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

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

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

$$0 = (\mathcal{L}^* \rho)(x) = \nabla \cdot (a(x) \nabla \rho(x) − b(x) \rho(x)).$$

($\mathcal{L}^*$ 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)

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

where $\sigma(x)$ satisfies $(\sigma^T \sigma)(x) = a(x)$ 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 $\beta$:

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

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

$$\dot{Q}(t) = M^{-1} P(t)$$

$$dP(t) = −\nabla V (X(t))dt − \gamma P(t)dt + \sqrt{2 \beta^{-1} \gamma M} dW(t),$$

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 $\tau_0 = H_{I, \mathcal{M}}^+(0)$. For each $t \geq \tau_0$, define $\Xi(t)$ to be the index of the last milestone hit by $X(t)$, i.e.,
\[
\Xi(t) = \text{index } i \in I \text{ such that } X(H_{I, \mathcal{M}}^-(t)) \in M_i.
\]
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 $\pm 1$. 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
\[
\alpha_n = \tau_n - \tau_{n-1}, \quad n \in \mathbb{N}.
\]
2. Main results. Having introduced the main objects used in milestoneing, 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 $\pi_i$ 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 viewpoint 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

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

where $\alpha_p^{(i,j)} = \tau_p^{(i,j)} - \tau_{p-1}^{(i,j)}$. 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 milestoneing 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

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

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

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

and $p_{i,j}$ is the probability that the trajectory hits the $j$th milestone after leaving the $i$th milestone, which can be defined as

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

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:

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

where

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

is the generator of the time-reversed diffusion process, and $\langle \cdot \rangle$ denotes the standard inner product on $\mathbb{R}^d$. (Note that $\mathcal{L}^L \neq \mathcal{L}$ 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 2.2 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.

Of optimal milestoning – that is, use isocommittor surfaces as milestones. We call this result as:

\[
\xi_n = \{ x : q^+(x) = z_n \}
\]

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

\[
\nu(x, B) = \mathbb{P}^x(X(\tau_1) \in B) = \mathbb{P}^x(Y_1 \in B),
\]

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. 5 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 $\mu$ satisfies

$$\mu(\cdot) = \int_{\cup_i M_i} \mu(dx) \nu(x, \cdot),$$

(11)

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

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

(12)

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

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

(13)

The distribution $\pi_i$ 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

$$\nu(x, dy) = \sum_{j \sim i} P_{i,j}(x) \nu_{i,j}(x, dy), \quad x \in M_i, \quad i \in I$$

(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.

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

(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

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

(16)

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

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

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

(17)

where

$$p_{j,i} = \int_{M_j} \mu_j(dx) P_{j,i}(x)$$

(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 $x \in \cup_{i \in I} M_i$ 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:

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

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

$$\tau(x) = T_j(x) - \int_{\cup_{i \neq j} M_i} \nu(x, dy) T_j(y), \quad x \in \cup_{i \neq j} M_j,$$

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 2.2, it is

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

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

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

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):

$$\int_{M_i} \mu_i(dx) \nu(x, B_j) = \mu_j(B_j), \quad \forall B_j \subset M_j, \quad \forall i \sim j$$

(22) \iff

$$\int_{M_i} \mu_i(dx) P_{i,j}(x) = \mu_j(B_j), \quad \forall B_j \subset M_j, \quad \forall i \sim j$$

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

$$t_i = T_{i,j} - \int_{M_i} \mu_i(dx) \sum_{k \sim i} P_{i,k}(x) \int_{M_k} \nu_{i,k}(x, dy) T_j(y)$$

(23) \iff

$$T_{i,j} = T_{i,j} - \sum_{k \sim i} p_{i,k} \int_{M_k} \mu_k(dy) T_j(y)$$

(by (22))

$$T_{i,j} = T_{i,j} - \sum_{k \sim i} p_{i,k} T_{k,j},$$

which is precisely (6). In other 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 (13) 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 $\int_{M_k} \nu(x, dy) T_j(y)$) 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 possible 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 milestoneing is based on the existence of invariant families of distributions, for which (22) hold. To introduce them precisely, let us define a shift operator $P^*$ of a probability measures $\mu$, associated with the first hitting chain $\{ Y_n : n \in \mathbb{N}\}$ by
\[
 (P^* \mu)(B) = \mathbb{P}_\mu [ Y_1 \in B ], \quad B \subset \cup_i I M_i,
\]
where we have used $\mathbb{P}_\mu$ to denote the law of the diffusion with the initial distribution $\mu$. Note that the kernel associated with $P^*$ is exactly $\nu(x, \cdot)$, the transition probability kernel of the first hitting chain used in (11), as
\[
 (P^* \mu)(B) = \int_{\mathcal{M}} \mu(\mathrm{d}x) \nu(x, B).
\]

