8-2-2016

Computing the optimal path in stochastic dynamical systems

Martha Bauver

Eric Forgoston
Montclair State University

Lora Billings
Montclair State University, billingsl@montclair.edu

Published Citation

Bauver, M., Forgoston, E., & Billings, L. (2016). Computing the optimal path in stochastic dynamical systems. Chaos, 26(8), 083101. doi:10.1063/1.4958926

Computing the optimal path in stochastic dynamical systems

Martha Bauver, Eric Forgoston, and Lora Billings

Citation: Chaos **26**, 083101 (2016); doi: 10.1063/1.4958926
View online: http://dx.doi.org/10.1063/1.4958926
View Table of Contents: http://scitation.aip.org/content/aip/journal/chaos/26/8?ver=pdfcov
Published by the AIP Publishing

Computing the optimal path in stochastic dynamical systems

Martha Bauver, Eric Forgoston,[a] and Lora Billings
Department of Mathematical Sciences, Montclair State University, 1 Normal Avenue, Montclair, New Jersey 07043, USA

(Received 7 January 2016; accepted 5 July 2016; published online 2 August 2016)

In stochastic systems, one is often interested in finding the optimal path that maximizes the probability of escape from a metastable state or of switching between metastable states. Even for simple systems, it may be impossible to find an analytic form of the optimal path, and in high-dimensional systems, this is almost always the case. In this article, we formulate a constructive methodology that is used to compute the optimal path numerically. The method utilizes finite-time Lyapunov exponents, statistical selection criteria, and a Newton-based iterative minimizing scheme. The method is applied to four examples. The first example is a two-dimensional system that describes a single population with internal noise. This model has an analytical solution for the optimal path. The numerical solution found using our computational method agrees well with the analytical result. The second example is a more complicated four-dimensional system where our numerical method must be used to find the optimal path. The third example, although a seemingly simple two-dimensional system, demonstrates the success of our method in finding the optimal path where other numerical methods are known to fail. In the fourth example, the optimal path lies in six-dimensional space and demonstrates the power of our method in computing paths in higher-dimensional spaces. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4958926]

Increasingly, stochastic dynamical systems are being used to model a wide variety of physical and biological phenomena. In these types of systems, one often sees rare transition events that are induced by noise which may be internal or external to the system. These noise-induced rare events may be associated with a desirable outcome, such as the extinction of an infectious disease outbreak, or an undesirable outcome, such as the sudden clustering of cancerous cells. One important feature of interest when studying noise-induced transitions is the optimal transition pathway of escape from a metastable state or the optimal transition pathway from one metastable state to another metastable state. Although there are many paths, the optimal path is the path that is most likely to occur. For most systems, it is impossible to analytically determine the optimal path to escape or switch. Therefore, the optimal path must be determined numerically. This article describes a novel methodology for computing the optimal path in higher-dimensional systems, where analytical results are not available. The methodology uses finite-time Lyapunov exponents (FTLE) to provide an initial guess for use in an Iterative Action Minimizing Method (IAMM) to compute the optimal path in general stochastic systems subjected to internal or external noise. Here, the methodology is demonstrated using the Susceptible-Infectious-Susceptible (SIS) and Susceptible-Exposed-Infectious-Recovered (SEIR) infectious disease models and an Allee population model. Two versions of the SIS model include a low-dimensional single population example as well as a careful examination in four dimensions using two populations. The SEIR model demonstrates the extension of the method to six dimensions. The Allee example represents a type of system where other numerical methodologies are prone to failure. Our methodology, although demonstrated with epidemic and population models, may be used to compute the optimal escape path from a metastable state or the optimal switching path between metastable states for general stochastic dynamical systems.

I. INTRODUCTION

It is well-known that noise can have a significant effect on deterministic dynamical systems. As an example, given an initial state starting in a basin of attraction, noise can cause the initial state to cross the basin boundary and move into another, distinct basin of attraction.[1–5] Many researchers have investigated how noise affects physical and biological phenomena at a wide variety of levels. In these systems, noise can induce spontaneous switching between coexisting stable states. Two examples from physical systems are switching between the magnetization states in magnets,[6] and voltage and current states in Josephson junctions.[7] In biology, noise can play a role in sub-cellular processes, tissue dynamics, large-scale population dynamics,[8] and genetic switching.[9] Two examples of spontaneous switching between stable states in biological systems are extinction of an epidemic and extinction of a species.[10]

Stochasticity manifests itself as either external or internal noise. External noise comes from a source outside the system being considered (e.g., population growth under the influence of climatic effects, or a random signal fed into a transmission line), and often is modeled by replacing an external parameter with a random process. Internal noise is inherent in the system itself and is caused by the random

[a]Author to whom correspondence should be addressed. Electronic mail: eric.forgoston@montclair.edu

interactions of discrete particles (e.g., individuals in a population, or chemical reactions).[11,12] Both types of noise can lead to a rare switching event between metastable states or a rare escape event from a metastable state. There are many possible escape/switching paths, but there is a path along which switching or escape is most likely to occur. We call this most likely path of escape or switching the optimal path. It is of great importance in a variety of applied problems to determine this optimal path, since knowledge of the path then enables the determination of the mean time to escape from a metastable state or to switch from one metastable state to another metastable state.

Mathematically, the effect of external noise is often described using a Langevin equation or the associated Fokker–Planck equation (though the dynamics of external noise may sometimes be described by a master equation[13]). Feynman noticed that each noise realization corresponds to a particular trajectory of the system, and therefore, the probability density of realizations of trajectories is determined by the probability density of noise realizations.[14] This idea can be used to formulate a variational problem that ultimately leads to a Hamiltonian system. One can solve the Hamiltonian, either analytically or numerically, for the most probable, or optimal, path of escape or switching.[11,12] The effect of internal noise due to the random interactions of individuals within the system is described mathematically using a master equation.

The master equation is a large, or even infinite, set of differential equations, and in general, it cannot be solved analytically. Therefore, one must resort to using a Wentzel–Kramers–Brillouin (WKB), or eikonal, approximation. The WKB method leads to the development of a Hamiltonian system, which can be solved for the optimal path.[15–21]

In summary, for either type of noise, the optimal path is found by transforming the original stochastic problem into a new deterministic system described by a Hamiltonian $\mathcal{H}(\mathbf{x}, \mathbf{p}) = 0$. The dimensions of the Hamiltonian are twice the dimensions of the original system due to the conjugate momenta variables $\mathbf{p}$. The method amounts to finding a zero-energy trajectory of an effective mechanical system, and at least one of the solutions to the zero-energy Hamiltonian is the optimal path. There may be other escape/switching/extinction paths, but the optimal path is the path that maximizes the probability of escape/switching/extinction. Even for simple problems, the optimal path may not have an analytical solution. The numerical computation of the optimal path trajectory has been achieved in the past using a shooting or other iterative methods.[22,23] However, since the procedure is very sensitive to boundary conditions, it is difficult to implement when analyzing paths far away from bifurcation points.[24,25] In Refs. 24 and 25, these numerical issues were overcome by computing finite-time Lyapunov exponents (FTLE). The method worked very well for low-dimensional problems, but in high dimensional problems, the FTLE results define a region on and around the optimal path, but fail to describe the optimal path itself in the necessary detail.

In this article, we describe a novel methodology that combines the use of two numerical methods to compute the optimal path trajectory for high-dimensional models. The method begins by computing finite-time Lyapunov exponents. FTLE computations provide a measure of stretching of initially close particles advected over a specified finite amount of time. Ridges of high FTLE values are of great interest, since it has been shown that a maximal FTLE ridge is equivalent to the optimal path.[24,25] Therefore, the FTLE provides a means to identify points in close proximity to the optimal path. We then use the FTLE results as an initial guess for the second numerical scheme.

This second scheme is an Iterative Action Minimizing Method (IAMM),[26] which applies a Newton's method process to the initial guess provided by the FTLE result. The IAMM is a minimum action method[27,28] and provides a method to converge rapidly to the optimal path. While the method is rather straightforward, there are difficulties in computing the trajectory of an unstable path in higher dimensions which lead us to explore other factors that may be influencing the computations. It is worth noting that in the absence of a good initial condition, the IAMM may not converge to the optimal path.

We illustrate our methodology using standard Susceptible-Infectious-Susceptible (SIS) and Susceptible-Exposed-Infectious-Recovered (SEIR) epidemic models, and an Allee model, all with intrinsic noise. All processes and methods are demonstrated using a low-dimensional single population model, as well as higher-dimensional two and three population models. Then, the method is refined to demonstrate its use for the case where the optimal path begins and ends on the deterministic path. Section II contains the mathematical theory, including the master equation formulation (Sec. II A), the finite-time Lyapunov exponent computations (Sec. II B), and the iterative action minimizing method computations (Sec. II C). The one-population (1D) and two-population (2D) SIS infectious disease models are described in Section III, along with the development of the Hamiltonians for both of these models. Section IV details the FTLE and IAMM computations used to obtain the optimal path results for the 1D model, with results verified by the analytical solution. In Section V, we describe the process to obtain the four-dimensional optimal path for the 2D SIS model. In both Secs. IV and V, we discuss the requirements of the methodologies for obtaining successful results. In Section VI, we describe the Allee model and, using a simplified system, present a scenario for the use of our methods to describe the optimal path of higher-dimensional systems when the optimal path begins and ends at critical points along the deterministic path. Section VII demonstrates the power of our method for a higher-dimensional 3D SEIR model. Section VIII contains the discussion and conclusions.

II. GENERAL THEORY

A. Master equation formalism

The method we present in this article is general and may be used to compute optimal escape paths, switching paths, or extinction paths for a wide variety of stochastic dynamical systems. However, to make the method transparent, this article focuses on computing the optimal path to extinction for

two types of epidemic models and a population model. The simplest deterministic epidemic compartmental models contain two steady states of differing stability: an extinct state where no infectious individuals are present and an endemic state where the infection is maintained. The stability of these steady states is determined by the value of the basic reproductive number R_0, as shown in Fig. 1. The reproductive number can be thought of as the average number of new infectious individuals that one infectious individual generates over the course of the infectious period, in an entirely susceptible population. When $R_0 < 1$, the extinct state is stable while the endemic state is unstable; when $R_0 > 1$, the extinct state is unstable while the endemic state is stable. It is worth noting that since the model is deterministic, a population at the attracting endemic state can never go extinct.

To capture extinction events, we must consider a stochastic model with internal noise that represents the random interactions of individuals in the population. Therefore, a master equation is used to describe the effect of stochasticity. Let X be a state variable that represents the number of individuals in a single population. The following theory will be developed for a single population, but it is easily extended to multiple populations X_1, X_2,…X_n by the use of a state vector $\mathbf{X}$.

