



Published in final edited form as:

Multiscale Model Simul. 2016 ; 14(1): 301–322. doi:10.1137/15M102157X.

A MATHEMATICAL FRAMEWORK FOR EXACT MILESTONING

DAVID ARISTOFF^{*}, JUAN M. BELLO-RIVAS[†], and RON ELBER[‡]

^{*}Department of Mathematics, Colorado State University, Fort Collins, CO

[†]Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX

[‡]Institute for Computational Engineering and Sciences, Department of Chemistry, University of Texas at Austin, Austin, TX

Abstract

We give a mathematical framework for Exact Milestoning, a recently introduced algorithm for mapping a continuous time stochastic process into a Markov chain or semi-Markov process that can be efficiently simulated and analyzed. We generalize the setting of Exact Milestoning and give explicit error bounds for the error in the Milestoning equation for mean first passage times.

Keywords

accelerated molecular dynamics; long-time dynamics; stationary distribution; semi-Markov processes

1. Introduction

Molecular Dynamics (MD) simulations, in which classical equations of motion are solved for molecular systems of significant complexity, have proven useful for interpreting and understanding many chemical and biological phenomena (for textbooks see [23, 12, 1]). However, a significant limitation of MD is of time scales. Many molecular processes of interest occur on time scales significantly longer than the temporal scales accessible to straightforward simulations. For example, permeation of molecules through membranes can take hours [6], while MD is usually restricted to the microsecond time scale. One approach to extend simulation times is to use faster hardware [24, 25, 21]. Other approaches focus on developing theories and algorithms for long time phenomena. The emphasis in developing algorithms has been on methodologies for activated processes with a single dominant barrier such as Transition Path Sampling [8, 19]. Approaches for dynamics on rough energy landscapes and for more general and/or diffusive behavior have also been developed [22, 7, 26, 20]. The techniques of Exact Milestoning [3] and Milestoning [10] belong to the latter category.

The acceleration in Milestoning is based on the use of a large number of short trajectories between certain hypersurfaces in phase space that we call milestones, instead of complete trajectories between reactants and products (see Figure 1). The use of short trajectories is trivial to parallelize, making the formulation efficient to use on modern computing resources. Moreover, using short trajectories initiated near bottlenecks of reactions allows for enhanced sampling of improbable but important events. The last enhancement can lead

to exponential speed up of the computations. A challenge is how to start the short trajectories and how to analyze the result to obtain the correct long time behavior.

Milestoning uses an approximate distribution for the initial conditions of the trajectories at the hypersurfaces. The results are then analyzed within the Milestoning theory. This approximation is typically the canonical distribution conditioned to be at the milestones. Exact Milestoning improves over the starting guess of the initial conditions by an iterative process that converges to the correct stationary distribution.

There are related approaches that use short trajectory information to obtain information on long time kinetics such as non-equilibrium umbrella sampling [28] and trajectory tilting [27]. However the overall scope of these methods is more limited than that of Exact Milestoning, which provides a complete statistical mechanical picture of the system.

While Milestoning is an approximate procedure, it shares the same philosophy and core algorithm as the Exact Milestoning approach. In both algorithms the state space Ω is coarse-grained by closed subsets called milestones. The short trajectories are run between nearby milestones and are terminated the first time they reach a milestone different from the one they were initiated on (see Figure 1). The short trajectories in time and space should be contrasted with long continuous trajectories from the reactant R to the products P in straight-forward MD.

In the theoretical analysis of the short trajectories, in Milestoning and Exact Milestoning emphasis is made on the absolute flux. The absolute flux is always positive and is the number of trajectories that pass through a milestone per unit time. While we obtained time-dependent solutions for the kinetics in Milestoning in the past [10], more recently and frequently we use stationary solutions [3, 17]. In stationary systems the flux is time independent and can be used to define a probability measure after a proper normalization. Hence, it is possible to reformulate the algorithms of Milestoning and Exact Milestoning in terms of widely used concepts and tools of probability theory. This formulation and a rigorous mathematical framework are provided in the present manuscript.

This article is organized as follows. In Section 2, we describe the setting for Exact Milestoning and introduce notation used throughout. In Section 3, we show existence of and convergence to a stationary flux under very general conditions. In Section 4 we state precisely the Exact Milestoning algorithm [3]. In Section 5, we establish conditions under which convergence to the stationary flux is consistent in the presence of numerical error (Lemma 5.2 and Theorem 5.3), and we give a natural upper bound for the numerical error arising in Exact Milestoning (Theorem 5.4). Finally, in Section 6 we consider some instructive examples.

2. Setup and notation

2.1. The dynamics and MFPT

In Milestoning we spatially coarse-grain a stochastic dynamics (X_t) . The basic idea is to stop and start trajectories on certain interfaces, called milestones, and then reconstruct functions

of (X_t) using these short trajectories, which can be efficiently simulated in parallel. We focus here on using Milestoning for the efficient computation of *mean first passage times* (MFPTs) of (X_t) , although similar ideas can be used to compute other non-equilibrium quantities.

To make our arguments we need some assumptions on the process (X_t) . We assume (X_t) is a time homogeneous strong Markov process with càdlàg paths taking values in a standard Borel space Ω . These assumptions allow us to stop and restart (X_t) on the milestones without knowing its history. In applications, usually (X_t) is Langevin or overdamped Langevin dynamics, and Ω is a subset of Euclidean space. More generally, (X_t) could be a diffusion and $\Omega = \mathbb{R}^d/\mathbb{Z}^d$ could be a torus, or (X_t) could be a continuous time Markov chain on a discrete set $\Omega = \mathbb{N}$.

We write \mathbb{P}, \mathbb{E} for all probability measures and expectations. \mathbb{P}^ξ and \mathbb{P}^x will represent probabilities for a process starting at a distribution ξ or point x , respectively. We write $\|\cdot\|_{TV}$ for total variation norm of signed measures. Our analysis below will mostly take place in an idealized setting where we assume infinite sampling on the milestones. In this setting, distributions are smooth (if state space is continuous) and the total variation norm is the appropriate one. We omit explicit mention of sigma algebras and filtrations, which will always be the natural ones, and we will use the words *distribution* and *probability measure* interchangeably.

Recall we are interested in computing the MFPT of (X_t) from a reactant set R to a product set P . Throughout we consider fixed disjoint product and reactant sets $P, R \subset \Omega$. When R is not a single point, we will start (X_t) from a fixed probability measure ρ on R . If R is a single point, $\rho = \delta_R$, the delta distribution at R . As discussed above, Milestoning allows for an efficient computation of the MFPT of (X_t) to P , starting at ρ . It is useful to think of P as a sink, and R as a source for (X_t) . More precisely, *we assume that when (X_t) reaches P , it immediately restarts on R according to ρ* . Obviously, this assumption has no effect on the MFPT to P . It will be useful, however, for computational and theoretical considerations.

2.2. The milestones and semi-Markov viewpoint

Milestoning uses a set $M \subset \Omega$ to spatially parallelize the computation of the MFPT. Each point $x \in M$ will belong to a *milestone* $M_x \subset M$. Thus, M is the union of all the milestones. We assume there are finitely many milestones, each of which is a closed set. Moreover, we assume that (X_t) passes through the intersection of two milestones with probability 0 – thus, (X_t) can only cross one milestone at a time. This can be accomplished for Langevin or overdamped Langevin dynamics by, for example, taking the milestones to be codimension 1 manifolds whose pairwise intersections are codimension 2 or larger. The sets P and R will be two of the milestones. Typically, the milestones are hyperplanes which approximate the level sets of a reaction coordinate, or they are faces of cell boundaries of a Voronoi tessellation obtained by taking points along a high temperature trajectory for (X_t) .

Let $X_0 \sim \xi$, where ξ denotes a generic probability measure on M . We now describe in detail the coarse-graining process. By following the sequence of milestones crossed by (X_t) , we obtain a stochastic process defined on M as follows. Let $\theta_0, \theta_1, \theta_2, \dots$ be stopping times for (X_t) defined by

if $X_{\theta_n} = x$, then $\theta_{n+1} := \inf\{t > \theta_n : X_t \in M_y \text{ for some } M_y \neq M_x\}$, $\theta_0 = 0$.