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 $\mu_i$ concentrated on $M_i$ such that the conditional distribution of $P^* \mu_i$ given $M_i$, if it makes sense, is $\mu_i$. 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 $\mathcal{M} = \{ M_i : 0 \leq i \leq N \}$ as
\[
 M_i = \{ x \in \mathbb{R}^d \setminus (A \cup B)^c : q^-(x) = z_i \}, \quad \text{for } i \in I,
\]
where $1 \geq z_0 > z_1 > \cdots > z_N \geq 0$. 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
\[
 Z_i = \int_{M_i} \frac{\rho(x)}{|\nabla q^-(x)|} \langle a(x) \nabla q^-(x), \nabla q^-(x) \rangle \mathrm{d}\sigma_{M_i}(x), \quad i \in I
\]
are finite. Then we are able to define on each $M_i$ a probability measure $\mu_i$ with the density function
\[
 \rho_i(x) = Z_i^{-1} \frac{\rho(x)}{|\nabla q^-(x)|} \langle a(x) \nabla q^-(x), \nabla q^-(x) \rangle, \quad i \in I
\]
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 $\rho_i$ defined in (28) is an invariant family of distributions associated with $\mathcal{M}$. Actually we have
\[
 P^* \mu_i = \sum_{j \in I} q_{i, j} \mu_j, \quad \text{for } i \in I
\]
where $q_{i, j}$ is given by
\[
 q_{i, j} = \mathbb{P}_\mu [ \xi_1 = j ] = \begin{cases} 
 \frac{z_i - z_{i+1}}{z_i - z_{i+1}} & \text{if } j = i - 1, \\
 \frac{z_{i-1} - z_{i+1}}{z_{i-1} - z_i} & \text{if } j = i + 1, \\
 \frac{z_{i-1} - z_{i+1}}{z_{i-1} - z_{i+1}} & \text{otherwise}.
\end{cases}
\]
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 $M = \{M_i : i \in I\}$. Then there exists an invariant distribution $\mu$ for the first hitting chain $\{Y_n : n \in \mathbb{N}_0\}$ such that each $\mu_i$ is the conditional distribution of $\mu$ given $M_i$.

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

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

where $p_{i,j} = (P^*\mu_i)(M_j) = P_{\mu_i}[\xi_1 = j]$. It is easy to check that this agrees with $p_{i,j}$ defined in (17). Let $(\pi_{i,j})_{i,j \in I}$ be an invariant distribution for the transition matrix $(p_{i,j})_{i,j \in I}$. Then it is straightforward to check that $\mu = \sum_{i \in I} \pi_{i,i}$ 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 $M = \{M_i : i \in I\}$. Let $\mu$ 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 $i_k \in I$, $0 \leq k \leq n$,

$$E_\mu[\alpha_n | \xi_k = i_k, 0 \leq k \leq n] = E_\mu[\alpha_1 | \xi_0 = i_{n-1}, \xi_1 = i_n].$$

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 $h_{i,j} = E_\mu[\tau_D | \xi_0 = i]$. Then $(h_{i,j})_{i,j \in I}$ is the unique solution to the following discrete Poisson problem

$$h_{i,j} = \sum_{k \in I} p_{i,k}h_{i,k} + \sum_{k \in I} p_{i,k}h_{k,j}, \quad i \in I, \quad i \neq j,$$

$$h_{j,j} = 0,$$
where \( p_{i,j} = \mathbb{P}_\mu [\xi_1 = j \mid \xi_0 = i] \) and \( t_{i,j} = \mathbb{E}_\mu [\alpha_1 \mid \xi_0 = i, \xi_1 = j] \).

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 (3), 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

\[
T_{i,j} = \mathbb{E}_{\mu^{(i,j)}} \left[ \alpha_1^{(i,j)} \mid Y_0^{(i,j)} \in M_i \right] = \mathbb{E}_{\mu^{(i,j)}} \left[ \alpha_1^{(i,j)} \right],
\]

where \( \mu^{(i,j)} \) 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 \( \mathbb{P}_{\mu^{(i,j)}} \),

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

as asserted in (32). \( \square \)

We are now ready to justify (6) rigorously.