The probability density $\rho(X, t)$ describes the probability of finding X individuals at time t. Each possible population-changing event (birth, death, infection, migration, etc.) is defined by a transition rate $W_r(X)$, where r is a positive or negative integer that defines an incremental change from state X to state $X+r$. Then, the master equation that provides the time evolution of $\rho(X, t)$ for a single population is[11,12]

$$\frac{\partial \rho(X,t)}{\partial t} = \sum_r \left[W_r(X-r)\rho(X-r,t) - W_r(X)\rho(X,t)\right]. \quad (1)$$

In general, it is not possible to analytically solve the master equation given by Eq. (1). Let X be scaled by N, the typical population size in the metastable state. Using $x = X/N$, the transition rate $W_r(X) = W_r(Nx)$ can be represented as the following expansion in N:

$$W_r(Nx) = Nw_r(x) + u_r(x) + \mathcal{O}(1/N), \quad (2)$$

where x and the scaled transition rates w_r and u_r are $\mathcal{O}(1)$.

For $N \gg 1$, we approximate the scaled master equation using the Wentzel–Kramers–Brillouin (WKB) approximation. To account for the rare possibility of extinction, we look for the probability distribution in the form of the WKB ansatz

$$\rho = e^{-N\mathcal{S}(x,t)}, \quad (3)$$

where $\mathcal{S}(x,t)$ is a quantity known as the action.[15–17,29] The WKB ansatz given by Eq. (3) is substituted into the scaled master equation, which is stated in terms of $w_r(x - r/N)$ and $\mathcal{S}(x - r/N, t)$, where r/N is small. A Taylor series expansion of these functions of $x - r/N$ is performed, and one arrives at a Hamilton–Jacobi equation $\mathcal{H} = -\partial\mathcal{S}/\partial t$. At leading order, the Hamilton–Jacobi equation has the form $\mathcal{H}(x,p) = 0$, where $\mathcal{H}$, known as the effective Hamiltonian, is given as

$$\mathcal{H}(x,p) = \sum_r w_r(x)(e^{pr} - 1); \qquad p = \frac{d\mathcal{S}}{dx}, \quad (4)$$

and p is the conjugate momentum. The solutions to $\mathcal{H}(x,p) = 0$ are the zero-energy curves of the system. At least one solution is the optimal path where the action $\mathcal{S}$ is minimized; this solution corresponds to the path that maximizes the probability of extinction.

Hamilton's equations

$$\dot{x} = \partial\mathcal{H}(x,p)/\partial p, \quad \dot{p} = -\partial\mathcal{H}(x,p)/\partial x \quad (5)$$

describe the system's dynamics and are easily found from the Hamiltonian given by Eq. (4). The x dynamics along the $p = 0$ deterministic line are described by

$$\dot{x} = \left.\frac{\partial\mathcal{H}(x,p)}{\partial p}\right|_{p=0}, \quad (6)$$

which is the rescaled mean-field rate equation associated with the original deterministic problem. For the simple models described by Fig. 1, the deterministic steady states are nodes. It is easy to show that the WKB method has transformed these steady state nodes in the original 1D deterministic system into steady state saddle points in the 2D set of Hamilton's equations. This allows for escape from the endemic state and provides a path to extinction that did not exist in the original deterministic model. In the epidemic models we will discuss, the optimal path leaves the endemic state and arrives at a new extinct point where at least one momentum p value is nonzero.

The complexity of exploring multi-population stochastic models becomes apparent when considering the solution to the generalized set of Hamilton's equations

$$\begin{aligned} \dot{\mathbf{x}}_\mathbf{j} &= \frac{\partial H(x_1,\ldots x_n,p_1,\ldots p_n)}{\partial p_j}, \\ \dot{\mathbf{p}}_\mathbf{j} &= -\frac{\partial H(x_1,\ldots x_n,p_1,\ldots p_n)}{\partial x_j}, \quad j = 1,\ldots n, \end{aligned} \quad (7)$$

where n is the number of distinct population groups included in a model. After employing the WKB formalism, the domain of the system is $\mathbb{R}^{2n}$, and a system of $2n$ equations must be solved to find steady states. Furthermore, since the Hamiltonian is now a function of $2n$ variables, the likelihood that analytic solutions to the zero-energy Hamiltonian

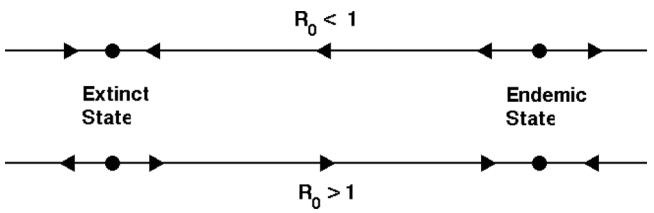

FIG. 1. The simplest deterministic compartmental models have two steady states. There is a bifurcation at $R_0 = 1$ that causes a change in stability. For $R_0 < 1$, the extinct state is attracting while the endemic state is repelling. For $R_0 > 1$, the extinct state is repelling while the endemic state is attracting.

equation can be found is greatly reduced. Therefore, we must consider alternate methodologies to compute the optimal path.

B. Finite-time Lyapunov exponents (FTLE)

The computation of finite-time Lyapunov exponents is often used to find coherent structures in fluid flows.[30–33] The FTLE provides a measure of how sensitively the system's future behavior depends on its current state. We have previously shown that the system displays maximal sensitivity near the optimal path trajectory,[24,25] which enables us to dynamically evolve toward the optimal path.

We consider a velocity field $\boldsymbol{v} : \mathbb{R}^{2n} \times I \to \mathbb{R}^{2n}$ given by Eq. (5) that is defined over a time interval $I = [t_i, t_f]$, and the system of equations

$$\dot{\mathbf{y}}(t; t_i, \mathbf{y}_0) = \boldsymbol{v}(\mathbf{y}(t; t_i, \mathbf{y}_0), t), \tag{8a}$$

$$\mathbf{y}(t_i; t_i, \mathbf{y}_0) = \mathbf{y}_0, \tag{8b}$$

where $\mathbf{y}, \mathbf{y}_0 \in \mathbb{R}^{2n}$, and $t \in I$. This dynamical system has quantities known as Lyapunov exponents that measure the growth rates of the linearized dynamics about the trajectory of the system. To find the finite-time Lyapunov exponents (FTLE), the Lyapunov exponents are computed on a restricted finite time interval.

To compute FTLE values, we choose a domain of interest and define it as an evenly spaced grid of $2n$ dimensional points $\mathbf{y} = (\mathbf{x}, \mathbf{p})$, with initial position $\mathbf{y}_0$ defined at the grid points. Then, using Hamilton's equations (Eq. (5)), all points are numerically integrated. The flow map ϕ determines the advection of the initial points as follows:[30–33]

$$\phi_{t_i}^{t_i+T} : \mathbf{y}_0 \to \phi_{t_i}^{t_i+T}(\mathbf{y}_0) = \mathbf{y}(t_i + T; t_i, \mathbf{y}_0). \tag{9}$$

Then, the FTLE can be defined as

$$\sigma(\mathbf{y}, t_i + T, T) = \frac{1}{|T|} \ln \sqrt{\lambda_{max}(\Delta)}, \tag{10}$$

where $\lambda_{\max}(\Delta)$ is the maximum eigenvalue of the right Cauchy–Green deformation tensor Δ, which is given as

$$\Delta(\mathbf{y}, t_i + T, T) = \left(\frac{d\phi_{t_i}^{t_i+T}(\mathbf{y}(t))}{d\mathbf{y}(t)} \right)^* \left(\frac{d\phi_{t_i}^{t_i+T}(\mathbf{y}(t))}{d\mathbf{y}(t)} \right), \tag{11}$$

with * denoting the adjoint.

For a given $\mathbf{y} \in \mathbb{R}^{2n}$ at initial time t_i, Eq. (10) gives the maximum finite-time Lyapunov exponent for some finite integration time T (forward or backward) and provides a measure of the sensitivity of a trajectory to small perturbations. The FTLE field given by $\sigma(\mathbf{y}, t_i, T)$ can be shown to exhibit ridges of local maxima in phase space. The ridges of the field indicate the location of attracting (backward time FTLE field) and repelling (forward time FTLE field) structures. In two-dimensional (2D) space, the ridge is a curve which locally maximizes the FTLE field, so that transverse to the ridge, one finds the FTLE to be a local maximum. What is remarkable is that the FTLE ridges correspond to the optimal path trajectories.[24,25] The basic idea is that since the optimal path is inherently unstable and observed only through many realizations of stochastic experiments, the FTLE shows that locally the path is also the most sensitive to initial data. Figure 2 shows a schematic that demonstrates why the optimal path corresponds to a maximal FTLE ridge. If one chooses an initial point on either side of the optimal path near the endemic state, the two trajectories will separate exponentially in time, since both extinct and endemic states are unstable saddle points.

For single population models, the optimal path is a curve in $\mathbb{R}^2$ with (x, p) coordinates. A contour map of σ values as a function of x and p is ideal for showing the maximal FTLE ridge that corresponds to the optimal path. However, for models involving multiple populations, the optimal path is a curve in a high-dimensional space which is not easily visualized. A methodology that can extract information about the optimal path from the FTLE field without relying on visual inspection is necessary.

Therefore, we use statistical tools to define a threshold cut-off value, so that points associated with the highest FTLE values can be identified. Additionally, points associated with deterministic structures, such as those occurring where $p = 0$, can be excluded from selection. In this way, points known to be in the vicinity of the optimal path are identified and selected for inclusion in an array of high

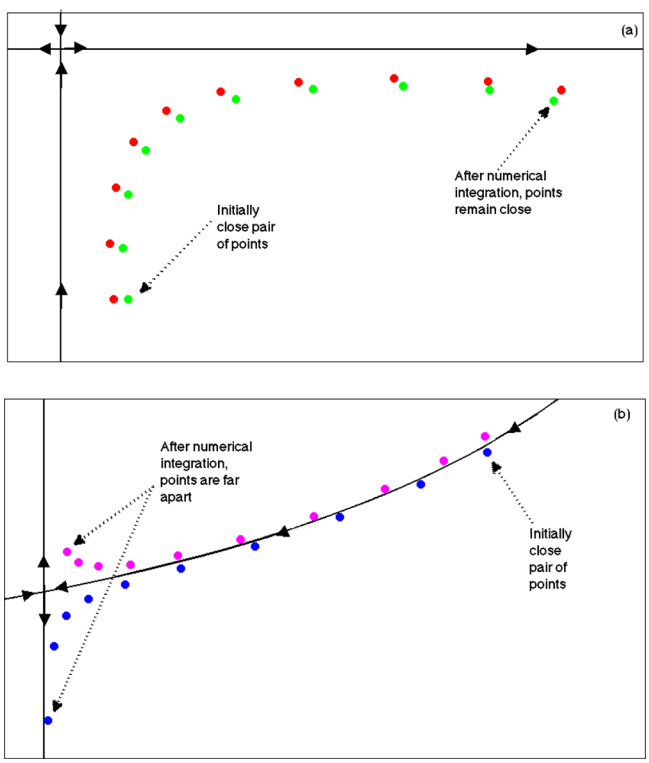

FIG. 2. (a) The numerical integration of two initially close points on one side of the manifold. After a finite time, the two points remain close to one another, and therefore, the corresponding FTLE value will be small. (b) The numerical integration of two initially close points on opposite sides of the manifold. After a finite time, the two points move a significant distance apart from one another, and therefore, the corresponding FTLE value will be large.