Informally, θ_n corresponds to the time at which (X_t) crosses a milestone for the n th time. It is very important to note that θ_n corresponds to the first time for (X_t) to hit the “next” milestone, that is, a milestone different from the one (X_t) hit at time θ_{n-1} . Thus, when we refer to milestone crossings, we mean *crossings of a milestone different from the last one crossed*.

Define also $\tau_0 = 0$ and for $n \geq 1$, $\tau_n = \theta_n - \theta_{n-1}$. Informally, τ_n is the time between the $(n-1)$ th and n th milestone crossings. For $n \geq 0$, define $J_n = X_{\theta_n}$ to be the sequence of milestone points visited. That is, J_n is the point on the n th milestone crossed by (X_t) . Note that each J_n is an element on some milestone – not the label of a milestone. Thus, (J_n) lives on the possibly uncountable set M .

As discussed above, due to our assumptions on M and (X_t) , the law of X_n and τ_n depends only on X_{n-1} , that is,

$$\mathbb{P}(J_n \in C, \tau_n \leq t | (J_0, \tau_0), \dots, (J_{n-1} = x, \tau_{n-1})) = \mathbb{P}(J_n \in C, \tau_n \leq t | J_{n-1} = x). \quad (2.1)$$

Let (Y_t) be the jump process on M defined by $Y_t = J_n$ if $\theta_n < t < \theta_{n+1}$. Since J_n is the n th milestone point crossed by (X_t) , and θ_n is the time at which (X_t) crossed this point, we see that (Y_t) equals (X_t) at each milestone crossing. Thus, (Y_t) can be considered a coarse-grained version of (X_t) which throws away the path of (X_t) between milestones, keeping only the endpoints. When (Y_t) has a probability density, it corresponds to the density $p(x, t)$ from [3] for the last milestone point passed. Due to (2.1), (Y_t) is called *semi-Markov*. It has transition kernel $K(x, dy, dt)$ given by

$$\mathbb{P}(J_n \in C, \tau_n \leq t | J_{n-1} = x) = \int_{[0, t]} \int_C K(x, dy, dt).$$

We adopt the convention that the innermost integral corresponds to the first variable of integration appearing in the integrand. *Note that if (Y_t) and (X_t) start at the same distribution ξ on M , then they have the same MFPT to P .* Thus, for the purposes of studying the MFPT, it suffices to consider the coarse-grained process (Y_t) . Since (X_t) restarts at ρ upon reaching P , so does (Y_t) , and so

$$\text{for all } x \in P, \quad K(x, dy, dt) = \delta_{t=0} \rho(dy). \quad (2.2)$$

Following standard terminology in probability theory, we call (J_n) the *jump chain* of (Y_t) . It is a Markov chain with transition kernel

$$K(x, dy) := \int_{[0, \infty)} K(x, dy, dt).$$

2.3. Stopping times

Note that the sojourn times τ_n between milestones depend on the position of (X_t) on the n th milestone crossed at time $t = \theta_{n-1}$. The following notation will thus be useful. For $x \in M$, we write

$$\tau_M^x = \inf\{t > 0: Y_t \in M_y \text{ for some } M_y \neq M_x\}.$$

This is the first time for (Y_t) to hit some milestone other than M_x , starting at $x \in M$. In particular, if $X_{\theta_{n-1}} = x$ was the $(n-1)$ th crossing point, then $\theta_n = \theta_{n-1} + \tau_M^x$ is the next crossing time. As we are concerned with first passage times to P , we will also use the notation

$$\tau_P = \inf\{t > 0: Y_t \in P\}.$$

In particular, the MFPT we are interested in is $\mathbb{E}^P[\tau_P]$. As discussed above, the MFPT will be estimated via short trajectories between milestones. An important ingredient is the correct starting distribution for these trajectories. Exact Milestoning makes use of a *stationary flux* of (X_t) on the milestones, which corresponds to the stationary distribution (shown to exist below) of the jump chain (J_n) . To study this distribution we need to define also the first hitting time of P for (J_n) :

$$\sigma_P = \min\{n \geq 0: J_n \in P\}.$$

Some assumption is required to guarantee the existence of a stationary flux. We adopt the following very general sufficient condition, which we assume holds through the remainder of the paper: $\mathbb{E}^\xi[\tau_P]$ and $\mathbb{E}^\xi[\sigma_P]$ are finite for all probability measures ξ on M . This ensures that (Y_t) reaches P in finite expected time and does not cross infinitely many milestones in finite time. In probabilistic language, the latter condition means that (Y_t) is non-explosive. This sufficient condition can be readily verified in the standard settings for milestoning discussed above. Under *only* this assumption, we give in Theorem 3.3 an *exact* milestoning equation for $\mathbb{E}^P[\tau_P]$. Moreover we show in Section (4) how the equation leads to an efficient algorithm for computing $\mathbb{E}^P[\tau_P]$. The exact equation we give is the same as in [3], except phrased in the semi-Markov framework defined above.

2.4. Dual action of kernels

We use the following notation for the expected value of a bounded function $f: M \rightarrow \mathbb{R}$ with respect to a probability measure ξ on M :

$$\xi f := \int_M f d\xi.$$

It will be useful to consider $K(x, dy)$ as a stochastic kernel which operates on probability measures from the left and bounded functions on the right. More precisely, by a *stochastic kernel* we mean an operator P such that

$$\xi P := \int_M \xi(dx) P(x, \cdot), \quad Pf := \int_M P(\cdot, dy) f(y),$$

are maps preserving, respectively, the complete metric space \mathcal{P} of probability measures on M with the total variation norm, and the Banach space \mathcal{B} of bounded functions on M with the sup norm. Informally, ξK is the distribution of J_n , and $\xi K f$ is the expected value of $f(J_n)$, given that J_{n-1} was distributed according to ξ . In particular, K^n is a stochastic kernel for each $n \geq 0$, with $\mathbb{P}^\xi(J_n \in \cdot) = \xi K^n$ and $\mathbb{E}^\xi[f(J_n)] = \xi K^n f$, where $K^0 = 1$ is the identity operator.

3. Invariant measure and MFPT

3.1. Stationary distribution on the milestones

The formula for the MFPT relies on a stationary distribution, which we denote by μ , for (J_n) . Our μ is the same as the appropriately normalized stationary flux q in other Milestoning papers. We use the notation μ instead of q to emphasize that μ is a probability measure on M that may not have a density. We show in Theorem 3.1 below that μ exists. The proof of existence requires only the above assumption of finite MFPTs to P and non-explosivity of (Y_t) , along with the special source and sink structure of the dynamics implied by (2.2).

Theorem 3.1— (J_n) has an invariant probability measure μ defined by

$$\mu(\cdot) := \mathbb{E}^\rho \left[\sum_{n=0}^{\sigma_P} \mathbb{1}_{\{J_n \in \cdot\}} \right] \mathbb{E}^\rho[\sigma_P + 1]^{-1}.$$

Proof: Define $\nu(\cdot) = \mathbb{E}^\rho \left[\sum_{n=0}^{\sigma_P} \mathbb{1}_{\{J_n \in \cdot\}} \right]$ and observe that

$$\nu(\cdot) = \sum_{n=0}^{\infty} \sum_{m=0}^n \mathbb{P}^\rho(J_m \in \cdot | \sigma_P = n) \mathbb{P}^\rho(\sigma_P = n) = \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_n \in \cdot, \sigma_P \geq n).$$

If $C \cap R = \emptyset$, by bounded convergence,

$$\int_M \nu(dx) K(x, C) = \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_{n+1} \in C, \sigma_P \geq n) = \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_n \in C, \sigma_P \geq n) = \nu(C),$$

where the second equality uses $\mathbb{P}^\rho(J_0 \in C) = 0$ and $J_{n+1} \notin R \Rightarrow \sigma_P \geq n$. If $C \subset R$,

$$\begin{aligned} \int_M \nu(dx) K(x, C) &= \sum_{n=0}^{\infty} \mathbb{P}^{\rho}(J_{n+1} \in C, \sigma_P = n) + \sum_{n=0}^{\infty} \mathbb{P}^{\rho}(J_{n+1} \in C, \sigma_P \geq n+1) \\ &= \rho(C) - \mathbb{P}^{\rho}(J_0 \in C, \sigma_P \geq 0) + \sum_{n=0}^{\infty} \mathbb{P}^{\rho}(J_n \in C, \sigma_P \geq n) = \nu(C). \end{aligned}$$

We will show below that (J_n) converges to μ under appropriate conditions. In that case μ is unique and we will call μ the *stationary measure*. We emphasize that the formula on which Exact Milestoning is based – the equation in Theorem 3.3 below – holds true whether or not μ is unique. However, a successful application of the Exact Milestoning algorithm will require some technique for sampling μ . The algorithm we present (Algorithm 1 below) is based on convergence of (J_n) to μ . We consider two types of convergence in Theorem 3.4 and 3.5 below. We also discuss a simple case where (J_n) may not converge following Theorem 3.4 below.