**Theorem 5.2.** Assume that \( \mathcal{M} = \{M_i : 0 \leq i \leq N\} \) 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 \( 0 \leq i \leq N \), let \( \mu_i \) be the probability measure concentrated on \( M_i \) with the density \( \rho_i \) given in (28). Also assume that \( \{Y_n^{(i,j)} : n \in \mathbb{N}_0\} \) is uniquely ergodic. Then the mean first passage times \( \{T_{i,j} \}_{i,j} \) is the unique solution to the following discrete Poisson problem:

\[
\begin{align*}
T_{i,j} &= t_i + \sum_{k \in I} q_{i,k} T_{k,j}, \quad i \in I, \quad i \neq j, \\
T_{j,j} &= 0.
\end{align*}
\]

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

\[
t_i = \sum_{j \in I} q_{i,j} \mathbb{E}_{\mu_i} [\alpha_1 | \xi_1 = j] = \mathbb{E}_{\mu_i} [\alpha_1].
\]
\textbf{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 $H_{M_i}^+(0) = D$ (recall $D$ from Lemma 4.5) due to $\lim_{n \to \infty} \tau_n = \infty$ (see the proof of Lemma 4.7). So by Proposition 5.1,

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

and the assertion follows immediately from Lemma 4.6. \square

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 $p_{i,j} = \mathbb{P}_{\mu_i} \left[ \xi_1 = j \right]$ and $t_i = \mathbb{E}_{\mu_i} \left[ \alpha_1 \right]$, where $\mu_i$ is the invariant family of distributions on the milestone $M_i$. In words, $t_i$ is the average time a trajectory initiated on the $i$th milestone randomly from $\mu_i$ takes before it hits another milestone, and $p_{i,j}$ is the probability that the next milestone hit by this trajectory (other than the $i$th milestone) is the $j$th 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 $p_{i,j} = \mathbb{P}_{\mu_i} \left[ \xi_1 = j \right]$ and $t_i = \mathbb{E}_{\mu_i} \left[ \alpha_1 \right]$ 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 $\mu_i$ 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 $\mu_i$, which we do not know \textit{a priori} (recall that in this section we do not assume that the milestones are optimal, i.e. the density of $\mu_i$ is not given by (28) in general). In the original milestoning procedure, it was assumed that each $\mu_i$ 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 $p_{i,j} = \mathbb{P}_{\mu_i} \left[ \xi_1 = j \right]$ and $t_i = \mathbb{E}_{\mu_i} \left[ \alpha_1 \right]$ 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 $\mu_i$ beforehand. In fact, as we will see below, this procedure permits to sample $\mu_i$, 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:

\textbf{Lemma 6.1.} We have

$$p_{i,j} = \mathbb{P}_{\mu_i} \left[ \xi_1 = j \right] = \mathbb{P}_\mu \left[ \xi_1 = j \mid \xi_0 = i \right]$$

$$= \lim_{n \to \infty} \frac{\sum_{k=1}^{n} \delta_{i,\xi_{k-1}} \delta_{j,\xi_k}}{\sum_{k=1}^{n} \delta_{i,\xi_{k-1}}} = \lim_{T \to \infty} \frac{N_{ij}^T}{N_i} = \lim_{T \to \infty} \frac{\sum_j N_{ij}^T}{\sum_j N_{ij}^T},$$

$$t_i = \mathbb{E}_{\mu_i} \left[ \alpha_1 \right] = \mathbb{E}_{\mu} \left[ \alpha_1 \mid \xi_0 = i \right]$$

$$= \lim_{n \to \infty} \frac{\sum_{k=1}^{n} \alpha_k \delta_{i,\xi_{k-1}}}{\sum_{k=1}^{n} \delta_{i,\xi_{k-1}}} = \lim_{T \to \infty} \frac{R_i^T}{N_i} = \lim_{T \to \infty} \frac{\sum_j R_{ij}^T}{\sum_j R_{ij}^T}.$$
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 milestoneing 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 \( f : \mathbb{R}^n \rightarrow [0, 1] \), 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 milestoneing, but with \( f \) not necessarily equal to \( q^- \)). Specifically, given \( 0 < z_0 < z_1 \cdots < z_N < 1 \), define

\[
M_i = \{ x : f(x) = z_i \}, \quad i \in I.
\]