FTLE points. For low-dimensional models, this array may produce a nearly complete optimal path. For high-dimensional models, however, such a nice result is less likely. Their known proximity to the optimal path, however, makes the point array highly desirable as initial data for the following method.

C. Iterative action minimizing method (IAMM)

As described in Section II A, the WKB formalism results in a Hamiltonian with a zero-energy curve that is the optimal path connecting two steady state saddle points. The Iterative Action Minimizing Method[26] (IAMM) is a numerical scheme based on Newton's method that computes optimal transition pathways in systems of stochastic differential equations. In particular, given an initial guess of high FTLE-valued points produced by the method described in Sec. II B, the IAMM is a useful method for determining the optimal path.

We consider the general situation where a path connecting steady states C_a and C_b starts at C_a at $t=-\infty$ and ends at C_b at $t=+\infty$. Then, a time parameter t exists such that $-\infty<t<\infty$. For this method, we require a numerical approximation of the time needed to leave the region of C_a and arrive in the region of C_b. Therefore, we define a time T_ϵ such that $-\infty<-T_\epsilon<t<T_\epsilon<\infty$. Additionally, $C(-T_\epsilon)\approx C_a$ and $C(T_\epsilon)\approx C_b$. In other words, the solution stays very near the equilibrium C_a for $-\infty<t\le -T_\epsilon$, has a transition region from $-T_\epsilon<t<T_\epsilon$, and then stays near C_b for $T_\epsilon<t<+\infty$.

The interval $[-T_\epsilon, T_\epsilon]$ is discretized into N segments using a uniform step size $h=(2T_\epsilon)/N$. Alternatively, one may map the $[-T_\epsilon, T_\epsilon]$ interval onto the [0, 1] interval via the linear transformation $t=2T_\epsilon\tau - T_\epsilon$ and use a step size of $h=1/N$ to discretize the interval. It is worth noting that for some problems, such as the Allee system described in Sec. VI C, it may be necessary to use a non-uniform step size to resolve sharp transition regions.[26]

Given a non-uniform time step h_k, then one has the time series $t_{k+1}=t_k+h_k$. The derivative of the corresponding function value $\mathbf{q}_k$ is approximated using central finite differences by the operator δ_h given as

$$\frac{d}{dt}\mathbf{q}_k \approx \delta_h\mathbf{q}_k \equiv \frac{h_{k-1}^2\mathbf{q}_{k+1}+\left(h_k^2-h_{k-1}^2\right)\mathbf{q}_k-h_k^2\mathbf{q}_{k-1}}{h_{k-1}h_k^2+h_kh_{k-1}^2}, \quad k=0,\ldots,N. \tag{12}$$

Clearly, if a uniform step size is chosen, then Eq. (12) simplifies to the familiar form given as

$$\frac{d}{dt}\mathbf{q}_k \approx \delta_h\mathbf{q}_k \equiv \frac{\mathbf{q}_{k+1}-\mathbf{q}_{k-1}}{2h}, \quad k=0,\ldots,N. \tag{13}$$

Thus, one can develop the system of nonlinear algebraic equations

$$\delta_h\mathbf{x}_k-\frac{\partial H(\mathbf{x}_k,\mathbf{p}_k)}{\partial\mathbf{p}}=0, \quad \delta_h\mathbf{p}_k+\frac{\partial H(\mathbf{x}_k,\mathbf{p}_k)}{\partial\mathbf{x}}=0, \quad k=0,\ldots,N, \tag{14}$$

which is solved using a general Newton's method. Note that Eq. (14) computes the difference between the central finite difference approximation of each partial derivative (see Eq. (7)) with the appropriate Hamilton's equation.

We let $\mathbf{q}_j(\mathbf{x},\mathbf{p})=\{\mathbf{x}_{1,j}\ldots\mathbf{x}_{N,j},\mathbf{p}_{1,j}\ldots\mathbf{p}_{N,j}\}^T$ be an extended vector of $2nN$ components that contains the j^{th} Newton iterate, where n is the number of populations. When $j=0$, $\mathbf{q}_0(\mathbf{x},\mathbf{p})$ provides the initial "guess" as to the location of the path that connects C_a and C_b. In our work, $\mathbf{q}_0$ comes from the results of FTLE computations. Given the j^{th} Newton iterate $\mathbf{q}_j$, the new $\mathbf{q}_{j+1}$ iterate is found by solving the linear system

$$\mathbf{q}_{j+1}=\mathbf{q}_j-\frac{\mathbf{F}(\mathbf{q}_j)}{\mathbf{J}(\mathbf{q}_j)}, \tag{15}$$

where $\mathbf{F}$ is the function defined by Eq. (14) acting on $\mathbf{q}_j$, and $\mathbf{J}$ is the Jacobian. Equation (15) is solved using LU decomposition with partial pivoting.

The IAMM, like any scheme involving Newton's method, is sensitive to the initial condition. A poor initial guess may lead to erroneous results including convergence to a curve that is not the optimal path or even no convergence. It is worth reiterating that path dynamics that include a fast, sharp transition may require the use of a non-uniform step size h, so that the transitional region has a finer mesh than areas away from the fast transition. Additionally, the choice of T_ϵ can affect the results. We discuss its importance in Secs. IV and V.

III. 1D AND 2D SIS MODELS

A. Introduction

The Susceptible-Infected-Susceptible (SIS) model describes a disease without immunity from re-infection. The population is composed of two compartments: Susceptibles S and Infectives I. An individual is born susceptible. Then, through contact, an individual may become ill and be classified as infectious. After a specified period, the individual has recovered and is returned to the susceptible compartment. Removal by death is possible from both compartments, but we assume no disease-related deaths in this model.

In the SIS compartmental model shown in Fig. 3, we assume a constant total population so that $N=S+I$. Rates are defined as follows: μ is a birth/death rate, β is the mass action contact rate, and γ is the recovery rate. The events affecting each compartment are used to formulate two mean-field equations that describe the system's behavior over time

$$dS/dt=\mu N-(\beta SI)/N+\gamma I-\mu S, \tag{16a}$$

$$dI/dt=(\beta SI)/N-\gamma I-\mu I. \tag{16b}$$

The mean-field system has two steady states given as

- $S=N, I=0$, the extinct state where no infection exists.
- $S=\frac{N(\mu+\gamma)}{\beta}, I=N(1-\frac{\mu+\gamma}{\beta})$, the endemic state of persistent infection.

As previously mentioned, the stability of the steady states depends upon the value of R_0. For this model,

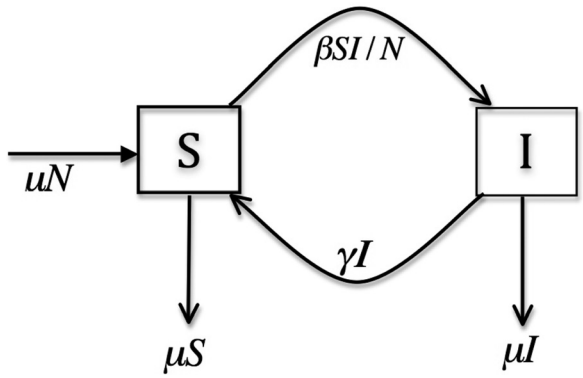

FIG. 3. The SIS model compartmental flow includes birth μN, infection $\beta SI/N$, recovery γI, death of susceptible individuals μS, and death of infected individuals μI.

$R_0 = \beta/(\mu + \gamma)$, so the endemic steady state can be rewritten in terms of R_0 as $S = N/R_0$, $I = N(1 - (1/R_0))$. When $R_0 > 1$, the endemic state is attracting and the extinct state is repelling.

B. 1D and 2D SIS stochastic models

Clearly, when $R_0 > 1$, a population at the endemic state can never go extinct in the deterministic SIS model formulated in Sec. III A. To understand how extinction can occur, we must include stochastic effects due to the random interaction of individuals in the population. We do this using the master equation and WKB formalism as described in Sec. II A. Using the state variable $\mathbf{X} = (S, I)$ to represent the two compartmental model, a list of transitions for all possible events is formulated, as shown in Table I. One can write down the master equation for this example using the transitions from Table I along with the general form of the master equation given by Eq. (1).

Scaling by the constant population size N produces new variables $s = S/N$ and $i = I/N$, so that $s + i = 1$. Then, the Hamiltonian arising from the WKB method for the 2D SIS model is

$$\begin{aligned} H(s, i, p_s, p_i) = {} & \mu(e^{p_s} - 1) + \beta si(e^{-p_s + p_i} - 1) \\ & + \gamma i(e^{p_s - p_i} - 1) + \mu s(e^{-p_s} - 1) \\ & + \mu i(e^{-p_i} - 1). \end{aligned} \tag{17}$$

If one assumes that births and deaths are negligible, then the total number N of individuals in the two population model is constant. As a result, the model effectively becomes a single infectious population with transitions and rates shown in Table II. This is known as the constrained SIS model, and a more detailed analysis can be found in Ref. 34.

TABLE I. 2D SIS transitions and rates.

Event	S Transitions	I Transitions	Scaled rate
Birth	$W_{+1} = \mu N$		μ
Infection	$W_{-1} = \beta SI/N$	$W_{+1} = \beta SI/N$	βsi
Recovery	$W_{+1} = \gamma I$	$W_{-1} = \gamma I$	γi
Death	$W_{-1} = \mu S$	$W_{-1} = \mu I$	μs, μi

As with the 2D SIS example, a master equation for the 1D SIS example can be formulated using Eq. (1) and the transitions in Table II. After scaling by N, and restating s as $1 - i$, the resulting Hamiltonian for the 1D SIS model is

$$H(i, p) = \beta(1 - i)i(e^p - 1) + \kappa i(e^{-p} - 1), \tag{18}$$

where we have omitted the p_i subscript, since there is only one momentum variable for a single population model.