It is worth noting that the proof of Theorem 3.1 leads to the following representation of μ as a Neumann series. The representation is given in Corollary 3.2 below. In principle, this representation can be used to sample μ even when (J_n) does not converge to μ . The Neumann series is written in terms of the sub-stochastic *transient kernel* \bar{K} defined by

$$\bar{K}(x, dy) = \begin{cases} K(x, dy), & x \notin P \\ 0, & x \in P \end{cases} \quad (3.1)$$

The transient kernel \bar{K} has the same right action as the stochastic kernel K , but its left action is on the larger Banach space of signed bounded measures with total variation norm. Informally, it corresponds to a version of (J_n) that is “killed” when upon reaching P .

Corollary 3.2—We have

$$\lim_{n \rightarrow \infty} \left\| \nu(M)^{-1} \sum_{i=0}^{n-1} \rho \bar{K}^i - \mu \right\|_{TV} = 0. \quad (3.2)$$

Proof: Recall that $\mu = \nu(M)$ where

$$\nu(\cdot) = \sum_{n=0}^{\infty} \mathbb{P}^{\rho}(J_n \in \cdot, \sigma_P \geq n) = \sum_{n=0}^{\infty} \rho \bar{K}^n.$$

Moreover,

$$\sup_{|f| \leq 1} \left| \nu(M)^{-1} \sum_{i=0}^{n-1} \rho \bar{K}^i f - \mu f \right| \leq \nu(M)^{-1} \sum_{i=n}^{\infty} \mathbb{P}^{\rho}(\sigma_P \geq i), \quad (3.3)$$

and the right hand side of (3.3) is summable since by assumption $\mathbb{E}^\rho[\sigma_P] < \infty$.

3.2. Milestoning equation for the MFPT

Equipped with an invariant measure μ , we are now able to state the Milestoning equation (3.4) for the MFPT. In Exact Milestoning, this equation is used to efficiently compute the MFPT. The algorithm is based on two principles: first, many trajectories can be simulated in parallel to estimate τ_M^x for various x ; and second, the stationary distribution μ can be efficiently estimated through a technique based on power iteration. (See the right hand side of equation (3.4).) The gain in efficiency comes from the fact that the trajectories used to estimate τ_M^x are much shorter than trajectories from R to P . The second principle above depends somewhat on whether we have a good “initial guess” of μ . When (X_t) is Langevin dynamics, we have found in some cases the canonical Gibbs distribution is a sufficiently good initial guess. See [3] and [4] for details and discussion.

Theorem 3.3—Let μ be defined as above. Then $\mu(P) > 0$ and

$$\mu(P)\mathbb{E}^\rho[\tau_P] = \int_M \mu(dx) \mathbb{E}^x[\tau_M^x] := \mathbb{E}^\mu[\tau_M]. \quad (3.4)$$

Proof: The assumption $\mathbb{E}^\rho[\sigma_P] < \infty$ shows that $\mu(P) > 0$. For any $x \in M$,

$$\begin{aligned} \mathbb{E}^x[\tau_P] &= \int_M \mathbb{E}^x[\tau_P | Y_{\tau_M^x} = y] K(x, dy) \\ &= \int_M \mathbb{E}^x[\tau_M^x | Y_{\tau_M^x} = y] K(x, dy) + \int_{M \setminus P} \mathbb{E}^x[\tau_P - \tau_M^x | Y_{\tau_M^x} = y] K(x, dy) \\ &= \mathbb{E}^x[\tau_M^x] + \int_{M \setminus P} \mathbb{E}^y[\tau_P] K(x, dy). \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}^\mu[\tau_P] &= \mathbb{E}^\mu[\tau_M] + \int_{M \setminus P} \int_M \mu(dx) \mathbb{E}^y[\tau_P] K(x, dy) \\ &= \mathbb{E}^\mu[\tau_M] + \int_{M \setminus P} \mu(dy) \mathbb{E}^y[\tau_P], \end{aligned}$$

and so

$$\mathbb{E}^\mu[\tau_M] = \int_P \mu(dy) \mathbb{E}^y[\tau_P] = \mu(P) \mathbb{E}^\rho[\tau_P].$$

In Section 4 below we present the Exact Milestoning algorithm (Algorithm 1) recently used in [3] and [4]. The algorithm uses a technique which combines coarse-graining and power iteration to sample μ . Consistency of power iteration algorithms are justified via Theorem 3.4 below, where we show $\xi K^n \rightarrow \mu$ as $n \rightarrow \infty$. Though we emphasize that there are a range of possibilities for sampling μ (for example, algorithms based on (3.2) or (3.7) below) we note that Algorithm 1 was shown to be efficient for computing the MFPT in the entropic barrier example of [3] and the random energy landscapes example of [4].

3.3. Convergence to stationarity

In this section we justify the consistency of power iteration-based methods for sampling μ by showing that ξK^n converges to μ in total variation norm as $n \rightarrow \infty$. The theorem requires an extra assumption – aperiodicity of the jump chain (J_n) .

Theorem 3.4—Suppose that (J_n) is aperiodic in the following sense:

$$g.c.d.\{n \geq 1: \mathbb{P}^p(\sigma_p = n-1) > 0\} = 1. \quad (3.5)$$

Then for all probability measures ξ on M ,

$$\lim_{n \rightarrow \infty} \|\mathbb{P}^\xi(J_n \in \cdot) - \mu\|_{TV} = 0. \quad (3.6)$$

In particular, μ is unique.

Proof: We use a simple coupling argument. Let (H_n) be an independent copy of (J_n) and let $J_0 \sim \xi$ and $H_0 \sim \mu$. For $n \geq 0$, let S_n (resp. T_n) be the times at which (J_n) (resp. (H_n)) hit P for the $(n+1)$ st time. Then $S_{n+1} - S_n, n \geq 0$, are iid random variables with finite expected value and nonlattice distribution, and $(S_{n+1} - S_n)_{n \geq 0} \sim (T_{n+1} - T_n)_{n \geq 0}$. It follows that $(S_n - T_n)_{n \geq 0}$ is a mean zero random walk with nonlattice step distribution. Thus, its first time to hit 0 is finite almost surely. So

$$\zeta := \inf\{n \geq 0: J_n \in P, H_n \in P\}$$

obeys $\mathbb{P}(\zeta < n) \rightarrow 0$ as $n \rightarrow \infty$. Note that $J_n \sim H_n$ whenever $\zeta < n$. Thus

$$|\mathbb{P}^\xi(J_n \in C) - \mathbb{P}^\mu(H_n \in C)| \leq 2\mathbb{P}(\zeta \geq n).$$

Since μ is stationary for (H_n) we have $\mathbb{P}^\mu(H_n \in C) = \mu(C)$. Now

$$\|\mathbb{P}^\xi(J_n \in \cdot) - \mu\|_{TV} = \sup_{C \subset M} |\mathbb{P}^\xi(J_n \in C) - \mu(C)| \leq 2\mathbb{P}(\zeta \geq n),$$

which establishes the convergence result. To see uniqueness, suppose ξ is another invariant probability measure for (J_n) ; then the last display becomes $\|\xi - \mu\|_{TV} \leq 2\mathbb{P}(\zeta < n)$. Letting $n \rightarrow \infty$ shows that $\xi \sim \mu$.