Then set

\[
\Omega_i = \{ x : z_{i-1} \leq f(x) \leq z_{i+1} \}, \quad i \in I
\]

where we interpret \( z_{-1} = -\infty \) and \( z_{N+1} = +\infty \). Thus, for \( i = 1, \ldots, N - 1, \Omega_i \) is the region comprised between \( M_{i-1} \) and \( M_{i+1} \), \( \Omega_0 \) is the region on the side of \( M_1 \) that contains \( M_0 \) and \( \Omega_N \) 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 \( \Omega_i \), 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 \( M(i) = \{ M_{i-1}, M_i, M_{i+1} \} \) for \( i = 1, \ldots, N - 1, M(0) = \{ M_0, M_1 \} \) and \( M(N) = \{ M_{N-1}, M_N \} \), and use the solutions of the SDE in \( \Omega_i \) with reflective boundary condition to construct the sequence \( \{ (\xi_n^{(i)}, \tau_n^{(i)}) : n \in \mathbb{N}_0 \} \) associated with \( M(i) \), then Lemma 6.1 still holds if we replace \( \xi_n \) by \( \xi_n^{(i)} \) and \( \alpha_n \) 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

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

\[
t_i = \lim_{n \to \infty} \frac{\sum_{k=1}^n \alpha_k^{(i)} \delta_{i,\xi_k^{(i)}}}{\sum_{k=1}^n \delta_{i,\xi_k^{(i)}}}
\]

**Proof.** The proposition is a consequence of the fact that the law of the solutions to the SDE in \( \Omega_i \) with reflective boundary condition on \( \partial \Omega_i \) is identical to that of the solutions to the original SDE pruned to \( \Omega_i \). \( \square \)

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 \( \Omega_i, i \in I \). 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 \( M(i) = \{ M_{i-1}, M_i, M_{i+1} \} \) for \( i = 1, \ldots, N - 1, M(0) = \{ M_0, M_1 \} \) and \( M(N) = \{ M_{N-1}, M_N \} \), and use the solutions of the SDE in \( \Omega_i \) with reflective boundary condition to construct the sequence
The key approximation in the string method is that the isocommittor surfaces are approximated locally in the reaction tube by the level sets of a smoothing kernel introduced to guarantee that the approximation 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 is parametrized by the smallest value. It is easily seen that 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 \( \gamma \) by

\[ q^{-}(x) \approx f(x) = \int_{\mathbb{R}^{n}} K_{\delta}(x-y)Q(s_{\gamma}(y))dy \]

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

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

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 be 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 $\Omega_j$ denote the open region enclosed by $M_i$ and $M_j$, i.e.,

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

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

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

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

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

Observe that

$$n(x) = \frac{\nabla q^-(x)}{\| \nabla q^-(x) \|}$$
is the unit normal vector of the surface $M_i$. Then for any $0 \leq i < j \leq N$, by the divergence theorem,

$$Z_j - Z_i = \int_{M_j} \langle \rho a \nabla q^- , n \rangle \, d\sigma_{M_j} - \int_{M_i} \langle \rho a \nabla q^- , n \rangle \, d\sigma_{M_i} = - \int_{\Omega_j^i} \nabla \cdot (\rho a \nabla q^-) \, dx$$

$$= - \int_{\Omega_j^i} \nabla \cdot (q^- J) \, dx = \int_{M_j} q^- J \cdot n \, d\sigma_{M_j} - \int_{M_i} q^- J \cdot n \, d\sigma_{M_i}$$

$$= z_j \int_{M_j} J \cdot n \, d\sigma_{M_j} - z_i \int_{M_i} J \cdot n \, d\sigma_{M_i}$$

$$= z_j \int_{\Omega_{N+1}^j} \nabla \cdot J \, dx - z_i \int_{\Omega_{N+1}^i} \nabla \cdot J \, dx = 0,$$

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

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

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

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

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

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

$$\begin{aligned}
(39) \quad \int_{M_i} u_i \rho_i \, d\sigma_{M_i} &= q_{i,i-1} \int_{M_{i-1}} u_i \rho_{i-1} \, d\sigma_{M_{i-1}} + q_{i,i+1} \int_{M_{i+1}} u_i \rho_{i+1} \, d\sigma_{M_{i+1}}
\end{aligned}$$

for each $0 \leq i \leq N$ 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 $\Omega_{i+1}^{i-1}$. By Lemma A.1 and definition of $\rho_i$ in (28), (39) reduces to