IV. 1D SIS RESULTS

We begin by finding the analytical zero energy solutions for the 1D SIS Hamiltonian given by Eq. (18). The solution $i = 0$ represents extinction; a second solution is $p = 0$, which corresponds to the deterministic dynamics. The third solution is the optimal path and is given by

$$p = \ln\left(\frac{\kappa}{\beta(1 - i)}\right). \tag{19}$$

Next, we find the analytical critical points of Hamilton's equations

$$\dot{i} = \beta(1 - i)ie^p - \kappa ie^{-p}, \tag{20a}$$

$$\dot{p} = \beta(2i - 1)(e^p - 1) - \kappa(e^{-p} - 1). \tag{20b}$$

The trivial solution $(i, p) = (0, 0)$ is associated with the deterministic extinct state $(S, I) = (N, 0)$. A second critical point $(i, p) = (1 - \kappa/\beta, 0)$ is associated with the endemic state. Note that the deterministic mean-field endemic (S, I) point was an attracting node for $R_0 > 1$, but the endemic (i, p) state found using Hamilton's equations is a saddle point, allowing for a path to escape from the endemic level. A third critical point $(0, \ln(\kappa/\beta))$, called the fluctuational extinction point, represents a new disease-free state with nonzero momentum, distinguishing it from the deterministic extinct point (0, 0). As shown in Fig. 4, all critical points for the SIS model are metastable saddle points.

We will now describe the use of the FTLE and IAMM numerical schemes to compute the optimal path. Our numerical results will be compared with the analytical solution given by Eq. (19). Although we know the analytic solution of the optimal path for the 1D SIS model, our objective is to present a useful methodology when no solution is known. In Sec. V, we will consider the 2D SIS model, where an analytic solution of the optimal path is not available.

A. Finite-time Lyapunov exponents (FTLE)

We compute the FTLE field using Hamilton's equations. There are two important components for producing an accurate FTLE field for any model: (1) a domain that is both

TABLE II. 1D SIS transitions.

Event	I Transitions	Scaled rate
Infection	$W_{-1} = \beta SI/N$	$\beta(1 - i)i$
Recovery	$W_{+1} = \kappa I$	κi

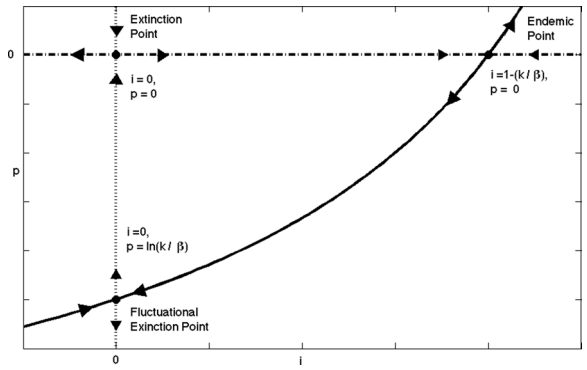

FIG. 4. General version of all solution curves for the 1D SIS model when $R_0 > 1$. The horizontal axis represents the infectious population i, while the vertical axis represents conjugate momentum p. The dashed, horizontal line denotes the deterministic dynamics when $p = 0$. The dotted, vertical line denotes the trivial solution when $i = 0$. The third curve, described by Eq. (19), connects the endemic point to the fluctuational extinction point, where the population is zero but the momentum is non-zero.

sufficiently large and sufficiently meshed must be chosen, and (2) a finite time T must be chosen large enough to capture the dynamics of the system. The critical points of Hamilton's equations are our primary source for boundary information. For the 1D SIS model, we use the critical points of Eqs. (20a) and (20b) to define domain boundaries of at least $0 \le i \le (1 - \kappa/\beta)$ and $\ln(\kappa/\beta) \le p \le 0$. We recommend that the FTLE domain extend a small distance beyond these boundaries to fully capture the system dynamics.

The numerical integration step size is given as δt. One must consider that a point within the domain may be integrated beyond the domain boundaries. We use hard domain boundaries for our FTLE computations. A point integrated outside of the domain is no longer integrated, and the FTLE value last assigned to it remains unchanged. Although some points are advected outside the domain on the initial integration step, using a relatively small δt value minimizes this, with the end result being a more accurate FTLE field.

B. High value FTLE point harvesting

We wish to harvest points along the high value FTLE ridge that forms a pathway from the endemic point to the fluctuational extinction point as shown in Fig. 6. These points will serve as an initial guess of the optimal path for IAMM computations. To harvest these points, we define a threshold that is three standard deviations above the mean of the 1D SIS FTLE data. Additional restrictions limit consideration to only those points occurring between the endemic and fluctuational extinction points. Specifically, for parameters $\beta = 5.0$ and $\kappa = 1.0$, we limit the range of i and p to $.001 < i < 0.8$ and $-1.6 < p < -0.001$. This also eliminates points in close proximity to the deterministic $p = 0$ line and the extinction $i = 0$ line from selection. The selection process yields an array of about 10 000 points that, when plotted, closely approximate the optimal path defined by Eq. (19).

Since the IAMM computation involves the inversion of a matrix with dimensions dependent on the number of array points, it is desirable to reduce the number of points. We evaluate the coordinates of each point in the Hamiltonian given by Eq. (18) and those yielding a value within a tolerance of 5.0×10^{-5} to $\mathcal{H} = 0$ are chosen. This produces the 994 point array shown in Fig. 5(a).

C. IAMM computations

For the 1D SIS model, it seems unnecessary to iterate the 994 point $\mathbf{q}_0$ initial condition using IAMM; the points appear to lie directly on the optimal path (Fig. 5(a)). However, performing the IAMM computation reveals information that is useful when finding the optimal path for higher dimensional models.

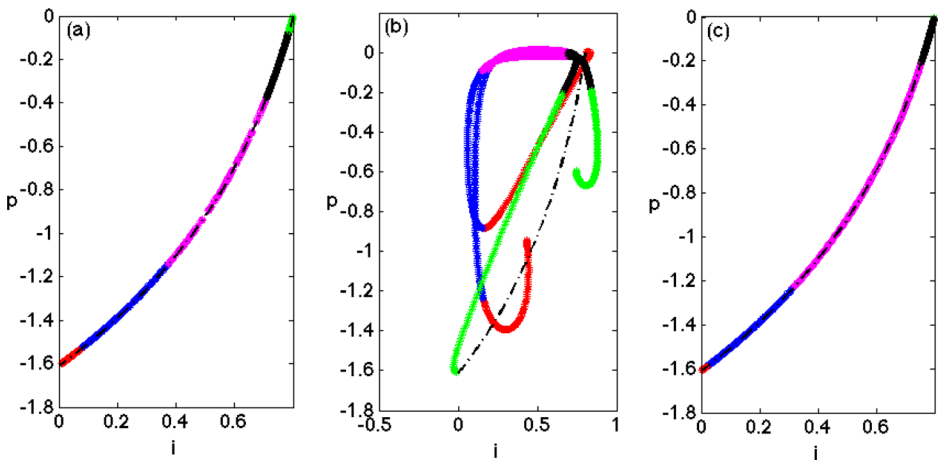

FIG. 5. The success of IAMM processing is dependent upon the ordering of the initial condition, as illustrated by these 1D SIS figures. Panel (a) contains the initial condition $\mathbf{q}_0$ with color-coded segments; each segment holds an equal number of points. The dashed black line underlying the points is the analytic optimal path given by Eq. (19). Note that the FTLE-generated initial condition is very accurate and thus obscures the position of the optimal path. Panels (b) and (c) contain the results after two Newton iterations of the IAMM. Since we have colored the initial condition, one can observe how the ordering of $\mathbf{q}_0$ affects the IAMM outcome. In (b), $\mathbf{q}_0$ was ordered from the fluctuational extinct state to the endemic state, whereas in (c), $\mathbf{q}_0$ was ordered from endemic state to fluctuational extinction. Panel (b) shows the shifting of array points as the IAMM attempts to reposition points far from their original position in order to achieve the desired point ordering from the endemic to the extinct state. The points move so far from the optimal path that, ultimately, the method does not converge to the optimal path. In contrast, (c) reflects minimal repositioning due to IAMM adjustments. All IAMM computations were done using $T_\epsilon = 2$.

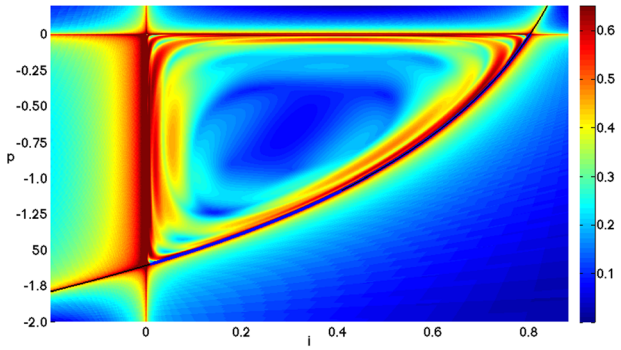

FIG. 6. Average of forward and backward FTLE fields for the 1D SIS model. The analytic solution for the optimal path appears as a black curve. The section between the endemic and fluctuational extinct states is overlaid in blue with the optimal path produced by IAMM computations. The FTLE computations were done using a fourth-order Runge–Kutta scheme with $\delta t = 0.02$, a finite time $T = 7$, and a grid resolution of 0.0005 for both i and p. Parameter values are $\beta = 5.0$ and $\kappa = 1.0$.

In particular, the order of the initial condition is of primary importance. The SIS optimal path flows from the endemic state to the fluctuational extinct state. Figure 5 illustrates the differing outcomes when $\mathbf{q}_0$ is ordered from the fluctuational extinct state to the endemic state (Fig. 5(b)) compared to when $\mathbf{q}_0$ mimics the correct flow (Fig. 5(c)). The extreme adjustments revealed in Fig. 5(b) reflect an attempt to reorder the points. Repositioning of points away from the path is fundamentally disruptive, making divergence more likely.

D. 1D SIS optimal path

Figure 6 demonstrates complete agreement of the FTLE results and the IAMM results with the analytical optimal path given by Eq. (19). Specifically, Fig. 6 shows an average of the forward and backward FTLE fields for the 1D SIS model. The three analytical zero-energy solution curves of the Hamiltonian agree well with the red, maximal FTLE ridges. Both the analytical optimal path and the final IAMM optimal path appear as overlays on the maximal FTLE ridge.

V. 2D SIS RESULTS

We begin by determining whether the Hamiltonian for the 2D SIS model, given by Eq. (17), has any zero-energy analytic solutions. One solution is the deterministic surface where both p_s and p_i are zero, and another is the non-infectious surface where i and p_s are both zero. Since neither of these solutions represents the optimal path, we must find it numerically.