We now consider a class of problems where there is a smooth one-dimensional *reaction coordinate* $\psi: \Omega \rightarrow [0, 1]$ tracking progress of (X_t) from R to P . In this case $\psi|_R \equiv 0, \psi|_P \equiv 1$, the milestones M_1, \dots, M_m are disjoint level sets of ψ , and $R = M_1, P = M_m$. The jump chain (J_n) can only hop between neighboring milestones, unless it is at P . That is, if $J_n \in M_i$ for $i \notin \{1, m\}$, then $J_{n+1} \in M_{i-1}$ or $J_{n+1} \in M_{i+1}$; if $J_n \in M_1$ then $J_{n+1} \in M_2$; and if $J_n \in M_m$,

then $J_{n+1} \in M_1$. Suppose that if $J_n \in M_j$ for $i \notin \{1, m\}$, then $J_{n+1} \in M_{i-1}$ with probability in $(0, 1)$. Such systems satisfy the aperiodicity assumption (3.5) if and only if m is odd. This is due to the fact that, starting at M_1 , (J_n) can reach M_m in either $m-1$ or $m+1$ steps with positive probability, and m and $m+2$ are coprime when m is odd. On the other hand, if m is even then the conclusion of Theorem 3.4 cannot hold. To see this, let m be even and start (J_n) at an initial distribution supported on an odd-indexed milestone. Then after any even number of steps (J_n) is supported on an odd-indexed milestone, while after any odd number of steps (J_n) is supported on an even-indexed milestone.

Theorem 3.4 establishes convergence of the time marginals of (J_n) , that is, the distributions of J_n at fixed times n . (J_n) also converges in a time-averaged sense, even if (3.5) does not hold. More precisely, we have the following version of the Birkhoff ergodic theorem:

Theorem 3.5—Let $J_0 \sim \xi$, with ξ a probability measure on M . For bounded $f: M \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(J_i) \stackrel{a.s.}{=} \int_M f d\mu \equiv \mu f. \quad (3.7)$$

Proof: Let S_n be the times at which (J_n) hits P for the $(n+1)$ st time, and define

$$f_n = \sum_{i=S_n+1}^{S_{n+1}} f(J_i).$$

Note that $f_n, n \geq 0$, are iid. Let $k(n) = \max\{k : S_k \leq n\}$ and write

$$\frac{1}{n} \sum_{i=1}^{n-1} f(J_i) = \frac{1}{n} \sum_{i=0}^{S_0} f(J_i) + \frac{1}{n} \sum_{i=0}^{k(n)-1} f_i + \frac{1}{n} \sum_{i=S_{k(n)}+1}^n f(J_i)$$

Since (J_n) hits P in finite time a.s., $n - S_{k(n)}$ and S_0 are finite a.s. Thus,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(J_i) = \lim_{n \rightarrow \infty} \frac{k(n)-1}{n} \frac{1}{k(n)-1} \sum_{i=0}^{k(n)-1} f_i.$$

Notice $R_n := S_{n+1} - S_n, n \geq 0$ are iid with finite expectation and

$$\frac{R_0 + \dots + R_{k(n)-1}}{k(n)} \leq \frac{n - S_0}{k(n)} \leq \frac{R_0 + \dots + R_{k(n)}}{k(n)}.$$

By the previous two displays and the law of large numbers,

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f(J_i) = \frac{\mathbb{E}[f_0]}{\mathbb{E}[R_0]} = \frac{1}{\mathbb{E}[\sigma_P]} \sum_{i=0}^{\sigma_P} \mathbb{E}^\rho[f(J_i)] = \frac{\nu f}{\nu(M)} = \mu f,$$

with ν defined as in Theorem 3.4.

Theorem 3.5 shows that problems in sampling μ arising from aperiodicity can be managed by averaging over time.

Markov chains for which the conclusion of Theorem 3.4 hold are called *Harris ergodic*. It is worth noting that Harris ergodicity, together with aperiodicity of the sojourn times, also leads to a limit for the time marginals of (Y_t) . More precisely, suppose (3.5) holds and for each $x \in M \setminus P$ and $y \in M$, $\mathbb{P}^x(\tau_1 \in \cdot \mid J_1 = y)$ is nonlattice. Then for any $C \subset M$ and μ -a.e. x ,

$$\lim_{t \rightarrow \infty} \mathbb{P}^x(Y_t \in C) = \frac{\int_C \mu(dy) \mathbb{E}^y[\tau_M^y]}{\int_M \mu(dy) \mathbb{E}^y[\tau_M^y]}. \quad (3.8)$$

See [2] for details and proof. When the right hand side of (3.8) has a density, it is the same as the stationary probability density $p(x)$ in [3] for the last milestone point passed.

4. Exact Milestoning algorithm

We now describe in detail an algorithm for sampling the stationary measure μ and the MFPT $\mathbb{E}^\rho[\tau_P]$, used successfully for instance in [3] and [4]. Let ξ be an initial guess for μ . For example, if (X_t) is Brownian or Langevin dynamics, we usually take ξ to be the canonical Gibbs distribution. We assume here that the conclusion of Theorem 3.4 holds. We write M_i for the distinct milestones, so that $M = \cup_i M_i$ and M_i, M_j have disjoint interiors when $i \neq j$. The algorithm will produce approximations

$$\xi \equiv \mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \dots$$

of μ . Let $\mu_i^{(n)}$ be the *non-normalized* restriction of $\mu^{(n)}$ to M_i , and define

$$\mathbb{E}^{\mu_i^{(n)}}[\tau_M] := \mu^{(n)}(M_i)^{-1} \int_{M_i} \mu_i^{(n)}(dx) \mathbb{E}^x[\tau_M^x],$$

(This is a slight abuse of notation, since $\mu_i^{(n)}$ is not a probability distribution: it is not normalized.) For $C \subset M_j$ we will also use the notation

$$a_{ij}^{(n)}(C) = \mu^{(n-1)}(M_i)^{-1} \int_{M_i} \mu_i^{(n-1)}(dx) K(x, C).$$

Below we think of $a_{ij}^{(n)}$ and $\mu_i^{(n)}$ as either distributions or densities. The $a_{ij}^{(n)}$ are obtained based on observations of trajectory fragments. We can conceive a simple Monte Carlo scheme for the purpose of approximating these distributions as follows. Let x_1, \dots, x_L be iid samples from the distribution $\mu_i^{(n-1)}/\mu^{(n-1)}(M_i)$. From each i sample point $x_\ell \in M_i$, we simulate (X_t) until it hits a neighboring milestone, say at the point $y_\ell \in M_j$. If we idealize by assuming the simulation of (X_t) is done exactly, then by Chebyshev's inequality,

$$\mathbb{P} \left(\left| a_{ij}^{(n)}(C) - \int_C \frac{1}{L} \sum_{\ell=1}^L \delta_{y_\ell}(dy) \right| > \varepsilon \right) \leq \frac{a_{ij}^{(n)}(C) - a_{ij}^{(n)}(C)^2}{L\varepsilon^2},$$

where δ_y is the Dirac delta distribution at y . We therefore write, for $y \in M$,

$$a_{ij}^{(n)}(y) \approx \frac{1}{L} \sum_{\ell=1}^L \tilde{\delta}_{y_\ell}(y), \quad (4.1)$$

where $\tilde{\delta}_z$ is either some suitable approximation to the identity at z , or simply a delta function at z . Thus, in Algorithm 1 we think of $a_{ij}^{(n)}$ and $\mu_i^{(n-1)}$ as either densities in the former case, or as distributions in the latter. The local mean first passage times, that is, the times between milestone crossings, are approximated by the sample means

$$\mathbb{E}^{\mu_i^{(n-1)}}[\tau_M] \approx \frac{1}{L} \sum_{\ell=1}^L \tau_M^{x_\ell}.$$

It is important to realize that we do not need to store the full coordinates of each y_ℓ in memory. Instead, it suffices to use a data-structure that keeps track of the pairs (y_ℓ, M_j) . The actual coordinates of each point can be written to disk and read from it as needed.

Algorithm 1

Exact Milestoning algorithm.