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

Let us insert the values of $q_{i,i-1}$ and $q_{i,i+1}$ given by (37) into the last equation, and multiply both sides by $z_{i-1} - z_{i+1}$, then we are left with checking

$$\begin{aligned}
(z_{i-1} - z_{i+1}) \int_{M_i} \langle u_i \rho a \nabla q^-, n \rangle \, d\sigma_{M_i} &= (z_{i-1} - z_{i+1}) \int_{M_{i-1}} \langle u_i \rho a \nabla q^- , n \rangle \, d\sigma_{M_{i-1}} \\
&\quad + (z_{i-1} - z_{i}) \int_{M_{i+1}} \langle u_i \rho a \nabla q^- , n \rangle \, d\sigma_{M_{i+1}}.
\end{aligned}$$
Moving the right hand side to the left, and regrouping these terms properly into three surface integrals, we obtain
\[ z_{i+1} \left( \int_{M_{i-1}} \langle u_i \rho a \nabla q^- \cdot n \rangle d\sigma_{M_{i-1}} - \int_{M_i} \langle u_i \rho a \nabla q^- \cdot n \rangle d\sigma_{M_i} \right) \\
+ z_{i-1} \left( \int_{M_i} \langle u_i \rho a \nabla q^- \cdot n \rangle d\sigma_{M_i} - \int_{M_{i+1}} \langle u_i \rho a \nabla q^- \cdot n \rangle d\sigma_{M_{i+1}} \right) \\
- z_i \left( \int_{M_{i-1}} \langle u_i \rho a \nabla q^- \cdot n \rangle d\sigma_{M_{i-1}} - \int_{M_{i+1}} \langle u_i \rho a \nabla q^- \cdot n \rangle d\sigma_{M_{i+1}} \right) = 0. \]

Applying divergence theorem to each surface integral yields
\[ z_{i+1} \int_{\Omega_i} \nabla \cdot \left( u_i \rho a \nabla q^- \right) dx + z_{i-1} \int_{\Omega_{i+1}} \nabla \cdot \left( u_i \rho a \nabla q^- \right) dx - z_i \int_{\Omega_i} \nabla \cdot \left( u_i \rho a \nabla q^- \right) dx = 0. \]

(40)

Now we calculate
\[ \nabla \cdot \left( u_i \rho a \nabla q^- \right) = \langle \nabla u_i, \rho a \nabla q^- \rangle + u_i \nabla \cdot \left( \rho a \nabla q^- \right), \]
\[ \nabla \cdot \left( q^- \rho a \nabla u_i \right) = \langle \nabla q^- \cdot \rho a \nabla u_i \rangle + q^- \nabla \cdot \left( \rho a \nabla u_i \right). \]

Since \( a \) is symmetric, we have
\[ \nabla \cdot \left( u_i \rho a \nabla q^- \right) = \nabla \cdot \left( q^- \rho a \nabla u_i \right) - q^- \nabla \cdot \left( \rho a \nabla \nabla u_i \right) + u_i \nabla \cdot \left( \rho a \nabla q^- \right). \]

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
\[ \mathcal{L} = \frac{1}{\rho^2} \left[ \nabla \cdot (\rho a \nabla J) + J \cdot \nabla \right], \]

a calculation similar to the derivation of (37) leads to
\[ \nabla \cdot (\rho a \nabla u_i) = -\nabla \cdot (u_i J). \]

(41)

Combining the above identities, we obtain
\[ \nabla \cdot \left( u_i \rho a \nabla q^- \right) = \nabla \cdot \left( q^- \rho a \nabla u_i + q^- u_i J \right). \]

(42)

Substituting this into (40) and applying the divergence theorem and \( q^- = z_i \) on \( M_i \), we get
\[ z_{i+1} \left( z_{i-1} \int_{M_{i-1}} \langle \rho a \nabla u_i + u_i J, n \rangle d\sigma_{M_{i-1}} - z_i \int_{M_i} \langle \rho a \nabla u_i + u_i J, n \rangle d\sigma_{M_i} \right) \\
+ z_{i-1} \left( z_i \int_{M_i} \langle \rho a \nabla u_i + u_i J, n \rangle d\sigma_{M_i} - z_{i+1} \int_{M_{i+1}} \langle \rho a \nabla u_i + u_i J, n \rangle d\sigma_{M_{i+1}} \right) \\
- z_i \left( z_{i-1} \int_{M_{i-1}} \langle \rho a \nabla u_i + u_i J, n \rangle d\sigma_{M_{i-1}} - z_{i+1} \int_{M_{i+1}} \langle \rho a \nabla u_i + u_i J, n \rangle d\sigma_{M_{i+1}} \right) = 0. \]
Regrouping these terms again into three new surface integrals and applying the divergence theorem again, we may convert the last equation into

