare events: Efficiency analysis. *J. Chem. Phys.* **124**(19), 194111 (2006).

[2] D. Aristoff, J. M. Bello-Rivas, and R. Elber. A Mathematical Framework for Exact Milestoning. *Multiscale Model. Simul.* **14**(1), 301-322 (2016).

[3] J. M. Bello-Rivas and R. Elber. Exact milestoning. *J. Chem. Phys.*, **142**, 094102 (2015).

[4] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability and low lying spectra in reversible Markov chains. *Comm. Math. Phys.* **228**(2), 219 (2002).

[5] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc.* **6**(4), 399 (2004).
[6] A. Bovier, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc.* **7**(1), 69 (2005).
[7] A. Dickson, A. Warmflash, and A. R. Dinner. Nonequilibrium umbrella sampling in spaces of many order parameters. *J. Chem. Phys.* **130**, 074104 (2009)
[8] N. Djurdjevac, M. Sarich, and C. Schütte. On Markov state models for metastable processes. in *Proceeding of the ICM 2010*.
[9] N. Djurdjevac, M. Sarich, and C. Schütte. Estimating the eigenvalue error of Markov state models. *Multiscale Model. Simul.* **10**(1), 61 (2012).
[10] W. E, W. Ren, and E. Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B* **66**, 052301 (2002).
[11] W. E, W. Ren, and E. Vanden-Eijnden. Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.* **413**, 242 (2005).
[12] W. E, W. Ren, and E. Vanden-Eijnden. Finite temperature string method for the study of rare events. *J. Phys. Chem. B* **109**, 6688 (2005).
[13] W. E and E. Vanden-Eijnden. Towards a theory of transition paths. *J. Stat. Phys.* **123**, 503 (2006).
[14] W. E and E. Vanden-Eijnden. *Transition path theory and path-finding algorithms for the study of rare events*. Extended version available from the ArXiv, (2009).
[15] W. E and E. Vanden-Eijnden. Transition-Path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* **61**, 391 (2010).
[16] R. Elber. A milestoning study of the kinetics of an allosteric transition: atomically detailed simulations of deoxy Scapharca hemoglobin. *Biophys. J.* **92**, L85 (2007).
[17] A. K. Faradjian and R. Elber. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* **120**, 10880 (2004).
[18] E. Guarnera and E. Vanden-Eijnden. Optimized Markov State Models for Metastable Systems. *J. Chem. Phys.* **145**, 024102 (2016).
[19] J. Lu and J. Nolen. Reactive trajectories and the transition path processes. *Probab. Theory Relat. Fields* **161**, 195–244 (2015).
[20] J. Lu and E. Vanden-Eijnden. Markov state modeling and metastability. In preparation.
[21] J. Lu and E. Vanden-Eijnden. Exact dynamical coarse-graining without time-scale separation. *J. Chem. Phys.* **141**, 044109 (2014).
[22] A. Ma and A. R. Dinner. Automatic method for identifying reaction coordinates in complex systems. *J. Phys. Chem. B* **109**, 6769 (2005).
[23] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **125**, 024106 (2006).
[24] P. Metzner, C. Schütte, and E. Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.* **125**, 084110 (2006).
[25] P. Metzner, C. Schütte, and E. Vanden-Eijnden. Transition path theory for Markov jump processes. *Multiscale Model. Simul.* **7**, 1192 (2009).
[26] D. Moroni, T. S. van Erp, and P. G. Bolhuis. Investigating rare events by transition interface sampling. *Physica A* **340**, 395-401 (2004).
[27] B. Peters and B. L. Trout. Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.* **125**, 054108 (2006).
[28] D. Shalloway and A. K. Faradjian. Efficient computation of the first passage time distribution of the generalized master equation by steady-state relaxation. *J. Chem. Phys.* **124**, 054112 (2006).
[29] C. Schütte, F. Noé, J. Lu, M. Sarich, and E. Vanden-Eijnden. Markov state models based on milestoning. *J. Chem. Phys.* **134**, 204105 (2011).
[30] M. Sarich, F. Noé, and C. Schütte. On the approximation quality of Markov state models. *Multiscale Model. Simul.* **8**, 1154 (2010).
[31] C. Valeriani, R. J. Allen, M. J. Morelli, D. Frenkel, and P. R. ten Wolde. Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling. *J. Chem. Phys.* **127**(11), 114109 (2007).
[32] E. Vanden-Eijnden. Transition Path Theory. In M. Ferrario, G. Ciccotti, and K. Binder, editors, *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, volume 1, pp. 439-478. Springer, 2006.
[33] E. Vanden-Eijnden and M. Venturoli. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* **130**, 194101 (2009).

[34] E. Vanden-Eijnden and M. Venturoli. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* **130**, 194103 (2009).
[35] E. Vanden-Eijnden and M. Venturoli. Exact rate calculations by trajectory parallelization and tilting. *J. Chem. Phys.* **131**, 044120 (2009)
[36] E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber. On the assumptions underlying milestoning. *J. Chem. Phys.* **129**, 174102 (2008).
[37] A. Warmflash, P. Bhimalapuram, and A. R. Dinner, Umbrella sampling for nonequilibrium processes. *J. Chem. Phys.* **127**, 154112 (2007).
[38] A. M. A. West, R. Elber, and D. Shalloway. Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide. *J. Chem. Phys.* **126**, 145104 (2007).

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012, USA
Current address: Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong
E-mail address: linglin@cityu.edu.hk

MATHEMATICS DEPARTMENT, DUKE UNIVERSITY, BOX 90320, DURHAM NC 27708, USA
E-mail address: jianfeng@math.duke.edu

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012, USA
E-mail address: eve2@cims.nyu.edu