Input: Milestones $M = \cup_{j=1}^m M_j$, initial guess ξ , and tolerance $\varepsilon > 0$ for the absolute error in the MFPT.

Output: Estimates for μ , local MFPTs $\mathbb{E}^{\mu}[\tau_M]$, and overall MFPT $\mathbb{E}^{\rho}[\tau_p]$.

$$\mu^{(0)} \leftarrow \xi$$

$$\mathcal{T}^{(0)} \leftarrow +\infty$$

for all $n = 1, 2, \dots$ **do**

for $i = 1$ **to** m **do**

 Estimate $a_{ij}^{(n)}$ and $\mathbb{E}^{\mu_i^{(n-1)}}[\tau_M]$

$$\mathbf{A}_{ij}^{(n)} \leftarrow a_{ij}^{(n)}(M_j)$$

end for

 Solve $\mathbf{w}^T \mathbf{A} = \mathbf{w}^T$ (with $\mathbf{A} = (\mathbf{A}_{ij}^{(n)}) \in \mathbb{R}_{\geq 0}^{m \times m}$ and $\mathbf{w} = (w_1, \dots, w_m) \in \mathbb{R}_{\geq 0}^m$)

for $j = 1$ **to** m **do**

$$\mu_j^{(n)} \leftarrow \sum_{i=1}^m w_i a_{ij}^{(n)}$$

end for

 Normalize $\mu^{(n)}$

$$\mathcal{T}^{(n)} \leftarrow \mu^{(n)T} \mathbb{E}^{\mu^{(n-1)}}[\tau_M]$$

if $|\mathcal{T}^{(n)} - \mathcal{T}^{(n-1)}| < \varepsilon$ **then**

break

end if

end for

return $(\mu^{(n)}, \mathbb{E}^{\mu^{(n-1)}}[\tau_M], \mathcal{T}^{(n)})$

The eigenvalue problem in Algorithm 1 involves a stochastic matrix $\mathbf{A} \in \mathbb{R}_{\geq 0}^{m \times m}$ that is sparse. Indeed, the i -th row corresponds to milestone M_i and may have only as many non-zero entries as the number of neighboring milestones M_j . In practice, to solve the eigenvalue problem we can use efficient and accurate Krylov subspace solvers [15] such as Arnoldi iteration [18] to obtain \mathbf{w} without computing all the other eigenvectors.

In Algorithm 1, if $\mathbf{w}_i := \mu_i^{(n-1)}(M_i)$ is used instead of the solution \mathbf{w} to $\mathbf{w}^T \mathbf{A} = \mathbf{w}^T$, then the algorithm approximates μ by simple power iteration. The reason for defining the weights as the solution to $\mathbf{w}^T \mathbf{A} = \mathbf{w}^T$ is practical: we have found that it gives faster convergence of the iterations, at no apparent cost to accuracy. It can be seen as a version of power iteration that uses coarse-graining. See [3, 4] for applications of the algorithm in Exact Milestoning and [15, 18] for related discussions.

5. Error analysis

5.1. Stationary distribution error

In practice, due to time discretization error, we cannot generate trajectories exactly according to the transition kernel K . Instead, we can generate trajectories according to a numerical approximation K_ε . We investigate here whether such schemes are consistent, that is, whether powers of K_ε of K converge to a distribution $\mu_\varepsilon \approx \mu$. We emphasize that, even though we account for time discretization here, we still assume infinite sampling, and thus for a given $x \in M$, $K_\varepsilon(x, dy)$ may be a continuous distribution. See Section 5.2 below for related remarks and a discussion of how time discretization errors affect the Exact Milestoning estimate of the MFPT.

The following theorem, restated from [11], establishes consistency of iteration schemes based on Theorem 3.4 when K_ε is sufficiently close to K and (J_n) is geometrically ergodic. After the theorem, in Lemma 5.2 and Theorem 5.3 we give natural conditions for geometric ergodicity of (J_n) .

Theorem 5.1—Suppose (J_n) is geometrically ergodic: there exists $\kappa \in (0, 1)$ such that

$$\sup_{x \in M} \|\delta_x K^n - \mu\|_{TV} = O(\kappa^n).$$

Let $\{K_\varepsilon\}$ be a family of stochastic kernels with $K_0 = K$, assumed to act continuously on \mathcal{B} , such that

$$\lim_{\varepsilon \rightarrow 0} \sup_{|f| \leq 1} \|K_\varepsilon f - K f\|_\infty = 0. \quad (5.1)$$

Then for each $\hat{\kappa} \in (\kappa, 1)$, there is $\delta > 0$ such that for each $\varepsilon \in [0, \delta)$, K_ε has a unique invariant probability measure μ_ε , and

$$\begin{aligned} \sup_{\varepsilon < \delta} \sup_{x \in M} \|\delta_x K_\varepsilon^n - \mu_\varepsilon\|_{TV} &= O(\hat{\kappa}^n), \\ \lim_{\varepsilon \rightarrow 0} \|\mu_\varepsilon - \mu\|_{TV} &= 0. \end{aligned}$$

Geometric ergodicity is inconvenient to check directly. We give two sufficient conditions for geometric ergodicity of (J_n) . The first condition is a uniform lower bound on the probability to reach P in N steps; see Lemma 5.2. We use this to obtain a strong Feller condition in Theorem 5.3. The latter is a very natural condition and is easy to verify in some cases, for instance when (X_t) is a nondegenerate diffusion and the milestones are sufficiently regular.

Lemma 5.2—Suppose that there exists $\lambda \in (0, 1)$ and $N \in \mathbb{N}$ such that for all $x \in M$, $\mathbb{P}^x(J_{N-1} \in P) \geq \lambda > 0$. Then (J_n) is geometrically ergodic:

$$\sup_{x \in M} \|\delta_x K^n - \mu\|_{TV} \leq \lambda^{-1} (1 - \lambda)^{\lfloor n/N \rfloor}.$$

Proof: Let $\xi_1, \xi_2 \in \mathcal{P}$, consider the signed measure $\xi = \xi_1 - \xi_2$ and compute

$$\begin{aligned} \|\xi_1 K^N - \xi_2 K^N\|_{TV} &= \sup_{|f| \leq 1} \left| \int_M \int_M \xi(dy) K^N(y, dz) f(z) \right| \\ &= \sup_{|f| \leq 1} \left| \int_M \int_M \xi(dy) (K^N(y, dz) - \lambda \rho(dz)) f(z) \right| \\ &= \sup_{|f| \leq 1} \left| \int_M \xi(dy) \int_M (K^N(y, dz) - \lambda \rho(dz)) f(z) \right| \\ &\leq (1 - \lambda) \sup_{|f| \leq 1} \left| \int_M \xi(dy) f(y) \right| = (1 - \lambda) \|\xi_1 - \xi_2\|_{TV}. \end{aligned}$$

The last line uses the fact that $K^N(y, dz) - \lambda \rho(dz)$ is a positive measure. This shows that K^N is a contraction mapping on \mathcal{P} with contraction constant $(1 - \lambda)$. Observe also that $\|\xi_1 K - \xi_2 K\|_{TV} \leq \|\xi_1 - \xi_2\|_{TV}$. The result now follows from the contraction mapping theorem. See for instance Theorem 6.40 of [9].

Note that the λ in Lemma (5.2) is a quantity that can be estimated, at least in principle, by running trajectories of (X_t) starting at x which cross $N - 1$ milestones before reaching P . However, this is likely impractical for the same reason direct estimation of the MFPT is impractical – the trajectories would be too long. One alternative would be to compute the probability $P^i(J_{N-1}^A \in P)$ for the Markov chain (J_n^A) on $\{1, \dots, m\}$ with transition matrix \mathbf{A} , and use the minimum over $i \in \{1, \dots, m\}$ as a proxy for λ . Even without a practical way to estimate λ , we believe the characterization of Lemma 5.2 is useful for understanding the convergence rate.

Lemma 5.2 leads to the following condition for geometric ergodicity of (J_n) .

Theorem 5.3—Suppose that M is compact and (J_n) is a strong Feller chain which is aperiodic in the sense of (3.5). Then (J_n) is geometrically ergodic.