\[
\begin{align*}
& z_{i-1} z_{i+1} \int_{\Omega_{i+1}} \nabla \cdot (\rho a \nabla u_i + u_i J) \, dx - z_i z_{i+1} \int_{\Omega_{i+1}} \nabla \cdot (\rho a \nabla u_i + u_i J) \, dx \\
& - z_{i-1} z_i \int_{\Omega_{i-1}} \nabla \cdot (\rho a \nabla u_i + u_i J) \, dx = 0.
\end{align*}
\]

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

**Appendix B. Proof of Lemma 4.3**

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

\[
P_{\mu} [\xi_k = i_k, \, 1 \leq k \leq n \mid \xi_0 = i_0] = P_{\mu_{i_0}} [Y_k \in M_{i_k}, \, 1 \leq k \leq n + 1]
\]

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

\[
\begin{align*}
&P_{\mu} [\xi_k = i_k, \, 1 \leq k \leq n + 1 \mid \xi_0 = i_0] = P_{\mu_{i_0}} [Y_k \in M_{i_k}, \, 1 \leq k \leq n + 1] \\
= &E_{\mu_{i_0}} 1_{\{Y_1 \in M_{i_1}\}} P_{\mu} [Y_{k-1} \in M_{i_{k-1}}, \, 2 \leq k \leq n + 1 \mid \mathcal{F}_{\tau_1}] \\
= &E_{\mu_{i_0}} [1_{M_{i_1}} (Y_1) P_{X(\tau_1)} [Y_{k-1} \in M_{i_{k-1}}, \, 2 \leq k \leq n + 1]],
\end{align*}
\]

where the second line follows from time homogeneity. Note that the distribution of \(Y_1 = X(\tau_1)\) relative to the probability law \(P_{\mu_{i_0}}\) 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

\[
= \int_M 1_{M_{i_1}} (x) P_x [Y_k \in M_{i_{k+1}}, \, 1 \leq k \leq n] \, (P^* \mu_{i_0})(dx)
\]

\[
= \sum_{j \in I} p_{i_0, j} \int_{M_{i_1}} P_x [\xi_k = i_{k+1}, \, 1 \leq k \leq n] \, \mu_j (dx)
\]

\[
= p_{i_0, i_1} \sum_{j \in I} P_{\mu} [\xi_k = i_{k+1}, \, 1 \leq k \leq n]
\]

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

\[
E_{\mu} [\alpha_n 1_{\{\xi_k = i_k, \, 1 \leq k \leq n\} \mid \xi_0 = i_0}] = E_{\mu_{i_0}} [\alpha_n \prod_{1 \leq k \leq n} 1_{M_{i_k}} (Y_k)]
\]

\[
= E_{\mu_{i_0}} [1_{M_{i_1}} (Y_1) E_{X(\tau_1)} [\alpha_{n-1} \prod_{2 \leq k \leq n} 1_{M_{i_k}} (Y_{k-1})]]
\]

\[
= p_{i_0, i_1} E_{\mu_{i_1}} [\alpha_{n-1} \prod_{1 \leq k \leq n-1} 1_{M_{i_{k+1}} (Y_k)}]
\]

where in the third equality we have used the assumption that under the law \(P_{\mu_{i_0}}\), 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 \(\mu_{i_1}\). On the other hand, by (i), we have

\[
P_{\mu} [\xi_k = i_k, \, 1 \leq k \leq n \mid \xi_0 = i_0] = p_{i_0, i_1} E_{\mu} [\xi_k = i_{k+1}, \, 1 \leq k \leq n-1 \mid \xi_0 = i_1].
\]
So we obtain
\[ \mathbb{E}_\mu[\alpha_n | \xi_k = i_k, 0 \leq k \leq n] = \mathbb{E}_\mu[\alpha_{n-1} | \xi_k = i_{k+1}, 0 \leq k \leq n - 1]. \]
This completes the inductive step. \(\square\)