We start by finding the analytical critical points of Hamilton's equations which are given by

$$\dot{s} = \mu e^{p_s} - \beta s i e^{-p_s+p_i} + \gamma i e^{p_s-p_i} - \mu s e^{-p_s}, \tag{21a}$$

$$\dot{i} = \beta s i e^{-p_s+p_i} - \gamma i e^{p_s-p_i} - \mu i e^{-p_i}, \tag{21b}$$

$$\dot{p}_s = -[\beta i(e^{-p_s+p_i} - 1) + \mu(e^{-p_s} - 1)], \tag{21c}$$

$$\dot{p}_i = -[\beta s(e^{-p_s+p_i} - 1) + \gamma(e^{p_s-p_i} - 1) + \mu(e^{-p_i} - 1)]. \tag{21d}$$

The point $(s, i, p_s, p_i) = (1, 0, 0, 0)$ represents the deterministic extinct state. The endemic state is located at $((\gamma+\mu)/\beta, 1 - ((\mu+\gamma)/\beta), 0, 0)$, while the fluctuational extinct state is located at $(1, 0, 0, \ln[(\mu+\gamma)/\beta])$. A stability analysis of these points shows that all three are saddle points. Two additional real critical points are also found. Their possible influence will be discussed later.

A. Finite-time Lyapunov exponents (FTLE)

The critical points of Eqs. (21a)–(21d) define domain boundaries of at least $(\gamma+\mu)/\beta \le s \le 1, 0 \le i \le (1 - [(\mu+\gamma)/\beta)$ and $\ln[(\mu+\gamma)/\beta \le p_i \le 0$. However, p_s is zero at both the endemic and fluctuational extinction steady states, revealing no information about its range. FTLE computations could be used in a "guess-and-check" manner to deduce the range of p_s, but this is a time-consuming choice. A quicker alternative is derived from our knowledge that the optimal path contains points that evaluate $\mathcal{H}$ to be zero. A four-dimensional grid based on the three known domains is defined using a broad guess for the range of p_s. Each grid point is evaluated in the Hamiltonian, and the points that meet a chosen tolerance are retained. A visual inspection provides the p_s range most likely to include points occurring along the optimal path.

For all 2D SIS computations, we use parameter values $\mu = 0.2$, $\beta = 104$, and $\gamma = 100$. Using a tolerance of $\mathcal{H} < 10^{-9}$, a minimum p_s value of -0.04 is determined by visual inspection. Table III contains the dimensions of the domain grid used for the 2D SIS FTLE calculations. Methods to view the resulting four-dimensional array are discussed at the end of this section.

B. High value FTLE point harvesting

Just as for the 1D SIS model, points along the high FTLE ridge must be selected to serve as an initial guess for the IAMM computation. First, an FTLE value threshold must be determined. Since the FTLE data contain over 10^8 values, the use of three standard deviations above the mean to define a threshold yields far too many points. Instead, we use an FTLE threshold value of 4.25, which reduces the total number of points to about 13 000. For this harvesting, we also eliminate all points on the deterministic surface, as well as points beyond the boundaries of the optimal path as described in Table III. Additionally, we eliminate high FTLE points within 0.001 of $i = 0$, since this interval contains a large number of points that provide no significant information.

Keeping in mind the matrix inversion done as part of IAMM processing, the number of points in $\mathbf{q}_0$ was further

TABLE III. 2D SIS domain definition.

	Lower limit	Upper limit	Grid dimension
s	0.954	1.0012	119
i	−0.002	0.038	101
p_s	−0.04	0.002	106
p_i	−0.04	0.0016	105

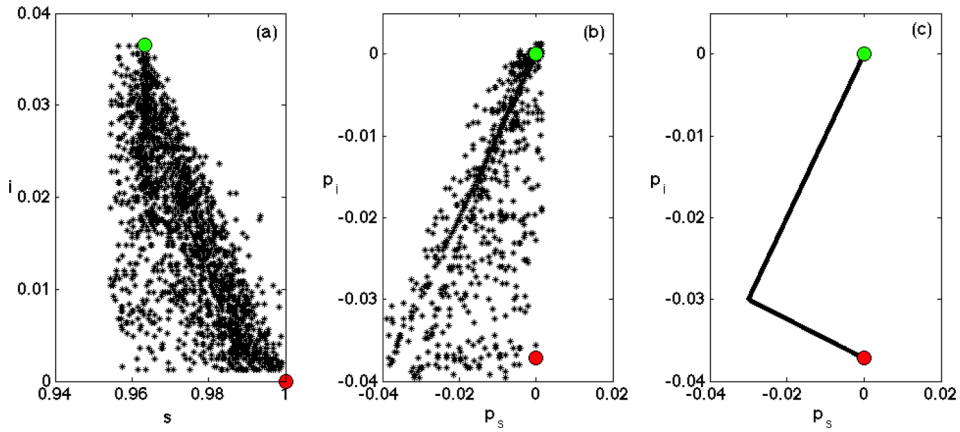

FIG. 7. Panel (a) contains the (s, i) coordinates and panel (b) contains the (p_s, p_i) coordinates of the 1752 points harvested from the FTLE results for use as the initial condition in the IAMM computations for the 2D SIS model. For best results, $\mathbf{q}_0$ should be ordered from the endemic state to the fluctuational extinct state. The harvested points cannot be ordered properly since p_s is not monotonic. We substitute the ordered array plotted in panel (c) for the actual (p_s, p_i) coordinates of panel (b), and pair it with a sorted (s, i) array plotted in panel (a) to obtain an ordered initial guess $\mathbf{q}_0$. The green circle denotes the endemic steady state, while the red circle denotes the fluctuational extinct steady state.

reduced by evaluating each point in the Hamiltonian given by Eq. (17). Those within a tolerance of 2.5×10^{-6} to $\mathcal{H} = 0$ were retained, and these 1752 points are shown in Figs. 7(a) and 7(b).

C. IAMM computations

As mentioned in Sec. IV C, successful IAMM results are more likely when $\mathbf{q}_0$ reflects the flow of the optimal path. For the 2D SIS model, the endemic and fluctuational extinct steady states suggest that s monotonically increases, and i and p_i monotonically decrease as the optimal path traverses the domain. Since $p_s = 0$ at both steady states, its value must decrease, and then increase, perhaps repeatedly. Therefore, the harvested points cannot be sorted into a single monotonic ordering that will fulfill all these conditions simultaneously.

Instead, a hybrid $\mathbf{q}_0$ is created for IAMM processing. For the (s, i) coordinates, the points shown in Fig. 7(a) are ordered with s increasing and i decreasing. For the (p_s, p_i) coordinates, the plot of the harvested points shown in Fig. 7(b) is examined, and an ordered array that approximates areas of dense points as a piecewise linear curve is created, as shown in Fig. 7(c). The section from $p_i = 0$ to $p_i = -0.3$ contains two-thirds of the points, and the remaining one-third populate the lower section. The points of Figs. 7(a) and 7(c) are combined to produce $\mathbf{q}_0$.

After twelve iterations using this initial condition, 90% of the array points meet a tolerance of 1.5×10^{-6} of $\mathcal{H} = 0$. After 50 iterations of the IAMM process, the optimal path shown in Fig. 8 is obtained. All points in this array meet a tolerance of 5×10^{-7}.

This successful result was only achieved after resolving another sensitivity of the IAMM scheme that involves the spacing of points in the initial array. Since the harvested points are not evenly spaced, it is likely that some areas are over-represented while others are under-represented. Our

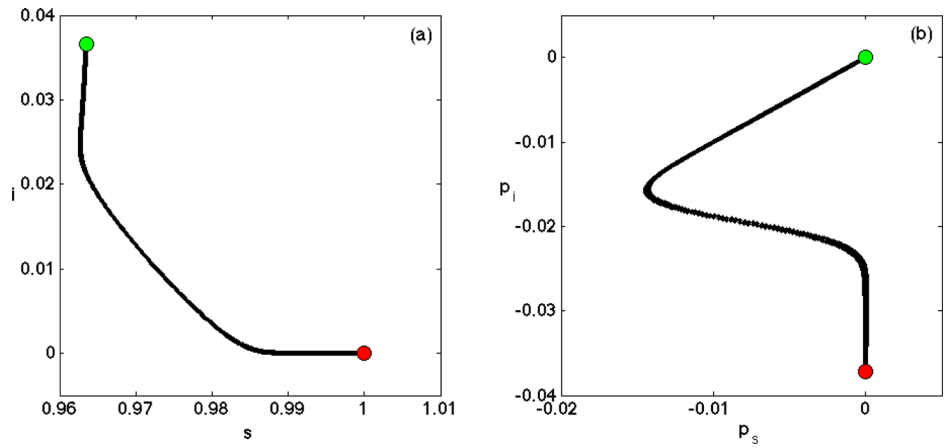

FIG. 8. The 2D SIS results from the IAMM, obtained using $T_\epsilon = 13$ for 50 iterations of the Newton method process. Panel (a) contains the final location of all (s, i) coordinates from Fig. 7(a). Panel (b) contains the final location of all (p_s, p_i) coordinates from Fig. 7(c). Together, they represent the optimal path in four-dimensional space. Although the points appear equally spaced, an examination of the array values shows that is not the case. The central section of the path, between two turns in both views, contains only about 200 of the 1752 point array. The other sections of the path, namely, the exit from the endemic state and the approach to the fluctuational extinct state, are each populated by half of the remaining points. This is due to the fact that very slow dynamics are present near either steady state, while the area of fast dynamics begins and ends with a transition that is, in fact, a change in the direction of the optimal path. The green circle denotes the endemic steady state, while the red circle denotes the fluctuational extinction steady state.

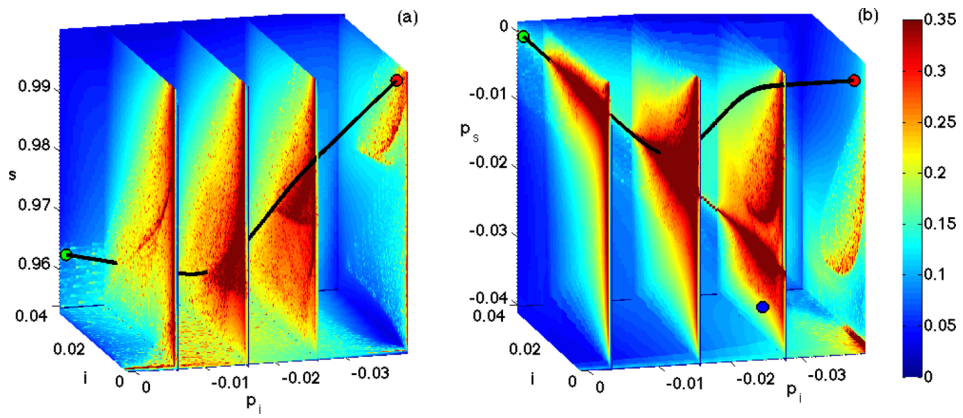

FIG. 9. Two projections of the IAMM and FTLE results for the 2D SIS model. The optimal path obtained from the IAMM computation is represented as a black curve. It begins at the endemic steady state (green circle) on the rear surface and ends at the fluctuational extinct steady state (red circle) in the upper right foreground. Both projections show FTLE values averaged over the missing dimension, so that (a) shows the mean FTLE values of all p_s slices and (b) shows the mean for all s slices. In projection (a), the vertical p_i slices contain areas of high FTLE values in red that closely correlate to the curve of the optimal path. Projection (b), however, exhibits a widening and splitting of high FTLE regions. On vertical slice $p_i = -0.0272$, one observes a path diverging from the optimal path. In the vicinity is a critical point, shown as a blue circle in the foreground of (b). This critical point is not part of the optimal path structure. A stability analysis reveals that the point is neither stable nor a saddle point, but from this projection and many others, it does create dynamics that may compete with those along the optimal path. All FTLE computations were done both forward and backward in time using a fourth-order Runge–Kutta scheme with $\delta t = 0.02$ and a finite time $T = 10$.

initial trials began with a smaller array of about 900 harvested points, and we used the same technique of pairing the actual s, i values with the approximated p_s, p_i values to create the initial array. While the array did converge to the optimal path, the results showed a double line of points in the middle portion of the optimal path. The points in this area were not accurate to a stringent tolerance, and additional iterations did not improve the result.