Proof: Let $\varepsilon \in (0, \mu(P))$. By Theorem 3.4, for each $x \in M$ there is $N_x \in \mathbb{N}$ such that $\mathbb{P}^x(J_n \in P) \geq \varepsilon$ for all $n \geq N_x$. Because (J_n) is strong Feller, the map $x \rightarrow \mathbb{P}^x(J_n \in P)$ is continuous. By compactness of M , it follows that for any $\lambda \in (0, \varepsilon)$ there is $N \in \mathbb{N}$ such that $\mathbb{P}^y(J_n \in P) \geq \lambda$ for all $y \in M$ and $n \geq N - 1$. Theorem 5.2 now yields the result.

5.2. MFPT error

As discussed above, Equation 3.4 can be used to estimate the MFPT $E^\mu[\tau_P]$ by sampling μ and local MFPTs τ_M^x . The error in this estimate has two sources. First, in general we only have an approximation $\tilde{\mu} \equiv \mu_\varepsilon$ of μ . The second source of error is in the sampling of τ_M^x , due to the fact that we can only simulate a time discrete version $(X_{n\delta t})$ of (X_t) . In Theorem 5.4 below we give an explicit formula for the numerical error of the MFPT in terms of these two

sources. We first need the following notation. Let $\tilde{\tau}_M^x$ be the minimum of all $n \delta t > 0$ such that the line segment between $X_{n\delta t}$ and $X_{(n+1)\delta t}$ intersects $M \setminus M_x$, and define

$$\mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M] := \int_M \tilde{\mu}(dx) \mathbb{E}^x[\tilde{\tau}_M^x].$$

We emphasize that Theorem 5.4 below gives an expression for the error in the *original* Milestoning as well as in Exact Milestoning.

Theorem 5.4—There exists a nonnegative function ϕ such that

$$|\mathbb{E}^\rho[\tau_P] - \tilde{\mu}(P)^{-1} \mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]| \leq c_1 |\mu(P)^{-1} - \tilde{\mu}(P)^{-1}| + \tilde{\mu}(P)^{-1} (c_2 \|\mu - \tilde{\mu}\|_{TV} + \phi(\delta t)), \quad (5.2)$$

where

$$c_1 := \mathbb{E}^\mu[\tau_M], \quad c_2 := \sup_{x \in M} \mathbb{E}^x[\tau_M^x].$$

Proof: Note that

$$\begin{aligned} & |\mathbb{E}^\rho[\tau_P] - \tilde{\mu}(P)^{-1} \mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]| = |\mu(P)^{-1} \mathbb{E}^\mu[\tau_M] - \tilde{\mu}(P)^{-1} \mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]| \\ \leq & |\mu(P)^{-1} \mathbb{E}^\mu[\tau_M] - \tilde{\mu}(P)^{-1} \mathbb{E}^\mu[\tau_M]| + |\tilde{\mu}(P)^{-1} \mathbb{E}^\mu[\tau_M] - \tilde{\mu}(P)^{-1} \mathbb{E}^{\tilde{\mu}}[\tau_M]| + |\tilde{\mu}(P)^{-1} \mathbb{E}^{\tilde{\mu}}[\tau_M] - \tilde{\mu}(P)^{-1} \mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]| \end{aligned}$$

where we have written $\mathbb{E}^{\tilde{\mu}}[\tau_M] := \int_M \tilde{\mu}(dx) \mathbb{E}^x[\tau_M^x]$. We may write

$$\phi(\delta t) = |\mathbb{E}^{\tilde{\mu}}[\tau_M] - \mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]|$$

for the term depending only on time stepping error. Note that

$$\begin{aligned} |\mathbb{E}^\mu[\tau_M] - \mathbb{E}^{\tilde{\mu}}[\tau_M]| &= \left| \int_M \mu(dx) \mathbb{E}^x[\tau_M^x] - \int_M \tilde{\mu}(dx) \mathbb{E}^x[\tau_M^x] \right| \\ &\leq \left(\sup_{x \in M} \mathbb{E}^x[\tau_M^x] \right) \|\mu - \tilde{\mu}\|_{TV}. \end{aligned}$$

Combining the last three expressions yields the result.

Recall that in the above we have ignored errors from finite sampling. We now discuss the implications of those errors. In the original Milestoning, μ is the canonical Gibbs distribution on the milestones. In that setting, we can typically sample independently from μ on the milestones. Thus, the central limit theorem implies that the *true* error in the Milestoning approximation of $\mathbb{E}^\rho[\tau_P]$ is bounded above with high probability by the right hand side of

(5.2) plus a constant times $1/\sqrt{N}$, where N is the number of samples. An analogous argument applies to Exact Milestoning if μ is sampled by simple power iteration. For our coarse-grained version of power iteration, however, we obtain samples of μ which are not independent, and thus a more detailed analysis would be required to determine the additional error from finite sampling.

We do not analyze the time discretization error $\phi(\delta t)$ and instead refer the reader to [14] and references therein. Here we simply remark that, if (X_t) is a diffusion process, then under certain smoothness assumptions on the drift and diffusion coefficients of (X_t) and on M , we have $\phi(\delta t) = \theta(\sqrt{\delta t})$ when $(X_{n\delta t})$ is the standard Euler time discretization with time step δt . See [13] for details and proof and see [5, 16] for numerical schemes that mitigate time discretization error in the MFPTs.

6. Illustrative examples

In this section we present two examples of Milestoning on the torus $\Omega = \mathbb{R}^2/\mathbb{Z}^2$. The process (X_t) will be governed by the Brownian dynamics equation,

$$\begin{cases} dX_t = -\nabla U(X_t) dt + \sqrt{2\beta^{-1}} dB_t, \\ X_0 \sim \rho \end{cases}$$

where $U: \Omega \rightarrow \mathbb{R}$ is a smooth potential energy function, $\beta > 0$ is the inverse temperature, and (B_t) is a standard Brownian motion. For our computations, we consider a uniformly spaced mesh of milestones with fixed product and reactant sets P, R ; see Figure 2.

6.1. Constant potential energy

In the case where U is a constant function, the corresponding canonical Gibbs distribution is also constant. However, the stationary distribution μ obtained in Milestoning is not constant. This is because every trajectory reaching the product state P immediately restarts at R according to ρ , as determined by (2.2). The resulting (density of) μ is shown in Figure 3.

6.2. Rough energy landscape

In this case, we consider U of the form

$$U(x_1, x_2) = \sum_{k_1=-N}^{N-1} \sum_{k_2=-N}^{N-1} C_{k_1, k_2} f_{k_1, k_2}(x_1, x_2), \quad (6.1)$$

where N is a parameter determining the maximum allowed frequency. The constants C_{k_1, k_2} are given by

$$C_{k_1, k_2} = \begin{cases} c_{k_1, k_2}, & \text{with probability } \frac{1}{2}, \\ 0, & \text{with probability } \frac{1}{2}, \end{cases}$$

with c_{k_1, k_2} being uniformly distributed in the interval $(-\frac{1}{2\pi}, \frac{1}{2\pi})$. The random functions f_{k_1, k_2} are defined as

$$f_{k_1, k_2}(x_1, x_2) = \begin{cases} \cos(2\pi k_1 x_1) \cos(2\pi k_2 x_2), & \text{with probability } 1/3, \\ \cos(2\pi k_1 x_1) \sin(2\pi k_2 x_2), & \text{with probability } 1/3, \\ \sin(2\pi k_1 x_1) \sin(2\pi k_2 x_2), & \text{with probability } 1/3. \end{cases}$$

Therefore, the potential energy function U is a sparse, periodic function on the two-dimensional torus. An example of a stationary distribution on a particular realization of the above potential energy functions is displayed in Figure 4, where we see that (the density of) μ is concentrated in a few basins of attraction.

Acknowledgments

This research was supported in part by a grant from the NIH GM59796 and from the Welch Foundation Grant No. F-1783.