**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 (43). Note that the event \( \{ D \geq n \} \) belongs to the \( \sigma \)-field generated by \( \xi_0, \ldots, \xi_{n-1} \). Then by Fubini’s Theorem and Lemma 4.5, we obtain

\[
(43) \quad h_i = \mathbb{E}_\mu[r_D] = \mathbb{E}_\mu\left[\sum_{n=1}^{D} \alpha_n \right] = \mathbb{E}_\mu\left[\sum_{n=1}^{\infty} \alpha_n 1_{\{ D \geq n \}} \right] = \sum_{n=1}^{\infty} \mathbb{E}_\mu[\alpha_n 1_{\{ D \geq n \}}] = \sum_{n=1}^{\infty} \mathbb{E}_\mu[\alpha_n 1_{\{ D \geq n \}} 1_{\xi_{n-1}, \xi_n}].
\]

Let us assume \( i \neq N \). Then \( D \geq 1 \) and the last-written sum can be split into

\[
\mathbb{E}_\mu[t_{i,\xi_1}] + \sum_{n=2}^{\infty} \mathbb{E}_\mu[1_{\{ D \geq n \}} t_{\xi_{n-1}, \xi_n}],
\]
where the first term amounts to \( \sum_{j \in I} t_{i,j} p_{i,j} \), while the second summation term is equal to

\[
\sum_{n=2}^{\infty} \sum_{j \in I} \mathbb{E}_\mu[1_{\{ D \geq n \}} t_{\xi_{n-1}, \xi_n} | \xi_0 = i, \xi_1 = j] \mathbb{P}_\mu[\xi_1 = j | \xi_0 = i] = \sum_{j \in I} p_{i,j} \sum_{n=2}^{\infty} \mathbb{E}_\mu[1_{\{ D \geq n \}} t_{\xi_{n-1}, \xi_n} | \xi_0 = i, \xi_1 = j].
\]

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

\[
\sum_{j \in I} p_{i,j} \sum_{n=2}^{\infty} \mathbb{E}_\mu[1_{\{ D \geq n \}} t_{\xi_{n-1}, \xi_n} | \xi_0 = i, \xi_1 = j] = \sum_{j \in I} p_{i,j} \mathbb{E}_\mu[t_{i,j}].
\]

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

\[
\sum_{j \in I} p_{i,j} \sum_{n=2}^{\infty} \mathbb{E}_\mu[1_{\{ D \geq n \}} t_{\xi_{n-1}, \xi_n} | \xi_0 = i, \xi_1 = j] = \sum_{j \in I} p_{i,j} \mathbb{E}_\mu[t_{i,j}].
\]

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 \( i \neq N \), we have

\[
y_i = \sum_{j \in I} p_{i,j} t_{i,j} + \sum_{j \in I} p_{i,j} y_j = \sum_{j \in I} p_{i,j} t_{i,j} + \sum_{j \neq N} p_{i,j} y_j.
\]
Using this identity to replace $y_j$ on the right hand side and by Lemma 4.5 we deduce
\[
y_i = \sum_{j \in I} p_{i,j} t_{i,j} + \sum_{j \neq n} p_{i,j} \left( \sum_{k \in I} p_{j,k} t_{j,k} + \sum_{k \neq n} p_{j,k} y_k \right)
\]
\[
= \sum_{j \in I} \mathbb{P}_\mu [\xi_1 = j \mid \xi_0 = i] \mathbb{E}_\mu [\alpha_1 \mid \xi_0 = i, \xi_1 = j] + \sum_{j \neq n} \sum_{k \in I} \mathbb{P}_\mu [\xi_1 = j, \xi_2 = k \mid \xi_0 = i] \mathbb{E}_\mu [\alpha_2 \mid \xi_0 = i, \xi_1 = j, \xi_2 = k] + \sum_{j \neq n} \sum_{k \neq n} p_{i,j} p_{j,k} y_k
\]
\[
= \mathbb{E}_{\mu_i} [\alpha_1] + \mathbb{E}_{\mu_i} [\alpha_2 1_{D \geq 2}] + \sum_{j \neq n} \sum_{k \neq n} p_{i,j} p_{j,k} y_k.
\]
By repeated substitution in the last term, we obtain after $n$ steps
\[
y_i = \sum_{m=1}^{n} \mathbb{E}_{\mu_i} [\alpha_m 1_{D \geq m}] + \text{Nonnegative Remainder} \geq \sum_{m=1}^{n} \mathbb{E}_{\mu_i} [\alpha_m 1_{D \geq m}].
\]
Sending $n \to \infty$, by (33), we find
\[
y_i \geq \sum_{m=1}^{\infty} \mathbb{E}_{\mu_i} [\alpha_m 1_{D \geq m}] = h_i,
\]
which completes the proof. \qed