This issue was resolved by lowering the harvest threshold to increase the number of initial points. In this instance, and in other problems we explored, the inclusion of more points to create a more densely packed initial guess increased the likelihood of converging to the optimal path.

D. 2D SIS optimal path

In order to validate the numerically computed IAMM optimal path, we consider a method for visualizing the results so that a comparison can be made with the FTLE simulations. The obvious choice is to develop a set of three-dimensional projections of the four-dimensional path. In the case of the IAMM result, one can simply omit one dimension of the final array and project the remaining three coordinate array. But in the case of the FTLE result, each four-dimensional point is associated with a unique FTLE value. Therefore, the omission of one dimension will produce multiple points with the same three-dimensional location but different FTLE values.

We therefore choose to average the FTLE values associated with the same three coordinate locations. For example, in Fig. 9(a), the contour plot is displayed on a grid of (i, p_i, s) points. Each location has an associated FTLE value that is the mean of the FTLE values assigned to the 106 points having coordinates i, p_i, and s in common. In other words, we have averaged through the p_s slices.

One flaw in averaging is that it may obscure the minimum and maximum FTLE values. However, a high FTLE ridge in four dimensions is more like a multi-dimensional rope than a single thread. Hence, the high FTLE values persist for many slices in one or more dimensions. Additionally, use of a finely meshed grid and a domain in the vicinity of the optimal path will control the loss of true FTLE information.

Figure 9 shows two projections of the FTLE field overlaid with the IAMM computed optimal path. The optimal path is shown as a black curve that begins on the rear face and flows to the front of both figures. Areas with high FTLE values appear on the vertical slices in dark red. One can see dominant curves on these red areas that mimic the curve of the optimal path. One can also observe that high FTLE areas in the vicinity of the optimal path persist, in varying degrees of intensity, through multiple slices of the domain.

The projection in Fig. 9(b) reveals an interesting aspect of the 2D SIS dynamics. The view, with p_s on the z axis, shows the optimal path fall and then rise. This change of direction occurs near $p_i = -0.015$, and the vertical slice in that location shows a widening of the high FTLE ridge. The vertical slice at $p_i = -0.027$ shows a split in the high FTLE ridge, with an upper part that curves with the optimal path and a lower part that curves downward away from the optimal path. As previously mentioned, there are two additional real-valued critical points. In Fig. 9(b), we include one of them as a blue circle positioned in the foreground at approximately $i = -0.008$, $p_s = -0.03$, $p_i = -0.02$. Its presence has led to the harvesting of many high FTLE points that "overshoot" the turn in the optimal path, such as those shown in Fig. 7(b) forming a diagonal line that extends well beyond $p_s = -0.015$. These same footprints then show up in the Fig. 9(b) projection and other projections not included in this article. Although the FTLE provides an initial guess with seemingly multiple paths, one must remember that one of these "paths" does not connect to an extinct point and is not a zero-energy solution of the Hamiltonian. Moreover, the IAMM uses the initial guess to converge to the correct optimal path.

VI. ALLEE POPULATION MODEL

The Allee model we present is a single population model. We choose to include it because it possesses an optimal path that is particularly difficult to find numerically. Just as with the SIS examples, the Allee optimal path diverges from the deterministic path at a steady state that has become a saddle point through the Hamiltonian formulation. The optimal path for the SIS model then ends at a new steady state with non-zero momentum. The optimal path for the Allee model, however, travels through a region of non-zero momentum and then rejoins the deterministic path. From there, the optimal path follows deterministic dynamics to extinction. This results in a dynamical system with two paths that traverse the region between the same two points in opposite directions.[21] Most numerical methods that have been used over the years to compute the optimal path fail in this instance. We show, however, that our numerical method successfully captures the optimal path.

A. Introduction

The Allee model was formulated as a result of the research of Warder Allee, who observed species in order to determine what contributed to their well-being.[35] The crux of his model is that population density has a direct effect on the growth of a population. His research showed that there exists an interval of population density in which a species exhibits cooperative behavior, and such behavior results in the growth of the population. He observed that a certain level of crowding was beneficial, and helped with combating external attacks such as increased predation or infection. Additionally, he noted that the resources the species requires, such as a healthy environment and food, must be available in sufficient quantities to support the population and engender cooperative behavior.

In contrast, Allee showed that lower population densities often result in a species with decreasing numbers which he attributed to a lack of cooperative behavior. He also noted that extinction was more likely in species with small populations. When he considered the other extreme, an overly dense population, he observed reduced lifespans and a lack of cooperative behavior, often caused by a competition for a stagnant amount of food, water, and other resources necessary for life. Allee concluded that the benefits of cooperation in a population break down when population density becomes too high or too low.

The dynamical model that describes Allee's conclusions divides the population density into three intervals. The middle interval, where population growth occurs, is delineated by a minimum value called the threshold and a maximum value called the carrying capacity. These deterministic dynamics are shown in Fig. 10 as a phase line. The steady state c_1 represents the threshold, and the steady state c_2 represents the carrying capacity. The steady state c_0 is the extinct state where the population is zero.

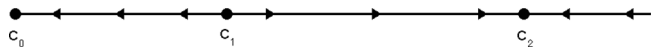

FIG. 10. Three steady states delineate the deterministic dynamics of the Allee model. The attracting point c_0 is the extinct steady state, the repelling point c_1 is the threshold value that separates a growing population from a decaying one, and the attracting point c_2 is the carrying capacity.

According to these dynamics, both c_0 and c_2 are attracting, while c_1 is repelling. Consideration of this scenario leads to the conclusion that population extinction is impossible whenever the population exceeds c_1. However, extinction of species does occur in nature. To capture these rare extinction events mathematically, one must consider a stochastic Allee model, where the noise is due to the random interactions of the individuals.

B. Stochastic Allee model

As mentioned earlier, we use the master equation and WKB formalism to find the Hamiltonian associated with the stochastic Allee problem. The stochastic Allee population model is represented by the transition processes and associated rates found in Table IV. The first two transitions are required to capture the Allee effect. The death rate of a low-density population is given by μ, and the growth rate of the population when the density is large enough is given by λ. The negative growth rate for an overcrowded population is provided by σ, and K is the carrying capacity of the population. In this particular example, all of the transitions are single-step transitions. Therefore, the increment r only takes on the values of ± 1. The scaled transition rates in Eq. 2 are also shown in the final column of Table IV.

Using the scaled transitions, we formulate the following two equations with leading terms of the form $Nw(r)$ as given in Eq. (2):

$$W_1(Kx) = K\left(\frac{\lambda}{2}x^2\right) - \frac{\lambda}{2}x, \quad \text{and}$$
$$W_{-1}(Kx) = K\left(\mu x + \frac{\sigma}{6}x^3\right) - \frac{\sigma}{2}x^2 + \frac{\sigma}{3K}. \tag{22}$$

Then, the effective Hamiltonian (Eq. (4)) for the Allee model is

$$\mathcal{H}(x,p) = \frac{\lambda}{2}x^2(e^p - 1) + \left(\mu x + \frac{\sigma}{6}x^3\right)(e^{-p} - 1). \tag{23}$$

TABLE IV. Allee transitions and rates.

Event	Transition expression	Transition	Scaled transition
Decay	$X \to \varnothing$	$W_{-1}(X) = \mu X$	$K\mu x$
Growth	$2X \to 3X$	$W_1(X) = (\lambda/2K)X(X-1)$	$(\lambda/2)(Kx^2 - x)$
Reduction	$3X \to 2X$	$W_{-1}(X) = (\sigma/6K^2)X(X-1)(X-2)$	$(\sigma/6)(Kx^3 - 3x^2 + \frac{2}{K})$

C. Allee results

Similar to the 1D SIS, the zero-energy solutions are the extinction line $x = 0$, the deterministic line $p = 0$, and the optimal path given by

$$p = \ln\left(\frac{6\mu + \sigma x^2}{3\lambda x}\right). \tag{24}$$

A set of Hamilton's equations for the stochastic Allee model are

$$\dot{x} = \frac{\lambda x^2}{2} e^p - \left(\mu x + \frac{\sigma x^3}{6}\right) e^{-p}, \tag{25a}$$

$$\dot{p} = -\left[\lambda x(e^p - 1) + \left(\mu + \frac{\sigma x^2}{2}\right)(e^{-p} - 1)\right], \tag{25b}$$

which have three steady states located at

$$x_0 = 0, \quad x_{1,2} = \frac{3\lambda \pm \sqrt{9\lambda^2 - 24\mu\sigma}}{2\sigma}. \tag{26}$$

The FTLE results, shown in Fig. 11, illustrate complete agreement with the analytical solution given by Eq. (24). High FTLE ridges appear along all the zero energy curves. Clearly, c_1 and c_2 are connected by two paths. On the deterministic line $p = 0$, the directional flow is from c_1 to c_2, while along the optimal path (in blue), the directional flow is from c_2 to c_1. For this particular example, existing numerical methods will converge to the deterministic piece rather than the optimal path of interest.

Just as with the 1D SIS, an extraction of the points associated with maximal FTLE values yields almost perfect agreement with the optimal path. Any attempt to improve the already excellent result using IAMM seems pointless. Instead, we present evidence of the efficacy of the IAMM process using a scenario that is analogous to a higher dimensional system. In such systems, maximal FTLE-valued points in the vicinity of the optimal path often define a multi-

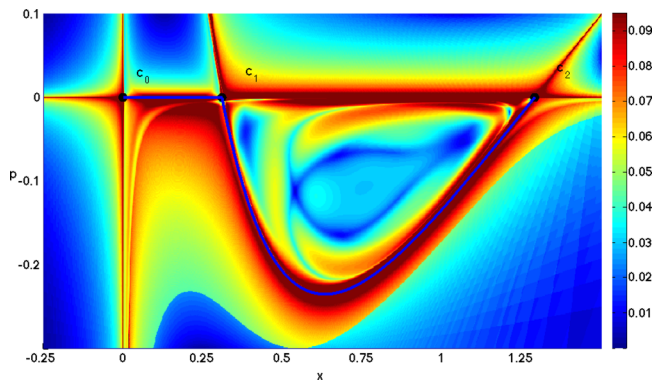

FIG. 11. Average of forward and backward FTLE fields for the Allee model. The optimal path, given by Eq. (24) and denoted by the blue curve overlay, leaves the deterministic line at the carrying capacity c_2 and rejoins the deterministic line at the threshold c_1. The optimal path then essentially runs deterministically to the extinct state c_0. All FTLE computations were done using a fourth-order Runge–Kutta scheme with $\delta t = 0.1$, a finite time $T = 40$, and a grid resolution of 0.002 in x and 0.001 in p. Parameter values are $\lambda = 1.6$, $\mu = 0.2$, $\sigma = 3.0$.