References

1. Allen, P.; Tildesley, DJ. *Computer Simulation of Liquids*. Oxford Science Publications, Clarendon Press; 1989.
2. Alsmeyer G. On the Markov renewal theorem. *Stochastic Processes and their Applications*. 1994; 50:37–56.
3. Bello-Rivas JM, Elber R. Exact milestoning. *The Journal of Chemical Physics*. 2015; 142:094102. [PubMed: 25747056]
4. Bello-Rivas JM, Elber R. Simulations of thermodynamics and kinetics on rough energy landscapes with milestoning. 2015
5. Bouchard B, Geiss S, Gobet E. First time to exit of a continuous Itô process: general moment estimates and L^1 -convergence rate for discrete time approximations. 2013:1–31.
6. Cardenas AE, DeLeon KY, Hegefled WA, Kuczera K, Jas GS, Elber R. Unassisted Transport of Block Tryptophan through DOPC Membrane: Experiment and Simulation. 2012
7. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *Journal of Chemical Physics*. 2007; 126
8. Dellago C, Bolhuis PG, Geissler PL. *Transition Path Sampling*. 2002; 123
9. Douc, R.; Moulines, E.; Stoffer, D. *Chapman & Hall/CRC Texts in Statistical Science*. Taylor & Francis; 2014. *Nonlinear Time Series: Theory, Methods and Applications with R Examples*.
10. Faradjian AK, Elber R. Computing time scales from reaction coordinates by milestoning. *Journal of Chemical Physics*. 2004; 120:10880–10889. [PubMed: 15268118]
11. Ferré D, Hervé L, Ledoux J. Regular perturbation of v -geometrically ergodic markov chains. *Journal of Applied Probability*. 2013; 50:184–194.
12. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*. 2002.
13. Gobet E, Menozzi S. Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme. *Stochastic Processes and their Applications*. 2004; 112:201–223.
14. Gobet E, Menozzi S. Stopped diffusion processes: Boundary corrections and overshoot. *Stochastic Processes and their Applications*. 2010; 120:130–162.
15. Golub, GH.; Van Loan, CF. *Matrix computations*. 4. Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press; Baltimore, MD: 2013.

16. Higham DJ, Mao X, Roj M, Song Q, Yin G. Mean Exit Times and the Multilevel Monte Carlo Method. *SIAM Journal on Uncertainty Quantification*. 1(1):2–18.
17. Kirmizialtin S, Elber R. Revisiting and computing reaction coordinates with Directional Milestoning. *The journal of physical chemistry A*. 2011; 115:6137–48. [PubMed: 21500798]
18. Lehoucq, RB.; Sorensen, DC.; Yang, C. ARPACK users' guide, vol. 6 of Software, Environments, and Tools. Society for Industrial and Applied Mathematics (SIAM); Philadelphia, PA: 1998.
19. Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Transition Path Theory for Markov Jump Processes. 2009.
20. Moroni D, Bolhuis PG, Van Erp TS. Rate constants for diffusive processes by partial path sampling. *Journal of Chemical Physics*. 2004; 120:4055–4065. [PubMed: 15268572]
21. Ruymgaart AP, Cardenas AE, Elber R. MOIL-opt: Energy-Conserving Molecular Dynamics on a GPU/CPU system. *Journal of chemical theory and computation*. 2011; 7:3072–3082. [PubMed: 22328867]
22. Sarich, M.; Noé, F.; Schütte, C. On the Approximation Quality of Markov State Models. 2010.
23. Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. 2010.
24. Shaw, DE.; Chao, JC.; Eastwood, MP.; Gagliardo, J.; Grossman, JP.; Ho, CR.; Lerardi, DJ.; Kolossváry, I.; Klepeis, JL.; Layman, T.; McLeavey, C.; Deneroff, MM.; Moraes, MA.; Mueller, R.; Priest, EC.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, SC.; Dror, RO.; Kuskin, JS.; Larson, RH.; Salmon, JK.; Young, C.; Batson, B.; Bowers, KJ. Anton, a special-purpose machine for molecular dynamics simulation. 2008.
25. Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K. Accelerating molecular modeling applications with graphics processors. *Journal of Computational Chemistry*. 2007; 28:2618–2640. [PubMed: 17894371]
26. Swenson DWH, Bolhuis PG. A replica exchange transition interface sampling method with multiple interface sets for investigating networks of rare events. *Journal of Chemical Physics*. 2014; 141
27. Vanden-Eijnden E, Venturoli M. Exact rate calculations by trajectory parallelization and tilting. *The Journal of chemical physics*. 2009; 131:1–7.
28. Warmflash A, Bhimalapuram P, Dinner AR. Umbrella sampling for nonequilibrium processes. *The Journal of chemical physics*. 2007; 127:154112. [PubMed: 17949137]

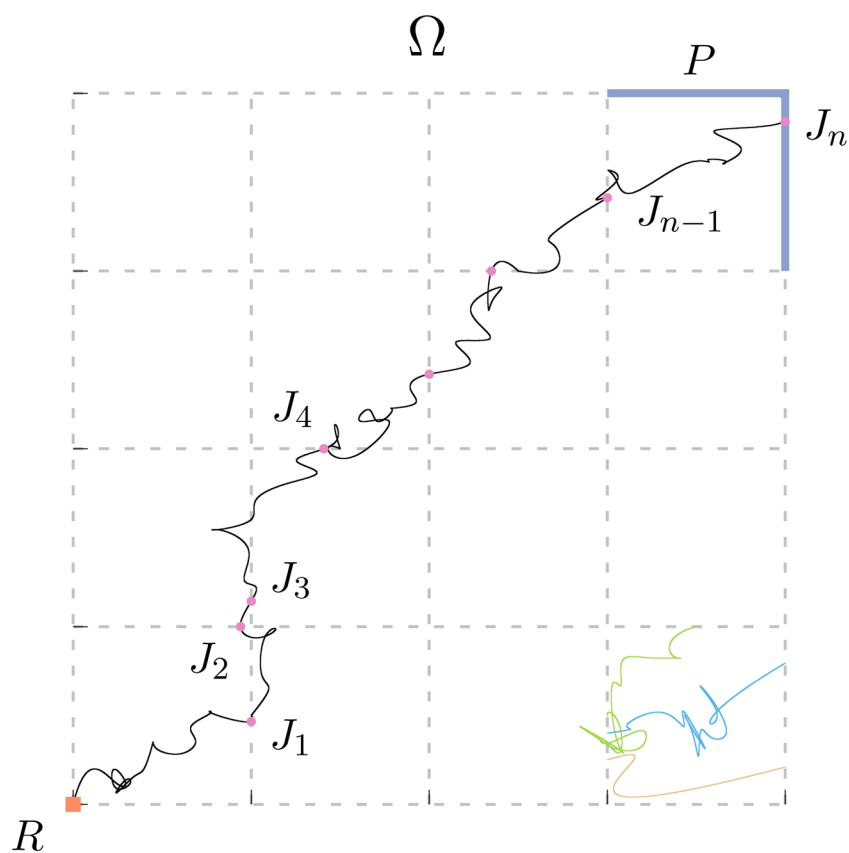


Fig. 1. Representation of the state space Ω and the milestones. Each milestone is an edge of one of the squares traced by dashed grey lines. The reactant state R is highlighted as a square dot in the left-bottom corner while the product state P is comprised of the two line segments shown in blue at the upper-right corner. A particular realization of a long trajectory appears as a continuous black line and the corresponding values of the process (J_n) are marked with circles. The individual trajectories shown in different colors near the bottom-right corner represent trajectory fragments as used in the Milestoning algorithm.