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 $E \subset M_j$,
\[
((\mathcal{P}')^* \mu_i)(E) = p'_{i,j} \mu_j(E), \quad \text{with} \quad p'_{i,j} = ((\mathcal{P}')^* \mu_i)(M_j).
\]
Fix $i$ and $j$ in $I'$. By definition, the left hand member is
\[
\mathbb{P}_{\mu_i} \left[ X \left( H_{M' \sim M'_j}^+ \right) \in E \right].
\]
Put $\eta = X \left( H_{M' \sim M'_j}^+ \right)$ and $J' = (I \setminus I') \cup \{i\}$. Using the strong Markov property of $X$ and the assumption that
\[
\mathcal{P}^* \mu_k = \sum_{\ell \in I} p_{k,\ell} \mu_\ell, \quad \text{with} \quad p_{k,\ell} = (\mathcal{P}^* \mu_k)(M_\ell),
\]
we obtain that for $k \in J'$,
\[
\mathbb{P}_{\mu_k} [\eta \in E] = \mathbb{E}_{\mu_k} [\mathbb{P}_{Y_1} [\eta \in E]] = \sum_{\ell \in I} p_{k,\ell} \mathbb{P}_{\mu_k} [\eta \in E]
\]
\[
= \sum_{\ell \in I' \setminus \{i\}} p_{k,\ell} \mathbb{P}_{\mu_k} [\eta \in E] + \sum_{\ell \in J'} p_{k,\ell} \mathbb{P}_{\mu_k} [\eta \in E]
\]
\[
= p_{k,j} \mu_j(E) + \sum_{\ell \in J'} p_{k,\ell} \mathbb{P}_{\mu_k} [\eta \in E].
\]
In the last step, we use the fact that under the law \( P_{\mu_k} \) for \( \ell \in I' \setminus \{ i \} \), we have \( H_{M', M_i}^+ = 0 \) and \( \eta = X(0) \) almost surely. Using this identity to replace the term \( P_{\mu_k}[\eta \in E] \) on the right hand side, we see that

\[
P_{\mu_k}[\eta \in E] = p_{k, j} \mathbb{P}_{\mu_j}(E) + \sum_{\ell \in J'} p_{k, \ell} p_{\ell, j} \mathbb{P}_{\mu_j}(E) + \sum_{\ell, m \in J'} p_{k, \ell} p_{\ell, m} P_{\mu_m}[\eta \in E] = P_{\mu_k}[\xi_1 = j] \mathbb{P}_{\mu_j}(E) + \sum_{\ell \in J'} p_{k, \ell} p_{\ell, m} P_{\mu_m}[\eta \in E].
\]

Continuing in the obvious way, we find

\[
P_{\mu_k}[\eta \in E] = \mu_j(E) \sum_{n=1}^{m} P_{\mu_k}[\xi_\ell \in J', 1 \leq \ell < n, \, \xi_n = j] + \text{Nonnegative Remainder}.
\]

Sending \( m \to \infty \) yields

\[
P_{\mu_k}[\eta \in E] \geq \mu_j(E) \sum_{n=1}^{\infty} P_{\mu_k}[\xi_\ell \in J', 1 \leq \ell < n, \, \xi_n = j].
\]

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

\[
P_{\mu_k}[H_{M', M_i}^+ = \tau_n \text{ for some } n \geq 0, \eta = X(H_{M', M_i}^+ \in M_j)].
\]

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

\[
P_{\mu_k}[H_{M', M_i}^+ = \tau_n \text{ for some } n \geq 0] = 1.
\]

Thus we obtain

\[
P_{\mu_k}[\eta \in E] \geq \mu_j(E) P_{\mu_k}[\eta \in M_j].
\]

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

\[
P_{\mu_k}[\eta \in M_j] = P_{\mu_k}[\eta \in E] + P_{\mu_k}[\eta \in M_j \setminus E] \geq \mu_j(E) P_{\mu_k}[\eta \in M_j] + \mu_j(M_j \setminus E) P_{\mu_k}[\eta \in M_j] = P_{\mu_k}[\eta \in M_j].
\]

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

\[
P_{\mu_k}[\eta \in E] = \mu_j(E) P_{\mu_k}[\eta \in M_j].
\]

In particular, for \( k = i \), this is exactly what was to be shown. The proof is complete. \( \square \)

References


**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