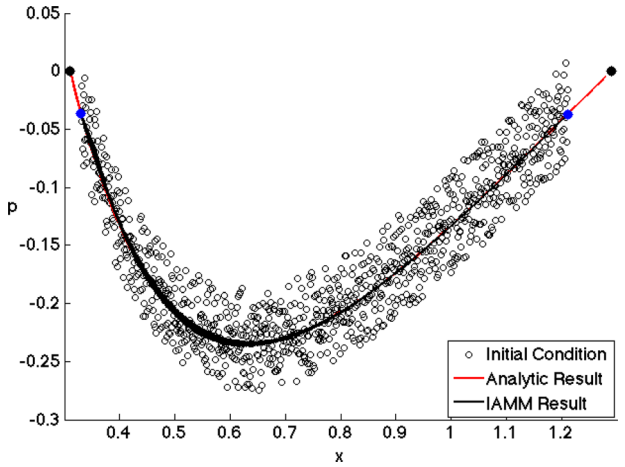

FIG. 12. IAMM results for the Allee model using 1000 points randomized about the analytical optimal path as the initial condition. After three Newton iterations, the point locations, shown as a black line, have markedly converged to the analytic optimal path (in red). The initial array contained nine points within 10^{-5} of $\mathcal{H}(x, p) = 0$, but after IAMM processing 60% of the points meet that tolerance. In order to lessen the effect of deterministic dynamics on IAMM convergence, artificial "endpoints," shown as blue dots, were defined at $x = 0.3101$ and $x = 1.2115$. Consequently, two small portions of the optimal path between these points and the analytic steady states (shown as black dots) are not found. A T_ϵ of 10.2 was used to produce these results.

dimensional region about the path rather than the true path. Therefore, using this two-dimensional system, we create such a region by randomizing points about the optimal path to replicate the location of high-FTLE points. These random points serve as the initial condition for IAMM processing. In order to mute the influence of deterministic dynamics, we do not include points in the initial condition that are in the immediate vicinity of the critical points, given by Eq. (26). This results in a "gap" between the steady state locations and the points chosen as the endpoints of the initial condition array. It is worth noting that the part of the path closest to the critical points (where the gap occurs) tends to be the easiest piece of the optimal path to find. More importantly, our method captures the central part of the path, where other methods are known to fail.

To achieve convergence in this problem, we employed a non-uniform step size as described in Sec. II C. In particular, we used an exponential stretching transformation given by $\tau = 2/(1 + e^{ct})$, where t lies in the interval [0, 1]. For the results shown in Fig. 12, c was set to 5.

VII. 3D SEI MODEL

To demonstrate the power of our method, we consider a higher-dimensional Susceptible-Exposed-Infected-Recovered (SEIR) model. In the mean-field system shown below, one can see that the R-equation is decoupled from the other 3 equations. This allows us to consider the 3D SEI system. Through the stochastic formulation described previously, the optimal path for this system will lie in 6-dimensional (6D) space. As far as the authors know, there are no published results with examples of computing optimal paths in dimensions higher than four.

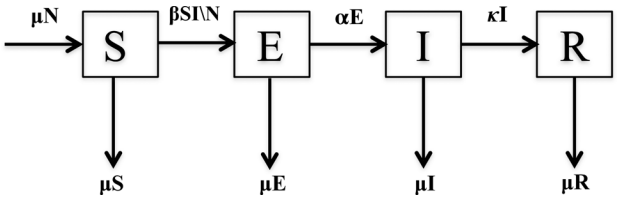

FIG. 13. The SEIR model compartmental flow includes birth μN, infection $\beta SI/N$, latency αE, recovery κI, death of susceptible individuals μS, death of exposed individuals μE, death of infectious individuals μI, and death of recovered individuals μR.

A. Introduction

The SEIR model typically is used to describe childhood diseases where the individual eventually recovers and obtains immunity from the disease. The population is composed of four compartments: Susceptibles S, Exposed E, Infectious I, and Recovered R. An individual is born susceptible. By contact with an infectious individual, a susceptible individual becomes exposed. An exposed individual is infected by the disease but is not yet infectious. After a specified latent period, the exposed individual becomes infectious and is now capable of transmitting the disease to a susceptible individual. After a specified period, the individual recovers and is immune to the disease. Removal by death is possible from all compartments, but we assume no disease-related deaths in this model.

In the SEIR compartmental model shown in Fig. 13, the rates are defined as follows: μ is a birth/death rate, β is the mass action contact rate, α is the exposure rate, and κ is the recovery rate. The events affecting each compartment are used to formulate the following mean-field equations that describe the system's behavior over time:

$$dS/dt = \mu N - (\beta SI)/N - \mu S, \tag{27a}$$

$$dE/dt = (\beta SI)/N - (\alpha + \mu)E, \tag{27b}$$

$$dI/dt = \alpha E - (\kappa + \mu)I, \tag{27c}$$

$$dR/dt = \kappa I - \mu R. \tag{27d}$$

As with the SIS models described in Sec. III, the SEIR model given by Eqs. (27a)–(27d) has analytical expressions for the extinct state and the endemic state.

B. 3D SEI stochastic model

As with the previous models, a population at the endemic state can never go extinct (for $R_0 > 1$) in the deterministic SEIR model formulated in Sec. VII A. To understand how extinction can occur, we must include stochastic effects due to the random interaction of individuals in the population. We do this by using the master equation and WKB formalism as described in Sec. II A.

Assuming that the average population size is N, we can constrain the population size so that $N = S + E + I + R$. Noting that R is overdetermined and decoupled from the rest of the system, we can consider the dynamics of the constrained SEIR model in terms of the S, E, and I individuals.

TABLE V. 3D SEI transitions and rates.

Event	S Transitions	E Transitions	I Transitions	Scaled rate
Birth	$W_{+1} = \mu N$			μ
Exposed	$W_{-1} = \beta SI/N$	$W_{+1} = \beta SI/N$		βsi
Infectious		$W_{-1} = \alpha E$	$W_{+1} = \alpha E$	αe
Recovery			$W_{-1} = \kappa I$	κi
Death	$W_{-1} = \mu S$	$W_{-1} = \mu E$	$W_{-1} = \mu I$	μs, μe, μi

Using the state variable $X = (S, E, I)$, a list of transitions for all possible events is formulated, as shown in Table V.

Scaling by the population size N produces new variables $s = S/N$, $e = E/N$, and $i = I/N$. Then, the Hamiltonian arising from the WKB method for the 3D SEI model is

$$\begin{aligned}\mathcal{H} = {} & \mu(\exp(p_s) - 1) + \beta si(\exp(-p_s + p_e) - 1) \\ & + \alpha e(\exp(-p_e + p_i) - 1) + \kappa i(\exp(-p_i) - 1) \\ & + \mu s(\exp(-p_s) - 1) + \mu e(\exp(-p_e) - 1) \\ & + \mu i(\exp(-p_i) - 1).\end{aligned} \tag{28}$$

C. 3D SEI results

The Hamiltonian given by Eq. (28) has no analytical solution for the optimal path. Therefore, we must find it numerically. As before, the process starts by finding the critical points of Hamilton's equations which are given by

$$\dot{s} = \mu \exp(p_s) - \beta se \exp(-p_s + p_e) - \mu s \exp(-p_s), \tag{29a}$$

$$\dot{e} = \beta si \exp(-p_s + p_e) - \alpha e \exp(-p_e + p_i) - \mu e \exp(-p_e), \tag{29b}$$

$$\dot{i} = \alpha e \exp(-p_e + p_i) - \kappa i \exp(-p_i) - \mu i \exp(-p_i), \tag{29c}$$

$$\dot{p}_s = -[\beta i(\exp(-p_s + p_e) - 1) + \mu(\exp(-p_s) - 1)], \tag{29d}$$

$$\dot{p}_e = -[\alpha(\exp(-p_e + p_i) - 1) + \mu(\exp(-p_e) - 1)], \tag{29e}$$

$$\begin{aligned}\dot{p}_i = {} & -[\beta s(\exp(-p_s + p_e) - 1) + \kappa(\exp(-p_i) - 1) \\ & + \mu(\exp(-p_i) - 1)].\end{aligned} \tag{29f}$$

The analytical endemic steady state is given as $p_s = p_e = p_i = 0$,

$$\begin{aligned} s &= \frac{(\alpha + \mu)(\kappa + \mu)}{\alpha\beta}, \\ e &= -\frac{\mu(\alpha + \mu)(\kappa + \mu) - \alpha\beta}{\alpha\beta(\alpha + \mu)}, \\ i &= -\frac{\mu(\alpha + \mu)(\kappa + \mu) - \alpha\beta}{\beta(\alpha + \mu)(\kappa + \mu)}, \end{aligned} \tag{30}$$

while the analytical fluctuational extinct state is given as $s = 1$, $e = i = p_s = 0$,

$$\begin{aligned} p_e &= \ln\left(\frac{\beta\mu + (\alpha + \mu)(\kappa + \mu)}{\beta(\alpha + \mu)}\right), \\ p_i &= \ln\left(\frac{(\alpha + \mu)(\kappa + \mu)}{\alpha\beta}\right). \end{aligned} \tag{31}$$