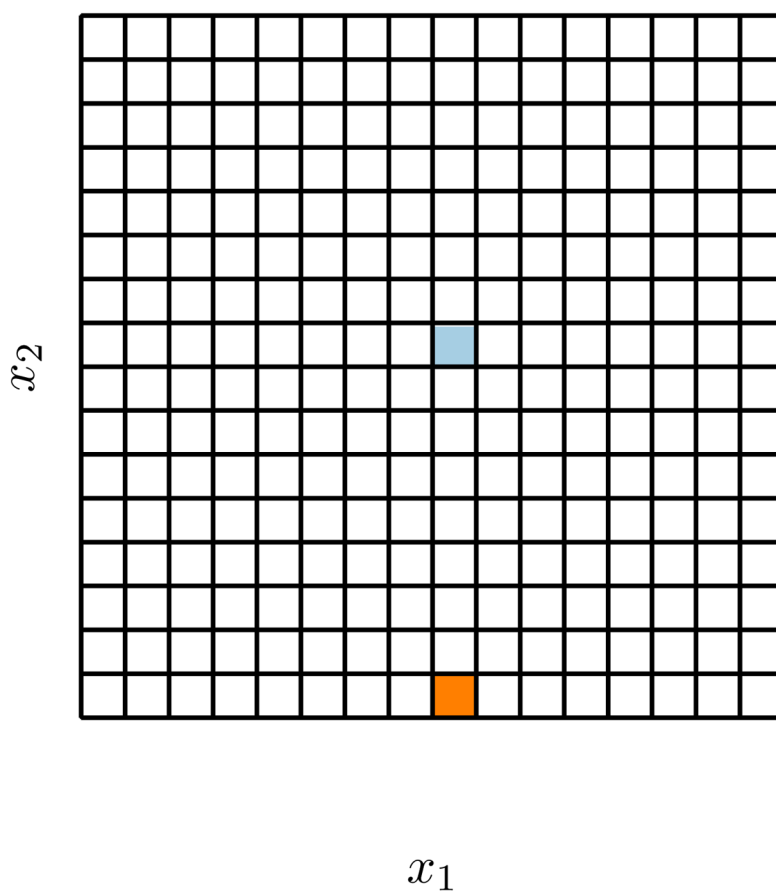


Fig. 2. Diagram showing the reactant state (orange square) and the product state (blue square) within the set of all milestones for the example in Section 6. Each milestone is an edge of one of the small squares in the diagram. (The total number of milestones has been decreased to enhance visibility.)

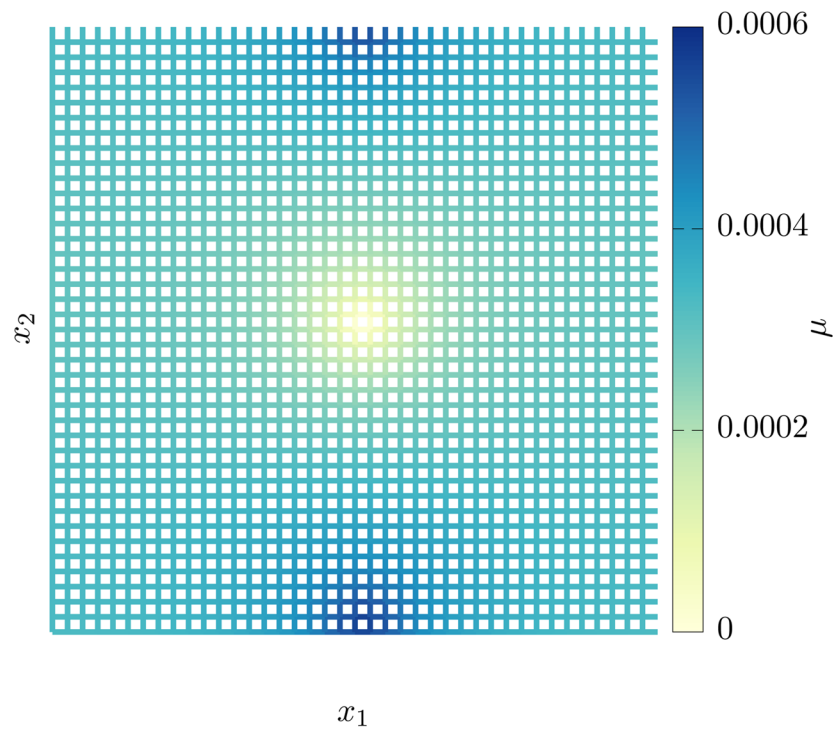


Fig. 3. Stationary distribution μ corresponding to a system with constant potential energy for the example in Section 6. There are $2 \times 40 \times 40$ total milestones.

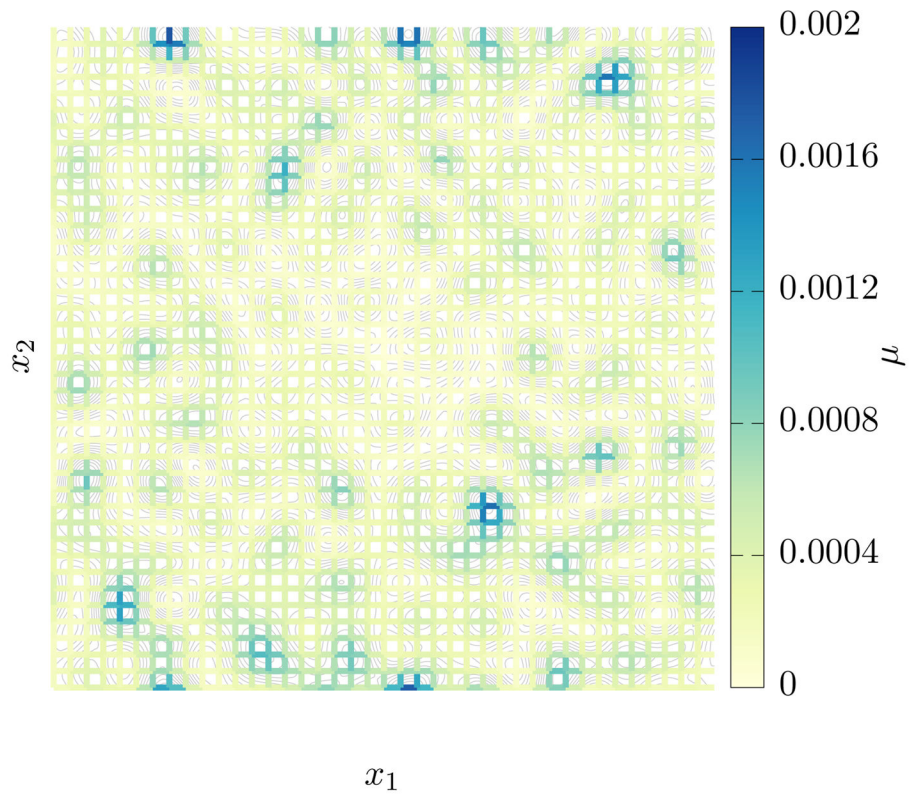


Fig. 4. Stationary density μ on a rough energy landscape for the example in Section 6. The contour lines are the level sets of U .

Algorithm 1

Exact Milestoning algorithm.

Input: Milestones $M = \cup_{j=1}^m M_j$, initial guess ξ , and tolerance $\varepsilon > 0$ for the absolute error in the MFPT.

Output: Estimates for μ , local MFPTs $\mathbb{E}^{\mu}[\tau_M]$, and overall MFPT $\mathbb{E}^{\rho}[\tau_p]$.

$$\mu^{(0)} \leftarrow \xi$$

$$\mathcal{T}^{(0)} \leftarrow +\infty$$

for all $n = 1, 2, \dots$ **do**

for $i = 1$ **to** m **do**

 Estimate $a_{ij}^{(n)}$ and $\mathbb{E}^{\mu^{(n-1)}}[\tau_M]$

$$\mathbf{A}_{ij}^{(n)} \leftarrow a_{ij}^{(n)}(M_j)$$

end for

Solve $\mathbf{w}^T \mathbf{A} = \mathbf{w}^T$ (with $\mathbf{A} = (\mathbf{A}_{ij}^{(n)}) \in \mathbb{R}_{\geq 0}^{m \times m}$ and $\mathbf{w} = (w_1, \dots, w_m) \in \mathbb{R}_{\geq 0}^m$)

for $j = 1$ **to** m **do**

$$\mu_j^{(n)} \leftarrow \sum_{i=1}^m w_i a_{ij}^{(n)}$$

end for

 Normalize $\mu^{(n)}$

$$\mathcal{T}^{(n)} \leftarrow \mu^{(n)}(P)^{-1} \mathbb{E}^{\mu^{(n-1)}}[\tau_M]$$

if $|\mathcal{T}^{(n)} - \mathcal{T}^{(n-1)}| < \varepsilon$ **then**

break

end if

end for

return $(\mu^{(n)}, \mathbb{E}^{\mu^{(n-1)}}[\tau_M], \mathcal{T}^{(n)})$