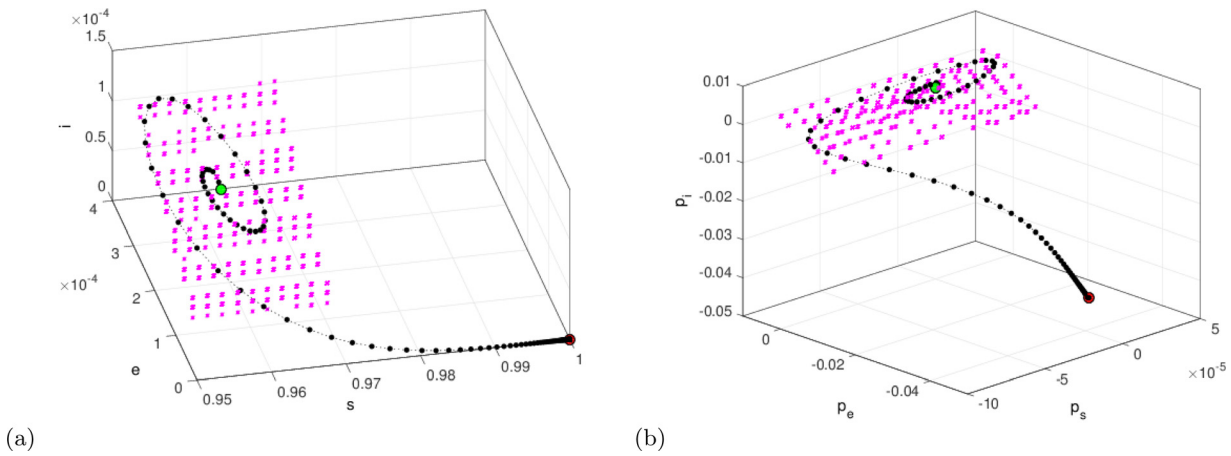

FIG. 14. The 3D SEI results using the FTLE initial conditions for the IAMM. Panel (a) contains the final location of all (s, e, i) coordinates. Panel (b) contains the final location of all (p_s, p_e, p_i) coordinates. Together, they represent the optimal path in six-dimensional space. The magenta stars represent the 300 closest initial conditions identified by the FTLE using the normalized Euclidean distance. The green circle denotes the endemic steady state, while the red circle denotes the fluctuational extinct steady state. Because of the spiral nature of the endemic steady state, a vector was constructed using the 300 FTLE points and 300 equidistant points connecting the endemic and the fluctuational extinct steady states linearly. A small perturbation of 0.001 was added to the p_s coordinate to move it away from the $p_s = 0$ hyperplane (to avoid the same sort of issue that can be found in the Allee problem and that was discussed in Sec. VI C). This vector of 600 points was used in the IAMM as the initial condition in a continuation method that increased the value by one from $T_\epsilon = 13.5$ to $T_\epsilon = 158.5$ for 20 iterations each. The final result has a maximum error for the Hamiltonian of 5.347×10^{-10}. Parameter values are $\alpha = 35$, $\beta = 105$, $\kappa = 100$, and $\mu = 0.2$.

The FTLE field was computed using Hamilton's equations (Eqs. (29a)–(29f)), and the high FTLE values were extracted based on Euclidean distance from the endemic state to use as an initial condition for the IAMM method. Spiral optimal paths are difficult to obtain by the IAMM or any method. In six dimensions, the FTLE identifies points in a small region of the endemic state sufficiently close to the optimal path for a good initial condition guess. Because this problem has initial and final points on the same $p_s = 0$ hyperplane, a small perturbation of 0.001 was added to the p_s coordinate to move it away from that invariant set (to avoid the same sort of issue that can be found in the Allee problem and that was discussed in Sec. VI C). The IAMM will converge to a part of the optimal path for small T_ϵ. Once we obtain a case of IAMM convergence, we can increase T_ϵ by continuation to stretch it along the path and fill out the spiral. Results are shown in Fig. 14.

VIII. DISCUSSION

In this article, we have developed a numerical methodology that allows one to compute the optimal path of escape from a metastable state or the optimal path of switching from one metastable state to another metastable state in a general stochastic dynamical system, whether the noise be external or internal. There are many physical and biological applications in which knowledge of the optimal path is of great importance to fully understand the system's dynamics. Since the optimal path typically has no analytical form, we must be able to numerically compute this path.

As described in the introduction, we transform the original stochastic problem into a new deterministic system described by a Hamiltonian that has dimensions twice that of the original system. The optimal path is one of the zero-energy solutions to this Hamiltonian. Hamilton's equations are found straightforwardly by differentiating the Hamiltonian, and these equations are used to compute the finite-time Lyapunov exponents. In previous work, we showed that the optimal path corresponds to the maximal ridge of FTLE, and used this to find the optimal path for several low-dimensional systems. While the theory holds for high-dimensional systems, the FTLE ridges are not so clear-cut, and therefore, it is difficult to extract the optimal path from FTLE in high-dimensional systems.

This article contains the detail required to use FTLE for the much more complicated higher-dimensional models. This work, taken by itself, illustrates a way to visualize the optimal path in higher dimensional models where typical methods, including shooting methods, are known to break down. Additionally, the visualization of the system's dynamics leads to greater understanding of the model under consideration.

Furthermore, we have developed a combined numerical scheme, wherein the FTLE results are used as an initial condition for an iterative action minimizing method. As mentioned in Ref. 26, one of the primary limitations of the IAMM is "…the finesse required in picking an initial guess that guarantees convergence…". Instead of making ad-hoc initial guesses that will often fail to converge, our use of FTLE to provide an initial guess is a constructive tool that enables the IAMM to converge to the optimal path.

Our scheme has been demonstrated using multiple types of stochastic epidemic models and a population model. However, the procedure is general and may be used to find the optimal escape or switching path for any stochastic dynamical system where noise-induced escape from a metastable state or noise-induced switching between metastable states occurs.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Science Foundation. M.B., E.F., and L.B. were supported by National Science Foundation Award No. CMMI-1233397. M.B. was additionally supported by the Margaret and Herman Sokol Graduate Summer Research Fellowship. This

material is based upon work while L.B. was serving at the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[1]M. I. Dykman, "Large fluctuations and fluctuational transitions in systems driven by coloured Gaussian noise: A high-frequency noise," Phys. Rev. A **42**, 2020 (1990).
[2]M. I. Dykman, P. V. E. McClintock, V. N. Smelyanski, N. D. Stein, and N. G. Stocks, "Optimal paths and the prehistory problem for large fluctuations in noise-driven systems," Phys. Rev. Lett. **68**, 2718 (1992).
[3]*Fluctuations and Order: The New Synthesis*, edited by M. Millonas (Springer-Verlag, 1996).
[4]D. G. Luchinsky, P. V. E. McClintock, and M. I. Dykman, "Analogue studies of nonlinear systems," Rep. Prog. Phys. **61**, 889 (1998).
[5]E. Forgoston and I. B. Schwartz, "Escape rates in a stochastic environment with multiple scales," SIAM J. Appl. Dyn. Syst. **8**, 1190 (2009).
[6]R. V. Kohn, M. G. Reznikoff, and E. Vanden-Eijnden, "Magnetic elements at finite temperature and large deviation theory," J. Nonlinear Sci. **15**, 223 (2005).
[7]T. A. Fulton and L. N. Dunkelberger, "Lifetime of the zero-voltage state in Josephson tunnel junctions," Phys. Rev. B **9**, 4760 (1974).
[8]L. S. Tsimring, "Noise in biology," Rep. Prog. Phys. **77**, 026601 (2014).
[9]M. Assaf, E. Roberts, and Z. Luthey-Schulten, "Determining the stability of genetic switches: Explicitly accounting for mRNA noise," Phys. Rev. Lett. **106**, 248102 (2011).
[10]O. Ovaskainen and B. Meerson, "Stochastic models of population extinction," Trends Ecol. Evol. **25**, 643 (2010).
[11]C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer-Verlag, 2004).
[12]N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, 2007).
[13]E. Roberts, S. Be'er, C. Bohrer, R. Sharma, and M. Assaf, "Dynamics of simple gene-network motifs subject to extrinsic fluctuations," Phys. Rev. E **92**, 062717 (2015).
[14]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, Inc., 1965).
[15]R. Kubo, K. Matsuo, and K. Kitahara, "Fluctuation and relaxation of macrovariables," J. Stat. Phys. **9**, 51 (1973).
[16]H. Gang, "Stationary solution of master equations in the large-system-size limit," Phys. Rev. A **36**, 5782 (1987).
[17]M. I. Dykman, E. Mori, J. Ross, and P. M. Hunt, "Large fluctuations and optimal paths in chemical-kinetics," J. Chem. Phys. **100**, 5735 (1994).
[18]V. Elgart and A. Kamenev, "Rare event statistics in reaction-diffusion systems," Phys. Rev. E **70**, 041106 (2004).
[19]D. A. Kessler and N. M. Shnerb, "Extinction rates for fluctuation-induced metastabilities: A real space WKB approach," J. Stat. Phys. **127**, 861 (2007).
[20]M. Assaf and B. Meerson, "Extinction of metastable stochastic populations," Phys. Rev. E **81**, 021116 (2010).
[21]G. Nieddu, L. Billings, and E. Forgoston, "Analysis and control of preextinction dynamics in stochastic populations," B. Math. Biol. **76**, 3122 (2014).
[22]D. M. Roma, R. A. O'Flanagan, A. E. Ruckenstein, A. M. Sengupta, and R. Mukhopadhyay, "Optimal path to epigenetic switching," Phys. Rev. E **71**, 011902 (2005).
[23]A. Kamenev and B. Meerson, "Extinction of an infectious disease: A large fluctuation in a nonequilibrium system," Phys. Rev. E **77**, 061107 (2008).
[24]I. B. Schwartz, E. Forgoston, S. Bianco, and L. B. Shaw, "Converging towards the optimal path to extinction," J. R. Soc. Interface **8**, 1699 (2011).
[25]E. Forgoston, S. Bianco, L. B. Shaw, and I. B. Schwartz, "Maximal sensitive dependence and the optimal path to epidemic extinction," B. Math. Biol. **73**, 495 (2011).
[26]B. S. Lindley and I. B. Schwartz, "An iterative action minimizing method for computing optimal paths in stochastic dynamical systems," Physica D **255**, 22 (2013).
[27]W. E, W. Ren, and E. Vanden-Eijnden, "Minimum action method for the study of rare events," Commun. Pure Appl. Math. **57**, 637 (2004).
[28]M. Heymann and E. Vanden-Eijnden, "The geometric minimum action method: A least action principle on the space of curves," Commun. Pure Appl. Math. **61**, 1052 (2008).
[29]A. D. Wentzel', "Rough limit theorems on large deviations for Markov stochastic processes, I," Theor. Probab. Appl. **21**, 227 (1976).
[30]G. Haller, "Finding finite-time invariant manifolds in two-dimensional velocity fields," Chaos **10**, 99 (2000).
[31]G. Haller, "Distinguished material surfaces and coherent structures in three-dimensional fluid flows," Physica D **149**, 248 (2001).
[32]G. Haller, "Lagrangian coherent structures from approximate velocity data," Phys. Fluids **14**, 1851 (2002).
[33]S. C. Shadden, F. Lekien, and J. E. Marsden, "Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows," Physica D **212**, 271 (2005).
[34]L. Billings, L. Mier-y-Teran-Romero, B. S. Lindley, and I. B. Schwartz, "Intervention-based stochastic disease eradication," PLoS One **8**, e70211 (2013).
[35]W. C. Allee, *Animal Aggregations, A Study in General Sociology* (University of Chicago Press, 1931).